Essays in empirical industrial organisation



Segye Shin

The Queen's College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity 2019

Acknowledgements

I am greatly indebted to Howard Smith, who has been a source of constant support throughout my time at Oxford and whose guidance and encouragement made this thesis possible.

Special thanks go to Oyvind Thomassen and Walter Beckert, who have been great examples to learn from throughout our time working together.

I also thank Abigail Adams, Mark Armstrong, Emmanuel Guerre, Meg Meyer, Andre Veiga and the seminar participants at Oxford for their very useful comments on various parts of this thesis.

Special thanks go to my parents, Mirae, Phil and Ethan for their love and support. My thanks also to all the wonderful people in OUDC and OADA.

I gratefully acknowledge financial support from Compass Lexecon for the MPhil portion of my degree and personal encouragement and support from my colleagues in Compass Lexecon London.

Contents

1 Repeated auctions with option of retaining incumbents: Case of the UK statutory audits market **4**

1.1 Introduction 4

1.2 Industry background and data 8

1.3 Model 15

1.4 Estimation 29

1.5 Results 37

1.6 Counterfactual experiment 41

1.7 Conclusion 46

2 Estimating the impact of endogenous seller entry on online marketplace prices **52**

2.1 Introduction 52

2.2 Data source and descriptive statistics 55

2.3 Empirical framework 62

2.4 Results 66

2.5 Conclusions 74

3 Impact of scale and competition on an online peer-to-peer platform **82**

3.1 Introduction 82

3.2 Description of the platform and data 86

3.3 Model 93

3.4 Estimation 102

3.5 Results 108

3.6 Counterfactual experiments and implications for the platform 114

3.7 Conclusion 123

Introduction

This thesis consists of three essays in the area of empirical industrial organisation, all of which relate to auctions and buyer-specific price determination.

Chapter 1: Repeated auctions with option of retaining incumbents: Case of the UK statutory audits market

Chapter 1 sets out a model of infinitely repeated auctions, where the ability of the winner of the auction to benefit from incumbency advantages in the following periods introduces a dynamic component to the auction equilibrium. The incumbency advantage comes in the form of the ability to make a take-or-leave offer to the buyer at the start of each period which allows both parties to avoid a costly auction process during that period. The equilibrium strategies in the auction game can be reduced to a set of 'incumbency valuation' parameters specific to each bidder, based on distributional assumptions that are similar to those already used in the empirical auction literature. This allows for a computationally feasible estimation procedure based on a two-stage approach widely used in other empirical studies of auctions.

The model is applied to the statutory audit industry in the UK, based on a dataset of prices compiled by UK regulators. The parameter estimates suggest that incumbency advantage is significant in this industry and that this results in lower rates of switching and higher levels of concentration. Counterfactual experiments suggest that policy interventions such as mandatory tendering would have the effect of lowering concentration marginally, but at the cost of increased prices as sellers bid less aggressively in

each auction. A simulation of market outcomes based on an assumed discount rate of 0.9 suggests that mandatory tendering with a maximum interval between tenders of 10 years, as has recently been introduced in the UK, would result in a slight decrease in the largest firm's share of 0.3%, largely to the benefit of smaller firms. At the same time, the expected auction prices increase by around 2%.

Chapter 2: Estimating the impact of endogenous seller entry on online marketplace prices

Chapter 2 examines the relationship between the number of bidders involved in the competition for a delivery task and the resulting prices from the bidding process on an online marketplace. Compared to the existing literature, the online marketplace setting examined in this study allows me to identify the variation in the number of competing sellers more directly, as well as include all transactions in the marketplace from a given period in a single price equation rather than restricting the analysis to a particularly homogenous and comparable set of products that are traded in the marketplace.

One of the main findings from the existing literature on online marketplace prices is that the participation decisions of buyers and sellers are endogenous, and that not controlling for this endogeneity can result in inconsistent estimates of the impact of competition on prices. I take a similar approach to the existing literature in using instrumental variables to address the endogeneity issue. Several unique characteristics of the online marketplace setting allow me to construct valid instruments that reflect the number of suitable alternative listings that were available to bidders at the time of bidding, which is shown to affect bidders' participation decisions in the expected directions. My results confirm that entry decisions of bidders in this setting are indeed endogenous, and that not accounting for the endogeneity can significantly underestimate the true impact of competition on prices.

Chapter 3: Impact of scale and competition on an online peer-to-peer platform

Chapter 3 draws its motivation from the findings from Chapter 2, which showed that the seller entry process to compete for individual buyer listings is influenced by other similar listings that are available on the platform as a whole. In comparison to Chapter 2 which used a reduced form approach to analyse prices on the online platform, this chapter sets out a structural model of the platform that explicitly considers the entry of sellers into individual listings. I construct a model of the matching process that determines the set of sellers that compete for each of the buyer listings available on the platform, and link this to an auction framework that determines the final prices as a function of the number of competing sellers.

The results show that the matching process for the platform exhibits increasing returns to scale. Holding other factors constant, increasing both the number of buyers and sellers makes it easier for the market participants to find suitable matches. This has an important policy implication for online platforms; under a free entry equilibrium for sellers, increasing the number of buyers using the platform brings benefits to the buyers in the form of lower prices and increased probability of finding a suitable seller. This is because the increase in the number of buyers results in additional sellers entering the platform, and the resulting increase in the overall scale of the platform makes the buyer-seller matching process more efficient. The implications of the model are in line with the theoretical literature on two-sided markets which emphasise the importance of scale and network effects.

Chapter 1

Repeated auctions with option of retaining incumbents: Case of the UK statutory audits market

1.1 Introduction

There are many examples of auctions in the real world that are repeated on an ongoing basis rather than being one-off in nature. For example, many procurement auctions are for the provision of services over a specific period, and the bidders involved in each auction can reasonably expect similar auctions to keep reoccurring for the foreseeable future. Repeated auctions can impact bidding behaviours if the outcome of one auction can influence the bidders' payoffs in future periods, introducing dynamic considerations to each bidder's bidding strategies. This can occur through various channels: bidders may face capacity constraints, such that winning one auction negatively affects their costs in future auctions (Jofre-Bonet and Pesendorfer 2003), or there may be incumbency advantages that lower the costs of the auction winner in future auctions (Froeb et al. 2009). Despite this, there are relatively few empirical studies of repeated auctions, especially infinitely repeated auctions.

This paper sets out an infinite horizon model where procurement auctions occur repeat-

edly, although not necessarily in every period. The winner of each auction becomes the incumbent seller for the next period, and can make a take-or-leave offer to the buyer which, if accepted, allows both parties to avoid a costly auction process for that period. Being an incumbent is therefore a valuable proposition for each seller, and this introduces a dynamic component to each seller's payoffs from winning the auction. In equilibrium, the dynamic effects of the incumbency advantage reduce each seller's bids below their short term costs and therefore lower the winning prices below the level that would be reached if each seller was behaving myopically.

Incorporating the dynamic effects into the estimation process for the seller's equilibrium bidding strategies can be computationally demanding. To address this, the estimation strategy makes use of the two-stage estimation approach set out in Guerre et al. (2000) for nonparametric estimation of first price auctions and applied to dynamic auctions in Jofre-Bonet and Pesendorfer (2000, 2003). The first stage of the estimation involves estimating the distribution of *pseudo* private values, which in this case equates to estimating the distribution of bids as in Jofre-Bonet and Pesendorfer (2003). The second stage then uses the estimated distributions of bids and an expression for each bidder's expected sum of future profits dependent on auction outcomes to infer the distribution of underlying costs. As explained in Guerre et al. (2000), the computational advantage of this approach is that it avoids the need to solve for the equilibrium bidding strategy at every step of the estimation process. This is particularly important in this case as the equilibrium strategies cannot be expressed explicitly and instead must be solved for using numerical methods.

The auction environment used in this paper is comparable to that of repeated auctions with incumbency cost advantage as set out in Froeb et al. (2009), although the source of the incumbency advantage in this paper is the ability to avoid a costly auction process rather than a direct auction cost advantage as assumed in Froeb et al. (2009). Following the logit auction distributional assumptions developed in Froeb et al. (1999) and also used in Froeb et al. (2009), I model the underlying costs of sellers as Gumbel-distributed with asymmetric location parameters varying by seller but a common scale

parameter. This ensures that the distributions of private costs and the distributions of bids for all bidders belong to the same family of Gumbel distributions and can each be represented by a set of location parameters, one for each seller. This in turn ensures that the second stage of the two-stage estimation process is exactly identified and can be solved by simple numerical methods. This paper is the first to empirically apply the repeated auction environment with incumbency advantages as set out in Froeb et al. (2009), which examined the repeated auction framework using simulations but did not apply the model empirically.

I estimate the parameters of the model based on data on the supply of statutory audits in the United Kingdom. The statutory audit industry is a particularly fitting setting for estimating models of infinitely repeated interactions because the industry features and the regulatory background mean that buyers and sellers must interact in a predictable and repeated way over extremely long periods of time. There are some examples of listed companies in the UK that have purchased statutory audit services from the same seller for over a century despite regulations that preclude long-term contracts and empirical evidence that companies that switch auditors tend to benefit from lower prices (Competition Commission 2013), which demonstrates both the long-term nature of the buyer-seller interactions and the value of incumbency for the sellers.

The industry is also of particular interest from a policy perspective. Competition authorities in the UK have recently completed detailed studies into the industry to examine perceived failings in the market such as high concentration and low rates of switching, and the financial regulator is currently in the process of rolling out a European-level policy intervention in the form of mandatory auditor tendering and switching for public companies.

The parameter estimates suggest that incumbency advantages in this industry have significant effects on market outcomes. The value of incumbency adds to the benefits of winning each tender to supply new audit customers, and has a downward impact on prices as all sellers bid more aggressively in the tender. Based on some reasonable discount rate assumptions, the parameter estimates suggest that sellers discount their

bids to 34% to 45% below their short-term costs. The degree of discounting is the largest for ex-ante lowest cost sellers, and the incumbency advantage therefore has the effect of magnifying existing asymmetries between sellers and increasing seller concentration. These findings hold some implications for the recently introduced regulations in the UK which require buyers to tender for a new auditor at least every ten years; the policy intervention is likely to marginally reduce market concentration, but at the cost of increasing auction prices by around 2%.

This paper contributes to the literature on dynamic auctions. Several studies in this literature concentrate on sequential procurement auctions where the outcomes of previous auctions affect the valuations of each bidder and therefore introduce a dynamic component to bidding incentives. As set out in Jofre-Bonet and Pesendorfer (2003), a common interpretation of this structure is that bidders are subject to capacity constraints and that winning one auction can reduce the available capacity for competing in subsequent auctions. This type of framework has also been extended to examine the impact of endogenous entry (Groeger 2014), subcontracting (Jeziorski and Krasnokutskaya 2016), information sharing (Asker et al. 2018), among others. Under this framework, the contract awarded in each auction is not necessarily the continuation of the contract awarded in the preceding auctions; instead, the set of sequential auctions are for distinct contracts which are linked by being subject to the same capacity constraints. Distinctly from this, Froeb et al. (2009) use a simulation approach to examine an infinitely repeated auction framework where the same buyer holds an auction in every discrete period to replace the successful bidder from the previous period. The winning bidder from the immediately preceding auction can therefore be thought of as the incumbent for the auction, and the dynamic aspect in this study arises from the fact that the incumbent potentially benefits from a cost advantage. The model presented in this paper has more in common with the model set out in Froeb et al. (2009), since the infinitely repeated auctions are renewals for provision of the same services, and the dynamic aspect arises from the benefits that the incumbent can expect to gain in future periods.

The findings from this paper also relate to the literature on incumbency advantages arising from switching costs (Klemperer 2005) and brand loyalty (Bronnenberg and Dubé 2017). A recent empirical contribution in this literature is Allen et al. (2019) which examines the incumbent market power arising from search costs and brand loyalty in the market for mortgages.

The two-stage estimation strategy used in the paper is equivalent to that set out in Guerre et al. (2000) for nonparametric estimation of static first-price auctions, and applied to other static auction settings (Campo et al. 2011, Elyakime et al. 1994). Jofre-Bonet and Pesendorfer (2000) applies this two-stage estimation strategy to dynamic auctions in a nonparametric setting, and Jofre-Bonet and Pesendorfer (2003) in a parametric setting. This paper is the first to apply this estimation strategy to a repeated auctions framework with incumbents as in Froeb et al. (2009).

This paper also follows a large empirical literature in accounting research that examine determinants of statutory audit prices. Dekeyser et al. (2019), Moon et al. (2019) and Zhang et al. (2018) are some recent examples.

1.2 Industry background and data

1.2.1 Market for statutory audits

Statutory audit is a mandatory review of a company's financial accounts by external auditors. Annual audit is a legal requirement for all public companies in the UK as well as private companies of a certain minimum size. The purpose of the audit is to ensure that the company's financial statements give a fair and accurate view of the company's current state to investors. The final output from the audit is a report by the auditor that indicates the auditor's opinions on whether the company's financial statements are fair and accurate, which is communicated to shareholders via the company's annual report.

Audits are carried out by accountancy firms registered with a Recognised Supervisory

Body (RSB) in the UK, such as the Institute of Chartered Accountants in England and Wales (ICAEW) or the Institute of Chartered Accountants of Scotland (ICAS). Despite the large number of acountancy firms that are qualified by this route to provide statutory audit services, audits of large companies are currently dominated by four accountancy firms collectively known as the 'Big Four': PricewaterhouseCoopers (PwC), Deloitte, KPMG and Ernst & Young (EY).

Due to financial regulation, accountancy firms cannot enter into long term contracts with companies for the supply of audit services. There are therefore no contractual constraints on companies preventing them from regularly switching to another auditor. In practice, however, companies typically reappoint the same accountancy firm by default, and switching is rare. This is despite the empirical evidence that companies can secure lower audit fees by switching auditors (Cameran et al. 2015). Some of the largest listed companies in the UK have stayed with the same auditor for many decades. An extreme example is that of Barclays plc, which recently switched its audit supplier to KPMG after having been audited by PwC since 1896, a tenure of around 120 years.

In cases where companies do decide against reappointing the incumbent auditor, the appointment of a new auditor is typically decided by a tender. A tender for a new auditor by a large listed company can potentially signal a significant change in market shares among auditors, and is a sufficiently rare occurrence to be well-publicised by financial news sources. In some cases companies may choose to hold a tender without firmly deciding to switch away from the incumbent auditor, and there are cases where the incumbent auditor from the previous year wins the competitive tender and continues to supply the company.

The nature of audits as a product means that prices vary widely and are specifically tailored to the company being audited. Total price of each audit can range from thousands of pounds for the smallest companies to tens of millions of pounds for the largest listed companies. Auditors are provided with information on the company and its financial accounts prior to a tender, and are also able to draw on experience of auditing the company or comparable companies to arrive at a highly tailored pricing

offer to each buyer.

1.2.2 Audit regulation

Until mid-2010s, the UK statutory audit market was not subject to any regulation requiring companies to tender for new auditors or switch auditors. This was in contrast to several other major economies that enforce mandated tendering (that is, requirement to hold a competitive tender at certain minimum intervals, but no requirement to switch auditors) or mandated rotation (that is, requirement to replace the incumbent auditor with a different auditor before a certain maximum tenure) for auditors of publicly listed companies.[1]

Competition and financial regulatory authorities in the UK have recently examined such policy interventions to address perceived failings in the market such as the high concentration and low rates of switching. The UK Financial Reporting Council started implementing such regulations in 2012, starting with changes to the UK Corporate Governance Code specifying that companies must hold a competitive tender at least every ten years or explain publicly why they are not doing so. The UK Competition Commission undertook a detailed market investigation starting from 2011, culminating in a report published in 2013 which also supported mandatory tendering regulation (Competition Commission 2013).

The latest implementation of regulations in the industry has resulted from changes in financial regulation at the European Union level. The European regulations now specify that companies must hold a competitive tender at least every ten years and switch auditors every twenty years (European Commission 2014). The change in regulations for the UK was announced in 2016 and are being implemented with a transitional arrangement in the UK, with the first of such enforced auditor changes due to start in 2020.

[1]For example, mandatory rotation of auditors for public companies has been enforced in Italy since the 1970s. Cameran et al. (2015) examine the impact of such policies on audit outcomes in Italy.

1.2.3 Data

The UK Competition Commission compiled a database of audit prices charged to UK-based public and private companies as a part of its market investigation between 2011 and 2013, and this forms the basis for the empirical analysis in this paper.[2] The database includes information on the identity of the company and the auditor, the prices charged for the audit, and a number of pertinent company characteristics which would be expected to influence the underlying costs of the audit. All data contained in the database is in the public domain, including information on audit prices which must be disclosed in company filings by all UK-based companies. However, the database is especially suited for examining audit pricing compared to other publicly available databases due to the fact that the Competition Commission worked jointly with all major UK auditors to review the database and ensure comparability of pricing measures across all companies.

The data consists of a panel dataset covering financial years 2000 to 2011, and therefore covers the period before the introduction of any mandatory audit firm rotation regulations in the UK. The data covers all companies that were either constituents of the FTSE 350 Index at any point during the sample period or constituents of the Top Track 100 during the period 2006 to 2011.[3] Companies are observed in the dataset for the entire period even if they fell into one of these two groupings for only a part of the sample period. The final dataset after cleaning contains 6,454 observations across 760 companies.

The database contains several measures of revenues for the accountancy firm supplying the statutory audit, including fees for services that are unrelated to the statutory audit itself.[4] For the purposes of estimation only the core audit fees are used as the measure of prices for the statutory audit.

[2]See Competition Commission (2013). All underlying data is publicly available and can be downloaded from the website of the UK Competition and Markets Authority.

[3]FTSE 350 includes the largest 350 companies listed on the London Stock Exchange main market by market capitalisation. Top Track 100 includes the 100 largest private companies in the UK by revenues.

[4]Reporting of such fees in company filings is part of the UK financial regulation.

The database also contains information on a number of company characteristics, the most complete of which are total annual revenues, total assets and company listing status (public or private). The database also shows the length of the relevant financial year for the audit, which equals one full year for the vast majority of observations but are significantly shorter or longer than one full year for a small number of observations. Such 'partial year' (defined as fewer than 325 days) and 'multi year' (defined as more than 405 days) audits each make up less than 1 percent of the data. Tables 1.1 and 1.2 provide the definitions and descriptive statistics for the buyer characteristics variables.

Variable	Description
Price	Core audit fees, GBP (log)
Revenues	Total company revenues, GBP (log)
Assets	Total company assets, GBP (log)
Listed	Dummy variable taking value 1 if company is a public listed company
Partial year	Audit reflecting partial year (<325 days)
Multi year	Audit reflecting longer period than one full year (>405 days)

Table 1.1: Buyer characteristics variables definitions

Variable	N	Mean	SD	P-25	Median	P-75
Price	6,454	5.78	1.48	4.84	5.70	6.68
Revenues	6,454	13.10	1.86	12.03	13.18	14.24
Assets	6,454	13.56	1.88	12.39	13.39	14.49
Listed	6,454	0.762				
Partial year	6,454	0.008				
Multi year	6,454	0.006				

Table 1.2: Buyer characteristic variables descriptives

Table 1.3 provides a summary of the auditor choices observed in the dataset. The market is highly concentrated, with the Big Four auditors accounting for around 90% of the observed 'stock' of existing clients as well as 'flow' of new clients. The rate of observed switching in the sample is low; of the total 6,454 observations, 289 were switches to an auditor other than the incumbent auditor, an overall switching rate of 4%.

Auditor	Switch	Non-switch	Total
PwC	58	1,946	2,004
Deloitte	104	1,242	1,346
KPMG	55	1,485	1,540
E&Y	40	874	914
Others	32	618	650
Total	289	6,165	6,454

Table 1.3: Observed auditor choices in dataset

The buyer characteristics and the auditor choices are both correlated with the observed audit prices in the dataset. Table 1.4 shows the results of a simple regression of prices against buyer characteristics and auditor choices based on ordinary least squares.

As noted in the regulatory review for which the dataset was collated (Competition Commission 2013), buyers that switch auditors are observed to pay significantly lower prices on average compared to buyers that choose the incumbent auditor, with the difference amounting to around 8%. There are also significant price differences between sellers, with the Big Four auditors observed to charge around 12% higher prices on average than smaller auditors. These pricing patterns are relevant to some of the modelling choices set out in the following sections, such as the existence of auction costs for the buyer and asymmetries between sellers.

Buyer characteristics also have large and significant effects on audit prices, as would be expected. The buyer's assets and revenues both have significant positive effects, both in the current period and with a one-period lag. Listed companies pay over 30% higher prices on average compared to non-listed companies. Finally, audits for financial years lasting significantly shorter (longer) duration than one year show significantly lower (higher) prices.

Appendix B contains additional descriptions of the dataset and explains the data cleaning process in more detail.

	(1) OLS Price (log)	(2) OLS Price (log)
Auditor switch	-0.0889***	-0.0867***
	(0.0223)	(0.0222)
Seller: PwC	0.0870*	
	(0.0360)	
Seller: Deloitte	0.0975**	
	(0.0333)	
Seller: KPMG	0.154***	
	(0.0389)	
Seller: E&Y	0.192***	
	(0.0427)	
Seller: Big Four		0.115***
		(0.0298)
Assets (log)	0.193***	0.193***
	(0.0143)	(0.0143)
Revenues (log)	0.197***	0.197***
	(0.0111)	(0.0111)
Assets (log) ($t-1$)	0.0760***	0.0754***
	(0.0137)	(0.0137)
Revenues (log) ($t-1$)	0.0312***	0.0309***
	(0.0105)	(0.0105)
Listed	0.317***	0.310***
	(0.0239)	(0.0238)
Partial year	-0.303***	-0.307***
	(0.0531)	(0.0531)
Multi year	0.323***	0.322***
	(0.0595)	(0.0595)
Year fixed effects	Yes	Yes
Buyer fixed effects	Yes	Yes
N	6,454	6,454
R^2	0.9548	0.9547

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 1.4: Price regression against buyer characteristics and auditor choices

1.3 Model

This section describes the dynamic model of infinitely repeated auctions. I first describe the overall structure of the model and then set out the sellers' intertemporal valuations that arise from the dynamic aspects of the auctions. I then describe the sellers' costs and decisions in more detail. Finally, I set out some distributional assumptions which, while not strictly necessary to identify the model, makes the empirical computation much more feasible.

1.3.1 Model structure

There are an infinite number of discrete periods indexed by t. There are I number of consumers indexed by i, who make exactly one purchase in each period from one of J sellers, indexed by j. All agents are risk-neutral. In each period, one of the sellers is the incumbent seller for each buyer, denoted h_{it}.

The buyer's utility in each period is determined by the price paid for the audit P_{it}, and a stochastic auction cost for the buyer T_{it} which is only incurred if the buyer holds an auction in that period:

$$U_{it} = -P_{it} - 1_{it}^{Auction} T_{it} \tag{1.1}$$

where $1_{it}^{Auction}$ is an indicator variable taking value 1 if the buyer has held an auction in that period.

Each seller j faces costs C_{ijt} of supplying consumer i in period t, the distribution of which is independent over time and will be defined in more detail in later sections. The distributions of all sellers' costs are known to all agents, but the realisation of the cost C_{ijt} only occurs partway through the period t and is known privately by seller j only. Seller j's payoff in period t relating to buyer i equals the price P_{it} minus the seller's costs C_{ijt} if seller j is the seller that supplies consumer i during the period, and equals zero for all other sellers.

Each period proceeds as follows.

1. In the *incumbent offer stage*, the buyer draws an auction cost T_{it} which is independent and identically distributed over time and observed by both the buyer and the incumbent. At this stage the incumbent's costs are not yet realised, but both the buyer and the incumbent know the distribution of the incumbent's costs. The incumbent gets the opportunity to make a take-or-leave offer to the buyer, denoted P_{it}^I.

   (a) If the buyer accepts this offer, then the incumbent h_{it} supplies the buyer during this period and remains the incumbent seller for the next period. The price charged is equal to the offer price: $P_{it} = P_{it}^I$.

   (b) If the buyer rejects the incumbent's offer or if the incumbent does not make an offer (which, given the complete information between the two parties, amounts to the same outcome), then the period moves to the *auction stage*.

2. In the *auction stage* (in the scenario that the buyer rejects the incumbent's offer), the costs of all sellers are realised. All sellers take part in a second-price procurement auction and submit bids denoted B_{ijt}. The auction yields an auction price (second lowest bid) denoted P_{it}^A which is the final price charged to the buyer: $P_{it} = P_{it}^A$. The winner of the auction is the supplier for the current period and becomes the incumbent for the next period.

The stochastic outcomes for each period consist of the buyer's auction cost T_{it} and the seller-specific costs C_{ijt}, both of which are independently distributed over time and therefore only affect the current period payoffs for both the buyer and the sellers. The only relevant state variable in the model is the identity of the incumbent in each period, h_{it}, and the stochastic outcomes in each period can only affect future payoffs through the identity of the incumbent. It is useful to note that the state variable only matters at the incumbent offer stage; once any given period reaches an auction stage, the identity of the incumbent no longer affects any of the outcomes. This is a useful

result since it implies that the distribution of bids are not conditional on the identity of the incumbent. This makes it possible to use the two-stage estimation procedure described in later sections to speed up the estimation considerably.

The model structure imposes the assumption that there are no informational asymmetries between the buyer and the incumbent at any stage. Both the buyer and the seller observe the buyer's auction costs in the incumbent offer stage, and neither the buyer nor the seller observe the seller's cost draw for that period during the incumbent offer stage, although the underlying distribution is known to both parties. These are strong assumptions in both cases, but the former may be justified since the buyer's auction costs can be interpreted as incorporating all non-price factors that may induce a buyer to hold a tender in that particular period, some of which may be very public in nature.[5] The latter assumption may be justified since the incumbent's initial offer may be formed a significant time before the start of the audit itself, and therefore not all aspects of the underlying costs may be known at the time of the incumbent's initial offer.

The auction is modelled as a second-price auction, which also simplifies the analysis since the standard result that bidders bid their true valuations still holds. Compared to a standard static second-price auction where the bidders' true valuations would equal their cost draws, the valuations in this dynamic setting equal the bidders' cost draws plus an additional *incumbency valuation* term which reflects the value that each bidder attaches to starting the next period as the incumbent. This incumbency valuation term therefore represents the equilibrium bidding strategies in the auction stage of the model, and is a key part of determining the sellers' equilibrium strategies at both stages of the model.

[5] As an extreme example, a significant fault by the incumbent auditor that requires a restatement of the accounts would be publicly known and would most likely result in a tender in the following year. Events such as this can be interpreted as a very low or negative draw of the auction cost. Other public and significant changes for the company such as mergers, listing or delisting can also increase the likelihood of a tender for reasons other than the incumbent's pricing. Furthermore, there is some evidence that incumbents' pricing offer decisions can also take into account client-specific non-public information gained through the audit itself (Hackenbrack et al. 2014), which reduces the degree of private information available to the buyer but not the seller.

1.3.2 Sellers' intertemporal valuations

The prospect of incumbency introduces a dynamic aspect to sellers' pricing and bidding decisions. All sellers are forward-looking and risk-neutral, and act to maximise their net present value of current and future profits rather than simply maximising their short-term profits myopically. This section describes the sellers' intertemporal valuations in more detail. The aim of the section is to characterise the *incumbency valuation* Φ_j, which is defined as the net present value of the gain in future payoffs for seller j that arises from starting the next period as the incumbent rather than as a non-incumbent.

Denote all stochastic variables that are realised in each period as $\boldsymbol{\theta} = \{\theta_j, \theta_{-j}\}$, where θ_j denotes all variables that are observed by seller j and θ_{-j} denotes all variables not observed by seller j.[6] Also denote a complete set of strategy profiles as $\boldsymbol{\alpha} = \{\alpha_j, \alpha_{-j}\}$, where α_j denotes seller j's strategies and α_{-j} denotes the strategies of all other sellers.

