The Extent of Internet Auction Markets:

An empirical study of the $2\log n$ rule on eBay

Oleg Tschernizki

Erasmus University Rotterdam

June 12, 2010

Abstract

Internet auctions attract many potential buyers, but not many of them tend to become active bidders. In *The Extent of Internet Auction Markets*, de Haan, de Vries, and Zhou (2007) develop a model accounting for this fact, with the $2\log n$ rule as its central prediction. The rule states that if the number n of potential bidders is large, the number of active bidders is approximately $2\log n$. Using a new dataset, this paper tests the rule's applicability to hard-close eBay auctions, and examines a few of the underlying conditions and assumptions more closely. For eBay, the rule is rejected in its most simple form; however, some results suggest that the model may in fact hold as long as the number of active bidders is kept sufficiently high, or when only auctions with nonzero starting prices are considered.

1. Introduction

Auctions, since ancient times, have been used to sell off the widest possible range of goods and services: anything from tulip bulbs and jewellery to radio frequency band permits and actual empires, lock, stock, and barrel[1]. The benefits offered by auctions to sellers are obvious: by attracting a multitude of potential buyers and having them bid against each other, one is more likely to find the highest valuation buyer, and extract the highest possible price. This *price discovery* advantage comes, or used to come, at a cost: the *transaction costs* involved in auctioning off an item can be prohibitive. Use of auctions was therefore reserved for those instances in which the need for price discovery (and the benefit derived from it) was high, i.e. when the seller was very uncertain about the actual value of the item in question.

The advent of online auctions has changed that, dramatically lowering transaction costs, while expanding the pool of potential buyers and thereby increasing the potential benefit of price discovery. Everyone with access to a computer and the internet can be a buyer or seller at (almost) no cost. A buyer can browse available auctions from the comfort of his home and at his leisure, and the seller gains access to hundreds of millions of potential buyers. The upshot of internet auctions is not one of convenience alone; Bulow and Klemperer (1996) showed that as long as the bidders' valuations of the item are independent and private (i.e. the IPVP, the Independent Private Values Paradigm, holds), enlarging the market makes the seller better off. Lusht (1996) tested this empirically in the market for middle- to high-priced houses, and came to much the same conclusion. Buyers are better off for similar reasons. Auctions, thus unchained from the constraints of auction houses and moved into a virtual domain, have become tremendously popular. At present, millions of items are traded on sites like eBay daily – ranging from the most mundane, to things like decommissioned aircraft carriers and ICBM bunkers.

Buyers and sellers aren't the only parties to benefit from the development. Online auctions, with their vast amounts of electronically recorded data, are fertile ground for research. This veritable cornucopia of data allows answering questions which have heretofore been unanswerable, and researchers have made eager use of these new possibilities. Unfortunately, though online auction sites like eBay have access to data on every aspect of their auctions, a lot of information is hidden from the public. The main contribution of this thesis is the use of a new data set with extra information on the bidders' behaviour.

[1] The Emperor's seat to the Roman Empire was auctioned off in 193 A.D. and bought by a fellow named Didius Julianus. He managed to enjoy his reign for a full two months before getting ousted and subsequently executed. One might be inclined to think he overpaid; the winner's curse, indeed.

An interesting aspect of online auctions is the fact that although the number of potential bidders is large, the number of actual or *active* bidders is generally rather small. For example, Ockenfels and Roth (2006) find in their study of 120 Amazon and 120 eBay laptop auctions that the average number of active bidders is 4.96 and 6.17, respectively. Similarly, a study of online auctions in Korea by Park and Bradlow (2005) finds an average of 5.8 bidders per auction. A study of eBay coin auctions by Bajari and Hortaçsu (2003) finds an average of 3.08 active bidders, with a standard deviation of 2.51 and a maximum of 14 active bidders.

To explain this phenomenon, de Haan et al. (2008) have proposed their so-called $2 \log n$ rule, which states that under the IPVP, and given the number of *potential* bidders n, the number of *active* bidders will be approximately equal to $2 \log n$. They confirmed this rule empirically using laptop auctions from a smaller Dutch auction site, taking the number of page views on a particular auction page as a proxy for the number of potential bidders for that item. Unfortunately, they were unable to do so for one of the more popular sites like Amazon (now defunct) or eBay, due to unavailability of data (i.e. page view counts). A second, indirect test performed on Yahoo! Auctions provided some supplementary evidence, but its validity relies on an additional, possibly onerous assumption.

The current paper offers an analysis o a more complete dataset collected from eBay auctions, which allows testing whether the $2 \log n$ hypothesis holds for eBay auctions, where strategic bidding is far more prevalent. Moreover, a record of the arrivals of new visitors over time makes it possible to verify the Poisson arrival process assumption, as well.

Though the new dataset is more comprehensive, it is not perfect either. eBay auctions do not align perfectly with the theoretical conditions used to derive the $2 \log n$ rule, in that they use fixed ending times and thus offer an incentive for *strategic bidding*. Auctions with fixed end times finish at the predesignated time, regardless of any bids done in the final moments. By contrast, auctions with a "soft close" automatically extend the end time whenever a last-minute bid is placed. When the ending time is known and fixed, it could make sense for a buyer to hold off on his bid, and try to place a bid that is lower than his valuation, as close as possible to the end of the auction. Due to the fixed ending, it is possible that no-one else has time to react, and the bidder wins the auction at a lower bid than he would have otherwise (or perhaps not at all). This strategy is known as *sniping*, and it can have an effect on the bid sequence vis-à-vis soft-close auctions. De Haan et al. (2008) derive and test the $2 \log n$ rule for soft-close auctions only. This paper offers some discussion as to the applicability of the rule to hard-close auctions, as well as the requisite empirical support.

The aims of the current paper are two-fold: first, to examine the applicability of the $2 \log n$ rule to eBay (fixed end) auctions, and to test this empirically; second, to (where possible) verify previous evidence, hopefully adding to its robustness, and to test one of the underlying assumptions (that of the Poisson arrival process). The paper is structured as follows. Section 2 examines the main characteristics of online auctions. Section 3 offers a simplified account of the derivation of the $2 \log n$ rule, as well as the existing empirical evidence for it. Section 4 discusses the data gathering process. Section 5 presents the in-depth analysis of the data and discusses the findings. Section 6 concludes.

2. The (Online) Auction

After an introduction to standard auction models, this section describes several characteristics specific to online auctions: termination rules and sniping, the bidding system, the reminder system, and the Buy-it-now option. Because the empirical research in this paper focuses on eBay auctions, this section will deal with eBay auction modalities predominantly.

2.1 Standard Auction Types, Bidder Behaviour, and Outcomes

Auctions are generally classified based on three distinct criteria; they are characterised by whether the bid sequence is ascending or descending, whether bids are sealed or open, and in the case of sealed bids, whether the first or second bid defines the price. For a more in-depth look at the various auction models as well as their mathematical underpinnings, see for example Menezes and Monteiro (2008).

The most well known by far are the open *ascending price* auctions (also known as English auctions). In English auctions, the price is raised until there is a single bidder left, and the winner pays the price at which his strongest competitor drops out. There are numerous ways in which the auctioneer can organise the price raising mechanism; bidders may raise the price themselves by announcing a new bid, but this could also be done by the seller, or continuously and automatically, in the form of a "price clock". In the first case, bidders actively raise the price until no-one is willing to do so, and the winner takes the item for the price of his last bid. In case of a price clock, bidders make the decision to drop out of the auction and make this known; this decision can be observed by other bidders, and the winner pays the price at which the last remaining bidder (other than himself) drops out. Both types may be employed when the seller announces prices. Almost all online auctions use the English auction model, with bidders announcing their bids. This can be easily explained by the convenience afforded by the mechanism; online auctions tend to run for days, and bidders may trickle in at any given moment. Anything else would be unacceptable (to buyers, sellers, or both), except perhaps in a few isolated cases.

Dutch auctions, named after their most well-known example (Dutch flower auctions) are open, *descending price* auctions. In this case, the price starts at a certain high value and decreases automatically and at a constant rate. Here, the first person to make a bid wins the auction, at the then prevailing price. This may have an important consequence for the buyer. In English auctions, the winner's price is determined by his last remaining competitor; in the Dutch variety, the winning price is determined entirely by the winner's maximum willingness to pay.

The two remaining standard models are the *first-* and *second-price sealed bid* auctions. In sealed bid auctions, bidders do not openly make their bids, and therefore cannot see what their competitors are bidding, either. In first-price sealed bid auctions, the highest bidder wins and pays a price equal to his bid; in second-price sealed bid auctions (also known as Vickrey auctions) on the other hand, the top bidder pays the bid of his closest competitor, i.e. the second-highest bid. Sealed-bid auctions are not generally used online, though perhaps classified ads on sites such as craigslist.com (which too has become extremely popular over the years) that invite potential buyers to send in bids can be seen as a form of sealed-bid auctions. Similarly, some smaller auction sites such as marktplaats.nl allow buyers to contact sellers with an offer directly; this too is akin to introducing a sealed-bid element to an otherwise standard English auction.

In addition to the four standard auction types, there are two models describing how bidders value an item: the *private-value* and the *common-value* models. In the former, bidders know their own, private valuation of the item (i.e. their maximum willingness to pay), but not that of the other bidders. In case of *independent* values (i.e. when the bidders' values are drawn independently from a particular distribution), we speak of the *independent private values* model, the benchmark model in auction theory. A consequence of the private nature of bidders' valuations is the fact that their valuations may differ; this, in contrast with the *common value* model, in which the value of the item is the same for each bidder (i.e. its "true value"), but different bidders may have different information regarding said value. In this model, each bid may be a signal to the other bidders of the item's actual worth. A bid by any one bidder may signal to the others that this particular bidder may possess information which leads him to value the item highly (or not), and they will modify their own valuations accordingly. Canonical examples include auctions for exploration rights of oil fields (or other natural resources) or auctions for antiques. Since bids depend on the information possessed by the bidders, common value auctions may give rise to a phenomenon known as the *winner's curse*; the fact that a bidder wins the auction may in itself be an indication that he has overpaid (because others had information which led them to lower bids).

Under the IPVP (Independent Private Value Paradigm), auction outcomes are relatively easy to predict. Let us assume risk-neutral bidders (an assumption which leads to the *symmetric independent private values model*). Take the ascending-price auction: it is easy to see that it is a dominant strategy for a bidder to stay in the auction until bids reach his valuation (i.e. maximum willingness to pay), and to drop out at that point. There is no strategy which will yield a higher payoff, regardless of the competitors' strategies; dropping out earlier would mean losing the auction, while dropping out later leads to a loss. The bidder with the maximum value will win the auction, and pay a price that is equal to the valuation of his closest competitor (who will drop out of the auction at that price) or, where bidding occurs in steps, a price that is one minimum increment above that. Bidders in second-price sealed-bid auctions will follow a similar logic; the dominant strategy being to bid one's value. Here too, as long as he bids his true valuation, the bidder with the maximum willingness to pay will win the auction, and because the price is determined by the *second* highest bid, he will only pay his closest competitor's bid.

The situation is markedly different when looking at Dutch and first-price sealed-bid auctions. Bidding one's valuation will lead to a payoff of zero (although one should assume that payoffs being equal, a bidder prefers winning an auction over losing it), yet it is the only way to ensure winning the auction if one's valuation is the highest. To receive a positive payoff, a bidder would have to bid less than his actual value, a practice known as *bid shading*. Bid shading comes at a cost, however; a chance at a positive payoff is traded for a risk of not winning the auction at all. Of course, actually knowing the other bidders' valuations simplifies matters considerably; the dominant strategy in this case (assuming one's valuation is the highest) is to simply bid the closest competitor's maximum willingness to pay (or the minimum increment above that). Therefore, when bidders' valuations are known, all four auction types are strategically equivalent and lead to similar outcomes. When values are hidden however, bidders face a trade-off between risk and return; their bids (and the degree of bid shading) will in that case depend on their beliefs about others' values, as well as their risk tolerance. It can be shown that under the previously made assumptions of risk neutrality and the IPVP, the equilibrium strategy for the bidder with the highest valuation is to bid his expectation of the value of his closest competitor. This means that the expected revenue of first-price sealed-bid and Dutch auctions is equal to that of second-price sealed-bid and English ones, a conclusion known as the *Revenue Equivalence Theorem*, first derived by Vickrey (1961).

The common values model complicates matters, since bidders gain an additional incentive to shade their bids, i.e. to escape the winner's curse. In common values auctions, having the highest valuation is a double-edged sword; it is necessary to win the auction in the first place, but also implies that one is probably overpaying, which is something a bidder will want to compensate for by underbidding. Nevertheless, the Revenue Equivalence Theorem can be shown to hold for common values auctions as well, as long as signals are independent.

A plethora of studies have been conducted to test the validity of these models, with mixed results. An overview of such studies that focuses on online auctions specifically can be found in Ockenfels et al. (2006). The remainder of this paper deals with the IPVP, except where stated otherwise. The $2 \log n$ rule is derived under the assumption of the IPVP, and an empirical test for it can be seen as a weak test for the paradigm. The empirical portion of this paper studies eBay auctions for iPhones, for which the private values model is a priori the more fitting one. As the true value of iPhones is known to all bidders, their bids aren't signals for it, but rather (only) of their own private values.

2.2 Auction Termination Rules

There are two main ways in which an online auction may end, each with different strategic implications. An auction may end at a predetermined time, the so-called "hard close", or it may face extension when certain conditions, usually the occurrence last-minute bidding, are met – the "soft close". The hard close system has been used by eBay since inception, and as eBay is the only large auction site left, fixed ending time auctions are currently by far the most commonly used. Amazon.com auctions (now defunct) used to offer auctions with a soft close, while Yahoo! auctions gave sellers the option to configure their auction in one of the two ways.

An eBay-style (i.e. hard close) auction will end at the time determined at the start of the auction (eBay auctions typically run for three, five, seven, or ten days). Although an auction may end early when a seller removes his listing or when a Buy-it-now option is used (on which more later), it cannot be extended past the indicated ending time. This gives bidders absolute certainty in regard to the final moments of the auction, and presents them with a new bidding strategy. A bidder may choose to postpone his bid until the very end and bid less than his valuation, in the hopes that his bid is registered by the system (there may be no time for repeat attempts), and that he won't be outbid in those few moments remaining by someone using the same strategy. This strategy is colloquially known as *sniping*, and its potential benefits to buyers are easy to see: when timed well (and with some luck), a bidder may get away with paying less than his valuation, or even less than that of his competitors. Other bidders may simply be following the auction less closely and have no time to react; other snipers on the other hand may fail to bid in time, or their attempted last-second bid might be lower.

Sniping is deemed unfair by some, and serves to lower the efficiency of auctions that permit it: the seller fails to extract a price equal to the bidders' second-highest valuation (as predicted by the Revenue Equivalence Theorem), and a bidder, who wouldn't normally have won the auction, nevertheless can do so. There exist several ways to mitigate the problem, the two main ones being proxy bidding (discussed in the next subsection), and soft close auctions. In soft close auctions (henceforth: Amazon-style auctions), the ending time is automatically extended whenever a bid is placed close to the ending time. On Amazon, placing a bid ten minutes or less from the end of the auction would extend it by an additional ten minutes, counting from the moment of said bid. This (recurring, if necessary) ten-minute delay should in principle give enough time to react to all but the most inattentive bidders, and thus diminish sniping considerably.

Sniping has been observed in several studies, perhaps most famously by Ockenfels and Roth (2002); they found that about 50% of eBay auctions in their sample showed bids in the last five minutes, 37% had bids in the last minute, and 12% in the final ten seconds (sniping of the more die-hard kind). By contrast, only about 3% of Amazon auctions showed bids in the final five minutes before the initially publicised ending time. Other studies show similar findings; a study of a large number of auctions by Hayne et al. (2003) found last-minute bidding in 25% of the 16.000 examined auctions. Sniping has become so widespread, that websites such as esnipe.com now offer to automate the sniping process for their customers. Even the mainstream media have taken notice.[2]

[2] e.g. http://www.usatoday.com/tech/science/columnist/vergano/2006-06-25-ebay-physics_x.htm, http://www.nytimes.com/2002/09/05/technology/news-watch-auctions-how-to-outbid-the-fanatics-software-that-never-sleeps.html

2.3 The Bidding System

Most online auction sites, eBay included, offer buyers two possible modes of bidding. When placing a *manual bid*, a potential buyer simply enters a bid higher than the current maximum, and as long as this new bid is larger than the previous top bid by at least the minimum increment (which is generally a certain percentage of the prevailing maximum bid), it becomes the new maximum bid. It's the simplest form of bidding in an ascending price auction, and the majority of bids are done in this way. In the final fifteen minutes of any eBay auction, a new type of manual bidding becomes available. This so-called "one-click bidding" allows bidders to bid one minimum increment on top of the then-current high bid with a single click, which speeds up the process of incremental bidding somewhat (and makes incremental sniping easier, too).

The second type of bidding is *proxy bidding*. When bidding by proxy, some of the bidding is done automatically by the system. The bidder enters a maximum value (which remains hidden from all other participants, including the seller) and the server does the rest of the bidding, up to the maximum determined by the bidder. Whenever the proxy is outbid, it will automatically raise its bid so as to come out on top (while observing the minimum increment), up until the maximum. In the event that two proxy bids collide, the one with the highest maximum will win, with a bid equal to the losing proxy bid's maximum (plus the required minimum increment). Proxy bidding relieves bidders of some of the drudgery of manual bidding, i.e. having to be online to react to an opponent's bid, keeping close tabs on the auction, etc. The benefits transcend mere convenience however, since proxy bidding can be used as a tool to combat sniping. Because the bids done by the proxy bidding system are instantaneous, it is impossible to successfully snipe against it (the proxy bid will always come in last), unless the sniping bid is higher than the proxy bid's maximum. eBay tries to make bidders aware of this fact in its user manual:

"Placing a high bid in the closing seconds of an auction-style listing is called "sniping" within the eBay community. Sniping is part of the eBay experience, and all bids placed before a listing ends are valid, even if they're placed one second before the listing ends. […] To protect yourself from being outbid at the last moment, enter the maximum amount you're willing to pay for an item up front, and eBay will bid automatically for you, making sure you're the high bidder until your maximum is reached."

Proxy bidding adds a Vickrey auction element to an otherwise normal English auction. The proxy bid mechanism can be seen as a sealed bid (the maximum bid is unknown to all), and the winner (the bidder with the highest proxy bid) pays the price of his closest competitor's proxy bid maximum plus the minimum increment. Furthermore, it follows from the Revenue Equivalence Theorem that the outcome of an auction would be the same whether proxy bidding is used or not. eBay explains it as follows:

"We encourage all members to use the proxy bidding system to bid the absolute maximum they are willing to pay for an item right from the start and let the proxy bidding system work for them. The proxy system will not take more of your bid than necessary to win the auction, thereby guaranteeing you the lowest winning price possible. This way, although it may be disheartening if you are outbid, you will have the satisfaction of knowing someone else was willing to pay more than you were."

One would therefore expect bidders to bid early and to bid their valuation using the proxy bidding system to avoid losing the auction through sniping. Interestingly, this does not happen too often; several reasons for this discrepancy have been proposed in literature. Ockenfels and Roth (2006) found that the more experienced eBay users are more likely to bid late. They proposed (and modelled) that late bidding is driven profit-seeking and the desire to outsmart so-called incremental bidders, i.e. bidders who keep coming back to outbid others (often by the smallest possible increment), sometimes going beyond their valuation (which results in bidding wars, often attributed to auction fever – an irrational desire to win the auction, even if it means having to pay overmuch). Esnipe.com explicitly states that avoidance of bidding wars is one of the benefits of its service: "*eSnipe reduces bidding wars by masking interest in auction items until the last possible moment. Because auctions on eBay take three to 10 days to close, emotional overbidding can start soon after the auction opens and last until the auction closes, instead of mere minutes as is the case in traditional offline or "outcry" auctions.*" Auction fever is a very interesting behavioural phenomenon, especially so because it predicts outcomes which are completely opposite to those of the standard models. Unfortunately, it falls outside of the scope of this paper.

Interestingly, Wintr (2008) finds that more experienced eBay users actually bid somewhat earlier. However, he identifies two separate groups of bidders (collectors and experts) with opposite behaviour. Collectors tend to bid high and early, because their main concern is to acquire the item; losing the auction and having to wait (sometimes for an extensive period of time) for another auction is experienced as a high cost. Experts on the other hand bid at the last moment to protect their information (i.e. to prevent signalling their true valuation to the other bidders), which can be especially important to bidders in common-value auctions.

