# 1

# INTRODUCTION

Online auctions have received an extreme surge of popularity in recent years. Websites such as *eBay.com*, *uBid.com*, or *Swoopo.com* are marketplaces where buyers and sellers meet to exchange goods or information. Online auction platforms are different from fixed-price retail environments such as *Amazon.com* since transactions are negotiated between buyers and sellers. The popularity of online auctions stems from a variety of reasons. First, online auction websites are constantly available, so sellers can post items at any time and bidders can place bids day or night. Items are typically listed for several days, giving purchasers time to search, decide, and bid. Second, online auctions face virtually no geographical constraints and individuals in one location can participate in an auction that takes place in a completely different location of the world. The vast geographical reach also contributes to the variety of products offered for sale—both new and used. Third, online auctions also provide entertainment, as they engage participants in a competitive environment. In fact, the social interactions during online auctions have sometimes been compared to gambling, where bidders wait in anticipation to win and often react emotionally to being outbid in the final moments of the auction.

Online auctions are relatively new. By an "online auction" we refer to a Web-based auction, where transactions take place on an Internet portal. However, even before the advent of Internet auctions as we know them today, auctions were held electronically via email messages, discussion groups, and newsgroups. David Lucking-Reiley (2000) describes the newsgroup `rec.games.deckmaster` where

Internet users started trading "Magic" cards (related to the game *Magic: the Gathering*) as early as 1995. He writes

> By the spring of 1995, nearly 6,000 messages were being posted each week, making `rec.games.tradingcards.marketplace` the highest-volume newsgroup on the Internet. Approximately 90 percent of the 26,000 messages per month were devoted to the trading of Magic cards, with the remaining 10 percent devoted to the trading of cards from other games.

Lucking-Reiley (2000) presents a brief history of the development of Internet auctions and also provides a survey of the existing online auction portals as of 1998. The first online auction websites, launched in 1995, went by the names of *Onsale* and *eBay*. Onsale (today *Egghead*) later sold its auction service to Yahoo! and moved to fixed-price retailing. Yahoo! and Amazon each launched their own online auction services in 1999. Within 10 years or so, both shut down their auction services and now focus exclusively on fixed-price operations. (At the time of writing, Yahoo! maintains online auctions in Hong Kong, Taiwan, and Japan.) Thus, from 1995 until today (i.e., 2010) the consumer-to-consumer online auction marketplace has followed the pattern of eCommerce in general: An initial mushrooming of online auction websites was followed by a strong period of consolidations, out of which developed the prominent auction sites that we know today: eBay, uBid, or Swoopo (for general merchandize), SaffronArt (for Indian art), or Prosper (for peer-to-peer lending).

Empirical research of online auctions is booming. In fact, it has been booming much more compared to traditional, brick-and-mortar auctions. It is only fair to ask the question: "Why has data-driven research of online auctions become so much more popular compared to that of traditional auctions?" We believe the answer is simple and can be captured in one word: data! In fact, the public access to ongoing and past auction transactions online has opened new opportunities for empirical researchers to study the behavior of buyers and sellers. Moreover, theoretical results, founded in economics and derived for the offline, brick-and-mortar auction, have often proven not to hold in the online environment. Possible reasons that differentiate online auctions from their offline counterparts are the worldwide reach of the Internet, anonymity of its users, virtually unlimited resources, constant availability, and continuous change.

In one of the earliest examinations of online auctions (e.g., Lucking-Reiley et al., 2000), empirical economists found that bidding behavior, particularly on eBay, often diverges significantly from what classical auction theory predicts. Since then, there has been a surge in empirical analysis using online auction data in the fields of information systems, marketing, computer science, statistics, and related areas. Studies have examined bidding behavior in the online environment from multiple different angles: identification and quantification of new bidding behavior and phenomena, such as bid sniping (Roth and Ockenfels, 2002) and bid shilling (Kauffman and Wood, 2005); creation of a taxonomy of bidder types (Bapna et al., 2004); development of descriptive probabilistic models to capture bidding and bidder activity (Shmueli et al., 2007; Russo et al., 2008), as well as bidder behavior in terms of bid timing and amount (Borle et al., 2006; Park and Bradlow, 2005); another stream of research focuses on the price evolution during an online auction. Related to this are studies on price dynamics

