Spatial aspects of choice and competition

Inauguraldissertation zur Erlangung des akademischen Grades eines Doktors der Wirtschaftswissenschaften der Universität Mannheim

Jan Robert Aue

Frühjahrs- und Sommersemester 2020

Abteilungssprecher: Prof. Volker Nocke, Ph.D
Vorsitzender der Prüfungskommission: Prof. Dr. Martin Peitz
Referentin: Prof. Michelle Sovinsky, Ph.D.
Ko-Referent: Prof. Nicolas Schutz, Ph.D.

Tag der mündlichen Prüfung: 07.08.2020

Contents

Acknowledgements

I dedicate this thesis to my parents to whom I owe so much.

I also want to express my sincere gratitude to Michelle Sovinsky for her patient and knowledgeable supervision that enabled me to write this thesis. Michelle was always available for discussions about the direction of my research or to give her detailed and encouraging advice. I thank Inken for a careful revision of the manuscript, and for many other things that have no place here. My thanks also go to Alessandra, Esteban, and Yihan for the many productive and often cheerful discussions that we had in our thesis group seminar. And I thank the market design group at the ZEW who have been great colleagues for the past two years during which I was able to work on many interesting research and applied projects.

This work was supported by the University of Mannheim's Graduate School of Economic and Social Sciences.

List of Figures

List of Tables

1. Introduction

This dissertation explores two main topics: first, it asks how retail firms interact with each other at close quarters; and second, it explores how students benefit from having more choice in the decisions to choose a secondary school. Both topics have a spatial dimension: retail firms compete with other retail firms who are nearby and, as we establish, students consider commuting times as an important selection criterion so they choose schools which are closer more frequently. These aspects motivate the reference to the spatial dimension of choice and competition in the title of this dissertation. This thesis consists of three independent chapters, which I summarize in turn.

1.1. Dynamic spatial entry

The first two chapters of this dissertation are concerned with the empirical analysis of dynamic spatial entry models. The first chapter examines the effect of price competition on the location choices of retail pharmacies in large cities. I exploit a regulatory change in 2004 that introduced price competition for non-prescription drugs to estimate the parameters of a dynamic spatial entry model. To this end, I use a comprehensive panel dataset of retail pharmacy locations that was especially constructed for this purpose. I estimate the dynamic model by means of a nested fixed point approach, because the asymmetric nature of the entry game renders conventional two-step estimators inapplicable. The computational burden of this approach is alleviated by tailoring the concept of an oblivious equilibrium to the spatial nature of the game, resulting in what I call a spatial oblivious equilibrium. I find that the regulatory change lead to more intense local competition and lower entry costs. The estimated structural model is then used to decompose the effects of the regulatory change on market structure and consumers' travel distances. I find that one third of the total decline in the number of pharmacies between 2004 and 2016 is attributable to increased local interaction, whereas this caused the consumers' distance to the nearest pharmacy to increase only marginally. This result suggests that price competition benefits consumers not only because it lowers retail prices, but also because it leads to a more efficient spatial distribution of retail pharmacies. These results also apply in other contexts where firms simultaneously decide on the placement of their horizontally differentiated products.

The second chapter revisits the spatial dynamic entry model developed in the first chapter, and takes a closer look at the estimation and identification aspects. I use a simplified version of the model that has only three regular parameters of interest to assess the properties of the nested fixed point, maximum likelihood estimator by means of a Monte Carlo study. I classify the model parameters in two categories: regular parameters that directly determine the payoff structure, and hyper-parameters which govern the nature of the strategic and spatial interaction. I find that the regular parameters can be consistently estimated and that they converge to a normal distribution as the size of the observed markets increases, provided that the model is specified correctly. This result is in line with standard results on extremum and maximum likelihood estimators, and it serves to illustrate that the method has been implemented correctly. I also find that the model's regular parameters can be consistently estimated even if some hyper-parameters of the model are unknown, and must be estimated via a grid search procedure. Furthermore, my results imply that the model's period return functions, but not the entry cost parameter, can still be estimated if the researcher does not observe all potential entry locations of the dynamic spatial entry game. This is a reassuring finding because in practice, the researcher often only observes rather short time horizons and so it is highly likely that some potential entry locations are never observed in this short time period. At last, the Monte Carlo simulations have also pointed to the limitations of nested fixed point procedures, because the computational burden becomes rather excessive even for the simplified strategic interactions that were developed in the first chapter. Thus, the results demonstrate the need for developing the concept of the spatial oblivious equilibrium further so as to accommodate computationally efficient k-step methods.

1.2. School district consolidation

The third chapter concerns a different topic altogether and is the result of a joint effort together with Josue Ortega, who is a lecturer at Queen's University in Belfast, and with Thilo Klein, who is a researcher at the ZEW and a professor at Pforzheim University. We study the welfare effects of school district consolidation, when both the de-centralized school markets as well as the consolidated school market determine the allocation of students to schools by means of the deferred acceptance algorithm. We show theoretically that there are expected welfare gains from district consolidation for all students, in particular for those who belong to smaller and over-demanded districts.

Using administrative data from the Hungarian secondary school assignment mechanism, we compute the actual welfare gains from district consolidation and compare these to our theoretical predictions. Hungary has a nationwide, consolidated school market which uses the deferred acceptance algorithm to assign students to schools. In this system, students submit their rank order lists of arbitrary length to a matching platform, and schools provide a ranking of their applicants. Then, the deferred acceptance algorithm is used to allocate school seats to students. We compare the outcome of this deferred acceptance algorithm when the market is consolidated, which corresponds to the status quo, to a counter-factual scenario in which the school market is split into several smaller districts. To this end, we us administrative data that includes students' rank order lists, their standardized test scores and socio-economic background, and schools' applicant rankings.

As an important building block of our empirical strategy, we describe a method to consistently estimate students' preferences across schools, and vice versa. Our method is based on the additive random utility framework and it corrects for the strategic reporting bias by using a combination of two identifying assumptions to construct latent feasible choice sets for students and for schools. Our method is implemented as an R package, and is made available for use by other researchers. We also conducted a Monte Carlo study to illustrate that the method works as intended and is indeed robust to strategic misreporting of students' preferences.

We find that students prefer schools that are nearby; they value schools with a high average academic achievement; and we also find evidence for assortative preferences. Because the status quo in the market under consideration is the consolidated school market, many students do not rank any school from their home district, which is why we use the estimated preferences to construct hypothetical complete rank order lists over all schools. We use these complete rank order lists to compute the deferred acceptance assignment in the consolidated and in the counter-factual district-level school markets. The additive random utility framework also allows us to determine the welfare gains and losses that students incur as a result of market consolidation. Because our utility specification includes a travel distance term, we can express the gains from district consolidation in terms of distance equivalence units. Our results imply large welfare gains from district consolidation for students, equivalent to attending a school five kilometres closer to the students' home address. We find that students with higher academic ability have significantly higher welfare gains from consolidation, but that there is also a large idiosyncratic component to explaining the consolidation gains. Students living in smaller markets, or in markets with less capacity, gain more from market consolidation than do other students. The secondary school market in Hungary is characterized by having a lot more nominal capacity than there are students, and so our results are not necessarily applicable to school markets which have less excess capacity. Indeed, if the educational market as a whole is balanced with little excess capacity, then the average welfare gains are much smaller, and the median student neither benefits nor looses from consolidation. In such cases, there may be districts in which the share of students who gain due to market consolidation is smaller than the share of students who loose due to market consolidation. These effects are highly dependent on the specific market circumstances, but they show that it can be difficult to obtain a majority consent for market consolidation.

1.3. Discussion

The main findings in this thesis are that more competition can improve the spatial allocation of retail pharmacies from a consumer perspective, and that bigger school choice markets can be beneficial for students because the positive choice effects outweigh the added competition effects. Taken together, these results may be interpreted as unequivocal support for market liberalization and consolidation. But far from it. The German retail pharmacy market is still regulated in many aspects and the introduction of price competition for non-prescription drugs has been a rather small shift towards market liberalization. I have merely shown that this small intervention has likely lead to a more efficient spatial allocation of pharmacies, but this does not imply that a more thorough liberalization, possibly including prices for prescription drugs, would be equally beneficial. On the contrary, it is quite possible that this would result in too much spatial differentiation (d'Aspremont et al., 1979), implying much higher travel costs. Also, I restricted the empirical analysis to urban markets and it need not generalize to rural markets where travel and hence competitive patterns are different. Likewise, although we find that the consolidated school market in Hungary benefits most students as opposed to the hypothetical unconsolidated market, this result may not translate to other settings. For instance, the Hungarian school market is characterized by substantial nominal excess capacity, but we found welfare gains to dwindle in a completely balanced market lacking excess capacity. Moreover, the society's objective function may include the students' aggregate welfare, but it could also include distributional or fairness motives. And so I prefer to draw a cautious conclusion from this thesis must remain cautious: In two important markets – allocating students to public schools in Budapest, and providing pharmaceuticals to the German public – competition and market consolidation can be socially beneficial, but this assessment depends on the specific circumstances, and on the objectives of the society.

2. Spatial effects of price regulation and competition

2.1. Introduction

Modern developed economies spend about a tenth of their national incomes on health care, and pharmaceuticals contribute a sizeable portion of this spending block. Figure 2.1 shows that the expenditure share of pharmaceuticals alone is close to two percent in major industrial nations. And while online pharmacies are pushing into this large market, the majority of prescription drugs is still sold in retail pharmacies:[1] their emblem is ubiquitous in many cities. This is also true in Germany, the world's fourth largest market for pharmaceuticals: In 2004, one fifth of all pharmacies were located closer than 110 metres apart from their nearest competitor.[2] But to the consumer (or patient) in need of a prescription drug, two adjacent pharmacies are in no way better than just one single pharmacy, because prescription drugs are subject to quality and price regulations. Therefore, a more dispersed spatial allocation that reduces travel costs would be preferable from a consumers' perspective. In this paper, I develop a spatial entry model to show that such a pattern has gradually emerged in Germany as a result of a regulatory change in 2004 which introduced price competition for non-prescription drugs. The topic is of current interest because the European Commission and the European Court of Justice have recently challenged the German system of fixed prices for prescription drugs. My contribution is twofold: First, I document a case where the introduction of price competition had a profound impact on the spatial market structure. Second, I develop a feasible method to estimate a dynamic spatial entry model with a large number of asymmetric agents and a very flexible notion of "space". I motivate my research with an illustrative theoretical model of space-then-price competition.

I illustrate the mechanism through which price competition affects location choices by means of a simple variant of the classical Hotelling model. In this model, retail firms typically face the trade-off between choosing a central location to attract high demand ("market share effect") and differentiating themselves from their competitors to increase their local market power ("market power effect"). The market share effect should lead to spatial clustering, while the market power effect should lead to spatial dispersion. If competition is mitigated due to price regulation, the market power effect should therefore become more dominant and lead to more clustering, and vice versa. As a result, the inter-firm distances

[1]The German statistical office puts the share of revenue from e-commerce in this large retail sector below 1. (destatis, 2019, table 45341-0001)

[2]Source: Deutscher Apotheker Verlag (2016), own calculations

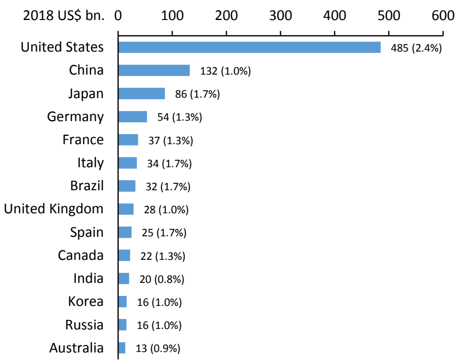


Figure 2.1.: Spending on pharmaceutical products in 2018, by country. Share of national GDP in parentheses. Source: author's representation based on IQVIA (2019).

should increase, while the consumers' average travel distances should decrease. This hypothesis will be examined for the retail pharmacy sector in Germany. Of course, the process of gradual re-locations that is at the heart of the Hotelling model cannot be observed empirically. It should rather be considered as an approximation to the dynamic process of entry and exit which forms the aggregate market patterns.

Until 2003, the German market was characterized by three distinctive institutional features: retail prices for prescription and non-prescription drugs alike were fixed, pharmacies were only allowed to operate as a single-store business, and no minimum distance regulations were imposed. A major health reform in 2004 changed the first two aspects: it introduced price competition for non-prescription drugs and allowed local pharmacy "chains" of up to four branches. I document that, following this reform, the number of pharmacies declined by about six percent, while the average consumer's distance to the nearest pharmacy only increased by a small amount, which suggests that the decline in the number of pharmacies was largely due to intensified competition between, and exit of, nearby or adjacent stores. In order to isolate and quantify the effect that the introduction of price competition had on pharmacies' location choices, I develop a dynamic structural model of spatial entry and fit it to the data on pharmacy locations. Using this model, I find that the period following the reform can be characterized by lower entry costs, lower period returns, and more intense competition among nearby competitors. I interpret the latter as a direct consequence of introducing price competition for non-prescription drugs and use simulations to isolate its effect on aggregate outcomes and consumer travel distances. The simulation results show that the price competition effect alone can explain one third of the observed decline in the number of pharmacies, but only one fifth of the observed increase in consumer travel distances. Therefore, the introduction of price competition contributed to a very consumer-friendly change in the spatial distribution of stores, with lower overall fixed costs and only marginally higher travel costs.

One particularly illustrative example is shown in figure 2.2. That figure shows the location of five pharmacies in the small town "Hagenow" with about twelve thousand inhabitants. What is striking about the picture – apart from the fact that such a small town can sustain five pharmacies – is the observation that they are all situated very close to what one may call the "city centre", an archetypical

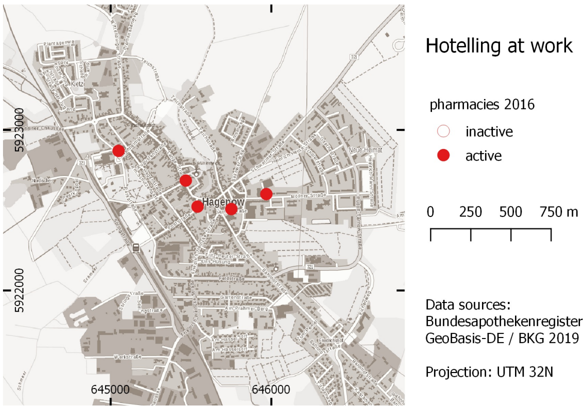


Figure 2.2.: Hotelling at work: agglomeration of pharmacies in a small town. Source: author's own representation.

outcome of the classical Hotelling model: competitors exhibit an "undue tendency [...] to imitate each other in quality of goods, in location, and in other essential ways" (Hotelling, 1929). I show that such inefficient market outcomes have become less prevalent due the health care reform of 2003.

I have structured this chapter as follows. Following a description of the institutional background, the relevant literature on spatial competition and location choice is reviewed in depth, and related to my approach. Section four develops a very stylized model of space-then-price competition to motivate the research question. Section five sets up a dynamic spatial entry model, section six describes my data and section seven applies the model to the data and estimates its structural parameters. The last section concludes.

2.2. Institutional background

This section briefly describes the German pharmacy market and then outlines the most relevant legislations. More detailed accounts of the pharmacy market in Germany can be found in Horvath (2010) or Coenen et al. (2011).

2.2.1. Business structure

In order to obtain a brief overview on the market under consideration, some key figures from two years, 2005 (the earliest year for which these data are available) and 2015, are compiled in table 2.1. The table shows that the industry has undergone some important changes from 2005 to 2015. First, the number of stores has declined by approximately six percent. After having reached a peak of 22 thousand stores in 2005, there are currently around 20 thousand pharmacies in Germany which amounts to 25 stores per 100,000 inhabitants. More importantly, the net profit margin – that is, the overall profitability of a store relative to its generated revenues – has halved from twelve to six percent. That figure is roughly

	2005	2015
Stores	21,968	20,639
Employees ('000)	157	210
Revenues per store ('000)	1,392	2,110
% prescriptions drugs	–	83
% from e-commerce	0.3	1.0
Number of packages sold (bn.)	–	1405
% prescriptions drugs	–	53
Gross earnings on sales per store† ('000)	402	509
% margin on revenues	29	24
Operating profits per store§ ('000)	159	125
% margin on revenues	12	6

Sources: Destatis (2017), ABDA (2016)
† revenues less wholesale cost
§ Earnings after costs, wages, taxes, rents

Table 2.1.: Business indicators of retail pharmacies.

comparable to the profitability of bakery shops in that period (Destatis, 2017). The table further shows that retail pharmacies are rather small, with revenues that average around 2m Euros in 2015 and roughly ten employees per store. Pharmacies sell medicine to patients that can be classified into prescription (Rx) and non-prescription, or over-the-counter (OTC) drugs. The table shows that prescription drugs account for the bulk of aggregate revenues, but make up only slightly more than half of the packages that were sold. The share of revenues generated online is still very small in this retail sector. However, online resellers from abroad, who are not subject to price regulation on prescription drugs[3] are pushing into the market meaning this market segment could become more important in the future.[4]

2.2.2. Regulatory framework

The retail pharmacy market is subject to a large body of regulations that sets standards for the operation of pharmacies. This regulatory framework consists of several separate laws which, taken together, determine who may operate a pharmacy, set standards for the establishment of one and govern the compensation schemes. The relevant regulations are summarized below.

First and foremost, the German pharmacies act (*Apothekengesetz, ApoG*) lays out the general conditions under which a pharmacy may be operated. It states that pharmacies are responsible to guarantee the "proper supply" of medication to the population. A pharmacy may only be operated by a certified pharmacist who has obtained a licence from the authorities. This licence expires if the business ceases to exist or if the operator dies. While a pharmacy may be operated jointly by more than one pharmacist (each of whom requires a licence), partnerships which make the compensation of one partner, indirectly or directly, contingent on profits or revenues, are in general not allowed. Neither may a pharmacy commit to exclusively sell the products of certain manufacturers, or strike special deals with physicians

[3] Ruling of the European Court of Justice, Case C-148/15, retrieved from http://curia.europa.eu on 5 June 2020.

[4] On July 08, 2017, the Swiss company "Zur Rose Group AG" has collected CHF 200m with its IPO and declared that it would use the proceedings to expand its German online business. (BZ, 2017)

to prescribe a certain range of products. All pharmacies are obliged to participate in a scheme which guarantees the provision of emergency services during night times or on public holidays. It is admissible for pharmacies to distribute products by post, although the numbers from Destatis (2017) suggest that this is a niche market. Since 2004, a licenced pharmacist may obtain permission to operate up to three subsidiary branches that must be in the same district as the main branch, or in an adjacent one. Each subsidiary branch must be operated by a licenced pharmacist, and fulfil all the requirements of a regular pharmacy with the exception that it need not have an own laboratory. Most importantly, the pharmacist is free to choose the location of his or her pharmacy, subject of course to residential zoning regulations but independent of the locations of other competitors.[5]

Further legislation is delegated to the ordnance on the operation of pharmacies (*Apothekenbetriebsordnung, ApoBetrO*) issued by the federal health ministry: first, the pharmacist who operates a pharmacy must do so in person, *i.e.* they cannot hire a manager to run the store. Every pharmacy must have a floorspace of at least 110 sq.m. and a laboratory that is fully equipped to produce custom medications, unless it is a subsidiary branch of another store in which case a laboratory is not mandatory. Stocks must be sufficient to cover the needs of the population for at least one "average" week, notwithstanding the obligation to always maintain further stocks of medications and vaccines for emergency purposes.[6]

In most European health insurance systems, patients pay some share of the costs of their prescribed medication out of their own funds (Panteli et al., 2016) and this is also the case in Germany. In what follows, the price that the patient sees under this cost-sharing rule will be referred to as the retail price. Usually, this retail price is a function of the list price, and it is in general the same for all members of a public health insurance. Unlike in most other retail markets, the pharmacy's variable profit per unit of prescription drug sold is not the difference between retail prices and list prices. Instead, markups are regulated directly, again as a function of the list price.[7] Figures 2.3a and 2.3b show how the implied markups and consumer retail prices as a function of list prices changed due to the reform in late 2003. The figure implies that prior to the reform, pharmacies had a strong incentive to sell expensive drugs, while consumers had no or little incentive to ask for cheap generics. The reform partially reversed this, as pharmacies now have a very small incentive to sell expensive drugs, while patients now have a stronger incentive to ask for a cheaper generic product. Until 2003, retail prices of OTC and Rx drugs were both regulated. The health care reform in late 2003 changed this: the price regulation scheme for non-prescription drugs was abandoned so that today, roughly half of all packages accounting for 15% of total revenues are sold competitively (see table 2.1). On the other hand, retail prices and markups of prescription drugs remain subject to regulation.[8]

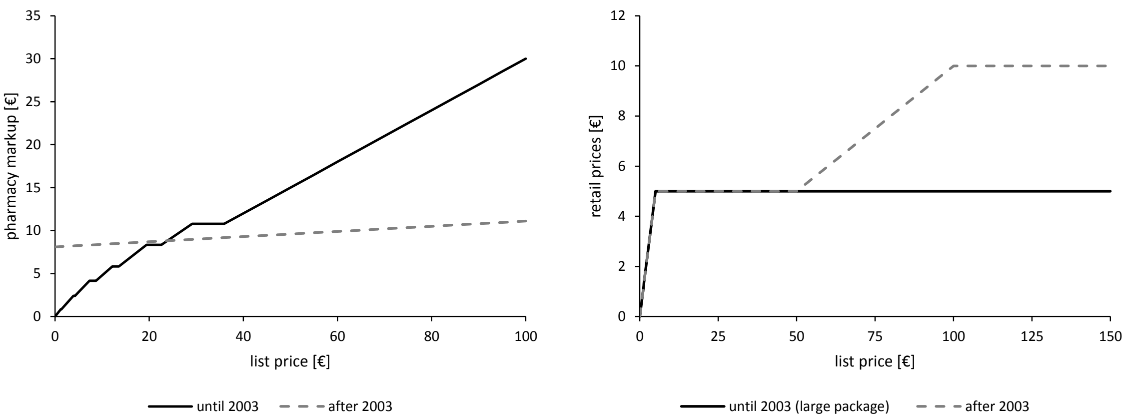
---

[5]§§1, 3, 8, 9, 10, 11, , 11a, 14, and 18 Apothekengesetz (ApoG), retrieved from gesetze-im-internet.de/apog on 28 June 2017

[6]§§2, 4 and 15, Apothekenbetriebsordnung (ApoBetrO), retrieved from gesetze-im-internet.de/apobetro_1987 on 28 June 2017

[7]*cf.* §3 Arzneimittelpreisverordnung (AmPreisV) as of 1 January 2002 and 11 May 2019; and §§31,61 SGB V as of 1 January 2003 and 1 January 2005. Retrieved from research.wolterskluwer-online.de on 24 July 2019.

[8]See Art. 1 (39,92,94) and 24 (1,3), GKV Modernisierungsgesetz. *Bundesgesetzblatt I*, 2003(55):2190–2258



(a) Regulated markups for prescription drugs

(b) Retail prices of prescription drugs

Figure 2.3.: Price regulation for prescription drugs, until and after 2003. Source: own representation based on the relevant legislative texts.

The European Commission (EC) has repeatedly called on its member states to liberalize their pharmacy markets, and was often supported in its view by the European Court of Justice (ECJ).[9] While the EC sees pharmacies as part of the retail sector and applies the rules of the common market to it, the prevailing view in some member states, manifested in their regulatory frameworks, regards pharmacies as part of the health care system where price competition should not play a role. Therefore, it is possible that the near future will see significant changes of the regulatory regimes in Europe.

To summarize, the health care reform of 2003 has changed the pharmacies' compensation scheme in Germany, it has likely changed the entry costs by permitting up to three subsidiary branches, and it has presumably increased competition among adjacent pharmacies by introducing price competition for non-prescription drugs. These considerations will guide my empirical approach.

2.3. Related Literature

2.3.1. Theoretical literature

For a long time, economists and social scientists have discussed the question of where and how economic agents locate. Nearly two centuries ago, Johann Heinrich von Thünen provided an economic

[9]For example, the European Commission (EC) has urged members to take action in the following areas: legislation restriction the freedom of establishment in Italy, Spain and Austria (EC press release IP/06/858, ECJ ruling C-367/12); legislation restricting the number of pharmacies that may be owned in Italy (EC press release IP/06/1789); and legislation concerning the delivery of pharmaceuticals (EC press release IP/09/438). In 2016, the ECJ ruled that online pharmacies that are located in a member state of the European Union outside Germany are not obliged to comply with the German price regulation scheme for prescription drugs if they ship to Germany (ECJ ruling C 148/15). In July 2019, the EC reiterated its request for Germany to abandon its price regulation scheme (EC press release MEMO-18-3446). In response, the German cabinet has drafted a law that makes adherence to the German price regulation scheme a precondition for medical expenses to be accounted for with the German public health insurers ("Gesetz zur Stärkung der Vor-Ort-Apotheken", currently under parliamentary revision, document no. 373/19, retrieved from bundestag.de on 5 June 2020).

explanation for different agricultural structures around cities. Weber and Pick (1909) discussed where production facilities should optimally be located, taking into account the locations and transport costs of different inputs and keeping market conditions constant. Next Christaller (1933) developed theories on the ideal spatial constellation of cities (or "central places"), and Lösch (1940) adopted a more general equilibrium approach which predicts that economic agents will be positioned such that each unit serves a hexagonal market area giving rise to a "honeycomb pattern" of market areas. Neither of these authors has considered the problem of price competition. A review of these classical approaches can be found in Kulke (2013) and Fischer (2011). An interesting extension to the works of Lösch and Christaller is Rushton (1972) who computes optimal market structures under non-uniform consumer distributions using a numerical approach. This leads to skewed point patterns, while the original hexagonal structure is still visible.

Taking into account that firms choose their locations strategically, and also compete on attributes other than their location complicates the analysis. Hotelling (1929) was the first to point out that profit maximizing competitive firms may choose to agglomerate, thus inflicting inefficiently high travel costs on the consumer side. He applies his finding to the political economy sphere:

> The competition for votes between the Republican and Democratic parties does not lead to a clear drawing of issues, an adoption of two strongly contrasted positions between which the voter may choose. Instead, each party strives to make its platform as much like the other's as possible. (Hotelling, 1929, p. 54)

Of course, there is no price competition in politics so what may be true for political parties must not necessarily hold for competitive firms. This was already apparent to Hotelling who noted that "Bertrand's objection applies" as soon as both competitors are in the same place (p. 52). An early extension of his work is Smithies (1941) who introduced elastic aggregate demand, thus disposing of the zero-sum nature of the original set up. Smithies finds that minimal differentiation is not a necessary outcome under this assumption. This finding has been confirmed by d'Aspremont et al. (1979) who even show that Hotelling's model with linear transport costs does not have an equilibrium at all, and, by assuming quadratic transport costs, derives a contrary outcome: that firms will optimally tend to locate at both extremes of the market. From there onwards, quadratic transportation costs have become a standard in the theoretical literature. For instance, Bester et al. (1996) is an in-depths game theoretic analysis of the location-then-prices model with quadratic transportation costs and deterministic consumers and Anderson et al. (1997) derive conditions for general, non-uniform population distributions under which an equilibrium exists.

The difficulties with establishing an equilibrium outcome are closely linked to firms' incentive to undercut each others' prices when both firms are located at the same point and perfect competition causes consumers to purchase the cheapest product (adjusted for travel costs) in a deterministic way. Thus, as soon as consumers are assumed to posses preferences over attributes other than the delivery price, price competition is softened and this problem can be expected to be alleviated. The first to note this were de Palma et al. (1985) who modelled consumer demand for two spatially differentiated firms by using the discrete choice framework that has now become standard in the literature on empirical industrial organization. Consumers care about travel costs and prices, but also posses idiosyncratic preference shocks over visiting different stores, and firms are assumed to sell their products at a given price. As a consequence, consumers may now purchase from a store that offers a higher travel cost adjusted retail price because this store appeals to the consumer in some other, unknown dimension. It will now be worthwhile for both firms to locate at the centre, because consumer heterogeneity "eliminates discontinuities in the profit function" (p. 771) so that any deviation from the market centre leads to lower demand. A similar conclusion is reached by Ben-Akiva et al. (1989) who introduce price competition and a second dimension along which products are differentiated, but which the firms do not choose strategically. This set up achieves the same effect in that it eliminates the incentive to undercut and thus makes the agglomeration a feasible equilibrium. Finally, (Anderson et al., 1992, p.343–392) combine logit demand, price competition and location choices along the real line. They show analytically that the location-then-prices game has a centralized equilibrium if consumers are sufficiently heterogeneous. They also solved their game numerically with two firms and locations on the real line, yielding both symmetric non-central equilibria and agglomeration as the outcome. They do not derive de-centralized equilibrium locations analytically.

The literature in the previous paragraph modelled space as a unidimensional line for the sake of analytic simplicity. Yet, most spatial patterns observed in the real world are two-dimensional. Eaton and Lipsey (1975) were the first to take spatial competition to the two-dimensional space, by simulating the movements of up to seventeen firms that sequentially re-locate so as to maximize their market shares *while keeping prices fixed.* They find that the honeycomb pattern of Lösch quickly breaks up and thus the authors "strongly suspect, but as yet cannot prove, the non-existence of any equilibrium configurations in the disc beyond $n = 2$" (Eaton and Lipsey, 1975, p. 44). Irmen and Thisse (1998) consider the case with quadratic transport costs and deterministic consumers who care about multiple product characteristics. Their finding is that firms will, in equilibrium, choose maximum differentiation in the dimension which consumers care about most, and minimum differentiation in the other dimensions. A similar result appears to have been simultaneously derived by Ansari et al. (1998) for the case of three product attributes. A more recent attempt to characterize equilibrium locations in a competitive environment is worked out by the four computer scientists Ottino-Loffler et al. (2017) who follow the approach of Eaton and Lipsey (1975) and find stable spatial patterns with up to seven firms, using deterministic consumers who care about travel costs. Yet, their pricing stage is not modelled explicitly and they employ a rather coarse grid such that their results should be considered with some caution. In a theoretical paper, Vogel (2008) studies location-then-price equilibria on the unit circle and finds that more productive firms locate in more isolated areas. Yet, his main contribution is a mathematical trick that establishes equilibrium existence in such a game despite using linear transportation costs. In a more recent paper, Allen and Arkolakis (2014) derive the existence of spatial equilibria in a general equilibrium setting. A continuum of consumers (who are also workers) equipped with CES preferences is distributed across a compact two-dimensional space. Each consumer also produces exactly one unit of the output good, and bilateral trade is then governed by a gravity equation. A spatial equilibrium is a distribution of consumers/workers such that their incomes equal their expenditures, and there are no profitable relocations. Finally, a recent review article by Biscaia and Mota (2013) reiterates that research on location-then-price competition in the two-dimensional plane is still not so abundant.