Finally, a factor which may prevent bidders from using proxy bids and bidding early, is the fact that although it is impossible to tell their proxy bid maximum, everyone is nevertheless aware that the automatic bidding system is at work. Sellers may be tempted to milk the proxy bid for all it's worth by placing small, incremental bids using another account, essentially turning a Vickrey auction into a first-price sealed-bid one. Although risky – one might overshoot the proxy bid maximum and inadvertently win the auction – and expressly forbidden by eBay's rules, this *shill bidding* occurs regularly. Though very difficult to spot, some studies of shill bidding have been attempted. Kauffman and Wood (2003) found that approximately 6% of the 10.000 auctions under examination showed signs of shill bidding; Engelberg and Williams (2006) estimated that almost 1,5% of all bids in the 40.000 auctions looked at were in fact shill bids.

For these and possibly other reasons, the vast majority of bids are of the simple, manual variety. Bidders need not keep track of the auction manually however, by virtue of the notification system.

2.4 The Reminder System

All auction sites offer an e-mail notification system to bidders, which automatically sends out an e-mail when they are outbid, as well as on a host of other triggers. eBay also offers the possibility to be notified through an instant messaging (IM) or short message service (SMS). This reminder system obviates the need for manual bidders to keep checking up on the auctions manually; this affords them a measure of convenience – eBay auctions tend to run for days. Furthermore, it allows them to respond to being outbid instantaneously, if they so wish – close to what the automatic bidding system would do. This is an important consideration for the derivation of the $2 \log n$ rule, on which more in section 3.

2.5 The Buy-now Option

The last modality of online auctions to be discussed here is the so-called "Buy-it-now" (BIN) option. Sellers at their discretion may include a price at which the item can be purchased immediately, bypassing the bidding process and the necessity to wait (sometimes a relatively long time) until the end of the auction. The buy-it-now price is temporary, however, and usually disappears as soon as the first bid is made. It gives bidders the choice to buy out early, or to bid in the hopes of winning at a lower price. A variation of the BIN option can be found on some smaller auction sites (such as marktplaats.nl), which allow bidders to contact a seller directly to negotiate a price outside of the auction, and, if successful, ending it prematurely.

Though the BIN option usually disappears as soon as a bid is made, there are two possible exceptions. Because there is some debate concerning the disappearance of the BIN option (the reasoning behind the disappearance being that bidders should not have to worry about the item disappearing after they've placed a bid, without them being able to do anything about it), eBay has selected several categories of items to run a test on: on these auctions (motor parts & accessories, tickets, clothing, shoes & accessories, cell phones & PDAs), the BIN option stays enabled until the current bid reaches 50% of the BIN-price. It is currently unknown how long this will remain the case, and whether other categories will follow. Second, the seller may choose to have a reserve price, i.e. a price which bids much reach for the item to be sold at all. When there is a reserve price, the BIN option stays on as long as this reserve is not met.

Several reasons have been proposed as to why buyers may or may not want to utilise a BIN option. For an overview of the recent literature, see for instance Ockenfels et al. (2006). They conclude that the BIN option is as of yet poorly understood. Unfortunately, it is not immediately clear which effect, if any, it should have on the standard models, or the validity of the $2\log n$ rule. But, it is something that occurs with some frequency.

3. The $2\log n$ Rule

The $2\log n$ rule, proposed by de Haan et al. (2008), simply states the following: when certain conditions hold, the number of bidders in an online auctions will approximate $2\log n$, where n is the number of potential bidders. This section offers an overview of the reasoning and the assumptions behind the model. For the in-depth look as well as the mathematical proofs, see the original paper. Whenever this papers refers to dHdVZ, "the authors", or similar, de Haan et al. are meant.

3.1 The Set-up

The $2\log n$ rule is a consequence of the insight that certain information about valuations can be gleaned from the bid sequence. Consider an online auction of the soft-close variety. This assumption is crucial, because of the strategic implications of a hard close, i.e. the incentive to delay one's bid (to snipe). Furthermore, we limit ourselves to auctions for which the IPVP holds; the bidders therefore do not reveal any valuable private information about the item's value by bidding their valuation (which removes another incentive for sniping). This perhaps can best be summarised as follows:

Assumption A: *The auction format does not reward late bidding*.

As outlined in Section 2, the (weakly) dominant strategy in such auctions is to simply bid one's value. When the only mode of bidding is by proxy, we should expect to see bidders come in and place their value as proxy bids. Because of the way proxy bids work, every new *recorded* bid will be the previously highest bid (plus a minimum increment), thus revealing the sequence of valuations up to the highest one.

Things change somewhat when manual bids are also allowed (which, as mentioned before, form the bulk of the bids in online auctions). Manual bidders do not generally bid their valuation, choosing instead to bid incrementally, and drop out when the current bid reaches their valuation. Incremental bidding by two manual bidders or by a manual-proxy bidder pair will therefore eventually "resolve": one will drop out, with his valuation as his last bid – thereby revealing the second highest valuation up to that point. However, if this incremental bidding does not happen quickly enough, there is a chance that a third bidder would intervene and bid in lieu of another, raising the standing bid, perhaps past the other's valuation – thus muddying the bid sequence. This is where the notification (reminder) system comes in: it allows manual bidders to respond immediately to being outbid. A second assumption, introduced by Song (2004), is therefore necessary:

Assumption B: *Each manual bidder returns to the auction immediately to respond to being outbid, as long as his valuation is higher than the prevailing price.*

As long as Assumption B holds, bidding "battles" between two bidders will always resolve before a third can intervene, revealing the valuation that is the lower of the two. New potential bidders who arrive at the auction will not bid unless their valuation is higher than the prevailing bid. Therefore, under the IPVP, and if both assumptions hold, the bid sequence can be seen as a sequence of (second-) highest valuations. This is summarised by the following proposition:

Proposition A: *In an internet auction with a hybrid system of manual and proxy bids, and under the IPVP, each active bidder's valuation is the highest or second-highest among all the valuations of the potential bidders who were active before.*

3.2 Bids as a Record Sequence and the Rule

From the previous subsection (Proposition A), it follows that bids can be seen as a sequence of *records* of the valuations of the bidders. Every time a new (potential) bidder comes in whose valuation is higher than the current standing bid, the sequence is updated with the second-highest valuation (the valuation of the new bidder or that of the previously highest one). dHdVZ use this fact to derive the $2 \log n$ rule. A simplified account follows.

Let $i = 1, 2, \cdots, n$ denote the order in which n potential bidders arrive at a given auction, and assume that the valuations of these potential bidders are independent and identically distributed (i.i.d.) random variables with distribution function $F(x)$, denoted by $X_1, X_2, \cdots, X_n$. The rank sequence for the valuations of the bidders $i = 1, 2, \cdots, n$ can then be defined as

$$R_i \equiv \sum_{k=1}^{i} 1_{\{X_k \geq X_i\}}, \tag{1}$$

where R_i is the rank of the valuation of the i-th potential bidder to arrive at the site, X_i being his valuation. Note that R_i is the i-th potential bidder's rank among those who arrived before him. For example, if bidders $\#1, 2, 3, 4, \ldots$ arrive at the auction one after another, and they value the item at $\$10, \$20, \$15, \$5, \ldots$, respectively, then the rank sequence is given by $1, 1, 2, 4$, and so on.

The valuation of the i-th potential bidder X_i is called a *record* if $R_i = 1$ (i.e. if his valuation is higher than that of the bidders who arrived before him), and a *2-record* if $R_i = 2$ (i.e. if his valuation is second-highest). The remainder of the derivation studies a particular *record sequence*, and relies on theory of records; for the formal derivations, see the original paper as well as Resnick (1987).

We can construct an index sequence of the records and 2-records, $\{J(j)\}_{j=1}^{m}$. This index sequence is to identify those potential bidders whose valuation is either highest or second-highest of the ones who've come before them, i.e. it identifies those potential bidders who become *active* bidders. The index sequence $1, 2, 4, 6$ says that of the first six potential bidders to arrive, only the first, second, fourth and sixth actually become active bidders. The valuation of the third is lower than that of the first two, which according to Proposition A means that he won't get to make a bid; the standing bid at that point will be higher than his valuation (the standing bid will be equal to the lower of the first two valuations). The same reasoning applies to the fifth bidder. The fourth bidder has a higher valuation than one of the first two (or both of them), so he will place a bid; either the first or the second bidder will at this point have dropped out of the race (which of the two this will be cannot be learned from this index sequence alone); the same goes for number six.

The index sequence can be formalised as follows:

$$J(1) = 1, J(2) = 2, \text{ and} \tag{2}$$

$$J(j + 1) = \min\{i > J(j) : R_i \leq 2\}, j = 2, 3, \cdots, m - 1, \tag{3}$$

where m is the number of active bidders; a number such that $R_i > 2$ for all $i > J(m)$.

The $2\log n$ rule can be derived from this index sequence. Using results derived in Resnick (1987), it can be shown that as long as the number of active bidders is large enough, the distance between them follows a specific distribution. Formally, given that $k \to \infty$, the sequence

$$\{\log J(k+j) - \log J(k+j-1)\}_{j=1}^{\infty}, \tag{4}$$

i.e. the difference between the logs of the index numbers of two consecutive active bidders, is asymptotically i.i.d. with exponential distribution and mean $1/2$. This implies that as long as the number of active bidders m is sufficiently large, $\log J(m)$ will be approximately equal to $m/2$.

Recall that m is the number of active bidders, while $J(m)$ denotes the number of potential bidders who have arrived before (and including) the m-th active bidder. It can therefore be said that if there are n potential bidders, the number of active bidders will be approximately $2\log n$.

3.3 The Arrival Process

The $2\log n$ is relatively easily testable by using a simple linear regression, as long as one knows the number of potential and active bidders. The number of active bidders is relatively easy to gather: most (if not all) auction sites publish bid data, and sometimes even report the actual number of bidders separately. The number of potential bidders is a much more nebulous statistic, however, and a proxy must be found. dHdVZ use the number of *page views* on a particular auction as a proxy for potential bidders. Though it is likely the best measure of potential bidders available, it comes with a caveat or two. First, it may not be an entirely accurate representation of potential bidders (more on this later); second, it is not reported by all auction sites. It is the latter problem that dHdVZ found themselves confronted by; while a small Dutch site (marktplaats.nl) did report page views at the time, Yahoo! (their other source of data) did not.

Although Marktplaats data might have sufficed to test the $2\log n$ hypothesis, Yahoo! was a priori a better fit due to certain peculiarities of markplaats auctions (more on this in the discussion subsection below). For more rigorous results, Yahoo! data had to be used as well. However, to do this, a way had to be found around the limitations of the dataset (i.e. the non-availability of page view counts). The authors got around it by introducing an additional assumption: a specific *Poisson arrival process* for potential bidders.

The Poisson process is a continuous time stochastic process in which the probability of a certain number of arrivals within a time increment follows the Poisson distribution, which depends on a particular arrival rate (the expected number of arrivals within a time increment). In a *homogeneous* Poisson process, this arrival rate is constant and the distribution is independent of the increment (i.e. its location in time). In other words, the probability of n arrivals in a certain period is fixed, and independent of when one looks. This seems unsatisfactory for online auctions; though bidders in soft-close auctions do not have an incentive to bid late, they can be deemed to prefer auctions with shorter ending times; all things being equal, they would prefer a shorter wait to win the item over a longer one. The authors therefore assume a *non-homogeneous* Poisson arrival process, in which the arrival rate grows over time (so that the expected number of arrivals in later periods is higher than in the earlier periods). The assumed function for the arrival rate $\lambda(t)$ is:

$$\lambda(t) = \lambda_0 e^{\theta t} \tag{5}$$

where the auction begins at $t = 0$, and θ is the time preference factor. This is summarised by the following assumption:

Assumption C: *Potential bidders arrive at the auction site following a non-homogeneous Poisson process with an increasing arrival rate over time, defined by* $\lambda(t) = \lambda_0 e^{\theta t}$.

As long as the arrival process assumption holds, it can be shown that that the arrival times of new active bidders $\{T(J(j))\}_{j=1}^{m}$ follow a specific distribution also. In particular, for $l \to \infty$, the distances in time between the arrivals of two consecutive active bidders, i.e. the sequence

$$\{T(J(l+j)) - T(J(l+j-1))\}_{j=1}^{\infty} \tag{6}$$

is asymptotically an i.i.d. sequence with exponentially distributed innovations with the mean $1/(2\theta)$.

Armed with this proposition, it is possible to test the $2\log n$ rule indirectly by using the arrival times of active bidders. If the $2\log n$ rule holds, then the distances between the entry times of new active bidders will be distributed in a particular manner; this is testable.

3.4 Empirical Testing

As mentioned earlier, dHdVZ use two different venues to empirically test the $2\log n$ rule: first, a direct test using linear regression on a dataset gathered on Marktplaats (a small Dutch auction website), and second, an indirect test using bidder arrival timing on a Yahoo! dataset.

3.4.1 Regression evidence from Marktplaats data

Perhaps a single test of (the more complete) Marktplaats data would have been enough, were it not for several peculiarities which set Marktplaats auctions apart from plain vanilla ones. First, bids on Marktplaats *aren't binding*; they are merely an indication for the seller (and other bidders) of a bidder's willingness to buy and his price. Second, bidders have the option to contact the seller with an offer directly; if the seller accepts the offer, the auction will end prematurely. This could be seen as a continuous Buy-it-now option with a hidden price.

These deviations from ideal theoretical conditions are likely to have an effect on the number of active bidders, yet it is not immediately obvious what this effect is to be. The ability to contact a seller directly will likely lower the number of bidders, as some choose to do so rather than place an open bid. On the other hand, due to the non-binding nature of bids, some bidders might bid even when the current bid surpasses their valuation of the item; they incur no risk, and can always decide at a later date whether or not they really want to pay the price. This would suggest that the number of bidders would be higher than under standard conditions. Other strategic considerations may play a role, as well, and might have an effect on the applicability of the $2\log n$ rule.

To account for the BIN feature, the model is modified slightly. The main idea is that out of n potential bidders, only a certain (fixed) percentage pn will actually place a bid; the rest will choose to contact the seller directly. The number of active bidders $N(n)$ becomes $N(pn)$, and should be approximately equal to $2\log pn$. The equation to be estimated is then:

$$\frac{N(pn)}{\sqrt{\log n}} = \frac{2\log pn}{\sqrt{\log n}} + \varepsilon_n = \beta_1\sqrt{\log n} + \frac{\beta_0}{\sqrt{\log n}} + \varepsilon_n, \qquad (7)$$

where $\beta_1 = 2$ and $\beta_0 = 2\log p$ and should be a negative number. Both sides of the equation are divided by $\sqrt{\log n}$ to ensure homoscedasticity of the error terms. Additionally, from the estimated β_0, one can calculate p, which is the estimated percentage of potential bidders who are in principle inclined to place an open bid rather than contacting the seller directly.

The Marktplaats regression results seem to support the $2\log n$ hypothesis, with an estimated β_1 of 1.98 (significant at the 5% level) and an estimated β_0 of -5.43 (significant at the 10% level). In addition, p is calculated to be 6.6%. However, the value of this particular result is questionable, due to the high std. error associated with β_0 (p is 100% when $\beta_0 = 0$, which is only 1.43 standard deviations away). The evidence seems encouraging: β_1 lines up neatly at 2, suggesting that the $2 \log n$ holds. One must however keep in mind the caveat of the non-binding nature of Marktplaats bids, which may very well be muddying up the underlying bidding process. Unfortunately, there does not seem to be a way to elicit those.

Coefficients:				
Parameter	Estimate	Std. Error	*t* Statistic	p-value
β_0	-5.43	3.80	-1.43	0.082
β_1	1.98	0.61	3.24**	0.970
R-Squared: 0.153		Adjusted R-Squared: 0.125		

**: Significant at the 5% level.

Table 3.1: Regression results from de Haan et al. (2008)

3.4.2 Bid timing evidence from Yahoo! data

The second empirical test of the $2 \log n$ rule involves the distribution of entering times of new active bidders, performed on soft-close Yahoo! Auctions. This test is done under the maintained assumption that the arrivals of new potential bidders follow a particular non-homogeneous Poisson arrival process (Assuption C). From this assumption combined with the $2 \log n$ rule, it follows that the distances between consecutive arrivals should be exponentially distributed with mean $1/(2\theta)$.

The testing methodology used here is two-fold: a graphical, QQ-plot based test (see Figure 3.1), as well as a Kolmogorov-Smirnov (KS) goodness-of-fit tests. Both seem to support the hypothesis (see Figure 3.1: the data points are more-or-less nicely lined up on the $y = x$ line), i.e. that the entry time intervals are exponentially distributed with mean $1/(2\theta)$, with θ estimated on a per-auction basis. Neither test manages to reject the $2 \log n$ hypothesis.

As with the Marktplaats regression test, there is a drawback; the Poisson arrival process, though necessary, may not hold at all. Unfortunately, a complete lack of data on Yahoo! auction page views means that the assumption could not be verified (of course, had the data been present, there would've been no need for the indirect test at all).

And so, evidence from both the regression and the indirect bidder entry time tests points towards the $2\log n$ rule being valid. Both sets of evidence have (possibly serious) drawbacks, however. In the first case, the somewhat exceptional nature of the forum (marktplaats.nl) raises the possibility that one is comparing apples with oranges; in the second, the cost is a possibly onerous assumption about the arrival process of new potential bidders.

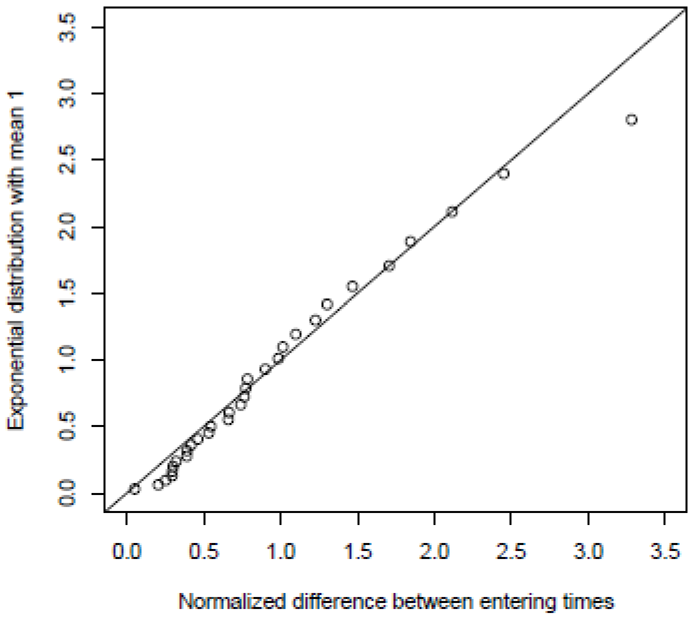

Figure 3.1: QQ-plot combined auction data, normalised entry time intervals.

The main thrust of the current paper is to utilise a fresh, more complete dataset gathered from eBay auctions to test the $2\log n$ rule as well as to verify some of the previous results (the Poisson arrival assumption in particular). Unfortunately, this new dataset is not without its own flaws. Specifically, the auctions examined are of the hard-close variety (the only kind available on eBay). This does a priori seem to violate an important condition for the validity of the model, and a way must be found to work around this limitation. It is to be hoped that from the analysis of this additional data, a more complete picture will emerge with regard to the $2\log n$ model.

4. The Dataset

To be able to subject the $2\log n$ rule to more thorough empirical testing, it is crucial to have a more complete dataset than the one dHdVZ had at their disposal. The choice was made to attempt to gather data from one of the big three auction sites (ebay.com, amazon.com, and yahoo.com), but by the time this paper had gotten into its planning stages, two of the three (Amazon auctions and Yahoo! auctions) were defunct (they now only offer classified ad services). The choice therefore fell on eBay quite naturally. Unfortunately, unlike the other two sites, eBay offers hard-close auctions only. On the one hand, this complicates matters as it is a departure from the theoretical conditions under which the $2 \log n$ is derived; on the other, it offers an opportunity to study the applicability of the rule in a somewhat different setting. It is therefore, at best, a double-edged sword. The study focuses on iPhone auctions, because they are quite plentiful and relatively homogenous. Though the main analysis to be used in this paper is to be performed on a per-auction basis, the freedom to make a homogeneity assumption (so as to be able to pool data across different auctions) may be convenient.