As defined in the earlier section, h denotes the incumbent for the period and is the only state variable in the model. Define ω, a set of transition probabilities for this state variable based on the ex-ante distributions of $\boldsymbol{\theta}$:

$$\omega_j\left(h, \boldsymbol{\alpha}\right) = \Pr\left(h_{t+1} = j | h, \boldsymbol{\alpha}\right). \tag{1.2}$$

Denote seller j's profits in period t as $\pi_j\left(h, \boldsymbol{\theta}, \boldsymbol{\alpha}\right)$, which is given by the difference between the price and seller j's costs if j is the chosen supplier for the period, and equals zero otherwise. Sellers act to maximise the discounted sum of the current and future values of these profits, choosing their own strategy α_j based on their observed stochastic outcomes for the current period θ_j and taking as given some set of beliefs about the likely strategies of other sellers, α_{-j}. This defines the following *value function* as an outcome of a dynamic optimisation problem by each seller j:

[6] Buyer subscripts i have been omitted for brevity, and time subscripts t have been omitted except for indicating periods other than t. All expressions can be understood to be with respect to a specific buyer for a specific period.

$$W_j(h, \theta_j, \alpha_{-j})$$
$$= \max_{\alpha_j} \left[\int \pi_j(h, \boldsymbol{\theta}, \boldsymbol{\alpha}) f(\theta_{-j}) d\theta_{-j} + \delta \sum_{k=1}^{J} \omega_k(h, \boldsymbol{\alpha}) E_{\theta_{j,t+1}} W_j(k, \theta_{j,t+1}, \alpha_{-j}) \right] \quad (1.3)$$

where δ is the discount factor and the expectations operator $E_{\theta_{j,t+1}}$ is with respect to the realisation of j-observed stochastic variables in period $t+1$. The fact that the identity of the incumbent is the only state variable and that the current period stochastic variables cannot affect the sellers' future profits except through the identity of the incumbent means that the transition probabilities for the state variable ω can be separated from the value functions W in the last term.

The value function can be simplified as follows. Firstly, define the *ex-ante value function* as the value function evaluated before the realisation of the stochastic variables in the current period:

$$V_j(h, \alpha_{-j}) = E_{\theta_j} W_j(h, \theta_j, \alpha_{-j}) \quad (1.4)$$

where the expectations operator E_{θ_j} is with respect to the realisation of j-observed stochastic variables in the current period.

Also assume an equilibrium set of strategies by all sellers including j, with each seller maximising the value function as defined above based on a consistent set of beliefs on the other sellers' strategies, yielding a set of equilibrium outcomes π^* and ω^*. This results in the following expression for the ex-ante value function:

$$V_j(h) = E_{\theta_j} \left[\int \pi_j^*(h, \boldsymbol{\theta}) f(\theta_{-j}) d\theta_{-j} + \delta \sum_{k=1}^{J} \omega_k^*(h) V_j(k) \right]. \quad (1.5)$$

So far, this ex-ante value function has been shown in a generic form and has not explicitly taken into account the two-stage structure of each period. This can now be added to the model as follows.

Firstly, define two sets of probabilities specific to each seller. Auction probabilities λ_j define the ex-ante probability of the incumbent's initial offer being rejected and an auction taking place when the incumbent j follows the equilibrium strategy as defined above. Conditional on an auction taking place, winning probabilities s_j define the ex-ante probability of the bidder j winning an auction when all bidders follow the equilibrium strategies as defined above.[7] The transition probabilities ω^* can be related to these two sets of probabilities as follows:

$$\omega_j^*(h) = \begin{cases} (1-\lambda_j) + \lambda_j s_j & \text{if } h = j \\ \lambda_j s_j & \text{if } h \neq j \end{cases}. \tag{1.6}$$

Similarly, define two sets of ex-ante expected profits. π_j^I is the ex-ante expected profits for seller j conditional on being an incumbent and making a successful initial offer to the buyer. π_j^A is the ex-ante expected profits for seller j conditional on taking part in an auction and winning. These relate to the equilibrium profits π^* as follows:

$$\pi_j^I = E_{\theta_j}\left[\int \pi_j^*(j, \boldsymbol{\theta})\, f(\theta_{-j})\, d\theta_{-j} | 1^{Auction}=0\right] \tag{1.7}$$

and

$$\pi_j^A = E_{\theta_j}\left[\int \pi_j^*(h, \boldsymbol{\theta})\, f(\theta_{-j})\, d\theta_{-j} | 1^{Auction}=1, j \text{ win}\right] \tag{1.8}$$

where $1^{Auction}$ is an indicator variable taking value 1 if the buyer has held an auction in that period.[8]

Based on these definitions, the ex-ante value functions can be expressed in the following

[7] Both probabilities will be defined more precisely below. For now, these can be interpreted as being ex-ante with respect to all stochastic variables $\boldsymbol{\theta}$.

[8] π_j^A does not need to be expressed as a function of the identity of the incumbent h since, conditional on the auction occurring, the identity of the incumbent has no impact on the auction outcomes.

way, depending on whether seller j is itself the incumbent or otherwise:

$$V_j(h) = \begin{cases} (1-\lambda_j)\,\pi_j^I + \lambda_j s_j \pi_j^A + \delta\left\{(1-\lambda_j)\,V_j(j) + \lambda_j \sum_{k=1}^J s_k V_j(k)\right\} & \text{if } h = j \\ \lambda_h s_j \pi_j^A + \delta\left\{(1-\lambda_h)\,V_j(h) + \lambda_h \sum_{k=1}^J s_k V_j(k)\right\} & \text{if } h \neq j \end{cases}. \quad (1.9)$$

Conditional on being the incumbent, seller j makes a successful initial offer with probability $(1-\lambda_j)$ and receives payoffs of π_j^I in expectation. In case of a successful initial offer, other non-incumbent sellers do not receive any payoffs. Auctions occur in each period with probability λ_h that varies with the identity of the incumbent, and each seller j has ex-ante probability s_j of winning the auction and receiving payoffs π_j^A in expectation. The continuation values in the form of one-period-discounted ex-ante value functions are also weighted by similar probabilities.

The ex-ante value functions can be used to define the *incumbency valuation* term for seller j, denoted Φ_j. This is the present value of the gain in future payoffs for seller j that arises from starting the next period as the incumbent rather than as a non-incumbent, where the identity of the incumbent in the latter case is probability-weighted based on other sellers' auction winning probabilities s_j:[9]

$$\Phi_j = \delta\left\{V_j(j) - \frac{\sum_{k\neq j} s_k V_j(k)}{\sum_{k\neq j} s_k}\right\}. \quad (1.10)$$

This incumbency valuation captures the dynamic incentives of the sellers in making their incumbent offer or bidding decisions and is therefore a crucial part of the estimation process. The following sections will discuss various parts of the model that more precisely define the terms that flow into the expression for the incumbency valuation, namely equilibrium profits π_j^I and π_j^A, auction probabilities λ_j, auction winning probabilities s_j and the discount rate δ.

[9]The auction winning probabilities are the relevant weightings to use here because at any stage where sellers are forming their pricing or bidding decisions, the auction winning probabilities represent the relative transition probabilities of each of the other sellers becoming the chosen supplier in the current period and becoming the incumbent in the next period.

1.3.3 Sellers' costs

The costs for seller j involved in supplying buyer i have the following structure:

$$C_{ijt} = \bar{c}_{it} - 1_{it}^{Offer}\kappa + \varepsilon_{ijt} \tag{1.11}$$

where 1_{it}^{Offer} is an indicator variable taking value 1 if the seller is supplying the buyer based on a successful initial incumbent offer rather than through an auction (i.e. $1_{it}^{Offer} = 1 - 1_{it}^{Auction}$). Each seller's costs are the sum of a common cost term $\bar{c}_{it}$ specific to the buyer for each period and perfectly observed by all sellers, an efficiency term κ which applies if the seller has avoided an auction in that period, and a seller-specific cost residual ε_{ijt}.

The common cost term reflects characteristics of the buyer and the time period which symmetrically affects the costs of all sellers. The seller-specific cost residuals represent the private cost draws for each of the sellers which determine the stochastic auction outcomes. These are independent and identically distributed across buyers and over time, but may have different distributions for different sellers reflecting any asymmetries in costs across sellers. Finally, the cost efficiency term κ represents the cost savings from avoiding an auction process as well as any cost efficiencies from the existing supplier relationship, which may be lost if the incumbent needs to go through a formal auction process again.

The inclusion of the common cost term is important as the costs of audits vary widely depending on the characteristics of the buyer. This includes both time-varying variables, such as the company revenues and assets, as well as variables that do not vary much over time or at all, such as intrinsic company-specific characteristics that make the company more or less costly to audit. Evidence from the audit industry suggest that such common costs are very significant and account for most of the variation observed in audit prices (Doogar et al. 2015). However, they are less important for the purposes of modelling the competition between sellers since they enter the total costs additively and do not influence the seller-specific cost residuals.

It is convenient to define two different cost terms, representing the relevant costs facing the seller when making the initial incumbent offer decision and the auction bidding decision. Each seller's underlying costs in an auction situation is denoted C^A_{ijt} and is equal to the cost expression above where the efficiency term does not apply:

$$C^A_{ijt} = \bar{c}_{it} + \varepsilon_{ijt}\,. \tag{1.12}$$

In an incumbent offer situation, the value of the cost residual is not yet realised. Since the incumbent is risk-neutral, the incumbent makes its offer decision based on its expected costs, which will be denoted C^I_{ijt} and can be expressed as:

$$C^I_{ijt} = E\left[C^A_{ijt}\right] - \kappa\,. \tag{1.13}$$

1.3.4 Bidding behaviour by suppliers in an auction

If an auction takes place, the sellers form their bids in order to maximise the expected net present value of the current and future payoffs taking into account the impact of current period auction outcomes on future period payoffs via the identity of the incumbent in the next period. Under the second price auction framework, the equilibrium bidding strategies can be expressed in a simple way using the incumbency valuation term Φ_j.

To do this, note that as in standard auction theory, the unique equilibrium in the second price auction game involves all bidders bidding their true valuations. In this case each bidder's true valuation of winning the auction incorporates their underlying costs C^A_{ijt} as in standard auction models, but also take into account the expected net present value of starting the next period as an incumbent rather than a non-incumbent. The incumbency valuation Φ_j therefore equals the amount by which each bidder shades their bids below their cost draws in the auction, since it represents the net present value of the incumbency benefit that the auction winner would obtain, in addition to the

immediate value in the present period from supplying the buyer. Denoting firm j's equilibrium bids in the auction as B_{ijt}:

$$B_{ijt} = C^A_{ijt} - \Phi_j \,. \tag{1.14}$$

The incumbency valuation term, in combination with the underlying costs, therefore expresses the equilibrium bidding strategies of the J bidders in the dynamic auction game.

Note that the incumbency valuation term Φ_j is time-invariant, and in particular is not influenced by firm j's specific cost draw for the current period. The cost draws for the current period will affect the firm's payoffs in the current period, but do not have lasting effects on the firm's payoffs for the next period except through auction outcomes in the current period. This means that the distributions of the bids B_{ijt} are shifted versions of the underlying firm-specific costs C^A_{ijt}, with the magnitude of the shift given by the values of Φ_j. This is a useful result for estimating the distribution of bids since it means that the distributional assumptions for underlying costs, which are not directly observed, imply similar distributions for bids, which are directly observed. If the asymmetric cost distributions of bidders are represented by a family of distributions with different location parameters for each individual bidder j, then this implies that the distributions of bids for each bidder also belong to the same family of distributions with a different set of location parameters for each bidder.

Characterising the equilibrium bidding strategy in this way means that π^A_j, the ex-ante expected profit for bidder j conditional on winning the auction, can be expressed in a more precise way. Firstly, as in standard auction theory, the expected value of the winning auction price conditional on bidder j being the winner is given by the expected value of the minimum bid among all bidders other than j, conditional on the bidder j's bid being lower than this minimum:

$$E\left[P^A | B^{min}_{-j} > B_j\right] = E\left[B^{min}_{-j} | B^{min}_{-j} > B_j\right] \tag{1.15}$$

where B_{-j}^{min} denotes the minimum bid among all bidders except bidder j, and i and t subscripts have been omitted for brevity. The ex-ante expected profit for bidder j conditional on winning the auction is given by the difference between this expected winning price and the expected costs of bidder j conditional on winning the auction:

$$\begin{aligned}\pi_j^A &= E\left[P^A | B_{-j}^{min} > B_j\right] - E\left[C_j^A | B_{-j}^{min} > B_j\right] \\ &= E\left[B_{-j}^{min} | B_{-j}^{min} > B_j\right] - E\left[B_j + \Phi_j | B_{-j}^{min} > B_j\right] \\ &= E\left[B_{-j}^{min} | B_{-j}^{min} > B_j\right] - E\left[B^{min}\right] - \Phi_j \,. \end{aligned} \tag{1.16}$$

where B^{min} denotes the minimum bid among all bidders, and the second and third lines express the same ex-ante expected profits in terms of bids rather than costs based on the equilibrium bid definition.

Note that the ex-ante profit term π_j^A represents the expected value of the current-period profits from winning the auction, and does not account for the additional payoffs that the winner of the auction would gain in the following periods from starting the next period as the incumbent. The total expected profit from winning the auction in net present value terms can be obtained by adding the incumbency valuation Φ_j to the above expression, which simply equals the difference between the bids of the winning bidder and the second-placed bidder as per the standard result for second price auctions.

1.3.5 Initial offer behaviour by the incumbent

At the point when the buyer is facing an initial offer by the incumbent, the auction price P_{it}^A is not yet realised. The buyer is willing to accept the incumbent's offer P_{it}^I as long as it satisfies:

$$P_{it}^I \leq E\left[P_{it}^A\right] + T_{it} \tag{1.17}$$

where $E\left[P_{it}^A\right]$ is the unconditional expected price that would be achieved in an auction

and T_{it} is the stochastic auction cost for the buyer. Positive values of auction costs produce an outcome where buyers generally pay lower prices via auctions compared to accepting the incumbent's initial offer, which matches the patterns observed in the data.

The incumbent j stands to gain some current period profits if its offer is accepted, amounting to the offer price minus its costs C^I_{ijt}. It also retains the incumbency status for the next period if its offer is accepted, giving it additional payoffs in the form of the incumbency valuation Φ_j. The incumbent is willing to make an initial offer to the consumer to bypass an auction process if the payoffs it expects to gain from this offer exceeds its expected payoffs from taking part in an auction, inclusive of the incumbency valuation in both cases:

$$P^I_{it} - C^I_{ijt} + \Phi_j \geq s_j \left(\pi^A_j + \Phi_j\right) . \tag{1.18}$$

The initial offer stage therefore proceeds in one of two ways. A credible offer which would be acceptable to both the consumer and the incumbent exists if:

$$E\left[P^A_{it}\right] + T_{it} - C^I_{ijt} + \Phi_j \geq s_j \left(\pi^A_j + \Phi_j\right) . \tag{1.19}$$

If this constraint is satisfied then the equilibrium outcome involves the incumbent making an offer at the maximum price that the consumer is willing to accept, $P^I_{it} = E\left[P^A_{it}\right] + T_{it}$, and the consumer accepting the offer. If the constraint is not satisfied, the incumbent does not make a credible offer (or equivalently, gives up the incumbency status) and the consumer opts to hold an auction, incurring the auction costs T_{it}.

All of the terms in the expression above are deterministic aside from the auction costs T_{it}. There is therefore a unique realisation of T_{it}, denoted T^*_{it}, where the constraint on the existence of a credible incumbent offer is exactly satisfied:

$$T^*_{ijt} = C^I_{ijt} - \Phi_j - E\left[P^A_{it}\right] + s_j \left(\pi^A_j + \Phi_j\right) . \tag{1.20}$$

If the draw of T_{it} is higher than this, then the incumbent makes a credible offer and retains the incumbency status. If the draw of T_{it} is lower than this, then there is no offer price that satisfies both the incumbent and the buyer, and an auction takes place. The distribution of T_{it} and the value of T_{it}^* define the probability that an auction occurs in any given period, as follows:

$$\lambda_j = F_T\left(T_{ijt}^*\right) \tag{1.21}$$

where F_T is the time-invariant distribution of T_{it}. The ex-ante expected profits for an incumbent conditional on a successful offer can be written as:

$$\begin{aligned} \pi_j^I &= E\left[P_{it}^I - C_{ijt}^I | T_{it} \geq T_{ijt}^*\right] \\ &= E\left[T_{it} | T_{it} \geq T_{ijt}^*\right] + E\left[P_{it}^A\right] - C_{ijt}^I \end{aligned} \tag{1.22}$$

1.3.6 Distributional assumptions

The model includes two stochastic random variables: the buyer's auction costs T_{it}, which are specific to the buyer but not the seller, and the seller-specific cost residuals ε_{ijt}. These could be estimated nonparametrically, but using the following distributional assumptions makes the estimation less computationally demanding.[10]

I assume that buyers' auction costs are normally distributed with mean τ and variance σ_τ^2, and independent and identically distributed over time and across buyers. This defines the distributions of various incumbent offer-related market outcomes set out in the previous section. For example, successful incumbent offers follow a truncated normal distribution, with the truncation driven by the value of the critical auction cost T_{it}^*. The probability of an auction occuring in any given period is:

[10]The distributional assumptions ensure that the auction outcomes have closed form likelihoods, which speeds up the numerical maximisation of the likelihood function. More importantly, the distributional assumptions ensure that the model can be estimated based on the two-stage estimation process described in the next section.

$$\lambda_j = F_{norm}\left(\frac{T^*_{it} - \tau}{\sigma_\tau}\right) \tag{1.23}$$

where F_{norm} is the standard normal distribution.

As in Froeb et al. (2009), I assume that the seller-specific cost residuals are Gumbel-distributed with seller-specific location parameters η^C_j and scale parameter μ which is equal for all sellers:

$$f_\varepsilon\left(\varepsilon; \eta^C_j, \mu\right) = \frac{1}{\mu} e^{\left(\varepsilon - \eta^C_j\right)/\mu} e^{-e^{\left(\varepsilon - \eta^C_j\right)/\mu}}. \tag{1.24}$$

Since Φ_j is deterministic with respect to the realisation of the cost residuals, this implies that the distributions of each seller's 'bid residuals' (i.e. bid values minus the common costs $\bar{c}$) also follow a Gumbel distribution $f_\varepsilon\left(\varepsilon; \eta^B_j, \mu\right)$, with the same scale parameter μ but a different set of location parameters η^B_j, where:

$$\eta^B_j = \eta^C_j - \Phi_j\,. \tag{1.25}$$

The standard logit properties of the Gumbel distribution then define the auction winning probabilities for each of the sellers as follows:

$$s_j = \frac{\exp\left(-\frac{\eta^B_j}{\mu}\right)}{\sum_{k=1}^{J} \exp\left(-\frac{\eta^B_k}{\mu}\right)}\,. \tag{1.26}$$

The Gumbel distributional assumption also gives a closed form expression for the distribution of the second-lowest bid conditional on the identity of the bidder with the lowest bid, or in other words the distribution of prices conditional on the auction winner:

$$f_{P^A|j\,win}\left(P\right) = \left(\frac{1}{s_j}\right) f_\varepsilon\left(P - \bar{c}_{it}; \eta^B_{\sim j}, \mu\right) - \left(1 - \frac{1}{s_j}\right) f_\varepsilon\left(P - \bar{c}_{it}; \eta^B_{max}, \mu\right)\,. \tag{1.27}$$

where $\eta^B_{\sim j} = -\mu \log\left(\sum_{k \neq j} \exp\left(-\frac{\eta^B_k}{\mu}\right)\right)$ and $\eta^B_{max} = -\mu \log\left(\sum_{k=1}^{J} \exp\left(-\frac{\eta^B_k}{\mu}\right)\right)$. This

distribution can also be used to find the ex-ante expected value of the auction price.[11]

Finally, the common cost term $\bar{c}_{it}$ is modelled as a linear combination of buyer characteristics X_{it}:

$$\bar{c}_{it} = X_{it}\beta\,. \tag{1.28}$$

Buyer characteristics X_{it} include both buyer (i) and time (t) fixed effects, to capture buyer and time-specific drivers of costs that are not captured by the observed buyer characteristics from the data.

1.4 Estimation

This section discusses the process for estimating the parameters of the model, in a setting where each observation includes the incumbent, the chosen supplier and the final price, but the occurrence of an auction and the bids of any of the individual bidders are not directly observed.[12]

The observed outcomes in the data are the chosen supplier in any given period (denoted Y_{it}) and the price paid to the supplier (P_{it}), conditional on the identity of the incumbent (h_{it}). The underlying parameters of the model are: the common cost covariate parameters (β); the expected value and standard deviation of the buyer-specific auction costs (τ, σ_τ); the incumbent cost advantage (κ); the scale parameter of the bid and cost distributions (μ); the seller-specific location parameters of the cost distributions ($\boldsymbol{\eta}^{\boldsymbol{C}}$); and the discount rate (δ).

The general framework for estimating the model is a two-step estimation strategy set out in Guerre et al. (2000) for static auctions and applied in Jofre-Bonet and Pesendorfer (2003) to dynamic auctions. The first stage of the estimation process proceeds

[11] All relevant derivations are shown in Froeb et al. (1999).

[12] This reflects the data availability in this particular setting. Directly observing the occurrence of auctions or the bids of individual bidders would simplify the estimation structure in places but generally would not affect the identification of the model.

similarly to estimation of a static second-price auction model, where the distributions of bids are treated as distribution of *pseudo* private values. This yields parameter estimates for the common cost covariate parameters (β), the expected value and standard deviation of the buyer-specific auction costs (τ, σ_τ), the incumbent cost advantage (κ) and the scale parameter of the bid and cost distributions (μ). In addition, while the seller-specific location parameters of the bid distributions ($\boldsymbol{\eta^B}$) are the outcomes of the equilibrium bidding strategies for the sellers and are therefore functions of model parameters rather than being parameters in themselves, the first stage estimation process treats these as the underlying parameters of the distribution of pseudo private values and yields numerical estimates for these parameters.

The second stage of the estimation process involves inferring the values of the seller-specific location parameters of the cost distributions ($\boldsymbol{\eta^C}$), based on the first stage parameter estimates and an assumption for the discount rate (δ). The discount rate itself is not identified by either stage of the estimation process and must therefore be based on an exogenous assumption.

1.4.1 Stage 1: Estimating the distribution of bids

The first stage of the estimation can be carried out through a maximum likelihood approach. Denote as Ψ the parameters to be estimated in the first stage of the estimation, including the seller-specific location parameters of the bid distributions ($\boldsymbol{\eta^B}$) which at this stage can simply be treated as the parameters of the distributions for pseudo private values rather than the outcome of a dynamic bidding game:

$$\Psi = \left\{\beta, \tau, \sigma_\tau, \kappa, \mu, \boldsymbol{\eta^B}\right\} . \tag{1.29}$$

Conditioning on an auction having taken place in any given period for a specific buyer, the observed supplier for the period Y_{it} must be the winner of the auction and the observed price P_{it} must be the winning auction price. As explained in the previous section, the Gumbel distributional assumption for the sellers' costs ensures that the

distribution of the auction price has a closed form. The conditional likelihood for the observed auction outcomes in the form of the auction winner and the auction price is therefore given by:

$$\begin{aligned}&\mathcal{L}\left(\Psi|Y_{it}=j,P_{it},1_{it}^{Auction}{=}1\right)\\&\quad=s_j\left\{\left(\frac{1}{s_j}\right)f_\varepsilon\left(P_{it}-X_{it}\beta;\eta_{\sim j}^B,\mu\right)-\left(1-\frac{1}{s_j}\right)f_\varepsilon\left(P_{it}-X_{it}\beta;\eta_{max}^B,\mu\right)\right\}\end{aligned}\tag{1.30}$$

where the term in the outermost bracket is based on the distribution of the second-lowest bid conditional on the identity of the bidder with the lowest bid as given in (1.27).

Conditioning on an auction not having taken place in a given period, the observed supplier must be the incumbent for that period and the observed price must the initial offer price from the incumbent. The conditional likelihood is given by a truncated normal distribution of the following form:[13]

$$\begin{aligned}&\mathcal{L}\left(\Psi|Y_{it}=j,P_{it},1_{it}^{Auction}{=}0\right)\\&\quad=\begin{cases}f_{norm}\left(\frac{P_{it}-E\left[P_{it}^A\right]-\tau}{\sigma_\tau}\right)/\left(1-\lambda_j\right) & \text{if } j=h_{it} \text{ and } P_{it}-E\left[P_{it}^A\right]\geq T_{it}^*\\0 & \text{otherwise}\end{cases}\end{aligned}\tag{1.31}$$

[13]This produces a likelihood function with a discontinuity at price $P_{it}=E\left[P_{it}^A\right]+T_{it}^*$, which can cause problems for numerical maximisation. For the purposes of estimation, this function is replaced with the following function:

$$\mathcal{L}^{\mathcal{S}}\left(Y_{it}=k,P_{it}|1_{it}^{Auction}=0\right)=\frac{1}{1+\exp\left(A\right)}f_{\mathcal{N}}\left(\frac{P_{it}-E\left[P_{it}^A\right]-\tau}{\sigma_\tau}\right)/\left(1-\lambda_{h_{it}}\right)$$

where:

$$A=\frac{T_{it}^*-P_{it}+E\left[P_{it}^A\right]}{\rho}.$$

This is a smoothed version of (1.31) with a logit-based smoother at the local point around the discontinuity point $P_{it}=E\left[P_{it}^A\right]+T_{it}^*$. This new function is continuous and twice-differentiable at all points which aids numerical maximisation of the likelihood function, and closely approximates (1.31) as $\rho\to 0$. The estimation uses ρ value of 0.05.

where f_{norm} is the density of the standard normal distribution.

In practice, whether an auction took place is only imperfectly observed, which means that these two conditional likelihoods cannot be directly applied to the data. This is a similar data limitation to that faced by Allen et al. (2019), and I take the same approach as these authors in forming a mixed likelihood based on the imperfectly observed switching behaviour. In periods where the data shows that the actual supplier is not the incumbent, it can be inferred with certainty that an auction took place. When the actual supplier is the incumbent, however, the incumbent may have either made an initial offer to the buyer which was accepted, or the incumbent may have been unsuccessful in making an initial offer but won the auction which subsequently occurred. The overall likelihood function given the observed outcomes in the data is therefore a mixture of the two conditional likelihoods given in (1.30) and (1.31), with the mixture weights given by $\lambda_{h_{it}}$, the ex-ante probability of an auction occurring depending on the identity of the incumbent h_{it}:

$$\begin{aligned}
&\mathcal{L}\left(\Psi|Y_{it}, P_{it}, h_{it}\right) \\
&= \begin{cases} \lambda_{h_{it}}\mathcal{L}\left(\Psi|Y_{it}, P_{it}, 1_{it}^{Auction}=1\right) + \left(1-\lambda_{h_{it}}\right)\mathcal{L}\left(\Psi|Y_{it}, P_{it}, 1_{it}^{Auction}=0\right) & \text{if } Y_{it} = h_{it} \\ \lambda_{h_{it}}\mathcal{L}\left(\Psi|Y_{it}, P_{it}, 1_{it}^{Auction}=1\right) & \text{if } Y_{it} \neq h_{it} \end{cases}.
\end{aligned} \tag{1.32}$$

1.4.2 Stage 2: Estimating the distribution of costs

Having established the distribution of bids in the first stage, the remaining part of the estimation is to back out the distributions of costs based on the estimated distributions of bids, or equivalently, to describe an equilibrium bidding strategy that produces the set of estimated bid distributions.