Specifically concerned with price regulation in the health care sector are theoretical papers by Brekke et al. (2006, 2011). Both add a quality dimension to firms' decision space letting firms effectively compete on locations and on quality. The theoretical paper (Brekke et al., 2006) is cast in a Hotelling duopoly environment with a linear choice of location and quality-specific investment. Representative consumers have a linear utility function and care about the price, quadratic transportation costs and quality of the service. The basic insight of their model is that, while price regulation leads to an agglomeration of firms, quality competition counteracts this force, leading to more spatial differentiation. A regulator who cares for social welfare would attempt to choose a price so as to maximize welfare, taking into account the subsequent quality and location decisions of firms. If location decisions are exogenous, the optimal quality level can be attained whereas if both location and quality choices are exogenous, the second best outcome would either entail too much spatial differentiation and too low quality or vice versa. In Brekke et al. (2011), the relation between quality choice and competition is examined theoretically, but the total number of hospitals and their location is taken as given.

2.3.2. Empirical literature

For the empirical analysis of spatial competition, it is necessary to model both consumers' spatial demand, and firms' location choices. Modelling spatial demand empirically does not pose substantial challenges. The standard ARUM model of consumer utility over spatially differentiated alternatives with random coefficients and logit choice probabilities can be found in papers such as Davis (2006), Ho and Ishii (2011), Crawford (2012) and the references therein and Aguirregabiria and Vicentini (2016). Most of the insights from Berry (1994) can be readily applied to the spatial context, using a distance function that enters a consumer's utility function. Further works who use this framework are Chisholm and Norman (2012) and Davis (2006) who examine the market for cinemas and Ho and Ishii (2011) who study retail banking.

It is far more difficult to model location choice in a tractable, yet plausible and flexible way. Some implications of the theoretical literature on multi-attribute competition in a spatial context have been tested empirically by Netz and Taylor (2002) who find evidence that gasoline stations respond to tougher competition with more spatial differentiation. Further, Thomadsen (2007) uses a structural pricing model to derive demand parameters and builds on (Anderson et al., 1992, p.343-392) to examine the optimal location decisions under different cost structures. Yet, his counter-factual simulations are restricted to spatial competition along one dimension. Contrary to the theoretical literature, empirical models of the supply side often assume that firms choose to enter in a discrete set of locations and derive equilibria under incomplete information. This approach was pioneered by Seim (2006) who uses relatively few locations and a reduced form profit function.

A few studies exist that are dedicated to competition in the retail pharmacy market. Horvath (2010) analyses the nexus of price regulation and quality and spatial competition on the German pharmacy market. However, he reviews models of quality and circular spatial competition and free entry which are less suited to study location choices because firms will usually locate at equidistant locations around the circle. His empirical results are based on rather coarse county-level data. Hence, an important extension of this work is to model spatial entry in a more detailed fashion, and to link it more tightly to detailed data sources. A related study published by Coenen et al. (2011) also features a very detailed institutional description of the pharmacy market in Germany. In a scenario-based approach, they compare different reform options with regard to their cost saving potential. The competitive reactions of firms in the market are not endogenously determined in their approach. In both aforementioned studies, the focus seems to be more on the equilibrium number of pharmacies rather than on their location choices. Similarly, Schaumans and Verboven (2008) set up a model of joint entry by physicians and pharmacies in small local markets but again, the focus is on entry and exit as opposed to the small-scale geographical distribution of economic units.

2.3.3. Dynamic empirical literature

The empirical approach that this paper presents relies on a dynamic entry game with a large number of players. These players interact with each other the more intensely, the closer they are together, but in principle, every agent interacts with every other agent. In solving dynamic discrete games with many players such as the one described below, a direct solution of the game becomes infeasible even for moderately-sized problems. The crucial point here is that the size of the state space grows exponentially in the number of players, which poses computational problems for two principal reasons: first, the large amount of computer storage required to store the entire value function, and second, the computational burden associated with computing an expectation over the future state space. The state space of the game presented above is of magnitude 2^N and so will be usually too large to estimate the full model using conventional methods – a typical city in Germany has around eighty pharmacies, and it is often not desirable to delineate ad hoc market boundaries.[10] The literature on dynamic discrete games has put forward a few approaches to modify the problem in such a way to be able to solve it, or at least to be able to estimate its key parameters. These approaches will be discussed in turn.

A first branch of the literature has evolved around the so-called two-step estimators that were initially proposed by Hotz and Miller (1993) and later refined by Aguirregabiria and Mira (2002) for single agent decision processes. Two-step (or k-step) estimators rely on obtaining non-parametric estimates of either the policy function, or the continuation values in a first stage, which are then used to compute the players' best responses and a likelihood function in a second stage. These ideas have been applied to

[10]To get an idea of the magnitudes, $2^{80} \approx 10^{24}$ so even storing the value function infeasible.

games with many players by Aguirregabiria and Mira (2007) who use estimates of the conditional choice probabilities, and by Pakes et al. (2007) who estimate continuation values in a first step. Bajari et al. (2007) extend these methods to allow for continuous choice variables. A recent addition to this literature is proposed by Aguirregabiria and Magesan (2019) who study dynamic entry when players' beliefs are not in equilibrium. Their approach also relies on obtaining non-parametric choice probabilities in a first stage. But while being computationally efficient and elegant, two-step estimators are unfortunately not applicable to my setting, for the following reasons. First, the size of the state space which is much larger than the number of observed decisions obviously prevents a direct application of the concept. Second, the nature of spatial competition is not symmetric: in a certain state of the world, one firm may face a lot more local competition than another firm and so their policy choices, or their strategies, would be very different. And even if one conditions on the local environment of each firm, certain configurations of their nearest competitors mean very different things to different firms due to differences in their relative spatial constellation to each other. Thus, in a spatial entry game, the asymmetry of players' strategies prevents a direct application of these two-step estimators.

The asymmetry of the problem at hand instead calls for a nested fixed point estimator where each firm solves a distinct dynamic problem and therefore has its own policy function. To alleviate the computational burden associated with solving such a dynamic discrete model, several approaches have been put forward that will be reviewed first, followed by a description of the approach followed in this paper.

First, Doraszelski and Judd (2012) set up a dynamic model in continuous time which significantly reduces the computational burden of calculating the expected future state. The expectation is easier to compute because it becomes increasingly unlikely that more than one player makes a move, as the time periods become shorter and eventually approach a continuum. Therefore, at each point in time there are only as many possible future states as there are players in the game. However, an empirical application of this setting requires that the precise timing of all entry and exit decisions be known, which is not the case in the data at hand.

Pakes and McGuire (2001) propose a stochastic algorithm to compute an approximation to the MPE. The algorithm relies on the fact that in many cases, the Markov chains that are induced by dynamic discrete games have a very large state space, but eventually wander into a smaller set of states that is known as the recurrent class. Their algorithm draws a new state in every iteration according to the current policy function, and updates the current state's continuation value with the new state's value. It reduces the number of states that are visited, and the computational burden associated with computing the expectation over future states, but adds a simulation error to the problem so that more iterations are necessary for the problem to converge. Their ideas cannot readily be applied to the setting at hand, mainly because the sampling procedure is not guaranteed to sample uniformly from the recurrent class of the game. The recurrent class of many dynamic discrete games encompasses the entire state space because the transition cost shock has, by assumption, full support, so that "anything goes" (albeit it may do so with very low probability). However, even if the MPE is unique, the transition dynamics in some games are likely to imply that the recurrent class is partitioned into a number of sub-classes which are almost recurrent in themselves. Once the Markov Chain that is induced by equilibrium play has wandered into one of these subclasses it is likely to remain there (although it is not certain that it does so). To make this point clearer, consider a dynamic entry model in which an even number of firms is located on a circle, and let the parameters be such that having two active direct nearest neighbours leads to negative (or very small) period returns, while the presence of an active indirect neighbour does not affect profits. If the market entry costs are chosen sufficiently large, it becomes equally likely that only firms with even numbers are active, or only firms with odd numbers. Small permutations of these configurations may arise, but it will be very unlikely to see a complete reversal of fortunes. Therefore, an unguided sampling procedure similar to the one described by Pakes and McGuire (2001) is likely to miss a large part of the state space that may be equally likely to occur as the one that was sampled, and so the procedure is not well suited for the empirical application at hand.

A different approach is taken by Weintraub et al. (2008, 2010) who develop an equilibrium concept in an entry game with quality investment that they call *oblivious equilibrium*, wherein individual firms condition their actions only on their own state and on the long-run average aggregate industry state. This aggregate industry state can be assumed to remain approximately constant over time if the number of firms and potential entrants is large so that individual decisions are averaged out. A requirement for being able to condition the decision process is that the period returns of each firm depend only on its own state, and on an aggregate state because all competitors are the same. Thus, the oblivious equilibrium is easier to compute because it greatly reduces the dimensionality of the problem. But in a spatial context, not all competitors are the same and it is difficult to reduce the spatial distribution of firms to an aggregate statistic with low dimensionality so that the concept of oblivious equilibrium is not readily applicable. Furthermore, the separation of the state space into an aggregate industry state that remains constant, and an individual state, effectively restricts the permissible state space very strongly. In my empirical spatial entry model, I will adapt and extend the oblivious equilibrium concept of Weintraub et al. to the spatial domain in order to address the aforementioned concerns. This extension of the oblivious equilibrium concept constitutes the main methodological contribution of this paper.

2.3.4. Contribution

This paper extends the literature on empirical dynamic entry models to incorporate a large number of heterogeneous, spatially interacting players without having to resort to simplifying symmetry considerations. I extend the ideas put forward in the concept of an oblivious equilibrium to develop a heuristic approach that reduces the dimensionality of the problem, and yet maintains the spatial and dynamic features of the model. As an empirical contribution, this paper presents evidence that the introduction of price competition in a retail market has profound effects on the spatial equilibrium distribution of retail firms' locations.

2.4. A stylized model

To illustrate the characteristic mechanisms and trade-offs of the market under consideration, and to motivate my research, I first set up a very simple modified Hotelling model of space-then-price competition on the real line. The analysis is related to the work of Brekke et al. (2006) who consider quality competition and location choices of hospitals.

Consider the case of two pharmacies competing on prices and locations: first, both pharmacies choose their location and next, taking location choices as given, they compete on prices. Denote the location of both stores by x_a and x_b and assume that admissible locations are restricted to the unit interval $[0, 1]$, and that $x_a \leq x_b$. Pharmacies sell prescription drugs (Rx) at a regulated price $\bar{r} = 0$ and non-prescription drugs (OTC) at price p_j, $j \in \{a, b\}$. To reflect the institutional characteristics of the retail pharmacy market, I assume that both pharmacies earn a regulated margin on each sold unit of a prescription drug. For simplicity, the regulated margin on prescription drugs is assumed to be equal to one. The marginal costs of non-prescription drugs are assumed to be zero, so that the margin on each unit of non-prescription drugs that is sold is equal to its retail price p_j. Hence, pharmacies' profits are given by

$$\pi_j = Q_j^{Rx} + p_j Q_j^{OTC}$$

where Q_j^{Rx} and Q_j^{OTC} are the demands for prescription and non-prescription drugs, respectively. A unit mass of consumers is distributed uniformly on this unit interval and indexed by their location $i \in [0, 1]$. Consumers incur quadratic travel costs and there is no outside option so that one consumer certainly purchases a product from one of the competitors. Each consumer purchases either a prescription drug or a non-prescription drug, but not both, from one of the stores. With probability α, the consumer purchases a prescription drug at the regulated price $\bar{r} = 0$. Her (normalized) utility from obtaining the drug at store j is

$$\nu_{ij} = -(i - x_j)^2$$

Therefore, consumers will always obtain their prescription drugs from the nearest pharmacy. On the other hand, a consumer purchases a non-prescription drug with probability $1 - \alpha$. Her utility when purchasing it from pharmacy j is given by

$$u_{ij} = -p_j - \tau(i - x_j)^2$$

I assume that $0 < \tau < 1$, as consumers purchase a prescription drug because they are seriously ill so they should have higher travel costs than those who are not. Since d'Aspremont et al. (1979) have shown that quadratic travel costs are sufficient to guarantee equilibrium existence, this has become a standard in the literature, and I abide by it.

2.4.1. Equilibrium

As a benchmark, consider first the trivial case were prices of non-prescription drugs are also regulated, so that $p_a = p_b = \bar{p}$, but firms choose their locations freely. This is the classical Hotelling model in which both firms tend to locate at the market centre so as to maximize their market shares, that is $x_a = x_b = 1/2$.

Next, I investigate whether the introduction of price competition for non-prescription drugs can lead to a market outcome that is preferable from a consumer's perspective. I focus on symmetric equilibria. The following proposition characterizes a symmetric space-then-price equilibrium of the model:

Proposition 1. *A symmetric space-then-price equilibrium of the model is given by location choices* $x_a = x^*$ *and* $x_b = 1 - x^*$ *with*

$$x^* = \begin{cases} 0 & \text{if } \alpha \leq \frac{\tau}{3+\tau}, \\ -\frac{1}{4} + \frac{3}{4}\frac{\alpha}{\tau(1-\alpha)} & \text{if } \frac{\tau}{3+\tau} < \alpha < \frac{\tau}{1+\tau}, \\ \frac{1}{2} & \text{if } \alpha \geq \frac{\tau}{1+\tau}, \end{cases} \tag{2.1}$$

and prices $p_a = p_b = p^*$ *with*

$$p^* = \begin{cases} \tau & \text{if } \alpha \leq \frac{\tau}{3+\tau}, \\ \frac{3}{2}\left(\tau - \frac{\alpha}{1-\alpha}\right) & \text{if } \frac{\tau}{3+\tau} < \alpha < \frac{\tau}{1+\tau}, \\ 0 & \text{if } \alpha \geq \frac{\tau}{1+\tau}. \end{cases} \tag{2.2}$$

The proof is standard, and it is given in appendix A.1 for completeness. This result encompasses two interesting polar cases: first, when the share of consumers purchasing price regulated prescription drugs, α, is equal to one, the locational equilibrium sees both firms located at the market centre – this is Hotelling's minimal differentiation result. Second, if pharmacies sell only non-prescription drugs ($\alpha = 0$), the market power effect dominates and both firms are located at the market boundaries. This corresponds to the principle of maximum differentiation of d'Aspremont et al. (1979). This result is interesting because it shows that the introduction of price competition for non-prescription drugs does not necessarily lead to an increase in spatial differentiation, as the centralized equilibrium may prevail for a large range of parameters. Whether this happens or not will depend crucially on local market circumstances that affect consumers' relative and absolute travel costs, and on the share of either consumer type.

2.4.2. Welfare analysis

Next, I compute consumer welfare and producer surplus in the competitive market, and compare these to the outcomes under complete price regulation, and to the social optimum. Throughout, I will assume that pharmacies can freely choose their location. Note first that the consumers' aggregate travel costs in any given symmetric locational equilibrium $x \in [0, 1/2]$, with $x_a = x$ and $x_b = 1 - x$, are given by $\tau T(x)$, where

$$T(x) = 2\int_0^{\frac{1}{2}} (i - x)^2 \mathrm{d}i = \frac{1}{12} - \frac{1}{2}x + x^2$$

Then, consumer welfare with retail prices of non-prescription drugs p and regulated prescription prices $\bar{r} = 0$ is

$$W(x, p) = -(1 - \alpha)p - (\alpha + (1 - \alpha)\tau)T(x)$$

and equilibrium profits are always given by $\pi^* = \frac{1}{2}(\alpha + (1 - \alpha)p^*)$.

Price regulation Consider first the case where both prescription and non-prescription drugs are regulated at prices $\bar{r} = 0$ and $\bar{p} \geq 0$, respectively, but firms can choose their locations freely. This case will be denoted as the regulatory benchmark in the discussion that follows. As was discussed above, the absence of price competition induces firms to locate at the market centre, and consumer welfare in this setting is given by

$$W\left(\tfrac{1}{2}, \bar{p}\right) = (1 - \alpha)\bar{p} - \frac{(\alpha + (1 - \alpha)\tau)}{12}$$

As is well known, the tendency of the two firms to bunch together at the market centre inflicts an inefficiently large amount of travel costs on consumers. The welfare maximizing location pattern would be the one that minimizes travel costs, with $x_a = \frac{1}{4}$ and $x_b = \frac{3}{4}$. Aggregate profits remain the same, but consumer welfare would increase to

$$W\left(\tfrac{1}{4}, \bar{p}\right) = -(1 - \alpha)\bar{p} - \frac{(\alpha + (1 - \alpha)\tau)}{48}$$

In fact, it is easy to see that any symmetric location pattern with $x_a \in (0, \frac{1}{2})$ and $x_b = 1 - x_a$ improves welfare compared to the laissez-faire case. Hence, a social planner could improve welfare by imposing a minimum distance regulation, but of course such a regulation could induce adverse effects that are not captured by this simple model: by effectively creating local monopolies, pharmacies have less incentives to invest in quality, or to expand opening hours – Brekke et al. (2006) discuss such aspects of quality competition in the context of hospital regulation in much greater detail.

Price competition Next, consider the case where pharmacies choose their locations first and then compete on prices for non-prescription drugs. For any given symmetric locational equilibrium $x^* \in [0, \frac{1}{2}]$ retail prices are given by $p^* = \tau(1 - 2x^*)$ where x^* is itself a function of α and τ. I am interested in conditions on α, τ, and $\bar{p}$ under which consumers' welfare in the location-then-prices equilibrium is larger than in the regulatory benchmark, *i.e.* when $W(x^*, p^*) \geq W(\frac{1}{2}, \bar{p})$. The change in welfare in the competitive equilibrium relative to the regulatory benchmark is given by

$$\Delta W(\bar{p}) = W(x^*, p^*) - W(\tfrac{1}{2}, \bar{p}) \tag{2.3}$$

where both x^* and p^* depend on parameters α, τ, and $\bar{p}$.

Welfare comparison To compare consumer welfare in both scenarios, I first consider the case where the non-prescription price in the regulatory benchmark is set to marginal costs (which are zero), $\bar{p} = 0$, so that prices must weakly *increase* under competition relative to the regulatory benchmark. Under this condition, I show that consumer welfare in the competitive equilibrium is weakly *smaller* than under the regulatory regime. Then I consider the case where the non-prescription price in the regulatory benchmark is initially larger than marginal costs. I argue that this opens up the possibility that consumer welfare is larger in the competitive equilibrium with free location choices.

The following proposition shows that consumer welfare in the competitive equilibrium is weakly smaller than in the regulatory benchmark if prices are initially regulated at marginal costs, which are zero:

Proposition 2. *Suppose that the regulated price of non-prescription drugs, $\bar{p}$, was initially zero. Then:*

1. *if α and τ are such that $x^* = \frac{1}{2}$, consumer welfare is the same in the competitive equilibrium as in the regulatory benchmark ($\Delta W(0) = 0$);*
2. *if α and τ are such that $x^* < \frac{1}{2}$, (a) the difference of consumer welfare in the competitive equilibrium relative to the regulatory benchmark increases strictly in the share of prescription consumers ($\frac{\mathrm{d}}{\mathrm{d}\alpha}\Delta W(0) > 0$), and it decreases strictly in the travel costs of non-prescription consumers ($\frac{\mathrm{d}}{\mathrm{d}\tau}\Delta W(0) < 0$); and (b) consumer welfare is smaller in the competitive equilibrium ($\Delta W(0) < 0$).*

The proof is delegated to appendix A.1. Now consider the case where the initial regulated price of the non-prescription drug, $\bar{p}$, was larger than zero. This leads to the possibility that consumer welfare in the competitive equilibrium is strictly larger than under the regulatory benchmark for a range of parameter values. To see this, note that price competition in the competitive location-then-prices equilibrium will drive down the price of the non-prescription drug to zero in such parameter constellations where firms choose to locate at the market centre. Under those parameter constellations, consumer welfare will therefore be larger in the competitive equilibrium than under the regulatory benchmark, as travel costs are the same, but non-prescription prices are lower. Now note that the difference between consumer welfare under the competitive equilibrium and consumer welfare under the regulatory benchmark at a non-prescription regulated price $\bar{p}$ is $\Delta W(\bar{p}) = \Delta W(0) + (1 - \alpha)\bar{p}$. By the above proposition, there are parameter constellations for which x^* is smaller than, but sufficiently close to $\frac{1}{2}$ so that $\Delta W(0)$ is large enough to obtain $\Delta W(\bar{p}) > 0$. This result is illustrated in figure 2.4 which shows the sets of α and τ, at which consumer welfare in the competitive location-then-prices equilibrium is equal to consumer welfare with regulated prices and free location choice, for different values of the regulated price $\bar{p}$. That figure confirms that welfare is strictly larger in the competitive equilibrium for a large range of parameter values if the regulated non-prescription price $\bar{p}$ is larger than zero. This insight is summarized in the following corollary:

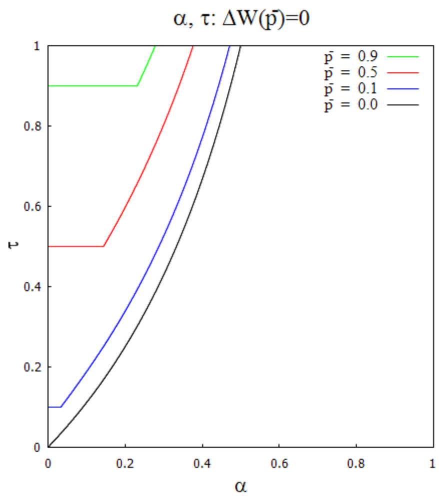


Figure 2.4.: Sets of α and τ, at which consumer welfare in the competitive location-then-prices equilibrium is equal to consumer welfare with regulated prices and free location choice, for different values of the regulated price $\bar{p}$. Consumer welfare in the competitive equilibrium is (weakly) larger than under the regulatory benchmark to the south east of the curves.

Corollary. *If the regulated price of prescription drugs was initially larger than zero, and the share of prescription consumers α is less than one, consumer welfare in the competitive equilibrium is strictly larger than in the regulatory benchmark, provided that the share of prescription consumers is sufficiently large, and the travel costs of non prescription consumers are sufficiently small.*

The theoretical analysis in this section is very stylized because it abstracts from entry and exit, assumes firm locations along a univariate line, and leaves the question open how the economy will transition from the regulatory to the competitive equilibrium. Still, this simple model serves to illustrate the fact that the introduction of price competition, apart from its impact on prices, can induce firms to enter in different locations so as to better serve the consumer. The model also illustrates that this re-location is not guaranteed to be beneficial for consumers: firms differentiate themselves in space because it increases their local market power, so they can raise their prices. Thus, in order to generate consumer welfare gains, the prices of non-prescription drugs must have been rather high initially, so that competitive prices decrease relative to the regulatory benchmark.

2.5. Empirical Model

This section develops a spatial dynamic entry model of pharmacy competition that will be fitted to data on retail pharmacy locations. It is designed to describe the process of entry and exit by which the spatial distribution of retail locations gradually responds to a regulatory change that introduced price competition for non-prescription drugs. I first outline a general framework of dynamic spatial entry that suffers from the curse of dimensionality. Then, I develop a new, computationally tractable approach to model the spatial entry dynamics in a large market. Finally, I describe how the model's structural parameters can be estimated. Importantly, I do not rely on symmetry or anonymity to ease the computational burden of the model.

The focus of this paper is on the spatial structure of the German pharmacy market, and how it reacts to regulatory changes, so one may ask why a dynamic model component is needed. The answer is twofold: first, the data on pharmacy locations spans a large time period of sixteen years, which makes it necessary to include a notion of time in the analysis. But more importantly, what is observed in these data are entry and exit decisions and such decisions are invariably dynamic in nature. Any decision maker who decides whether to set up a pharmacy, or whether to close it, will necessarily try to make some educated guess about the future, and most importantly, about the decisions of his current or future competitors. The interplay of these considerations is what drives the spatial industry structure over time. Therefore, a model needs to include both a spatial and a dynamic aspect. A more detailed argument is given in section 2.5.4.

2.5.1. Dynamic entry decisions

The economy consists of N firms (or potential entrants) indexed by j, which are located at a fixed location x_j. In each discrete time period t, each of these firms can either be active or inactive, indicated by $a_{jt} \in \{0, 1\}$. Let $\mathbf{a}_t = (a_{jt})_{j \in N}$ be called the state of the economy at time t, and denote the entire state space as $A = \{0, 1\}^N$. The usual notation is adopted where $\mathbf{a}_{-jt} = (a_{it})_{i \neq j}$ denotes the states of all firms but firm j. Sometimes, the tuple $(a_{jt}, \mathbf{a}_{-jt})$ is used as an alternative way of writing the aggregate state $\mathbf{a}_{jt}$. Each firm earns a period return $\pi_j(\mathbf{a}_t)$ when the aggregate state of the economy is $\mathbf{a}_t$, and it is assumed that $\pi_j(\mathbf{a}_t) = 0$ if $a_{jt} = 0$. The timing is as follows: at the beginning of every period t, every firm earns its period return $\pi_j(\mathbf{a}_t)$, and all firms learn the realization of a private information idiosyncratic random variable ξ_{jt} that follows a distribution function F with full support over $\mathbb{R}$. Upon learning this value, each potential entrant decides whether it should enter the market in period $t + 1$ and incur an entry cost of $\theta^x + \xi_{jt}$. Similarly, every incumbent decides whether it should stay in the market, or leave the market in which case it receives a sell-off value $\theta^e + \xi_{jt}$. The introduction of privately known transition costs ξ_{jt} is a common assumption in the literature on dynamic discrete games because it guarantees equilibrium existence, see e.g. Seim (2006), Doraszelski and Satterthwaite (2010), or Aguirregabiria and Magesan (2019). I do allow for re-entry because it is observed in the data. In the empirical application, it will be assumed that there are M independent large markets indexed by $m = 1, \ldots, M$, and T observed periods indexed by t, but the market subscripts will often be omitted for convenience.

The solution concept employed here is a Markov Perfect Equilibrium (MPE) where all players base their decisions solely on the current state of the economy (including their own activity status) and on a privately known disturbance to their transition costs. Let the strategy of firm j be denoted by

$$\sigma_j : (a_{jt}, \mathbf{a}_{-jt}, \xi_{jt}) \mapsto a_{jt+1} \in \{0, 1\} \tag{2.4}$$

and collect all strategies into $\sigma = (\sigma_j)_{j=1}^N$. Note that the behaviour of firm j is deterministic conditional on its latent variable ξ_{jt}, but stochastic from the point of view of its competitors.[11] From the perspective of another firm k, and of the econometrician, the probability that firm j chooses to be active in period $t+1$ is given by $q_j(\mathbf{a}_t) = \int \sigma_j(a_{jt}, \mathbf{a}_{-jt}, \xi)\mathrm{d}F(\xi)$. Taking the strategies of their competitors and the realization of the transition cost shock as given, each firm j decides whether to be active or not in the next period. Let the value function $V_j^\sigma(a_{jt}, \mathbf{a}_{-jt}, \xi_{jt})$ denote the expected discounted future profits of firm j when the state is $(a_{jt}, \mathbf{a}_{-jt})$ and the transition cost shock is ξ_{jt}. The value function is given by

$$\begin{aligned} V_j^\sigma(a_{jt}, \mathbf{a}_{-jt}, \xi_{jt}) = \pi_j(\mathbf{a}_t) + \max \Big\{ & a_{jt}\left(\theta^x + \xi_{jt}\right) + \beta \mathbb{E}[V_j^\sigma(0, \mathbf{a}_{-jt+1}, \xi_{jt+1}) | \mathbf{a}_t], \\ & -\left(1 - a_{jt}\right)\left(\theta^e + \xi_{jt}\right) + \beta \mathbb{E}[V_j^\sigma(1, \mathbf{a}_{-jt+1}, \xi_{jt+1}) | \mathbf{a}_t] \Big\} \end{aligned} \tag{2.5}$$

The expectation operator integrates over the distribution of all future states that is induced by σ, and over the future idiosyncratic shocks ξ. Since there is a one-to-one mapping from players' actions to states (unlike in many dynamic investment games, where idiosyncratic and market-specific shocks affect the success of an investment), this amounts to integrating over all conceivable actions of firm j's competitors, taking their strategies as given. The σ-superscript was added to clarify this dependence on the other firms' strategies.