For a direct test of the $2 \log n$ rule, two metrics are vital: measures of potential and active bidders. Additionally, for an attempt at testing the Poisson arrival process assumption, one needs a record of the arrivals of potential bidders over time. Active bidders are published by (almost) all auction sites, making data ubiquitous; this isn't the case for potential bidders. The problems start with the fact that there is no direct way to measure them, which means that a proxy must be found. The best available proxy is the number of page views for a given auction, which is an indication of how many people have seen the auction in detail (though it too is not perfect, see below). Unfortunately, in contrast with the number of active bidders, the number of page views is not published on most auction sites (which is why dHdVZ had to resort to an indirect test of the $2 \log n$ rule on Yahoo! data).

Thankfully, as it turns out, some eBay auctions *do* sport a page views counter. On eBay, sellers are allowed to create (or customise) their own auction pages, and have the choice to include a views counter (or use their own). This means that some of the auctions have a page views counter embedded in them; the trick is to extract the necessary data. Although visible at a glance (see Figure 4.1) when accessing a specific auction's page, the counter isn't conveniently accessible, nor is there a record of the page views over time (which is necessary to test the arrival process assumption).

Figure 4.1 Example of what a page views counter may look like

At first, an attempt was made to gather data manually, by accessing each auction's page and copying the page views by hand. The process proved exceedingly cumbersome, and the necessary rigour as well as the long auction runtime (auctions run for up to ten days) meant that a decent time series could not be constructed, and that the number of auctions observed would be extremely limited by necessity. This attempt was subsequently dropped. The process needed to be automated. Unfortunately, it did not seem to lend itself to automation easily. The views counter, though visible when viewing a page in a regular browser, could not be accessed otherwise. Luckily, on closer inspection, it turned out that *some* of the auction pages had the counter embedded in their source code (which got updated every time the counter changed value). This allowed for relatively easy access to the counter, and a program could be developed which was to load the page, parse it for the counter string, and record the value along with a timestamp. This could be done at arbitrary time intervals; for the sake of convenience, the choice was made to record the views every 5 minutes. At this stage it was also decided to concentrate on 10-day auctions, so as to make the process uniform for all examined auctions (and to preserve homogeneity where possible).

Data gathering began in earnest on April 12, 2009. A computer was set up to run the program 24/7, recording the page views for a number of auctions. The number of auctions monitored was kept low (batches of 10-20) as a precaution; too many "hits" in too short a time (and over a period of weeks, months) might've been interpreted by eBay as malicious activity. Several problems were encountered along the way: internet access was not as stable as one would have liked; some auctions were removed prematurely; layouts were changed; counters went from readable to unreadable. Data collection continued (despite the hitches) until June, when eBay suddenly changed its auction page layout, rendering the previous method for ripping the views counter unusable.

Fortunately, a second method presented itself after the somewhat serendipitous discovery of the script used by eBay to generate the page views counter. By plugging an auction id number (which itself is retrievable from the auction page) into the script, one gets a readout of the views counter for that auction. Unlike the previous method, this one does not depend on whether or not the seller opts to include the counter in his auction, or the specific layout of the page. This made the data gathering process both easier and more stable. The final version of the program used (written in Java) is included as Appendix A. With the new method, data collection could be resumed in July and continued through October. For an example of the resultant time series, see Figure 4.2.

As mentioned previously, page views aren't a perfect measure of potential bidders. It may be useful to expand on that here. There are two main reasons for this: first, the page views registered by a counter might not all represent *unique individuals*, and second, not all potential bidders may in fact be registered by the page views counter.

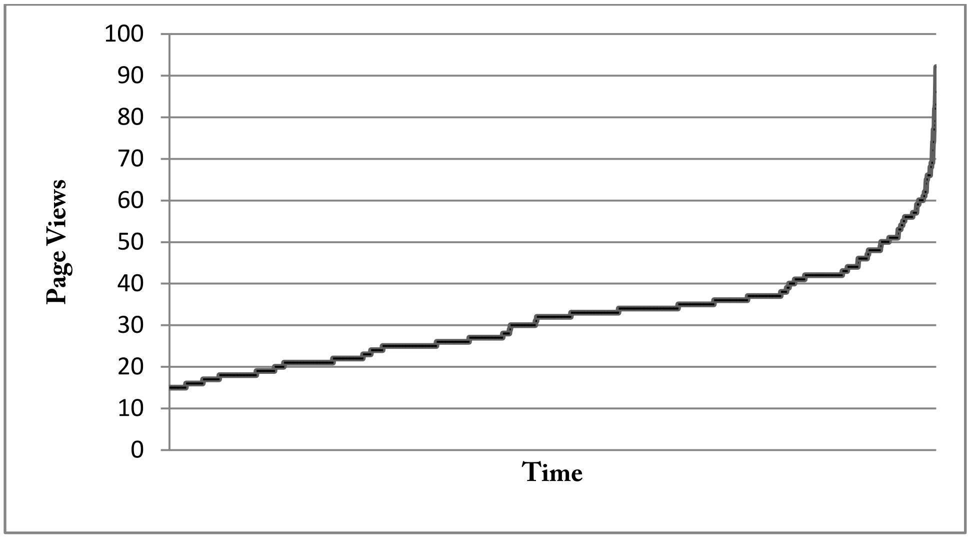

Figure 4.2 Example of a page views time series (10-day auction); a very typical picture

To understand the first problem, it is important to know how the page views counter works: it registers page visits from *unique IP-addresses*. An IP address is a numerical string assigned to a device (connection) by an internet provider. It is either *static* if it does not change when a user disconnects from and reconnects to the internet, or *dynamic* if a provider assigns a random address within a certain range when a user connects. The existence of dynamic IPs makes it quite conceivable that a single user (potential bidder) is registered more than once as he visits a page, then disconnects and reconnects to the internet, and visits it again. However, the effect users with dynamic IPs have on the page views count should be rather small in practice. First, static IPs are the norm for broadband connections, and we should expect that eBay users who are interested in iPhones are relatively gadget-savvy and unlikely to be on a slow dial-up connection. Second, in the case of broadband connections, users generally stay connected even when their computers are off (their IP thus remaining unchanged) by virtue of their modems staying on (and online) at all times. Third, as evident from Figure 4.2, most page views occur in a relatively short time towards the end of an auction, which makes it unlikely that many users will have had their IPs changed in that period.

The second problem may be more insidious. It arises from the fact that certain information about an auction (price and the number of bids) is visible without accessing the actual auction (and thus being detected by the views counter). The output of the eBay search engine (Figure 4.3) does not merely give links to the auctions themselves, but also shows the number of bids and the price. Potential bidders whose valuation of the item is lower than the current highest bid will likely never click on the auction, and will therefore never be registered by the views counter. The number of page views is thus likely to be lower than the actual number of potential bidders, and it is impossible to tell by how much.

Figure 4.3 Example of eBay search results

Despite these shortcomings, page views seem to be the best *available* measure of potential bidders. Perhaps with (far) more extensive access to eBay's data, a better proxy could be constructed. For example, it might include the number of (unique) searches for iPhones within a certain period (for example, while an auction is somewhere at the top of the search list). Such a measure would have its own pitfalls. In any case, one must make do with what one has, and the page views are the best (and possibly only) indicator of potential bidders available.

Other information is readily available and significantly easier to collect. The number of bidders, the bid sequences, and several characteristics of the auctions under examination were gathered after the auctions in question had ended. The collected information included the starting and ending times of the auctions, the number of bids, the starting price, the winning bid, the buy-out price (where applicable), and a few defining characteristics of the item in question (condition of the iPhone, model, whether it ships internationally). A bid sequence includes a full list of bids (including proxy bids), the bidders' names (though made anonymous by eBay) and their feedback scores.

5. Empirical Analysis

This section deals with the main meat of the paper: the empirical analysis. After a brief presentation of some basic statistics, I make an attempt to replicate the results from de Haan et al. (2008), and discuss the hard-close nature of eBay auctions and what effect this element should have. An in-depth examination of the page views time series follows, with testing the veracity of the Poisson arrival process assumption as the main goal. An attempt is made at bringing all of the results together.

5.1 Basic Statistics

Of the 312 attempts to gather bid and page views information from specific auctions, 294 resulted in at least *some* data for some type of analysis. For 51 of these, there is no bid data available at all, mostly due to premature listing removal (generally, bid data is available for 90 days after an auction's end). Views data was gathered, however. Of the 243 left, 41 ended early because of a utilised buy-it-now option. This is a considerable number; unfortunately, due to the way the data was gathered, there is no way of telling how many of the auctions actually had a buy-it-now option set up by the seller. Eight auctions were cancelled before their slated end times. Six auctions do not have their detailed bid sequence available (they were set to "private" by the sellers), but their totals for bidders and page views are known, so they can be included in the direct test of the $2\log n$ rule. The same goes for the five that ended without the reserve price having been met.

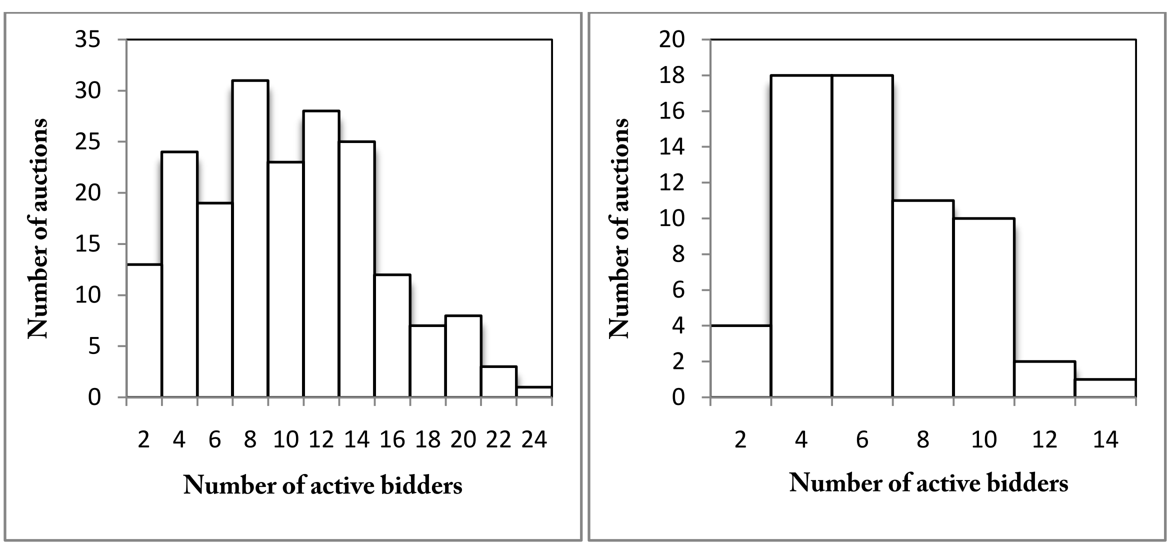

Figure 5.1: Histograms of the number of active bidders in the eBay (left) and Marktplaats (right) datasets

In total, 194 data points are available for the direct test; 188 of these have a full time series of page views as well as a complete bid sequence. For the full set of 194 auctions, the number of active bidders is distributed as shown in the left panel of Figure 5.1. At first glance, the distribution seems to be far wider than observed by dHdVZ; their histogram is reprinted here in the right-side panel of Figure 5.1. Values far in excess of 14 are observed in the eBay sample. The average observed in the eBay sample is considerably higher as well: 9.6 vs. 6.5.

This difference should give one pause, because the predicted number of potential bidders which follows from the $2 \log n$ rule rises exponentially with additional active bidders. At 24, the number of potential bidders predicted by the rule is over 160,000; this is more than an order of magnitude greater than anything observed (see Figure 5.2).

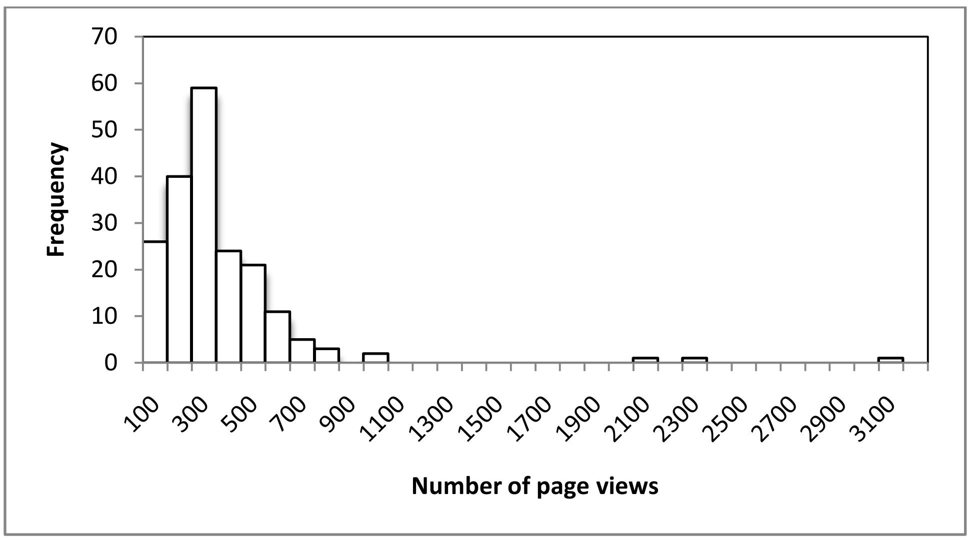

Figure 5.2: Histogram of the number of page views (eBay); entire dataset

Variable	Mean	Std. Dev	Min	Max	Median
Potential bidders (page views)	314.3	323.0	19	3008	246
Active bidders	9.6	5.1	1	23	10
Bids	21.0	13.1	1	59	20
Starting price (US Dollars)	114.1	153.7	0.01	799	50
Highest bid (US Dollars)	346.9	181.2	40	1589	318.3

Table 5.1: Summary statistics of some auction characteristics; entire dataset

Finally, to give the reader some idea of the range of a few characteristics of the auctions in the gathered sample, some statistics have been collected in Table 5.1. These include the number of potential and actual bidders, the number of bids, the auction starting prices, as well as the height of the winning bids (although one needs to keep in mind that in some auctions no actual exchange occurred because the winning bid did not reach the reserve price).

5.2 Simple Regression Analysis

Regression analysis is used here in a fashion identical to dHdVZ; the number of active bidders is regressed on $\log n$, with the slope coefficient equal to 2 if the $2 \log n$ rule is true. For a first (naive) stab at a direct test, the following equation is estimated:

$$N(n) = \beta \log n + \varepsilon, \tag{8}$$

where $N(n)$ is the number of active bidders, and β, the slope coefficient, should equal 2. ε is an error term. The hypothesis to be tested is whether the slope coefficient equals two or not:

$$H_0\colon \beta = 2 \text{ versus } H_1\colon \beta \neq 2.$$

Auctions with at least five bidders are selected (of which there are 156), as per the original paper. This is necessary because what is being tested is an approximate relationship which is supposed to apply only when m and n are sufficiently large. The main statistics of the auctions that qualified can be seen in Table 5.2; for comparison, statistics of the dataset used by dHdVZ are provided also.

Variable	Mean	Std. Dev	Min	Max	Median
	eBay sample statistics				
Potential bidders	337.2	272.1	70	2297	276.5
Active bidders	11.2	4.2	5	23	11
	Marktplaats sample (de Haan et al., 2008)				
Potential bidders	603.7	326.4	164	1331	552
Active bidders	6.9	2.3	5	15	6

Table 5.2: Statistics of the data; five bidders or more

The differences between the eBay sample and the Marktplaats sample used by dHdVZ are rather stark; they cut at both ends, with significantly fewer potential bidders on the one hand, and more actual bidders on the other. This would seem to not bode well for the $2\log n$ rule. In fact, the statistics for views in the eBay sample are skewed by two uncharacteristically high values, as also seen in Figure 5.2. Removing the two extreme values drops the mean and standard deviation further, to 313.2 and 171.6, respectively. The eBay auctions under study therefore seem on average to have fewer potential bidders, while at the same time attracting more actual bidders.

The regression results are summarised in Table 5.3. At first glance, the results seem to offer strong support for the $2\log n$ hypothesis. It is tempting to leave it at that, yet there are two reservations to be made. First, the graphical output (Figure 5.3) at the very least suggests that the model could use an intercept coefficient. It looks like this should increase the slope coefficient considerably, as well. Second, it seems rather inexplicable that the hard-close nature of eBay auctions should have no effect on the applicability of the $2\log n$ rule at all. In fact, sniping happens, and it happens often; that is a clear deviation from the theoretical conditions under which the rule was derived.

Parameter	Estimate	Std. Error	*t* Stat.	p-value for H_0
β	2.00	0.0523	0.033	0.974
R^2: 0.223				Obs: 156

Table 5.3: Empirical test on (8)

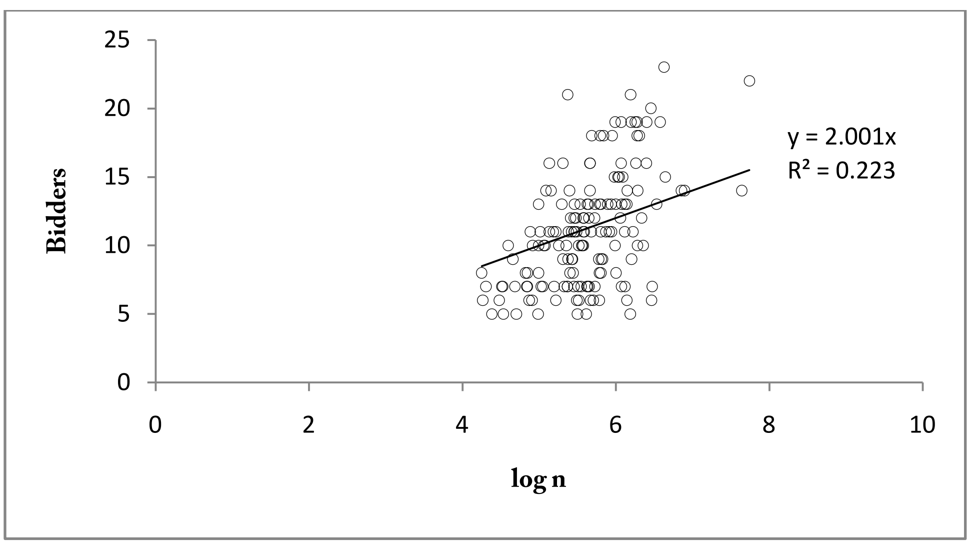

Figure 5.3: Regression output of (8); active bidders on $\log n$, no intercept

As expected, adding an intercept coefficient to the equation changes the results dramatically. The new equation to be estimated here is:

$$N(n) = \alpha + \beta \log n + \varepsilon, \tag{9}$$

where α is the intercept coefficient. The regression results are given in Table 5.4, while the graphical output has been relegated to Appendix B (Figure B-2.1). The intercept is negative and highly significant, and the new slope coefficient deviates significantly from 2, so that the $\beta = 2$ hypothesis is rejected rather convincingly. In addition, the inclusion of an intercept coefficient raises the R^2 from 0.223 to 0.285 (0.281 adjusted); overall, a model with an intercept seems a better fit for the data.

Parameter	Estimate	Std. Error	t Stat.	p-value for H_0
α	-9.83	2.69	-3.65**	
β	3.73	0.476	3.64**	0.0003
R^2: 0.285		**: Significant at the 5% level		Obs: 156

Table 5.4: Empirical test on (9)

This presents a problem: after all, there is no interpretation in the $2 \log n$ model for an intercept coefficient. Likewise, the model does not take into account the strategic aspect of sniping at all, i.e. the fact that some bidders will choose to wait until the very last moment to enter their bid. Perhaps these two aspects can be rolled into one.

5.3 Regression Analysis: Strategic Bidding

The empirical model used by dHdVZ did in fact include an intercept, and maybe the same principle can be made to work for eBay auctions specifically. Recall the estimated equation used for the Marktplaats dataset:

$$\frac{N(pn)}{\sqrt{\log n}} = \frac{2 \log pn}{\sqrt{\log n}} + \varepsilon_n = \beta_1 \sqrt{\log n} + \frac{\beta_0}{\sqrt{\log n}} + \varepsilon_n, \tag{7}$$

where $\beta_1 = 2$ and $\beta_0 = 2 \log p$, p being the proportion of potential bidders who choose to place an actual bid rather than contact the seller directly. There are therefore two groups of bidders: pn, who place bids, and $(1-p)n$ who choose to negotiate with the seller. Only pn potential bidders are assumed to translate into actual bids placed. Perhaps an analogy can be drawn to eBay and sniping, if one could split the total population of eBay bidders into two: those who snipe, and those who do not. Seen this way, there would be pn potential bidders who place bids throughout the entirety of an auction's lifespan, and $(1-p)n$ potential bidders who place bids in the closing moments of an auction only.