As set out in the earlier sections, the equilibrium bidding strategy can be expressed as a set of J incumbency valuations Φ_j, which satisfies the system of J equations of the

following form:

$$\Phi_j = \delta \left\{ V_j(j) - \frac{\sum_{k \neq j} s_k V_j(k)}{\sum_{k \neq j} s_k} \right\}. \tag{1.33}$$

Earlier sections also defined the ex-ante value functions as follows:

$$\begin{aligned} & V_j(h) \\ &= \begin{cases} (1-\lambda_j)\pi_j^I + \lambda_j s_j \pi_j^A + \delta \left\{ (1-\lambda_j) V_j(j) + \lambda_j \sum_{k=1}^J s_k V_j(k) \right\} & \text{if } h = j \\ \lambda_h s_j \pi_j^A + \delta \left\{ (1-\lambda_h) V_j(h) + \lambda_h \sum_{k=1}^J s_k V_j(k) \right\} & \text{if } h \neq j \end{cases}. \end{aligned} \tag{1.34}$$

Rearranging (1.34) and incorporating (1.33) gives the following expressions for the ex-ante value functions as functions of equilibrium profits π_j^I and π_j^A, auction probabilities λ_j, auction winning probabilities s_j and the discount rate δ:[14]

$$V_j(h) = \begin{cases} \frac{1}{1-\delta} \left\{ (1-\lambda_j)\pi_j^I + \lambda_j s_j \pi_j^A - \lambda_j (1-s_j) \Phi_j \right\} & \text{if } h = j \\ \frac{1}{1-\delta+\delta\lambda_h} \left\{ \lambda_h s_j \pi_j^A + \lambda_h \frac{V_j(j)}{(1+r)} - \lambda_h (1-s_j) \Phi_j \right\} & \text{if } h \neq j \end{cases}. \tag{1.35}$$

This expression then can be substituted back into (1.33) to yield an expression for the incumbency valuation which does not include the ex-ante value functions, only the equilibrium profits π_j^I and π_j^A, auction probabilities λ_j, auction winning probabilities s_j and the discount rate δ. Of these, the discount rate was not identified by the first stage estimates and therefore must be based on an exogenous assumption. The auction winning probabilities s_j are defined by the distributions of bids as estimated in the first stage of the esitmation process. The equilibrium profits π_j^I and π_j^A and auction probabilities λ_j were defined in earlier sections as functions of the underlying cost distributions which are unknown at this stage of the estimation, but all of the definitions can also be written as functions of bid distributions and the incumbency

[14] Where the first line can be substituted into $V_j(j)$ shown in the second line. The derivation is set out fully in Appendix A.

valuation Φ_j based on the definition of the equilibrium bids. The definitions are set out again below for reference:

$$\begin{aligned}\pi_j^A &= E\left[P^A | B_{-j}^{min} > B_j\right] - E\left[C_j^A | B_{-j}^{min} > B_j\right] \\ &= E\left[B_{-j}^{min} | B_{-j}^{min} > B_j\right] - E\left[B_j + \Phi_j | B_{-j}^{min} > B_j\right] \\ &= E\left[B_{-j}^{min} | B_{-j}^{min} > B_j\right] - E\left[B^{min}\right] - \Phi_j \end{aligned} \tag{1.36}$$

$$\begin{aligned}\pi_j^I &= E\left[P_{it}^I - C_{ijt}^I | T_{it} \geq T_{ijt}^*\right] \\ &= E\left[T_{it} | T_{it} \geq T_{ijt}^*\right] + E\left[P_{it}^A\right] - C_{ijt}^I\end{aligned}$$

$$\lambda_j = F_T\left(T_{ijt}^*\right) \tag{1.37}$$

where:

$$T_{ijt}^* = C_{ijt}^I - \Phi_j - E\left[P_{it}^A\right] + s_j\left(\pi_j^A + \Phi_j\right) \tag{1.38}$$

and:

$$\begin{aligned}C_{ijt}^I &= E\left[C_{ijt}^A\right] - \kappa \\ &= E\left[B_{ijt}\right] + \Phi_j - \kappa\,.\end{aligned} \tag{1.39}$$

Substituting these definitions into (1.35) yields expressions for the ex-ante value functions $V_j(h)$ and therefore the incumbency valuation Φ_j which do not rely on the distributions of the underlying costs and only include parameters estimated in the first

stage of the estimation process. This still does not yield an explicit solution for the incumbency valuation Φ_j, since Φ_j enters the expressions for the ex-ante value functions as shown in (1.35) both directly and indirectly via other terms such as the equilibrium profits. Instead, (1.33) becomes a system of J equations with J unknowns ($\boldsymbol{\eta}^C$) which can be solved by numerical methods.[15]

The fact that the second stage of the estimation process can be reduced to a system of J equations with J unknowns is a result of the distributional and modelling assumptions, and demonstrates the usefulness of the chosen approach in making the estimation process more computationally feasible. Restricting the cost asymmetries between sellers to the location parameters only and imposing a common scale parameter across sellers plays an important role, since this ensures that only J parameters remain to be inferred in the second stage. The modelling assumptions that lead to the result that the identity of the incumbent does not affect the bidding strategies in the auction stage are also crucial, since they ensure that the equilibrium strategies can be expressed as a set of J expressions.[16]

1.4.3 Identification

The auction portion of the first stage of the estimation as shown in (1.30) is equivalent to estimating the underlying distributions of costs in a standard independent private values (IPV) second price auction setting using data on winning prices only. This is a well-established problem and the underlying cost distributions are identified even without any distributional assumptions (Athey and Haile 2002). Equivalently, in the current setting the distributions of bids are identified by the observed auction prices, including both the firm-specific bid distribution parameters η_j^B and μ as well as the parameters for the common cost covariates β.

[15] Computational experiments appear to show that the system is identified for a wide range of inputs and can be solved very quickly with standard algorithms such as trust-region methods.

[16] As an illustration of a model where this would not hold, Froeb et al. (2009) assume a real cost advantage for the incumbent in their auctions which means that the identity of the incumbent does affect each seller's equilibrium bidding strategy. This turns the incumbency valuation into a system of J^2 equations and makes it impossible to exactly identify the distributions of costs based on observed distributions of bids.

The fact that the occurrence of auctions is not perfectly observed adds a complication to the structure of the likelihood function in the first stage. However, any observable outcome that is informative of an auction having taken place (even on a probabilistic basis) is enough to identify the different distributions for observed outcomes resulting from auctions relative to observed outcomes resulting from incumbents' initial offers. In this particular setting, any observation where the buyer switches supplier means that an auction must have taken place, and this ensures that both auction and non-auction outcomes are identified. The differences in price distributions between the 'switching' and 'non-switching' observations identifies the distribution of the auction costs for the consumer, since this is the main determinant of the price differential. The observed rate of switching then identifies the magnitude of the cost efficiency for the incumbent, which affects the probability of initial incumbent offers being successful but does not impact prices either at the incumbent offer stage or the auction stage.

The second stage of the estimation involves a system of J equations in J unknowns as shown in (1.33), which yields the firm-specific cost parameters $\boldsymbol{\eta}^{\boldsymbol{C}}$ as a function of all parameters estimated in the first stage and the discount factor based on numerical methods. Computational experiments appear to show that this system of equations is exactly identified for a wide range of inputs.

The discount factor itself is not identified in the estimation process and must be fixed based on some exogenous value. Depending on the discount factor assumed, the model can yield a wide range of outcomes. For example, as δ becomes small, firms discount the future more heavily and the incumbency valuation Φ_j approaches zero for all firms, moving each firm's bids closer to their costs and auction outcome towards the standard static auction outcome. Conversely, as δ becomes closer to unity, the incumbency valuation increases and firms bid even lower below their costs. Non-identification of the discount rate is a common feature of dynamic models (Rust 1994, Jofre-Bonet and Pesendorfer 2003).

1.5 Results

1.5.1 ML estimates of bid distributions

Table 1.5 reports the maximum likelihood estimates for the first stage of the estimation for the distributions of bids. Seven variables are used as covariates for the common cost component for each company-year observation: current and one-period lagged revenues, current and one-period lagged assets, a dummy variable for public listed companies and two dummy variables for audit periods significantly shorter or longer than one year. The common costs are also modelled with buyer and year-level fixed effects.

Parameter	ML Estimate	Standard error
Bid distribution scale parameter (μ)	1.2013	0.0822
Seller-specific bid location parameters: (η^B)		
PwC	-0.4684	0.0802
Deloitte	-0.5647	0.0721
KPMG	-0.4967	0.1133
EY	-0.4840	0.0741
Others*	-0.3190	0.1057
Auction cost expected value (τ)	0.2284	0.0663
Auction cost standard deviation (σ_τ)	0.2510	0.0420
Incumbent cost advantage (κ)	0.9619	0.0690
Common cost covariates: (β)		
Assets	0.1958	0.0403
Turnover	0.1356	0.0304
Assets (t-1)	0.0485	0.0376
Turnover (t-1)	0.0746	0.0339
Listed	0.2609	0.0598
Partial year	-0.2801	0.1701
Multi-year	0.2507	0.2055
Buyer fixed effects	Included	
Year fixed effects	Included	
N	6,454	
Companies	760	

Table 1.5: First stage ML estimates

The estimated seller-specific bid location parameters show the expected pattern, with the Big Four sellers showing a close grouping while all other sellers are shown to have higher costs than any of the Big Four.[17] The estimates therefore appear to confirm the dominance of the Big Four sellers in the industry.

The cost advantages from avoiding the auction are shown to be very significant for both the buyer and the seller. The estimated value of the auction cost parameter τ suggests

[17]The inclusion of the fixed effects mean that the absolute values of the bid location parameters do not have an interpretation and only the relative values are meaningful. The bid location parameter estimate for 'other' firms shown above should be interpreted as the location parameter for the distribution of the lowest bid among all sellers other than the Big Four.

that buyers are willing to pay a premium of around 25% on average in order to avoid an auction, while the estimated value of the incumbent cost advantage parameter κ suggests that the incumbents' ability to retain its incumbency status is equivalent to a cost reduction of well over 50%.

The common cost covariate parameters all have expected signs and generally have low standard errors. The company's total revenues and assets are shown to have positive impacts on audit costs, both for the current period and (to a lesser extent) with a one-period lag. Audits for financial years lasting significantly shorter (longer) durations than one year have lower (higher) costs compared to other audits, as would be expected. Finally, audits for public listed companies are shown to be significantly more expensive compared to audits for private companies, which may reflect the higher levels of disclosure and transparency required for public listed companies that result in higher required standards of audits.

1.5.2 Distributions of underlying costs implied by the estimated distributions of bids

Table 1.6 reports the cost distributions and the incumbency valuations for each bidder as implied by the bid distributions obtained in the first stage maximum likelihood estimates, fixing the discount factor at a value of $\delta = 0.9$.

Bidder	Bid locations (η^B)	Incumbency valuation (Φ_j)	Cost locations (η^C)	Lowest cost probability	Win probability
PwC	-0.4684	0.5255	0.0571	0.2002	0.1999
Deloitte	-0.5647	0.6017	0.0370	0.2035	0.2165
KPMG	-0.4967	0.5478	0.0511	0.2012	0.2046
EY	-0.4840	0.5378	0.0537	0.2007	0.2025
Others	-0.3190	0.4112	0.0922	0.1944	0.1765

Table 1.6: Second stage estimates of cost distributions

The estimates suggest that the dynamic aspects of the auction game have very significant impacts on bids. The value of the incumbency status leads all bidders in the

auction to bid well below their cost draws in the auction, with the ex-ante most efficient bidder (Deloitte) bidding 45% below their short-term costs and the tail of 'other' bidders bidding 34% below their short-term costs.

The estimates also suggest that incumbency benefits have the effect of magnifying any existing asymmetries between sellers, as the more cost efficient firms tend to place higher valuations on being the incumbent and therefore discount their bids more heavily compared to their costs. For example, the ex-ante most efficient bidder is estimated to have 20.3% probability of being the lowest cost in any given auction, but 21.7% probability of placing the lowest bid in the auction under the dynamic auction equilibrium. Conversely, the smaller firms other than the Big Four collectively have 19.4% probability of being the lowest cost seller in any given auction, but only 17.6% probability of placing the lowest bid in an auction under the dynamic auction equilibrium. Incumbency benefits therefore have the effect of increasing the seller concentration in the industry, relative to the shares that would be implied by the underlying cost distributions.

1.5.3 Goodness of fit

This section sets out some goodness of fit tests for the estimated parameters. Table 1.7 presents a comparison of the predicted auction winning probabilities for each firm against the observed winning probabilities (with the sample limited to the subset of the data where auctions are certain to have taken place, i.e. observed switches) and a comparison of the predicted and observed switching rates depending on the identity of the incumbent firm.

Bidder / Incumbent	Winning probability as bidder:		Switching rate as incumbent:	
	Predicted	Observed	Predicted	Observed
PwC	0.1999	0.2007	0.0759	0.0333
Deloitte	0.2165	0.3599	0.0434	0.0259
KPMG	0.2046	0.1903	0.0648	0.0262
EY	0.2025	0.1384	0.0696	0.0385
Others	0.1765	0.1107	0.1618	0.1557

Table 1.7: Observed vs. predicted auction winning probabilities and switching rate by bidder/incumbent

Overall, the model tends to predict too little variation in the predicted shares among the sellers compared to the data. The data shows share of 36% for the largest of the Big Four sellers and 11% for the tail of 'other' sellers. The parameter estimates instead show sellers' shares varying between 22% and 18%. However, the parameters do generally predict the relative ranking of the sellers well compared to the data, in particular correctly identifying the seller with the largest share and the smallest share.

The model tends to slightly overpredict the rates of switching across the board compared to the rates observed in the data. The data shows that buyers switch away from the largest of the Big Four sellers 2.6% of the time compared to 4.3% as predicted by the model, and switch away from the tail of 'other' sellers 15.6% of the time compared to 16.2% as predicted by the model. However, the model successfully reproduces the patterns observed in the data which show that buyers tend to switch away from smaller sellers much more frequently than from larger sellers.

1.6 Counterfactual experiment

This section sets out a counterfactual experiment to predict the impact of the recently introduced mandatory tendering regulations in the UK on market outcomes. The regulations stipulate that companies must put their audit contracts out to tender at least once every ten years.

The counterfactual experiment uses the parameter estimates set out in the previous section, including in particular the cost distribution parameters ($\boldsymbol{\eta^C}$) estimated in the second stage of the estimation process. Compared to the bid distributions ($\boldsymbol{\eta^B}$) that were numerically estimated in the first stage of the estimation process, the policy intervention changes the equilibrium of the dynamic auction game and therefore produces a new set of bid distributions based on the same underlying cost distributions. The new bid distributions then impact the market outcomes by changing the outcomes of the auction process.

Expressions previously given for the ex-ante value functions $V_j(h)$ in (1.35) now need to be adjusted to account for the impact on each sellers' payoffs from mandatory tenders as well as 'voluntary' tenders that occur before the incumbent reaches a tenure of ten years. I use a simulation approach to estimate the ex-ante value functions and therefore the new equilibrium auction bids. First, define a simulated version of the ex-ante value function, $V_j^{sim}\left(h, \boldsymbol{\Phi'}\right)$. $\boldsymbol{\Phi'} = \left\{\Phi'_1, ..., \Phi'_J\right\}$ is a candidate set of incumbency valuation terms which define the sellers' bid distributions based on the underlying cost distributions. This simulated value function is calculated by simulating a large number of random seller choice and price outcomes for a buyer over an extended period of time subject to an initial incumbent h, bidding strategies defined by $\boldsymbol{\Phi'}$ and the policy intervention of mandated tendering every ten years, and then taking the averages of each seller's net present value payoffs across all simulated outcomes.[18] These simulated ex-ante value functions then define the incumbency valuation in an analogous way to (1.10). The new equilibrium of the auction game subject to the mandated tendering policy intervention can be approximated by a set of incumbency valuation terms $\boldsymbol{\Phi^*} = \{\Phi_1^*, ..., \Phi_J^*\}$ that satisfy:

$$\Phi_j^* = \delta \left\{ V_j^{sim}\left(j, \boldsymbol{\Phi^*}\right) - \frac{\sum_{k \neq j} s_k V_j^{sim}\left(k, \boldsymbol{\Phi^*}\right)}{\sum_{k \neq j} s_k} \right\}. \tag{1.40}$$

[18] The results presented in this section are based on generating 200 years of market outcomes across 10,000 simulations for each possible initial incumbent h. As with the second stage of the estimation process, I use a discount rate value of 0.9 when calculating the net present values.

Solving this system of equations is a similar problem to the second stage of the two-stage estimation process discussed in previous sections, and can be carried out with the same numerical methods.[19]

Tables 1.8, 1.9 and 1.10 set out the results of this counterfactual experiment based on the simulation methodology: the impact of mandated tendering on sellers' bids and auction outcomes; the impact on occurrences of tenders; and the impact on prices.

		No mandated tendering	Mandated tendering	Impact (logs)	Impact
Bid location parameters (η_j^B)	PwC	-0.4684	-0.4538	0.0146	+1.5%
	Deloitte	-0.5647	-0.5273	0.0374	+3.8%
	KPMG	-0.4967	-0.4713	0.0254	+2.6%
	EY	-0.4840	-0.4624	0.0217	+2.2%
	Others	-0.3190	-0.3284	-0.0094	-0.9%
Win probability (s_j)	PwC	0.1999	0.2006		+0.0007
	Deloitte	0.2165	0.2132		-0.0033
	KPMG	0.2046	0.2035		-0.0011
	EY	0.2025	0.2020		-0.0005
	Others	0.1765	0.1807		+0.0042
HHI*		0.1697	0.1679		-0.0018

* Herfindahl-Hirschman Index calculated over auction winning probabilities (s_j) assuming that the group of 'other' bidders are atomistic.

Table 1.8: Impact of mandated tendering: Sellers' bids and auction outcomes

[19] It is, however, computationally much more demanding since the simulation to calculate the ex-ante value functions need to be repeated at every iteration of the numerical algorithm.

		No mandated tendering	Mandated tendering	Impact
Voluntary tender probabilities by incumbent (λ_j)*	PwC	0.0948	0.0929	-0.0020
	Deloitte	0.0554	0.0618	0.0064
	KPMG	0.0814	0.0845	0.0031
	EY	0.0872	0.0887	0.0015
	Others	0.1965	0.1733	-0.0232
Overall tender probability including mandated tenders		0.0856	0.1505	+0.0650

* Probability of auction occurring in given period due to buyer rejection of incumbent offer rather than mandated tender

Table 1.9: Impact of mandated tendering: Tenders

		Impact (logs)	Impact (%)
Expected auction price		+0.0186	+1.9%
Expected auction price conditional on winner	PwC	+0.0194	+2.0%
	Deloitte	+0.0161	+1.6%
	KPMG	+0.0179	+1.8%
	EY	+0.0185	+1.9%
	Others	+0.0221	+2.2%
Overall price impact including impact of more frequent tenders		+0.0045	+0.4%

Table 1.10: Impact of mandated tendering: Prices

For most sellers, the mandated tendering regulation has the effect of decreasing the incumbency valuations and therefore increasing the sellers' bids in an auction. Mandated tendering limits the length of time over which each seller can benefit from its incumbency position before having to compete for it again in an auction, and therefore decreases the net present value of winning an auction. This effect is greatest for the sellers with the ex-ante lowest cost distributions, since they are the sellers that would have been most likely to hold onto their incumbency status for the longest period in the absence of mandatory tendering. The simulation results show that the bids of the Big Four sellers increase by between 1.5% and 3.8% as a result of mandated tendering.

Only the tail of small 'other' sellers show lower bids as a result of mandated tendering, which occurs due to the general increase in prices resulting from mandated tendering which makes incumbency more valuable for all sellers.

Since mandated tendering raises the bids of the ex-ante lowest cost sellers relative to higher cost sellers, the overall impact on auction outcomes is to reduce the disparity in winning probabilities across the sellers. However, the magnitude of the impact is minor; the simulation results show that the ex-ante lowest cost seller's probability of winning each auction decreases by 0.3%, while the probability of the tail of small 'other' sellers winning each auction increases by 0.4%. The impact on other sellers is even more minor. As a result, the Herfindahl-Hirschman Index of market concentration calculated over the auction winning probabilities decreases by 18 points from 1,697 to 1,679.

Mandated tendering has some minor impacts on the probability of 'voluntary' tenders. In particular, in cases where the incumbent seller is one of the small 'other' sellers and the mandatory tendering regulation is not yet binding, the ex-ante probability of an auction occuring in each period decreases from 19.6% to 17.3%. However, by far the largest impact of the policy is that it will introduce a significant number of mandatory tenders. Overall, the average occurrence rate of tenders increases from 8.6% to 15.1%.

As would be expected from the increased bids of the majority of sellers, mandatory tendering increases average auction prices by around 1.9%. The magnitude of the expected price increase is relatively similar regardless of the winner of the auction. Importantly, this increase in expected auction prices directly translates into increased incumbent offer prices, since the maximum incumbent offer that a buyer is willing to accept is impacted one-to-one by the expected price that would be achieved from rejecting the offer and holding an auction. Therefore, mandatory tendering also has the effect of increasing the average prices offered by incumbent sellers during the incumbent offer stage of each period. This price increase is partially mitigated by the fact that the frequency of tenders increases, since the auction process produces lower prices on average compared to the incumbents' initial offers. Even after taking this effect into

account, however, mandatory tendering increases average prices over the long run by 0.4%.

One caveat to these results is that the exact magnitudes of each of the effects are dependent on the discount rate assumption. However, the qualitative results would seem to clearly indicate that the impact of mandated tendering regulation would be mixed; while it may achieve the regulators' aims of more frequent tenders for audit clients, the impact on market concentration may be limited and there may be some unintended consequences in the form of higher prices.

Several industry sources that have examined the impact of policy interventions in this market would seem to confirm these qualitative results. Commenting on the earlier implementation of mandatory tendering in the UK, Agnew (2016) suggests that while the regulation has increased the number of tenders, this has resulted in a 'merry-go-round' effect between the Big Four auditors and has had little impact on the market shares of smaller firms. Past reports by financial regulators in the US on the potential impact of mandatory auditor rotation also raised concerns that such policy interventions would result in increased prices (GAO 2003). These sources therefore provide some external validation of the results of the counterfactual experiment.

1.7 Conclusion

This paper set out an infinite horizon model of repeated auctions, where the incumbent seller in each period has the option to make an initial price offer to the buyer which, if accepted, means that the parties can avoid an auction in that period. The auction is usually costly for both of the parties, with the buyer incurring some auction costs and the seller losing cost advantages built up as a result of the incumbency. However, the buyer's auction costs are stochastic and ensures that auctions do occur with a positive probability. The ability of the incumbent to make an initial offer to the buyer to avoid an auction introduces a dynamic component to the auction game, as the bidders in the auction must take into account not just the payoffs from the current period but also

the value of being the incumbent in the next period.

The model is applied to the statutory audit industry in the UK. The results show that the dynamic aspect of the auction is an important part of the auction game and has the effect of lowering each seller's bids below their short-term costs. With an assumed discount rate of 0.9, the model finds that the bids in the auction are lower than the bidders' underlying costs by 34% to 45%. The dynamic aspects of the auction game also has the effect of magnifying asymmetries between bidders, as the ex-ante most efficient bidders tend to value the incumbency status more, and therefore tend to discount their bids more heavily in the auction.

The model has some policy implications, particularly with regards to the recently introduced regulations in the UK which mandate a tender for company auditors at least every ten years. While this would ensure that more auctions occur in the market and would marginally reduce market concentration, this also limits the value of the incumbency status and therefore has the effect of raising most sellers' bids in the auction process, ultimately raising the auction prices. With an assumed discount rate of 0.9, the model predicts that auction prices would increase by around 2%. This increase is partially, but not fully, mitigated by the increased frequency of tenders which tend to produce lower prices than incumbents' offers to bypass tenders.

One caveat to the model and the policy implications set out above is that the exact magnitudes of the policy experiments are dependent on the assumed discount rate, which the estimation does not separately identify. The non-identification of the discount rate is a common problem in dynamic models, as found in Rust (1994) and Jofre-Bonet and Pesendorfer (2003). In this particular case, the discount rate is not identified because the data used to estimate the model is entirely from the period prior to any policy interventions that would induce a reaction from bidders that are dependent on the bidders' discount rates. However, the recently introduced regulations in the UK are the exact type of interventions that would induce this type of reaction, and it is possible that further data from the UK after the full implementation of the regulations will identify the model more completely.

Appendix A: Derivation of the incumbency valuation

This appendix sets out the derivation of the expression for the ex-ante value functions as given in (1.34). Start with the incumbency valuation defined in terms of the valuation functions:

$$\Phi_j = \delta \left\{ V_j(j) - \frac{\sum_{k \neq j} s_k V_j(k)}{\sum_{k \neq j} s_k} \right\} \tag{1.41}$$

Rearranging yields:

$$\begin{aligned} \Phi_j &= \delta \frac{V_j(j)(1-s_j) - \sum_{k \neq j} s_k V_j(k)}{(1-s_j)} \\ &= \delta \frac{\nu_j(j) - \sum_{k=1}^{J} s_k V_j(k)}{(1-s_j)} \end{aligned}$$

$$\delta \sum_{k=1}^{J} s_k V_j(k) = \delta V_j(j) - (1-s_j)\Phi_j \tag{1.42}$$

Substituting into the definition of $V_j(h=j)$ as given in the first part of (1.34):

$$\begin{aligned} V_j(j) &= (1-\lambda_j)\pi_j^I + \lambda_j s_j \pi_j^A + \delta \left\{ (1-\lambda_j) V_j(j) + \lambda_j \sum_{k=1}^{J} s_k V_j(k) \right\} \\ &= (1-\lambda_j)\pi_j^I + \lambda_j s_j \pi_j^A + \delta(1-\lambda_j) V_j(j) + \delta\lambda_j V_j(j) - \lambda_j(1-s_j)\Phi_j \end{aligned}$$

$$\begin{aligned} V_j(j) - \delta(1-\lambda_j) V_j(j) - \delta\lambda_j V_j(j) &= (1-\delta) V_j(j) \\ &= (1-\lambda_j)\pi_j^I + \lambda_j s_j \pi_j^A - \lambda_j(1-s_j)\Phi_j \end{aligned}$$

$$V_j(j) = \frac{1}{1-\delta}\left\{(1-\lambda_j)\pi_j^I + \lambda_j s_j \pi_j^A - \lambda_j(1-s_j)\Phi_j\right\} \tag{1.43}$$

Which yields the first part of (1.35). Substituting (1.42) into the definition of $V_j(h \neq j)$ as given in the second part of (1.34):

$$\begin{aligned} V_j(h \neq j) &= \lambda_h s_j \pi_j^A + \delta\left\{(1-\lambda_h)V_j(h) + \lambda_h \sum_{k=1}^{J} s_k V_j(k)\right\} \\ &= \lambda_h s_j \pi_j^A + \delta(1-\lambda_h)V_j(h) + \delta\lambda_h V_j(j) - \lambda_h(1-s_j)\Phi_j \end{aligned}$$

$$\begin{aligned} V_j(h) - \delta(1-\lambda_h)V_j(h) &= (1-\delta+\delta\lambda_h)V_j(h) \\ &= \lambda_h s_j \pi_j^A + \lambda_h \frac{V_j(j)}{(1+r)} - \lambda_h(1-s_j)\Phi_j \end{aligned}$$

$$V_j(h) = \frac{1}{1-\delta+\delta\lambda_h}\left\{\lambda_h s_j \pi_j^A + \lambda_h \frac{V_j(j)}{(1+r)} - \lambda_h(1-s_j)\Phi_j\right\} \tag{1.44}$$

Which yields the second part of (1.35).

Appendix B: Dataset description and cleaning

The source for the dataset used in the estimation is the database of audit fees compiled by the UK Competition Commission ("UKCC") during its investigation into the statutory audit market which closed in 2014.[20]

The information on audit fees for UK-based companies is in the public domain and can be found in company filings or commercial databases. However, the UKCC found

[20]The database and and other case documents can be found at: https://www.gov.uk/cma-cases/statutory-audit-services-market-investigation

during its investigation that these data sources were typically not reliable as they did not show audit fees on a like-for-like basis. The UKCC therefore constructed its own database based on inputs from all major UK auditors. The database covers all companies that were either constituents of the FTSE 350 Index at any point during the period 2001 to 2011 or constituents of the Top Track 100 during the period 2006 to 2011, and contains data on financial years 2000 to 2011 for all companies covered.[21] In total, the database contains 7,959 observations on 856 companies.

Several company characteristics shown in the database formed the buyer-level covariates for the audit costs: total assets; total turnover; whether the company was a public listed company at the time of the audit; and whether the audit reflected a time period significantly shorter or longer than one year. The database contained some additional company characteristics which may have been suitable controls for costs such as total company equity, inventories and more detailed breakdowns for assets and turnover. However, these variables suffered from large numbers of missing observations and were therefore excluded from the model.

The database shows audit fees under several different headings: Company audit fees; Subsidiary audit fees; Audit-related fees; Total audit and audit-related fees; and Non-audit fees. These reflect the categories under which various fees paid to the auditing accountancy firm may be recorded in company filings, although not all companies break down the fees to these categories and some may classify fees differently from others. Company audit fees and subsidiary audit fees reflect the 'cost audit' services of auditing the financial statements of the parent company and the subsidiaries respectively. Audit-related fees are fees for services that do not fall into the definition of core audit but are nevertheless necessary to fulfil the statutory duties of the auditor. Total audit and audit-related fees are defined as the sum of company audit fees, subsidiary audit fees and audit-related fees.