To write this problem in a more compact form, I will follow Aguirregabiria and Vicentini (2016) and integrate out the idiosyncratic error term ξ_{jt}: first, let

$$\overline{V}_j^\sigma(a_{jt}, \mathbf{a}_{-jt}) = \int V_j^\sigma(a_{jt}, \mathbf{a}_{-jt}, x)\mathrm{d}F(x) \tag{2.6}$$

be the integrated value function of firm j. Sometimes I will write $\overline{V}_j^\sigma(\mathbf{a}_t)$ as a shorthand. Further, let the choice-specific integrated value function

$$v_j^\sigma(1, \mathbf{a}_t) = \pi_j(\mathbf{a}_t) - (1 - a_{jt})\theta^e + \beta \mathbb{E}^\sigma\left[\overline{V}_j^\sigma(1, \mathbf{a}_{-jt+1}) | \mathbf{a}_t\right] \tag{2.7}$$

denote the expected value of choosing to be active in the next period, and let

$$v_j^\sigma(0, \mathbf{a}_t) = \pi_j(\mathbf{a}_t) + a_{jt}\theta^x + \beta \mathbb{E}^\sigma\left[\overline{V}_j^\sigma(0, \mathbf{a}_{-jt+1}) | \mathbf{a}_t\right] \tag{2.8}$$

[11] Or, as Doraszelski and Satterthwaite (2010, p.216) put it: "Although a firm formally follows a pure strategy in making its entry/exit decision, the dependence of its entry/exit decision on its randomly drawn, privately known setup cost/scrap value implies that its rivals perceive the firm as though it were following a mixed strategy."

denote the expected value of choosing to be inactive in the next period, where the current activity status $a_{jt} \in \{0, 1\}$ governs whether the firm incurs any transition costs or period returns.[12] Then, the value function can be written more compactly as

$$V_j^\sigma(a_{jt}, \mathbf{a}_{-jt}, \xi_{jt}) = \max\left\{a_{jt}\xi_{jt} + v_j^\sigma(0, \mathbf{a}_t), -(1 - a_{jt})\xi_{jt} + v_j^\sigma(1, \mathbf{a}_t)\right\} \tag{2.9}$$

and the integrated value function is

$$\overline{V}_j^\sigma(a_{jt}, \mathbf{a}_{-jt}) = \int \max\left\{a_{jt}x + v_j^\sigma(0, \mathbf{a}_t), -(1 - a_{jt})x + v_j^\sigma(1, \mathbf{a}_t)\right\} \mathrm{d}F(x) \tag{2.10}$$

Given strategies σ and conditional on its own transition cost shock ξ_{jt}, firm j decides to be active in the next period if the second part of the above equation (2.9) is greater than the first, i.e. if

$$\begin{aligned} -(1 - a_{jt})\xi_{jt} + v_j^\sigma(1, \mathbf{a}_t) &\geq a_{jt}\xi_{jt} + v_j^\sigma(0, \mathbf{a}_t) \\ \Leftrightarrow \xi_{jt} &\leq v_j^\sigma(1, \mathbf{a}_t) - v_j^\sigma(0, \mathbf{a}_t) \end{aligned}$$

The above equation implicitly defines cutoff values which govern each firm's behaviour conditional on its observed private information shocks. The probability of being active next period can then conveniently be expressed as

$$\Pr(a_{jt+1} = 1 | \mathbf{a}_t) = F\left(v_j^\sigma(1, \mathbf{a}_t) - v_j^\sigma(0, \mathbf{a}_t)\right) =: q_j(\mathbf{a}_t) \tag{2.11}$$

where F is the distribution function of the latent errors ξ. Define $\mathbf{q}(\mathbf{a}_t) \equiv \{q_j(\mathbf{a}_t)\}_{j=1}^N$. These conditional choice probabilities (CCPs) are a best response probability to other firms following the strategy σ in state $\mathbf{a}_t$, and they "contain all the information about competitors' strategies that a firm needs to construct its best response" (Aguirregabiria and Vicentini, 2016, p.726). The reason is that the value functions depend on the strategy only through the CCPs that feed into the expectation operator $\mathbb{E}^\sigma$:

$$\underbrace{\begin{aligned} &\mathbb{E}^\sigma\left[\overline{V}_j(a_{jt+1}, \mathbf{a}_{-jt+1}) | a_{jt+1}, \mathbf{a}_t\right] = \\ &\sum_{\mathbf{a}_{-jt+1} \in \{0,1\}^{N-1}} \overline{V}_j(a_{jt+1}, \mathbf{a}_{-jt+1}) \prod_{i \neq j} q_i(\mathbf{a}_t)^{a_{it+1}} (1 - q_i(\mathbf{a}_t))^{1 - a_{it+1}} \end{aligned}}_{\equiv \mathbb{E}^{\mathbf{q}}\left[\overline{V}_j(a_{jt+1}, \mathbf{a}_{-jt+1}) | a_{jt+1}, \mathbf{a}_t\right]} \tag{2.12}$$

[12] Note that the arguments of the integrated value function, a_{jt} and $\mathbf{a}_{-jt}$, are different from the arguments of the choice-specific value functions, a_{jt+1} and $\mathbf{a}_t$.

Therefore, the problem of finding equilibrium strategies σ is equivalent to determining equilibrium CCPs $\mathbf{q}$ that satisfy the following equations at all states $\mathbf{a}_t$, and for all firms j (Aguirregabiria and Mira, 2007, p.11):

$$\begin{aligned} q_j(\mathbf{a}_t) &= F\left(v_j^{\mathbf{q}}(1,\mathbf{a}_t) - v_j^{\mathbf{q}}(0,\mathbf{a}_t)\right), \text{ where} \\ v_j^{\mathbf{q}}(1,\mathbf{a}_t) &= \pi_j(\mathbf{a}_t) - (1-a_{jt})\theta^e + \beta\mathbb{E}^{\mathbf{q}}\left[\overline{V}_j^{\mathbf{q}}(1,\mathbf{a}_{-jt+1})|a_{jt+1}=1,\mathbf{a}_t\right] \\ v_j^{\mathbf{q}}(0,\mathbf{a}_t) &= \pi_j(\mathbf{a}_t) + a_{jt}\theta^x + \beta\mathbb{E}^{\mathbf{q}}\left[\overline{V}_j^{\mathbf{q}}(0,\mathbf{a}_{-jt+1})|a_{jt+1}=0,\mathbf{a}_t\right] \\ \overline{V}_j^{\mathbf{q}}(a_{jt},\mathbf{a}_{-jt}) &= \int \max\left\{a_{jt}\xi + v_j^{\mathbf{q}}(0,\mathbf{a}_t), -(1-a_{jt})\xi + v_j^{\mathbf{q}}(1,\mathbf{a}_t)\right\} \mathrm{d}F(\xi) \end{aligned} \tag{2.13}$$

The function $q_j(\mathbf{a}_t)$ embodies firm j's best response, given all other firms' CCPs. By Brower's theorem, this system of best response functions is guaranteed to have a fixed point, as it defines a continuous mapping from the compact space $[0,1]^N$ onto itself. This motivates using an iterative procedure to solve for an equilibrium vector of CCPs, as will be outlined in more detail below. However, it must be noted that the above mentioned problem suffers from a "curse of dimensionality" that prevents its computation in all but the simplest problems. This is a common problem encountered in dynamic discrete games, as was outlined in section 2.3.3. According to Pakes and McGuire (2001), the computational burden of finding an equilibrium in such problems is principally determined by the size of the state space, the time it takes to compute the expectation, and the time to convergence. In the above problem, the computational burden to check whether a given vector $\mathbf{q}$ is in equilibrium grows exponentially in the number of firms.[13] Therefore, without further simplifying assumptions, the model is of little practical use when it comes to examining real-world economies with many firms.

Contrary to many empirical applications, I will not restrict attention to symmetric and anonymous equilibria. An equilibrium is symmetric if the equilibrium strategies are the same for all firms, i.e. $q_j(\mathbf{a}) = q_k(\mathbf{a})$ for every state $\mathbf{a} \in A$ and for all firms j,k, and it is anonymous if the equilibrium strategies are invariant to arbitrary permutations of the vector of its competitors' states $\mathbf{a}_{-j}$ (Doraszelski and Pakes, 2007). While being very convenient from a computational point of view, symmetry is not a good assumption in the context of spatial competition because the payoffs of a certain firm in any given state depend crucially on its location relative to its competitors, i.e. $\pi_j(\mathbf{a})$ is in general different from $\pi_k(\mathbf{a})$ and therefore, the equilibrium CCPs differ too. For the same reason, players' period return functions are in general not anonymous which leads to strategies that are not anonymous. I consider asymmetry and non-anonymity to be crucial characteristics of spatial dynamic interaction processes.

[13] The state space is of size 2^N so that the memory requirements to store CCPs, conditional and unconditional value functions for each firm are of order $4N2^N$. The computational burden of evaluating (2.13) for a given vector $\mathbf{q}$ is determined by the expectation operator in each of two choice-specific integrated value functions that integrates over the entire state space A_{-j} of size 2^{N-1}, and by evaluating and integrating the distribution function F. Thus, abstracting from the costs of memory look-up operations, the time to compute (2.13) for all firms and states is proportional to $N \times 2^N \times (2+2^N)$, so the computational burden is $\mathcal{O}(N2^{2N})$.

2.5.2. A spatial oblivious equilibrium model

To obtain a computable equilibrium model while maintaining asymmetry and non-anonymity, I develop a procedure that is close in spirit to the oblivious equilibrium concept of Weintraub et al. (2008), but adapted to fit the spatial, asymmetric structure of my data. My approach can be summarized as follows: firms principally assume that the spatial market structure remains constant, except in a close neighbourhood around their own location. Why firms are restricted in their strategic reasoning in this way is not specified; but it could be a rational decision to do so if planning ahead per se is costly. Indeed, given the sheer size of the unrestricted state space, it would be unreasonable to assume that any firm can accurately form and store expectations for all possible spatial market structures.

To be more concrete, I will restrict the strategy space by assuming that firms follow a time-varying heuristic strategy

$$\tilde{\sigma}_{jt} : (a_{jt}, \tilde{\mathbf{a}}_{-jt}, \xi_{jt}) \mapsto \{0, 1\}$$

where $\tilde{\mathbf{a}}_{jt} = (a_{jt}, \tilde{\mathbf{a}}_{-jt})$ is a member of what I call the "oblivious state space" $\tilde{A}_{jt}$ of firm j at time t that is composed of the observed actual state outside a close neighbourhood around firm j, and all possible market configurations of firms within a neighbourhood around firm j. In what follows, let $\hat{\mathbf{a}}_t$ denote the observed state at time t, and let $\hat{a}_{jt}$ denote the observed status of firm j at time t. Then, the oblivious state space of firm j at time t is defined as

$$\tilde{A}_{jt} \equiv \left\{ \Big(a_i : a_i \in \{0,1\} \text{ if } i \in nn_j^k, \text{ else } a_i = \hat{a}_{it} \Big)_{i \in N} \right\}$$

That is, in any given period t, firm j takes the state of its competitors beyond the range of k "strategic nearest neighbours" nn_j^k [14] as given, and heuristically assumes that only its k nearest neighbours will ever change their state. Note that the magnitude of the oblivious state space is only 2^k. Figure 2.5 illustrates the idea. The model is solved analogously to the full problem described above by finding equilibrium entry probabilities that satisfy (2.13), but using the restricted state space $\tilde{A}_{jt}$ for all firms and observation periods.

One problem that arises in computing the equilibrium entry probabilities is that the oblivious state spaces of any two firms j and i will often be different, because these two firms have different strategic neighbourhoods, i.e. $\tilde{A}_{jt} \neq \tilde{A}_{it}$. This implies that the expectation in (2.12) is not well defined: consider firm j and some state $\tilde{\mathbf{a}}_{jt} \in \tilde{A}_{jt}$ such that $\tilde{\mathbf{a}}_{jt} \notin \tilde{A}_{it}$. Since this state is not in firm i's oblivious state space, its strategy $\tilde{\sigma}_{it}$ is not defined at that point, so that firm j cannot form the conditional expectation in (2.12). However, an appropriate interpretation of the oblivious state space still allows for a coherent formation of this expectation in the following sense: For any firm i at time t, define the mapping $M_{it} : A \to \tilde{A}_{it}$ as

$$M_{it}(\mathbf{a}) = \Big(\tilde{a}_s : \tilde{a}_s = a_s \text{ if } s \in nn_i^k, \text{ else } \tilde{a}_s = \hat{a}_{st} \Big) \tag{2.14}$$

[14] I specify that $j \in nn_j^k$ so that each firm is its own nearest neighbour. But this is only . In principle, one could also define the neighbourhood based on the inter-firm distance, but using a fixed number of k neighbours has the advantage of allowing the usage of equally sized matrices in the computational implementation.

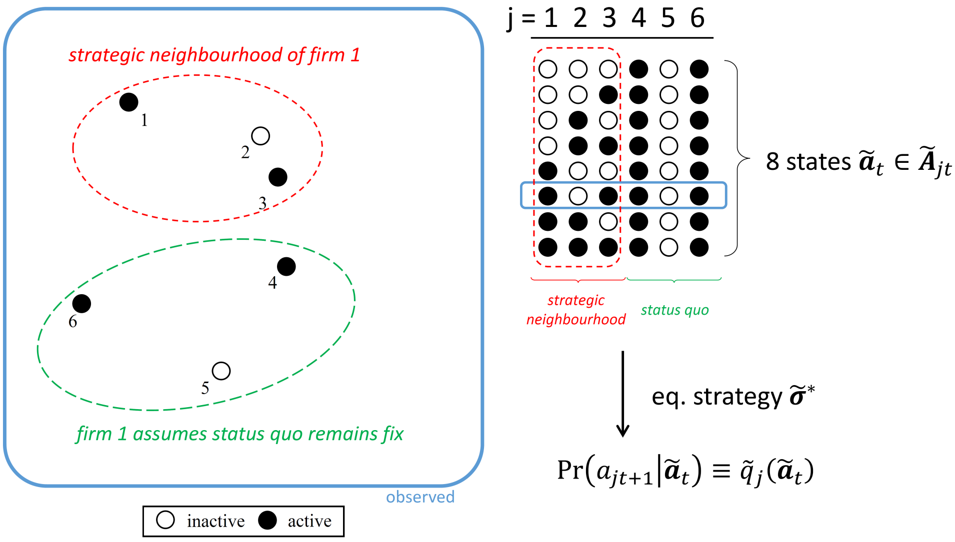


Figure 2.5.: The left side of the figure shows a sample market with six firms, four of which are active (indicated by a solid dot) and two of which are inactive (hollow dot). Firm one assumes that all firms except its two nearest neighbours will remain in their current status quo, and so the oblivious state space of firm two consists of only eight distinct states which are illustrated in the right part of the figure. In contrast, the unrestricted state space encompasses ($2^6 = 64$) distinct states.

For any state $\mathbf{a} \in A$, this mapping extracts the relevant local state as seen from the perspective of firm i, and uses the actual state $\hat{\mathbf{a}}$ for any firm outside the strategic neighbourhood nn_i^k. This idea is illustrated in figure 2.6. Using this mapping, I define the conjecture that firm j has about firm i's behaviour as follows:

$$_j\tilde{q}_{it}(\mathbf{a}) \equiv \begin{cases} \tilde{q}_{it}(\mathbf{a}) \text{ if } \mathbf{a} \in \tilde{A}_{it} \\ \tilde{q}_{it}\left(M_{it}(\mathbf{a})\right) \text{ else.} \end{cases} \tag{2.15}$$

where $\tilde{q}_{it}$ are the CCPs that are implied by firm i's strategy $\tilde{\sigma}_{it}$. Note that $M_{it}(\mathbf{a}_t)$ differs from $\mathbf{a}_t$ only for firms that are outside the strategic neighbourhood of firm i, and so the conjecture about firm i's strategy that other firms may have can be assumed to be reasonably close to the actual strategy of a firm i.

I assume that firms use these conjectures about their competitors' strategies to form approximate expectation over the future states. For a firm j that chooses an action a, this approximate expectation is:

$$\begin{aligned} &\mathcal{E}^{\tilde{\mathbf{q}}}\left[\overline{V}_j(a_{jt+1}, \mathbf{a}_{-jt+1})|a_{jt+1} = a, \mathbf{a}_t\right] = \\ &\sum_{\substack{\tilde{\mathbf{a}}_{t+1} \in \tilde{A}_{jt} \\ \tilde{a}_{jt+1} = a}} \overline{V}_j(a, \tilde{\mathbf{a}}_{-jt+1}) \prod_{i \in nn_j^k \setminus \{j\}} {_j\tilde{q}_{it}(\mathbf{a}_t)}^{\tilde{a}_{it+1}} (1 - {_j\tilde{q}_{it}(\mathbf{a}_t)})^{1-\tilde{a}_{it+1}} \end{aligned} \tag{2.16}$$

for all states $\mathbf{a}_t \in \tilde{A}_{jt}$. The above expression differs from that in equation (2.12) in two important ways: first, firm j assumes that all firms outside of its k-neighbourhood neither enter nor exit, and

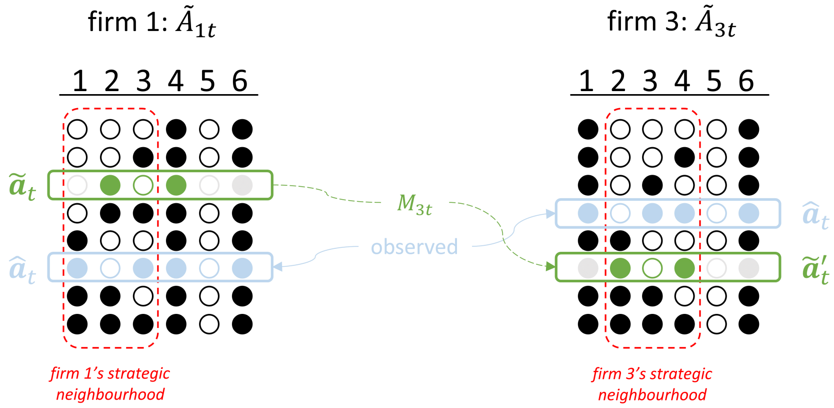


Figure 2.6.: This figure shows the oblivious state spaces of two firms that have only one common neighbour, so that their two oblivious state spaces differ. By construction, the observed state $\hat{a}_t$ (blue) is contained in both oblivious state spaces, but another state $\tilde{a}_t$ is not. Instead, firm 1 maps this state to another similar state $\tilde{a}'_t$ that is identical to $\tilde{a}_t$ for all firms that are contained in firm 3's strategic neighbourhood.

second, firm j forms a conjecture about the strategies of those firms inside its k-neighbourhood that is consistent with that assumption. Because firm j assumes that the future state will remain in its oblivious state space, which is of magnitude 2^k, the expectation $\mathcal{E}$ can be computed very quickly. Using the above expression (2.16), I define a spatial oblivious equilibrium to be a vector of CCPs $\tilde{\mathbf{q}}^*$ such that, for all firms j, at each time t, it holds for every state $\tilde{\mathbf{a}}_t \in \tilde{A}_{jt}$ that

$$\begin{aligned}
\tilde{q}_j^*(\tilde{\mathbf{a}}_t) &= F\left(v_j^{\tilde{\mathbf{q}}}(1, \tilde{\mathbf{a}}_t) - v_j^{\tilde{\mathbf{q}}}(0, \tilde{\mathbf{a}}_t)\right), \text{ where} \\
v_j^{\tilde{\mathbf{q}}}(1, \tilde{\mathbf{a}}_t) &= \pi_j(\tilde{\mathbf{a}}_t) - (1 - \tilde{a}_{jt})\theta^e + \beta \mathcal{E}^{\tilde{\mathbf{q}}}\left[\overline{V}_j^{\tilde{\mathbf{q}}}(1, \mathbf{a}_{-jt+1}) | a_{jt+1} = 1, \tilde{\mathbf{a}}_t\right] \\
v_j^{\tilde{\mathbf{q}}}(0, \tilde{\mathbf{a}}_t) &= \pi_j(\tilde{\mathbf{a}}_t) + \tilde{a}_{jt}\theta^x + \beta \mathcal{E}^{\tilde{\mathbf{q}}}\left[\overline{V}_j^{\tilde{\mathbf{q}}}(0, \mathbf{a}_{-jt+1}) | a_{jt+1} = 0, \tilde{\mathbf{a}}_t\right] \\
\overline{V}_j^{\tilde{\mathbf{q}}}(\tilde{a}_{jt}, \tilde{\mathbf{a}}_{-jt}) &= \int \max\left\{\tilde{a}_{jt}\xi + v_j^{\tilde{\mathbf{q}}}(0, \tilde{\mathbf{a}}_t), -(1 - \tilde{a}_{jt})\xi + v_j^{\tilde{\mathbf{q}}}(1, \tilde{\mathbf{a}}_t)\right\} \mathrm{d}F(\xi)
\end{aligned} \tag{2.17}$$

Again, Brower's fixed point theorem can be applied so that an equilibrium $\tilde{\mathbf{q}}^*$ is guaranteed to exist.

2.5.3. Estimation and identification

The structural parameters of the model are estimated using a nested fixed point approach to accommodate the asymmetric and non-anonymous nature of spatial strategic interactions. I assume that the private information shocks ξ_{jt} follow a standard normal distribution Φ so that the conditional choice probabilities in equation (2.13) can be expressed in convenient form as:

$$\Pr(a_{jt+1} = 1 | \tilde{\mathbf{a}}_t) = \tilde{q}_j(\tilde{\mathbf{a}}_t) = \Phi\left(v_j^{\tilde{\mathbf{q}}}(1, \tilde{\mathbf{a}}_t) - v_j^{\tilde{\mathbf{q}}}(0, \tilde{\mathbf{a}}_t)\right) \tag{2.18}$$

The integrated value function can then be computed as

$$\overline{V}_j^{\tilde{\mathbf{q}}}(a_{jt}, \tilde{\mathbf{a}}_{-jt}) = \phi\left(v_j^{\tilde{\mathbf{q}}}(1, \tilde{\mathbf{a}}_t) - v_j^{\tilde{\mathbf{q}}}(0, \tilde{\mathbf{a}}_t)\right) + (1 - \tilde{q}_j(\tilde{\mathbf{a}}_t)) \cdot v_j^{\tilde{\mathbf{q}}}(0, \tilde{\mathbf{a}}_t) + \tilde{q}_j(\tilde{\mathbf{a}}_t) \cdot v_j^{\tilde{\mathbf{q}}}(1, \tilde{\mathbf{a}}_t) \quad (2.19)$$

where ϕ is the probability density function of the standard normal distribution.[15] For the empirical application, the reduced form profit equation is parametrized as follows:

$$\pi_j(\mathbf{a}_t) = \begin{cases} \underbrace{\left(\alpha \mathbf{X}_j + \frac{\beta Y_j}{1 + N_{jt}^d}\right)}_{\text{local demand}} \underbrace{\left(1 - \delta N_{jt}^d)\right)}_{\text{local competition}}, \text{if } a_{jt} = 1 \\ 0 \ \text{else.} \end{cases} \quad (2.20)$$

where Y_j is the local population around store j that is divided by the number of active firms in firm j's d-neighbourhood, N_{jt}^d (and including firm j). X_j is a vector of profit shifters of store j that is common knowledge to all players. The first composite term captures local demand at a given location in market state $\mathbf{a}_t$, and the second term captures the market power effect of local competition. The parameter δ measures the relative reduction of a pharmacy's profitability due to the presence of one additional active competitor within a distance d. In order to test the main hypothesis of the paper, namely whether spatial competition has increased after 2004, δ was estimated separately before and after 2004 (δ^{pre} and δ^{post}). The spatial co-variates include measures of the local residential population within a radius x around the firm's location and measures such as the number of nearby supermarkets or doctors. A summary of these variables can be found in table 2.4. As part of a robustness check, the population variable was scaled with aggregate municipality-level population growth rates, and further time-varying and municipality-level covariates were included. However, for the sake of simplicity this time dependence is not explicitly modelled: in the model, agents act as if all variables are time-constant while they actually change over time. This naturally introduces some error, but I believe that this error is small because the time-varying variables change rather slowly so that the assumption of time-constant co-variates may be a good approximation to actual decision processes.

Using the parametric profit function in (2.20), the conditional expectation in (2.16), the conditional choice probabilities (2.18) and the integrated value function (2.19), the system of equations in (2.13) for some market m and time period $t < T$ can be computed as follows:

Procedure 1 (Fixed point algorithm)**.**

0. *initialization:*

    a) *set* $k \leftarrow 0$

    b) *compute* $\pi_j(\mathbf{a})$ *for all* $\mathbf{a} \in \tilde{A}_{jt}$, *and for all firms* $j \in 1, \ldots, N$

[15]This follows from the last equation in (2.17) and from the fact that the expectation of a truncated normally distributed random variable x is given by $\mathbb{E}[x|x \leq y] = \frac{-\phi(y)}{\Phi(y)}$, where Φ is the normal c.d.f. and ϕ is the normal p.d.f..

c) for all firms $j = 1, \dots, N$, initialize the vectors $\mathbf{q}_j$, $\mathbf{v}_j^0$, $\mathbf{v}_j^1$ and $\bar{\mathbf{V}}_j$ of length 2^k to zero, representing the corresponding functions $q_j(\cdot)$, $v_j^{\mathbf{q}}(0, \cdot)$, $v_j^{\mathbf{q}}(1, \cdot)$ and $\overline{V}_j^{\mathbf{q}}(\cdot)$ evaluated at each state $\mathbf{a}_{jt} \in \tilde{A}_{jt}$. Collect all firm-specific vectors into $\mathbf{q}, \mathbf{v}^0, \mathbf{v}^1$ and $\overline{\mathbf{V}}^{\mathbf{q}}$.

1. *at step k:*

   a) set $\Delta \leftarrow 0$, and for all firms j do:

   - $\mathbf{v}_j^{a_{jt+1}} \leftarrow \left(v_j^{\mathbf{q}}(a_{jt+1}, \tilde{\mathbf{a}}) \right)_{\tilde{\mathbf{a}} \in \tilde{A}_{jt}}$ *for* $a_{jt+1} \in \{0, 1\}$ *using* (2.8), (2.7), (2.20) *and* (2.16).
   - $\bar{\mathbf{V}}_j \leftarrow \left(\overline{V}_j^{\mathbf{q}}(\tilde{\mathbf{a}}) \right)_{\tilde{\mathbf{a}} \in \tilde{A}_{jt}}$ *using* (2.19)
   - $\Delta_j \leftarrow \left\| \mathbf{q}_j - (\tilde{q}_j(\tilde{\mathbf{a}}))_{\tilde{\mathbf{a}} \in \tilde{A}_{jt}} \right\|_\infty$ *and* $\mathbf{q}_j \leftarrow \frac{1}{2} \left(\mathbf{q}_j + (\tilde{q}_j(\tilde{\mathbf{a}}))_{\tilde{\mathbf{a}} \in \tilde{A}_{jt}} \right)$ *using* (2.18)
   - $\Delta \leftarrow \max\{\Delta, \Delta_j\}$

   b) if $\Delta \geq \epsilon$ go back to step 1 with $k \leftarrow k + 1$, else STOP.

Technically, this is a Gauss-Seidel algorithm because the updates in step k are always computed using the most recent available conditional choice probabilities from either of steps k or $k-1$. Doraszelski and Pakes (2007) write that this leads to faster convergence in many cases, and I found that this applied also to my setting. The dampening that is introduced in updating the conditional choice probabilities was found to significantly improve convergence speeds, as is also suggested by Doraszelski and Pakes (2007).[16] In my examples, with more than two hundred firms per market, the algorithm usually converged within one hundred iterations to a tolerance of $\epsilon = \sqrt{3 \cdot 10^{-10}}$, irrespective of the starting values. In each market, I computed the equilibrium only up to time period $T - 1$ because the equilibrium in period T makes predictions about transition patterns in period $T + 1$ which are not observed.[17]

The model's parameters of interest are α, δ, θ^e and θ^x. As outlined above, the spatial interaction parameter δ is estimated separately before and after the reform in 2003. I also estimated θ^e separately, and included a dummy variable for $t \geq 2004$ in the profit equation. I assume that the decision making units do not anticipate this change in parameters, i.e. the parameter change in 2004 comes as a surprise to them, and the market then transitions into a new steady state that is consistent with new parameter values. In Monte Carlo simulations, I found that I could not identify all parameters separately, so I

[16] I implemented this model in Python (version 3.7) using the Numpy (version 1.15.4), Scipy (version 1.1.0) and Numba (version 0.42.0) libraries. The actual estimation was conducted on a fast compute node of the bwHPC cluster, using the multiprocessing library.

[17] The oblivious state space is of size 2^k so that the memory requirements to store CCPs, conditional and unconditional value functions for each firm in every period are of order $4NT2^k$. The computational burden of evaluating (2.13) for a given vector $\tilde{\mathbf{q}}$ is determined by the expectation operator $\mathcal{E}$ in each of two choice-specific integrated value functions that integrates over the oblivious state space $\tilde{A}_{-jt}$ of size 2^{k-1} in each time period t, and by evaluating and integrating the distribution function F. Thus, abstracting from the costs of memory look-up operations, the time to compute (2.13) for all firms and states is approximately $\mathcal{O}(NT2^k(2 + 2 \cdot 2^{k-1})) = \mathcal{O}(NT2^k(2 + \cdot 2^k)$. The actual time that it takes to compute an equilibrium will also crucially depend on the number of iterations that it takes to converge to $\tilde{\mathbf{q}}^*$.

chose to normalize the market exit value to unity, i.e. $\theta^x = 1$.[18] Thus, the parameters of interest can be summarized in a vector $\theta = (\alpha, \delta_{pre}, \delta_{post}, \theta^e_{pre}, \theta^e_{post})$ with $\theta^x = 1$ and α including a pre/post dummy. In order to estimate θ, for each market m, and conditional on parameters θ, the equilibrium $\tilde{\mathbf{q}}^*_{mt}$ is computed for all years $t = 1, \ldots, T-1$. Thus, the log likelihood of the observed market outcomes $\{\mathbf{a}_{mt}\}_{t=2}^T$, conditional on the market states $\{\mathbf{a}_{mt}\}_{t=1}^{T-1}$ is computed as follows:

$$ll_m(\theta) = \sum_{t=2}^{T} \sum_{j=1}^{N_m} a_{mjt} \log \tilde{q}^*_{mjt-1} + (1 - a_{mjt}) \log(1 - \tilde{q}^*_{mjt-1})$$

These market likelihoods are then aggregated to form the total log-likelihood

$$ll(\theta) = \sum_{m=1}^{M} ll_m(\theta)$$

The likelihood is optimized with respect to parameters θ using the BFGS algorithm.[19] Standard errors are computed using the estimated Hessian matrix that is returned by the BFGS algorithm.

Economic agents in the model that is described above are heterogeneous with respect to the realization of their private information shocks, and with respect to their spatial configuration relative to each other. As in Seim (2006), the multiplicity of observed outcomes that may arise in a pure strategy equilibrium due to the presence of a spatial interaction effect is circumvented by modelling entry and exit probabilities. Thus, players form their expectations ex ante and they may eventually end up in a state which is not an equilibrium outcome ex post. Yet, Berry and Reiss (2007, p.1878) note that the entry probabilities in such a model may not be unique: as the variance of the unobserved error decreases so that the model approaches one of perfect information, multiple equilibrium entry rates that mirror the multiple equilibria in pure strategies can arise. Thus, for the model to predict unique equilibrium entry and exit rates it is necessary that the private information shocks, relative to the observed component of the player's payoffs, are sufficiently important. I believe that this is the case in my empirical application, because the Gauss-Seidel algorithm that is described above always converged to the same equilibrium probabilities regardless of the initial conditions.

Given that the conditional entry and exit rates are uniquely determined, the model is primarily identified by matching the predicted transition rates between observed market states to the observed transition rates. Period returns and entry costs are only identified relative to the variance of the errors, which is set to unity. The magnitude of the entry costs is identified by matching the degree of turnover in the data: larger entry costs imply higher persistence, i.e. fewer firms enter and leave the market. Larger exit values have the opposite effect, and this is why I constrain the parameter θ^x to unity so as to circumvent near collinearity issues. The magnitude of the period returns are identified – relative to the

[18]This problem was alleviated when payoffs were assumed to depend on market-level variates, and one had many markets. But although my model does include market-level covariates, it is by no means certain that these are the correct ones, so I decided to normalize the scrap values nonetheless.

[19]as implemented in the Scipy (version 1.1.0) library

logit error scale – by the average number of active firms in the market: Larger period returns alone imply that more firms will be active, on average. The parameters that govern local demand vary across locations, and are thus identified by the spatial variation in entry and exit rates. Finally, the spatial interaction parameter δ is identified through the effect that every additional active firm has on entry and exit rates in nearby locations.

2.5.4. Remarks

Before proceeding to describe the data, and the estimation results of this paper, I discuss certain aspects of the method and the model by means of illustrative examples. In particular, I will outline why a structural model that incorporates dynamic strategic interactions is needed, rather than a reduced form model.