(Wang et al., 2008a,b; Bapna et al., 2008b; Dass and Reddy, 2008; Reddy and Dass, 2006; Jank and Shmueli, 2006; Hyde et al., 2008; Jank et al., 2008b, 2009) and the development of novel models for dynamically forecasting auction prices (Wang et al., 2008a; Jank and Zhang, 2009a,b; Zhang et al., 2010; Jank and Shmueli, 2010; Jank et al., 2006; Dass et al., 2009). Further topics of research are quantifying economic value such as consumer surplus in eBay (Bapna et al., 2008a), and more recently, online auction data are also being used for studying bidder and seller relationships in the form of networks (Yao and Mela, 2007; Dass and Reddy, 2008; Jank and Yahav, 2010), or competition between products, between auction formats, and even between auction platforms (Haruvy et al., 2008; Hyde et al., 2006; Jank and Shmueli, 2007; Haruvy and Popkowski Leszczyc, 2009). All this illustrates that empirical research of online auctions is thriving.

## 1.1 ONLINE AUCTIONS AND ELECTRONIC COMMERCE

Online auctions are part of a broader trend of doing business online, often referred to as *electronic commerce*, or *eCommerce*. eCommerce is often associated with any form of transaction originating on the Web. eCommerce has had a huge impact on the way we live today compared to a decade ago: It has transformed the economy, eliminated borders, opened doors to innovations that were unthinkable just a few years ago, and created new ways in which consumers and businesses interact. Although many predicted the death of eCommerce with the "burst of the Internet bubble" in the late 1990s, eCommerce is thriving more than ever. eCommerce transactions include buying, selling, or investing online. Examples are shopping at online retailers such as Amazon.com or participating in online auctions such as eBay.com; buying or selling used items through websites such as Craigslist.com; using Internet advertising (e.g., sponsored ads by Google, Yahoo!, and Microsoft); reserving and purchasing tickets online (e.g., for travel or movies); posting and downloading music, video, and other online content; postings opinions or ratings about products on websites such as Epinions or Amazon; requesting or providing services via online marketplaces or auctions (e.g., Amazon Mechanical Turk or eLance); and many more.

Empirical eCommerce research covers many topics, ranging from very broad to very specific questions. Examples of rather specific research questions cover topics such as the impact of online used goods markets on sales of CDs and DVDs (Telang and Smith, 2008); the evolution of open source software (Stewart et al., 2006); the optimality of online price dispersion in the software industry (Ghose and Sundararajan, 2006); the efficient allocation of inventory in Internet advertising (Agarwal, 2008); the optimization of advertisers' bidding strategies (Matas and Schamroth, 2008); the entry and exit of Internet firms (Kauffman and Wang, 2008); the geographical impact of online sales leads (Jank and Kannan, 2008); the efficiency and effectiveness of virtual stock markets (Spann and Skiera, 2003; Foutz and Jank, 2009); or the impact of online encyclopedia Wikipedia (Warren et al., 2008).

Broad research questions include issues of privacy and confidentiality of eCommerce transactions (Fienberg, 2006, 2008) and other issues related to mining Internet transactions (Banks and Said, 2006), modeling clickstream data (Goldfarb and Lu,

2006), and understanding time-varying relationships in eCommerce data (Overby and Konsynski, 2008). They also include questions on how online experiences advance our understanding of the offline world (Forman and Goldfarb, 2008); the economic impact of user-generated online content (Ghose, 2008); challenges in collecting, validating, and analyzing large-scale eCommerce data (Bapna et al., 2006) or conducting randomized experiments online (Van der Heijden and Böckenholt, 2008); as well as questions on how to assess the causal effect of marketing interventions (Rubin and Waterman, 2006; Mithas et al., 2006) and the effect of social networks and word of mouth (Hill et al., 2006; Dellarocas and Narayan, 2006).