The estimation in the main body of the paper used the core audit fees as the measure

[21] The FTSE 350 Index includes the largest 350 companies on the London Stock Exchange main market by market capitalisation. Top Track 100 includes the 100 largest private companies in the UK by revenues.

of audit pricing, due to the fact that this measure refers to well-defined and relatively comparable services across companies and over time. A minority of observations in the database showed missing values for the individual components of the core audit fees but did contain information on the total audit and audit-related fees. In these cases, the total audit and audit-related fees were used as a proxy for the core audit fees.

Chapter 2

Estimating the impact of endogenous seller entry on online marketplace prices

2.1 Introduction

Online marketplaces that act as platforms for buyers and sellers to find each other are important channels of distribution for many different types of goods and services. Platforms for bringing buyers and sellers together are of course not specific to the Internet and predates it. However, online marketplaces are potentially capable of bringing together larger numbers of participants at lower search costs compared to more traditional technologies. For consumers, this can potentially mean a wider selection of firms to choose from, as well as lower prices due to the stronger competition between firms.

This study empirically examines one potential form of such consumer benefit by quantifying the impact of the number of competing sellers on the prices faced by buyers, in an online marketplace environment where prices are determined through a decentralised searching and bidding process by sellers. The data for the analysis is sourced from AnyVan, an online marketplace for delivery and removals services.

The AnyVan platform provides a unique advantage for examining the impact of supplier competition on prices in that it is an environment where varying numbers of sellers engage in price bidding activity for individual buyers rather than vice versa. This is in contrast to most other major online marketplaces which operate as product auctions (such as eBay), use posted prices (Airbnb) or use centralised prices (Uber). This means that the variation in the degree of competition between sellers for each consumer can be observed directly, rather than having to be imperfectly inferred based on the presence of other comparable sellers on the platform. Compared to most online marketplaces, this online marketplace also deals wth a relatively homogeneous set of services rather than a wide range of different products. This allows me to analyse all observed listings from the platform in a single pricing equation rather than having to restrict the analysis to a specific subset of listings for homogeneous products. This gives more confidence that the results are representative of the online marketplace as a whole, rather than only a specific type of products or services within it.

I estimate the relationship of interest between the number of competing sellers and prices based on a reduced form model. Under the assumption that the number of competing sellers is random or exogenous, a simple ordinary least squares (OLS) estimates of a price regression would give consistent estimates of the impact of the number of sellers on prices. In practice, the empirical literature finds that the entry decisions by buyers or sellers to enter the bidding process for online products or services are typically endogenous (Reiley 2005, Bajari and Hortaçsu 2003). Previous empirical studies have found that OLS estimates that do not account for the endogenous nature of entry can give severely biased estimates of the true relationship between competition and prices (Newberry 2015, Onur and Velamuri 2014).

To account for endogenous entry by sellers, I take advantage of the fact that my dataset includes practically all listings available on the platform during the sample period. This allows me to construct variables that reflect the number of similar alternative listings that the sellers could have entered instead of the relevant listing, which I use as instruments in a two-stage least squares (2SLS) estimation. The two instruments I construct

are the total number of similar listings (defined by delivery type and geography) created on the platform on the same day, and the amount of time between the creation of the relevant listing and the creation of the next listing of the same delivery type within the same geography. Both instruments would be expected to affect the number of sellers competing for each listing, while being independent of unobserved listing-level characteristics that could effect sellers' entry decisions.

The 2SLS estimates show significant and negative impacts of additional sellers on prices, equivalent to a price reduction of 30 to 50 percent as the number of sellers increases from 1 to 5, and a price reduction of 55 to 65 percent as the number of sellers increases from 1 to 10. In contrast, simple OLS estimates imply negligble impacts of additional sellers on final prices, confirming the importance of accounting for the endogenous nature of entry. The qualitative results are robust to changes in functional form and the exact delineation of the instruments used.

This paper is related to the existing literature on the impact of competition on online marketplace prices, the majority of which relies on data from product auctions on the online auction site eBay. Onur and Velamuri (2014) examine the relationship between the number of bidders and prices in a sample of eBay auctions for calculators, using a 2SLS approach to account for endogenous entry. Newberry (2015) also uses a 2SLS approach to examine the impact of changes in the number of eBay listings (i.e. sellers) on prices, using a sample of eBay auctions for Chevrolet Corvettes. Lucking-Reiley et al. (2007) examine the impacts of various eBay auction characteristics on final prices, and while the authors do not directly examine the impact of the number of bidders, they note that the entry of bidders is endogenous and could be the mechanism by which various auction characteristics affect final prices. Giray et al. (2009) and Bajari and Hortaçsu (2003) estimate structural models of auctions which explicitly incorporate the endogenous entry processes of bidders, again based on a sample of eBay auctions for a narrowly defined set of products. Other recent empirical studies of price determination using eBay data include Einav et al. (2016b) and Einav et al. (2015), which look at the impact of eBay sales formats on prices. Notably, unlike the previously

mentioned studies, these authors work with extremely large scale datasets that include all eBay auctions from an extended period of time, although their identification strategy uses a smaller subset of products for which multiple sales formats are observed on a comparable basis. Hasker and Sickles (2010) and Bajari and Hortaçsu (2004) provide reviews of the existing literature on eBay auctions.

Aside from eBay, there are a smaller number of empirical studies which examine decentralised price determination in other online marketplace settings, mostly under posted-price systems. Farajallah et al. (2016) examine determinants of prices on a car-sharing platform which uses a posted-price system rather than a bidding system. The authors find that the level of experience of the seller has a significant downward impact on prices, and also that the ethnicity of the driver effects prices. Jolivet et al. (2016) examine the impact of seller reputation on prices of books, CDs, video games and DVDs in a French e-commerce platform which also uses a posted-price system.

The remainder of the paper is organised as follows. Section 2.2 sets out a description of the AnyVan platform and descriptive statistics on the dataset. Section 2.3 explains the empirical framework for estimating the relationship of interest between the number of competing sellers and prices. Section 2.4 presents the results, and Section 2.5 concludes.

2.2 Data source and descriptive statistics

2.2.1 Description of the AnyVan platform

AnyVan is an online marketplace which acts as a platform for bringing together buyers and sellers of delivery and removals services. The majority of the services traded on the platform are for generic deliveries of relatively small packages which can be fulfilled by smaller scale 'man and van' type delivery providers. However, the platform also accommodates more specialised transport services such as commercial freight, vehicle deliveries and livestock deliveries. The platform originally launched for the UK market in October 2009, and had attracted around 24,000 job listings in total by the end of

2010. Since then it has grown rapidly in size, attracting around 500,000 listings in total during 2011 and reaching annual listing numbers above one million by 2014. Since its launch it has also spread to several other European countries.

The buyers on the platform who require delivery services create a delivery job listing, which contains the details of the services required including the origin and destination locations, the time requirements for the delivery and a text description of other aspects of the required delivery service. Unlike other marketplace platforms such as eBay, the buyers do not set a starting price or a reserve price. The sellers on the platform browse through the job listings and are free to submit bids on any listing. All sellers can observe the identities of the other sellers who have already bid on any particular listing and the bids that they have submitted. The platform does not operate as a formal auction, and there are very few restrictions on the sellers' bidding decisions. While the listing is active on the platform, sellers can submit bids at any price level including at a level above the current lowest bid. Sellers can also update any of their previous bids at any point in time in response to the bidding activities of other sellers, or withdraw their bids. Despite the relative lack of structure on the bidding process, the bids in many listings follow a process that roughly resembles an English auction, with multiple rounds of decreasing bids as sellers undercut each other in prices.

The buyers are also not restricted in their purchase choice and can choose to accept any active bids at any point after the listing is created, or choose not to accept any bid. If a buyer accepts any of the existing bids, however, both parties are then committed to the transaction. The platform discourages buyers and sellers from reneging on an agreed transaction by charging the buyer a deposit for accepting a bid, and punishing any sellers who do not deliver on their bids.

There are no fees for buyers or sellers to sign up to the platform, create listings or submit bids. The platform derives its revenues by charging a percentage-based fee for successful transactions only, which is applied uniformly for all sellers and across all types of deliveries. Bids on the platform are displayed inclusive of such fees, and can therefore be understood as the sellers' pricing offer plus a percentage fee for the

platform itself.

2.2.2 Data collection process

The AnyVan website allows users to browse through all historical listings that have been created on the platform, which makes it possible to collect data for the analysis directly from the website. The dataset for the analysis was collected using an automated web-scraping program which operated by loading each individual listing from a sample period, identifying a number of key variables for each listing and saving them into a database. The web-scraping process recovers data on approximately 98 percent of all listings during the period, and therefore can be treated as being representative of all listings on the platform during the sample period.[1] This method was used to collect data on all types of listings on the platform for the month of April 2015.

While this data collection method can access data on nearly all of the listings on the platform, it does have some limitations. In particular, the data which can be collected is confined to the information on the listing which is publically displayed on the AnyVan website. This does include all information relating to the delivery which would be observed by the sellers when they are submitting their offers, such as the origin and destination locations and the text descriptions of the delivery written by the buyer. The sellers who submitted offers on any of the listings can also be individually identified. However, the publicly accessible portion of the AnyVan website only provides information on the current lowest bid for each listing, rather than providing the full bidding history for all bidders. Since the data was collected several months after the creation of the listings by which time all bidding activities had concluded, this means that for each listing only the single lowest bid value is observed. Furthermore, the publicly accessible portion of the website does not give information on the buyer's final purchase choice. The lowest observed bids therefore may not equal the final transaction prices, if the buyer chose to accept a bid other than the lowest bid (in which case the

[1]The program fails to recover the data if the listing is missing information in some key fields, particularly the origin and destination locations.

final transaction price is higher) or if the buyer did not accept any of the bids (in which case there is no transaction price). However, since submitting a bid contractually binds the seller to carrying out the delivery if it is accepted, the observed bids can still be interpreted as credible pricing offers. Moreover, for most types of deliveries found in the data the scope for quality differentiation would seem to be limited, therefore it seems unlikely that the buyers would accept a bid which is significantly higher than another available bid. I therefore take the observed lowest bid values as reasonable approximations to pricing offers or final transaction prices, and refer to them as prices throughout for brevity.

2.2.3 Dataset and descriptive statistics

I used the data collection methodology detailed above to gather data on 64,653 listings for the sample period. Of these, deliveries other than domestic UK deliveries were excluded for the purposes of the analysis. Furthermore, listings which attracted no bidders and therefore did not have any information on prices were also excluded from the analysis. After some additional exclusions for missing data issues, the final dataset contained 58,152 listings.[2] For each listing, the dataset includes information on the date and time the listing was created, the delivery type category as selected by the buyer, the origin and destination locations, the number of unique bidders and the price.

Among the product categories observed in the data, the largest category is 'Furniture delivery' which accounts for the majority of the observations. The text descriptions that accompany the listings suggest that this category not only includes deliveries of furniture but also acts as a catch-all category for other listings that do not fall into a more specialised category. Aside from furniture delivery, the largest product categories include 'Home move', 'Cars / Vehicles' and 'Motorcycle delivery'. Converting the origin location variable to UK NUTS 1 region categories shows that all major UK regions are

[2]Listings with no bidders were not included in the dataset used for the pricing regressions but were taken into account when constructing measures of alternative listings for use as instruments, the calculation of which is explained below. Listings with missing data issues were also included in the construction of instrumental variables whenever possible.

reasonably well represented in the data.[3] Table 2.1 shows the number of observations in the data divided by product categories and the origin locations.

	N	%	UK population %*
Whole sample	58,152	100	
Distance:			
Less than 10 miles	14,960	26	
10 to 100 miles	21,042	36	
More than 100 miles	22,150	38	
Category:			
Furniture delivery	40,105	69	
Home move	7,328	13	
Cars / Vehicles	4,130	7	
Motorcycle delivery	1,985	3	
Other			
Origin UK region (UK NUTS 1):			
North East	1,951	3	4
North West	6,365	11	11
Yorkshire and the Humber	4,423	8	8
East Midlands	3,932	7	7
West Midlands	5,177	9	9
East of England	5,399	9	9
London	11,350	20	13
South East	8,527	15	14
South West	5,598	10	8
Wales	1,801	3	5
Scotland	3,298	6	8
Northern Ireland	331	1	3

* UK NUTS1 total population, Eurostat

Table 2.1: Sample distribution

Table 2.2 presents descriptive statistics on the data. The mean price in the data is around 120 GBP, with the distribution showing a long right tail of higher-priced listings up to a maximum of around 6,000 GBP. Some major delivery type categories show significantly different pricing patterns compared to the remaining sample. In particular, home moves are shown to be much more expensive than other types of

[3]The Nomenclature of Territorial Units for Statistics (NUTS) system is defined by the European Union and divides the UK into 12 geographical regions as shown in Table 2.1.

deliveries, with a mean price of 213 GBP and a higher standard deviation. As would be expected, longer delivery distances are associated with higher prices, although this effect is very slight when comparing deliveries of distances below 10 miles (mean price 89 GBP) against deliveries of distances between 10 and 100 miles (mean price 95 GBP). For deliveries of distances above 100 miles, the mean price is much higher at 162 GBP.

The average number of bidders per listing observed in the data (conditional on attracting at least one bidder) is 6.9. Again, the major delivery type categories show some variation, with home moves attracting an average of 14 bidders and car deliveries attracting an average of 3.4 bidders. Distance is also shown to be negatively correlated with number of bidders, with deliveries below 10 miles attracting 10.4 bidders on average while deliveries above 100 miles attracting only 4.7 bidders on average.

Sample / Variable	N	Mean	SD	Min	P25	Median	P75	Max
Prices (GBP)								
Whole sample	58,152	118.7	129.8	16	50	76	141	5,958
Furniture deliveries	40,105	91.1	92.4	16	45	64	100	5,228
Home moves	7,328	213.4	213.0	20	103	168	252	5,958
Cars / Vehicles	4,130	216.8	137.6	21	119	193	277	1,808
Motorcycle delivery	1,985	122.6	76.0	25	76	112	147	1,580
Other product categories	4,604	119.4	128.9	19	50	82	135	1,808
Less than 10 miles	14,960	89.1	104.3	17	39	56	114	5989
10 to 100 miles	21,042	94.8	88.1	17	49	68	110	3,434
More than 100 miles	22,150	161.5	163.0	16	68	110	200	3,038
Bidders								
Whole sample	58,152	6.9	5.5	1	3	5	9	36
Furniture deliveries	40,105	6.3	4.5	1	3	5	8	32
Home moves	7,328	14.0	7.1	1	8	14	19	36
Cars / Vehicles	4,130	3.4	2.0	1	2	3	5	12
Motorcycle delivery	1,985	3.8	2.0	1	2	4	5	14
Other product categories	4,604	5.1	3.5	1	3	4	7	25
Less than 10 miles	14,960	10.4	6.9	1	5	9	15	35
10 to 100 miles	21,042	6.8	4.8	1	3	6	9	36
More than 100 miles	22,150	4.7	3.4	1	2	4	6	30
Delivery distance (miles)	58,152	89.5	93.0	0.0	9.4	61.6	145.2	635.2
Listings in market	58,152	128.6	103.1	1	31	119	191	415
Listing time interval	58,137	92.2	530.4	1.0	2.2	6.6	26.2	32,599.8

Table 2.2: Descriptive statistics on sample

The dataset allows me to look at the relationship of interest between the number of bidders and prices in a descriptive sense before attempting to control for other factors affecting prices. Figure 2.1 plots the average prices observed in the data conditional on the number of bidders, dividing the sample into three categories based on the delivery distance (less than 10 miles, between 10 and 100 miles, over 100 miles). As shown in the figure, instead of a negative relationship which would be expected, the data exhibits a U-shaped relationship between the number of bidders and prices. For bidder

numbers lower than 5, a higher number of bidders is associated with lower prices. For bidders numbers greater than 5, however, higher number of bidders is associated with higher prices, often higher than the average prices observed for listings with only a single bidder. This counterintuitive finding is likely indicative of endogeneity problems in the bidder number variable, which highlights the importance of accounting for the endogenous nature of entry when estimating the true shape of this relationship.

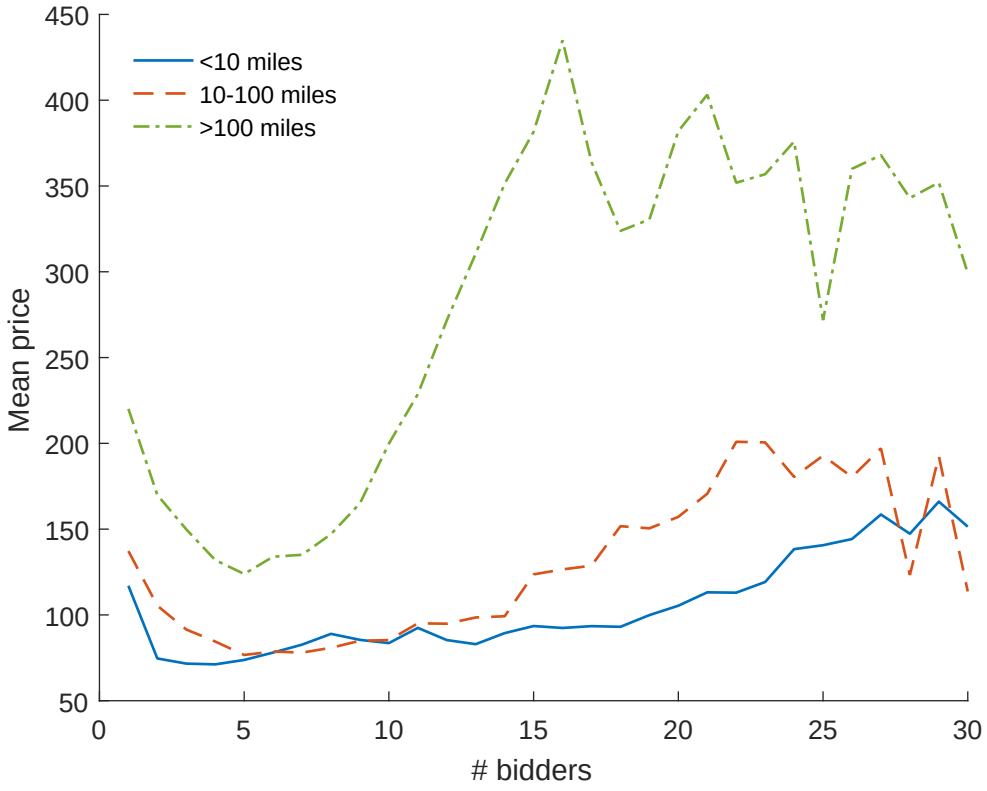


Figure 2.1: Average prices by number of bidders

2.3 Empirical framework

The relationship between the price and the number of bidders could be estimated based on a regression model of the following form:

$$ln\left(P_i\right) = f\left(N_i\right) + \mathbf{X_i}\beta + \varepsilon_i \tag{2.1}$$

where P_i is the price, N_i is the number of bidders, $\mathbf{X_i}$ includes all other observable

listing characteristics that could influence the price, and $f(\cdot)$ is the relationship of interest.

The dataset contains several variables that would be expected to affect the underlying costs of fulfilling the listing, and should therefore be included in $\mathbf{X_i}$. I include the delivery distance and its squared value to account for increasing costs of completing deliveries over longer distances. I also include a dummy variable indicating deliveries from Great Britain to Northern Ireland or vice versa to account for the higher costs that would be expected for such deliveries. Finally, I also include product category and origin region fixed effects.

In the context of online product auctions, Onur and Velamuri (2014) used a simple linear specification for $f(\cdot)$, the relationship between the number of bidders and prices (in logarithms). Newberry (2015) also used a simple linear specification when examining the relationship between number of competing sellers (number of available auctions for similar products) and prices. I include the same linear specification in my regressions and also add an additional specification with a logarithmic relationship for the number of bidders, which accounts for potential decreasing impacts of additional bidders on prices. In addition, Appendix A sets out a robustness check based on a quadratic specification for the number of bidders which is intended to capture non-linear effects more generally. The qualitative results remain unchanged.

Regardless of the specification used, a simple OLS regression of the relationship between the number of bidders and prices can potentially give inconsistent estimates of the true relationship. This is because the decisions by individual sellers to participate in the auction is likely to be affected by various auction characteristics, and therefore the number of observed bidders is likely to be endogenous and correlated with both observed and unobserved characteristics of the individual auctions. Several previous empirical studies have confirmed that the number of bidders in online auctions is endogenous and correlated with observed auction characteristics, such as reserve prices, minimum bids, book values and seller feedback scores (Reiley 2005, Bajari and Hortaçsu 2003). Previous reduced-form studies of the relationship between the number

of eBay participants and prices have also confirmed that entry decisions into online auctions by both buyers and sellers are endogenous, and that a simple OLS regression that does not account for endogenous entry can significantly understate the true effect of number of bidders or sellers on prices (Newberry 2015, Onur and Velamuri 2014). Evidence from auctions in more traditional industries also confirm that participation decisions in auctions are likely to be endogenous. Groeger (2014), Athey et al. (2011) and Li and Zheng (2009) all examine data from US government procurement auctions, and present auction models that include endogenous entry decisions by potential bidders to enter each auction subject to an entry cost. Each of these studies find that the decision by potential entrants to participate in auctions is endogenous with respect to various auction and bidder characteristics.

The nature of the services traded on the AnyVan platform would further seem to suggest that the number of observed bidders could suffer from endogeneity. The delivery listings on the AnyVan platform are very heterogeneous in terms of delivery requirements and likely costs, and it is likely that the set of potential bidders who can bid on each observed listing varies widely depending on the product category, the location and the delivery requirements. This is in contrast to many of the existing empirical studies of online auctions, which typically examine auctions for a narrowly defined set of homogeneous products. Furthermore, the decision to submit a bid on a listing is costly for the bidder, both in terms of the time taken to review each listing and estimate the likely costs involved in fulfilling the delivery requirements as well as the commitment value of having to fulfill the delivery if the offer is accepted. The bidder numbers shown in Table 2.2 confirm that differences in delivery characteristics can have significant effects on the average observed number of bidders, which suggests that endogenous entry is an important factor on the AnyVan platform.

I take advantage of the fact that the dataset contains an almost complete observation of all listings on the platform for the relevant period to construct instruments for the number of bidders based on measures of the availability of suitable alternative listings. Firstly, I use as instruments the total number of listings in the same market as the

listing, which I define as the combination of date, product category and the UK NUTS 1 geographic region for the origin location. This is intended to be a broad measure of the number of alternative listings with similar characteristics that a transport supplier could bid on instead of the relevant listing. Secondly, I use the listing time interval, defined as the amount of time between the creation of the listing and the creation of the next listing within the same product category and UK NUTS 1 geographic region. This captures the narrower impact on bidder entry when an alternative listing with similar characteristics appears on the website around the same time as the relevant listing.[4] Both instruments are used in the 2SLS estimation in logarithms. Table 2.2 sets out the descriptive statistics on the two instruments.

The two instruments would be expected to be correlated with the number of bidders for each listing since they reflect the number and closeness of alternative listings that the sellers could have chosen to enter instead of their actual entry choices. The total number of listings in the market would be expected to have a negative impact on the number of bidders for each listing, since the sellers would likely face capacity constraints that prevent them from fulfilling all of the delivery jobs, or alternatively may have limited attention or time that prevent them from browsing through all listings. The listing time interval captures the same type of constraints in a narrower time window; a listing which is far removed in time from other similar listings may attract more attention from sellers, increasing the number of bidders. Furthermore, the amount of time between each listing and the next listing with similar characteristics would be expected to have a disproportionate impact on the number of bidders, due to the fact that the AnyVan website lists the available listings in reverse chronological order. Therefore, the listing time interval as defined reflects the amount of time for which the listing would be the highest ranked listing (out of the set of listings with similar characteristics) that sellers would see when browsing. Previous empirical studies confirm that the placements of listings in such search rankings have significant effects on the number of matches that

[4] A conceptually similar instrument was introduced in Onur and Velamuri (2014), which used the time interval between the closing times of eBay auctions and the closing times of previous auctions for similar items.

they attract (Ursu 2018, Goldman and Rao 2014).

The exogeneity assumption for the instruments requires that they are uncorrelated with any unobserved characteristics of the individual listings that influence the final prices. This would seem reasonable on the basis that both of the instruments are defined not by the decisions or characteristics of the buyer creating each listing, but the decisions of other buyers on the platform. One potential channel by which the exogeneity assumption may be violated is if buyers' decisions to create listings are based on the the observed decisions by other buyers with similar delivery requirements to create their listings. This would seem unlikely, however, since the AnyVan platform does not offer potential buyers any easy way of checking the overall level of activities on the platform on that day or checking for the existence of another buyer with similar delivery requirements to themselves.

2.4 Results

Specifications (1) and (2) in Table 2.3 show the results of estimating the relationship shown in Equation 2.1 via OLS under two different specifications for the impact of bidder numbers on prices. Both specifications find a statistically significant effect of bidder numbers on prices, but the relationship suggested by the estimated coefficients do not seem economically plausible. For easier comparison, Table 2.4 and Figure 2.2 show the relationship between bidder numbers and prices implied by the OLS estimates, with prices expressed in relative terms compared to the predicted prices when there is only a single bidder. The OLS estimates under the linear specification (1) imply that an increase in bidder numbers has a minor *positive* impact on prices, while the estimates under the logarithmic specification (2) imply that even a large increase in bidder numbers from 1 to 10 only decreases prices by around 10 percent.

	(1) OLS Price (log)	(2) OLS Price (log)	(3) 2SLS Price (log)	(4) 2SLS Price (log)
Endogenous variables: (Instruments: Listings in market (log), Listing time interval (log))				
Bidders	0.00400***		-0.0885**	
	(0.000617)		(0.0286)	
Bidders (log)		-0.0495***		-0.456***
		(0.00405)		(0.136)
Exogenous variables:				
Distance (log)	0.0204***	0.0190***	0.000278	0.0144**
	(0.00444)	(0.00442)	(0.00794)	(0.00474)
Distance (log) ^ 2	0.0255***	0.0247***	0.0197***	0.0200***
	(0.000778)	(0.000775)	(0.00198)	(0.00175)
NI-GB delivery	0.605***	0.597***	0.584***	0.539***
	(0.0312)	(0.0314)	(0.0353)	(0.0402)
Category fixed effects	Yes	Yes	Yes	Yes
Region fixed effects	Yes	Yes	Yes	Yes
N	58152	58152	58137	58137
R^2	0.431	0.432	0.173	0.304

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2.3: OLS and 2SLS price regressions

Aside from the bidder number variables, other control variables show more sensible estimates. Delivery distance is shown to have a significant positive effect on prices as expected, with the estimates showing a quadratically increasing relationship between (logarithm of) delivery distances and (logarithm of) prices. Deliveries from Great Britain to Northern Ireland or vice versa are shown to be significantly more expensive than other deliveries even after accounting for the typically longer distances, which is as expected given the likely additional costs of transport.