Two questions present themselves. First, can one reasonably speak of two distinct groups of bidders, or do they overlap? And second, can bids placed by these groups be distinguished in practice? The answer to the second question relies heavily on that to the first; if it is in fact possible to speak of two separate groups, then all one needs to do in order to construct a sample in which only non-sniping bids are expressed is to discard auctions that have bids in their final moments. There is of course no clear definition of sniping to be had; here, I'll define sniping as placing a bid within the final 10 minutes of an auction. This is admittedly arbitrary, and probably too wide; real, hard-core sniping happens in the closing *seconds* of an auction. Yet, 10 minutes seems to be a fitting criterion for ensuring that none of the remaining auctions exhibit sniping. In addition, 10 minutes was used in soft-close auctions as the cut-off point (trigger) for auction extension (for example on Amazon).

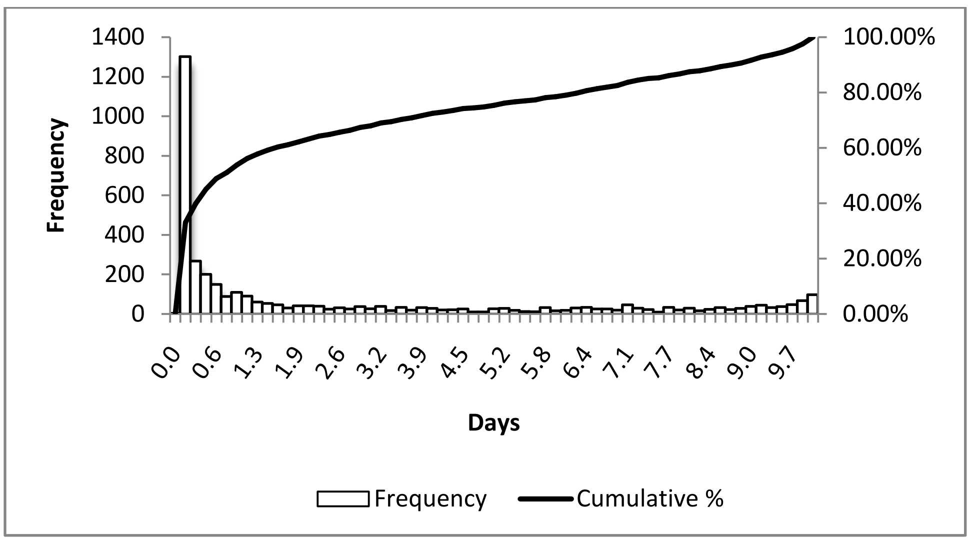

Figure 5.4: Histogram and cumulative percentage graph of bid times; 10-day period

What remains now is to actually establish whether or not the view of eBay buyers as comprised of two distinct and separate groups is a realistic one. Fortunately, the availability of bid sequence data allows one to form some inkling of an idea. The distribution of bid times alone reveals quite a lot; see Figure 5.4 for a histogram and cumulative percentage graph of bid times for the entire sample (over a 10-day run time).

Though it is undoubtedly skewed towards the last hours of the auctions, there is nevertheless a significant amount of bidding in the days prior. In fact, one observes a similar distribution at several zoom levels (see Appendix B, Figures B-3.1 – B-3.6), in a fractal-like fashion. 407 out of a total of 3880 bids take place less than 10 minutes before the closing time: slightly less than 10%. It is a relatively small portion of the whole. It is also interesting to note that the very start of auctions seems to attract more bids than the days that follow; this is in line with observations made by Shmueli et al. (2004).

The aforementioned fractal pattern breaks down at the 10-second level and seems to be centred around 5 seconds; perhaps that is the timeframe that snipers aim for. This is corroborated by the recommendations made to its users by esnipe.com:

"Our tests show the ideal buffer time to be about 4 to 8 seconds. […] If you are concerned about eBay being slow during a very important auction, we recommend 10 seconds to play it safe."

It is important to note that bidding early is of absolutely no strategic value to a sniper; one risks unleashing a bidding war, or announcing one's arrival thereby making other bidders aware of one's presence. The vast majority of bids therefore would be made by non-sniping bidders. Does this distinction hold in reality, however?

This question can be answered by using relatively simple metrics gleaned from the (final moments of the) bid sequence. Table 5.5 shows a tally of bids placed by old and new bidders within the final 10 minutes of an auction, split into bidders who entered before the 10-minute mark and those who entered after. The bid sequence tells that of the 238 bidders active in the final 10 minutes, about 81% had not placed a bid earlier in the auction. Likewise, new bidders are responsible for 76% of all the bids placed in those final 10 minutes. Though 19% might seem like a sizeable chunk, one must keep in mind that many of these old bidders would have been responding to the newcomers' late bids, and not attempting to snipe.

It is informative to see what happens when one zooms in closer on the finishing line. Of the 68 auctions with (winning) bids in the final 10 seconds, 36 were won by entirely new bidders, then entering the fray for the very first time; these, one could say, are the hard-core snipers. Of the 32 remaining, 17 would have been won by such hard-core snipers, had it not been for an earlier placed proxy bid by another bidder (which goes to show that proxy bidding can be a valuable tool in combating sniping; the reason it is not used more often is likely due to fear of shill bidding). Of the last 15 left, 14 were won after a short period of frantic bidding initiated by new bidders, responded to by old ones, and so on. In these cases, the final bids were preceded by bids placed by new bidders, which were subsequently challenged by others; this suggests that they were not in fact attempts at sniping (which is the deliberate aiming of a bid at the last possible moment), but rather normal bidding activity in the only time still available. Finally, there remains a single instance (out of 68) where an old bidder came back after a protracted period (days in this case) to issue a last-second bid. With but a single such case, one may justifiably regard the entire population of bidders as consisting of two (almost) entirely distinct groups: snipers and non-snipers. This also means that a "clean" sample with only non-sniping active bidders can be produced by leaving out auctions with last-minute bids.

	Old	Old %	New	New %	Total
New vs. old bidders	45	19%	193	81%	238
Bids by new vs. old bidders	99	24%	308	76%	407

Table 5.5: Bidding in the final 10 minutes

As before, only auctions with five bidders or more are counted. There are 46 such auctions in total. (see Table 5.6 for the statistics). Here, too, the distribution is skewed by an observation with an unusually large number of potential bidders (2081 page views); with it removed, the mean and standard deviation become 282.2 and 161.5, respectively.

Variable	Mean	Std. Dev	Min	Max	Median
Potential bidders	321.3	309.6	80	2081	247
Active bidders	9.9	4.1	5	23	9

Table 5.6: Statistics of the no-sniping subsample; five bidders or more

The $2\log n$ model should fare better when tested using this sample, as it is a closer fit to the theoretical conditions under which the rule was derived. The methodology followed here is the same as used by dHdVZ in the Marktplaats case. It is assumed that pn potential bidders will not attempt to snipe and will simply bid normally, leaving $(1-p)n$ snipers. Of course, pn is not observed directly; n, the total number of potential bidders, is. However, if the $2\log n$ rule holds, then the number of active bidders in auctions where no last-moment bidding occurs should be approximately equal to $2\log pn$. With this in mind, the equation to evaluate becomes:

$$N(pn) = 2\log pn + \varepsilon_n = 2\log n + 2\log p + \varepsilon_n = \beta_1 \log n + \beta_2 + \varepsilon'_n, \qquad (10)$$

where, if the rule holds, $\beta_1 = 2$, and $\beta_2 = 2\log p \leq 0$. The hypotheses to be tested on these coefficients are thus:

$H_{0,0}: \beta_1 = 2$ versus $H_{0,1}: \beta_1 \neq 2$, and

$$H_{1,0}: \beta_2 > 0 \text{ versus } H_{1,1}: \beta_2 \leq 0.$$

The regression output can be found in Table 5.7; the graphical representation has been included in Appendix B (Figure B-2.2).

Parameter	Estimate	Std. Error	t Stat.	p-value for $H_{i,0}$
β_1	3.24	0.78	1.60	0.12
β_2	-8.03	4.33	-1.85*	0.07
R^2: 0.285	Adjusted R^2: 0.267	**: Significant at 5%; *: 10%		Obs: 50

Table 5.7: Empirical test on (10); no sniping

As expected, the $2\log n$ does fare somewhat better; the $\beta_1 = 2$ hypothesis cannot be rejected at the 10% confidence level. Unfortunately, this is in large part due to the higher standard error. $H_{1,0}$ (i.e. that the intercept coefficient is above 0) can be rejected at the 5% but not at the 10% level. Using the fact that $\beta_2 = 2\log p$ in the model, it is possible to calculate $p = e^{\beta_2/2} = 0.018$. According to the regression results, only about 1.8% of the entire eBay user population should fall into the non-sniping category. This seems a gross understatement, and is likely to be incorrect; of course, due to the large standard error, it would be unwise to put stock into an exact measure (it's a single standard error away from being 15.7%, and the percentage increases exponentially as the coefficient gets closer to 0). These results seem to be rather ambiguous; nevertheless, they do not reject the $2\log n$ rule outright.

In an attempt to solidify these findings, a second regression is performed using a sample of data where sniping is likely to have occurred. This time, only auctions with winning bids within 60 seconds of the finishing time are considered; though sniping has been defined before as placing a bid within the final 10 minutes, this definition is tightened somewhat here. The murky area in between 10 minutes and 60 seconds where one might or might not justifiably speak of sniping has been left out. The equation used is, as before, (10). This time however, only $H_{0,0}$ is tested for (versus $H_{0,1}$), because the intercept has no direct interpretation in a sample where sniping is present. The results are found in Table 5.8. The graphical output has been relegated to Appendix B, as Figure B-2.3.

Though the difference between these results and those of the previous test is not overwhelming, it is enough to make it so that the $2\log n$ rule is rejected at the 5% level here, whereas this could not be done with the non-sniping sample. However, neither of the coefficients differ significantly from each other, and sadly, no unambiguous conclusions about the veracity of the $2\log n$ rule can be drawn.

Parameter	Estimate	Std. Error	t Stat.	p-value for $H_{i,0}$
β_1	3.70	0.73	2.33**	0.0226
β_2	-9.96	4.09	-2.43**	
R^2: 0.242	Adjusted R^2: 0.233	**: Significant at the 5% level		Obs: 82

Table 5.8: Empirical test on (10); sniping present

Further testing raises some doubts about these results, however. Recall that the 10-minute mark for sniping bids was more or less arbitrarily chosen. In fact, the fractal (self-scaling) pattern of the timing of the bids suggests that perhaps the 10-second cut-off point is a better one. A similar regression using only auctions without bids in the final 10 seconds rejects the $2\log n$ convincingly (β_1 = 3.40; std. error: 0.57; p-value for $H_{1,0}$: 0.015). The previous result therefore may well have been spurious.

Finally, some attention should be given to the problem of heteroskedasticity. An examination of the plotted error terms from the regression on (9) reveals strong heteroskedasticity (see Appendix B, Figure B-4.1). This is not at all surprising; it is quite reasonable to believe that the variance of the number of actual bidders will increase with the number of potential bidders. In fact, the $2\log n$ model predicts heteroskedasticity, with a theoretical standard deviation of actual bidders equal to $\sqrt{2\log n}$. dHdVZ use this prediction to correct their regression for heteroskedasticity; an OLS regression on (7) is equivalent to a weighted least squares regression on (9), with weights equal to $1/\sqrt{2\log n}$ (the least weight is given to observations with the highest variance). The same method was used in an attempt to correct for heteroskedasticity in the current dataset as well, but to no avail: the plot of the errors from the regression on (7) still shows considerable heteroskedasticity (Appendix B, Figure B-4.2). This may be interpreted as additional evidence indicating that the $2\log n$ rule does not fit the eBay data very well. Since adjusting by the theoretical weights does not seem to resolve the problem, two supplementary methods are used throughout this paper to correct for heteroskedasticity: residualisation[3] on the independent variable (i.e. $\log n$), and heteroskedasticity-consistent standard errors as

[3] The residualisation precedure works as follows. The squared errors of the regression to be corrected are themselves regressed on the independent variable ($\log n$). The inverse of the square root of the predicted values of this regression ($1/\sqrt{pred}$) are subsequently used as weights in a new Weighted Least Squares regression, reducing the influence of extreme values on the estimated coefficients.

per White (1980). However, these corrections do not lead to markedly different results in any of the tests performed, and shall not be reported here.

5.4 The Potential Bidders Time Series: Structural Breaks

So far, nothing has been said about the gathered page views (potential bidder) time series; perhaps examining those more closely might reveal something of use. This subsection shall attempt to combine insights gained from page views data with regression analysis from the previous sections; Section 5.6 deals with the Poisson arrival process assumption made by dHdVZ for their indirect test.

Examination of the page views series reveals an interesting feature: the number of potential bidders rises slowly but steadily over the course of days until just a few hours before the end of the auction, where the arrival speed changes rather suddenly. For an example of this, see Figure 5.5; many auctions show a break point that is visible to the naked eye, though sometimes the change is more gradual (for an example of this, see Figure 4-2 in the previous section).

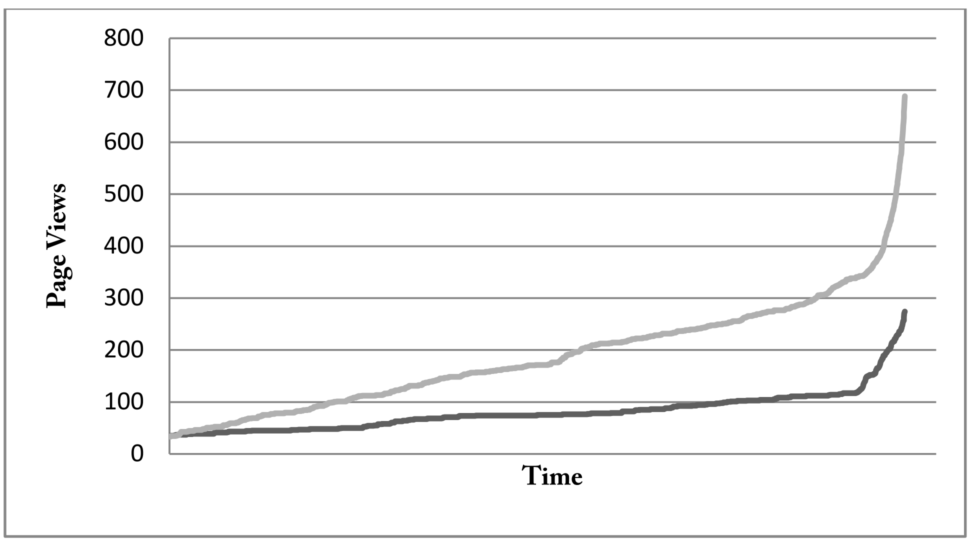

Figure 5.5: Page views time series of two auctions, full 10-day runs

Perhaps this sudden influx of new (potential) bidders corrupts the orderliness of the bidding process; mounting time pressures and heightened activity may lead to a situation in which bidding "battles" between two bidders are no longer allowed to "resolve" before other bidders enter, leading to intervening bids and throwing off the applicability of the $2 \log n$ rule.

To see whether this is in fact the case, it is first necessary to locate this break point. To that end, an iterative least-squares regression is used; a procedure reminiscent of the Chow test. Each series of n observations is divided into two groups of u and $n - u$ observations. The number of page views is subsequently regressed on time for both groups, and the sum of squared residuals for both regressions added up. This is done for every $1 < u < n$ in an effort to minimise the total sum of squared errors S_t. The u that leads to the lowest S_t is deemed to be the break point.

This method seems to work reasonably well; however, in cases where the increase in page views in the final hours is less pronounced, and there are aberrations in the earlier parts of the time series, the procedure may pick up on those, instead (an example of such a series can be seen in Appendix B, Figure B-5.1; the exceedingly quick ascent at the end combined with the hump in the middle leads to a false break point estimate). I try to mitigate the problem by truncating the dataset to the last 500 observations only. This seems justified, because of interest here is the quick change of pace at the very end, and 500 data points gathered at 5 minute intervals still cover a 2500-minute (i.e. 41-hour) period. Figure 5.6 shows the distribution of break points thus calculated (the distribution of break points calculated using the entire sample can be found in Appendix B, Figure B-5.2).

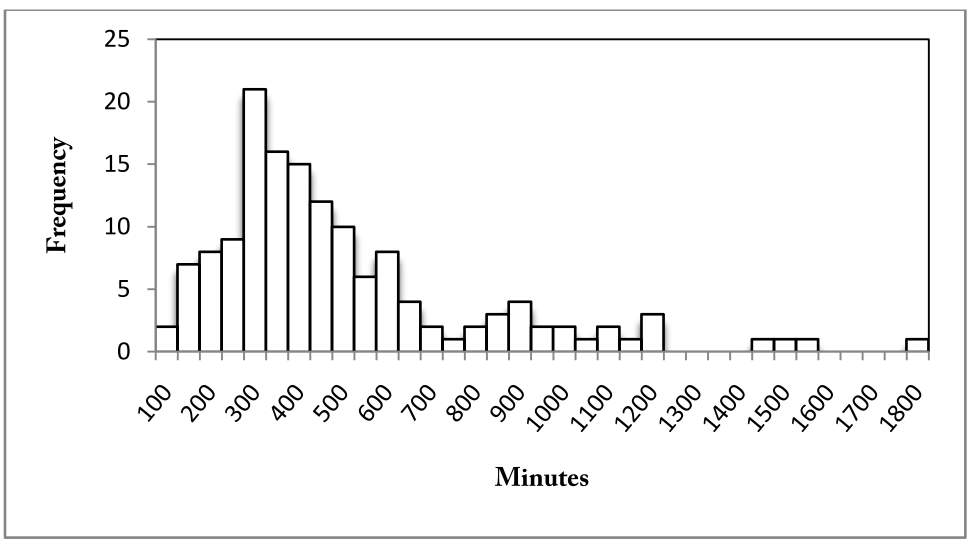

Figure 5.6: Histogram of page views time series break points, truncated dataset (500 obs.)

Again, the method does pick up some points relatively early on in the auctions, but a visual inspection of the series shows that the upswing at the end, though somewhat less pronounced, is present in those auctions also. The distribution seems to be centred on 350 minutes. One possible explanation for this (often very sudden) rush of new potential visitors is that around this time, auctions start appearing on the first page of the search results for that particular item (when exactly this happens depends on the actual number of open auctions at that time). This leads to much higher visibility, because many people don't bother clicking through to other pages (the number of pages can be quite large depending on the item in question), and results are often ordered by the time remaining.

In an effort to provide some empirical support for this hypothesis, a small sample of the times at which auctions jump onto the first search page has been gathered as well. The number of results per page is customisable, and can be set by users to 200, 100, or 50. The 200 setting is used here, because that is the earliest point at which users begin seeing auctions on the first page of their search results. The actual number of users who employ this setting is not known, however. The average of these observations is 302, with a standard deviation of 142; the explanation offered would therefore seem to be a reasonable one. It is an interesting issue; however, to expound on it further would be to go beyond the scope of this paper.

The applicability of the $2\log n$ rule should not depend on the arrival rate of potential bidders in principle; a higher arrival rate simply means more potential bidders, which should translate into more actual bidders. However, one of the underlying conditions for the derivation of the $2\log n$ rule is that bidders respond immediately when they are outbid. In practice, this assumption will undoubtedly fail to hold. This needn't be onerous; when the arrival rate of potential bidders is slow enough, there may be sufficient time for overbid bidders to react without interference from others, with the same end result (i.e. as though reaction times were instantaneous). Perhaps it is this assumption that is violated more gravely in the last hours of an auction after the break point in the page views, as more and more new potential bidders come pouring in. If this is true, then the $2\log n$ rule should fare better in the period before the break. To test this, once again linear regression on (10) is used, but this time using the number of potential and actual bidders at the calculated break point. As before, only auctions with 5 bidders or more qualify (there are 88 such auctions). Because the breaks occur (relatively) long before the end, there should be no bids by snipers; both hypotheses can therefore be tested. Regression results are in Table 5.9; graphical output in Appendix B, Figure B-2.4.