Internet advertising is another area where empirical research is growing, but currently more so inside of companies and to a lesser extent in academia. Companies such as Google, Yahoo!, and Microsoft study the behavior of online advertisers using massive data sets of bidding and bidding outcomes to more efficiently allocate inventory (e.g., ad placement) (Agarwal, 2008). Online advertisers and companies that provide services to advertisers also examine bid data. They study relationships between bidding and profit (or other measures of success) for the purpose of optimizing advertisers' bidding strategies (Matas and Schamroth, 2008).

Another active and growing area of empirical research is that of *prediction markets*, also known as "information markets," "idea markets," or "betting exchanges." Prediction markets are mechanisms used to aggregate the *wisdom of crowds* (Surowiecki, 2005) from online communities to forecast outcomes of future events and they have seen many interesting applications, from forecasting economic trends to natural disasters to elections to movie box-office sales. While several empirical studies (Spann and Skiera, 2003; Forsythe et al., 1999; Pennock et al., 2001) report on the accuracy of final trading prices to provide forecasts, there exists evidence that prediction markets are not fully efficient, which brings up interesting new statistical challenges (Foutz and Jank, 2009).

There are many similarities between the statistical challenges that arise in the empirical analysis of online auctions and that of eCommerce in general. Next, we discuss some of these challenges in the context of online auctions; for more on the aspect of eCommerce research, see, for example, Jank et al. (2008a) or Jank and Shmueli (2008a).

## 1.2   ONLINE AUCTIONS AND STATISTICAL CHALLENGES

A key reason for the booming of empirical online auctions research is the availability of data: lots and lots of data! However, while data open the door to investigating new types of research questions, they also bring up new challenges. Some of these challenges are related to data volume, while others reflect the new structure of Web data. Both issues pose serious challenges for the empirical researcher.

In this book, we offer methods for handling and modeling the unique data structure that arises in online auction Web data. One major aspect is the *combination of temporal and cross-sectional* information. Online auctions (e.g., eBay) are a point in case. Online auctions feature two fundamentally different types of data: the bid history and

the auction description. The bid history lists the sequence of bids placed over time and as such can be considered a time series. In contrast, the auction description (e.g., product information, information about the seller, and the auction format) does not change over the course of the auction and therefore is cross-sectional information. The analysis of combined temporal and cross-sectional data poses challenges because most statistical methods are geared only toward one type of data. Moreover, while methods for panel data can address some of these challenges, these methods typically assume that events arrive at equally spaced time intervals, which is not at all the case for online auction data. In fact, Web-based temporal data that are user-generated create nonstandard time series, where events are not equally spaced. In that sense, such temporal information is better described as a process. Because of the dynamic nature of the Web environment, many processes exhibit *dynamics* that change over the course of the process. On eBay, for instance, prices speed up early, then slow down later, only to speed up again toward the auction end. Classical statistical methods are not geared toward capturing the change in process dynamics and toward teasing out similarities (and differences) across thousands (or even millions) of online processes.

Another challenge related to the nature of online auction data is capturing *competition between auctions*. Consider again the example of eBay auctions. On any given day, there exist tens of thousands of identical (or similar) products being auctioned that all compete for the same bidders. For instance, during the time of writing, a simple search under the keywords "Apple iPod" reveals over 10,000 available auctions, all of which vie for the attention of the interested bidder. While not all of these 10,000 auctions may sell an identical product, some may be more similar (in terms of product characteristics) than others. Moreover, even among identical products, not all auctions will be equally attractive to the bidder due to differences in sellers' perceived trustworthiness or differences in auction format. For instance, to bidders that seek immediate satisfaction, auctions that are 5 days away from completion may be less attractive than auctions that end in the next 5 minutes. Modeling differences in product similarity and their impact on bidders' choices is challenging (Jank and Shmueli, 2007). Similarly, understanding the effect of misaligned (i.e., different starting times, different ending times, different durations) auctions on bidding decisions is equally challenging (Hyde et al., 2006) and solutions are not readily available in classical statistical tools. For a more general overview of challenges associated with auction competition, see Haruvy et al. (2008).