Why a dynamic model is needed

Reduced form estimates lead to inconsistent results of the interaction parameter because the competitors' actions are endogenous with respect to own actions, which renders reduced form estimators inconsistent. A structural model such as the one put forward in this paper can alleviate this problem by imposing appropriate behavioural assumptions. To explore this issue further, I set up a dynamic entry and exit model with two firms. The notation and timing structure is the same as in the full model above. Their period returns are given by

$$\pi_i(a_{it}, a_{-it}) = \begin{cases} \frac{1}{2} - \delta a_{-it}, \text{ if } a_{it} = 1 \text{ with } \delta \in [0, 1] \\ 0, \text{ else} \end{cases}$$

Entry costs are given by $\theta^e = 4$ and scrap values are given by $\theta^x = 1$, and both are subject to a standard normally distributed shock ξ_{it} as in the full model above. The state space of this small illustrative model encompasses only four distinct states so that it can easily be solved exactly. A Markov Perfect Equilibrium (MPE) of this model is a set of conditional choice probabilities (CCPs), denoted by $q_i^*(a_{it}, a_{-it})$, for $i \in \{1, 2\}$ and $a_{it}, a_{-it} \in \{0, 1\}^2$ such that the system of equations (2.13) holds. Suppose only the interaction parameter δ is to be estimated from a sequence of observed market states $\{\mathbf{a}_t\}_{t=1}^T$. This can be achieved by choosing δ such that the model-implied CCPs match the observed transition rates as closely as possible. I computed the equilibrium CCPs for a range of parameter values δ, ranging from zero to one. These CCPs are shown in figure 2.7, in terms of entry and exit probabilities for different states of the competitor (solid and dotted blue lines). Alongside the equilibrium CCPs I also plotted CCPs that are derived non-strategically, i.e. with players that assume that their competitors do not change their state (dashed red lines).

The interaction parameter is identified if there is a unique mapping from entry and exit probabilities to parameter values, and the presence of multiple equilibria may prevent this. Figure 2.7 shows that the model generates an asymmetric equilibrium (dotted blue lines) for very large values of δ in addition to a symmetric equilibrium (solid blue lines), in line with what Berry and Reiss (2007, p.1878) write. While a multiplicity of equilibria does not necessarily lead to non-identification, in this case it does lead to a non-unique mapping from CCPs to parameter values which can prevent identification. To alleviate this problem, I will assume that the interaction parameter is sufficiently small compared to the period returns so as to admit a unique equilibrium. Figure 2.7 also shows that an identification of the interaction parameter comes mainly from matching entry and exit rates in the presence of an active competitor (top right and top left panel) because in these cases, the CCPs exhibit monotonicity over a large range of parameters δ.

If agents ignore the strategic reactions of their competitors, their optimal entry and exit decisions will differ from the ones obtained by a model of strategic decision making. This can be seen in the two panels on the right hand side of figure 2.7. The figure shows that a potential entrant is less likely to enter if it disregards the possibility that the incumbent will leave the market, and that a duopolist is more likely to leave the market if it disregards the reactions of its competitor. Thus, strategic play leads to more entry and delayed exit. This implies that any estimation procedure that attempts to match entry and exit rates without modelling the strategic interaction will underestimate the interaction parameter δ if agents act strategically, and it is the reason why a structural model of dynamic forward-looking decision making is needed. Of course, it could be that agents do not behave in this manner and merely take their competitors' actions as given. This could indeed be a rational strategy to follow if entry and exit rates are rather small, and the costs of strategic planning are large. In markets with more than two firms, intermediate cases of strategic decision making are likely to occur, where decision makers do not pay attention to very distant competitors. This is precisely the idea behind the spatial oblivious equilibrium concept; and my empirical approach will allow me to determine the degree to which strategic dynamic decision making is important.

Why two-step estimators cannot be used

The spatial aspect of the entry game at hand introduces non-anonymity (Doraszelski and Pakes, 2007) in firms' best response functions that render conventional two-step estimators of dynamic games inapplicable. These estimators typically rely on some consistent first-stage estimates of firm's CCPs. However, such estimates are hard to obtain in the current setting because CCPs depend crucially on the precise spatial configuration of firms' competitors: Even if the period return function depends only on the number of active neighbours, but not on their identity or spatial configuration, the same does not hold for firms' CCPs. Therefore, the first stage estimates would have to be estimated conditionally on the spatial configuration. But this configuration is drawn from a high dimensional space, which precludes the usage of simple non-parametric estimators.

This point shall be highlighted in an illustrative example. Consider a linear market of length $2d$ with four firm locations as illustrated in figure 2.8. Suppose that period returns of an active firm are given by

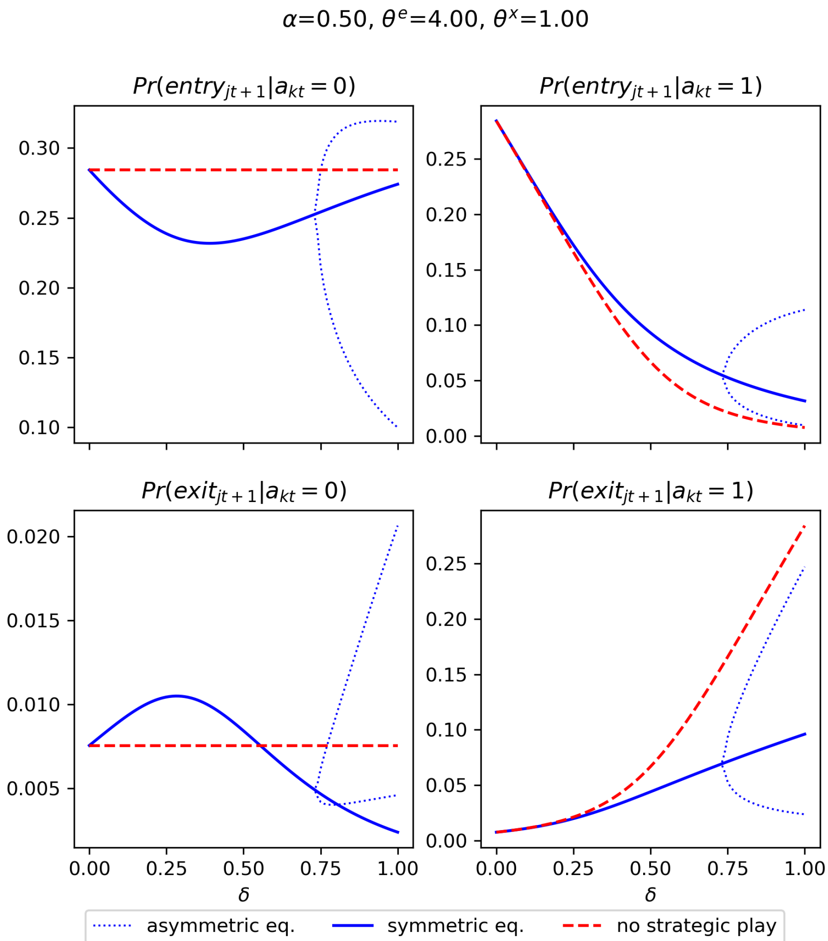


Figure 2.7.: Conditional choice probabilities in a two-firm model of dynamic entry for different interaction parameters. The figure shows the CCPs of firm j conditional on firm k's current status.

(a) $\mathbf{a}_t = (1,1,1,0), q_0^*(\mathbf{a}_t) = 0.46$ **(b)** $\mathbf{a}_t = (1,0,1,1), q_0^*(\mathbf{a}_t) = 0.59$

Figure 2.8.: Non-anonymity. Solid circle: active firm. Firm zero is less likely to remain active if its two neighbours to the left are active, than if one firm on either side is active. The equilibrium CCPs were computed for $\alpha = 1$, $\delta = \frac{1}{2}$, $\theta^e = 4$ and $\theta^x = 1$.

$\pi_{jt} = 1 - \frac{1}{2}N_{jt}^d$ where N_j^d is the number of firm j's active neighbours within a radius d. The remaining structure of the game is as described above, with $\theta^e = 4$ and $\theta^x = 1$, both being subject to privately known random perturbations. I am interested in the equilibrium behaviour of the central firm 0. Note that this firm's profits are negatively affected if any of its three neighbours is active. Conversely, firms 1 and 2 have only two potential direct competitors, and firm 3 has only one such neighbour. This means that firm 3 is a stronger competitor for firm 0 than firms 1 and 2 are. Keeping the number of firm 0's active neighbours fixed at two, firm 0 is more likely to remain active if its weaker competitor 2 is the potential entrant (figure 2.8b), than if its stronger competitor 3 is a potential entrant (figure 2.8a). The same mechanisms are at play in spatial entry games more generally, and because of this inherent non-anonymity, it is generally not possible to consistently estimate firms' CCPs conditional on some low-dimensional market characteristic such as the number of active competitors, as would be required for a two-step estimator.

Properties of the oblivious approximation to the MPE

The natural question is how the spatial oblivious equilibrium in (2.17) relates to the full MPE defined in (2.13) above. Of course, if the number of strategic nearest neighbours is equal to the total number of firm locations in the market ($k = N$), then the two equilibria are the same but the whole idea of the spatial oblivious equilibrium is that the number of strategic neighbours (k) is smaller than the number of firms (N). Thus, it is interesting to know whether the two quantities become closer as k increases and eventually approaches N. A theoretical answer to this question along the lines of Weintraub et al. (2008)[20] is work to be done in the future. To approach this question from a computational point of view, I created a sample of $N = 10$ firm locations in a square market of length one thousand. The profit of an active firm is given by $\pi_{jt} = 1 - \frac{1}{3}N_j^d$ where N_j^d is the number of firm j's active neighbours within a radius $d = 400$. This radius is chosen such that, on average, five neighbour locations fall within it.[21] Entry costs are given by $\theta^e = 3$, and scrap values are given by $\theta^x = 1$, both being subject to randomly drawn, normally distributed disturbances. Using the model-implied CCPs, the industry's long run distribution across states was computed, and the state with the largest long-run probability

[20] A direct application of their results is not possible because of the asymmetric structure of the problem at hand.

[21] Since firm locations were generated randomly, the expected number of firms within radius d is $\lambda\pi d^2$, where $\lambda = N/A$ is the density of locations on the area A.

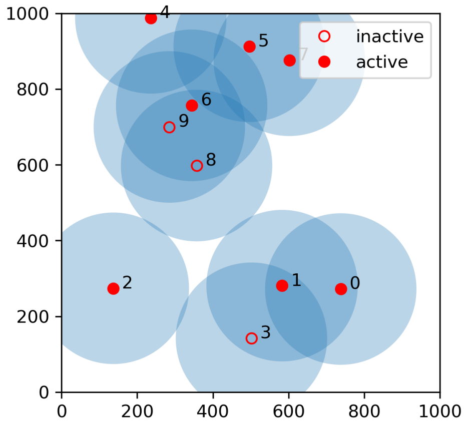


Figure 2.9.: The example market. The blue circles are drawn to represent half the interaction radius. Hence, if the circles around to adjacent locations overlap, these two locations interact directly through the profit equation.

of occurrence was selected as the industry's initial state.[22] This market configuration is denoted as $\hat{\mathbf{a}}$ and is shown in figure 2.9. At this initial industry state, I computed the equilibrium CCPs $\tilde{\mathbf{q}}^{(k)}(\hat{\mathbf{a}})$ that satisfy (2.17) for different choices of k, ranging from $k = 1$ (no strategic interaction) to $k = 10$ (full strategic interaction; nine strategic nearest neighbours and self). For each value of k, I computed two distance measures of the firms' CCPs between $\tilde{\mathbf{q}}^{(k)}(\hat{\mathbf{a}})$ and $\tilde{\mathbf{q}}^{(10)}(\hat{\mathbf{a}})$: the maximum absolute difference ($|\cdot|_\infty$), and the root of the mean squared difference ($|\cdot|_2$).

The results, presented in table 2.2, indicate that both distance measures become very small as k increases, while at the same time the computational time increases rapidly. For this particular market, the oblivious MPE for $k = 6$ offers a good approximation of the "true" MPE with $k = N$ at an acceptable computational cost. Of course, this depends crucially on the particular choice of the interaction range d that determines how many of a firms' k nearest neighbours have a direct effect on its period returns. In the extreme, if the interaction is such that every firm is completely isolated, then any choice $k \geq 1$ would obviously lead to the same result, and one did not need to worry about strategic interactions at all. On the other hand, if all firms interacted with every one of their competitors, then any choice of $k < 10$ would be unlikely to produce a correct result. In this case, one could instead directly apply the oblivious equilibrium concept of Weintraub et al. (2008) without any modifications. If the interaction parameter is so large that multiple equilibria occur, it could even be that an approximation with $k > 1$ does worse than the myopic MPE with $k = 1$. Furthermore, the quality of the approximation depends crucially on firm turnover in the market: if turnover is very low, then the status quo is a good prediction of the future state, and so a smaller strategic neighbourhood might suffice. In future work, it would thus be desirable to investigate the relationship between the spatial structure of firms' potential locations and the properties of the approximative equilibrium in greater detail, and in more general terms.

[22]Given the CCPs, one can build the transition matrix Q that determines the probability of transitioning from each market state $\mathbf{a}$ to any other state $\mathbf{a}'$, with $Q(\mathbf{a}', \mathbf{a}) = \Pr(\mathbf{a}'|\mathbf{a})$. Then, the steady state distribution across market states is a vector $\mathbf{p}^*$ that satisfies $Q\mathbf{p}^* = \mathbf{p}^*$ and $\sum_s p_s^* = 1$.

firm j	$\tilde{q}_j^{(1)}$	$\tilde{q}_j^{(2)}$	$\tilde{q}_j^{(3)}$	$\tilde{q}_j^{(4)}$	$\tilde{q}_j^{(5)}$	$\tilde{q}_j^{(6)}$	$\tilde{q}_j^{(7)}$	$\tilde{q}_j^{(8)}$	$\tilde{q}_j^{(9)}$	$\tilde{q}_j^{(10)}$
0	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98
1	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98
2	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99
3	0.07	0.07	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08
4	0.69	0.82	0.86	0.83	0.88	0.85	0.85	0.85	0.85	0.85
5	0.69	0.80	0.87	0.93	0.89	0.85	0.85	0.85	0.85	0.85
6	0.69	0.68	0.68	0.70	0.78	0.85	0.85	0.85	0.85	0.85
7	0.69	0.80	0.87	0.82	0.81	0.85	0.85	0.85	0.85	0.85
8	0.00	0.00	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
9	0.02	0.03	0.03	0.06	0.05	0.09	0.09	0.09	0.09	0.09
$\lvert\tilde{\mathbf{q}}^{(k)} - \tilde{\mathbf{q}}^{(10)}\rvert_\infty$	0.161	0.166	0.168	0.147	0.071	0.000	0.000	0.000	0.000	
$\lvert\tilde{\mathbf{q}}^{(k)} - \tilde{\mathbf{q}}^{(10)}\rvert_2$	0.322	0.192	0.181	0.175	0.103	0.001	0.000	0.000	0.000	
runtime [s]	0.03	0.04	0.04	0.04	0.04	0.07	0.30	0.82	9.24	54.82

Table 2.2.: Convergence of the spatial oblivious equilibrium to the MPE. This table shows the equilibrium CCPs that were derived using different levels of strategic sophistication in a generic market of ten firms, at a particular industry state. See the main text for a description of how the market was constructed, and see figure 2.9 for a graphical depiction.

2.6. Data

2.6.1. Data sources

In order to address the research questions of this paper, extensive data on the German pharmacy market were collected. The addresses of all active pharmacy locations[23] were taken from sixteen editions of the *Bundesapothekenregister* (Deutscher Apotheker Verlag, 2016) and were geocoded.[24] An overview of the resulting panel data set is given in table 2.3. The number of new establishments and firm exits displays substantial variability over time, which may be due to the fact the the data source is issued quarterly, but it was not always possible to obtain the same issue in each year. Lacking a unique firm identifier, I attempted to identify firms whose location changed from year to year using fuzzy string matching techniques. But because the number of firm re-locations is rather small compared to the number of entries and exits, I decided to abstract from firm re-locations in the empirical model, and these figures are not reported here. Hence, a re-location is treated as simultaneous entry and exit in the same year in two different locations. In total, there are 26,964 distinct locations.

For the empirical analysis, attention is restricted to eighty German cities (excluding Berlin, Hamburg, Munich and Cologne). This set of urban markets was chosen in order to create a homogeneous sample in which the nature of spatial interaction is fairly similar. Rural areas, on the other hand, are likely to exhibit very different spatial interaction mechanisms due to different travel and commuting patterns. Also, the largest German cities are larger by an order of magnitude than the majority of other cities

[23]For this empirical analysis, a pharmacy is one outlet but may belong to a group of pharmacies

[24]This was done using an academic licence for Bing Spatial Data Services. Care was taken to obtain accurate locations, and the results were double-checked manually where the geocoding API indicated that its result was imprecise.

	Germany			80 city sample		
year	*active*	*entries*	*exits*	*active*	*entries*	*exits*
2001	16,286			5,449		
2002	16,272	175	189	5,438	57	68
2003	16,300	120	92	5,442	48	44
2004	16,328	110	82	5,437	25	30
2005	16,184	313	457	5,307	91	221
2006	16,226	1,042	1,000	5,234	334	407
2007	16,307	420	339	5,233	129	130
2008	16,372	409	344	5,229	161	165
2009	16,342	101	131	5,209	33	53
2010	16,273	460	529	5,178	161	192
2011	16,168	266	371	5,126	84	136
2012	16,051	256	373	5,067	96	155
2013	15,845	266	472	4,953	93	207
2014	15,807	54	92	4,933	11	31
2015	15,645	181	343	4,856	62	139
2016	15,519	192	318	4,771	56	141
Total	26,964	4,365	5,132	6,741	1,441	2,119

Table 2.3.: Number of active firms, entries, and exits across years for the entire dataset, and for the sample of eighty large German cities that is used in the subsequent analysis.

and so they were excluded from the analysis. A list of the eighty sample cities is shown in table A.1 in the appendix.

In the base line analysis, I only use locations where an active pharmacy has been observed at some point as potential entry locations. In a robustness check, I extend the set of potential entry locations in two different ways. On the one hand, I use the locations of bakery shops.[25] Since bakery shops and pharmacies have approximately the same size, these locations therefore represent an appropriate set of potential entry locations. Furthermore, they also respect local entry restrictions that may result from zoning laws. As a second alternative, I generated random entry locations within the administrative boundaries of each city. These random entry locations, of course, do not respect zoning laws, and they also probably do not correspond to actual feasible entry locations. Therefore, these results should be interpreted cautiously.

To model pharmacies' variable reduced form profits (see equation (2.20)), data on local demand and supply conditions was obtained form various sources. First, data from the German census in 2011 (Zensus, 2011) was used which shows the spatial residential population distribution on a fine grid with a spacing of just one hundred metres. This allows me to compute the local residential population, within a radius of 500 metres around each potential entry location. This variable proxies local residential demand at each location and corresponds to the variable Y_j in equation (2.20). Moreover, the share of people aged 65 and older was computed for every potential entry location.

[25]These locations were extracted from OpenStreetMap (OpenStreetMap contributors, 2017).

Second, I obtained the locations and outlines, respectively, of doctors, supermarkets, train stations, pedestrian zones, and main roads from an OpenStreetMap data base (OpenStreetMap contributors, 2017).[26] Using GIS software, the distances to the nearest doctor, supermarket, main road, etc. were computed for each pharmacy location in the sample.[27] The aforementioned variables are specific to each distinct location, but do not vary over time because the census data are available only for 2011, and because I used only a single year of OpenStreetmap data. To capture temporal changes in the profitability of stores, I allow the intercept, entry costs, and the interaction term, to differ before and after the reform period. Table 2.4 shows summary statistics for these co-variates in the selected sample of cities, and for the three sets of point locations – pharmacies, bake shops, and random dummy locations. The table shows that pharmacy locations, and the bake shop locations are fairly similar in terms of their observable statistics, although a formal t-test rejects the null of equal means for all variables (see table A.2). In comparison, the random locations differ substantially in their observable characteristics.

[26]The data were downloaded from download.geofabrik.de, and processed using the command line tool Osmosis

[27]This was done in the QGIS environment.

		pharmacies ($N = 6,741$)				bake shops ($N = 7,382$)				dummy points ($N = 10,033$)			
Variable	*Explanation*	*mean*	*SD*	*min*	*max*	*mean*	*SD*	*min*	*max*	*mean*	*SD*	*min*	*max*
Local residential population as of 2011 census (Y_i, between variation)													
local population	within 500m radius (in 10k)	0.495	0.269	0.000	1.678	0.467	0.284	0.000	1.809	0.142	0.194	0.000	1.530
elderly share	share over 65 within $500m$ radius	0.198	0.062	0.000	0.634	0.194	0.066	0.000	1.000	0.171	0.124	0.000	1.000
Local spatial covariates from Open StreetMap (time invariant, between variation)													
pedestrian zone	$\leq 50m$ from pedestrian zone	0.216	0.411	0.000	1.000	0.212	0.408	0.000	1.000	0.008	0.088	0.000	1.000
main road	$\leq 50m$ from main road	0.331	0.471	0.000	1.000	0.263	0.440	0.000	1.000	0.064	0.245	0.000	1.000
doctor nearby	physician within $100m$	0.217	0.412	0.000	1.000	0.141	0.348	0.000	1.000	0.006	0.076	0.000	1.000
supermarket	supermarket within $100m$	0.231	0.421	0.000	1.000	0.298	0.458	0.000	1.000	0.007	0.086	0.000	1.000
trainstation	trainstation within $250m$	0.045	0.208	0.000	1.000	0.061	0.239	0.000	1.000	0.007	0.086	0.000	1.000

Table 2.4.: Summary statistics of variables that determine local demand, data from 80 large German cities (excluding Berlin, Hamburg, Munich, Cologne). Author's own calculations based on data from Zensus (2011), OpenStreetMap contributors (2017), and Deutscher Apotheker Verlag (2016).

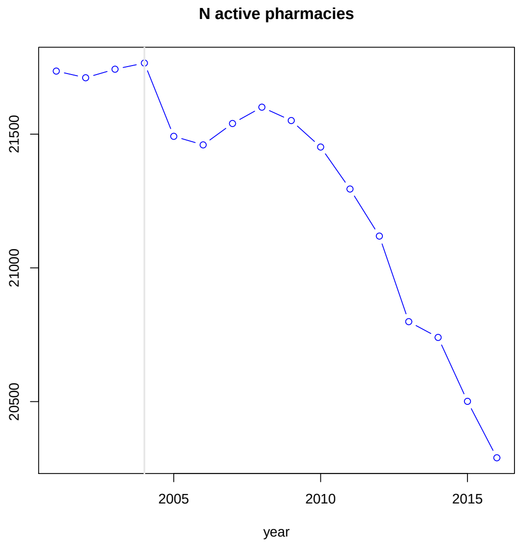
Variable	*Explanation*	*Variation*	*mean*	*SD*	*min*	*max*
vacancy rate	among all buildings, (2011)	total	0.043	0.023	0.016	0.136
commuters (in)	share of total population (2012)	total	0.091	0.114	-0.056	0.452
sq. m. price	annual price of construction land	total	0.220	0.155	0.002	1.236
	per square metre, in 1,000€	between		0.140	0.027	0.792
		within		0.067	-0.098	0.793
unemployment	annual unemployment rate	total	0.099	0.035	0.031	0.237
		between		0.029	0.042	0.169
		within		0.019	0.038	0.172
income	annual real disposable income	total	2.069	0.281	1.569	4.337
	per capita, in 10,000€	between		0.270	1.638	3.335
		within		0.082	1.088	3.071
income growth	annual growth rates	total	0.002	0.018	-0.100	0.100
	in per capita units	between		0.004	-0.006	0.020
		within		0.018	-0.109	0.106
population growth	annual population growth	total	0.003	0.008	-0.043	0.058
		between		0.005	-0.008	0.013
		within		0.006	-0.046	0.055

Table 2.5.: Summary statistics of variables that determine demand at the municipal level. Data from 80 large German cities (excluding Berlin, Hamburg, Munich, Cologne). The between and within variations are only computed for variables which actually vary over time. Author's own calculations based on data from Statistische Ämter des Bundes und der Länder (2018).

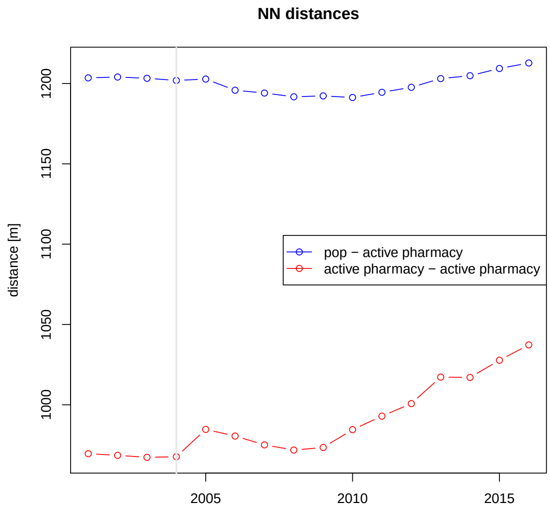
Third, further economic municipality-level co-variates from the German statistical offices (Statistische Ämter des Bundes und der Länder, 2018) were assigned to each location based on municipal boundaries (Bundesamt für Kartographie und Geodäsie, 2018). The local population data were scaled with municipality level population growth rates. As a robustness check, I also estimated the model with unscaled population data. Table 2.5 shows the summary statistics of these variables across cities, and across observational periods.

2.6.2. Descriptive analysis

The most interesting information contained in the panel data set are the pharmacies' relative locations, and how they change over time. The analysis of these spatial data is complicated by the fact that individual observations are usually not independent of each other, so that classical statistical concepts cannot be used. For this reason, a specialized branch of statistics has evolved that is concerned with the analysis of such spatial point patterns. Although the methods from spatial point pattern statistics lack a direct economic interpretation, they are nonetheless useful to describe the observed data appropriately. This subsection describes the spatial data set using methods from point pattern statistics. A good introduction to point pattern statistics is Diggle (2014).



(a) No. of active pharmacies



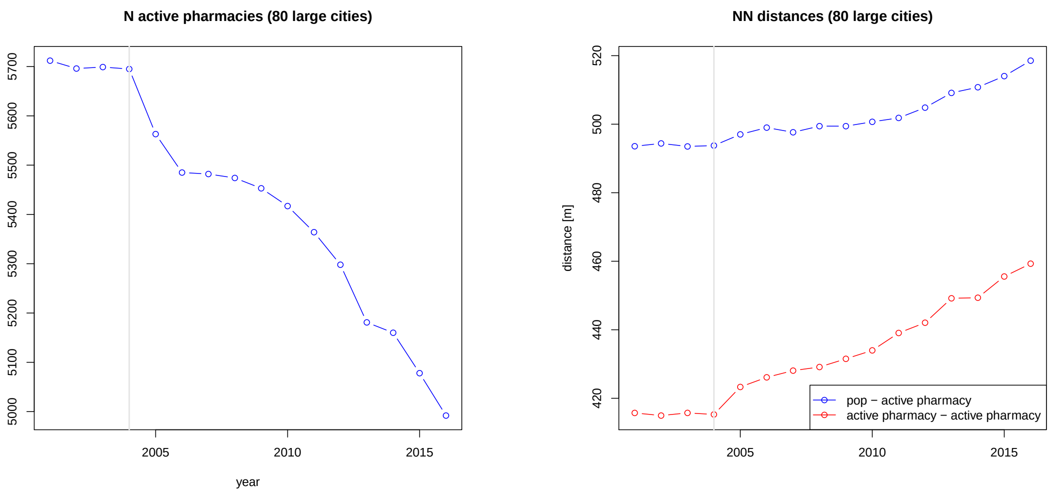
(b) Nearest neighbour metrics

Figure 2.10.: Spatial configuration of pharmacies in Germany

First, the data are described in terms of nearest neighbour distances. For each active pharmacy in every year, the distance towards its nearest active competitor was computed. The average nearest competitor distance is plotted in figure 2.10b along with the average population weighted distance from consumer cells (Zensus, 2011) to closest active pharmacies. That metric can be thought of as a crude welfare measure that can be used to evaluate the importance of changes in the spatial distribution of stores from a consumer perspective. The figure shows that the average nearest pharmacy distance changes very little over time, in the magnitude of only a few metres. On the other hand, there has been a marked increase by 70 metres in the nearest competitor distance since 2004. This change is not a large one in absolute terms, but it is still remarkable because changes in the spatial equilibrium configuration are naturally expected to be a rather slow process.[28] The structural empirical estimation below will use a subset of data from eighty large German cities (see table A.1). Figure 2.11 below shows the development of the number of active pharmacies, as well as the nearest competitor and nearest pharmacy distances over time for this sub-sample of the data. The patterns shown in figures 2.10 and 2.11 are qualitatively very similar, although the magnitudes of the nearest neighbour distances are naturally much smaller. The descriptive evidence so far is consistent with the hypothesis of increased competition among nearby competitors as a result of the introduction of price competition to the retail pharmacy market.

A key question that is addressed in spatial point pattern statistics is whether the observed events occur independently of each other, or not. In particular, researchers are often interested in whether the data generating process exhibits a tendency to produce clustered or, on the contrary, regular point patterns. As a benchmark hypothesis, it is often assumed that the points are generated by a spatial

[28]Note that it is not straightforward to test for yoy changes because the data are not independent, and do not converge to a normal distribution. That is why (Diggle, 2014, p.19) suggests using bootstrap tests in spatial point statistics.