This test, too, rejects the $2\log n$ rule. It is interesting to note that the slope coefficient here is closer to 2 than it had been in previous tests; in that sense, it fits more closely with the $2\log n$ model. However, due to the lower standard error, it ends up failing in this instance also.

Parameter	Estimate	Std. Error	t Stat.	p-value for $H_{i,0}$
β_1	3.02	0.46	2.22**	0.029
β_2	-7.28	2.49	-2.92**	0.0044
R^2: 0.333	Adjusted R^2: 0.325	**: Significant at the 5% level		Obs: 88

Table 5.9: Empirical test on (10); no sniping

5.5 Regression Analysis: Extension

So far, the regression results have not been kind to the $2\log n$ rule. The model faces rejection in all tests but one, and while the creation of an adjustment for last-minute bidding does introduce some ambiguity, the results (i.e. the rejection of the rule) are dependent on the precise definition of sniping. With the sniping adjustment, eBay auctions seem to be in line with the theoretical conditions assumed by dHdVZ. Perhaps then it is the special nature of Marktplaats auctions which led to the failure to reject the model in the original study. In addition, it is useful to look at the asymptotic nature of the derived rule; therein may lie the reason why it seems to fail in the preceding tests.

Marktplaats: non-binding bidding

Marktplaats is different from sites like eBay in that its auctions are not *auctions*, per se. They have much more in common with negotiations than with run-of-the-mill auctions: bids aren't binding (neither to the bidder nor to the seller) and a final consensus is necessary before any transaction is to be made. This is markedly different for auction sites like eBay, where bids are legally enforceable and binding to all parties involved, often even when they are erroneous. Perhaps this element can be used to explain the apparent differences between the outcomes of the two studies.

The non-binding nature of Marktplaats bids means that bidders with extremely low valuations will likely never get to bid on the item, due to the fact that sellers must approve the transaction afterwards. Such bidders will simply have no hopes of winning, especially when an external valuation is available (such as the store price). This is different in the case of eBay auctions, where, as long as no other bidders show up and the seller does not set a reserve price, the bidder ends up winning the item. This may well be what some bidders aim for, as many eBay auctions exhibit extremely low bids early on, placed by one-time bidders, possibly hoping to make a large (though low-probability) kill. Conversely, one would not expect to see (very) low starting bids on Marktplaats: while bidding the minimum price occurs often in regular auctions, a non-credible bid in a negotiation may come across as insulting to the counterparty, or signal that the buyer isn't really serious about the purchase.

Perhaps filtering out the lowest bids is a way to replicate the Marktplaats results using the eBay sample. One way of doing this is to only look at auctions with starting prices higher than a certain threshold. The exact criterion cannot be chosen but arbitrarily, however. Unfortunately, the choice is dictated in large part by the availability of data, as the selection criteria are two-fold: a specific starting price (most auctions in the sample do not have a minimum price set up, and start out at $0.01) as well as the absence of sniping.

To exclude sniping, the 10-second mark is used, following the break-down of self-similarity in that period (see Appendix B, Figures B-3.1 – 3.6). Ideally, one would prefer to be consistent; unfortunately, the limited amount of data precludes the use of the 10-minute mark (too few auctions would remain for a meaningful regression analysis). Similarly, the limited availability of data narrows down the choice of starting price. Due to the fact that not all iPhones under consideration are exactly alike, a relative measure of starting auction price is used here, i.e. a certain percentage of the winning bid: a 10% starting bid seems reasonable. As before, only auctions with more than five bidders are considered; this leaves a total 47 auctions for the regression. Equation (10) and both hypotheses are reused.

Parameter	Estimate	Std. Error	t Stat.	p-value for $H_{i,0}$
β_1	1.92	0.84	-0.099	0.921
β_2	-1.66	4.66	-0.36	0.723
R^2: 0.102	Adjusted R^2: 0.0827	**: Significant at the 5% level		Obs: 47

Table 5.10: Empirical test on (10); starting price above 10% of winning bid, no sniping (10 seconds)

The results are found in Table 5.10, and the graphical output once again in Appendix B (Figure B-2.5). Interestingly, the results are more in line with those shown by dHdVZ, although the model's explanatory power suffers from the adjustment (the R-squared takes a big hit). In addition, the average number of bidders in this subsample is 8.9: lower than the previous 11.2 (see Table 5.2), and much closer to the 6.9 seen in the original study.

This seems to call the Marktplaats analysis into question: after all, it appears that the Marktplaats bidding mechanism excludes certain bidders from ever placing bids. However, it can also be argued that it is in fact eBay auctions that fail to fit the theoretical conditions. Perhaps (very) low bidders on eBay are simply trying to game the system by bidding as low as they can on as many auctions as possible, hoping that one of them will go unnoticed by other bidders, netting them a profit. This implies that they underbid by as much as they are able, bidding lower than their actual valuation of the item – which in turn means that the bid sequence fails to reveal their true valuations. In principle, the $2 \log n$ rule should apply even when some bidders do not bid their full valuation, and never had the intention to: in that case, their final bid may as well be their true valuation. However, if part of their strategy is to arrive at the auction *early*, then the model assumptions are violated, because the valuations of potential bidders are no longer random: early potential bidders would be more likely to have a lower valuation, leading to more bids overall than predicted by the model. This seems likely for another reason, as well: the strategy rewards bidding in as many auctions as one is able, including auctions which have just begun – this in contrast with bidders who place credible bids and are likely to stick to one auction at a time (winning multiple auctions would likely lead to suboptimal outcomes when they're simply looking to buy a single item). The format of Marktplaats auctions on the other hand does not seem to leave any room open for strategies of this kind.

It is impossible to tell which of these two arguments holds more water, but with them, an explanation for the difference between the results of the two studies has been proposed. Perhaps a comparison of bid sequences from the two sites could reveal more; it shall be left to future studies.

Asymptotic properties

One aspect of the $2 \log n$ rule has so far remained unexamined: its asymptotic nature. Recall (see Section 3) that the rule states that *as* $m \to \infty$, the number of active bidders shall be approximately equal to $2 \log n$. It is for this reason that dHdVZ limited their regression analysis to those auctions which attracted 5 or more bidders ($m \geq 5$). However, the choice of the threshold of 5 is somewhat arbitrary (as any such choice will be), and five may be too low a number to bring out the asymptotic rule. dHdVZ could not choose a higher threshold value, because the Marktplaats data (with its maximum of 14 bidders) did not allow for it. Fortunately, the greater variability in the eBay sample allows for a weakening of the restriction.

To this end, a series of regressions is performed in the same manner as before, with a changing minimum bidder threshold value. Sniping is again defined using the 10-second threshold, to maximise the available data. The results are summarised in Figure 5.7, which shows the slope coefficients with corresponding confidence intervals (two standard errors wide).

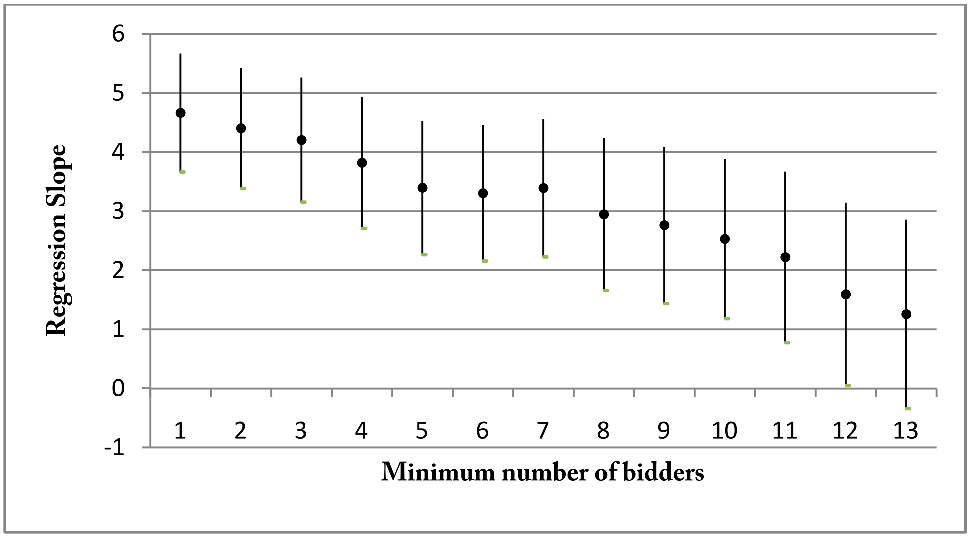

Figure 5.7: Varying minimum m: regression results (slope coef. with 2σ confidence intervals)

Interestingly, the slope coefficients show an unmistakably declining trend as m increases, while the sparser data also widens the confidence intervals. After the m reaches 8, one can no longer reject the $2\log n$ rule. This seems to fit with the idea that the $2\log n$ rule holds when m is sufficiently large. There is no indication however that the slope coefficient stabilises at 2; far more data would be needed to find out whether or not this is the case.

The preceding results also indicate that the regression outcomes are heavily dependent on the exact threshold chosen. This raises the possibility that choosing a different minimum value (in either direction) would have led dHdVZ to different results, perhaps significantly so. There is a capriciousness about the results arrived at in this manner which cannot be resolved easily, or perhaps at all; one should be mindful of this when interpreting them.

5.6 The Potential Bidders Time Series: the Poisson Arrival Process

Until now, this study has focused on replicating and expanding on the first type of evidence provided by dHdVZ: the direct test of the $2\log n$ model by means of linear regression of the number of active bidders on $\log n$. As has been outlined in Section 3.4 however, the evidence found is two-fold: a direct test on Marktplaats data, and an indirect one using data gathered from Yahoo! auctions. This subsection deals with the latter.

The indirect test is based on an extension of the $2\log n$ model, from which follows that under the assumption of a specific Poisson arrival process (Assumption C), and as $m \to \infty$, the differences in the arrival timing of new bidders will follow a specific distribution. This allows for an indirect test of the rule, using the bid timing sequence rather than page views (and bidder) data. dHdVZ do indeed find evidence for the $2\log n$ rule using this method, but for the results to be meaningful, the specific Poisson arrival process assumption must hold. The availability of the page views time series in the eBay dataset makes a direct test of this assumption possible.

Assumption C in Section 3 states that arrivals of new potential bidders to the auction follow a Poisson distribution with the time-dependent expected arrival rate $\lambda(t)$:

$$\lambda(t) = \lambda_0 e^{\theta t}, \tag{5}$$

where λ_0 is a constant and θ is the so-called *time preference parameter*. To verify whether or not the page views follow the Poisson arrival process posited in Assumption C, it is necessary to estimate (5). For this purpose, 5-minute arrival rates are used: $\xi(t) = v(t) - v(t-5)$, where $v(t)$ is the number of page views at time t. This is done to fully utilise the available data (page views were gathered at 5-minute intervals).

In principle, it is possible to estimate (5) using linear regression by taking its logarithm:

$$\log \xi(t) = \log \lambda_0 + \theta t + \varepsilon_t = \lambda_0^{'} + \theta t + \varepsilon_t, \tag{5'}$$

where $\xi(t)$ is the actual 5-minute change in page views from time $t-5$ to time t, $\lambda_0^{'}$ is a constant, and ε_t an error term. There is a serious drawback to this method, however: the actual number of arrivals within most 5-minute periods is zero, in which case $\log \xi(t)$ is undefined.

Instead, Poisson regression (see for example Colin Cameron and Trivedi, 1998) is the preferred method to estimate the following equation (for each auction i):

$$\lambda_i(t) = \lambda_{i,0} e^{\theta_i t} = e^{c_i + \theta_i t}, \tag{11}$$

where c_i is a constant and $e^{c_i} = \lambda_{i,0}$. Unlike OLS, Poisson regression does not minimise the squared errors, but maximises the likelihood function, i.e. the probability of a particular sequence of observations given a specific distribution (in this case, Poisson):

$$L(\theta_i, c_i) = \prod_{j=1}^{k} f(\xi_{i,j} | t_j, \theta_i, c_i), \tag{12}$$

where for each observation j, $\xi_{i,j}$ is the observed number of potential bidders who arrive at auction i in the five minutes preceding time t_j. t_j is the amount of time in minutes that the auction has been running for when observation j was taken. $f(\xi_{i,j}|t_j, \theta_i, c_i)$ is the Poisson probability density function:

$$f(\xi_{i,j}|t_j) = \frac{e^{-\lambda_{i,j}}\lambda_{i,j}^{\xi_{i,j}}}{\xi_{i,j}!}, \tag{13}$$

where the time-dependent mean parameter $\lambda_{i,j}$ (the expected arrival rate) is given by (11), ie. $\lambda_i(t) = e^{c_i+\theta_i t_j}$.

The time preference parameter θ_i estimated in this manner from the page views data offers an additional way to test the model. The indirect evidence found by dHdVZ relies on the derived result which states that the distances between the arrival times of bidders are (asymptotically) exponentially distributed with the mean $1/(2\theta_i)$. Because this too is an asymptotic relationship, only the last 2/3 of the bidder entry times are used to calculate the bid sequence θ_i (following the example set by dHdVZ), and once again only auctions with more than 5 active bidders are considered. There are therefore two ways to calculate θ_i which, if the model holds, should lead to the same outcome. In addition, the time preference parameter implies a "half-life" of an auction, which can be found by solving the equation $e^{-\theta t} = 0.5$; the resulting half-life indicates how much time must pass before potential bidders' incentive to check the auction doubles (recall that the incentive to look at an auction grows over time, as bidders prefer auctions that are closer to their finish).

Table 5.11 shows a comparison of the summary statistics of the time preference parameters arrived at using the two methods outlined above, with the corresponding half-life times. Several extreme values have been removed from the page views series (e.g. a half-life time of 135,064 minutes, or 94 days), because these skew the statistics. A very high half-life time (θ_i close to zero) means that the arrival rate remains (almost) constant for the entirety of the auction's duration. It also implies a large average time between the arrival times of bidders, sometimes in the order of weeks (as in the aforementioned case). Clearly, the 10-day running time of the eBay auctions precludes such observations from ever being made; the highest average observed is 58 hours (vs. 67 days implied by the minimum page views θ_i). One should therefore keep in mind that the resulting bid sequence thetas have a positive bias.

The results show quite a few differences. Perhaps most dramatically, the average θ_i calculated using the bidder arrival times is almost an order of magnitude greater than the θ_i estimated from the page views time series. It is also far more volatile, which is not too surprising: in most cases, it is calculated using only five data points or less. Although the bid sequence theta reaches much higher values, it does not drop quite as low, due to the aforementioned positive bias and possibly other reasons. The pattern is reversed for the half-life times, due to their relationship with the time preference parameters: a low θ_i implies a high half-life time, and vice versa.

Parameter	Mean	Std. Dev	Min	Max	Median
θ_i (page views)	0.000436	0.000388	0.0000585	0.00350	0.000346
θ_i (bid sequence)	0.00416	0.00846	0.000153	0.0819	0.00117
Half-life time in minutes (page views)	2604	1976	198	11855	2004
Half-life time in minutes (bid sequence)	1012	1067	8	4535	595

Table 5.11: Summary statistics of θ_i and the corresponding half-life times

If the $2 \log n$ model and the Poisson arrival process assumption hold, the two different thetas (and their corresponding half-life times) should be equal. To see if this is (more or less) the case, they have been plotted in Figure 5.8 (a few extreme observations have been left out for the sake of better oversight; unaltered plots can be found in Appendix B, Figures B-6.1 and B-6.2). If the model is true, one would expect the observations to line up on the $y = x$ line, but these clearly do not.

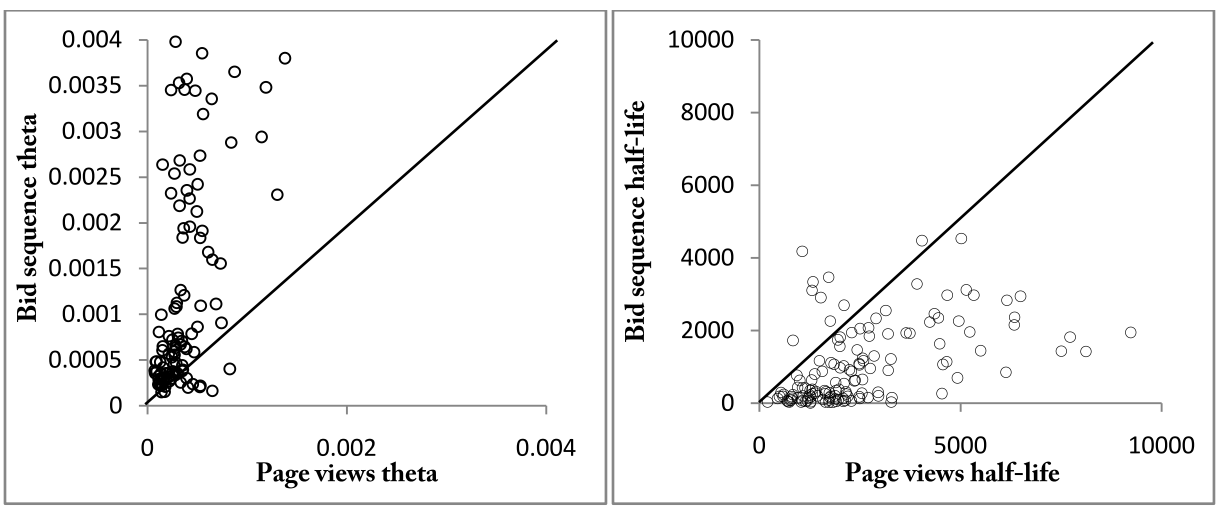

Figure 5.8: Plots of θ_i and half-life times estimated from the page views vs. those from the bid sequence

There is some positive correlation to be sure (ρ = 0.261 for θ_i and 0.424 for the half-life times), but this is to be expected: an auction that gets busier more quickly is also more likely to have bids follow each other up in more rapid succession. One cannot definitively say whether or not there is enough correlation to support the model's prediction; although the bid sequence θ_i appears to be much higher than one would expect given the page views paremeter (the linear regression slope coefficient is 5.7 instead of the expected 1), this may be due to the positive bias mentioned earlier.

It may very well be, however, that the Poisson arrival process assumption (Assumption C) simply does not hold, in which case the two methods of calculating the time preference parameters need not converge at all. The availability of page views time series data lets us test the assumption.

With the θ_i estimated from the page views time series, one can construct an "ideal" Poisson arrival process, i.e. one for which $\xi(t) = \lambda(t)$. This simulation should provide a decent first indication of whether or not the posited Poisson arrival process is a good approximation of the actual arrival of potential bidders onto the auction page. Figure 5.9 shows an example. The problem is clear: the Poisson arrival process seems to be ill-equipped to deal with the sudden surge of page views at the very end of an auction's lifetime.

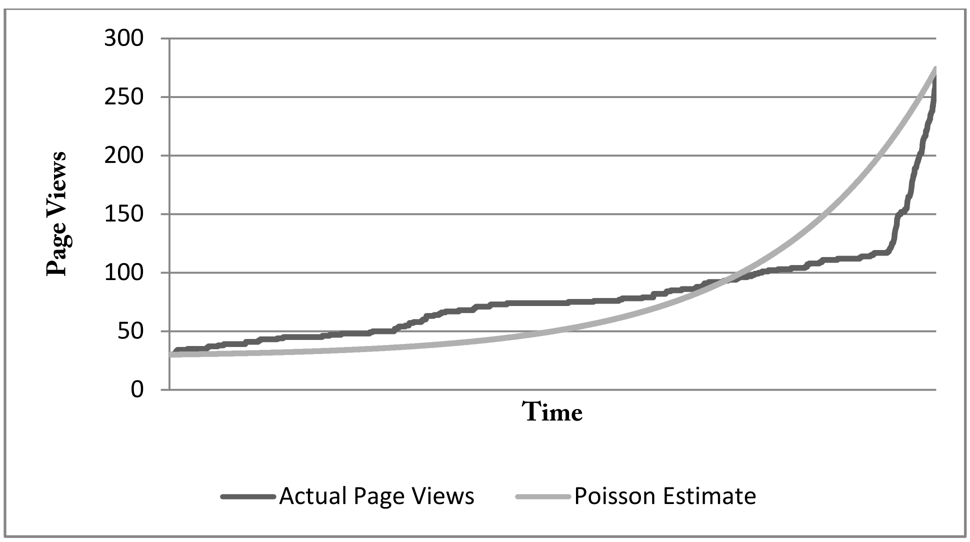

Figure 5.9: Simulated Poisson arrival process vs. the actual page views time series

Testing whether or not potential bidder arrivals are Poisson distributed is relatively straightforward (see for example Gelman and Hill, 2007). One of the characteristics of the Poisson distribution is the fact that its expected value equals its variance. Or, in the context of our model, $E(\xi_{i,j}) = e^{c_i+\theta_i t}$ and $\mathrm{sd}(\xi_{i,j}) = \sqrt{e^{c_i+\theta_i t_j}}$. This implies that if the Poisson model is true, the standardised residuals of the previously executed Poisson regression,

$$z_{i,j} = \frac{\xi_{i,j} - \hat{\xi}_{i,j}}{\mathrm{sd}\,(\hat{\xi}_{i,j})} = \frac{\xi_{i,j} - e^{c_i + \hat{\theta}_i t_j}}{\sqrt{e^{c_i + \hat{\theta}_i t_j}}}, \tag{14}$$

should be approximately independent (not exactly independent, as $\hat{\theta}_i$ is used to compute all of them), with mean 0 and standard deviation 1. If the arrival rates are indeed Poisson distributed, then the sum of k standardised residuals $\sum_{j=1}^{k} z_{i,j}^2$ should follow a χ_{k-l}^2 distribution, where l is the number of estimated parameters (in our case, l = 2).