Specifications (3) and (4) in Table 2.3 show the results of estimating the same rela-

	(1) OLS	(2) OLS	(3) 2SLS	(4) 2SLS
Bidder variable used:	Bidders	Bidders (log)	Bidders	Bidders (log)
Number of bidders:			Relative prices:	
1	1.00	1.00	1.00	1.00
2	1.00	0.97	0.92	0.73
3	1.01	0.95	0.84	0.61
4	1.01	0.93	0.77	0.53
5	1.02	0.92	0.70	0.48
6	1.02	0.92	0.64	0.44
7	1.02	0.91	0.59	0.41
8	1.03	0.90	0.54	0.39
9	1.03	0.90	0.49	0.37
10	1.04	0.89	0.45	0.35

Table 2.4: Implied impact of bidder numbers on prices implied by OLS and 2SLS results

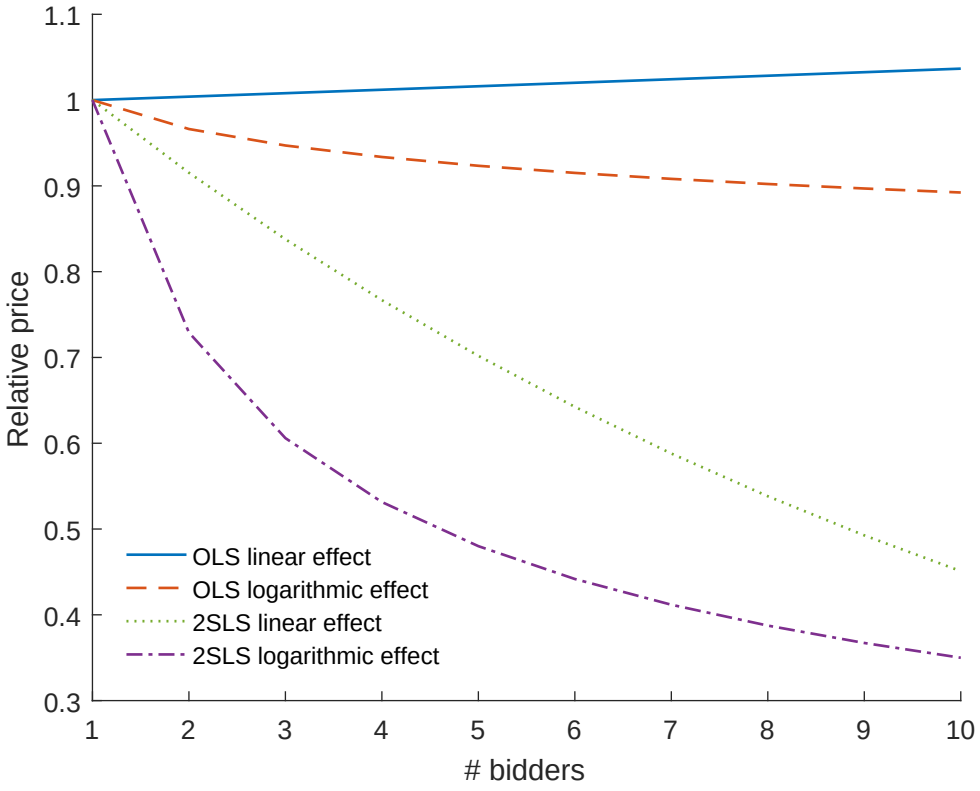


Figure 2.2: Implied impact of bidder numbers on prices implied by OLS and 2SLS results

tionship via 2SLS, using the number of listings in the market and listing time intervals as instruments. As before, both specifications find statistically significant effects of bidder numbers on prices. However, the relationships between bidder numbers and prices implied by the estimated coefficients now seem much more economically plausible. As shown in Table 2.4, the estimates imply that increases in bidder numbers have a sizeable negative impact on prices, with an increase in bidder numbers from 1 to 5 resulting in a reduction in price of 30 to 50 percent, and an increase in bidder numbers from 1 to 10 resulting in a price reduction of 55 to 65 percent, depending on the specification. Notably, the change in the implied relationship between bidder numbers and prices is similar in direction and magnitude under both specifications, which gives some confidence that the results are robust to the exact specification of the relationship. The estimated effects of explanatory variables other than the bidder numbers on prices remain relatively similar between the OLS and 2SLS results.

Table 2.5 sets out specification tests on the 2SLS regressions. Firstly, Wooldridge's tests of endogeneity reject the null hypotheses that the bidder number variables are exogenous, and suggest that the 2SLS is the more suitable approach. The first stage regression results confirm that the two instruments used (number of listings in the market, listing time intervals) have statistically significant and economically plausible impacts on bidder numbers. The total number of listings in the market has a negative impact on bidder numbers for individual listings which is significant at 5 percent confidence level, while the listing time interval has a positive impact on bidder numbers which is significant at 0.1 percent confidence level. The signs of both of these effects are as expected given the impact that the existence of similar alternative listings would have on sellers' decisions to bid on the given listing. The first stage F-statistics on the two instruments are also large, and confirm that the instruments together have significant impact on bidder numbers. Finally, the tests of over-identifying restrictions do not reject the null hypothesis that the instruments are exogenous.

	(3) 2SLS	(4) 2SLS
Bidder variable used	Bidders	Bidders (log)
Tests of endogeneity on bidder variable:		
Wooldridge's score test chi-squared	15.141	10.922
P-value	(0.0001)	(0.0010)
Wooldridge's regression-based test F	15.141	10.922
P-value	(0.0001)	(0.0010)
First stage regression summary statistics:		
Coefficient estimates:		
Listings in market (log)	-0.116 *	-0.0186*
Standard error	(0.0563)	(0.00917)
Listing time interval (log)	0.0618***	0.0125***
Standard error	(0.0131)	(0.00211)
First stage F statistic	16.823	23.937
P-value	(0.0000)	(0.0000)
Test of over-identifying restrictions on instruments:		
Wooldridge's score test chi-squared	0.0191	0.1346
P-value	(0.8900)	(0.7137)

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2.5: Specification tests for 2SLS

Together, the specification test results suggest that bidder entry decisions are indeed endogenous in this context, leading to biased estimates under OLS. The chosen instruments used in the 2SLS analysis appear to be suitable, being both exogenous to listing-level unobserved variables that influence prices while being significant drivers of bidder entry. A comparison of the OLS and 2SLS estimates shows that while the OLS estimates of the effects of bidder numbers on prices seem either economically incorrect or implausibly small, the 2SLS estimates are economically plausible in both direction and magnitude, as well as being consistent across the two specifications.

Bidder number variable	Marginal effect on log(price)	
	OLS	2SLS
Furniture delivery: ($n = 40105$)		
Bidders	-0.00529***	-0.309**
	(0.000825)	(0.0967)
Bidders (log)	-0.0924***	-1.074***
	(0.00488)	(0.215)
Home moves: ($n = 7328$)		
Bidders	0.0262***	0.00710
	(0.00124)	(0.0260)
Bidders (log)	0.255***	0.0694
	(0.0157)	(0.243)
Vehicle deliveries: ($n = 6115$)		
Bidders	-0.0539***	-0.0874*
	(0.00237)	(0.0364)
Bidders (log)	-0.192***	-0.302*
	(0.00808)	(0.123)

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2.6: Bidder number marginal effects on prices, by product category

Instrument	Marginal effect on bidder number variable	
	Bidders	Bidders (log)
Furniture delivery: ($n = 40105$)		
Listings in market (log)	-0.177	-0.0730***
	(0.0923)	(0.0164)
Listing time interval (log)	0.0387**	0.00878**
	(0.0148)	(0.00273)
Home moves: ($n = 7328$)		
Listings in market (log)	-0.583*	-0.0638**
	(0.228)	(0.0235)
Listing time interval (log)	0.115*	0.0120*
	(0.0495)	(0.00492)
Vehicle deliveries: ($n = 6115$)		
Listings in market (log)	-0.124	-0.0426
	(0.0721)	(0.0222)
Listing time interval (log)	0.0719***	0.0201***
	(0.0167)	(0.00520)

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2.7: 2SLS first stage regression results by product category

The qualitative results from this section continue to hold when the major product categories within the data are considered separately. Table 2.6 shows the marginal effects of bidder number variables (number of bidders or logarithm of number of bidders) on prices according to separate OLS and 2SLS regressions for three major product categories found in the data: furniture deliveries, home moves and vehicle deliveries (i.e. car and motorcycle deliveries). Table 2.7 shows the first stage regression results for the marginal effects of the instruments on bidder number variables (number of bidders or logarithm of number of bidders). For all three subgroups, the qualitative results from

the main body of the paper continue to hold; the OLS results tend to understate the impact of the number of bidders on prices, i.e. produce more positive estimates, while the 2SLS results imply stronger or more negative relationships between the number of bidders and prices. Moreover, the estimated directional impact of the instruments on the number of bidders is consistent for all three groups, with higher number of listings in the market reducing the number of bidders per listing, and longer time interval between listings increasing the number of bidders per listing. This gives some confidence that the same drivers of endogenous entry is being picked up across all types of deliveries listed on the platform.

However, there are some differences between the three major product categories in the magnitudes of the estimated impacts. For home moves, the OLS results imply much larger positive impacts of additional bidders on prices, while the 2SLS results imply smaller positive impacts which are not significant. For vehicle deliveries, the OLS results already find a significant negative relationship between bidder numbers and prices, and the magnitude of the difference to the 2SLS estimates are modest compared to other categories. These findings may suggest that the degree to which the endogenous nature of entry impacts the estimated parameters differs between the delivery categories; the listings in the home moves category appear to suffer relatively more from the endogeneity issue, while the vehicle delivery category suffers much less. This is in line with what would be expected from the nature of the different types of deliveries. Home moves are likely to have many other important cost drivers which are not picked up by the observed variables (e.g. the size of the dwelling), some of which may also be correlated with the entry decision and therefore increase the bias caused by not accounting for endogenous entry. Costs involved in deliveries of vehicles, on the other hand, are comparatively homogenous and well-defined by the observed variables, which decreases the degree to which the endogeneity of entry can bias the price regression estimates.

The consistent finding that the bias in the OLS estimates tends to be in the direction of underestimating the impact of competition on prices rather than overestimating the

impact is also in line with previous studies which have estimated similar relationships via 2SLS. Newberry (2015) examined the impact of the number of competing sellers on eBay product auction prices, and found that the OLS estimates implied an implausibly small and statistically insignificant impact of competition on prices while 2SLS estimates showed significant negative impacts of additional sellers on prices, with the magnitude of the impact depending on the exact delineation used to define the number of competing sellers across product space and over time. The results presented in this section match these findings, but use a more direct measure of the variation in the number of competing sellers, which removes the ambiguity involved in defining the measure of competing sellers. Also using eBay data, Onur and Velamuri (2014) examined the impact of the competing number of bidders (i.e. buyers) on product auction prices, and also found that OLS underestimated the positive impact of bidder numbers on prices by a factor of three compared to 2SLS.

2.5 Conclusions

This study set out an analysis of the relationship between the number of competing sellers and prices in the context of an online marketplace for delivery services. The results confirm that an increase in the number of sellers has a significant downward impact on prices. For example, listings with 5 competing sellers, the median number of sellers observed in the sample, are predicted to have prices that are around 30 to 50 percent lower compared to listings where only one seller is bidding. The analysis also confirms the importance of controlling for endogenous entry decisions when examining the impact of competition on prices. In line with previous empirical studies, the results show that not controlling for endogenous entry can severely understate the true impact of seller competition on prices or even reverse the estimated direction of the relationship.

The online marketplace setting examined in this analysis uses a decentralised price determination system where the bidding activities of sellers determine the final prices,

and involves transactions for a fairly homogenous set of services. This is in contrast to the settings for the majority of previous studies such as eBay, in which buyers undertake bidding activities and an extremely wide range of goods and services are traded. This is an advantage in examining the impact of seller competition on prices, since the variation in the degree of competition between sellers can be directly observed rather than imperfectly inferred from the presence of other sellers on the platform. Moreover, unlike the majority of previous studies based on eBay data, I am able to include all observed listings from the platform in a single pricing equation rather than having to restrict the analysis to a specific subset of listings for comparable products. This gives some confidence that the results are representative of the way that competition functions on the platform in general, rather than a smaller segment of the platform. The conclusions could also be extended to other online marketplaces with similar decentralised systems for price determination.

The empirical model used in this study took a reduced form approach to examining the relationship between bidder numbers and prices, rather than imposing any assumptions on the nature of the competition between sellers. While this does mean that the results are independent of any structural assumptions, it does also limit the power of the model to generate counterfactuals or policy implications for the online platform. It also imposes a linear or log-linear specification on the relationship between the number of bidders and prices, which is not justified by an economic model.

The reduced form approach also did not impose any structural assumptions on the relationship between bidders' entry decisions and the two instruments, the number of listings in the market and the listing time interval, and simply utilised the observed correlation between the instruments and the number of bidders. However, the fact that the number of bidders entering a specific listing is correlated in a predictable way with the availability of similar alternative listings is suggestive of a decision making process by sellers who optimise over the entry decisions for multiple listings simultaneously.

Following on from these points, Chapter 3 will deal with these shortcomings by presenting a structural model which encompasses both the competition between sellers at

the level of individual listings and the process which determines the number of sellers competing in each listing. The downside of this approach is that it requires many more structural assumptions compared to the analysis set out in this chapter. However, as will be discussed in Chapter 3, the model allows for a much wider set of counterfactuals, and can be used to derive richer implications about the nature of online platforms.

Appendix A: Robustness checks

Quadratic specification for bidder number variable

The regressions presented in the main body of the paper accounted for potential diminishing marginal effects of additional bidders on prices by including a specification that used a logarithm of the number of bidders. An alternative approach that would capture potential non-linear effects is to include a quadratic term for the bidder number variable. Compared to the logarithmic specification, the quadratic specification makes fewer assumptions about the shape of the non-linear relationship. Two downsides of this approach are that the estimates of the bidder number coefficients are greatly influenced by observations with unusually high values for bidder numbers, and that the specification introduces an additional endogenous variable which means that the 2SLS regression is no longer overidentified.

Tables 2.8 and 2.9 present the OLS (5) and 2SLS (6) estimates under the quadratic specification and the implied impacts of bidder numbers on prices. The OLS estimates imply very small negative impacts of additional bidders on prices for lower bidder numbers, the magnitude of which is similar to the effects implied by the OLS results under the logarithmic specification as shown in Table 2.4. The 2SLS estimates imply much larger negative impacts of additional bidders on prices, which has similar magnitude as the 2SLS linear effect shown in Table 2.4. However, the estimated coefficients for the bidder number variables are not statistically significant. This appears to be due to the fact that the quadratic functional form has a tendency to produce unrealistic predicted prices for extremely large bidder numbers. Specification (7) sets out the 2SLS estimates under the same regression structure excluding observations with 20 or more bidders. This produces statistically significant estimates of the impact of bidder numbers on prices, and the implied relationship between bidder numbers and prices remains of similar magnitude to the 2SLS results shown in Table 2.4 for suitably low bidder numbers.

	(5) OLS Price (log)	(6) 2SLS Price (log)	(7) 2SLS Price (log)
Endogenous variables:			
(Instruments: Listings in market (log), Listing time interval (log))			
Bidders	-0.0247***	-0.0758	-0.455**
	(0.00146)	(0.0980)	(0.143)
Bidders ^ 2	0.00131***	-0.000651	0.0285*
	(0.0000571)	(0.00475)	(0.0113)
Exogenous variables:			
Distance (log)	0.0233***	-0.00147	0.0484**
	(0.00434)	(0.0149)	(0.0170)
Distance (log) ^ 2	0.0248***	0.0199***	0.0240***
	(0.000763)	(0.00275)	(0.00437)
NI-GB delivery	0.594***	0.589***	0.428***
	(0.0316)	(0.0504)	(0.0715)
Category fixed effects	Yes	Yes	Yes
Region fixed effects	Yes	Yes	Yes
N	58152	58137	55711
R^2	0.436	0.158	.

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2.8: OLS and 2SLS price regressions

	(5) OLS	(6) 2SLS	(7) 2SLS
Number of bidders:	Relative prices:		
1	1.00	1.00	1.00
2	0.98	0.93	0.69
3	0.96	0.85	0.51
4	0.95	0.79	0.39
5	0.93	0.73	0.32
6	0.93	0.67	0.28
7	0.92	0.62	0.26
8	0.91	0.56	0.25
9	0.91	0.52	0.26
10	0.91	0.47	0.28

Table 2.9: Implied impact of bidder numbers on prices implied by OLS and 2SLS results

Alternative specifications for instruments

The two instruments set out in the main body of the paper used the UK NUTS1 regions as geographical delineations, which reflected potentially local nature of competition for delivery services. Tables 2.10 and 2.11 present the 2SLS estimates and specification tests based on an alternative instruments that instead use the UK as the geographical delineation. Therefore, the two instruments used are the total number of listings created within the same product category during the same day, and the listing time interval, now defined as the amount of time between the listing and the next listing within the same product category located anywhere in the UK. The qualitative results remain unchanged.

	(1)	(2)	(8)	(9)
	OLS	OLS	2SLS	2SLS
	Price (log)	Price (log)	Price (log)	Price (log)
Endogenous variables:				
(Instruments: Listings in market (log), Listing time interval (log))				
Bidders	0.00400***		-0.164***	
	(0.000617)		(0.0357)	
Bidders (log)		-0.0495***		-0.770***
		(0.00405)		(0.132)
Exogenous variables:				
Distance (log)	0.0204***	0.0190***	-0.0165	0.0107*
	(0.00444)	(0.00442)	(0.00971)	(0.00500)
Distance (log) ^ 2	0.0255***	0.0247***	0.0149***	0.0164***
	(0.000778)	(0.000775)	(0.00245)	(0.00173)
NI-GB delivery	0.605***	0.597***	0.570***	0.496***
	(0.0312)	(0.0314)	(0.0410)	(0.0440)
Category fixed effects	Yes	Yes	Yes	Yes
Region fixed effects	Yes	Yes	Yes	Yes
N	58152	58152	58152	58152
R^2	0.431	0.432	.	.

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2.10: OLS and 2SLS price regressions

	(8)	(9)
	2SLS	2SLS
Bidder variable used	Bidders	Bidders (log)
Tests of endogeneity on bidder variable:		
Wooldridge's score test chi-squared	56.829	49.982
P-value	(0.0000)	(0.0000)
Wooldridge's regression-based test F	56.668	49.848
P-value	(0.0000)	(0.0000)
First stage regression summary statistics:		
Coefficient estimates:		
Listings in market (log)	-0.418 ***	-0.0992***
Standard error	(0.0563)	(0.0134)
Listing time interval (log)	0.0628**	0.0102**
Standard error	(0.0232)	(0.00360)
First stage F statistic	18.132	35.593
P-value	(0.0000)	(0.0000)
Test of over-identifying restrictions on instruments:		
Wooldridge's score test chi-squared	1.4950	0.6032
P-value	(0.2214)	(0.4374)

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2.11: Specification tests for 2SLS

Chapter 3

Impact of scale and competition on an online peer-to-peer platform

3.1 Introduction

Online peer-to-peer platforms which bring together large numbers of small scale buyers and sellers to facilitate transactions have seen tremendous growth in recent years.[1] Online platforms have spread into many existing markets where buyers and sellers need to find each other (eBay, Amazon) and have also brought many smaller fringe suppliers into markets previously dominated by more traditional suppliers (Airbnb, Uber). Such platforms depend on the participation of two distinct groups of users to operate, and as such can be seen as archetypal examples of two-sided markets. Given the importance that the economic literature places on the scale effects of two-sided markets (Ellison and Fudenberg 2003, Evans and Schmalensee 2010), an empirical examination of the impact of scale on how these platforms function is of particular interest from a policy perspective.

[1] The existing literature uses several similar terms such as digital marketplaces, online marketplaces, online platforms and e-commerce platforms to refer to platforms that use new technologies to bring together buyers and sellers. As noted in Fradkin (2017), there is no universally accepted definition and delineating these firms from 'traditional' firms is not necessarily simple. I use the term 'online peer-to-peer platform' in this chapter to emphasise the fact that the buyer-seller matching and price determination examined in this paper occurs in a decentralised way via interactions between individual buyers and sellers.

This study presents a model of an online peer-to-peer platform and uses it to examine how the market outcomes on the platform are affected by entry of buyers and sellers and the resulting changes in the scale of the platform. The structure of the model reflects the observed features of AnyVan, an online platform for transactions of delivery and removals services, which also forms the data source for the estimation of the model parameters. The model sets out an environment where sellers engage in bidding competition within listings created by buyers to be the chosen supplier for a specific delivery job, and prices are determined in a decentralised way by the bidding process. This sets it apart from some other major online platforms such as Uber, where the platform itself controls prices in a centralised way with the aim of balancing supply and demand (Hall et al. 2016, Chen and Sheldon 2015). The structure of the platform is comparable to several other online platforms, especially those for transactions of less standardised services, such as TaskRabbit which was examined in Cullen and Farronato (2018).

I model a three-stage process that determines the matching between buyers and sellers and the terms of the resulting transactions. In the first stage, a large number of undifferentiated potential sellers decide on whether to enter and participate on the overall platform itself, which depends on their expected profits as a function of the number of buyers on the platform and the characteristics of the market. In the second stage, a stochastic process determines a set of linkages ('matches') between the buyers and sellers participating on the platform. These matches determine the set of sellers who submit bids on the listings created by each buyer. I use a Cobb-Douglas function to represent the match-generating process, with buyers and sellers participating in the market forming the 'inputs'. In the final stage, the sellers matched to each buyer compete in a second-price procurement auction which determines the chosen supplier and the price. This model structure allows me to examine the scale effects of the matching process, which has a large bearing on how the market outcomes are affected by the entry decisions of buyers and sellers.

I estimate the parameters of all three stages of the model jointly in a single step

by maximum likelihood, based on a dataset of bidding activities on vehicle delivery listings on the AnyVan platform. The results show that the platform-level matching process exhibits increasing returns to scale; increasing the size of the platform by increasing the numbers of participating buyers and sellers while holding their relative proportions constant results in a greater-than-proportional increase in the number of matches between buyers and sellers. The parameter estimates suggest that doubling both the numbers of buyers and sellers, for instance, would result in a 9 percent increase in the average number of sellers bidding on each listing, resulting in a greater proportion of fulfilled listings and lower prices for buyers.

The results have significant implications on how the market outcomes are affected by changes in the scale of the platform over the longer term. Under a free entry equilibrium for sellers, an exogenous increase in the number of buyers somewhat counterintuitively *reduces* prices and *increases* the proportion of buyers who will successfully find a seller. This is because sellers react to increased demand by entering the platform, and the resulting increase in scale generates a greater-than-proportional increase in the number of matches generated by the platform. These conclusions are in line with the positive scale and network effects often assumed in the theoretical literature on two-sided platforms (Weyl 2010, Rysman 2009, Rochet and Tirole 2006, Armstrong 2006), and show how these type of scale and network effects could arise naturally from a matching process between buyers and sellers.

This study is related to several strands of literature concerning online platforms. The paper is closest in spirit to Cullen and Farronato (2018), which examines the matching process between buyers and sellers in TaskRabbit, an online platform for small-scale labour services. The authors model the platform as a market for homogenous labour, and abstract away from sellers' bidding activities at the level of individual listings in favour of examining market-level prices and quantities. In comparison to this study, I model the competition between sellers at the level of individual listings explicitly, which allows me to make counterfactual price predictions based on a structural model as well as address the large heterogeneity in the service requirements of individual buyers.

This study is part of a growing literature that examines search, matching and price determination in online platforms. An overview of the literature can be found in Einav et al. (2016a). Some recent examples include Fradkin (2019) which examines the searching and matching process on the accommodation platform Airbnb, Dinerstein et al. (2018) which examines a change in the design of buyers' search processes on eBay, and Hall et al. (2016) and Chen and Sheldon (2015) on the dynamic pricing on the Uber platform and the seller entry responses to "surge" pricing. Farronato and Fradkin (2018) also examine the endogenous participation decisions of sellers on Airbnb and the resulting impact on market outcomes both on Airbnb and in competing hotel markets.

This study is also related to a large empirical literature on auctions, many of which also take advantage of the large amount of high quality data available from online platforms, in particular eBay (Backus and Lewis 2019, Giray et al. 2009, Nekipelov 2007, Bajari and Hortaçsu 2003).[2] Many empirical studies also examine the auction entry processes using data from more traditional industries (Groeger 2014, Athey et al. 2011, Li and Zheng 2009, Li 2005). These studies are typically based on a setting where an exogenous number of potential bidders make entry decisions over a number of independent auctions, which produces an endogenous outcome for the number of actual bidders for each auction. The model set out in this study can be interpreted as an extension of this structure where the relevant decisions of the sellers are their decisions to participate in the platform itself rather than enter individual auctions, and a separate stochastic process incorporating the sellers' decisions then determines the number of actual bidders for each auction within the platform. Incorporating the sellers' platform entry decisions into the model allows me to draw richer conclusions regarding the scale effects of the platform on auction outcomes, and also links the model's conclusions to the existing theoretical literature on two-sided markets, which typically examine the decisions of consumers to participate in the platform as a whole.

The remainder of this paper is organised as follows. Section 3.2 sets out a description of the platform and the nature of the bidding process, which informs the modelling

[2]See also the literature review given in Chapter 2.

decisions set out in later parts. Section 3.3 presents a structural model of an online platform. Section 3.4 discusses the data and the empirical methodology used to estimate the structural model. Section 3.5 sets out the estimation results. Section 3.6 discusses the implications of the results on the performance and growth of online platforms, and Section 3.7 concludes.

3.2 Description of the platform and data

3.2.1 Searching and matching on the AnyVan platform

AnyVan is an online marketplace which acts as a platform for bringing together buyers and sellers of delivery and removals services.[3] The buyers on the AnyVan platform create listings which provide information on their service requirements, such as the delivery origin and destination locations and a description of the objects to be delivered. The information on the origin and destination locations are required inputs when creating a listing, and the platform automatically calculates measures of delivery distance and estimated travel time and makes these available to the sellers. Many of the other types of delivery characteristics which would be expected to drive costs, however, are provided in the form of text descriptions in an unstructured way. These text descriptions are important because the platform allows for a very wide range of delivery requirements to be listed, and much of the crucial information on the delivery requirements would not be captured by standardised inputs alone. This means that unlike platforms for more standardised services such as taxis, AnyVan cannot use a centralised system to match buyers and sellers. Instead, AnyVan relies on decentralised searching activities by individual buyers to enable transactions to take place. The searching mechanism for buyers consists of a simple browsing system which lists all available listings in reverse chronological order. Sellers use this system to find suitable listings on which to submit bids.

[3] A general introduction to the AnyVan platform was given in Chapter 2, and this section will focus on some specific features of the platform which are particularly relevant to the modelling decisions in this chapter.

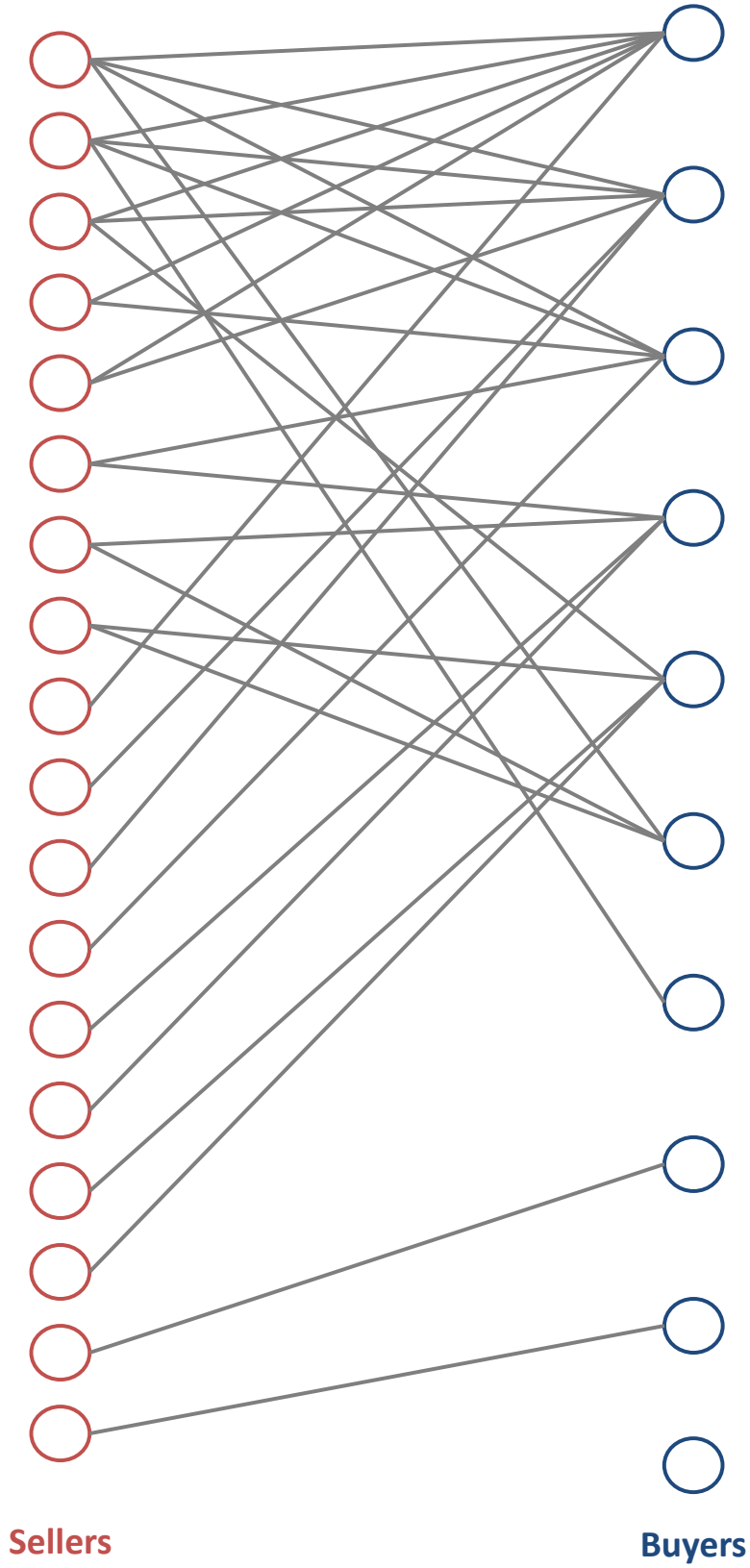
Figure 3.1 shows an example outcome of this searching process. The figure shows all observed sellers and buyers (listings) for car deliveries in the UK North-East region on 3 April 2015. Each buyer and seller is shown with a linkage if the seller submitted a bid on the buyer's listing. The figure shows that there is some variation in the number of listings that each seller chooses to bid on, and in the number of bidders that each listing attracts. In the sample, the observed sellers submitted bids on between 1 and 4 listings,[4] while the listings attracted between 0 and 6 bidders. Furthermore, the figure shows that the majority of the observed participants in the sample were interconnected by a series of bidding relationships. This justifies the modelling approach followed in this study, which will use the same combinations of delivery types, geography and date to define markets composed of buyers and sellers that are subject to the same process that generates matches between buyers and sellers.