(a) No. of active pharmacies

(b) Nearest neighbour metrics

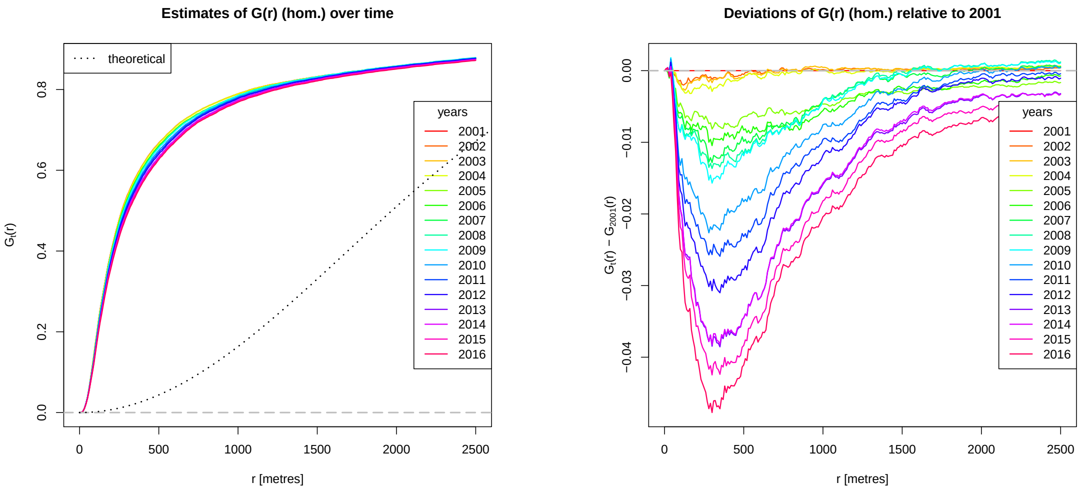
Figure 2.11.: Spatial configuration of pharmacies in 60 large German cities

Poisson point process. Poisson processes generate point patterns that are characterized by complete spatial randomness (CSR). Their theoretical properties are used to construct formal tests against the null hypothesis of CSR. The interpretation of such tests is however complicated by the fact that it is very difficult to formally distinguish whether a point process exhibits inherent clustering, or whether the observed clustering is an artefact of some unobserved spatial heterogeneity (Diggle, 2014, chapters 2 and 4). One frequently used statistic to describe the properties of spatial point processes is the cumulative distribution function of the nearest neighbour distances between points ("events"), called the G-function.[29] It can be used to construct a test against CSR, but also more generally to test whether two different point patterns share the same distributive properties. The tests are usually constructed as exact Monte-Carlo tests (Diggle, 2014, p.19).

The left panel of figure 2.12 shows the estimated G-functions for all years from 2001 through to 2016. The dotted line represents the theoretical G-function under CSR. The figure clearly shows that the empirical G-functions differ from their theoretical counter-part under CSR, which can be formally confirmed by means of a bootstrap Monte Carlo test. But as discussed above, this is does by no means imply that pharmacies have an inherent tendency to cluster together. Instead, it is far more likely that the observed tendency to cluster is the result of spatial heterogeneity in local demand conditions, which can be unobserved (the "attractiveness" of a location) or observed (such as the local residential population).[30] To ascertain whether the spatial pattern of pharmacies indeed exhibits clustering, one would have to control for all factors that affect the probability to open up a pharmacy at a given location. However, since the prime goal of this paper is whether the 2004 health care reform has *changed* the

[29]In Diggle (2014), the K-function is discussed as an alternative measure. It has the appealing property that it does not depend on the total number of observations, but is also less intuitive to explain and thus requires a more extensive discussion.

[30]There are methods to compute the G-function in the presence of spatial heterogeneity, but these methods are highly susceptible to obtaining an initial estimate of the spatial intensity function (Diggle, 2014).



(a) $G_t(d)$ **(b)** $G_t(d) - G_{2001}(d)$

Figure 2.12.: The distribution of nearest-neighbour distances (G-function) over time

spatial pattern of locations over time, the panel structure of the data allows me to examine this more directly. To make the changes over time more visible, the difference $G_t(d) - G_{2001}(d)$ is shown in the right panel of figure 2.12. That figure clearly shows that the G-function has decreased over time and moved gradually closer to the theoretical function under CSR. Therefore, the amount of clustering has decreased over time. It seems plausible that this change has occurred as a consequence of price competition among nearby competitors, because the most pronounced changes occur in the range from zero to four hundred metres. In order to test whether the observed changes of the G-function are statistically significant, an exact Monte Carlo test for the equality of the G-function in two subsequent years was constructed. Further, a variation of the test which uses the differences between G_y and G_{2001}, $y = 2002, \ldots, 2016$ was conducted in order to test whether the cumulative changes relative to the base year are statistically significant.[31] The test statistics of these tests are shown in table 2.6, and the 0.95 Monte Carlo critical value was found to be 0.019. The null hypothesis that two distributions are the same is rejected if the test statistic exceeds the critical value. So the table shows that the year-over-year changes are never statistically significant at the 5% level, but that the cumulative changes relative to 2001 are significant from 2010 onwards.

The statistical analysis in this subsection has shown show that the spatial distribution of pharmacies exhibits pronounced changes over time. The structural model in the next section will explore the causes of this change from an economic perspective, and relate it to the introduction of price competition in the retail pharmacy market.

[31]The test was conducted with the methods described in Diggle (2014, p.19), with the exception that the *difference* between the G-functions of two point patterns is used as a test statistic. The distribution of this test statistic under CSR was computed by repeatedly simulating random point patterns.

t	Test statistics $G_t - G_{t-1}$	$G_t - G_{2001}$
2002	0.00006	0.00006
2003	0.00001	0.00011
2004	0.00005	0.00023
2005	0.00378	0.00487
2006	0.00035	0.00708
2007	0.00065	0.00752
2008	0.00014	0.00908
2009	0.00007	0.01034
2010	0.00361	0.02543*
2011	0.00138	0.03801*
2012	0.00134	0.05337*
2013	0.00453	0.08719*
2014	0.00001	0.08622*
2015	0.00146	0.10886*
2016	0.00161	0.13525*

H_0: G-functions are the same.
0.95 critical value: $q_{95} = 0.01933$

Table 2.6.: Test statistics for the equality of the distribution of nearest neighbour distances, G, in two subsequent years (first column) and relative to the base year 2001 (second column). The Monte Carlo critical value is based on 199 simulations of two random spatial point patterns.

2.7. Empirical results

The maximum likelihood estimator presented in section 2.5.3 is used to estimate the principal structural parameters of the dynamic entry game in a set of large German cities. I am mainly interested in the spatial interaction parameter δ in equation (2.20) that governs by how much the variable profits of a pharmacy change due to the presence of a nearby competitor. In order to assess the effect of the 2004 reform, this parameter is allowed to differ before and after 2004. Following the discussion about the institutional details of the market, I also allow for the entry costs to differ before and after the regulatory change in 2004. The empirical analysis proceeds as follows. In a first step, the spatial interaction radius d is chosen by means of a simplified model without strategic interaction. Then, the model that allows for strategic interaction at the local level of k nearest neighbours is brought to the data, and an appropriate value for k is chosen. The results of this model are discussed, followed by a model validation exercise. Lastly, I use the estimated model parameters to isolate the effect of more intense local competition on market outcomes, and also perform a counter-factual exercise to assess the effect of a hard geographic entry restriction. As a robustness check, I re-estimate the model with additional potential entry locations and with additional co-variates.

2.7.1. Specification search

In order to choose an appropriate spatial interaction radius d that determines whether two adjoining pharmacies have a direct effect on each others' profits (c.f. the profit equation (2.20)), a stripped

$d[m]$	$\hat{\delta}_{pre}$	$\hat{\delta}_{post}$	ll	$d[m]$	$\hat{\delta}_{pre}$	$\hat{\delta}_{post}$	ll
100	0.051	-0.004	-14837.8	1100	0.003	0.009	-14783.7
200	0.023	0.022	-14833.9	1200	0.003	0.009	-14788.2
300	0.018	0.018	-14823.6	1300	0.003	0.009	-14789.8
400	0.011	0.014	-14810.7	1400	0.002	0.009	-14794.3
500	0.008	0.013	-14798.9	1500	0.002	0.009	-14797.4
600	0.005	0.011	-14789.4	1600	0.002	0.009	-14801.4
700	0.004	0.010	-14787.8	1700	0.002	0.009	-14805.6
800	0.003	0.009	-14786.6	1800	0.002	0.009	-14810.0
900	**0.002**	**0.009**	**-14780.3**	1900	0.002	0.009	-14814.4
1000	0.002	0.009	-14780.5	2000	0.002	0.008	-14820.3

Table 2.7.: Estimates of the spatial interaction parameter and the log-likelihood for different values of the spatial interaction radius d, in a simple model without strategic interaction. A maximal log-likelihood is attained at $d = 900m$.

down version of the model that excludes any strategic interaction is estimated. This model is obtained by setting the number of strategic nearest neighbours to one (so firms assume that no competitor changes its status), and solving the resulting single-agent dynamic decision problem. Pharmacies' profits are given by equation (2.20), with variables as given in table 2.4, and their decision was modelled as outlined in section 2.5.2 with $k = 1$, i.e. no strategic neighbours are considered. Table 2.7 shows the estimated spatial interaction parameters, and the resulting log-likelihood value for different choices of the interaction parameter. It can be seen that the log-likelihood attains a maximal value for $d = 900$, and I use this value in the subsequent analysis. Also, the table shows that the estimates of the interaction terms δ remain approximately constant as the interaction radius increases further. Since the number of parameters is the same across all specifications, the Akaike information criterion would therefore lead to the same model selection.

Next, I use a similar procedure to determine the number of strategic nearest neighbours that corresponds to the size of the local state space which every firm forms their expectations about. For a spatial interaction radius of nine hundred metres, I estimate the model for $k = 1, 2, \ldots, 6$ and record the maximum likelihood value. Table 2.8 shows that the maximum likelihood is attained at $k = 5$. The table also shows that there is initially a large improvement in the log-likelihood when just one strategic nearest neighbour is added (going from $k = 1$ to $k = 2$). For larger values of k the log-likelihood remains approximately flat, and parameter estimates change only marginally but the computational time increases rapidly. Therefore, I am confident that larger values of k beyond of what is computationally feasible would not lead to great improvements or changes in the model.

The full parameter estimates are shown in table A.3 in the appendix. An interesting insight of this table is the apparent invariance of the estimated coefficients with regard to the choice of the parameter k. A formal test of one specification against the other is work to be done in the future[32] but the log-likelihood, and the estimated coefficients are virtually the same for any $k > 1$. The specification without strategic interaction ($k = 1$) attains a somewhat smaller log-likelihood, and estimated parameter values that are

[32]This is complicated by the fact that the models are not nested into each other

k	$\hat{\delta}_{pre}$	$\hat{\delta}_{post}$	ll	runtime [s]
1	0.0024	0.0089	-14780.27	730.00
2	0.0021	0.0087	-14769.41	587.00
3	0.0020	0.0086	-14768.09	1170.00
4	0.0020	0.0087	-14768.42	2186.00
5	**0.0020**	**0.0087**	**-14767.65**	**4067.00**
6	0.0020	0.0088	-14768.27	15673.00

Table 2.8.: Estimates of the spatial interaction parameter and the log-likelihood for different sizes of the strategic neighbours k, at a spatial interaction radius $d = 900m$. A maximal log-likelihood is attained at $k = 5$. The estimation with $k = 1$ had a larger runtime because the parameters were not initialized in that specification.

different, albeit probably not significantly so. This could indicate that very little strategic interaction takes place among pharmacies and that players only take into account the actions of their immediate nearest neighbouring location. On the other hand, the kind of strategic interaction that is built into the model by means of adding "strategic neighbours" is of anticipatory nature, and so another explanation is that firms do not engage in anticipating their neighbours' actions and instead base their decisions solely on the currently observed market state. This behaviour is already captured by the panel structure of the data.

Does this invalidate the chosen approach of building a dynamic entry model with strategic interaction? The answer is no. For although the result is that anticipatory strategic interactions play a relatively minor role in determining firms' behaviour in the industry under consideration, the answer could be a different one in a different industry. The strength of the approach presented here is that it allows the researcher to test to what degree agents make strategic anticipatory decisions. These insights can also guide modelling approaches in other contexts with regard to how much emphasis is placed on dynamic strategic interactions of agents.

2.7.2. Main result

Table 2.9 below shows the main estimation result that was obtained from a sample of eighty large German cities with more than six thousand pharmacy locations, using $d = 900m$ and $k = 5$ as outlined above. Standard errors were computed using the estimated inverse Hessian matrix that is returned by the BFGS optimization routine. The exit value θ^x was normalized to unity. Further results are deferred to table A.4 in appendix A.2.5.

The estimated spatial interaction parameter δ is smaller and insignificant before the reform, and it is more than four times as large in the post-period, and significantly different from zero. This confirms one key hypothesis of the paper, namely that the introduction of price competition for non-prescription drugs has stiffened competition among nearby pharmacies, and so increased the tendency of firms to locate further away from each other. The estimates imply that one additional active competitor within a radius of 900m reduces profit margins by about 0.9%. The total effect on variable profits is larger than

Local demand		
intercept	0.2941***	*(0.0092)*
post reform	-0.0324***	*(0.0065)*
local population†	0.1682***	*(0.0260)*
elderly share	0.0968***	*(0.0258)*
pedestrian zone	0.0163***	*(0.0042)*
mainroad	-0.0024	*(0.0032)*
doctor nearby	0.0272***	*(0.0039)*
supermarket nearby	0.0323***	*(0.0038)*
trainstation nearby	0.0187**	*(0.0074)*
Local competition		
δ_{pre}	0.0020	*(0.0025)*
δ_{post}	0.0087***	*(0.0017)*
Entry costs		
θ^e_{pre}	4.9005***	*(0.0393)*
θ^e_{post}	4.3318***	*(0.0179)*
log-likelihood	-14,767.7	
N locations	6,741	
T periods	16	

$^{*} < 0.1$; $^{**} < 0.05$; $^{***} < 0.01$

Table 2.9.: Estimates of the spatial entry model with a strategic neighbourhood of size five (self and four nearest neighbours), and a spatial interaction radius of nine hundred metres. N=6741 firms, T=16 time periods in 80 large German cities. Exit values are normalized to 1. †local population is the residential population within 500 metres of the store's location, divided by the number of active competitors in the respective time period or future state. Standard errors in parentheses, computed from estimated Hessian matrix.

this, because local demand decreases as the number of active neighbours increases. Since the number of active nearest neighbours within that radius can be quite large in urban areas, this implies that competition has now sizeable effects on pharmacies' profitability, whereas it was virtually nil before the reform. Because the estimated interaction parameters are also rather small in comparison to the magnitudes of the other parameters, the model is likely to possess only one single equilibrium, as is also discussed in section 2.5.3. Entry costs are smaller in the post-reform period, and the post-reform dummy is negative which points to smaller period returns in the post period. Lower entry costs and smaller period returns together imply that there is more turnover. The size of the local population, divided by the number of active stores within the interaction radius has a significantly positive effect on period returns, as well as the local share of the population that is older than 65 (elderly share). The coefficients for proximity to public transport, supermarkets and physicians generally have the anticipated signs, whereas proximity to a main road seems to have a negative effect on profits, albeit insignificant.

2.7.3. Model validation

In order to assess whether the model can replicate key trends that are observed in the data (see figure 2.10), I simulated a large number of counter-factual market outcomes, starting from the observed market state in the year 2004 and using either the pre-reform estimates (i.e. post reform dummy set to zero, $\delta = \hat{\delta}_{pre}$, and θ^e_{pre}), or the corresponding post-reform estimates. The simulated market outcomes

	year	Number of stores	distance (metres) pharmacy to nearest competitor	consumer to nearest pharmacy
observed				
	2004	5,437	423	514
	2016	4,770	469	539
	Δ	**-667**	**+45**	**+25**
simulated				
pre-reform parameters	2016	5,466	407	526
post-reform parameters	2016	4,850	441	558
	Δ	**-615**	**+34**	**+32**

Table 2.10.: Comparison of actual changes throughout the post-reform period, and the model-implied differences between the pre- and the post-reform periods.

are analysed along three dimensions: (1) number of active stores, (2) average distance to the nearest competitor and (3) average consumer distance to the nearest pharmacy.

The simulation results are shown graphically in figure A.1 in appendix A.2.4. The top panel in figure A.1a shows that the model, when set up with the pre-reform parameter values, predicts that the total number of active stores remains approximately constant, as desired. When using the post-reform parameter values, shown in the top panel of figure A.1b, this number exhibits a downward sloping trend that does follow the observed number of active stores quite closely. On the other hand, the bottom two panels in figures A.1a and A.1b show that the simulated average store-to-store distance is smaller, and the average simulated consumer travel distance is larger than what is observed prior to the reform, and after the reform, respectively. Apparently, the model-implied competition among nearby stores is too small to fully replicate the observed behaviour. This could also be due to the linear way in which I modelled the competition among nearby stores, where every store within a radius d has the same negative effect δ on its competitors' profit margins. It seems plausible that stores which are closer also exhibit more competitive pressure, but modelling this in a reduced form manner would require an arbitrary choice of a functional form. Instead, one could integrate a truly spatial demand model in the estimation routine, which would automatically capture such effects.[33] This is work to be done in the future.

Despite the discrepancies between the observed and the model-implied simulation results, I argue that the *difference* between the two simulation results (pre- and post-reform) accurately reflects the changes that have occured due to the reform. This is supported by table (2.10). This table shows that the model-implied differences between the pre- and the post-reform period are rather close to the observed changes from 2004 to 2016, and it therefore confirms that the changes of the structural parameters, when suitably interpreted, can indeed explain a good part of what was observed in the post reform period.

[33] On the other hand, such a structural demand model, possibly combined with a price equilibrium, is much harder to interpret than the single interaction parameter δ.

	post2004	entry costs	interaction	outcome of interest
(A) compare	0	$\hat{\theta}^e_{pre}$	$\hat{\delta}_{post}$	$t(0, \theta^e_{pre}, \delta_{post})$
against	0	$\hat{\theta}^e_{pre}$	$\hat{\delta}_{pre}$	$-t(0, \theta^e_{pre}, \delta_{pre})$
				$= \Delta_A t$
(B) compare	1	$\hat{\theta}^e_{post}$	$\hat{\delta}_{post}$	$t(1, \theta^e_{post}, \delta_{post})$
against	1	$\hat{\theta}^e_{post}$	$\hat{\delta}_{pre}$	$-t(1, \theta^e_{post}, \delta_{pre})$
				$\Delta_B t$

Table 2.11.: Counterfactual simulations to quantify the effect of stiffer price competition

2.7.4. Quantifying the effect of price competition

What can be seen in the data as well as in the simulations in figure A.1b is the total effect of three changes: first, lower entry costs; second, lower per-period profits; and third, increased local competition due to price competition. To isolate the effect of increased local competition, the idea is to differentiate the model's predictions with respect to the spatial interaction parameter δ. Since no closed form solutions are available, this is done numerically, as outlined in the following. The effect of increased price competition can be evaluated using either the pre-reform parameter estimates as a starting point, or by using the post-reform estimates as a starting point. More precisely, consider a simulated time series of market outcomes that was generated using the parameters $post\_2004 \in \{0, 1\}$, θ^e and denote an arbitrary aggregate statistic that was computed from this simulated data as $t(post\_2004, \theta^e, \delta)$. Then, the partial effect of changing δ from δ_{pre} to δ_{post} can be computed as either $\Delta_A = t(0, \theta^e_{pre}, \delta_{post}) - t(0, \theta^e_{pre}, \delta_{pre})$ or as $\Delta_B = t(1, \theta^e_{post}, \delta_{post}) - t(1, \theta^e_{post}, \delta_{post})$. These two estimates of the partial effect will in general be different, but it turns out that they are quite close to each other. Table 2.11 summarizes the procedure.

Table 2.12 shows the results of this decomposition exercise for the three aggregate statistics that are also shown graphically in figure A.1, using simulated data runs from 2004 to 2016. The table shows that increased local competition due to changing the interaction parameter from δ_{pre} to δ_{post} can explain about one third of the decline in the number of pharmacies. The table also shows that about one third of the observed increase in the inter-firm nearest neighbour distance can be attributed to this change of parameters, whereas only one sixth to one seventh of the total increase in consumer travel distances are attributable to this factor.

Thus, from a consumer perspective, while increased price competition has lead to a substantial reduction of the number of pharmacies, it did not lead to much greater travel distances, presumably because it has caused the exit of retail pharmacies that were located very close to another competitor which can offer the same services and products. If the aim of the 2003 health care reform was to reduce the costs of the health care system, introducing a modest degree of price competition into the retail pharmacy sector has thus been a very consumer-friendly way of reducing the number of pharmacies and, thereby, the total fixed costs of the health care system. One should note that a full welfare analysis is not possible due to the lack of detailed price data. However, it seems reasonable to assume that prices did not

		distance (metres)	
	Number of stores	pharmacy to nearest competitor	consumer to nearest pharmacy
Total change 2004-2016			
actual	-667	+45	+25
simulated	-616	+34	+32
of which: competition effect			
(A) $t(0, \theta^e_{pre}, \delta_{post}) - t(0, \theta^e_{pre}, \delta_{pre})$	-224	+15	+4
(B) $t(1, \theta^e_{post}, \delta_{post}) - t(1, \theta^e_{post}, \delta_{post})$	-215	+16	+4

Table 2.12.: The effect of increased price competition on aggregate outcomes in 2016 using model (5) in table 2.9, average over 100 simulations, 80 large Germany cities.

increase as a result of more intense price competition, and so the cost savings due to the lower number of retail pharmacies probably outweigh the small increase in consumer travel distances.

2.7.5. Policy experiment: geographical entry restriction

The theoretical model in section 2.4 has shown that free location choice in the absence of spatial competition leads to inefficient, Hotelling-style clustering, thus inflicting inefficiently large travel costs on consumers. This is precisely the reason why many European countries have or had minimum distance regulations in place whereby a minimum distance $\bar{d}$ between any two active pharmacies must be maintained at all times (see section 2.2.2). The estimated model allows me to evaluate the effect of imposing such a regulation on the German pharmacy market, by changing the period return structure to include a “penalty” parameter p that is subtracted from firms’ profits if any one of their active neighbours is closer than the regulated minimum distance. Thus, new period returns are given by

$$\pi_j^{reg} = \begin{cases} \pi_j - p \, , \text{if } \exists k : d_{jk} \leq \bar{d} \\ \pi_j \, , \text{else.} \end{cases}$$

where π_j is specified in (2.4). I chose to set $p = \theta^e$, so that entry is unprofitable already after the first period, as desired. Then, two hundred independent market configurations are simulated, starting from the market configuration in 2004.

The average statistics across two hundred independent simulations for a minimum distance of one hundred metres are shown in figure A.3a in appendix A.2.4. That figure shows that such a regulation would immediately reduce the number of active pharmacies by around six hundred stores, or by around ten percent, followed by a further decline that is due to the changed profit and competition conditions after the 2004 reform. As an immediate consequence, the nearest competitor distance would have instantaneously increased by roughly seventy metres, followed by a further increase. Consumer to nearest pharmacy distances would have increased, too.

		distance (metres)	
	Number of stores	pharmacy to nearest competitor	consumer to nearest pharmacy
Total simulated change 2004-2016			
post-reform parameters	−616	+34	+32
post-reform + distance regulation	−1036	+93	+46
distance regulation effect	**−421**	**+60**	**+14**

Table 2.13.: The effect of a minimum distance regulation on key market outcomes. The first line shows the simulated changes from 2004 to 2016, using the post-reform parameter estimates. The second line shows the simulated changes using the post-reform parameter estimates and a minimum distance regulation of 100 metres. The last line denotes the additional simulated effect of a minimum distance regulation. Numbers are averages over 100 simulations.

To interpret these simulation results, it is important to keep in mind that neither the nearest competitor distance, nor the nearest pharmacy distance, have been used as direct estimation moments. Therefore, the model predictions for these statistics are biased, as was shown in section 2.7.3. Thus, to assess the effects of a minimum distance regulation it is necessary to determine how these model predictions change as such a policy change is implemented. The idea behind this is to "differentiate" the model predictions with respect to a policy change by means of simulation, akin to the procedure used in the previous section. To this end, I compared the simulated outcomes under the distance regulation to those without such a regulation in table 2.13. That table shows that the total number of stores in 2016 would have been smaller by about four hundred stores had such a regulation been in place since 2004. The nearest competitor distance would have been larger by sixty metres, whereas the nearest pharmacy distance would have increased by only fourteen metres. A graphical depiction of the difference between the simulated market outcomes with and without a minimum distance regulation is shown in figure A.3b. The top panel of that figure shows that the number of active firms immediately decreases relative to the base line scenario, but then the difference remains relatively constant. Also, the nearest competitor distance remains approximately constant relative to the base line scenario. The most important insight from this analysis is that the nearest pharmacy distances do not increase much more than in the baseline scenario.

One apparent caveat of this exercise lies in the fact that I am only using observed locations, whereas such a drastic regulatory measure may actually lead to new locations becoming feasible. Section A.2.5 in the appendix addresses this concern by including a larger set of potential entry locations. The results however remain qualitatively the same. A second concern is that pharmacies compete along a quality dimension, so such a regulatory scheme could lead to lower service quality because it actually creates local monopolies. More generally speaking, since the period returns of the dynamic entry model are modelled in a reduced form, counter-factual analyses are in principle subject to the Lucas' critique (Lucas, 1976) in that the estimated reduced form coefficients tell us little about the agent's reactions to such a drastic change in the economic environment. But to a lesser extent, this would also be true for a more elaborate model with a "structural" profit equation. Any model, be it reduced form or structural, can only inform us about those aspects of agents' decision making that are built into it: a structural model of price competition can make predictions about price reactions to the ownership structure only if the ownership structure is part of the model. Similarly, a model of dynamic spatial entry can inform us about the responses to minimum distance regulations, but not how pharmacies would change their business model, their opening hours, or their service quality.

2.8. Conclusion

I have documented pronounced qualitative and quantitative changes in the spatial distribution of pharmacies in Germany over time. Motivated by a simple theoretical model, I developed a structural dynamic entry model and used it to estimate the key parameters that govern the process of spatial entry and exit. These parameter estimates indicate that local competition has indeed increased after 2004, most likely due to a large health care reform that introduced price competition for non-prescription drugs. A simulation exercise shows that increased competition can explain one third of the total change in the number of pharmacies, but only a small share of the increase in consumer's travel distances. This suggests that more price competition can lead to more efficient spatial store configurations in that the total number of stores is reduced, which implies lower fixed costs, while consumers do not have to travel much farther. Thus, even abstracting from the fact that prices are likely to be smaller due to price competition, increased price competition can lead to better market outcomes. I have also examined the likely effects of introducing a geographic entry barrier that prevents stores from locating very close to each other. My results show that the effects of such a regime are similar to those that are generated by the introduction of price competition in that the total number of stores decreases, but consumers' travel costs do not increase very much. But because such a regulatory regime amounts to establishing local monopolies, it probably has detrimental effects on consumer welfare that are not captured in the model. Therefore, the introduction of price competition is the more efficient regulatory measure.

The analysis has a number of shortcomings which are left for future research. First of all, it restricts the analysis to urban markets. An extension of the analysis to rural markets is possible, but because consumers' travel patterns and, in consequence, the range of spatial interaction among pharmacies, are likely to be very different in those markets, this calls for a separate analysis, perhaps using the isolated market paradigm of Bresnahan and Reiss (1991). Second, the reduced form profit equation could be replaced by a structural revenue model with spatial demand and endogenous prices, but this is currently infeasible in the context of a dynamic model due to the large additional computational burden. Yet, my results indicate that the anticipatory strategic component does not play a large role, and so a simpler model of dynamic decision making could well be used to that effect.

On the methodological front, I have established a method to compute and estimate a spatial dynamic entry model with a large number of asymmetric heterogeneous agents. The method has proves to work well with thousands of potential entry locations, and could be extended to include more sophisticated "structural" period return functions. Due to the flexible way in which the size of the strategic neighbourhood is specified, the model can be used to examine in how far strategic anticipatory motives play a role in dynamic decision making. In principle, this model is applicable to a wide range economic questions, but the main application lies in retail markets where spatial interaction and strategic decision making are important factors.

A. Appendix to Chapter 2

A.1. Proofs

First, consider the proof of proposition 1, repeated here for convenience:

Proposition. *The symmetric space-then-price equilibrium of the model described above is characterized by location choices $x_a = x^*$ and $x_b = 1 - x^*$ with*

$$x^* = \begin{cases} 0 & \text{if } \alpha \leq \frac{\tau}{3+\tau}, \\ -\frac{1}{4} + \frac{3}{4}\frac{\alpha}{\tau(1-\alpha)} & \text{if } \frac{\tau}{3+\tau} < \alpha < \frac{\tau}{1+\tau}, \\ \frac{1}{2} & \text{if } \alpha \geq \frac{\tau}{1+\tau}, \end{cases}$$

and prices $p_a = p_b = p^*$ with

$$p^* = \begin{cases} \tau & \text{if } \alpha \leq \frac{\tau}{3+\tau}, \\ \frac{3}{2}\left(\tau - \frac{\alpha}{1-\alpha}\right) & \text{if } \frac{\tau}{3+\tau} < \alpha < \frac{\tau}{1+\tau}, \\ 0 & \text{if } \alpha \geq \frac{\tau}{1+\tau}. \end{cases}$$

Proof. With quadratic travel costs, the consumer who is just indifferent between purchasing a price regulated prescription drug from either pharmacy is located at $\bar{\imath} = (x_a + x_b)/2$, and the demand for prescription drugs is hence given by $Q_a^{Rx} = \alpha\bar{\imath}$ and $Q_b^{Rx} = \alpha(1 - \bar{\imath})$. The indifferent consumer who purchases a non-prescription drug is located at

$$\breve{\imath} = \frac{x_a + x_b}{2} + \frac{1}{2\tau}\frac{p_b - p_a}{x_b - x_a}$$

so that demand for non-prescription drugs is similarly given by $Q_a^{OTC} = (1-\alpha)\breve{\imath}$ and $Q_b^{OTC} = (1-\alpha)(1-\breve{\imath})$.

Consider first the price equilibrium, given location choices x_a and x_b. Because there are no complementarities across products, firms will essentially set the price for non-prescription drugs so as to maximize variable profits from that product category. Maximizing the firms' profit equations, and re-arranging the first-order conditions, it is easy to derive the following price equilibrium:

$$\begin{aligned} p_a^* &= \frac{\tau}{3}(x_b - x_a)(2 + x_a + x_b) \\ p_b^* &= \frac{\tau}{3}(x_b - x_a)(4 - x_a - x_b) \end{aligned} \tag{A.1}$$

Note that, in the symmetric location case where $x_b = 1 - x_a$, I have $p_a^* = p_b^* \equiv p_{symm}^* = \tau(1 - 2x_a)$. The location of firm a is a measure of centrality, and prices are lower when the two firms are located closer to the market centre (and therefore closer to each other).