Another challenge to statistical modeling is the existence of *user networks* and their impact on transaction outcomes. Networks have become an increasingly important component of the online world, particularly in the "new web," Web 2.0, and its network-fostering enterprises such as Facebook, MySpace, and LinkedIn. Networks also exist in other places (although less obviously) and impact transaction outcomes. On eBay, for example, buyers and sellers form networks by repeatedly transacting with one another. This raises the question about the mobility and characteristics of networks across different marketplaces and their impact on the outcome of eCommerce transactions. Answers to these questions are not obvious and require new methodological tools to characterize networks and capture their impact on the online marketplace.

## 1.3 A STATISTICAL APPROACH TO ONLINE AUCTION RESEARCH

In this book, we provide empirical methods for tackling the challenges described above. As with many books, we present both a description of the problem and potential solutions. It is important to remember that our main focus is statistical. That is, we discuss methods for collecting, exploring, and modeling online auction data. Our models are aimed at capturing empirical phenomena in the data, at gaining insights about bidders' and sellers' behavior, and at forecasting the outcome of online auctions. Our approach is pragmatic and data-driven in that we incorporate domain knowledge and auction theory in a less formalized fashion compared to typical exposés in the auction literature. We make extensive use of nonparametric methods and data-driven algorithms to avoid making overly restrictive assumptions (many of which are violated in the online auction context) and to allow for the necessary flexibility in this highly dynamic environment. The online setting creates new opportunities for observing human behavior and economic relationships "in action," and our goal is to provide tools that support the exploration, quantification, and modeling of such relationships.

We note that our work has been inspired by the early research of Lucking-Reiley et al. (2000) who, to the best of our knowledge, were the first to conduct empirical research in the context of online auctions. The fact that it took almost 9 years from the first version of their 1999 working paper until its publication in 2007 (Lucking-Reiley et al., 2007) shows the hesitation with which some of this empirical research was greeted in the community. We believe though that some of this hesitation has subsided by now.

## 1.4 THE STRUCTURE OF THIS BOOK

The order of the chapters in this book follows the chronology of empirical data analysis: from data collection, through data exploration, to modeling and forecasting.

We start in Chapter 2 by discussing different ways for obtaining online auction data. In addition to the standard methods of data purchasing or collaborating with Internet businesses, we describe the currently most popular method of data collection: Web crawling and Web services. These two technologies generate large amounts of rich, high-quality online auction data. We also discuss Web data collection from a statistical sampling point of view, noting the various issues that arise in drawing data samples from a website, and how the resulting samples relate to the population of interest.

Chapter 3 continues with the most important step in data analysis: data exploration. While the availability of huge amounts of data often tempts the researcher to directly jump into sophisticated models and methods, one of the main messages of this book is that it is of extreme importance to first understand one's data, and to explore the data for patterns and anomalies. Chapter 3 presents an array of data exploration methods and tools that support the special structures that arise in online auction data. One such structure is the unevenly spacing of time series (i.e., the bid histories) and their combination with cross-sectional information (i.e., auction details). Because many of the models presented in the subsequent chapters make use of an auction's price

evolution, we describe plots for displaying and exploring curves of the price and its dynamics. We also discuss curve clustering, which allows the researcher to segment auctions by their different price dynamics.

Another important facet is the concurrent nature of online auctions and their competition with other auctions. We present methods for visualizing the degree of auction concurrency as well as its context (e.g., collection period and data volume). We also discuss unusual data structures that can often be found in online auctions: semicontinuous data. These data are continuous but contain several "too-frequent" values. We describe where and how such semicontinuous data arise and propose methods for presenting and exploring them in Chapter 3.