Testing against the chi-square distribution in the manner outlined, all but two of the 147 auctions fail the Poisson distribution hypothesis at the 5% confidence level. The two that do not, are characterised by a very mild rise in the page views count in their final hours – not at all like most other auctions under study. Assumption C therefore fails, at the very least for the eBay dataset. There is no reason however to believe that the eBay potential bidder arrival process is significantly different from that of other auction sites. Though it is true that bidders may choose to *bid* late due to the strategic benefits associated with sniping, this does not mean that they will choose to arrive at the very last moment to do so. The converse may very well be true: a sniper may wish to research an item and its real worth before attempting to snipe. In addition, setting up an automatic bidding system (such as esnipe.com) takes time. There is therefore little reason to believe that the Poisson arrival process would fit Yahoo! auctions better, especially given the fact that Yahoo! used to provide auctions of both types (hard-close as well as soft-close). Still, in the absence of data, this cannot be confirmed; and unfortunately, with the website now defunct, no new data can be gathered.

The fact that the eBay data does not follow a Poisson arrival process may have implications for the indirect test performed by dHdVZ, which relied on a confluence of the $2\log n$ model and the Poisson arrival process assumption for its interpretation. It is of course possible that the arrival processes in eBay and Yahoo! auctions differ significantly, i.e. that the latter follows a Poisson arrival process while the former does not. However, if dHdVZ's results can be replicated using the eBay dataset (where, as has been established, Assumption C does not hold), it would be reasonable to believe that the same may be true for the Yahoo! sample as well, and that the observed exponentially distributed bid arrival series isn't necessarily indicative of the $2\log n$ rule.

Recall that the $2\log n$ model, in conjunction with Assumption C, predicts that the distances between the bidders' entry times will be asymptotically exponentially distributed with mean $1/(2\theta_i)$. Alternatively, one can standardise the entry time distances to mean 1 by multiplying all observations by $2\theta_i$; these standardised differences should also be asymptotically exponentially distributed with mean 1. This is the core of dHdVZ's indirect test, and is precisely what they find in the Yahoo! data (see Section 3.4.2 of this paper). For an attempt at replicating these results using the eBay dataset, it is probably best to stick as close as possible to the auction selection criteria used in the original paper. Due to the asymptotic nature of the rule, only the last 2/3 of the bidder entry times are used for the analysis, and only auctions with at least 4 active bidders and 25 bids qualify. In addition, auctions with bids in the final 10 minutes are excluded from the analysis, to account for sniping. The 10-minute mark is used here (instead of the previously used 10-second cut-off) to approach, as best one can, the properties of soft-close Yahoo! auctions. 16 such auctions fit the bill, for a total of 143 data points.

Figure 5.10 is a QQ-plot of the standardised entry time differences vs. an exponential distribution with mean 1. It seems to fit reasonably well, though by no means perfectly. The Kolmogorov-Smirnov produces a p-value of 0.0017. The difference, though relatively minor to the naked eye, is nevertheless large enough for the KS test to reject the hypothesis that the normalised differences between entering times are asymptotically exponentially distributed with mean 1 convincingly.

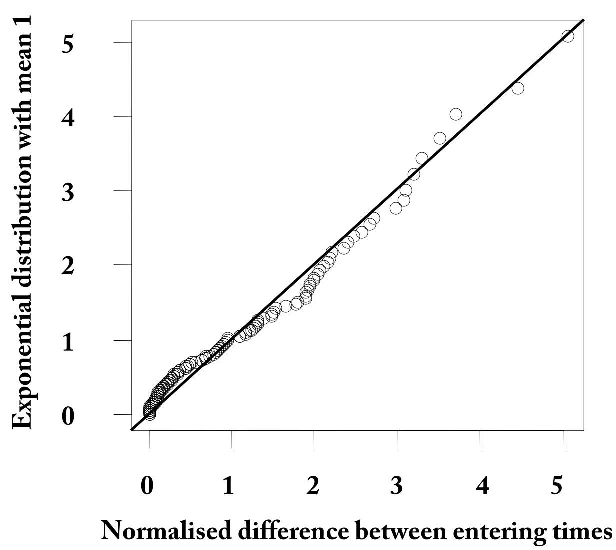

Figure 5.10: QQ-plot for eBay bid series data

This result is quite different from the one seen in dHdVZ's analysis of Yahoo! data. Unfortunately, it appears that it is at least in part due to the larger size of the eBay sample. If one restricts the sample size to that of the original study (31 observations), the results become far more ambiguous: a KS-test on a random sample of 31 observations (from a total of 143) leads to a p-value of 0.05 or higher in more than 70% of the cases (based on 10,000 random draws). It is therefore impossible to formulate unequivocal conclusions: one can say that the exponential distribution hypothesis is rejected for the eBay dataset (and with it the the $2 \log n$ model in conjunction with Assumption C, although the latter has already been disproved), but it isn't apparent whether or not parallels may be justifiably drawn to dHdVZ's study, due to the apparent size effect.

Given the outcome of these tests, it is hard to say just how different the Yahoo! dataset is from the eBay one (corrected for sniping). It is quite possible that Yahoo! auctions follow the Poisson arrival process, and that the $2 \log n$ model applies to them but not to eBay. The results of the two studies are diametrically opposed, and it is not clear whether this is because of differences in auction characteristics.

It is possible that the model can be made to work with an "overdispersed Poisson" arrival process (e.g. a "quasipoisson" distribution, where the variance is a multiple of its mean); it is however not clear whether or not such an adjustment is possible, and what its implications would be. A simpler correction may suffice. It seems reasonable that a simple Poisson arrival process cannot perfectly capture the structural breaks in the page views time series discussed in Section 5.4: both the arrival rate as well as the time preference parameter appear to change suddenly. It is likely that two separately fit Poisson functions would describe the actual arrival process much more closely.

To see whether this is the case, I split the time series of 10 auctions into two parts, using the break points calculated in Section 5.4 (recall that the break points are hypothesised to happen when an auction reaches the front page of the site's search engine). Subsequently, Poisson regression is used to estimate (11) for each segment separately, in the manner outlined above. Figure 5.11 shows the resulting expected arrival process for one of these auctions.

To the naked eye, this split Poisson process (the dashed line) seems to fit remarkably well. 12 out of 20 segments still fail the formal overdispersion test at the 5% confidence level, however. Nevertheless, this is a great improvement over the previous 145/147, and is perhaps not all too surprising: the break points found in Section 5.4 are relatively rough approximations, calculated using linear regression. The differences in the time preference parameter before and after the structural break are enormous: θ_i is on average 69 times higher after the break point. In terms of half-life times, the average goes down from 2.7 *days* to 2.1 *hours* (a factor of 77).

What does this mean for the $2 \log n$ rule, and the indirect test using the bid sequence? Theoretically, one can look at the periods before and after the break point as two separate auctions; the timing of bidder entries changes from one to the other due to the changing time preference parameter, but the applicability of the model and the predicted results should not. It should therefore be possible to test the $2\log n$ rule in the exact same manner as before, by splitting the bid series in two and using different θ_i to normalise the bidder entry time differences. However, in practice, splitting the bid series means that even fewer data points are available for calculating each θ_i. In essence, one deals with two auctions instead of one, each left with only a fraction of the bidders. Under these circumstances, the applicability of the $2\log n$ model cannot but suffer, given its asymptotic nature.

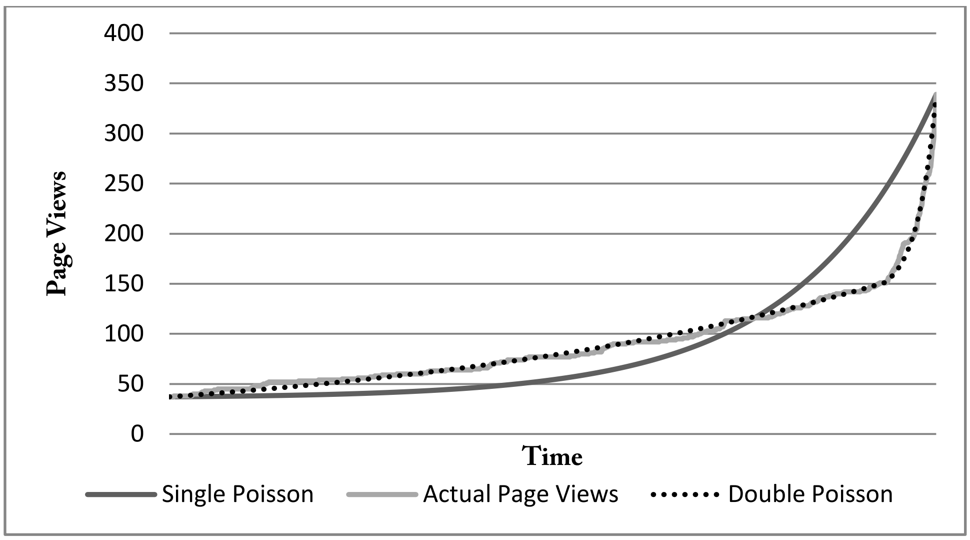

Figure 5.11: Poisson arrival process (split time series and otherwise) vs. the actual page views

Do Yahoo! auctions share eBay's structural break characteristic? The question is impossible to answer with the available data, but if answered affirmatively, it indicates a potential flaw in dHdVZ's methodology for the indirect test on Yahoo! data.

I shall refrain from repeating the indirect test using the break point split here: given the low number of bidders per auction, and the previously discussed size effect when dealing with the Kolmogorov-Smirnov test, its results would not be very meaningful. A larger dataset might be able to counter at least some of the problems – I leave it therefore up to future research.

6. Conclusion

dHdVZ's $2\log n$ rule offers a new, elegantly derived way of thinking about the extent of internet auctions, and formulates a central prediction that can be tested empirically. The previous study found evidence largely in favour of the rule, but had to overcome some imperfections in the available data. The primary concern of the current paper has been to replicate dHdVZ's findings using a new, more complete dataset gathered from eBay, and to explore several factors which may have an effect on the application of the model.

The evidence presented by dHdVZ is two-fold. For a small Dutch auction site called Marktplaats, data on potential and actual bidders was readily available, enabling the authors to perform a direct test of the rule. The same is true for the eBay dataset used here. For the second test, dHdVZ used data from Yahoo!, one of the larger auction sites. Even though Yahoo! did not publish the number of potential bidders (page views) for its auctions, an additional assumption (a specific Poisson arrival process for potential bidders) made it possible to test the $2 \log n$ rule indirectly, using the bid sequence only. Thanks to the availability of time series page views data, this additional assumption can now be tested as well (for eBay auctions, at least).

The results are mixed. Direct tests on eBay data using the same test parameters as seen in dHdVZ reject the $2 \log n$ hypothesis, even after an adjustment is made to the model to accommodate sniping. Two findings temper this conclusion. First, raising the minimum active bidder threshold changes the outcome of the test, and the rule can no longer be rejected. This is entirely in line with the asymptotic nature of the $2 \log n$ model, which is supposed to hold only when the number of bidders is sufficiently high. Some caution is warranted, however: though the model is not rejected when the active bidder thresholds are relatively high, there is some indication that the rule may not hold for very high active bidder thresholds, either. Unfortunately, testing this would require a (much) larger dataset. Second, the rule is not rejected for auctions with higher starting prices, and the test results are similar to those found by dHdVZ for Marktplaats. Perhaps this is because Marktplaats auctions, though more reminiscent of negotiations, actually capture the basic assumptions of the $2 \log n$ model more closely in some respects. And so, even though straightforward tests of the $2 \log n$ rule on eBay data seem to fail, there is reason to believe that this may not be the case if one discounts potentially assumption-violating early bidding, or auctions with relatively few bidders.

Analysis of page views data sheds some new light on the indirect test results achieved by dHdVZ. The hypothesis that potential bidders follow a Poisson arrival process is rejected, apparently due to the existence of structural breaks in the page views time series late into the auction. Once the structural breaks are taken into account, a (double) Poisson process seems to capture reality reasonably well. The causes of the breaks are unknown, though it is hypothesised here that breaks occur when auctions reach the first page of the website's search engine results for a particular item; a preliminary test seems to confirm this. If this explanation is correct, then structural breaks of this kind are likely to exist in Yahoo! auctions as well, calling into question the interpretation of dHdVZ's test results. These results could not be replicated using the eBay data; they are fairly similar however, and it cannot be ruled out that the difference is caused by the larger size of the eBay sample. It is therefore not entirely clear which conclusions, if any, should be drawn from the outcome of the original paper's indirect test. One simply cannot state with confidence that bid arrivals on Yahoo! differ significantly from those on eBay; sadly, this can no longer be confirmed empirically.

Though Yahoo! may be gone, other (smaller, local) soft-close auction sites remain, such as the New Zealand-based trademe.co.nz. It should be in principle possible to extend data collection method used here to other sites as well, perhaps on a larger scale. A multitude of questions can then be answered. It would be interesting, for example, to compare the number of bidders as well as potential bidder arrivals across different (types of) auction sites. Such a comparison should also reveal whether eBay is representative of most online auctions, or if it is the odd duck out.

There is far more research to be done on eBay data, as well. It would be of eminent interest to see how well the $2 \log n$ rule does across various product groups. If the hypothesis made here concerning starting prices holds, then one would expect the rule to fare better in the case of low-cost items, i.e. auctions which are less likely to be trawled by people trying to make a killing. An additional test of this hypothesis might involve cross-auction tracking of bidders: the expectation is that certain bidders are far more active during the opening hours (or the first few bids) of an auction than others. Finally, a much larger sample may also make it possible to explore the margins, and to confirm that the model holds better when the number of active bidders is high.

While no longer directly related to testing the validity of the $2 \log n$ rule (because it can be tested directly), page views time series data offers new opportunities for research. The structural breaks found in the eBay data may be pinpointed more accurately by using the fact that the arrivals are Poisson distributed instead of the linear regression method used in this paper. The exact location of the breaks could perhaps be explained using various product or auction site characteristics. The same could be done with the time preference parameters (before and after the break point), as originally suggested by dHdVZ.

Many questions remain unanswered; the current contribution has merely scratched the surface. Whether or not the $2 \log \mathrm{n}$ rule truly holds is still to be decided. The previous study as well as this one have looked at two different auction sites, neither of which fit the model assumptions entirely. It is not clear yet which of the departures are more egregious. It seems likely that the data and data gathering methodology introduced here can be used in further fruitful research. Hopefully, this will be the case.

Appendix A (Java program used to gather page views data)

```
import java.awt.event.ActionEvent;
import java.awt.event.ActionListener;
import java.io.*;
import java.net.*;
import java.text.SimpleDateFormat;
import java.util.Date;
import java.util.GregorianCalendar;
import java.util.Locale;
import java.util.TimeZone;
import java.util.Timer;
import java.util.TimerTask;

	// “XXX” is the identifier for a particular auction (a three-letter tag was used for the
	study)

public class XXX
{
	public int numTimesAccessed = 0;
	public int timerIntervalSec = 300;
	public long lastUpdateTime = 0;

	public final SimpleDateFormat TIME_DATE_FORMAT =
	 new SimpleDateFormat("M/d/yyyy kk:mm:ss");

	// Use the same time-zone as eBay
	public GregorianCalendar cal = new GregorianCalendar(TimeZone.getTimeZone("PST"),
	Locale.US);

	public String myURL, filename;
	public PrintWriter outfile;
	public Timer timer;

	public ALJ(String url, String filename) {
		TIME_DATE_FORMAT.setCalendar(cal);
		this.myURL = url;
		this.filename = filename;
		timer = new Timer();
	}


	public void start() {
		try {
			outfile = new PrintWriter( new FileWriter(filename) );
		} catch(Exception e) {
			e.printStackTrace();
		}

		timer.schedule(new RetrieveData(), 0, 1000 * this.timerIntervalSec);
	}

	class RetrieveData extends TimerTask {
		public void run() {
			retrieveData();
		}
	}

	public void retrieveData() {
		while(true) {
		try {
			StringBuffer sb=new StringBuffer();
			URL url = new URL(myURL);
			URLConnection ucon = url.openConnection();
			DataInputStream dip=new DataInputStream(ucon.getInputStream());
			while(true) {
				String line = dip.readLine();
				if(line == null) break;
```

```
                        sb.append(line);
                }
                String s = sb.toString();
                int counterIndex = s.indexOf("vicount");
                if(counterIndex < 0) {
                        System.err.println("No counter found");
                } else {
                        String timestamp = TIME_DATE_FORMAT.format(new Date());
                        String parsedString = s.substring(counterIndex+9,counterIndex+13);
                        numTimesAccessed++;
                        System.out.println(timestamp + "\t" + parsedString + "\t" +
                        numTimesAccessed);
                        if(!parsedString.contains("--</s")) {
                                outfile.println(timestamp + "\t" + parsedString +"\t" +
                                numTimesAccessed);
                                outfile.flush();
                                        break;
                        }

                }
        }
        catch(Exception e) {
                e.printStackTrace();
        }
        }
    }

        // The url string for the counter can be is a combination of the counter script
        url, “http://cgi.ebay.com/ws/eBayISAPI.dll?VICounter&item=”, and the auction’s ID
        number, “250509413207”, which can be found on the auction’s web page.

    public static void main(String args[]) {
        String url = "http://cgi.ebay.com/ws/eBayISAPI.dll?VICounter&item=250509413207";
        String outfile = "C:\\Thesis\\XXX.txt";
        if(args.length < 2) {
                System.out.println("Requires two aruments: URL OutputFile; using
                defaults");
        } else {
                url = args[0];
                outfile = args[1];
        }

        System.out.println("Attempting to retrieve data from " + url);
        System.out.println("Saving to: " + outfile);
        XXX test = new XXX(url, outfile);
        test.start();
    }
}
```

Appendix B (Figures)

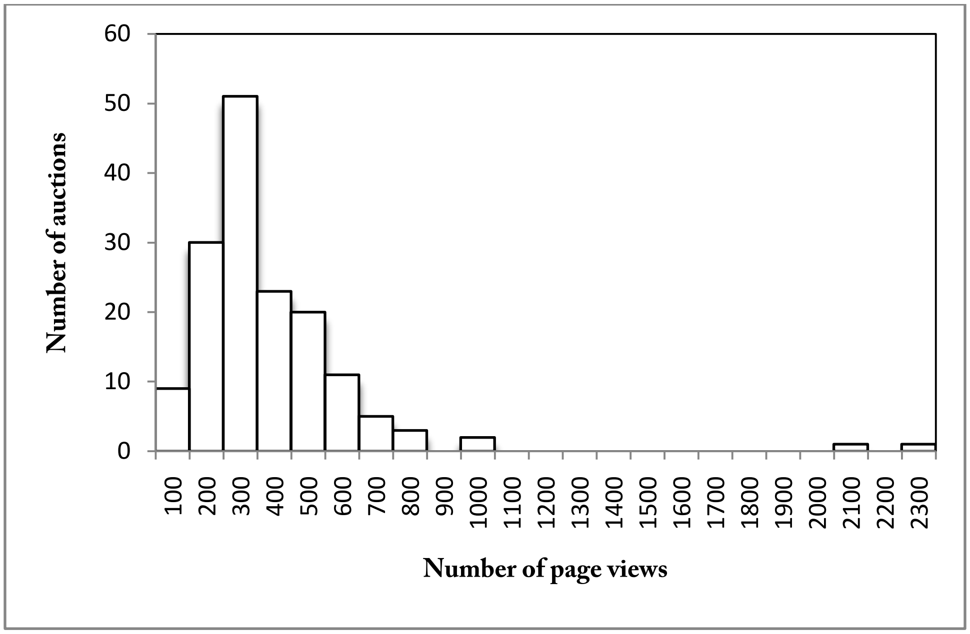

Figure B-1: Histogram of the number of page views (eBay sample); auctions with over five bidders