Next turn to the equilibrium location decisions of both firms. Anticipating the equilibrium effect on prices, firms now choose their locations simultaneously so as to maximize their profits. In doing so, they must trade off the market share effect of being located closer to the market centre against the market power effect of being located more distantly from their competitor:

$$\frac{\mathrm{d}\pi_j}{\mathrm{d}x_j} = \underbrace{\alpha \frac{\partial Q_j^{Rx}}{\partial x_j} + (1-\alpha)p_j \frac{\partial Q_j^{OTC}}{\partial x_j}}_{\text{market share effect}} + \underbrace{(1-\alpha)p_j \frac{\partial Q_j^{OTC}}{\partial p_{-j}} \frac{\mathrm{d}p_{-j}^*}{\mathrm{d}x_j}}_{\text{market power effect}} \tag{A.2}$$

The equilibrium effect on own prices is cancelled out in equilibrium, so that the market power effect only includes the equilibrium effect of one's re-locations on the prices of other firms. I am looking for a symmetric equilibrium x^* where $x_a = x^*$ and $x_b = 1 - x^*$. The price equilibrium then implies $p_a^* = p_b^* = \tau(1 - 2x^*)$. Using results from above, equation (A.2) can be used to compute a candidate equilibrium:

$$\left.\frac{\mathrm{d}\pi_a}{\mathrm{d}x_a}\right|_{x^*} = 0 \Leftrightarrow x^* = -\frac{1}{4} + \frac{3}{4}\frac{\alpha}{\tau(1-\alpha)}$$

With locations restricted to the unit interval, and $x_a \leq x_b$, the locational equilibrium must satisfy $x^* \in \left[0, \frac{1}{2}\right]$. Hence, an equilibrium location is an interior location whenever the share of consumers purchasing price regulated prescription drugs α is sufficiently large, but not too large:

$$x^* = \begin{cases} 0 & \alpha \leq \frac{\tau}{3+\tau} \\ -\frac{1}{4} + \frac{3}{4}\frac{\alpha}{\tau(1-\alpha)} & \frac{\tau}{3+\tau} < \alpha < \frac{\tau}{1+\tau} \\ \frac{1}{2} & \alpha \geq \frac{\tau}{1+\tau} \end{cases} \tag{A.3}$$

The ensuing price equilibrium is then given by

$$p^* = \begin{cases} \tau & \alpha \leq \frac{\tau}{3+\tau} \\ \frac{3}{2}\left(\tau - \frac{\alpha}{1-\alpha}\right) & \frac{\tau}{3+\tau} < \alpha < \frac{\tau}{1+\tau} \\ 0 & \alpha \geq \frac{\tau}{1+\tau} \end{cases} \tag{A.4}$$

□

This section continues with the proof for the proposition 2 in section 2.4, split up into two lemmas for convenience.

Lemma 1. *Suppose that the regulated price of prescription drugs* $\bar{p}$ *was zero, and that parameters are such that* $x^* < \frac{1}{2}$ *in the competitive equilibrium. Then, the difference of consumer welfare in the competitive equilibrium relative to the regulatory benchmark increases strictly as the share of prescription consumers* α *increases* ($\frac{\mathrm{d}}{\mathrm{d}\alpha}\Delta W(0) > 0$).

Proof. One can distinguish two cases here. *Case 1:* Suppose that $\alpha \leq \frac{\tau}{3+\tau}$ so that $x^* = 0$. Then, $\Delta W = (\alpha - 1)\tau < 0$ and the change in consumer welfare is obviously increasing strictly in α, as $\tau > 0$ by assumption. *Case 2:* If $0 < x^* < \frac{1}{2}$ it must hold that $\frac{\tau}{3+\tau} < \alpha < \frac{\tau}{1+\tau}$. Because $0 \leq \tau < 1$, this implies $0 < \alpha < 1$. Note first that the marginal change of the firms' equilibrium locations with respect to α is given by

$$\frac{\mathrm{d}x^*}{\mathrm{d}\alpha} = \frac{3}{4\tau(1-\alpha)^2}$$

Because this expression is increasing strictly in α for any $\alpha \in (0,1)$, and because $\alpha > \frac{\tau}{3+\tau}$ it can be bounded below by the function $f(\tau) = \frac{1}{12}\frac{(3+\tau)^2}{\tau}$. This term, in turn, is strictly decreasing for $\tau \in [0,1]$ and hence, its minimum is attained at $\tau = 1$ with $f(1) = \frac{4}{3}$. Therefore,

$$\frac{\mathrm{d}x^*}{\mathrm{d}\alpha} \geq \frac{4}{3} \tag{A.5}$$

with a strict inequality if $\tau < 1$. Also, note that $\alpha < \frac{\tau}{1+\tau}$ is equivalent to $\tau(1-\alpha) > \alpha$ which will be convenient below. The change in consumer welfare in (2.3) can equivalently be written as

$$\Delta W = (\alpha - 1)\tau + \frac{1}{2}\left(5\tau + \alpha(1-5\tau)\right)x^* - (\tau + \alpha(1-\tau)){x^*}^2$$

The marginal change of ΔW with respect to changes in α is

$$\frac{\mathrm{d}\Delta W}{\mathrm{d}\alpha} = \tau + \underbrace{\left[\frac{1-5\tau}{2} - (1-\tau)x^*\right]}_{\equiv A} x^* + \underbrace{\left[\frac{1}{2}(\alpha + 5\tau(1-\alpha)) - 2x^*(\alpha + \tau(1-\alpha))\right]}_{\equiv B}\frac{\mathrm{d}x^*}{\mathrm{d}\alpha} \tag{A.6}$$

First, consider term A. Substituting for the expression of equilibrium location choices, one obtains

$$\begin{aligned}
A &= \frac{1}{2}(1-5\tau) - \left(-\frac{1}{4} + \frac{3}{4}\frac{\alpha}{\tau(1-\alpha)}\right)(1-\tau) \\
&= -\frac{3}{4}\frac{\alpha(1-\tau)}{\tau(1-\alpha)} + \frac{1}{4}(1-\tau) + \frac{1}{2}(1-5\tau) \\
&= -\frac{3}{4}\underbrace{\frac{\alpha(1-\tau)}{\tau(1-\alpha)}}_{<1-\tau} + \frac{3}{4} - \frac{11}{4}\tau \\
&> -\frac{3}{4}(1-\tau) + \frac{3}{4} - \frac{11}{4}\tau = -2\tau.
\end{aligned}$$

Next, consider term B. Because $x^* < \frac{1}{2}$ and with $\tau(1-\alpha) > \alpha$, it can be bounded below as follows:

$$\begin{aligned} B &> \frac{1}{2}(\alpha + 5\tau(1-\alpha)) - (\alpha + \tau(1-\alpha)) \\ &= -\frac{1}{2}\alpha + \frac{3}{2}\underbrace{\tau(1-\alpha)}_{>\alpha} > \alpha \end{aligned}$$

Finally, substituting for the lower bounds of A, B, $\frac{\mathrm{d}x^*}{\mathrm{d}\alpha}$, and x^* in (A.6), one can derive the following inequality

$$\begin{aligned} \frac{\mathrm{d}\Delta W}{\mathrm{d}\alpha} &> \tau + \frac{4}{3}\alpha - 2\tau\left(-\frac{1}{4} + \frac{3}{4}\frac{\alpha}{\tau(1-\alpha)}\right) \\ &= \frac{3}{2}\tau + \frac{4}{3}\alpha - \frac{3}{2}\frac{\alpha}{1-\alpha} \end{aligned}$$

So, a sufficient condition for $\frac{\mathrm{d}\Delta W}{\mathrm{d}\alpha} > 0$ is

$$\begin{aligned} \frac{3}{2}\underbrace{\tau(1-\alpha)}_{>\alpha} + \frac{4}{3}\alpha(1-\alpha) - \frac{3}{2}\alpha \geq 0 \\ \Leftarrow \frac{4}{3}\alpha(1-\alpha) \geq 0 \end{aligned}$$

which is true for all $\alpha \in [0, 1]$. □

Lemma 2. *Suppose that the regulated price of prescription drugs $\bar{p}$ was zero, and that parameters are such that $x^* < \frac{1}{2}$ in the competitive equilibrium. Then, the difference of consumer welfare in the competitive equilibrium relative to the regulatory benchmark decreases strictly as the share of prescription consumers τ increases ($\frac{\mathrm{d}}{\mathrm{d}\tau}\Delta W(0) < 0$).*

Proof. Case 1: Again, one can first consider the trivial cases where $x^* = 0$ so that $\Delta W = (\alpha - 1)\tau$ which is strictly decreasing in τ since $\alpha < 1$ is implied (see above). *Case 2:* If $0 < x^* < \frac{1}{2}$ it must hold that $\frac{\tau}{3+\tau} < \alpha < \frac{\tau}{1+\tau}$. Note first that the marginal change of the firms' equilibrium locations with respect to τ is given by

$$\frac{\mathrm{d}x^*}{\mathrm{d}\tau} = -\frac{3}{4}\frac{\alpha}{\underbrace{\tau^2(1-\alpha)}_{>\frac{1}{\tau}}} > -\frac{3}{4\tau}$$

The change in welfare (2.3) compared to the baseline case where prices are regulated can also be written as

$$\Delta W = (\alpha - 1)\tau + \frac{1}{2}\left(\alpha + 5\tau(1-\alpha)\right)x^* - (\alpha + \tau(1-\alpha))x^{*2},$$

the marginal change of which with respect to τ is

$$\frac{\mathrm{d}\Delta W}{\mathrm{d}\tau} = -(1-\alpha)+\underbrace{\left(\frac{5(1-\alpha)}{2} - x^*(1-\alpha)\right)}_{\equiv C}x^*+\underbrace{\left(\frac{1}{2}(\alpha+5\tau(1-\alpha)) - 2x^*(\alpha+\tau(1-\alpha))\right)}_{\equiv B}\frac{\mathrm{d}x^*}{\mathrm{d}\tau}. \tag{A.7}$$

Note that term B in the above equation is the same as term B in equation (A.6), and so I already know that $B > \alpha$. Substituting for x^* in C and observing that $\alpha > \frac{\tau}{1+\tau}$ is equivalent to writing $\alpha/\tau > 1-\alpha$, I obtain

$$\begin{aligned} C &= \frac{5(1-\tau)}{2} - \left(-\frac{1}{4} + \frac{3}{4}\frac{\alpha}{\tau(1-\alpha)}\right)(1-\alpha) \\ &= \frac{11}{4}(1-\alpha) - \frac{3}{4}\underbrace{\frac{\alpha}{\tau}}_{>1-\alpha} < 2(1-\alpha). \end{aligned}$$

Noting that $0 < x^* < \frac{1}{2}$ and putting together the established lower bounds for B, C, and $\frac{\mathrm{d}x^*}{\mathrm{d}\tau}$ I can derive that

$$\frac{\mathrm{d}\Delta W}{\mathrm{d}\tau} < -(1-\alpha) + (1-\alpha) - \frac{3}{4}\frac{\alpha}{\tau} = -\frac{3}{4}\frac{\alpha}{\tau} < 0$$

because $\alpha > 0$ is implied, and $\tau > 0$ by assumption. □

Lemmas 1 and 2 now allow me to prove the validity of proposition 2, repeated below for the reader's convenience:

Proposition. *Suppose that the regulated price of non-prescription drugs, $\bar{p}$, was initially zero. Then:*

1. *if α and τ are such that $x^* = \frac{1}{2}$, consumer welfare is the same in the competitive equilibrium as in the regulatory benchmark ($\Delta W(0) = 0$);*
2. *if α and τ are such that $x^* < \frac{1}{2}$, (a) the difference of consumer welfare in the competitive equilibrium relative to the regulatory benchmark increases strictly in the share of prescription consumers ($\frac{\mathrm{d}}{\mathrm{d}\alpha}\Delta W(0) > 0$), and it decreases strictly in the travel costs of non-prescription consumers ($\frac{\mathrm{d}}{\mathrm{d}\tau}\Delta W(0) < 0$); and (b) consumer welfare is smaller in the competitive equilibrium ($\Delta W(0) < 0$).*

Proof. Point (1) is trivial. Point (2a) follows immediately from lemmas 1 and 2. To see point (2b), pick any point $(\alpha', \tau') \in [0,1]^2$ with $\alpha' < \frac{\tau'}{1+\tau'}$ so that $x^*(\alpha', \tau') < \frac{1}{2}$. Choose a different point (α'', τ'') with $\alpha'' = \frac{\tau''}{1+\tau''}$, $\alpha'' > \alpha'$ and $\tau'' < \tau'$. Such a point, by construction, always exists. Then,

$$\Delta W(0)_{\alpha',\tau'} = \underbrace{\Delta W(0)_{\alpha'',\tau''}}_{=0} + \int_{\alpha''}^{\alpha'} \underbrace{\Delta W(0)_{\alpha,\tau''}\mathrm{d}\alpha}_{>0\,\forall\alpha<\alpha''} + \int_{\tau''}^{\tau'} \underbrace{\Delta W(0)_{\alpha',\tau}\mathrm{d}\tau}_{<0} < 0$$

which concludes the proof (note again that $\alpha'' > \alpha'$ and $\tau'' < \tau'$). □

A.2. Supplementary empirical material

A.2.1. List of cities

	city	AGS8	N		city	AGS8	N
1	Düsseldorf	05 111 000	237	41	Oberhausen	05 119 000	66
2	Frankfurt a.M.	06 412 000	236	42	Ludwigshafen	07 314 000	64
3	Hannover	03 241 001	203	43	Hamm	05 915 000	63
4	Essen	05 113 000	200	44	Würzburg	09 663 000	61
5	Stuttgart	08 111 000	199	45	Heidelberg	08 221 000	61
6	Dortmund	05 913 000	197	46	Potsdam	12 054 000	60
7	Nürnberg	09 564 000	190	47	Paderborn	05 774 032	59
8	Bremen	04 011 000	179	48	Mülheim (Ruhr)	05 117 000	59
9	Leipzig	14 713 000	174	49	Darmstadt	06 411 000	57
10	Duisburg	05 112 000	157	50	Herne	05 916 000	54
11	Dresden	14 612 000	146	51	Leverkusen	05 316 000	54
12	Bonn	05 314 000	135	52	Neuss	05 162 024	53
13	Bochum	05 911 000	133	53	Koblenz	07 111 000	53
14	Münster	05 515 000	121	54	Solingen	05 122 000	51
15	Mannheim	08 222 000	118	55	Pforzheim	08 231 000	51
16	Wuppertal	05 124 000	109	56	Trier	07 211 000	50
17	Augsburg	09 761 000	109	57	Göttingen	03 152 012	50
18	Bielefeld	05 711 000	106	58	Ulm	08 421 000	49
19	Halle (Saale)	15 002 000	105	59	Erlangen	09 562 000	48
20	Karlsruhe	08 212 000	102	60	Recklinghausen	05 562 032	46
21	Wiesbaden	06 414 000	102	61	Ingolstadt	09 161 000	45
22	Gelsenkirchen	05 513 000	99	62	Zwickau	14 524 330	45
23	Mönchengladbach	05 116 000	96	63	Kaiserslautern	07 312 000	45
24	Aachen	05 334 002	96	64	Salzgitter	03 102 000	45
25	Braunschweig	03 101 000	93	65	Offenbach	06 413 000	44
26	Kiel	01 002 000	91	66	Heilbronn	08 121 000	44
27	Lübeck	01 003 000	89	67	Bremerhaven	04 012 000	43
28	Freiburg	08 311 000	87	68	Hildesheim	03 254 021	43
29	Krefeld	05 114 000	87	69	Wolfsburg	03 103 000	42
30	Magdeburg	15 003 000	83	70	Fürth	09 563 000	41
31	Mainz	07 315 000	82	71	Bamberg	09 461 000	40
32	Chemnitz	14 511 000	82	72	Flensburg	01 001 000	39
33	Kassel	06 611 000	81	73	Remscheid	05 120 000	39
34	Osnabrück	03 404 000	74	74	Cottbus	12 052 000	39
35	Erfurt	16 051 000	71	75	Gütersloh	05 754 008	39
36	Saarbrücken	10 041 100	70	76	Worms	07 319 000	38
37	Rostock	13 003 000	69	77	Siegen	05 970 040	38
38	Hagen	05 914 000	69	78	Bergisch Gladbach	05 378 004	38
39	Regensburg	09 362 000	68	79	Wilhelmshaven	03 405 000	37
40	Oldenburg i.O.	03 403 000	66	80	Esslingen	08 116 019	37

Table A.1.: List of eighty large German cities that were used in the empirical application, their unique identifier codes, and the number of unique locations per city, across all years.

A.2.2. Comparison of pharmacy and bake shop locations

	bake shops	pharmacies	difference
local population	0.467	0.495	0.028***
	(0.284)	(0.269)	(0.005)
elderly share	0.194	0.198	0.004***
	(0.066)	(0.062)	(0.001)
pedestrian zone	0.212	0.216	0.004
	(0.408)	(0.411)	(0.007)
mainroad	0.263	0.331	0.068***
	(0.440)	(0.471)	(0.008)
doctor	0.141	0.217	0.076***
	(0.348)	(0.412)	(0.006)
supermarket	0.298	0.231	-0.068***
	(0.458)	(0.421)	(0.007)
trainstation	0.061	0.045	-0.015***
	(0.239)	(0.208)	(0.004)
Observations	7,382	6,741	14,123

Table A.2.: Comparison of local spatial co-variates for pharmacy locations, and bake shops.

A.2.3. Choosing the strategic neighbourhood

The following table shows the estimates that were obtained using different sizes of the strategic neighbourhood in greater detail than table 2.8 does. Overall, the changes between the different specifications are largest when increasing k from one to two. For greater values of k, the changes become virtually nil.

	(1)	(2)	(3)	(4)	(5)	(6)
Local demand						
intercept	0.298***	0.295***	0.294***	0.294***	0.294***	0.294***
	(0.009)	*(0.009)*	*(0.009)*	*(0.009)*	*(0.009)*	*(0.009)*
post reform	-0.032***	-0.032***	-0.032***	-0.033***	-0.032***	-0.032***
	(0.007)	*(0.006)*	*(0.007)*	*(0.007)*	*(0.007)*	*(0.007)*
local population†	0.134***	0.162***	0.166***	0.167***	0.168***	0.167***
	(0.044)	*(0.026)*	*(0.026)*	*(0.026)*	*(0.026)*	*(0.026)*
elderly share	0.098***	0.096***	0.096***	0.097***	0.097***	0.097***
	(0.021)	*(0.025)*	*(0.026)*	*(0.026)*	*(0.026)*	*(0.025)*
trainstation	0.018**	0.018**	0.019**	0.019**	0.019**	0.019**
	(0.007)	*(0.007)*	*(0.007)*	*(0.007)*	*(0.007)*	*(0.008)*
pedestrian zone	0.015***	0.016***	0.016***	0.016***	0.016***	0.016***
	(0.004)	*(0.004)*	*(0.004)*	*(0.004)*	*(0.004)*	*(0.004)*
supermarket	0.032***	0.032***	0.032***	0.032***	0.032***	0.032***
	(0.004)	*(0.004)*	*(0.004)*	*(0.004)*	*(0.004)*	*(0.004)*
mainroad	-0.002	-0.002	-0.002	-0.002	-0.002	-0.002
	(0.004)	*(0.003)*	*(0.003)*	*(0.003)*	*(0.003)*	*(0.003)*
doctor	0.027***	0.027***	0.027***	0.027***	0.027***	0.027***
	(0.005)	*(0.004)*	*(0.004)*	*(0.004)*	*(0.004)*	*(0.004)*
Local competition						
δ_{pre}	0.002	0.002	0.002	0.002	0.002	0.002
	(0.003)	*(0.003)*	*(0.003)*	*(0.003)*	*(0.003)*	*(0.003)*
δ_{post}	0.009***	0.009***	0.009***	0.009***	0.009***	0.009***
	(0.002)	*(0.002)*	*(0.002)*	*(0.002)*	*(0.002)*	*(0.002)*
Entry costs						
θ^e_{pre}	4.908***	4.902***	4.901***	4.901***	4.901***	4.900***
	(0.049)	*(0.039)*	*(0.039)*	*(0.039)*	*(0.039)*	*(0.039)*
θ^e_{post}	4.335***	4.333***	4.332***	4.332***	4.332***	4.332***
	(0.027)	*(0.018)*	*(0.018)*	*(0.018)*	*(0.018)*	*(0.018)*
interaction radius [m]	900	900	900	900	900	900
strategic neighbours§	1	2	3	4	5	6
total entry locations	6,741	6,741	6,741	6,741	6,741	6,741
runtime [s]	730	587	1,170	2,186	4,067	15,673
log-likelihood	-14,780.3	-14,769.4	-14,768.1	-14,768.4	-14,767.7	-14,768.3

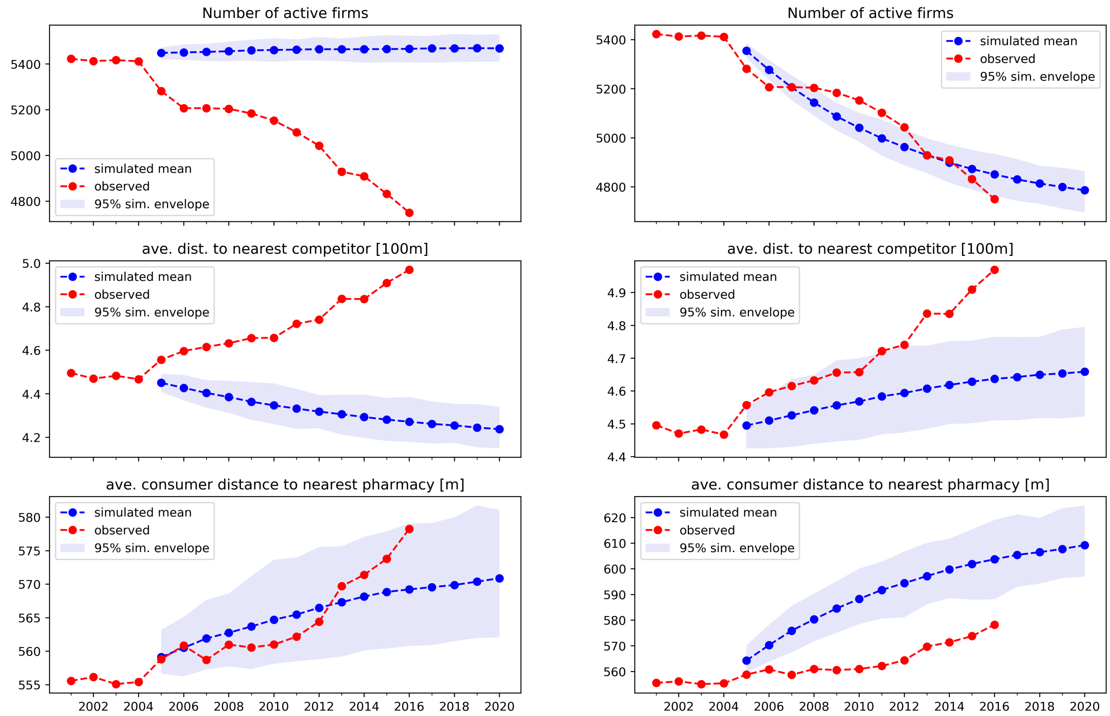
$^{*} < 0.1$; $^{**} < 0.05$; $^{***} < 0.01$

†Local residential population ÷ # local active stores

§ Including "self"

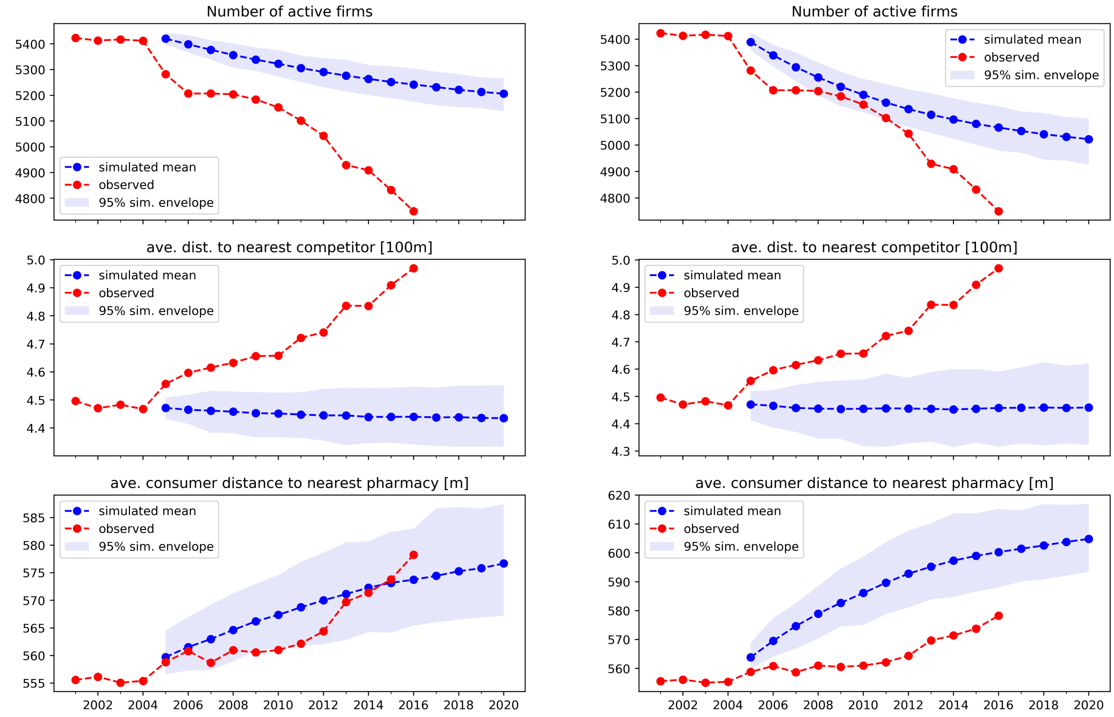
Table A.3.: Estimates of the spatial entry model with various sizes of the strategic neighbourhood (including self). N=6741 firms, T=16 time periods in 80 large German cities. Exit values are normalized to 1. Standard errors in parentheses, computed from estimated Hessian matrix.

A.2.4. Simulation results



(a) pre-reform parameters **(b)** post-reform parameters

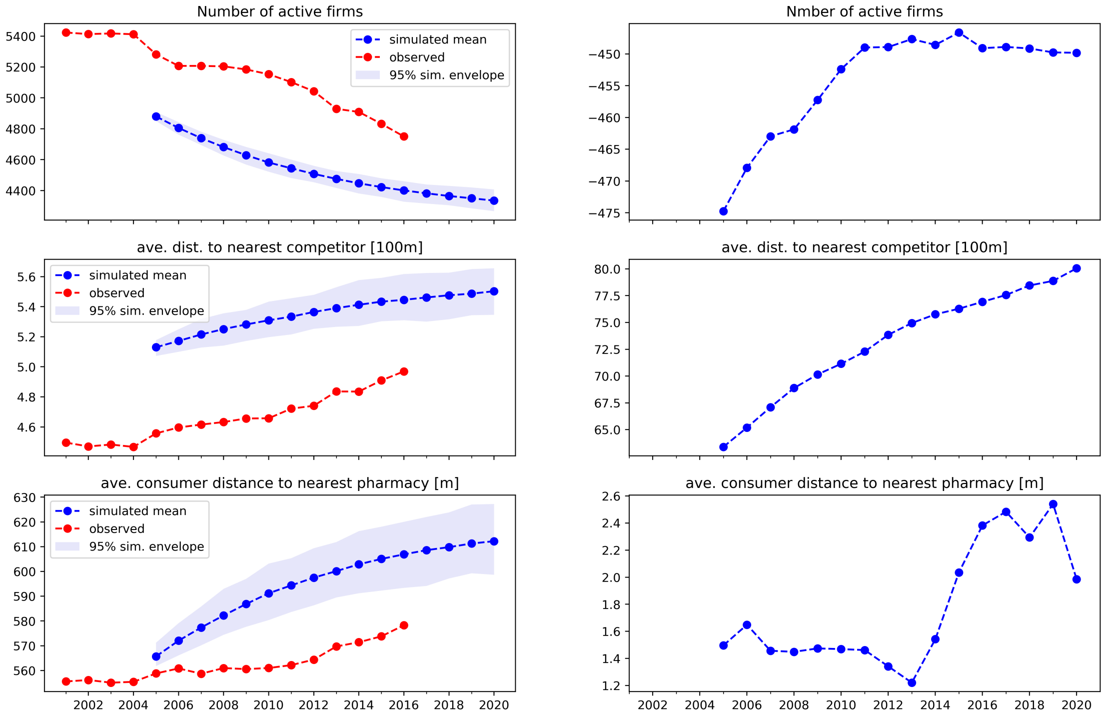
Figure A.1.: Simulation results using the parameters in table 2.9, starting from the observed market state in 2004 and simulating forward with two hundred parallel samples. The blue shaded areas represents the 95% simulation envelope.



(a) Scenario A

(b) Scenario B

Figure A.2.: Incremental effects of increased local competition, model (3) in table 2.9. See table 2.11 and section 2.7.4.



(a) Simulated counter-factual outcomes

(b) Difference to baseline (see figure A.2b)

Figure A.3.: The effects of a mandated 100m minimum distance between store locations, remaining parameters as in model (5) in table 2.9

A.2.5. Robustness checks

	$d = 800m$ (1)	$d = 1000m$ (2)	const. pop. (3)	more cov. (4)	bake shops (5)	random loc. (6)
Local demand						
intercept	0.295***	0.295***	0.293***	0.259***	0.293***	0.201***
	(0.009)	*(0.009)*	*(0.010)*	*(0.022)*	*(0.007)*	*(0.006)*
post reform	-0.033***	-0.031***	-0.032***	-0.033***	-0.036***	-0.042***
	(0.007)	*(0.007)*	*(0.007)*	*(0.007)*	*(0.005)*	*(0.004)*
local population†	0.136***	0.180***	0.172***	0.171***	0.066***	0.397***
	(0.021)	*(0.023)*	*(0.032)*	*(0.027)*	*(0.015)*	*(0.017)*
elderly share	0.102***	0.096***	0.098***	0.096***	0.075***	0.116***
	(0.027)	*(0.025)*	*(0.025)*	*(0.026)*	*(0.017)*	*(0.013)*
trainstation	0.018**	0.019**	0.019**	0.019**	0.005	0.022***
	(0.008)	*(0.008)*	*(0.008)*	*(0.008)*	*(0.005)*	*(0.004)*
pedestrian zone	0.017***	0.015***	0.016***	0.016***	0.008***	0.028***
	(0.006)	*(0.004)*	*(0.004)*	*(0.004)*	*(0.003)*	*(0.003)*
supermarket	0.033***	0.032***	0.032***	0.032***	0.011***	0.051***
	(0.004)	*(0.004)*	*(0.005)*	*(0.004)*	*(0.002)*	*(0.003)*
mainroad	-0.003	-0.002	-0.002	-0.003	0.004*	0.020***
	(0.003)	*(0.003)*	*(0.004)*	*(0.003)*	*(0.002)*	*(0.002)*
doctor	0.028***	0.027***	0.027***	0.028***	0.024***	0.042***
	(0.004)	*(0.004)*	*(0.004)*	*(0.004)*	*(0.003)*	*(0.003)*
vacancies				0.048		
				(0.066)		
sq. m. price				-0.014		
				(0.010)		
unemployment				-0.021		
				(0.061)		
income				0.019**		
				(0.008)		
income growth				-0.008		
				(0.090)		
in commuters				-0.005		
				(0.016)		
Local competition						
δ_{pre}	0.003	0.002	0.002	0.002	0.005	-0.021***
	(0.003)	*(0.003)*	*(0.003)*	*(0.003)*	*(0.003)*	*(0.003)*
δ_{post}	0.009***	0.009***	0.009***	0.009***	0.012***	-0.021***
	(0.002)	*(0.002)*	*(0.002)*	*(0.002)*	*(0.002)*	*(0.002)*
Entry costs						
θ^e_{pre}	4.903***	4.902***	4.900***	4.898***	5.643***	5.435***
	(0.032)	*(0.039)*	*(0.018)*	*(0.040)*	*(0.032)*	*(0.029)*
θ^e_{post}	4.334***	4.332***	4.332***	4.332***	5.106***	4.929***
	(0.018)	*(0.018)*	*(0.019)*	*(0.018)*	*(0.015)*	*(0.016)*
interaction radius	800	1000	900	900	900	900
strategic neighbours§	5	5	5	5	5	5
total entry locations	6,741	6,741	6,741	6,741	14,150	16,728
.. of which pharmacies	6,741	6,741	6,741	6,741	6,741	6,741
runtime [s]	4,800	3,678	4,930	5,188	14,604	13,894
log-likelihood	-14,774.5	-14,768.3	-14,766.7	-14,764.0	-17,374.4	-16,821.8

$^{*} < 0.1$; $^{**} < 0.05$; $^{***} < 0.01$

Table A.4.: Robustness checks. Exit values are normalized to 1. †local population is the residential population within 500 metres of the store's location, divided by the number of active competitors in the respective time period or future state. Except for columns three, the local residential population was scaled with municipality-level population growth rates. Standard errors in parentheses, computed from estimated Hessian matrix. § The number of strategic nearest neighbours always includes the decision maker.