3.2.2 Bidding process and prices on the platform

Once a buyer has created a listing for a delivery job on the platform, any of the sellers browsing the platform can submit a bid to fulfill the delivery service. The bidding process is not structured as a formal auction, and the platform places very few restrictions on the bidding process. Sellers can submit bids at any price level at any time, including bids that are higher than or equal to any previous bids, and can also update their previous bids at any point. Buyers not obligated to choose the lowest-priced bid, and can instead choose an alternative bid or choose not to accept any of the bids.

Despite the lack of structure imposed by the platform, the pattern of bidding actually observed in the data suggests that a second price auction may be a good approximation to the observed behaviour of bidders. As an illustration, Table 3.1 shows the bids submitted for an example listing for a car delivery which was created on 2 April 2015. For this listing, 6 sellers submitted a total of 21 bids over a period of two days, reaching a

[4]Note that sellers that do not submit any bids are unobserved by definition. This poses a problem for estimation, and will be discussed in later sections.



Car deliveries, UK North-East, 3 April 2015

Figure 3.1: Example of buyer-seller matches within a closely defined market

final lowest bid of 156 GBP compared to the first bid of 220 GBP. As shown in the table, the majority of the bids submitted consisted of matching or marginal undercutting of the existing bids by other sellers, with sellers' reactions to other sellers' bids occuring sometimes within seconds.

#	Seller	Date	Time	Bid (GBP)
1	A	2 April 2015	10:25	220
2	B	2 April 2015	10:30	217
3	C	2 April 2015	11:13	216
4	B	2 April 2015	11:15	211
5	C	2 April 2015	11:15	211
6	D	2 April 2015	12:01	223
7	E	2 April 2015	23:34	200
8	F	3 April 2015	18:14	198
9	E	3 April 2015	18:14	186
10	F	3 April 2015	18:14	185
11	F	3 April 2015	18:17	182
12	E	3 April 2015	18:17	180
13	F	3 April 2015	18:18	178
14	E	3 April 2015	18:25	174
15	F	3 April 2015	18:25	172
16	E	3 April 2015	18:26	168
17	F	3 April 2015	18:26	166
18	E	3 April 2015	18:26	162
19	F	3 April 2015	18:26	161
20	F	3 April 2015	18:26	158
21	E	3 April 2015	18:26	156

Table 3.1: Pattern of bidding for an example listing

This pattern of multiple rounds of undercutting between sellers is typical of the listings on the platform. In the sample of vehicle delivery listings used in this study, the average number of bidders among listings that attracted multiple bidders was 4, compared to average number of bids of 7. This suggests that in cases where multiple sellers are in competition for the same listing, there are typically multiple rounds of bidding as sellers submit lower bids in reaction to their competitors' bids. This produces a pattern of bidding which is similar in structure to an English auction.

One drawback of the data source and the data gathering process is that while the bidding activity of sellers can be observed, the final purchase choice of the buyer is not.

Therefore, while the lowest bid reached by the bidding process is observed, this value may not equal the actual transaction price, either because the buyer chose to accept a bid which was not the lowest bid, or because the buyer did not accept any of the bids. However, given the nature of the delivery service which is relatively homogenous, it would not seem likely that any buyer would accept a bid which is significantly higher than an alternative bid. Moreover, since any of the bids submitted would, if accepted, contractually oblige the seller to fulfill the delivery requirement, they can be treated as credible pricing offers that are informative of the underlying costs of the sellers. Based on these considerations, I treat the observed lowest bid values as resulting from a second price auction for the purposes of estimation, and refer to them as prices throughout for brevity.

The platform derives revenues by charging a percentage-based fee on successful transactions, which is taken out of the final price reached via the bidding process. Therefore, the prices should be understood as the seller's pricing offer plus a percentage fee for the platform itself. Since this fee applies uniformly across all sellers and is known to the sellers at the time of bidding, it does not impact the bidding process in a meaningful way and simply becomes another part of the costs of fulfilling the delivery for the sellers.

3.2.3 Data and descriptive statistics

The data used to estimate the model was collected from the AnyVan website using an automated web-scraping program.[5] The dataset used for the estimation includes all listings for vehicle deliveries listed on the AnyVan platform during the month of April 2015. These include listings for deliveries of cars or other motor vehicles as well as listings for deliveries of motorcycles. I concentrate on this delivery category in particular because the service offering in this category is relatively homogenous compared to other major delivery types such as furniture deliveries or removals, and as a result the observable delivery characteristics in the data capture the main factors

[5]See Chapter 2 for a description of the data gathering process.

affecting the underlying costs of delivery much better. At the same time, vehicle deliveries form a reasonably large proportion of the listings created on the platform, compared to other more specialised categories.

The listings in the dataset are grouped into individual markets, which will define the set of listings over which sellers make their entry decisions as well as the set of buyers and sellers that impact the matching process for each listing. I define markets based on a combination of delivery type, the date of the creation of the listing and the geographic region. The delivery type is based on the category assigned by the buyer during the listing process among the options offered by the platform, which includes cars or other motor vehicles and motorcycles as two distinct categories. The geographic region is based on the delivery origin location provided in the listing,[6] which was coded to UK NUTS 1 geographic regions.[7]

The final dataset consists of 6,790 domestic UK listings of motor vehicles ("listing-level dataset"), divided into 698 distinct markets based on a combination of delivery type (car or motorcycle delivery), the date of listing and the origin UK region ("market-level dataset"). Tables 3.2 and 3.3 present descriptive statistics on the two datasets.

[6]The data does also include the destination location for each delivery, which provides some alternative ways of defining the geographic region such as unique origin-destination pairs used in some previous studies of transport markets (Berry 1992). The baseline results presented in the main body of the paper are based on defining geographic regions in terms of the origin location only, since this results in relatively fewer markets with wider variation in sizes which are more useful for examining scale effects. Nevertheless, Appendix A presents a robustness check which uses unique origin-destination pairs to define geographic regions. The qualitative results are unaffected.

[7]The Nomenclature of Territorial Units for Statistics (NUTS) system is defined by the European Union and divides the UK into 12 geographical regions.

Sample / Variable	N	Mean	SD	Min	P25	Median	P75	Max
Prices (GBP)								
Whole sample	6,115	186.2	128.9	21	98	155	245	1,808
Cars	4,130	216.8	137.6	21	119	193	277	1,808
Motorcycles	1,985	122.6	76.0	25	76	112	147	1,580
Bidders								
Whole sample	6,790	3.2	2.2	0	2	3	5	14
Cars	4,721	3.0	2.2	0	1	3	4	12
Motorcycles	2,069	3.7	2.1	0	2	4	5	14
Conditional on Bidders>0	6,115	3.5	2.0	1	2	3	5	14
Cars	4,130	3.4	2.0	1	2	3	5	12
Motorcycles	1,985	3.8	2.0	1	2	4	5	14
% with no bidder								
Whole sample	6,790	0.10						
Cars	4,721	0.13						
Motorcycles	2,069	0.04						
Delivery distance (miles)								
Whole sample	6,790	137.5	100.0	0	57.8	122.9	189.9	522.7
Cars	4,721	131.3	99.2	0	51.3	113.5	184.2	496.1
Motorcycles	2,069	151.5	100.5	0	74.2	141.1	203.1	522.7

Table 3.2: Listing-level descriptive statistics

Sample / Variable	N	Mean	SD	Min	P25	Median	P75	Max
Listings								
Whole sample	698	10.0	7.0	1	5	8	14	39
Cars	356	13.6	13.6	1	8	13	18	39
Motorcycles	342	6.2	6.2	1	3	6	8	19
Sellers (observed)								
Whole sample	698	19.5	10.4	0	12	19	26	54
Cars	356	22.8	10.9	0	16	23	30	54
Motorcycles	342	16.2	8.5	0	9	16	22	43
Matches								
Whole sample	698	32.0	22.3	0	15	28	44	127
Cars	356	40.5	24.9	0	23	39	54	127
Motorcycles	342	23.1	14.7	0	11	21	31	87

Table 3.3: Market-level descriptive statistics

3.3 Model

3.3.1 Overall structure

The overall structure of the model of online marketplace platforms is as follows. The platform is composed of I distinct markets indexed by i, where each market can be understood to be delineated by variables such as product type, geography and time. Within each market there are B_i heterogeneous buyers or listings,[8] the number of which is determined exogenously. For each market there exists a large number of homogeneous potential entrant sellers, S_i of which enter the market and become active sellers. The entry decisions of the potential entrants are endogenous and depend on the ex-ante expected profits from participating in the online marketplace, taking into account some fixed costs of entry.

The platform generates linkages between buyers and sellers, which will be denoted as 'matches'. Each match links a single buyer to a single seller. The set of matches will typically be incomplete; that is, not all buyers will be matched to all sellers, and vice versa. Some buyers may not be matched to any sellers, and conversely some sellers may not be matched to any buyers. I model the generation of such matches as a stochastic process over the number of matches generated rather than directly modelling the searching processes of sellers on the platform.

Once the platform generates the matches between the buyers and sellers, the actual supplier for each buyer and the transaction price is decided in one of two ways. For the cases where the buyer is matched to more than one seller, a second-price procurement auction determines the final supplier and the transaction price. Buyers that are matched to only one seller face a take-or-leave offer from the sole seller. The demands from buyers that are not matched to any seller remain unfulfilled.

The entire model of the platform can therefore be broken down into three distinct

[8] On the AnyVan platform, the terms 'buyer' and 'listing' are not synonymous since a single buyer may create multiple listings. This does not make a difference to the analysis however, and therefore I will treat each listing as being representative of a unique buyer and will use the terms 'buyer' and 'listing' interchangeably.

processes: the seller entry process that determines the number of sellers in each market, the match-generating process that determines the number of bidders in the auction for each buyer, and the auction process that determines the actual supplier and the price for each buyer. The remainder of this section examines each of these processes, starting from the auction process and working backwards to the entry decision.

3.3.2 Buyer-level auction

The auction process for each buyer is modelled as a second price, independent private values (IPV) procurement auction.

Indexing individual bidders and buyers with j and k respectively, bidder j faces the following cost of fulfilling buyer k's requirements:

$$c_{jk} = \bar{c}_k - \varepsilon_{jk} = \gamma \mathbf{Y_k} + \xi_k - \varepsilon_{jk} \,. \tag{3.1}$$

The total cost c_{jk} is comprised of a common cost component $\bar{c}_k$ which is symmetrical for all bidders, and a bidder-specific cost residual ε_{jk} which is only realised at the start of the auction process. The common cost $\bar{c}_k$ can be further decomposed into a linear combination of buyer-specific covariates $\mathbf{Y_k}$ which are observed by the econometrician, and a buyer-specific cost term ξ_k which is not observed by the econometrician but is observed by all auction participants. I model the buyer-specific cost term ξ_k as being normal-distributed with zero mean and fixed variance σ_ξ^2. The common costs represent all drivers of costs that would be evident to all bidders based on the information available in the listing and affect all bidders symmetrically. The buyer-specific covariates $\mathbf{Y_k}$ capture the aspects of common costs that are observed by the econometrician, while the buyer-specific cost term ξ_k captures cost drivers that are either not observed by the econometrician or cannot be incorporated into the model in a simple linear way, such as the text descriptions accompanying the listing. The bidder-specific cost residuals ε_{jk} represent other cost factors that would affect bidders in an asymmetric way.

As in standard auction theory, the final transaction prices p_k are determined by the second lowest realisation of costs among the bidders. Denoting the number of bidders for buyer k as n_k and the second order statistic of ε_{jk} among all n_k bidders as $\varepsilon^{(2:n_k)}$:

$$p_k\left(n_k > 1\right) = \bar{c}_k - \varepsilon^{(2:n_k)} = \gamma \mathbf{Y_k} + \xi_k - \varepsilon^{(2:n_k)}\,. \tag{3.2}$$

Under the IPV framework, distribution of the bidder-specific cost term ε_{jk} is identified by the pricing outcomes non-parametrically without any distributional assumptions. This is because the observed prices identify the distribution of the second-order statistics $\varepsilon^{(2:n_k)}$, and this distribution uniquely identifies the underlying distribution of ε_{jk} (Athey and Haile 2002). However, imposing further distributional assumptions on this term can ensure a closed-form likelihood function for the auction model, which greatly speeds up the estimation. Consistent with previous empirical studies of second price IPV auctions with asymmetric valuations (Allen et al. 2019, Brannman and Froeb 2000) I model the bidder-specific cost residual ε_{jk} as following a Gumbel distribution with location parameter η_k and scale parameter μ_k, with density given by:

$$f_\varepsilon(\varepsilon;\mu,\eta) = \mu e^{-\mu(\varepsilon-\eta)} e^{-e^{-\mu(\varepsilon-\eta)}}\,. \tag{3.3}$$

As shown in Froeb et al. (1999), this functional form for the bidder-specific cost residuals ensures that the second order statistic $\varepsilon^{(2:n)}$ has the following closed-form distribution:

$$f_{\varepsilon(2:n)}(\varepsilon;\mu,\eta) = n\cdot f_\varepsilon(\varepsilon;\mu,\eta_{n-1}) - (n-1)\cdot f_\varepsilon(\varepsilon;\mu,\eta_n) \tag{3.4}$$

where $\eta_{n-1} = \eta + \frac{log(n-1)}{\mu}$ and $\eta_n = \eta + \frac{log(n)}{\mu}$.

I model the Gumbel scale parameter μ_k as a linear function of listing characteristics $\mathbf{Z_k}$:

$$\mu_k = \delta \mathbf{Z_k}\,. \tag{3.5}$$

I model the Gumbel location parameter η_k as a simple constant since this term affects the underlying costs of all bidders symmetrically, and any covariance between this term and observed listing characteristics is already captured by the parameterisation of the common cost term $\bar{c}_k$ as shown in Equation (3.1).[9]

The pricing equation (3.2) and the distributional assumptions for residual terms ξ_k and ε_{jk} fully specify the predicted distributions for the observed prices for listings with at least two bidders.

3.3.2.1 Monopoly listings

The platform may also generate matching outcomes where only a single seller is matched to a buyer. For these cases, the second-price auction framework set out above cannot generate a predicted price. I model the pricing outcomes in these cases as a take-or-leave offer from the seller, which is determined by the seller's realised costs plus a fixed markup.[10] As in the auction framework, the sole seller's total costs of fulfilling the buyer's requirements is comprised of common and seller-specific cost components, the distributional assumptions for which were discussed above. The final transaction prices equal the sum of this realised cost of the seller and a fixed mark-up ψ:

$$p_k\left(n_k = 1\right) = c_{jk} + \psi = \gamma \mathbf{Y_k} + \xi_k - \varepsilon_{jk} + \psi\,. \tag{3.6}$$

This pricing equation in combination with the distributional assumptions for residual terms ξ_k and ε_{jk} specify the predicted distributions for the observed prices for listings with only one bidder.

[9] η_k can therefore also be interpreted as the constant term in $\mathbf{Y_k}$. As such, $\mathbf{Y_k}$ will not include a constant term during estimation.

[10] During the estimation steps I will take prices in logarithms rather than in absolute values. Using a fixed markup term in this part of the model therefore implies a proportional markup in actual currency units.

3.3.3 Buyer-seller matching process

Within each market i, the platform generates matches between each of the B_i buyers and the S_i sellers which determines the number of bidders for each buyer. Denote the number of bidders matched to each buyer k as n_k.

I assume that the number of matches n_k follows a Poisson distribution with buyer-specific Poisson parameter λ_k:

$$Prob\left(n_k = n; \lambda_k\right) = f_{Poisson}\left(n; \lambda_k\right) = \frac{\lambda_k^n e^{-\lambda_k}}{n!}\,. \tag{3.7}$$

The Poisson distributional assumption provides a parsimonious way of modelling the distribution of the count data based on a single parameter, and has been used in previous studies to model the endogenous distribution of number of bidders in an auction entry context (Bajari and Hortaçsu 2003, Giray et al. 2009). Nekipelov (2007) also models entry of bidders into auctions as a Poisson process in a continuous time setting.

As noted in Bajari and Hortaçsu (2003), the Poisson assumption in this context is also consistent with game theoretical models of auction entry such as Levin and Smith (1994). The Poisson distribution for the number of matches can be interpreted as resulting from an entry game where a large number of identical sellers are making entry decisions over a set of auctions. A symmetric mixed strategy equilibrium in this game involves each potential entrant entering each auction with a fixed probability, which produces a binomial distribution for the total number of bidders (since each buyer-seller pairing has a fixed probability of producing a bidder). With large numbers of participants in the market, the binomial distribution approximates to a Poisson distribution.

I assume that λ_k is determined by the total number of buyers in the market, the number of sellers in the market and listing characteristics via a Cobb-Douglas production function. For listing k in market i:

$$\lambda_k = A_k B_i^{\beta_1} S_i^{\beta_2} = \alpha \mathbf{X_k} B_i^{\beta_1} S_i^{\beta_2} \,. \tag{3.8}$$

The Cobb-Douglas 'productivity' term A_k reflects any listing-specific factors that influence the number of bidders that the listing attracts. For simplicity I model this as a linear combination of observed listing characteristics $\mathbf{X_k}$. β_1 and β_2 reflect the marginal impact of additional buyers and sellers participating in the market on the expected number of bidders per listing, and would be expected to be negative and positive respectively. The use of a Cobb-Douglas function in this context is in line with previous work on matching in online platforms (Cullen and Farronato 2018).[11] It offers a simple functional form that satisfies several common sense modelling requirements, such as producing zero matches whenever the numbers of buyers or sellers is zero, while still allowing for estimation of the relative contributions and returns to scale from the two factors.

The Cobb-Douglas functional form and the Poisson distribution assumptions together give an intuitive way to examine the scale effects for the matching process for the market as a whole, in line with the usual interpretation of Cobb-Douglas production functions. To see this, note that the total number of matches generated in the market equals the sum of the number of matches generated for each of the buyers in the market, and is therefore itself Poisson-distributed. Denoting the set of buyers in market i as $\mathcal{B}_i$, the distribution of the total number of matches generated on the market is given by:

$$\sum_{k \in \mathcal{B}_i} n_k \sim Poisson\left(\bar{\lambda}_i\right) \tag{3.9}$$

where:

[11] The Cobb-Douglas function is also widely used in other labour matching contexts (Petrongolo and Pissarides 2001).

$$
\begin{aligned}
\bar{\lambda}_i &= \sum_{k \in \mathcal{B}_i} \lambda_k \\
&= \sum_{k \in \mathcal{B}_i} \alpha \mathbf{X_k} B_i^{\beta_1} S_i^{\beta_2} \\
&= \alpha \left(\sum_{k \in \mathcal{B}_i} \mathbf{X_k} \right) B_i^{\beta_1} S_i^{\beta_2} \\
&= \alpha \bar{\mathbf{X}}_\mathbf{i} B_i^{\beta_1 + 1} S_i^{\beta_2}
\end{aligned} \tag{3.10}
$$

and:

$$
\bar{\mathbf{X}}_\mathbf{i} = \frac{1}{B_i} \sum_{k \in \mathcal{B}_i} \mathbf{X_k} \,. \tag{3.11}
$$

$\bar{\mathbf{X}}_\mathbf{i}$ can be interpreted as the average or representative listing characteristics for the listings in market i. The Cobb-Douglas parameters in the market-level match distribution as given in (3.10) therefore provide an easy way to examine the scale effects of the matching process at the level of the market. If the two Cobb-Douglas parameters in (3.10) add to unity ($1 + \beta_1 + \beta_2 = 1$, i.e. $\beta_1 + \beta_2 = 0$), the matching process exhibits constant returns to scale and increasing the numbers of buyers and sellers by the same proportion results in an increase in the number of matches by the exact same proportion. If the two parameters add to greater than unity ($\beta_1 + \beta_2 > 0$), the matching process exhibits increasing returns to scale and an increase in the numbers of buyers and sellers results in a greater-than-proportional increase in the number of matches. If the two parameters add to less than unity ($\beta_1 + \beta_2 < 0$), the matching process exhibits decreasing returns to scale or crowding effects.

3.3.4 Seller entry decision

The buyer-level auction model and the match generating process set out above can be used to derive the ex-ante expected margin for the seller from entering the market as

a function of the number of buyers and sellers in the market.

First, define $\pi(n_k)$, the ex-ante expected margin for the actual supplier to an individual buyer k conditional on n_k, the number of sellers matched to buyer k. The actual supplier is either the winner of the auction (if two or more sellers were matched to the buyer) or the monopoly seller (in case of only one matched seller). Due to the additive functional form for the seller-specific cost residuals, this value is independent of the common cost component $\bar{c}_k$ and is only a function of the number of bidders n_k:

$$\pi(n_k) = \begin{cases} \varepsilon^{(1:n_k)} - \varepsilon^{(2:n_k)} & \text{if } n_k > 1 \\ \psi & \text{if } n_k = 1 \end{cases}. \tag{3.12}$$

The total ex-ante expected margin for all sellers in a market is given by summing the values of $\pi(n_k)$ over all probability-weighted values of n_k and across all buyers in the market. Since the sellers are ex-ante homogenous, this total margin can be divided by the number of sellers to derive the ex-ante expected margin for an individual seller. This yields an expression for V_i, the ex-ante expected margin for a potential seller from entering market i, as follows:[12]

$$V_i(S_i, B_i, \mathbf{I_{\mathcal{B}_i}}) = \frac{1}{S_i} \sum_{k \in \mathcal{B}_i} \sum_{n=1}^{\infty} \{Prob(n_k = n | \lambda_k, S_i, B_i)\ \pi(n)\}\ . \tag{3.13}$$

$\mathbf{I_{\mathcal{B}_i}}$ represents all buyer characteristics for the buyers in market i which affect sellers' expected margins, and includes $\mathbf{X_k}$ (which affects the expected margins via the expected number of bidders per listing) and $\mathbf{Z_k}$ (which affects the expected margins via the variance of bidder-specific costs across bidders).

I assume that the potential sellers face symmetric fixed costs T_i to enter market i. This gives the following ex-ante expected profit function for a seller participating in market

[12] Unmatched buyers ($n_k = 0$) do not generate any margins for sellers, and therefore can be excluded from the summation. The summation over the possible values of n_k is shown as an infinite summation here in line with the Poisson distribution assumption, although in reality the number of matches for a buyer is bounded by the number of sellers in the market. In practice, the Poisson distribution approaches zero for values much lower than the typical values for S_i in the data, and therefore this does not impact the results.

i, which I denote Π_i:

$$\Pi_i\left(S_i\right) = V_i(S_i, B_i, \mathbf{I_{\mathcal{B}_i}}) - T_i\,. \tag{3.14}$$

Similar to the static entry models set out in Bresnahan and Reiss (1990, 1991), I assume that there are a large number of homogeneous and risk-neutral potential entrants who enter the market as sellers as long as the expected profit from participating in the market is positive. The unique Nash equilbrium of this entry game is reached when all sellers in the market have positive ex-ante (i.e. before the matching process is realised) expected profits and no additional potential sellers can enter ex-ante profitably:

$$V_i(S_i, B_i, \mathbf{I_{\mathcal{B}_i}}) - T_i \geq 0 \tag{3.15}$$

and

$$V_i(S_i + 1, B_i, \mathbf{I_{\mathcal{B}_i}}) - T_i < 0\,. \tag{3.16}$$

The level of T_i therefore uniquely determines the number of sellers in the market in equilibrium S_i^* according to the following condition:

$$V_i(S_i^* + 1, B_i, \mathbf{I_{\mathcal{B}_i}}) < T_i \leq V_i(S_i^*, B_i, \mathbf{I_{\mathcal{B}_i}})\,. \tag{3.17}$$

Under simple functional form assumptions for the distribution of T_i, the observed number of entries S_i can identify the parameters of the distribution of T_i under the assumption that the observed outcome is consistent with this Nash equilibrium outcome. For example, Bresnahan and Reiss (1990, 1991) assume a normal distribution fo the distribution of entry costs, which turns the model into an ordered probit model. I follow a similar approach but assume a lognormal distribution for T_i conditional on market-level characteristics $\mathbf{W_i}$, which accounts for the fact that the entry costs must be strictly positive. In particular, I assume the following empirical specification for T_i:

$$T_i = exp\left\{\tau \mathbf{W_i} + e_i\right\} \tag{3.18}$$

where e_i is a market-level residual component of the entry costs which is normally distributed with zero mean and variance σ_τ^2. This turns the entry model into an ordered probit model and defines the probability of observing any non-negative number of sellers in the market S_i as follows:

$$\begin{aligned} &Prob\left(S_i | B_i, \mathbf{W_i}, \mathbf{I_{\mathcal{B}_i}}\right) \\ &\quad = \Phi\left[\frac{log\left(V_i(S_i, B_i, \mathbf{I_{\mathcal{B}_i}})\right) - \tau \mathbf{W_i}}{\sigma_\tau}\right] - \Phi\left[\frac{log\left(V_i(S_i+1, B_i, \mathbf{I_{\mathcal{B}_i}})\right) - \tau \mathbf{W_i}}{\sigma_\tau}\right]. \end{aligned} \tag{3.19}$$

This entry process assumes that the sellers know the number and the composition of buyers in each market before making their entry decisions, such that in equilibrium they are never ex-post regretful of their entry choices due to the number of buyers or the types of buyers present in the market (though they may be ex-post regretful of their entry decision due to the realisation of the number of matches via the Poisson matching process). This could be justified on the basis that the sellers are likely to have some experience with using the platform, are likely to be able to predict with reasonable accuracy the number and likely compositions of the buyers who would be present on the platform on a day-to-day basis, and are able to make entry and exit decisions quickly if their expectations for the numbers of buyers and sellers in the market turn out to be incorrect (e.g. on a 'slow day' they may choose to not spend any further time participating on the platform).

3.4 Estimation

This section sets out the process for estimating the parameters of the model. The observed outcomes in the data can be split into market-level outcomes and listing-

level outcomes. Market-level outcomes are the numbers of sellers in each market (S_i). The listing-level outcomes are the number of observed bidders (n_k) and prices (p_k). The observed data also includes market-level buyer numbers (B_i) and covariates for the entry costs ($\mathbf{W_i}$), and the listing-level covariates for the Cobb-Douglas matching process, the common costs and the cost scale parameter ($\mathbf{X_k}$, $\mathbf{Y_k}$, $\mathbf{Z_k}$). The parameters to be estimated are the seller entry covariate and variance parameters (τ, σ_τ), the matching process Cobb-Douglas parameters (α, β), the common cost covariate and scale parameters (γ, σ_ξ), the seller-specific cost residual parameters (η, δ) and the monopoly markup (ψ).

The three component parts of the model define the probability for each of the observed market outcomes, with the seller entry decision determining the number of sellers in each market, the buyer-seller matching process determining the number of bidders per listing and the buyer-level auction process determining the prices. This yields a joint probability for all observed market outcomes of the following form:

$$\prod_{i=1}^{I} Prob\left(S_i|B_i, \mathbf{W_i}, \mathbf{X_k}, \mathbf{Y_k}, \mathbf{Z_k}\right) \left\{ \prod_{k \in \mathcal{B}_i} Prob\left(n_k|S_i, B_i, \mathbf{X_k}\right) Prob\left(p_k|n_k, \mathbf{Y_k}, \mathbf{Z_k}\right) \right\}. \tag{3.20}$$

The model is estimated by maximum likelihood in a single step, using a joint likelihood function that reflects this joint probability for all observed outcomes. The following sections explain the likelihood functions for each of the three parts of the model in more detail.