The chapter continues with another prominent feature of online auction data: data hierarchies. Hierarchies arise due to the structure of online auction websites, where listings are often organized in the form categories, subcategories, and subsubcategories. This organization plays an important role in how bidders locate information and, ultimately, in how listings compete with one another.

Chapter 3 concludes with a discussion of exploratory tools for interactive visualization that allow the researcher to "dive" into the data and make multidimensional exploration easier and more powerful.

Chapter 4 discusses different statistical models for capturing relationships in auction data. We open with a more formal exposition of the price curve representation, which estimates the price process (or price evolution) during an ongoing auction. The price process captures much of the activity of individual bidders and also captures interactions among bidders, such as bidders competing with one another or changes in a bidder's bidding strategies as a result of the strategies of other bidders. Moreover, the price process allows us to measure all of this change in a very parsimonious matter—via the price dynamics. Chapter 4 hence starts out by discussing alternatives for capturing price dynamics and then continues to propose different models for price dynamics. In that context, we propose *functional regression models* that allow the researcher to link price dynamics with covariate information (such as information about the seller, the bidders, or the product). We then extend the discussion to *functional differential equation models* that capture the effect of the process itself in addition to covariate information.

We then discuss statistical models for auction competition. By competition, we mean many auctions that sell similar (i.e., substitute) products and hence vie for the same bidders. Modeling competition is complicated because it requires the definition of "similar items." We borrow ideas from spatial models to capture the similarity (or dissimilarity) of products in the associated feature space. But competition may be more complex. In fact, competition also arises from temporal concurrency: Auctions that are listed only a few minutes or hours apart from one another may show stronger competition compared to auctions that end on different days. Modeling temporal relationships is challenging since the auction arrival process is extremely uneven and hence requires a new definition of the traditional "time lag."

Chapter 4 continues with discussing models for bidder arrivals and bid arrivals in online auctions. Modeling the arrival of bids is not straightforward because online auctions are typically much longer compared to their brick-and-mortar counterparts

and hence they experience periods of little to no activity, followed by "bursts" of bidding. In fact, online auctions often experience "deadline effects" in that many bids are placed immediately before the auction closes. These different effects make the process deviate from standard stochastic models. We describe a family of stochastic models that adequately capture the empirically observed bid arrival process. We then tie these models to bidder arrival and bid placement strategies. Modeling the arrival of bidders (rather than bids) is even more challenging because while bids are observed, the entry (or exit) of bidders is unobservable.

Chapter 4 concludes with a discussion of auction networks. Networks have become omnipresent in our everyday lives, not the least because of the advent of social networking sites such as MySpace or Facebook. While auction networks is a rather new and unexplored concept, one can observe that links between certain pairs of buyers and sellers are stronger than others. In Chapter 4, we discuss some approaches for exploring such bidder–seller networks.

Finally, in Chapter 5 we discuss forecasting methods. We separated "forecasting" from "modeling" (in Chapter 4) because the process of developing a model (or a method) that can predict the future is typically different from retroactively building a model that can describe or explain an observed relationship.

Within the forecasting context, we consider three types of models, each adding an additional layer of information and complexity. First, we consider forecasting models that only use the information from within a given ongoing auction to forecast its final price. In other words, the first—and most basic—model only uses information that is available from within the auction to predict the outcome of that auction. The second model builds upon the first model and considers additional information about other simultaneous auctions. However, the information on outside auctions is not modeled explicitly. The last—and most powerful—model explicitly measures the effect of competing auctions and uses it to achieve better forecasts.

We conclude Chapter 5 by discussing useful applications of auction forecasting such as automated bidding decision rule systems that rely on auction forecasters.

## 1.5   DATA AND CODE AVAILABILITY

In the spirit of publicly (and freely) available information (and having experienced the tremendous value of rich data for conducting innovative research firsthand), we make many of the data sets described in the book available at `http://www.ModelingOnlineAuctions.com`. The website also includes computer code used for generating some of the results in this book. Readers are encouraged to use these resources and to contribute further data and code related to online auctions research.