Model specification

The spatial interaction radius was chosen in a non-strategic version of the model, with $k = 1$. To ensure that the choice of the interaction radius remained optimal after having chosen the size of the strategic neighbourhood, I re-estimated the model with $k = 5$ and for two different interaction radii, eight hundred and one thousand metres. These results are shown in columns one and two of table A.4. The table shows that the estimated coefficients are very close to the ones obtained with an interaction radius of nine hundred metres. Also, the log-likelihood in either case is smaller than the one obtained for an interaction radius of nine hundred metres; so the choice of the interaction radius remains optimal.

The population data are derived from the 2011 census and also, the spatial co-variates that are derived from OpenStreetmap reflect only one particular point in time, as the data were downloaded in 2016. Unobserved temporal variation of these co-variates may lead to biased estimates but unfortunately, there are no comparable data sources for earlier years. I controlled for unobserved population growth by scaling the spatial population distribution with observed municipality-level growth rates, but this obviously leaves the spatial variation unchanged. It would be desirable to obtain the spatial population distribution for at least one additional time period, so that a population growth rate could be computed at every given point in space. Currently, such data is unavailable, but a new census is planned for 2021[1] so that this could, in principle, be achieved in the future. For the current analysis, I assessed the robustness of the estimation results only with respect to the temporal aggregate variation of the population data, by re-estimating the model's parameters with the unscaled, constant, local population data. These results are shown in column three of figure A.4. That column shows that the estimated population coefficient (0.172) is not statistically different from the coefficient that is derived with the annually scaled population data (0.168, see table 2.9). Also, the other coefficients remain largely the same.

As a further robustness check, I added municipality level co-variates that are described in table 2.5 to the model. These results are shown in column four of table A.4. The column shows that all added coefficients are insignificant, except for municipality level income per capita. All other coefficients retain their sign, magnitude, and significance. Apparently, city-level heterogeneity is unlikely to have a large effect on the results.

Potential entry locations

The analysis in section 2.7.2 is based on the assumption that the set of potential entry locations can be approximated well by the set of locations where a pharmacy has been active at least once. In this section, I will assess the robustness of the main result with respect to this assumption by including additional entry locations. As explained in section 2.6.1, the additional entry locations are generated

[1] See `www.zensus2021.de`, accessed 03/18/2020

from two sources: first, I used the locations of bake shops in Germany and second, I generated a set of uniformly distributed random entry locations in all cities of my sample.

Column five in table A.4 shows the estimation results with bake shop locations, and column six displays the estimation results with additional random locations. First, consider column five. These results are quantitatively very similar to the main estimation results in table 2.9. The estimated coefficient for the local population density is substantially smaller than in the base line results, and the estimated entry costs are larger than in the base line case. The spatial interaction coefficients remain about the same, which is also true for the estimated coefficients for the local elderly share, and for the proximity to a physician, a supermarket, a trainstation, a pedestrian zone, or a main road. An explanation for this is that the inclusion of many locations that are not pharmacies has forced the model to attribute a greater weight to locations that are specifically important to pharmacies, as opposed to being just favourable to small retail stores in general. This increases the external validity of the model and therefore it can be a useful tool to conduct out-of-sample predictions and simulations. I repeated the model validation and counterfactual simulations described in section 2.7, using the additional set of entry locations. These results are shown in table A.5. The table shows that the simulated change from 2004 to 2016 does not match the actual change quite as well as does the base line model. The magnitudes of the partial competition effect are comparable to those derived in the base line model, and the effect of a minimum distance regulation is estimated to be larger than under the base line scenario.

The time needed to estimate the parameters increased more than proportionally with respect to the number of entry locations. This could be due to the fact that it takes longer for the Gauss-Seidel algorithm to reach the MPE. Still, the increase in computational time is very modest, and so the oblivious spatial equilibrium has proven to be a viable approach to estimate spatial dynamic entry models with a large number of agents.

Next, consider the results that were obtained by using the random set of dummy locations in column six. The estimated coefficients differ substantially from the base line results. In general, all estimated coefficients that govern local demand, except for the intercept and the post reform dummy, are larger than in the base line result. Moreover, the estimated interaction coefficients are now negative, and do not differ in the pre and post reform periods. This can be explained as follows. Pharmacy locations, and also the locations of bake shops, do exhibit a substantial amount of spatial clustering, as was shown in section 2.6.2. This is most likely an artefact of unobserved spatial heterogeneity, and not due to an inherent tendency of firms to cluster. On the other hand, the dummy locations were generated at random, and so do not exhibit any spatial clustering; and further, these dummy locations are never "active". Therefore, it is natural for the model to explain the higher likelihood that a "true" pharmacy location is active relative to a randomly generated dummy location by a positive point-to-point interaction, which is reflected in a negative interaction coefficient (recall that the interaction coefficient denotes by how much variable profits *decrease* due to the presence of a nearby competitor). Another interesting observation is the fact that the interaction coefficient is estimated to be the same in both the pre- and the post-periods. This is rather unintuitive, because the pattern of randomly generated entry locations

	Number of stores	distance (metres) pharmacy to nearest competitor	consumer to nearest pharmacy
observed outcomes			
2004	5,437	423	514
2016	4,770	469	539
Δ	**-667**	**+45**	**+25**
simulated outcomes			
pre-reform, 2016	5,403	403	536
post-reform, 2016	4,982	423	571
Δ	**−511**	**+20**	**+35**
competition effect			
(A) $t(0, \theta^e_{pre}, \delta_{post}) - t(0, \theta^e_{pre}, \delta_{pre})$	-226	+17	+4
(B) $t(1, \theta^e_{post}, \delta_{post}) - t(1, \theta^e_{post}, \delta_{post})$	-267	+20	+6
minimum distance regulation			
simulated change 2004-2016	−680	+81	+30

Table A.5.: Comparison of actual changes throughout the post-reform period; incremental effect of increased spatial competition; and the effect of a minimum distance regulation.

does not change over time, whereas the pattern of pharmacy locations does (as is reflected in the larger interaction coefficients in column five, and in table 2.9). The most likely explanation is an inaccuracy of the estimation procedure which prematurely stopped the algorithm before having reached the true global optimum.

How should these two additional results be interpreted, and compared to each other? Many of the randomly generated dummy locations would fall in residential zones, or even in uninhabited areas where it is either not permitted, or not possible to open a store. Thus, these locations can hardly be considered to be "potential entry locations". On the other hand, many bakery shops have a size that is similar to that of a pharmacy, and so their locations can arguably be a potential entry location. At the same time, the underlying behavioural assumption of the dynamic entry model is that each potential entrant plays an entry game (in entry probabilities, essentially) with each of its $k - 1$ strategic nearest neighbours. The model therefore presupposes that at every entry locations sits a potential competitor which could conceivably become an active firm. But as discussed above, this is not true for many of the randomly generated locations and so the estimation results obtained with random entry locations may not accurately reflect the true parameters, and should be interpreted cautiously.

3. Oblivious estimates of spatial dynamic entry games with many firms: Monte Carlo evidence

3.1. Introduction

Dynamic entry games are complex phenomena because decision makers must speculate about what the future will bring, conditional on their own actions and on those of their competitors. The necessity to consider a large number of possible future outcomes in these games means that their theoretical treatment, and their empirical analysis alike, is challenging. However, the difficulties in the theoretical and in the empirical literature are of a different nature altogether. Whereas the theorist is very much concerned with formally establishing that a certain game, under certain conditions, has desired properties and possesses an equilibrium, the large number of possible future states poses no substantial challenge as these are rarely spelled out explicitly. On the contrary, the empiricist who ventures to build a structural model of dynamic decision making will have to quantify the value of being in one state of the game, or another. He will thus almost certainly face the situation where his own computer is not able to store, let alone compute, all the different states of the dynamic model which looked rather innocent in the beginning. The reason for this is commonly called the curse of dimensionality, and it effectively prevents a direct computation of the equilibrium in all but the simplest dynamic games. A number of approaches have been developed to overcome this problem, and some of these have been shown to perform rather well. The problem is that none of these solutions lends itself to a spatial context, because they all assume some form of symmetry, or require a non-parametric first stage.

In chapter 2, I presented a feasible method to add a spatial dimension to an empirical dynamic entry model at an acceptable computational cost, and applied it to the German pharmacy market. This chapter examines the properties of that estimator in greater detail by means of a Monte Carlo study. In particular, I will provide evidence that the estimator is consistent and asymptotically normal if the likelihood is specified correctly. I will also examine its properties in cases where the researcher does not know certain aspects of the dynamic decision making process, and may thus have to approximate the likelihood. Most importantly, it follows from my results that it is possible to obtain good estimates of a subset of the model parameters even if other parameters are unknown, or are known only approximately.

3.2. Related literature

This paper relates to various strands of literature. First, it is related to the theoretical study of dynamic and spatial entry and exit processes. Second, I relate to the empirical treatment of dynamic entry games, and to the methods that have been proposed therein to reduce the curse of dimensionality. More specifically, this paper is related to small literature on spatial dynamic entry games.

In the theoretical literature, there has been a dichotomy between static models of location choice, and dynamic models of entry. The literature on location choices started with the seminal work of Hotelling (1929) who found that free location choice leads to an inefficient amount of spatial clustering, which was termed as the principle of minimum differentiation. This finding was re-examined by d'Aspremont et al. (1979) who found on the opposite that price competition may actually lead to maximum spatial differentiation. They also showed that a game in which firms choose their locations first, and subsequently compete on prices, may not possess a pure-strategy subgame-perfect Nash equilibrium unless the consumer's travel costs increase quadratically in distance. From there onwards, quadratic travel costs have become the de facto standard in the theoretical literature on location-then-price competition. Most theoretical work is constrained to location choices on a line, or on a circle. Exceptions are e.g. Economides (1986), or Irmen and Thisse (1998), but the insights are very stylized at best. The most fundamental insight that can be gained from the theoretical literature on location choices is that firms trade-off the benefits of reaching more customers in a central location against the added value by increasing their market power in a more remote location. These two motives form the basis for building models of spatial entry where the profitability of a particular location depends on local demand characteristics, and on the distance to any active competitors.

On the other hand, dynamic entry games are often modelled without explicit regard to a spatial aspect. Seminal articles include Ghemawat and Nalebuff (1985) who model dynamic exit in shrinking markets, and Hopenhayn (1992) who studies equilibrium and the steady state dynamics in a model with a continuum of firms that produce a homogeneous product, subject to idiosyncratic productivity shocks. Ericson and Pakes (1995) have presented a model to describe the dynamics in markets with idiosyncratic random shocks to the profitability of an investment, and with increasing external competitive pressure. The equilibrium concept that is most commonly used in dynamic games is the Markov Perfect equilibrium (MPE), due to Maskin and Tirole (1988), where decision makers condition their choices only on the immediate past so that the dynamics give rise to a Markov Process. Doraszelski and Satterthwaite (2010) have contributed significantly to simplifying the treatment of dynamic entry games. They showed that the addition of small perturbations to entry and exit costs leads to an equilibrium in pure strategies, that can be computed relatively easily.

Dynamic games are notoriously hard to tackle numerically, because the value of choosing one action, or another, depends on its implications for the future course of affairs through their direct effect on outcomes, and through the indirect effect via the competitors' actions. Dynamic programming techniques are commonly used to attach a value to a particular action, and to compute the MPE. These techniques require the evaluation of a value function for every possible state of affairs, and are usually solved iteratively by a contraction mapping scheme, because analytical closed form solutions are seldomly available. Because dynamic games often possess state spaces that are too large to fit into a physical computer, researchers have considered various options to reduce the computational burden of finding an equilibrium in these games. As Pakes and McGuire (2001) pointed out, this computational burden is primarily determined by three factors:

1. the number of different states for which a value function must be computed,
2. the computational complexity of each single operation, and
3. the number of iterations until convergence is reached.

Most approaches rely on reducing the size of the state space (1.) but is is also important to keep points (2.) and (3.) in mind in any practical application. For instance, Pakes and McGuire (1994) developed an algorithm that relies on the firms' symmetry to compute the MPE of the Ericsson and Pakes entry and investment model. The same authors developed a stochastic algorithm to compute the MPE in the same model that works when the dynamic game eventually wanders into a small absorbing subset of the state space, so that many states can be ignored because they are unlikely to occur, which greatly reduces the computational demands (Pakes and McGuire, 2001). Weintraub et al. (2008, 2010) have introduce the important concept of an "oblivious equilibrium" (OE) in which agents consider the aggregate state to be constant, because thy face a large number of competitors and so their actions cancel out in equilibrium. That concept can, without further modifications, only be applied in situations where the market is in a stationary equilibrium.

The econometric treatment of dynamic discrete games started with the seminal article of John Rust (1987) who estimated a single agent optimal stopping problem by means of a nested fixed point likelihood. These estimators require that the equilibrium strategies of all economic agents be computed for every candidate value in the numerical parameter search, and thus are relatively inefficient from a computational standpoint. Their advantage is a close link to theory, and their flexibility. To ease the computational burden, two-step estimators (Hotz and Miller, 1993) and K-step estimators were developed (Aguirregabiria and Mira, 2002). These estimators build on the optimal strategy response of dynamic decision makers to a consistent non-parametric assessment of choice and transition probabilities, rather than solving the optimal programming problem explicitly for every candidate parameter value. In the case of the K-step estimator, the initial non-parametric estimate of the players' best response probabilities is updated iteratively, which gives rise to what is called a nested pseudo likelihood estimator. Aguirregabiria and Mira (2007) extended this approach to dynamic games with more than one player. A similar approach was developed by Pakes et al. (2007) who, instead of relying on non-parametric estimates of choice probabilities, used non-parametric estimates of players' continuation values to elicit their optimal strategy responses. Based on Monte Carlo simulations, Pakes, Ostrovsky, and Berry argue that their estimator may outperform several other estimation approaches, including the nested fixed point and two-step approaches. Berry and Reiss (2007) and Aguirregabiria and Mira (2010) provide a comprehensive treatment of the empirical entry literature.

The synthesis of the empirical dynamic entry literature with the spatial entry literature is still not very well developed. A first pass to an empirical analysis of spatial entry models was made by Seim (2006) who estimated a static spatial entry game among retail video rental stores. She reduces the state space by using a static setting with ex-ante identical potential entrants and asymmetric information. Her estimation approach is a nested fixed point approach. Because players are identical up to the realization of a private-value location-specific idiosyncratic profitability shock, and because all players are entrants, their strategies are symmetric, and are conditioned only on the realization of their profitability shock. This leads to fairly small computational burden of finding the fixed point of players' strategies. However, this method does not extend easily to a dynamic entry game because (i) entrants and incumbents have different strategies; (ii) players condition their strategy on the observed state and so the strategy space quickly explodes in the number of players; and (iii) points one and two imply that players' strategies are no longer symmetric. Aguirregabiria and Vicentini (2016) were the first to develop a framework of a spatial dynamic entry game. In their model, a multistore retailer decides on where to locate its branches in an oligopolistic market. Unfortunately, they do not apply their concept to observational data. They use value function interpolation and a restriction of the entrants' action spaces to reduce the dimensionality of the problem, and to obtain a computationally feasible algorithm to compute the equilibrium. That approach has two main drawbacks. First, the authors do not specify how the support points for the value function interpolation are chosen. But the size of the state space grows exponentially in the number of players and or entry locations. Therefore, any computationally feasible subset of the state space can only capture a minuscule proportion of the entire state space, even for dynamic games with a moderate number of potential entry locations. Thus, the interpolation becomes very sparse, and also probably quite inaccurate. Second, the desire to compute the "true" MPE on the full state space may be leading nowhere, because it seems unreasonable to assume that any decision maker could solve a problem, the very size of which would overwhelm any currently existing computer. Therefore, to analyse spatial dynamic entry games it seems necessary to reduce the state space in a way that reflects the manner in which real-work agents think about the future, and I have proposed a possible method to do so in chapter 2.

In chapter 2 I showed how a spatial dynamic entry game in the spirit of Aguirregabiria and Vicentini (2016) can be applied to the OE concept of Weintraub et al. (2008). The result is a "spatial oblivious equilibrium" which can be used to estimate structural parameters of the model by means of a nested fixed point estimator. The degree of simplification is governed by a single intuitive parameter. The aim and the contribution of this paper is to determine the statistical properties of that estimator under various conditions. First, I provide evidence that the estimator is consistent, and root-n asymptotically normal – if the likelihood and all the hyper-parameters are specified correctly. This is to be expected, and so the confirmation of these properties merely serves as a proof that the method was implemented correctly. Second, I show that some of the model parameters can still be estimated if the model's hyper-parameters are unknown. Finally, I investigate how the presence of unobserved potential entry locations affects the estimation results.

3.3. Methodology

The purpose of this paper is to learn more about the properties of the maximum likelihood estimator outlined above. This is achieved by means of a Monte Carlo study, the basic workflow of which is as follows: a data-generating process is set up with certain “true” parameters, and generates market observations. Then, a maximum likelihood estimator is used to estimate these parameters, and the deviance of the estimated parameters relative to the true parameters is measured. This is repeated a many times, so that the empirical distribution of parameter estimates approximates their theoretical distribution, which is the object of interest. The data generating process is based on a simplified version of the dynamic entry model that was developed in chapter 2. Also, I discuss how the model can be used to estimate parameters via nested fixed point approach, and which data are necessary for identification. Finally, I describe the data generating process and some technicalities of the Monte Carlo study in some detail and I discuss some general theoretical properties of maximum likelihood estimators.

3.3.1. Model and estimation

The basic framework consists of N potential entrants, each one of which is indexed by j and equipped with a fixed location $x_j \in X \subset \mathbb{R}^2$. The market area X is a subset of the Euclidean space, and so it is equipped with a distance norm. There is an infinite horizon of future time periods $t = 1, \dots, \infty$ and in each period, an entrant can be either active, or inactive. The status of firm j in time t is denoted by $a_{jt} \in \{0, 1\}$ and the aggregate state of all firms is $\mathbf{a}_t = (a_{t1}, \dots, a_{tN}) \in \mathbb{A} = \{0, 1\}^N$. $\mathbb{A}$ denotes the state space of the game, and it encompasses 2^N distinct states. Without any further modifications, the state space is too large to be of practical significance. Period returns of an active firm in time period t and state $\mathbf{a}_t$ are given by

$$\pi_j(\mathbf{a}_t) = \alpha \left(1 - \delta \mathcal{N}_j^d(\mathbf{a}_t)\right) \tag{3.1}$$

where $\mathcal{N}_j^d(\mathbf{a}_t)$ denotes the number of active competitors around firm j's location, given the state vector $\mathbf{a}_t$. Inactive firms receive a payoff of zero. Thus, δ governs the relative decrease in period returns for every additional active competitor within a distance d, and α is a monopoly profit. This profit function should be thought of as a reduced form of a more elaborate profit function, and it is chosen for clarity here. The equilibrium concept that follows does not rely on a simple functional form, and could indeed accommodate any kind of profit function.

The timing is as follows. At the beginning of every period, each active firm receives its current per-period return according to equation (3.1). Also, every firm learns the realization of a private information, idiosyncratic, and normally distributed random variable ξ_{jt} with distribution function Φ. Upon learning this value, all potential entrants decide whether they choose to enter the market by paying an entry fee $\theta^e + \xi_{jt}$. Similarly, all incumbent firms decide whether they continue their business, or exit the market and receive a pay-off value ξ_{jt}. After these decisions have been made, all entry cost and exit fee payments are carried out, and the state evolves to the next period accordingly. The future is discounted by a common discount factor β. This sequence continues ad infinitum, and multiple re-entries are a possibility in this framework.

Following Doraszelski and Satterthwaite (2010), the introduction of a private information transition cost shock greatly simplifies the analysis because firms can compute the expected value of their actions by integrating out their competitors' cost shocks. Then, they choose an optimal action based on the realization of their own cost shock. From the point of view of their competitors, their actions seem stochastic because they depend on the realization of ξ_{jt} and so they can equivalently be expressed by their conditional choice probabilities (CCPs) $\mathbf{q}$, where $q_j(\mathbf{a}_t)$ is the probability that firm j's cost shock realization is such that it chooses to be active in period $t+1$, given the current state and the CCPs of its competitors $\mathbf{q}_{-jt}$. Then, an MPE of the dynamic entry game is a set of CCPs for each firm that constitute mutual best responses. These equilibrium CCPs can be found as follows: Let $\overline{V}_j(\mathbf{a}_t)$ denote firm j's expected value of being in state $\mathbf{a}_t$ – that is, before the realization of ξ_{jt} is known. Following the notation of Aguirregabiria and Vicentini (2016), the ex-ante expected value of being active next period for that firm j is

$$\bar{\nu}_j^1(\mathbf{a}_t) = \pi_j(\mathbf{a}_t) - (1 - a_{jt})\theta^e + \mathbb{E}\left[\overline{V}_j(\mathbf{a}_{t+1}) | a_{jt+1} = 1, \mathbf{a}_{-jt}\right] \tag{3.2}$$

and similarly, the expected value of being inactive is given by

$$\bar{\nu}_j^0(\mathbf{a}_t) = \pi_j(\mathbf{a}_t) + \mathbb{E}\left[\overline{V}_j(\mathbf{a}_{t+1}) | a_{jt+1} = 0, \mathbf{a}_{-jt}\right] \tag{3.3}$$

The term $(1 - a_{jt})\theta^e$ reflects the fact that a new entrant must pay the entry fee, but an incumbent does not. The competitors' CCPs induce a distribution over the entire future state space $\mathbb{A}$, based on which the expectation operator returns the expected future value. Note that the ex-ante expected values $\bar{\nu}$ do not include the random shock ξ_{jt} which must be added to the entry costs for an entrant, and to the exit value for an incumbent. Thus, upon learning the realization of the transition cost cost ξ_{jt}, the ex-interim expected value of choosing to be active next period is $\nu_j^1 = \bar{\nu}_j^1 - (1 - a_{jt})\xi_{jt}$, and the corresponding expected value of being inactive next period is $\nu_j^0 = \bar{\nu}_j^0 + a_{jt}\xi_{jt}$. Then, firm j decides to be active in the next period if, and only if

$$\bar{\nu}_j^1(\mathbf{a}_t) - (1 - a_{jt})\xi_{jt} \geq \bar{\nu}_j^0(\mathbf{a}_t) + a_{jt}\xi_{jt}$$

the probability of which to happen is

$$q_j(\mathbf{a}_t) = \Phi\left(\bar{\nu}_j^1(\mathbf{a}_t) - \bar{\nu}_j^0(\mathbf{a}_t)\right). \tag{3.4}$$

With ξ_{jt} being distributed normally, the ex ante expected value function can be expressed as[1]

$$\overline{V}_j(\mathbf{a}_t) = \phi\left(\bar{\nu}_j^1(\mathbf{a}_t) - \bar{\nu}_j^0(\mathbf{a}_t)\right) + q_j(\mathbf{a}_t)\bar{\nu}_j^1(\mathbf{a}_t) + (1 - q_j(\mathbf{a}_t))\bar{\nu}_j^0(\mathbf{a}_t). \tag{3.5}$$

where ϕ is the density of the standard normal distribution.[2] Equations (3.2), (3.3), (3.4), and (3.5) define a system of highly non-linear equations in CCPs that can be solved for equilibrium CCPs $\mathbf{q}^*$ by means of a fixed point algorithm (Aguirregabiria and Vicentini, 2016).[3]

Note that the value functions and CCPs are defined on, and the expectation operator in equations (3.2) and (3.3) integrates over the entire state space $\mathbb{A}$, the magnitude of which is 2^N. Therefore, this equilibrium concept suffers from the curse of dimensionality and cannot be solved explicitly due to numerical constraints when the number of firms is large.[4] At the same time, the researcher typically only observes a very small set of states, and so one possible solution is to compute the MPE only for states that are "close" to the observed ones. This is closely related to the idea of the oblivious equilibrium due to Weintraub et al. (2008), where firms take the aggregate state as given, and make their choices based solely on their own individual state. However, due to the spatial nature of this game, it is not reasonable to draw a clear distinction between the aggregate state and the individual state. Instead, a subset of the unbounded state space must be selected to achieve a distinction between states which the decision maker should care about, and other states which fall into oblivion. A structured approach to select such a subset of states was presented in chapter 2 and I repeat the main idea below.

A *spatial oblivious equilibrium* of order k, denoted by SOE(k),[5] is a set of CCPs $\tilde{\mathbf{q}}^*$ such that $\tilde{q}_j^*(\tilde{\mathbf{a}}_j)$ is the best-response probability with which firm j chooses to be active in the next period, conditional on the game being in state $\tilde{\mathbf{a}}_j$, and all other firms playing according to $\tilde{\mathbf{q}}^*_{-j}$. Importantly, these states $\tilde{\mathbf{a}}_j$ are contained in a firm-specific subset of the state space, called the *oblivious state space* and denoted by $\tilde{\mathbb{A}}_j \subseteq \mathbb{A}$. Given a realized state $\hat{\mathbf{a}}$, this oblivious state space for firm j is constructed under the assumption that all other firms remain in their currently observed state $\hat{a}_j$, except for the k firms that are closest to firm j (including firm j itself). This set of firms is called the *strategic neighbourhood*. Thus, an SOE is intricately linked to, and always contains, an observed market state, which will be important for constructing a likelihood function below. The equilibrium concept basically remains as in the MPE, but since all functions are only defined for the oblivious state spaces of the respective firms, the state spaces of two firms at the same time period may differ, because these two firms may have different strategic neighbourhoods. That is why some adjustments have to be made in order to still allow for a coherent formation of the firms' expectations, the details of which can be found in chapter 2. If the size of the strategic neighbourhood is smaller than the total number of firms ($k < N$), then the oblivious state space is a strict subset of the unrestricted state space and the dimension of the state

[1]See chapter 2.

[2]See chapter 2 for a detailed derivation.

[3]For instance, a Gauss-Seidel algorithm can be used, see also chapter 2 and the references therein.

[4]Depending on the type of the system, one could possible store the required matrices for problems up to $N = 30$, but the time to compute the equilibrium would probably be prohibitive much earlier.

[5]The reference to k is often dropped where this causes no confusion.

space reduces from 2^N to 2^k. As a result, an SOE can often be computed in markets where a full MPE would be impossible to compute due to the curse of dimensionality.

Besides being a convenient computational device, the SOE also serves as a behavioural model that actually describes how individuals make decisions in an overly complex game theoretical setting. The parameter k governs how sophisticated these individuals are in terms of dynamic strategic decision making. Two polar cases can be distinguished: first the completely spatially myopic case ($k = 1$) where firms assume that no other firm changes their status, and second, the full MPE with $k = N$. However, it is an open question whether the $SOE(k)$ converges to the MPE if either k or N grows large, and if so, how.

If one assumes that the SOE is a good model for how firms make their dynamic entry decisions, then its equilibrium CCPs can easily be used to compute the conditional likelihood of transitioning from one observed state to another one. Given a set of observed market states $\{\hat{\mathbf{a}}_t\}_{t=1}^T$, this likelihood can then be used to estimate the structural parameters of the entry model. In this simple version of the entry model, there are only three parameters, α, δ, and θ^e, which are subsumed under the vector $\boldsymbol{\gamma}$. I call $\boldsymbol{\gamma}$ *regular parameters.* For convenience, I define firm j's equilibrium CCP in the observed state at time period t as $\tilde{q}_{jt}^* \equiv \tilde{q}_j^*(\hat{\mathbf{a}}_t)$. Note that this probability is always well-defined, because the currently observed state forms the basis, and is always part of, the current oblivious state space. Then, the conditional log-likelihood is given by

$$ll(\boldsymbol{\gamma}) = \sum_{t=1}^{T-1} \sum_{j=1}^{N} \hat{a}_{jt+1} \log \tilde{q}_{jt}^* + (1 - \hat{a}_{jt+1}) \log \left(1 - \tilde{q}_{jt}^*\right) \tag{3.6}$$

This expression is called a nested fixed point likelihood because it is based on a nested fixed point algorithm to determine the transition probabilities (Rust, 1987). It is important to point out that this expression is constructed from a sequence of SOEs, because a new SOE must be computed for every time period. Each SOE will be slightly different from its predecessor, because as the observed state changes, so does the oblivious states space. The maximum likelihood estimator (MLE) for $\boldsymbol{\gamma}$ is the parameter which maximizes the above expression. In many empirical applications, the researcher has data from distinct markets, which have no strategic connection to each other. In that case, the aggregate likelihood is simply the sum of the individual market-level likelihoods.