3.4.1 Buyer-level auction

I include the following buyer-level variables in the common cost covariates $\mathbf{Y_k}$: motorcycle dummy variable, (logarithm of) delivery distance and its squared value, a dummy variable reflecting deliveries from Northern Ireland to Great Britain or vice versa, and the (logarithm of) region-average labour costs for the delivery origin and destination

regions.[13] The separate inclusion of the Gumbel location parameter η_k means that a constant term is not required in $\mathbf{Y_k}$. For cost scale covariates $\mathbf{Z_k}$ I use a contant term, a motorcycle dummy variable and (logarithm of) delivery distance.

The pricing equations 3.2 and 3.6 with the associated distributional assumptions give the following mixed likelihood function for prices conditional on whether the number of bidders is equal to or greater than 1:

$$\mathcal{L}_k^{auction}(\gamma, \sigma_\xi, \eta, \delta, \psi | \mathbf{Y_k}, \mathbf{Z_k}, p_k, n_k) = \begin{cases} \int f_{\varepsilon(2:n_k)}(\gamma \mathbf{Y_k} - p_k + \xi; \mu_k, \eta) \phi\left(\frac{\xi}{\sigma_\xi}\right) d\xi & \text{if } n_k > 1 \\ \int f_{\varepsilon}(\gamma \mathbf{Y_k} - p_k + \psi + \xi; \mu_k, \eta) \phi\left(\frac{\xi}{\sigma_\xi}\right) d\xi & \text{if } n_k = 1 \end{cases} \tag{3.21}$$

where ϕ is the density of the standard normal distribution. The combination of the Gumbel distributional assumption and the second price IPV auction framework ensures that this likelihood function has infinite support, therefore the concerns regarding the asymptotics of maximum likelihood estimators for auction models discussed in Donald and Paarsch (1993, 1996) do not arise.

3.4.2 Buyer-seller matching process

I include the following buyer-level variables in the Cobb-Douglas matching term covariates $\mathbf{X_k}$: a contant term, motorcycle dummy variable, (logarithm of) delivery distance and region population densities for the delivery origin and destination regions.[14]

Given the Cobb-Douglas function form for λ_k, the Poisson match-generating process set out in equation 3.7 gives the following likelihood function for the number of observed matches for each listing n_k in market i:

[13] Eurostat region average labour costs for NUTS1 regions
[14] Eurostat region average population densities for NUTS1 regions

$$\mathcal{L}_k^{matching}(\alpha, \beta | \mathbf{X_k}, B_i, S_i, n_k) = f_{Poisson}\left(n|\lambda_k\right) = \frac{\lambda_k^n e^{-\lambda_k}}{n!} \tag{3.22}$$

where $\lambda_k = \alpha \mathbf{X_k} B_i^{\beta_1} S_i^{\beta_2}$.

The use of the market-level seller numbers S_i in this likelihood function requires some further data assumptions. This is because while the sellers in the dataset are observed individually, they are only observed through their bidding behaviour. This means that any sellers who made the decision to enter the market but did not get matched to any listings through the matching process would not be observed in the data, and the observed number of sellers is a biased measure of the true number of sellers as a result.[15] I account for this bias by generating a measure of the number of such unobserved unmatched bidders from their activities elsewhere on the platform on the same day. More specifically, for each market, I calculate the number of sellers that satisfy two criteria: (i) they were observed bidding on a listing of the same delivery type and geographic area at any time during the sample period of April 2015, and (ii) they were not observed within the market concerned, but they were observed bidding on another listing on the platform on the same day. These sellers are those who have recently showed that they have the practical capacity to carry out a similar delivery and were also actively browsing the platform on the day in question, and may therefore be a suitable approximation to the set of unmatched sellers.

3.4.3 Seller entry decision

I include a constant term and a motorcycle dummy variable in the entry cost covariates $\mathbf{W_i}$.

The probability given in ordered probit equation (3.19) gives the following likelihood:

[15] This problem is not present on the buyer side, since all buyers/listings are observed including those that do not attract any bidders. The data on listings shows that around 10 percent of the listings in the sample attract zero bidders, which suggests that the bias from not counting unmatched buyers or sellers is potentially significant.

$$\mathcal{L}_k^{entry}(\theta|S_i, B_i, \mathbf{W_i}, \mathbf{I}_{\mathcal{B}_\mathbf{i}})$$
$$= \Phi\left[\frac{log\left(V_i(S_i, B_i, \mathbf{I}_{\mathcal{B}_\mathbf{i}})\right) - \tau\mathbf{W_i}}{\sigma_\tau}\right] - \Phi\left[\frac{log\left(V_i(S_i+1, B_i, \mathbf{I}_{\mathcal{B}_\mathbf{i}})\right) - \tau\mathbf{W_i}}{\sigma_\tau}\right]. \quad (3.23)$$

θ includes all parameters from all three parts of the model and reflects the fact that the ex-ante expected margin function V_i includes parameters from both the auction and matching parts of the model.

3.4.4 Joint estimation

The model yields three likelihood functions for the three distinct sets of observed outcomes: two likelihoods shown in (3.21) and (3.22) for the matching and auction outcomes of each observed listing, and one likelihood shown in (3.23) for the market entry outcomes of each observed market. The likelihoods are summed across all observations of listings and markets, consistently with the joint probability given in (3.20). This yields the following combined log-likelihood:

$$\ell(\theta) = \sum_{k=1}^{K} \ell_k^{auction}(\gamma, \sigma_\xi, \eta, \psi, \delta) + \sum_{k=1}^{K} \ell_k^{matching}(\alpha, \beta) + \sum_{i=1}^{I} \ell_i^{entry}(\theta) \quad (3.24)$$

where each ℓ represents the log-likelihood corresponding to the likelihoods set out above. The structure of the likelihood function reflects the nature of the sample, which comprises observations on I separate markets and K listings which fall within the I markets. Simply summing all three likelhoods across K observations of listings would therefore include many market-level observations multiple times in the likelihood function and arbitrarily give higher weightings to observations for markets with large numbers of listings. The appropriate combined log-likelihood is therefore a summation of the matching and auction likelihoods across all K observations of listings and the market entry likelihoods across all I observations of markets, as per above.

While the two likelihoods for the listing-level matching and auction outcomes ($\ell_k^{auction}$,

$\ell_k^{matching}$) do not share any common parameters, the market-level likelihood for the market entry outcomes (ℓ_i^{entry}) share parameters with both the matching and auction likelihoods (via the ex-ante expected margin V_i). Therefore each of the estimated parameters can influence the estimation of all other parameters. This combined likelihood function is therefore used to estimate all model parameters jointly by maximum likelihood.

3.4.5 Identification

Since all parameters are estimated jointly, the estimation of each of the parameters are potentially influenced by all aspects of the data. However, certain variations observed in the data drive the identification of each parameter more than others. The observed covariance between price levels and listing characteristics from the listing-level data identifies the common cost covariate parameters γ. For listings with more than one bidder, the second-price auction framework and the distributional assumption for the underlying costs specify the relationship between the number of bidders and the first two moments of prices, which identifies the Gumbel cost location parameter η and scale parameters δ. Any deviation in the second moment of prices from this specified relationship identifies the residual variance parameter σ_ξ. The first moment of observed prices for listings with only one bidder then identifies the monopoly seller markup parameter ψ. The observed numbers of buyers and sellers from the market-level data, along with the observed numbers of bidders from the listing-level data and their correlation with listing characteristics, identify the matching parameters α and β. Finally, the variation in observed numbers of sellers across different types of markets identifies the market entry parameters τ and σ_τ.

The theoretical literature on platforms and two-sided markets have previously emphasised the potential multiple-equilibria outcomes that can result when both sides of the market are making their participation decisions endogenously, which can pose problems for identification in empirical settings. This issue does not arise here due to the assumption that the participation decisions of buyers are exogenous. This is a reasonable

approach for the setting given that the model is estimated based on data collected over a period of a month, which is likely too short a time period for consumer habits over the choice of platforms to be influenced by the market outcomes on the platform itself. Any observed variation in the numbers of buyers there would likely reflect exogenous variation in demand across product categories, different geographies and on different days, rather than the performance of the platform itself. This is further supported by the fact that there are no easy ways for buyers to check the market outcomes being reached on the platform prior to creating their listing, short of spending large amounts of time browsing through the listings themselves. Given the low costs of creating a listing on the platform and the fact that creating a listing does not commit the buyer to a transaction, it would seem more likely that the variation in numbers of buyers in the data are determined by external factors such as exogenous fluctuations in demand for deliveries, rather than endogenously.

3.5 Results

Table 3.4 presents the maximum likelihood estimates of the model based on a joint estimation of the combined likelihood function (3.24).

The distributional parameters for the auction cost residuals appear to be well estimated with relatively small standard errors. The estimates for Gumbel scale covariates δ imply typical scale parameter μ values of 4.9 to 5.1, which translates into standard deviations for seller-specific cost residuals of 0.25 to 0.26.[16] This is significantly larger than the estimated standard deviation for the common cost residual of 0.085. This suggests that a significant proportion of the observed variation in prices can be explained by the observed variation in the number of bidders, and a relatively small proportion of the variation in prices is left to be explained by unobserved listing characteristics. This gives some confidence that the auction model is properly capturing the effect of the competition between bidders on prices.

[16]Scale parameter μ and standard deviation values for car and motorcycle deliveries calculated at the median delivery distance of 122.9 miles.

Parameter	ML estimate	S.E.
Gumbel cost location parameter (η)	-2.338	0.395
Common cost covariates (γ):		
Motorcycle dummy variable	-0.572	0.011
Distance (miles) (log)	-0.050	0.020
Distance (log)^2	0.061	0.003
NI-GB deliveries	0.335	0.031
Region avg labour costs, Origin (log)	0.111	0.030
Region avg labour costs, Destination (log)	0.062	0.056
Gumbel cost scale covariates (δ):		
Constant	3.164	0.098
Motorcycle dummy variable	0.205	0.163
Distance (miles) (log)	0.358	0.015
Common cost standard deviation (σ_ξ)	0.085	0.003
Monopoly seller markup (ψ)	0.276	0.014
Cobb-Douglas factor parameters:		
Buyers (β_1)	-0.060	0.017
Sellers (β_2)	0.343	0.030
Cobb-Douglas productivity parameters: (α)		
Constant	0.199	0.056
Motorcycle deliveries dummy	0.109	0.022
Region pop density, Origin (logs)	0.018	0.009
Region pop density, Destination (logs)	0.029	0.008
Distance (miles) (log)	0.047	0.005
Fixed cost (log) covariates (τ)		
Constant	-4.358	0.039
Motorcycle deliveries dummy	-1.031	0.051
Fixed cost (log) s.d. (σ_T)	0.544	0.022

Table 3.4: Maximum likelihood estimates

The estimate for the monopoly seller markup ψ is significantly positive as expected, with a small standard error. The ML estimates imply monopoly price-cost markup of 28 percent, which would seem to be plausible in magnitude.

The estimated parameters for the common cost covariates γ again show relatively small standard errors. Motorcycle deliveries are shown to have significantly lower underlying costs, with the ML estimate of -0.57 suggesting around 43 percent lower costs for motorcycles compared to all other types of larger motor vehicles. All other common cost covariates show expected signs: Delivery distance (in logarithms) is shown to have a positive and marginally increasing effect on costs (in logarithms). Deliveries from Great Britain to Northern Ireland or vice versa are shown to be significantly more costly compared to other deliveries. Finally, region-average labour costs for the delivery origin and destination regions are shown to have positive effects on costs.

As expected, the buyer-seller matching process parameters show that larger numbers of buyers have a negative impact on the number of matches for each buyer, while larger numbers of sellers have a larger positive impact. The estimates suggest that doubling the number of buyers in the market while holding all other factors constant would result in a 6 percent decrease in the number of matches per buyer, while doubling the number of sellers while holding all other factors constant would result in a 34 percent increase in the number of matches. Taken together, the match generating process exhibits increasing returns to scale; doubling both the number of buyers and sellers would result in a 28 percent increase in the number of matches per buyer, which equates to a larger-than-proportional increase in the total number of matches being generated in the market.[17]

The estimated parameters for the Cobb-Douglas productivity covariates α show that markets for motorcycle deliveries generate more matches on average than other vehicle deliveries, with a given motorcycle delivery market generating 11 percent more matches than an equivalent car delivery market with the same number of buyers and sellers.

[17] This is in line with the theoretical literature on two-sided markets, although at odds with Cullen and Farronato (2018) which found that buyer and seller numbers exhibited constant returns to scale in producing buyer-seller matches based on data from another online platform.

Regional population densities for both the origin and destination regions are also shown to have positive and significant effects on the number of matches generated. Finally, the delivery distance is also shown to have a significant and positive effect on the number of matches for each listing.

The maximum likelihood estimates for the seller entry process also appear to be closely estimated with low standard errors, and suggest that motorcycle delivery markets have lower entry costs compared to markets for larger vehicles. The estimated values of fixed costs covariates τ suggest an average entry cost of 2.76 GBP for car delivery markets and a smaller entry cost of 0.56 GBP for motorcycle delivery markets.[18]

Some goodness of fit tests show that the model parameters reasonably fit the observed distributions in the data in terms of both prices and matches. Figure 3.2 shows the medians and inter-quartile ranges of the prices in the data after controlling for the observed listing characteristics $\mathbf{Y_k}$ (i.e. the graphs reflect distributions of observed prices minus $\gamma\mathbf{Y_k}$) for different numbers of bidders, compared to the values predicted by the model as determined by the distributions of the cost residuals and the monopoly seller markup. While the model predicts a slightly tighter distribution of prices compared to the data, overall the model appears to capture the impact of the number of bidders on prices reasonably well, especially for lower bidder counts of around 5 or below which make up the majority of the sample (see Table 3.2).

[18]Note that these costs do not include the variable costs of fulfilling the deliveries, which are taken into account during the auction process, nor the fixed costs of the transport equipment which would be sunk costs at the time of the sellers' decisions to enter the platform. These costs include only the marginal cost of the decision by the seller to engage with the platform on a particular day, and could be understood as attention costs or the opportunity costs of searching for delivery jobs through other channels.

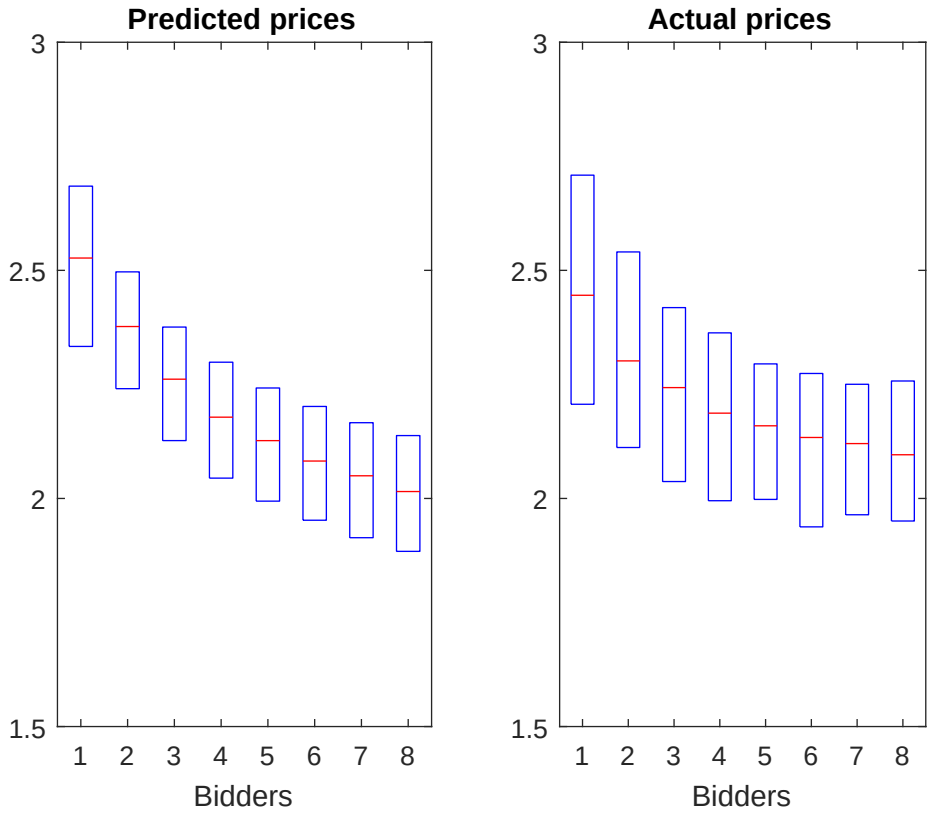


Figure 3.2: Predicted vs actual distributions of price residuals, Medians and interquartile ranges

Figure 3.3 shows the predicted distributions for the number of matches per buyer based on the Poisson distribution assumption, and the observed distributions from the data. The two distributions are shown in six different groupings for the value of λ_k so that markets that would be expected to generate low number of matches are shown separately from other markets, and so on. The frequency counts for the bars are included on the y-axes to indicate the relative sizes of the six groupings.[19] As shown in the first two figures, the model tends to over-predict the number of matches for listings with lower estimated λ_k, i.e. those with characteristics linked to lower number of matches. The model also has a tendency to under-predict the proportion of listings which receive zero matches. However, for listings with intermediate levels of estimated λ_k which make up the vast majority of the sample, the Poisson distribution

[19]The predicted distributions show the exact Poisson distributions for the indicated value of λ_k. The observed distributions have been divided between the six groupings by rounding the estimated values of λ_k to the nearest 0.5, i.e. the observed distribution for the group $\lambda_k = 2.5$ includes all observations with estimated values of λ_k between 2.25 and 2.75.

appears to do a reasonable job of matching the observed distributions for the number of matches. Moreover, the model appears to capture the shifts of the overall distribution across the six groupings reasonably well, which gives some confidence that the model is incorporating the listing characteristics which are the main drivers of the matching process.

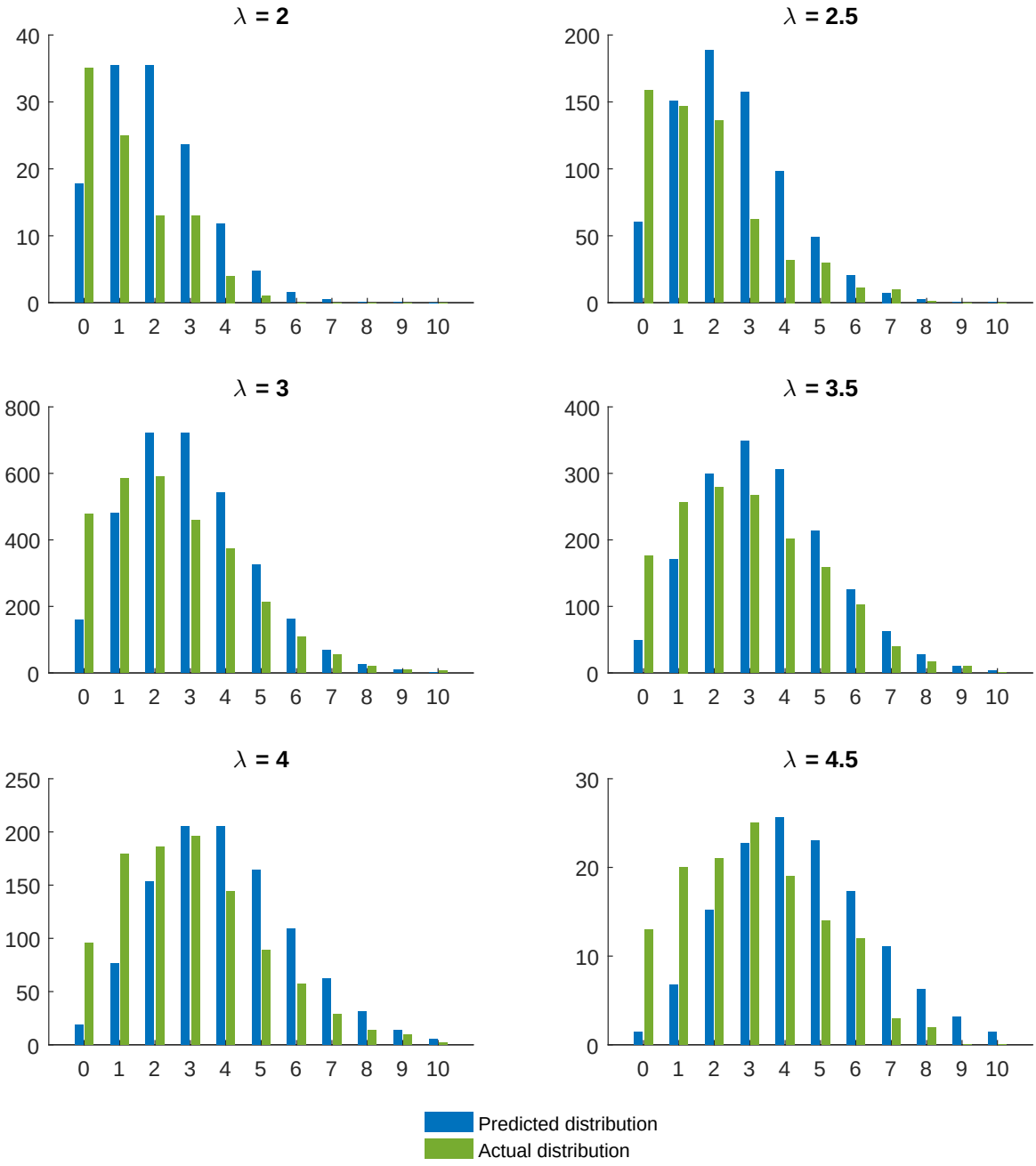


Figure 3.3: Predicted vs actual distribution of bidder numbers

3.6 Counterfactual experiments and implications for the platform

3.6.1 Impact of changes in number of sellers on prices

The estimates of the model parameters from the previous section allow for some conclusions on the functioning of the platform.

Firstly, the parameter estimates from the auction model allow me to revisit the main question from Chapter 2: What is the impact of the number of sellers bidding on each listing on the final price? The analysis in Chapter 2 examined this question using reduced-form models without making any structural assumptions. Within the second-price auction model presented above, the impact of bidder numbers on prices is determined by the combination of the monopoly seller markup and the competition between multiple bidders in the second-price auction.

Figure 3.4 presents the predicted relationship between the number of bidders and the expected price, with the price level for a listing with a monopoly seller normalised to 100. The same predicted relationships from the 2SLS estimates for vehicle listings in Chapter 2 are also presented for comparison.[20]

[20]These are based on the 2SLS results for the vehicle deliveries data only as presented in Chapter 2.

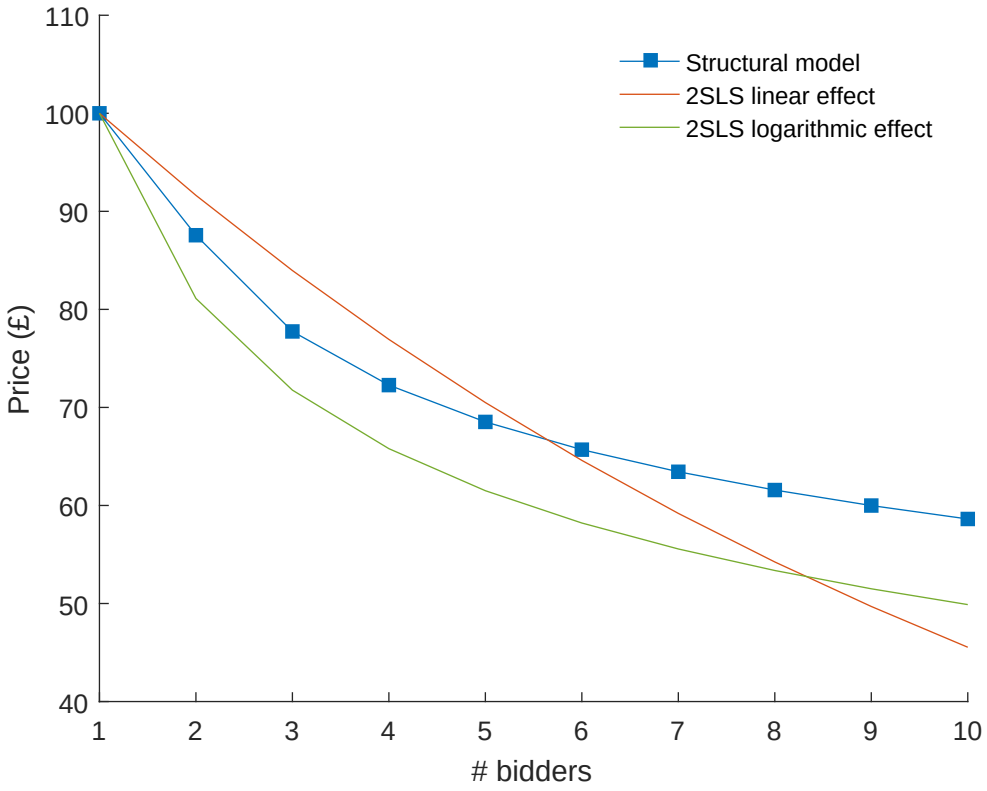


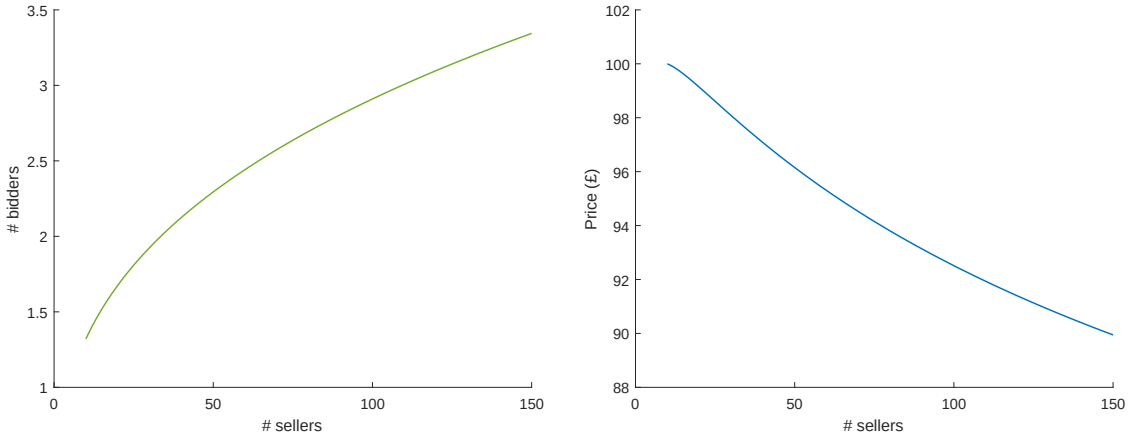
Figure 3.4: Predicted relationship between number of bidders and listing prices

As expected, the predicted relationship between the number of bidders and prices is sizeable and negative. The structural model predicts that relative to a monopoly seller, the addition of one additional bidder to a listing has the effect of lowering prices by around 10 percent. Increasing the number of bidders from one to five lowers prices by around 30 percent. The predicted relationship is roughly in line with the relationship predicted by the 2SLS estimates in Chapter 2, especially for listings with lower bidder numbers of five or fewer which make up the majority of the sample.

In addition to examining the relationship between number of sellers and prices at the narrower level of individual listings, the structural model presented in this chapter has the benefit of being able to address a broader question: What is the impact of changes in the number of sellers entering the platform at the level of the market on prices? This differs from the narrower question addressed in Chapter 2 since it is determined by a combination of the buyer-seller matching process occurring at the level of the platform as well as the competition between bidders at the level of individual listings. As the

number of sellers on the platform changes, it raises the average number of bidders that are matched to each buyer through the matching process. This change in the number of bidders in turn reduces the average prices reached for each listing as shown above.

Figure 3.5 presents a simulation which illustrates this two-step process. The simulation is based on an example market for car deliveries with 10 buyers, in a market with an average level of population density for the UK. Figure 3.5(a) shows the impact of changes in the number of sellers participating in the market on the average number of bidders per listing, varying the number of sellers between 10 and 150. As expected, increasing the number of sellers participating in the market increases the average number of bidders per listing. However, the figure shows that the magnitude of the effect is relatively small, with the average bidder numbers increasing from 1.3 to 3.3 as the total number of sellers in the market increases from 10 to 150. Correspondingly, as shown in Figure 3.5(b), the downward impact on average prices in the market is also minor; the same increase in the number of sellers from 10 to 150 results in a price decrease of around 10 percent.