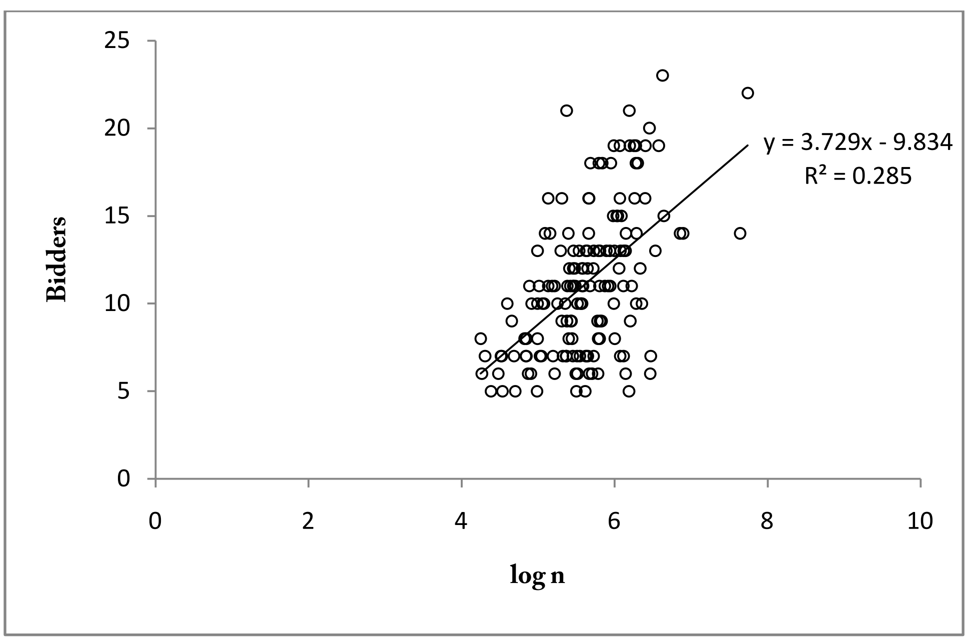

Figure B-2.1: Regression output of (9); active bidders on $\log n$, with intercept

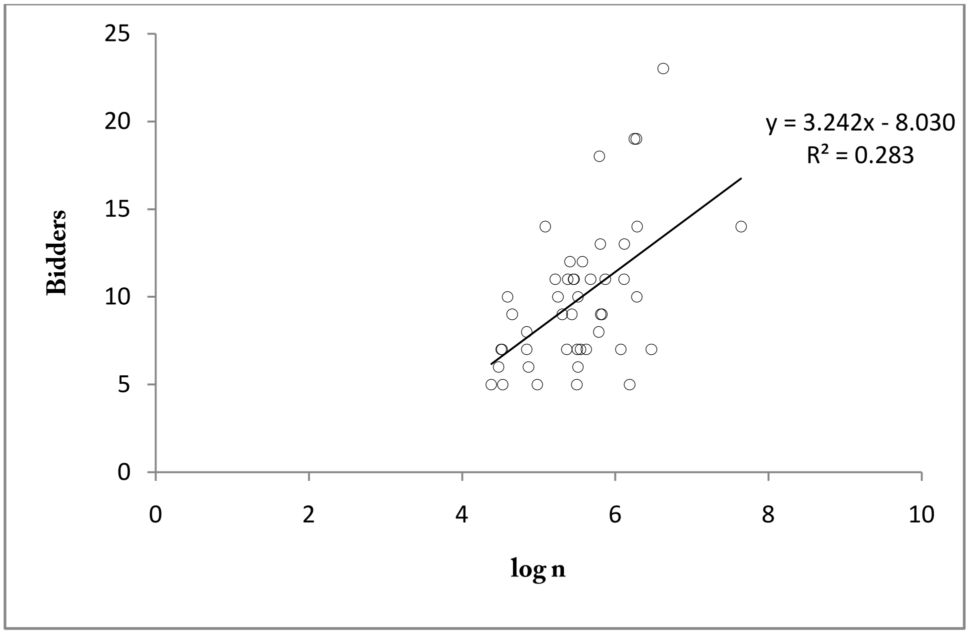

Figure B-2.2: Regression output of (10); non-sniping active bidders on $\log n$, with intercept

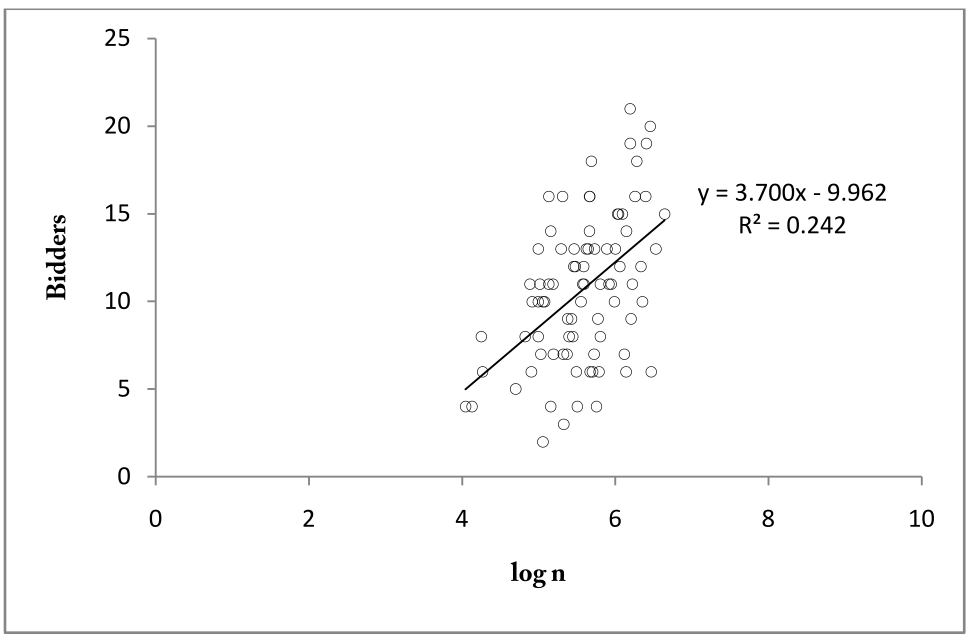

Figure B-2.3: Regression output of (10); auctions with (possible) sniping, with intercept

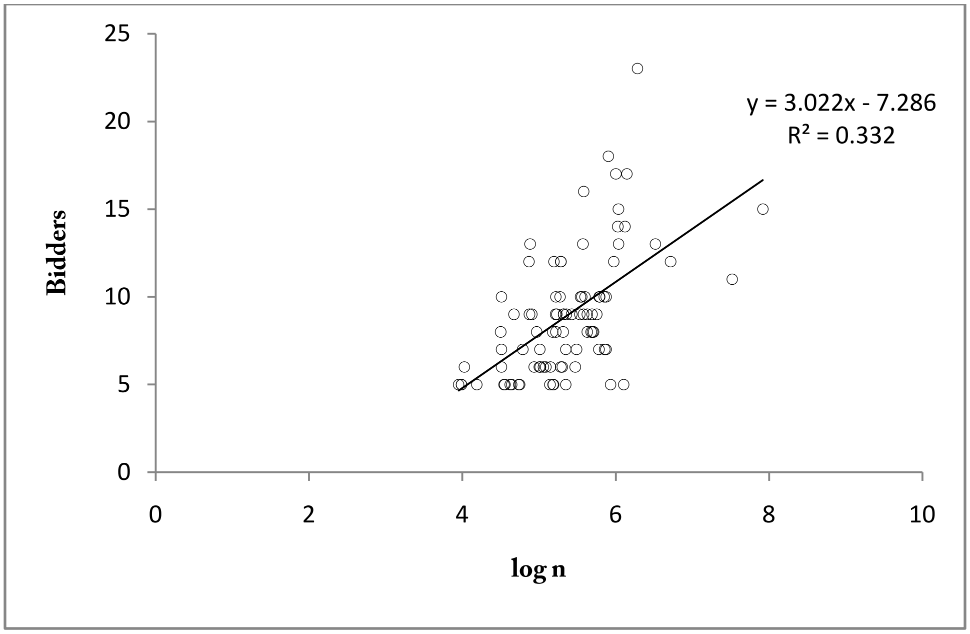

Figure B-2.4: Regression output of (10); viewers and bidders before the break-point

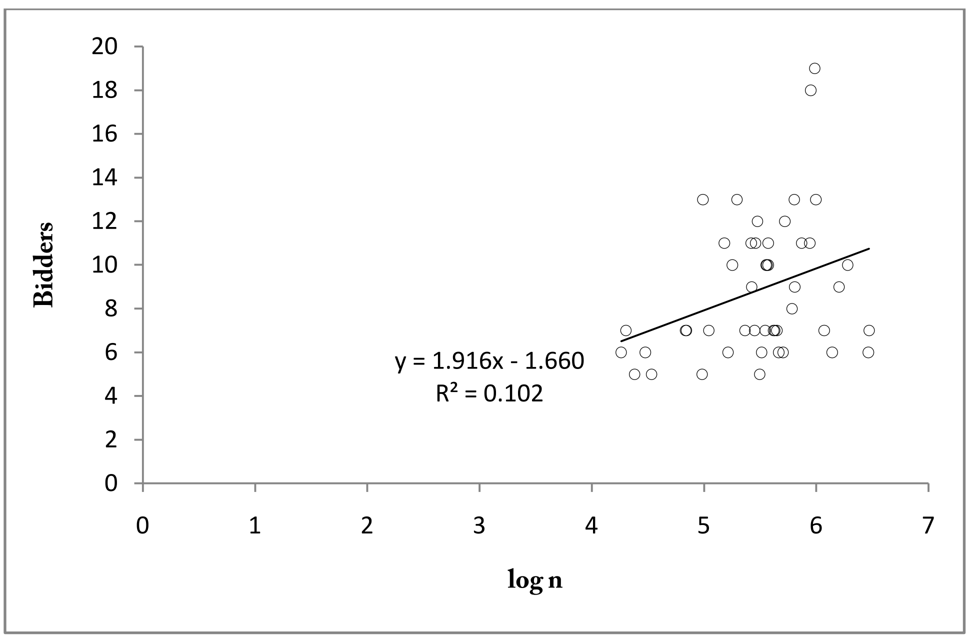

Figure B-2.5: Regression output of (10); starting price limited to 10% of final bid, no sniping (10 seconds)

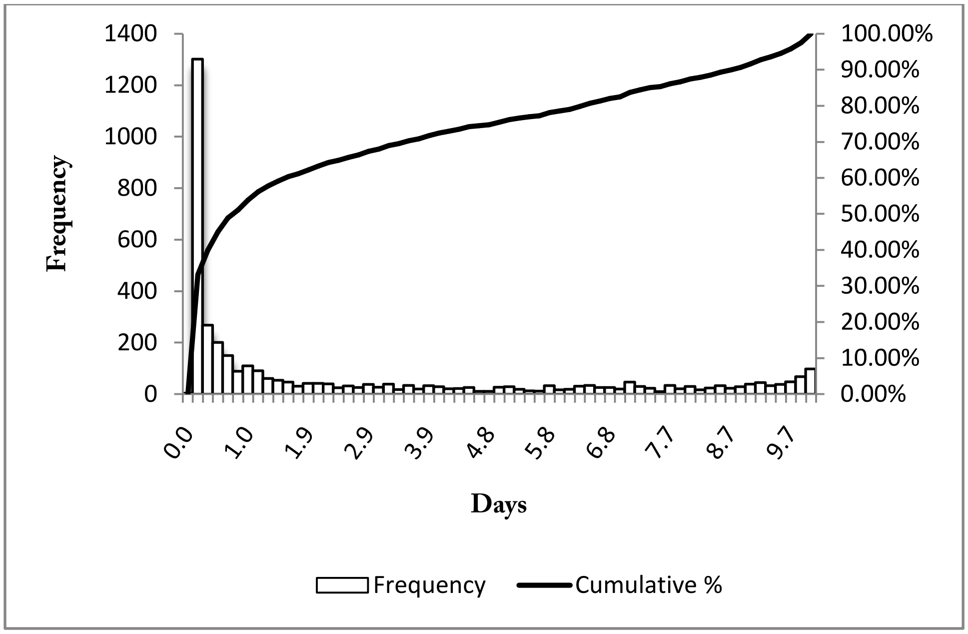

Figure B-3.1: Histogram and cumulative percentage graph of bid times; ten-day period

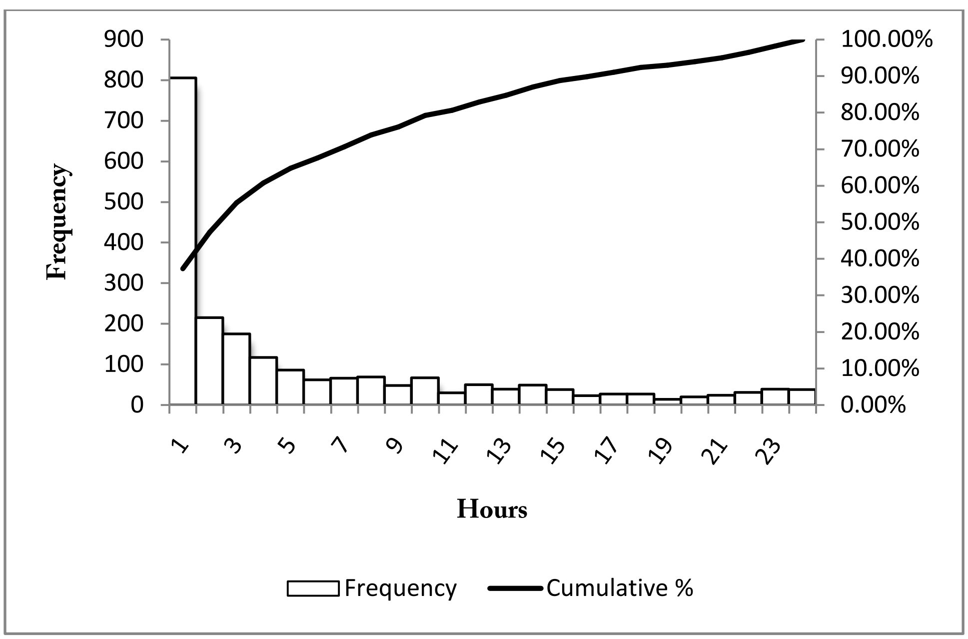

Figure B-3.2: Histogram and cumulative percentage graph of bid times; final 24 hours

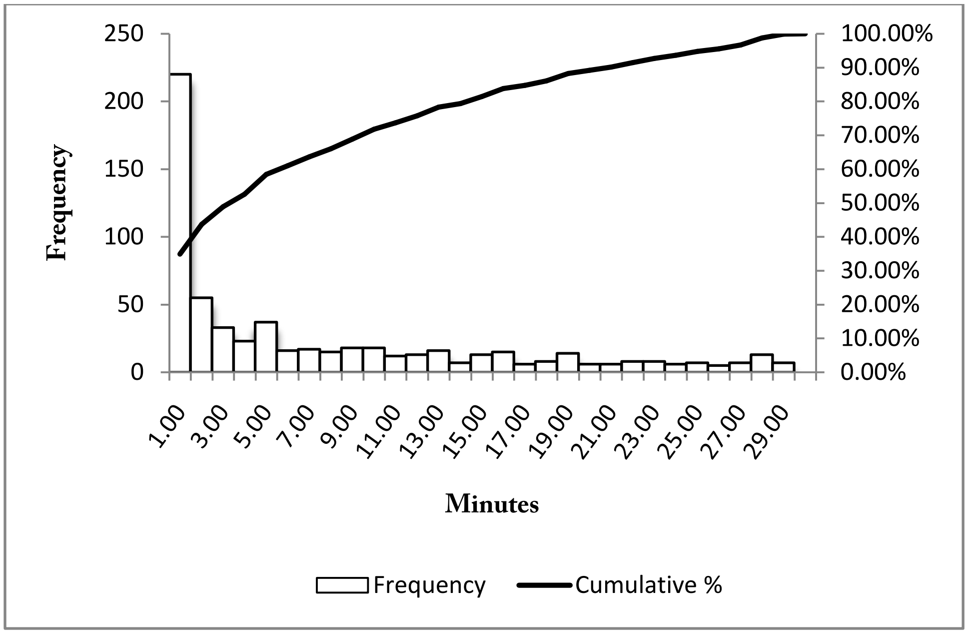

Figure B-3.3: Histogram and cumulative percentage graph of bid times; final thirty minutes

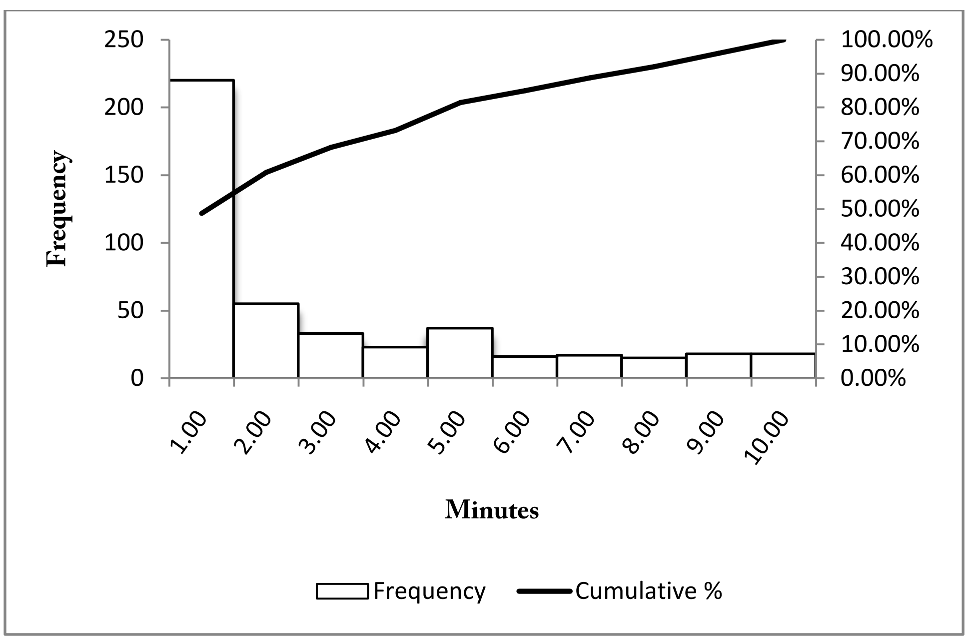

Figure B-3.4: Histogram and cumulative percentage graph of bid times; final 10 minutes

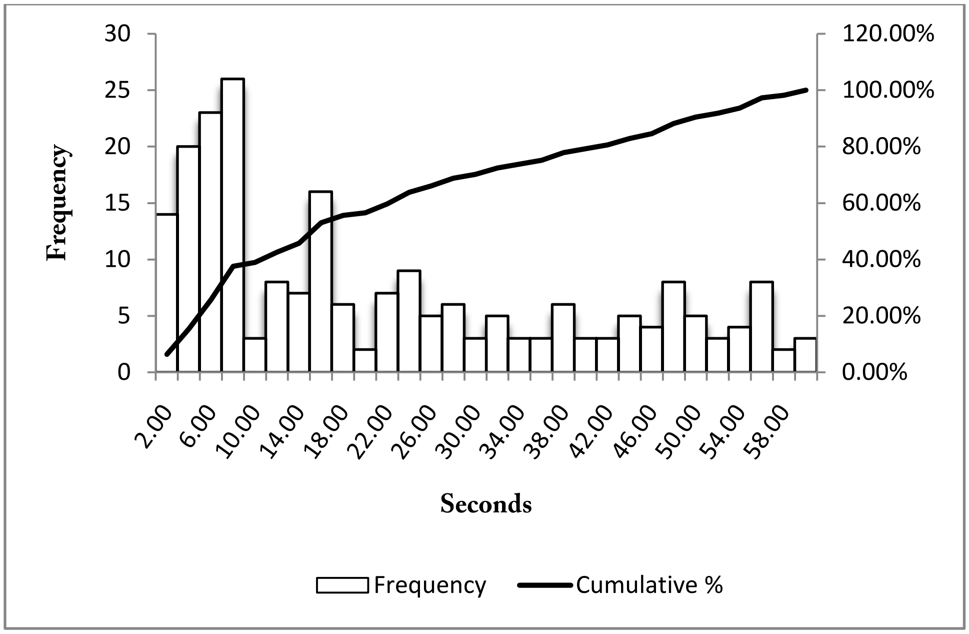

Figure B-3.5: Histogram and cumulative percentage graph of bid times; final 60 seconds