(a) Average bidders per listing

(b) Expected price

Figure 3.5: Impact of increase in number of sellers in market

This result is driven by the parameters of the buyer-seller matching process. As presented in Table 3.4, while the number of buyer-seller matches generated by the platform is positively impacted by increases in the number of sellers, the magnitude of this impact is much smaller than unity, at around 0.34. This implies that when the number

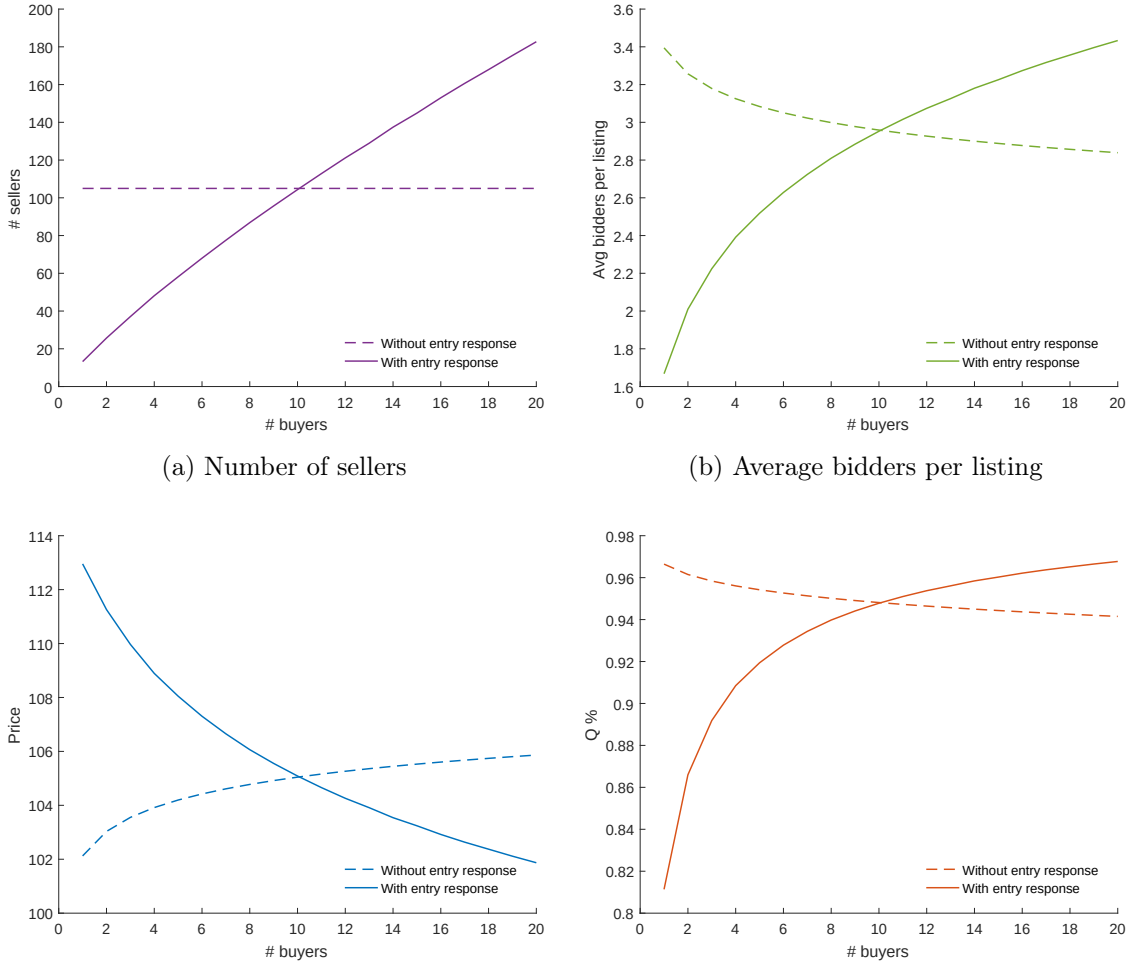
of sellers increases without any corresponding changes in the number of buyers or the efficiency of the matching process, the additional sellers run into congestion problems and do not readily translate into active bidders. Accordingly, even with significant increases in the number of sellers, the final impact on prices is relatively modest. This example highlights the importance of the matching process when examining the scale effects of online platforms. From the point of view of a consumer, an efficient process for matching buyers and sellers is as important as the number of sellers that are signed up to a platform when it comes to fully benefitting from the competitive impacts of online platforms.

3.6.2 Impact of changes in the numbers of buyers with seller entry response

The previous section examined a counterfactual change in the number of sellers with the number of buyers held constant, based on the parameter estimates of two of the three main parts of the model: listing-level auction process and the buyer-seller matching process. The final part of the structural model, the seller entry process, endogenously determines the number of sellers that would enter the market in response to the observed number of buyers and other market characteristics. This section examines the impact of this endogenous seller entry on market outcomes, by looking at impacts of demand fluctuations on market outcomes under a free-entry equilibrium for sellers.

Figure 3.6 presents a simulation of the model which illustrates the impact of varying the level of demand in the market. The number of buyers in the market are varied between 1 and 20, and the remaining market outcomes are calculated based on an example market for car deliveries with average listing and market characteristics. The market outcomes examined are the number of sellers entering the market (a), the average number of bidders per listing in the market (b), the average price (c) and the expected proportion of buyers that are matched to at least one seller ('fulfilled listings') (d). Since it is useful to distinguish between the direct impact of the change in the number

of buyers and the indirect impact via the seller entry response, all panels show the 'direct' effect of the change in the number of buyers with the number of sellers held constant in dotted lines, as well as the 'full' effect including the seller entry response in solid lines.



(a) Number of sellers

(b) Average bidders per listing

(c) Average price

(d) Proportion of fulfilled listings

Figure 3.6: Predicted relationship between number of bidders and prices

As shown in Figure 3.6(a), under a free entry equilibrium a change in the number of buyers has a roughly proportional effect on the equilibrium number of sellers. The direct effect on seller entry is zero by definition. Figure 3.6(b) shows that in the absence of seller entry response, an increase in the number of buyers results in a decrease in the average number of bidders per listing. This is to be expected since, as shown in the parameter estimates for the matching process, the total number of buyers has a negative impact on the number of matches per buyer. However, Figure 3.6(b) also shows that

once seller entry response is taken into account, the direction of this effect reverses; an increase in the number of buyers results in an increase in the average number of bidders per listing. The intuition for this is again provided in the parameter estimates for the matching process, which shows increasing returns to scale for the numbers of buyers and sellers in the generation of matches. Therefore, while an exogenous increase in demand has a proportional impact on the equilibrium number of sellers in the market, the resulting increase in the numbers of both buyers and sellers means that the platform becomes more efficient in matching up the buyers and sellers in the market, increasing the number of bidders matched to each of the listings.

Figures 3.6(c) and 3.6(d) show the impacts on the average prices and the expected proportions of fulfilled listings. In the absence of seller entry response, an increase in the number of buyers without a corresponding increase in the number of sellers means that the average prices in the market increase and the proportion of fulfilled listings decreases. However, when seller entry response is taken into account, the signs of the effects reverse, such that an increase in the number of buyers lowers prices and increases the chances for each buyer of finding at least one seller.

The model therefore suggests that any exogenous change in the number of buyers has two types of effects on the market outcomes for buyers in terms of expected prices and the likelihood of finding a seller. An increase in the number of buyers, for example, has a direct effect of increasing expected prices and decreasing the likelihood for each buyer of finding a seller, as the number of buyers relative to the number of sellers increases. However, the increase in the number of buyers also has an indirect effect of increasing the number of sellers entering the market in equilibrium, which has the opposite effect of decreasing expected prices and increasing the likelihood for each buyer of finding a seller. The above results showed that based on the parameter estimates for the model, the indirect effect dominates; on balance, an increase in the number of buyers improves the market outcomes for buyers.

This finding that the entry response of sellers on the platform can be large enough to reverse the direct effects of demand fluctuations may be counterintuitive. However,

the significant and timely entry response by small scale sellers which drives this effect has been empirically confirmed in several other studies of online platforms. Hall et al. (2016) note that the sellers on Uber are extremely responsive to the 'surge pricing' system used on the platform for balancing supply and demand, and that this mostly cancels out the negative impact on the matching rate which would otherwise be observed during peak demand periods. Similarly, Cullen and Farronato (2018) find that demand fluctuations on the TaskRabbit platform are mostly absorbed by responses in seller supply, with minimal impacts on prices and matching probabilities.

The finding that consumers benefit through lower prices and higher matching probabilities from entry of sellers onto the platform also gives empirical support to the cross-network effects described in the theoretical literature on two-sided platforms (Weyl 2010, Rochet and Tirole 2006, Armstrong 2006, among others). The theoretical models described in the literature typically describe the utilities of one group of users served by the platform as an increasing function of the number of users on the other side of the platform. The results of this section show how these types of utilities could arise naturally from a matching process between buyers and sellers.

3.6.3 Implications for long term growth of online platforms

The model presented in this study treated the entry decisions of buyers as exogenous, due to the fact that the time period covered by the data was too short for consumer habits to be influenced to a significant degree by the market outcomes being reached on the platform itself. In the long term, however, it is plausible that consumers' decisions to use the platform is influenced by the prices and matching probabilities that they would receive on the platform. This, in combination with the scale effects discussed above, holds some key implications for the long term growth of the platform. For instance, the scale effects suggest a "chicken and egg" problem for achieving growth in online platforms (Caillaud and Jullien 2003). Buyers may only be attracted to the platform if they would receive lower prices and higher match probability, but as shown above this would only occur if there is already a large user base of sellers. The sellers

would in turn only be attracted to the platform if there is a large user base of buyers using the platform already. The scale effects also suggest that the long term success of the platform requires achieving a critical mass user base (Evans and Schmalensee 2010). Below this critical mass, the quality of service offered to the buyers in the form of prices and match probabilities may be too low for the platform to retain its user base, and the shrinking user base would further lower the quality of service, leading to a downward spiral in platform size. Above this critical mass, however, the platform may see explosive growth as the larger user base enables a higher quality of service, which in turn attracts more users to the platform.

To examine this effect further, I used the same data collection methodology described in earlier sections to gather data on listings created on the AnyVan platform during the period October 2009 to March 2010, the first six full months of the platform's operation, when the scale of the platform was much smaller than the period reflected in the main sample of April 2015. Table 3.5 compares the market outcomes from the earlier period to the market outcomes in April 2015.

	Oct 2009-Mar 2010	April 2015
Number of listings per market:		
Furniture delivery	1.6	117.5
Home move	1.1	22.9
Vehicle delivery	1.1	9.9
Others	1.1	3.0
All categories	1.4	20.2
Mean bidders per listing:		
Furniture delivery	3.7	6.1
Home move	3.7	13.9
Vehicle delivery	3.4	4.8
Others	3.1	3.2
All categories	3.6	6.7
Proportion of listings fulfilled:		
Furniture delivery	0.947	0.961
Home move	0.956	0.988
Vehicle delivery	0.901	0.914
Others	0.939	0.940
All categories	0.942	0.956
Average price (GBP):		
Furniture delivery	67.0	91.1
Home move	132.8	215.0
Vehicle delivery	126.2	184.7
Others	102.1	119.9
All categories	83.6	118.6

Table 3.5: Comparison of buyer-seller matching outcomes, October 2009-March 2010 and April 2015

As expected, the average number of listings in each market (defined in terms of delivery category, geography and date) is very low in the earlier sample compared to the sample from April 2015. The number of listings per market as defined by the combination of date, geography and product category averages between 1 and 2 in the earlier period, orders of magnitude lower than the number of listings observed during April 2015. The model would predict that this would result in a lower number of bidders per listing and a higher proportion of unfulfilled listings, as the smaller scale means that the matching process works less efficiently. The data shows that this is indeed the case. Across all

major product categories, the buyers in the earlier period attracted fewer sellers per listing on average, and higher proportions of listings failed to attract any sellers.[21]

Taken together, the evidence on the historical growth of the AnyVan platform would seem to be in line with the implications of the model. During the earlier period initially following the launch of the AnyVan platform, the number of buyers using the platform was much smaller, and this appears to have resulted in worse matching outcomes for buyers. As with many other platform businesses, once the platform achieved a critical mass of users the market outcomes were much more advantageous for the buyers, and this may have led to the rapid growth of the platform.

3.7 Conclusion

This study set out a structural model of an online platform and used it to examine the scale effects of the platform on market outcomes such as prices and probability of successful match. The parameters of the model were estimated using data from AnyVan, an online peer-to-peer platform for delivery and removals services. The estimated parameters of the model imply that changes in the numbers of buyers or sellers using the platform in the absence of entry responses by the users on the other side of the platform have the expected directional impacts on market outcomes; an increase in the number of sellers lowers prices and increases the probability of match, while an increase in the number of buyers raises prices and decreases the probability of match. However, the parameters also imply that under a free entry equilibrium for sellers, this effect reverses direction; an increase in the number of buyers reduces prices and increases the probability of match. This is driven by the platform's process for matching buyers and sellers, which exhibits increasing returns to scale. An increase in the number of buyers causes an entry response by sellers which is roughly in proportion with the increased

[21]The data also shows that despite the lower number of sellers bidding on each listing, the prices in the earlier period were significantly lower on average compared to the sample from April 2015, which is at odds with the empirical findings from this chapter and Chapter 2. The magnitude of the change is too large to be explained by inflation. This finding may instead be due to significant changes in the typical characteristics of the delivery jobs listed on the platform as the platform has matured.

demand, and this in turn causes a greater-than-proportional increase in the number of buyer-seller matches on the platform. The end result is that each buyer receives bids from more sellers on average, which allows them to benefit from lower prices as well as lower chances of failing to find any suitable sellers. The implication that scale is crucial for a well-functioning platform is in line with the theoretical literature on two-sided markets (Evans and Schmalensee 2010, Caillaud and Jullien 2003).

The findings imply that the sellers on the platform are responsive to daily demand fluctuations in making their entry decisions, which is in line with the findings of other recent empirical studies which examined small-scale sellers on online platforms for services (Hall et al. 2016, Cullen and Farronato 2018). Limitations of the data mean that it is not practical in this case to examine the entry responsiveness of buyers, which seem likely to be less responsive to daily fluctuations in market conditions (due to slower changes in consumer habits) as well as being more influenced by exogenous factors (seasonal fluctuations in demand, marketing activities of the platform). Incorporating buyer entry as an endogenous factor would capture the scale effects of the platform more completely, and would be a valuable area for future work.

Appendix A: Robustness checks

Alternative definition of the geographic market

The model estimates presented in the main body of the paper were based on dividing listings into separate markets based on, among other dimensions, the delivery origin location defined by one of 12 UK NUTS1 regions. Some previous studies of entry into transport markets have used unique origin-destination pairs as the basis for defining geographic markets, for example Berry (1992) on air travel.

Since the destination information is also available from the AnyVan data, a similar geographic market definition based on unique origin-destination pairs can also be constructed for the model. This means that the 12 UK NUTS1 regions now produce 78 unique geographic markets based on unique origin-destination pairs. The total number of unique markets in the data defined by product, geography and time increases from 698 to 2,880.

Table 3.6 shows the maximum likelihood estimates for the model based on this alternative definition of the geographic market. Many of the parameter estimates are not significantly changed and the main qualitative results with regards to the matching process continue to hold; the number of buyers in the market has a negative effect on the number of matches for each buyer, the number of sellers in the market has a larger positive impact, and the two factors taken together show increasing returns to scale. The estimated values of the Cobb-Douglas matching parameters β, however, are significantly larger in magnitude compared to the base results.[22] Nevertheless, the results appear to confirm that the results from the main body of the paper are not sensitive to the exact definition of the geographic market delineations.

[22]This may be to be expected, given that the Cobb-Douglas matching parameters are now identified by variation in the numbers of buyers and sellers who are much more similar in terms of delivery characteristics compared to the base results. For instance, the base results imply that an addition of a single buyer to the platform reduces the number of sellers bidding on other buyers who share the same origin location by around 6 percent. The new results imply that an addition of a single buyer reduces the number of sellers bidding on other buyers who share the same origin *and* destination locations by around 33 percent.

Parameter	Base results		Robustness check	
	ML estimate	S.E.	ML estimate	S.E.
Gumbel cost location parameter (η)	-2.338	0.395	-2.405	0.536
Common cost covariates (γ):				
Motorcycle dummy variable	-0.572	0.011	-0.557	0.010
Distance (miles) (log)	-0.050	0.020	-0.033	0.016
Distance (log)^2	0.061	0.003	0.059	0.002
NI-GB deliveries	0.335	0.031	0.321	0.034
Region avg labour costs, Origin (log)	0.111	0.030	0.080	0.038
Region avg labour costs, Destination (log)	0.062	0.056	0.079	0.036
Gumbel cost scale covariates (δ):				
Constant	3.164	0.098	3.845	0.118
Motorcycle dummy variable	0.205	0.163	0.307	0.159
Distance (miles) (log)	0.358	0.015	0.396	0.016
Common cost standard deviation (σ_ξ)	0.085	0.003	0.094	0.002
Monopoly seller markup (ψ)	0.276	0.014	0.422	0.012
Cobb-Douglas factor parameters:				
Buyers (β_1)	-0.060	0.017	-0.329	0.013
Sellers (β_2)	0.343	0.030	1.077	0.019
Cobb-Douglas productivity parameters: (α)				
Constant	0.199	0.056	0.021	0.003
Motorcycle deliveries dummy	0.109	0.022	-0.005	0.001
Region pop density, Origin (logs)	0.018	0.009	-0.001	0.000
Region pop density, Destination (logs)	0.029	0.008	0.002	0.000
Distance (miles) (log)	0.047	0.005	0.003	0.000
Fixed cost (log) covariates (τ)				
Constant	-4.358	0.039	-6.045	0.029
Motorcycle deliveries dummy	-1.031	0.051	-0.691	0.036
Fixed cost (log) s.d. (σ_T)	0.544	0.022	0.757	0.010

Table 3.6: Model ML estimates, geographic markets based on origin-destination pairs

Bibliography

Agnew, H. (2016). Audit merry-go-round increases competition but limits choice. *Financial Times*, 17 January 2016.

Allen, J., Clark, R., and Houde, J.-F. (2019). Search frictions and market power in negotiated-price markets. *Journal of Political Economy*, Forthcoming.

Armstrong, M. (2006). Competition in two-sided markets. *RAND Journal of Economics*, 37(3):668–691.

Asker, J., Fershtman, C., Jeon, J., and Pakes, A. (2018). A computational framework for analyzing dynamic procurement auctions: The market impact of information sharing. *Working Paper*.

Athey, S. and Haile, P. (2002). Identification of standard auction models. *Econometrica*, 70(6):2107–2140.

Athey, S., Levin, J., and Seira, E. (2011). Comparing open and sealed bid auctions: Evidence from timber auctions. *Quarterly Journal of Economics*, 126:207–257.

Backus, M. and Lewis, G. (2019). Dynamic demand estimation in auction markets. *Working Paper*.

Bajari, P. and Hortaçsu, A. (2003). The winner's curse, reserve prices, and endogenous entry: Empirical insights from ebay auctions. *RAND Journal of Economics*, 34(2):329–355.

Bajari, P. and Hortaçsu, A. (2004). Economic insights from internet auctions. *Journal of Economic Literature*, 42:457–486.

Berry, S. (1992). Estimation of a model of entry in the airline industry. *Econometrica*, 60(4):889–917.

Brannman, L. and Froeb, L. (2000). Mergers, cartels, set-asides, and bidding preferences in asymmetric oral auctions. *Review of Economics and Statistics*, 82(2):283–290.

Bresnahan, T. and Reiss, P. (1990). Entry in monopoly markets. *Review of Economic Studies*, 57(4):531–553.

Bresnahan, T. and Reiss, P. (1991). Entry and competition in concentrated markets. *Journal of Political Economy*, 99(5):977–1009.

Bronnenberg, B. and Dubé, J.-P. (2017). The formation of consumer brand preferences. *Annual Review of Economics*, 9:353–382.

Caillaud, B. and Jullien, B. (2003). Chicken & egg: competition among intermediation service providers. *RAND Journal of Economics*, 34(2):309–328.

Cameran, M., Francis, J. R., Marra, A., and Pettinicchio, A. (2015). Are there adverse consequences of mandatory auditor rotation? evidence from the italian experience. *Auditing: A Journal of Practice & Theory*, 34(1):1–24.

Campo, S., Guerre, E., Perrigne, I., and Vuong, Q. (2011). First-price auctions with risk-averse bidders. *Review of Economic Studies*, 78:112–147.

Chen, M. K. and Sheldon, M. (2015). Dynamic pricing in a labor market: Surge pricing and flexible work on the uber platform. *Working Paper*.

Competition Commission (2013). Statutory audit services market investigation. Report and supporting documents available from https://www.gov.uk/cma-cases/statutory-audit-services-market-investigation.

Cullen, Z. and Farronato, C. (2018). Outsourcing tasks online: Matching supply and demand on peer-to-peer internet platforms. *Working Paper*.

Dekeyser, S., Gaeremynck, A., and Willekens, M. (2019). Evidence of industry scale effects on audit hours, billing rates, and pricing. *Contemporary Accounting Research*, 36(2):666–693.

Dinerstein, M., Einav, L., Levin, J., and Sundaresan, N. (2018). Consumer price search and platform design in internet commerce. *American Economic Review*, 108(7):1820–1859.

Donald, S. and Paarsch, H. (1993). Piecewise pseudo-maximum likelihood estimation in empirical models of auctions. *International Economic Review*, 34:121–148.

Donald, S. and Paarsch, H. (1996). Identification, estimation, and testing in parametric empirical models of auctions with the independent private values paradigm. *Econometric Theory*, 12(3):517–567.

Doogar, R., Sivadasan, P., and Solomon, I. (2015). Audit fee residuals: Costs or rents? *Review of Accounting Studies*, 20(4):1247–1286.

Einav, L., Farronato, C., and Levin, J. (2016a). Peer-to-peer markets. *Annual Review of Economics*, 8:615–635.

Einav, L., Farronato, C., Levin, J., and Sundaresan, N. (2016b). Auctions versus posted prices in online markets. *Journal of Political Economy*, 126(1):178–215.

Einav, L., Kuchler, T., Levin, J., and Sundaresan, N. (2015). Assessing sales strategies in online markets using matched listings. *American Economic Journal: Microeconomics*, 7(2):215–247.

Ellison, G. and Fudenberg, D. (2003). Knife-edge or plateau: When do market models tip? *Quarterly Journal of Economics*, 118(4):1249–1278.

Elyakime, B., Laffont, J. J., Loisel, P., and Vuong, Q. (1994). First-price sealed bid auctions with secret reservation prices. *Annales d'Economie et de Statistique*, 34:115–141.

European Commission (2014). Regulation (eu) no 537/2014 of the european parliament and of the council of 16 april 2014 on specific requirements regarding statutory audit of public-interest entities and repealing commission decision 2005/909/ec. *Official Journal of the European Union*, L158/77.

Evans, D. and Schmalensee, R. (2010). Failure to launch: Critical mass in platform businesses. *Review of Network Economics*, 9(4).

Farajallah, M., Hammond, R., and Penard, T. (2016). What drives pricing behaviour in peer-to-peer markets? evidence from the carsharing platform blablacar. *Working Paper*.

Farronato, C. and Fradkin, A. (2018). The welfare effects of peer entry in the accommodation market: The case of airbnb. *Working Paper*.

Fradkin, A. (2017). Digital marketplaces. In *New Palgrave Dictionary of Economics*. Palgrave Macmillan.

Fradkin, A. (2019). Search, matching and the role of digital marketplace design in enabling trade: Evidence from airbnb. *Working Paper*.

Froeb, L., Shor, M., and Tschantz, S. (2009). Mergers in auctions with an incumbent advantage. *Working Paper*.

Froeb, L., Tschantz, S., and Crooke, P. (1999). Mergers among asymmetric bidders: A logit second-price auction model. *Working Paper*.

GAO (2003). Public accounting firms: Required study on the potential effects of mandatory audit firm rotation. *United States General Accounting Office*, November 2003.

Giray, T., Hasker, K., Jiang, B., and Sickles, R. (2009). Estimating consumer surplus in ebay computer monitor auctions. *Working Paper*.

Goldman, M. and Rao, J. (2014). Experiments as instruments: Heterogeneous position effects in sponsored search auctions. *Working Paper*.

Groeger, J. (2014). A study of participation in dynamic auctions. *International Economic Review*, 55(4):1129–1154.

Guerre, E., Perrigne, I., and Vuong, Q. (2000). Optimal nonparametric estimation of first-price auctions. *Econometrica*, 68(3):525–574.

Hackenbrack, K., Jenkins, N. T., and Pevzner, M. (2014). Relevant but delayed information in negotiated audit fees. *Auditing: A Journal of Practice & Theory*, 33(4):95–117.

Hall, J., Kendrick, C., and Nosko, C. (2016). The effects of uber's surge pricing: A case study. *Working Paper*.

Hasker, K. and Sickles, R. (2010). ebay in the economic literature: Analysis of an auction marketplace. *Review of Industrial Organization*, 37:3–42.

Jeziorski, P. and Krasnokutskaya, E. (2016). Dynamic auction environment with subcontracting. *RAND Journal of Economics*, 47(4):751–791.

Jofre-Bonet, M. and Pesendorfer, M. (2000). Bidding behavior in a repeated procurement auction: A summary. *European Economic Review*, 44:1006–1020.

Jofre-Bonet, M. and Pesendorfer, M. (2003). Estimation of a dynamic auction game. *Econometrica*, 71(5):1443–1489.

Jolivet, G., Jullien, B., and Postel-Vinay, F. (2016). Reputation and prices on the e-market: Evidence from a major french platform. *International Journal of Industrial Organization*, 45:59–75.

Klemperer, P. (2005). Switching costs. In Durlauf, S. N. and Blume, L. E., editors, *The New Palgrave Dictionary of Economics Online*. Palgrave-Macmillan.

Levin, D. and Smith, J. L. (1994). Equilibrium in auctions with entry. *American Economic Review*, 84(3):585–599.

Li, T. (2005). Econometrics of first-price auctions with entry and binding reservation prices. *Journal of Econometrics*, 126:173–200.

Li, T. and Zheng, X. (2009). Entry and competition effects in first-price auctions: Theory and evidence from procurement auctions. *Review of Economic Studies*, 76:1397–1429.

Lucking-Reiley, D., Bryan, D., Prasad, N., and Reeves, D. (2007). Pennies from ebay: The determinants of price in online auctions. *Journal of Industrial Economics*, 40:223–233.

Moon, J., Shipman, J., Swanquist, Q., and Whited, R. (2019). Do clients get what they pay for? evidence from auditor and engagement fee premiums. *Contemporary Accounting Research*, 36(2):629–665.

Nekipelov, D. (2007). Entry deterrence and learning prevention on ebay. *Dissertation, Duke University*.

Newberry, P. (2015). The effect of competition on ebay. *International Journal of Industrial Organization*, 40:107–118.

Onur, I. and Velamuri, M. (2014). Competition, endogeneity and the winning bid: An empirical analysis of ebay auctions. *Information Economics and Policy*, 26:68–74.

Petrongolo, B. and Pissarides, C. (2001). Looking into the black box: A survey of the matching function. *Journal of Economic Literature*, 39:390–431.

Reiley, D. (2005). Experimental evidence on the endogenous entry of bidders in internet auctions. In Rapoport, A. and Zwick, R., editors, *Experimental Business Research, Economic and Managerial Perspectives, Volume II*. Springer.

Rochet, J.-C. and Tirole, J. (2006). Two-sided markets: A progress report. *RAND Journal of Economics*, 37(3):645–667.

Rust, J. (1994). Structural estimation of markov decision processes. In Engle, R. F. and McFadden, D. L., editors, *Handbook of Econometrics, Vol 4*, pages 2111–3155. North Holland.

Rysman, M. (2009). The economics of two-sided markets. *Journal of Economic Perspectives*, 23(3):125–143.

Ursu, R. (2018). The power of rankings: Quantifying the effect of rankings on online consumer search and purchase decisions. *Marketing Science*, 37(4):507–684.

Weyl, E. G. (2010). A price theory of multi-sided platforms. *American Economic Review*, 100:1642–1672.

Zhang, M., Xu, H., Tong, L., and Ye, T. (2018). International evidence on economic policy uncertainty and asymmetric adjustment of audit pricing: Big 4 versus non-big 4 auditors. *Journal of Business Finance & Accounting*, 45:728–756.