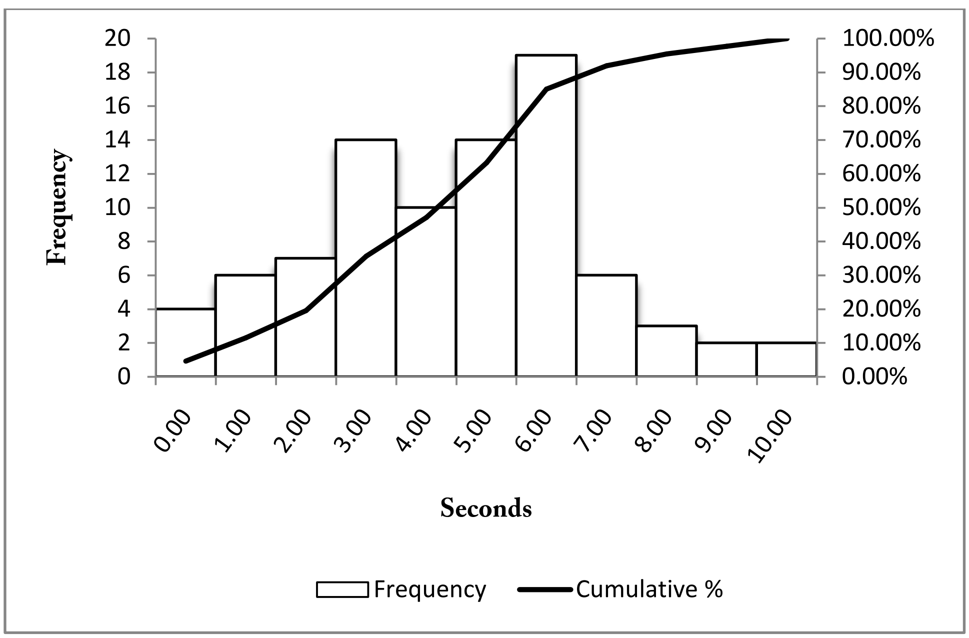

Figure B-3.6: Histogram and cumulative percentage graph of bid times; final 10 seconds

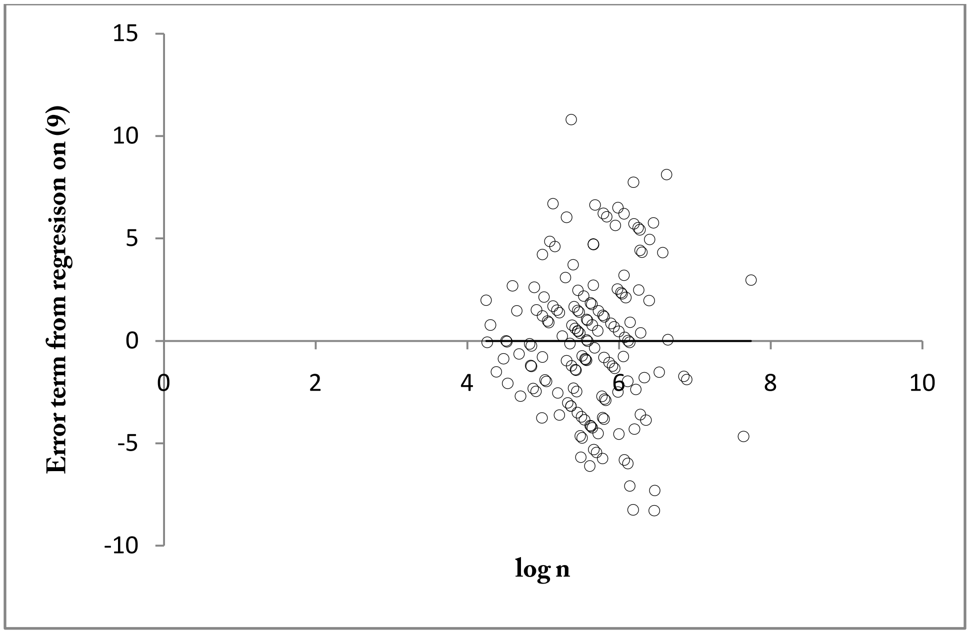

Figure B-4.1: Plot of $\log n$ and error term from the regression on (9)

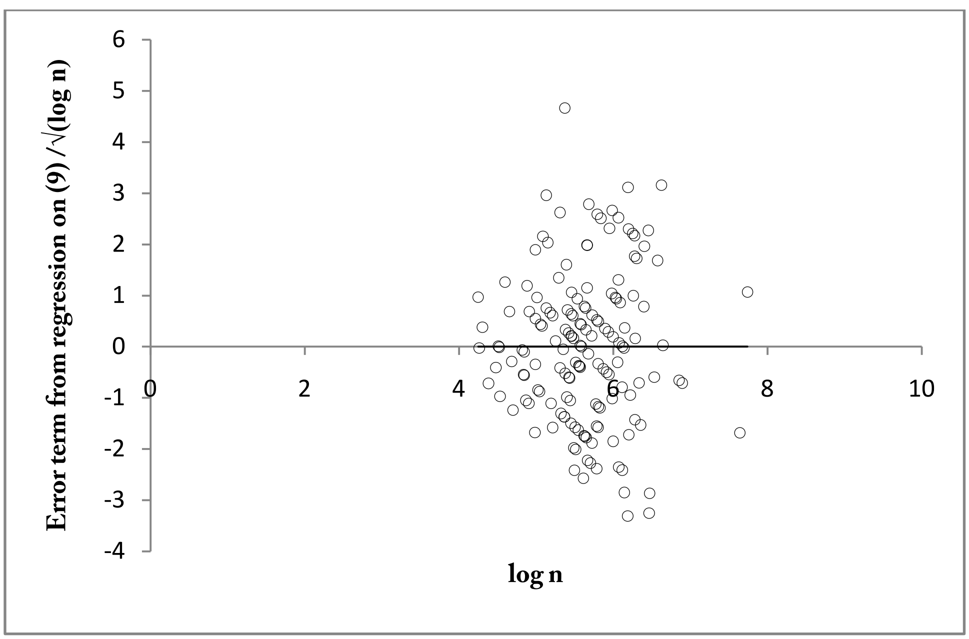

Figure B-4.2: Plot of $\log n$ and error term from the regression on (9), divided by $\sqrt{\log n}$

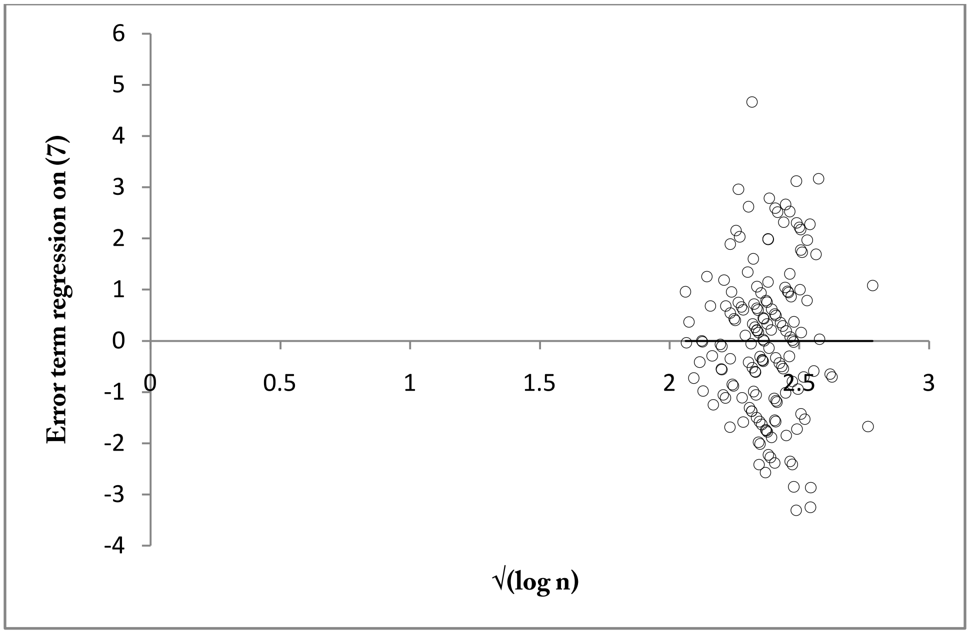

Figure B-4.3: Plot of $\sqrt{\log n}$ and error term from the regression on (7)

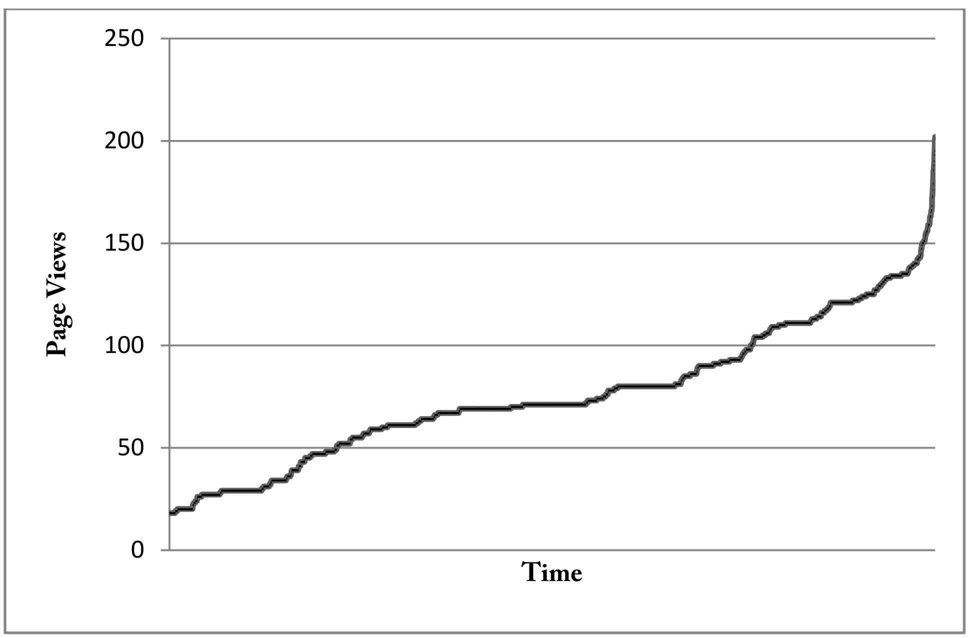

Figure B-5.1: Page views time series with a problematic break point calculation

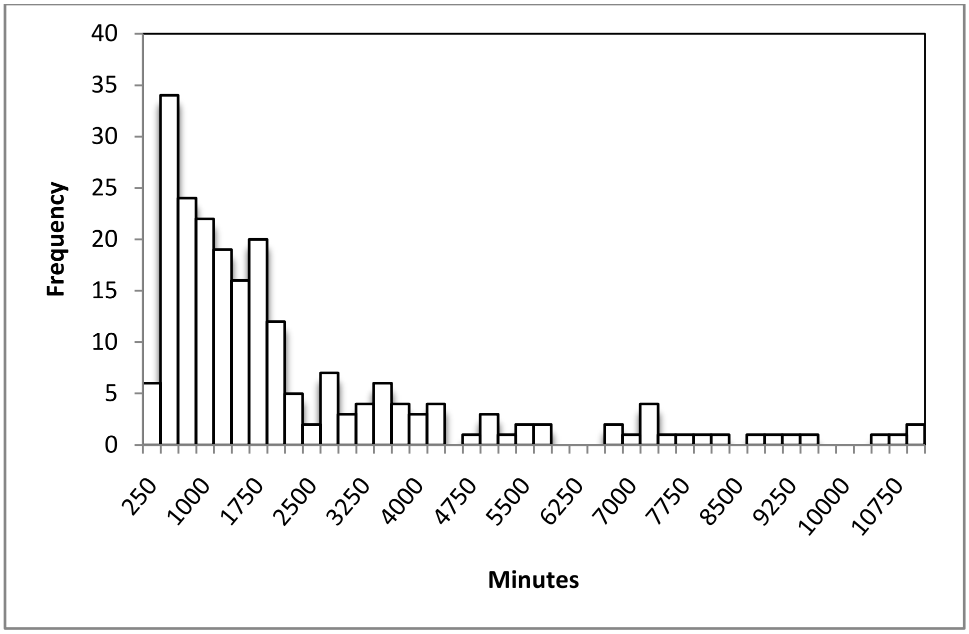

Figure B-5.2: Page views times series break point distribution, full dataset

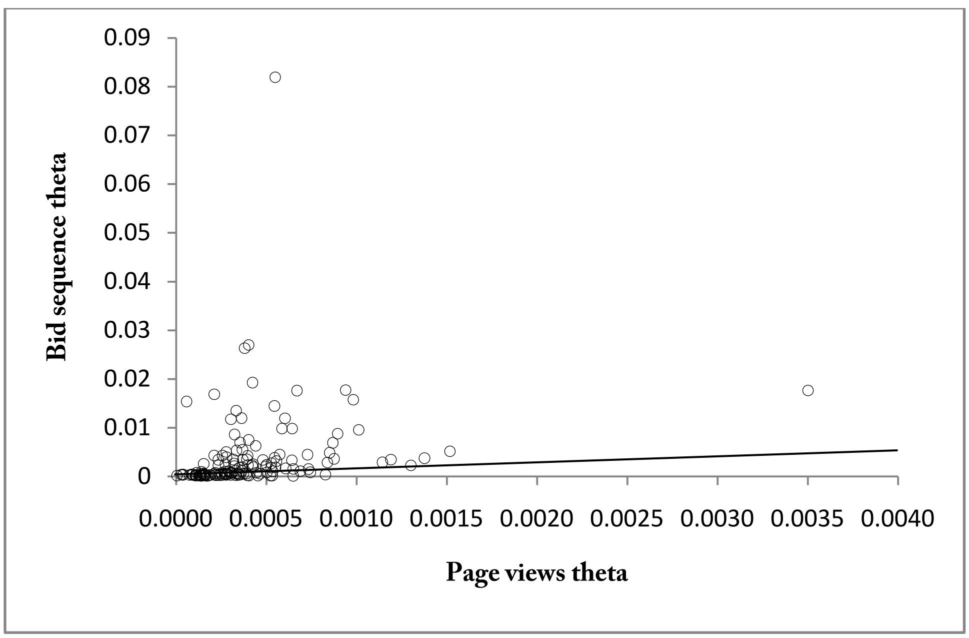

Figure B-6.1: Plot (unaltered) of θ_i estimated from the page views vs. those from the bid sequence

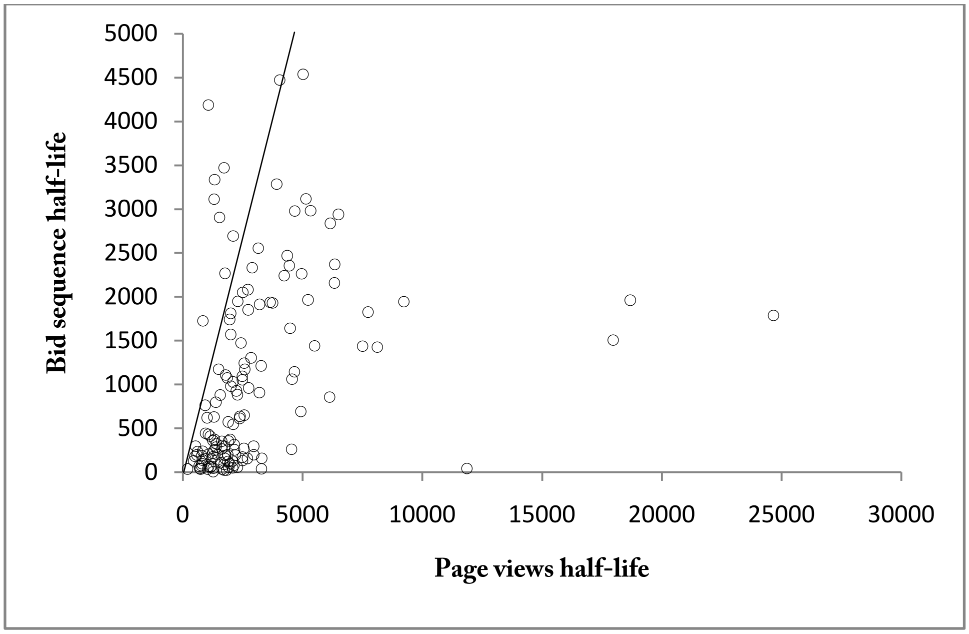

Figure B-6.2: Plot (unaltered) of half-life times estimated from the page views vs. those from the bid sequence

References

P. Bajari and A. Hortaçsu, *Winner's curse reserve prices and endogenous entry: Empirical insights from ebay auctions*, RAND Journal of Economics **34** (2003), 329-355

J. I. Bulow and P. D. Klemperer, *Auctions vs. negotiations*, American Economic Review **86** (1996), 180-194

A. Colin Cameron and P. K. Trivedi, *Regression analysis of count data*, Cambridge University Press (1998)

J. Engelberg and J. Williams, *eBay's proxy system: a license to shill*, Working paper, Kellogg School of Management (November 2006), Available at SSRN: http://ssrn.com/abstract=616782

A. Gelman and J. Hill, *Data analysis using regression and multilevel/hierarchical models*, Cambridge University Press (2007)

L. de Haan, C. G. de Vries, and C. Zhou, *The extent of internet auction markets*, (December 2007), Available at SSRN: http://ssrn.com/abstract=1123402

S. C. Hayne, C. A. P. Smith, and L. R. Vijayasarathy, *Sniping on eBay: a cultural analysis*, Working paper (2003), University of Colorado

R. J. Kauffman and C. A. Wood, *Running up the bid: detecting, predicting, and preventing reserve price shilling in online auctions*, Working paper (February 2003), available at http://misrc.umn.edu/workingpapers/fullPapers/2003/0304_022003.pdf

K. M. Lusht, *A comparison of prices brought by English auctions and private negotiations*, Real Estate Economics **24** (1996) No.4, 517-530

F. Menezes and P. Monteiro, *An introduction to auction theory*, Oxford University Press (2008)

A. Ockenfels and A. E. Roth, *The timing of bids in internet auctions: market design, bidder behavior, and artificial agents*, Artificial Intelligence Magazine (2002), 79-87

A. Ockenfels and A. E. Roth, *Late and multiple bidding in second price internet auctions: Theory and evidence concerning different rules for ending an auction*, Games and Economic Behavior **55** (2006), 297-320

A. Ockenfels, D. Reiley, A. Sadrieh, *Online auctions,* NBER Working paper No. W12785 (December 2006), Available at SSRN: http://ssrn.com/abstract=953208

Y. H. Park and E. T. Bradlow, *An integrated model for bidding behavior in internet auctions: Whether, who, when, and how much,* Journal of Marketing Research **42** (2005), no. 4, 470-482

S. I. Resnick, *Extreme values, regular variation, and point processes*, Springer-Verlag (1987)

G. Shmueli, R. P. Russo, and W. Jank, *Modeling bid arrivals in online auctions*, Working paper (2004), Robert H. Smith School of Business

U. Song, *Nonparametric estimation of an eBay auction model with an unknown number of bidders*, Working paper (2004), University of British Columbia, see http://faculty.arts.ubc.ca/usong/eBay.pdf

W. Vickrey, *Counterspeculation, auctions and competitive sealed tenders*, Journal of Finance **16** (1961), 8-37

H. White, *A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity*. Econometrica **48** (1980), no. 4, 817–838

L. Wintr, *Some Evidence on Late Bidding in Ebay Auctions*, Economic Inquiry **46** (2008), 369-379