As outlined in chapter 2 regular parameters are identified follows. First, all parameters are only identified up to the scale of the error variance, as in most variants of empirical discrete choice models. The entry cost parameter governs how much turnover there is in the market, that is, whether firms change their status frequently. Technically speaking, higher entry costs introduce a hysteresis, i.e. a tendency to cling to the status quo. Thus, based on an observed set of market states, the entry cost parameter can be identified by matching the model-implied turnover rates to the empirically observed ones. The constant payoff parameter α governs how profitable firms are, and so it determines how many firms are active, on average. This parameter is identified by bringing the average number of observed active firms in line with its model-implied counter-part. Finally, the local competition parameter δ determines how strongly firms' profits are affected by the presence of an active nearby competitor. It is identified by the differential entry and exit rates of firms with differing numbers of active competitors. Point identification may not be possible in certain cases. First, it could be that the SOE is non-unique in which case the likelihood would not be well-defined. In chapter 2 I have provided some evidence that this is rather unlikely if the local interaction parameter δ is not too large. Another problematic case arises when there is no observed entry or exit, because the maximum likelihood for this to happen is attained in the limiting case where θ^e approaches infinity, and so the entry costs are not (point) identified. Also, it may be the case that no single firm is ever active, or that all firms are active at all times. This would also lead to α being non-identified (or rather, being identified at plus and minus infinity, respectively). At last, if all firms had exactly the same number of neighbours within the interaction distance d, or if no firm had any adjacent neighbour within that radius, then the identification of δ would be impossible because no differential entry and exit patterns between firms with many, and those with few competitors, could be observed and matched to the model-implied entry and exit rates. Therefore, identification requires that these identifying features[6] be present in the data, and that the data does not exhibit the aforementioned pathological patterns.

Besides the regular parameters, two additional parameters are important in this model: the distance term d, which determines the radius within which firms exerts a negative effect on each other, and the size of the strategic neighbourhood k, which determines the level of strategic sophistication. I call d and k *hyper-parameters*, and subsume them into the vector $\boldsymbol{\psi} = (d, k)$. Because the likelihood function is not continuous in either of these hyper-parameters and since k can take only positive integer values, it is not possible to include the continuous search for these parameters in the MLE for $\boldsymbol{\gamma}$, which typically makes use of routines to find the maximum of a continuous function. Two alternative approaches to handle this are possible. On the one hand, one could specify some ad hoc hyper-parameters which seem reasonable, possibly supported by suggestive evidence. Alternatively, it is possible to conduct a grid search for these hyper-parameters. In the next section, I will assess the robustness of these approaches.

3.3.2. Theoretical properties of ML estimators

The nested fixed point algorithm belongs to the class of maximum likelihood estimators (MLE) which have attractive theoretical properties in that they are consistent, asymptotically normal, and among those estimators that share these properties, they are also efficient – if they are specified correctly (Cameron and Trivedi, 2005, chapter 5.6). In what follows, I will briefly discuss these concepts, and explore whether their prerequisites apply to my model at hand.

The following conditions, taken from Cameron and Trivedi (2005, p.142), are sufficient to ensure that a generic MLE for some parameter θ is both consistent and asymptotically normal:

[6] I call them features, but they could also be called moments.

Proposition 3 (Distribution of the MLE)**.** *(Cameron and Trivedi, 2005, proposition 5.5) Make the following assumptions:*

1. *The data generating process is the conditional density $f(y_i|\mathbf{x}_i, \theta_0)$ used to define the likelihood function.*
2. *The density function $f(\cdot)$ satisfies $f(y, \theta^{(1)}) = f(y, \theta^{(2)})$ iff $\theta^{(1)} = \theta^{(2)}$.*
3. *The matrix*
$$\mathbf{A}_0 = \text{plim}\frac{1}{n}\left.\frac{\partial^2 \mathcal{L}_n(\theta)}{\partial\theta\partial\theta'}\right|_{\theta_0}$$
exists and is finite non-singular.
4. *The order of differentiation and integration of the log-likelihood can be reversed.*

Then, the maximum likelihood estimator $\hat{\theta}$ is consistent for θ_0, and

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N\left(0, -\mathbf{A}_0^{-1}\right).$$

The notation is as follows: y_i is a dependent variable, $\mathbf{x}_i$ is a set of independent variables, and θ_0 is the true parameter value. The dynamic entry model fits into this framework by setting $y_i = a_{it+1}$ and $\mathbf{x}_i = \mathbf{a}_t$. The parameter to be estimated is $\theta = \boldsymbol{\gamma}$, and the hyper-parameter $\boldsymbol{\psi}$ is required to be known. Note that since all entry and exit decisions are carried out simultaneously, and because the transition cost shocks are independent and identically distributed, the observations of each firm's future status, conditional on the previous market state, are also independent. Essentially, the log-likelihood specified in equation 3.6 is that of a binary probit model, albeit with a rather complicated and highly non-linear link function.

The first condition requires that the likelihood be specified correctly, so that it is based on the same conditional density function as the true data generating process. When applied to the current setting, this condition simply demands that the model is correct, and the hyper-parameters be known precisely. However, even if the likelihood is misspecified it may still be possible to obtain consistent estimates of the "pseudo-true" value of $\boldsymbol{\gamma}$ via the quasi maximum likelihood approach (Cameron and Trivedi, 2005, p.147). For example, this could be the case if the distribution of transition cost shocks ξ_{jt} does not follow the standard normal distribution, contrary to what is assumed. Therefore, it should be kept in mind that the proposed MLE can only identify parameters up to the distribution of the structural transition cost shocks, and that any departure from the assumed distribution results in a different, pseudo-true parameter estimates. A violation of this assumption would arise if the transition cost shocks were not iid, or had some observable component in which case the equilibrium concept would break down. The second condition demands that the parameter $\boldsymbol{\gamma}$ be identifiable. This was discussed in some detail above, and is explored in greater depth in chapter 2. In particular, this condition could fail if the local interaction parameter δ is large because this may give rise to multiple equilibria. The third condition is necessary in order to derive the asymptotic distribution, and it states that the likelihood function be twice differentiable in θ, and that its Hessian matrix be non-singular. Because the likelihood function is a continuous function and twice differentiable in the model's CCPs, this condition is satisfied if the CCPs of the underlying SOE are differentiable twice in the model parameters $\boldsymbol{\gamma}$. That, however, is not easily established because no closed form solutions of the CCPs exist. The fourth condition is a technical one, and according to (Wooldridge, 2010, p.395) it may fail "if the conditional support of y_i depends on the parameters θ_0". This is however not the case in the proposed model, because the transition cost shocks have full support over $\mathbb{R}$ and so a transition to *any* market state is possible, independent of the current market state or on the parameter vector $\boldsymbol{\gamma}$. So it seems probable that this condition is satisfied in the proposed model.

In summary, it seems hard to establish from first principles that the MLE for $\boldsymbol{\gamma}$ is consistent *and* asymptotically normal. However, the above conditions are rather strong, and not necessary if one is merely concerned with consistency. The following proposition, replicated from Wooldridge (2010), states milder sufficient conditions for consistency alone:

Proposition 4 (Consistency of the MLE). *(Wooldridge, 2010, based on Theorem 13.1) Let* $\{(\mathbf{x}_i, \mathbf{y}_i) : i = 1, 2, \ldots\}$ *be a random sample with* $\mathbf{x}_i \in \mathcal{X} \subset \mathbb{R}^K, \mathbf{y}_i \in \mathcal{Y} \in \mathbb{R}^P$. *Let* $\Theta \subset \mathbb{R}^P$ *be the parameter set and denote the parametric model of the conditional density as* $\{f(\cdot|\mathbf{x};\theta) : \mathbf{x} \in \mathcal{X}, \theta \in \Theta\}$. *Define* $ll_i(\theta) \equiv ll(\mathbf{y}_i, \mathbf{x}_i, \theta) = f(\mathbf{y}_i|\mathbf{x};\theta)$. *Assume that*

1. $f(\cdot|\mathbf{x};\theta)$ *is a true density with respect to the measure for all* $\mathbf{x}$ *and* θ *so that it integrates to one;*
2. *for some* $\theta_0 \in \Theta$, $p_0(\cdot|\mathbf{x}) = f(\cdot|\mathbf{x};\theta_0)$, *all* $\mathbf{x} \in \mathcal{X}$, *and* θ_0 *is the* unique *maximizer of* $E[ll_i(\theta)]$;
3. Θ *is a compact set;*
4. *for each* $\theta \in \Theta$, $ll(\cdot, \theta)$ *is a Borel measurable function on* $\mathcal{Y} \times \mathcal{X}$;
5. *for each* $(\mathbf{y}, \mathbf{x}) \in \mathcal{Y} \times \mathcal{X}$, $ll(\mathbf{y}, \mathbf{x}, \cdot)$ *is a continuous function on* Θ*; and*
6. $|ll(\mathbf{w}, \theta)| \leq b(\mathbf{w})$, *all* $\theta \in \Theta$, *and* $E[b(\mathbf{w})] < \infty$.

Then the MLE $\hat{\theta} = \arg\max \frac{1}{n}\sum_{i=1}^{n} \log f(\mathbf{y}_i|\mathbf{x}_i;\theta)$ *exists and* $\operatorname{plim} \hat{\theta} = \theta_0$.

The likelihood estimator for the dynamic entry model in (3.6) can be put in this framework as follows. Consider first the case where only one transition from state $\mathbf{a}_t$ to a future state $\mathbf{a}_{t+1}$ is observed. Let $\mathcal{X} = \mathcal{Y} \equiv \mathbb{A}$ and define $\mathbf{y}_i \equiv a_{it+1}$, and $\mathbf{x}_i \equiv \mathbf{a}_t$. The model parameters are $\theta \equiv \boldsymbol{\gamma}, \theta \in \Theta \subset \mathbb{R}^3$. Then, the conditional density f is defined as the binomial density over the future state space,

$$f(\mathbf{y}_i|\mathbf{x}_i;\theta) = \prod_{i=1}^{N} \tilde{q}_{it}^{*a_{it+1}}(1 - \tilde{q}_{it}^*)^{1-a_{it+1}}$$

where $\tilde{q}_{it}^*$ is defined as in equation (3.6). Since $\tilde{q}_{it}^* \in [0, 1]$ for all $\mathbf{a}_t$ and $\boldsymbol{\gamma}$, this density clearly integrates to one and so the first condition is satisfied. The second condition requires that this conditional density is equal to the true density of the DGP, here denoted p_0, at some value θ_0. This requirement is therefore weaker than the second condition in proposition (3), because it only refers to one particular point in the parameter space. Further, it is required that this particular value θ_0 maximizes the expected likelihood, as defined in equation (3.6). Similarly, this second assumption cannot be tested, as the likelihood is derived directly from an underlying theoretical model, and so the conditional density of the DGP is assumed to be identical to the one used to formulate the likelihood. At the very least, this requires that the hyper-parameters are known, or can be estimated, with sufficient accuracy. Conditions three, four, and six are technical conditions that do not need to be checked in practice (Wooldridge, 2010, p.391). Thus, the remaining key assumption is that the likelihood function be continuous in θ. This is considerably weaker than what is required in proposition (3) for asymptotic normality, and for this condition to hold it would be sufficient that the MPE of the game be "regular" in the sense outlined by Doraszelski and Escobar (2010), who also show that this applies to most dynamic discrete games.

In summary, while it seems probable that the structural likelihood in equation (3.6) satisfies the requirements for consistency, provided that the model is specified correctly, it is not certain that the requirements for asymptotic normality, which are stricter, are satisfied. Therefore, a Monte Carlo study is needed to examine whether and under which conditions the MLE for θ is indeed consistent, and possibly asymptotically normal.

3.3.3. Monte Carlo approach

For the purpose of the Monte Carlo study, I set up a data generating process (DGP) that produces a sequence of observed market states in M independent markets with N firms each, given a set of (hyper-)parameters. Then, the (hyper-)parameters are estimated via the nested fixed point maximum likelihood algorithm (NFXP). The empirical distribution of these estimates approximates the theoretical distribution of the estimator and is used to study its properties under various assumptions about the DGP and the NFXP. The DGP is constructed as follows:

Procedure 2 (Data generating process DGP)**.**
Required parameters*:* N,T,M,$\boldsymbol{\gamma}$,$\boldsymbol{\psi}$, seed, *and* burnin.
Seed the random number generator with seed. *For each market* $m \in \{1, \dots, M\}$, *draw a uniformly random set of* N *firm locations with* $x_{jm} \in [0, 100]^2$ *for* $j \in \{1, \dots, N\}$*; initialize an empty list of observed market states* $\hat{\mathbf{A}}_m$*; and an initial market state* $\mathbf{a}_{mt} = (0, 0, \dots, 0)$. *Set* $i = t = 0$*, and*

1. *compute the SOE(*k*) with cutoff distance* d *around the market state* $\mathbf{a}_{mt}$*, resulting in a set of CCPs denoted by* $\tilde{\mathbf{q}}^*_m$*;*
2. *draw a future state* $\mathbf{a}'$ *from the distribution over* $\mathbb{A}$ *that is induced by the choice probabilities* $\tilde{q}^*_{jm}(\mathbf{a}_{mt})$*;*
3. *update* $\mathbf{a}_{mt} \leftarrow \mathbf{a}'$*;*

4. *if* $i >$burnin*, append* $\mathbf{a}_{mt}$ *to* $\hat{\mathbf{A}}_m$ *and increment* t *by one;*

5. *if* $t < T$*, increment* i *by one and go to step 1, else stop.*

Return the collection of simulated market state sequences $\hat{\mathbf{A}} = \{\hat{\mathbf{A}}_m\}_{m=1}^M$ *and the simulated firm locations* $\mathbf{X} = \{x_{jm}\}$.

Given the returned data on firm locations and market states, the parameters and possibly the hyper-parameters can be estimated via maximum likelihood. Importantly, the SOE underlying the NFXP may differ from the SOE underlying the DGP. For example, the data may be generated with a large strategic neighbourhood, and the NFXP may be based on a less sophisticated type of reasoning with a smaller strategic neighbourhood. Also, the distance cutoff d may be different, and even the firm locations need not be the same, because the DGP could lead to some firms being inactive in all periods, so that they would be unobserved for the researcher. Then, the NFXP would be constructed by using a smaller set of potential entry locations. And of course, if the hyper-parameters are not estimated via grid search, these could differ, too. The procedure, due to Rust (1987), is formalized as follows.

Procedure 3 (Nested fixed point alorithm NFXP)**.**
Required data*: observed market states* $\hat{\mathbf{A}}$ *and firm locations* $\mathbf{X}$.
Required parameters*: initial value* $\boldsymbol{\gamma}_0$*, and a set of hyper-parameters* Ψ
Initialize an empty list L*. For each* $\boldsymbol{\psi} \in \Psi$*:*

1. *set up the aggregate likelihood constructed from a sequence of SOEs for each observed market state, using the hyper-parameters* $\boldsymbol{\psi}$ *and firm locations* $\mathbf{X}$*;*

2. *obtain maximum likelihood estimates* $\hat{\boldsymbol{\gamma}}_{\boldsymbol{\psi}}$*, starting the optimization routine at the initial values* $\boldsymbol{\gamma}_0$.

3. *append the triple* $\left(\boldsymbol{\psi}, \hat{\boldsymbol{\gamma}}_{\boldsymbol{\psi}}, ll(\hat{\boldsymbol{\gamma}}_{\boldsymbol{\psi}})\right)$ *to* L

Determine $\hat{\boldsymbol{\psi}} \in \Psi$ *that maximizes* $ll(\hat{\boldsymbol{\gamma}}_{\boldsymbol{\psi}})$*; and return* $\hat{\boldsymbol{\psi}}$ *and* $\hat{\boldsymbol{\gamma}} \equiv \hat{\boldsymbol{\gamma}}_{\hat{\boldsymbol{\psi}}}$.

Of course, Ψ could be a singleton and so the last grid search step would be trivial. Having laid out the DGP and the MLE, the Monte Carlo routine can be described as follows:

Procedure 4 (Monte Carlo Simulation)**.** *For each of* S *samples,*

1. *run the DGP with* N,T,M*, and true parameters* $\boldsymbol{\gamma}_0$*, and* $\boldsymbol{\psi}_0$*;*

2. *if desired, identify firms which are never active, and delete the corresponding locations and market states from* $\hat{\mathbf{A}}$ *and* $\mathbf{X}$*;*

3. *obtain estimates* $\hat{\boldsymbol{\gamma}}_s, \hat{\boldsymbol{\psi}}_s$ *from the MLE that is set up with market states* $\hat{\mathbf{A}}$ *and firm locations* $\mathbf{X}$*, the initial value* $\boldsymbol{\gamma}_0$*, and a set of hyper-parameters* Ψ*;*

Note that the model makes predictions about state-to-state transitions, so that the total number of observations is $n = N(T-1)M$, rather than NTM. This fairly general description of the Monte Carlo simulation encompasses many different cases that can be used to examine the small-sample properties of the MLE as well as its limiting behaviour under different assumptions.

As was outlined above, identification of the regular parameters requires that the data does not exhibit pathological behaviour, and contains the required features. Thus, it is necessary to select the (hyper-)parameters for the DGP such that these requirements are met. To this end, I followed an informed trial and error process to obtain simulated market sequences that resemble the data that were used in chapter 2. First, I chose a value for the spatial interaction radius d. Because firm locations are simulated on a square with side length one hundred, the expected number of nearest neighbours for any firm is approximately[7]

$$\mathbb{E}\left[\mathcal{N}_j^d(\mathbf{a})\right] \approx \frac{N\pi d^2}{100^2} \tag{3.7}$$

With $d = 35$ and $N = 10$ firms, this expectation evaluates to 3.8. I keep this number constant by adjusting d whenever I change the number of firms, so that a larger number of firms is equivalent to a larger market, but not to a denser market. With $\mathbb{E}\left[\mathcal{N}_j^d(\mathbf{a})\right] = 3.8$, I chose $\delta = 0.1$ so that firms experience substantive competitive pressure, as their period returns are decreased by a share of 0.38, on average. Depending on the spatial constellation of all firms, this will be quite different for various firm locations.[8] Next the entry cost parameter was determined such that the simulated data exhibit a small amount of turnover. Also, I chose α to be equal to 0.3 because this lead to a large number of active locations. The strategic neighbourhood was chosen to be of size $k = 5$ because it is about the same as the number of nearby competitors within the radius d, 3.8, plus one. Also, it still allows for a fairly quick computation of the SOE, which is an important aspect for the feasibility of the Monte Carlo study. To summarize, the default parameters for the DGP are as follows:

$$\boldsymbol{\psi}_0 : d = \sqrt{\frac{10 \cdot 35^2}{N}}; k = 5 \tag{3.8}$$

$$\boldsymbol{\gamma}_0 : \alpha_0 = 0.3; \delta_0 = 0.1; \theta_0^e = 2 \tag{3.9}$$

Because I expect that most real-world data are generated by a stationary process in the sense that the number of active and inactive firms remains roughly the same, I initialized the DGP with an initial state where all firms are inactive, and then discarded the first thirty samples. However, the NFXP does not, in general, rely on the observational data coming from a stationary distribution. An example of the market patterns that are generated by the DGP with the aforementioned parameters and twenty firms is shown in figure 3.1. Initially, all firms are inactive, and no burn-in samples were discarded for illustrative purposes. The figure encompasses all the data that are necessary to obtain maximum likelihood estimates of the structural parameters – firm locations, and activity status for a number of successive periods.

[7]This is an approximate figure because edge effects are not taken into account.

[8]For $N > 10$ firms, I adjusted the profit equation (3.1) to $\pi_j(\mathbf{a}_t) = \alpha \max\left\{0, \left(1 - \delta\mathcal{N}_j^d(\mathbf{a}_t)\right)\right\}$.

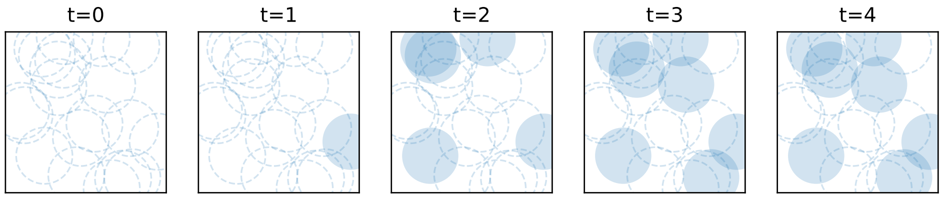


Figure 3.1.: Five simulated market states in a market with twenty firms. Active firms indicated by filled circles; inactive firms indicated by hollow circles. Circles around each firm location are drawn with half the interaction radius d, so that firms interact if their circles overlap.

Consistency and asymptotic normality To assess whether the MLE for the dynamic model outlined in section 3.3.1 is consistent and asymptotically normal, I initialized the DGP with varying numbers of firms per market, time periods, and number of markets, and then determined whether the MLE approaches the true parameter values (3.9). I assumed that the econometrician observes all potential entry locations, and knows the hyper-parameters $\boldsymbol{\psi}_0$. As a base line scenario, I set up the DGP in a very small market with only ten firms, eleven observed time periods (i.e., ten observed state transitions), and a single market. I then let increased the number of firms to $N \in \{20, 50, 100\}$, the number of time periods to $T \in \{21, 51, 101\}$, and the number of independent markets to $M \in \{2, 5, 10\}$, all the while keeping the other dimensions fixed. Throughout, I used $S = 100$ Monte Carlo samples.In addition, I conducted a specification with an even larger data set ($N = 10, T = 21, M = 50$) in order to better study the limiting behaviour. Recall that the regular model parameters are contained in the vector $\boldsymbol{\gamma}$ with $\boldsymbol{\gamma}_k, k = 1, 2, 3$ denoting its individual elements α, δ, and θ^e, respectively.

There are different ways to define how a stochastic variable converges. In theorems 3 and 4, it is defined as convergence in probability. A stronger concept is convergence in mean squared error (MSE), which states that a stochastic variable θ_n converges to θ if $\mathbb{E}[(\theta_n - \theta)^2)] \to 0$ as $n \to \infty$. MSE convergence implies convergence in probability (Cameron and Trivedi, 2005, p.946). To see whether the proposed estimator for $\hat{\boldsymbol{\gamma}}$ exhibits MSE convergence, I constructed the empirical analogue of the mean squared estimation error across all Monte Carlo samples, and inspected whether this measure decreases monotonically as the sample size is increased.

Testing whether the MLE is asymptotically normally distributed is less straightforward, and so I used a variety of approaches. First, I plotted the quantiles of the re-scaled Monte Carlo estimates $t_{k,s} \equiv \sqrt{n}(\hat{\boldsymbol{\gamma}}_{k,s} - \boldsymbol{\gamma}_{k,0})$ against the quantiles of the normal distribution (QQ plot).[9] As a formal test for normality of the scalar estimates for α, δ, and θ^e, I used the test developed by Shapiro and Francia (1972). In addition, the method by Doornik and Hansen (2008) was employed to test whether the parameter estimate $\hat{\boldsymbol{\gamma}}$ converges to a multivariate normal distribution. In both cases, the null hypothesis is that

[9]The QQ-plot as implemented in Stata was used. According to Stata's documentation, it is constructed as follows: Given a set of data points $\{x_s\}_{s=1}^S$, the data is brought in ascending order $x_{(1)} < \ldots < x_{(S)}$. For each $s = 1, \ldots, S$, the the coordinate $(x_{(s)}, q_s\hat{\sigma} + \hat{\mu})$ is plotted, where $q_s = \Phi^{-1}(p_s)$ and $p_s = \frac{S}{N+1}$.

the data are normally distributed, while for the purpose at hand it would be better if this was the alternative hypothesis so as to minimize the risk of falsely concluding that a distribution is normal. With the available tests, a failure to reject the null hypothesis may be due to a lack of power against the alternative, rather than being a definite proof that the distribution is indeed normally distributed. To alleviate this concern, I used tests that are specifically tailored for tests against a normal distribution.[10] If the estimates converge to a normal distribution, I expect that the p-values of the tests become larger, so that the null hypothesis is rejected less frequently. Thus, increasing p-values are taken as an indicator for convergence towards a normal distribution. Both tests were conducted using the statistical software Stata (Stata Corp., 2017, pp. 590–597; 2712–2717).

Hyper-parameter search I further tested whether it is possible to estimate the model's hyper-parameters by using a maximum likelihood grid search. In the baseline case, the DGP was initialized with the parameters given in (3.9) and (3.8), and I set up a market with ten firms, eleven observed time periods, and one observed independent market. In order to see whether the result of the grid search improves as more data becomes available, I increased the number of observed time periods 51 and 101.[11] The search grid for the is visualized in figure 3.2. I ran a total of one hundred Monte Carlo iterations for each scenario.

Convergence in probability requires that the probability, with which a stochastic variable deviates from its probability limit by more than an arbitrary amount, approaches zero. For instance, this implies that we should expect the estimates for the hyper-parameters to fall within the region delineated by the dashed box in figure 3.2 more frequently in our Monte-Carlo samples, as the number of observed time periods is increased. The dashed box is subsequently called the "ϵ-box". Of course, the converse is not true; if the proportion of estimates falling into the ϵ-box increases, this does by no means imply that the probability of the estimator to fall in *any* region around the true values increases and approaches one. But it would still lend support to the hypothesis that the estimator for the hyper-parameters $\boldsymbol{\psi}$ is consistent. To formalize this idea, I conducted an approximate two-sample t-test, with the null hypothesis being that the proportion of estimates falling into the ϵ-box is larger in the scenario with 51 or 101 observed time periods, than in the scenario with only eleven observed time periods. The tests were conducted using the statistical software Stata (Stata Corp., 2017, pp. 2070–2079). In addition, I computed the mean squared error (MSE) for $\hat{k}$ and $\hat{d}$ in the three Monte Carlo scenarios, to see whether it decreases as the number of observed time periods increases, as would be consistent with MSE convergence.

Robustness with respect to unobserved entry locations Lastly, I assessed whether it is important to know all potential entry locations even if some are never observed to be active. For this purpose,

[10] The Kolmogorv-Smirnov test that is more commonly known does not possess this desired property, and is thus less powerful against alternative distributions (Shapiro and Francia, 1972).

[11] I found that increasing the number of time periods is the least time consuming way to obtain more data, because the equilibria across subsequent time periods can be re-used if market states occur repeatedly.

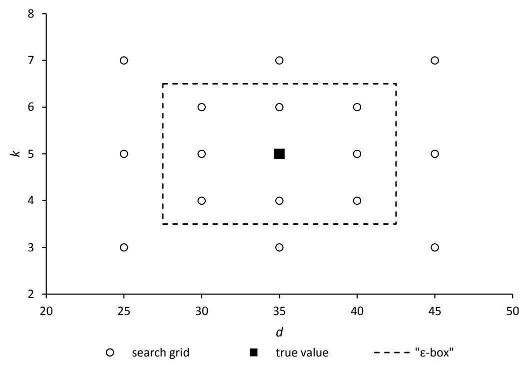


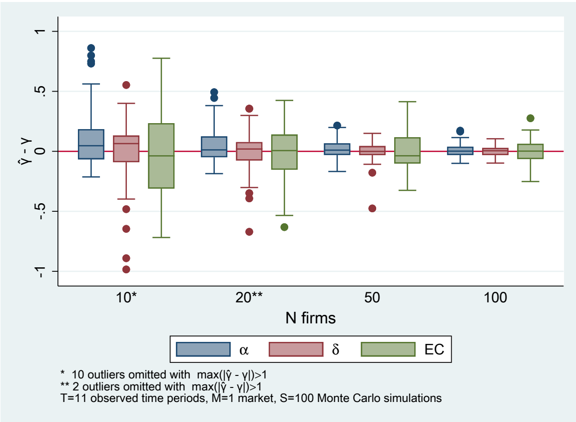
Figure 3.2.: Search grid for the hyper-parameters $\boldsymbol{\psi}$.

I initialized a DGP with twenty firms, twenty-one observed time periods, and one observed market. The true parameter values are as given in (3.8) and (3.9), with the exception that I did not adjust the local interaction radius, and decreased the period return parameter to $\alpha_0 = 0.2$ so that fewer firms will be active, on average. I also increased the entry costs slightly to $\theta^e = 3$ so that there is a higher chance that some firms will never enter. I then set up the MLE to use only those firm locations that are observed to be active at least once during the observed time horizon. The hyper-parameters were assumed to be known.

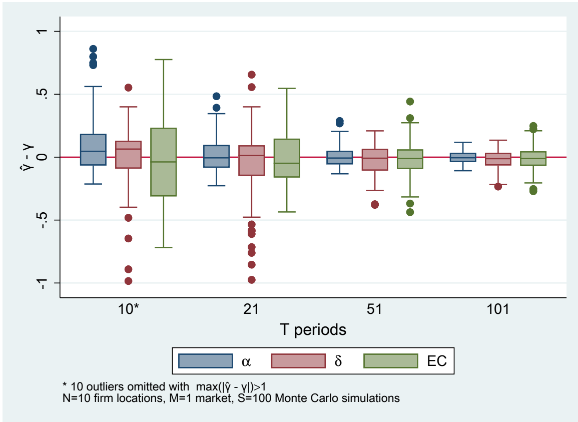
3.4. Results

3.4.1. Consistency and asymptotic normality

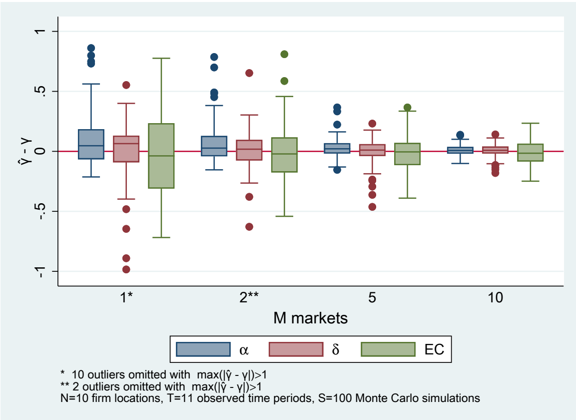
Graphical results from the Monte Carlo study are shown in figure 3.3. Each panel of that figure depicts a box plot of the estimation errors for each of the parameters (α, δ, and $\theta^e = EC$), as one dimension of the data is increased. In the top left panel, $T = 11$ time periods and $M = 1$ market is held constant, and the number of firms is increased. In the top right panel, the number of observed time periods is increased, and in the bottom right panel, the number of independently observed markets is increased. For each data set, the total number of observed state transitions at the firm level is $n = N(T-1)M$, and the bottom right panel shows the evolution of the estimation error as n increases. Notably, all panels show that the estimation errors for all three structural parameters are rather small, and that there is no discernible bias for sample sizes of two hundred state-firm transitions and more. Also, the variance of the estimation errors decreases as more data becomes available. I also computed the MSE for each of α, δ, and θ^e as the total number of observed firm-state transitions n increases, shown in table 3.1. That table shows that the MSE for all parameters decrease monotonically, albeit at a decreasing rate, as more data becomes available to the MLE. Taken together table 3.1 and figure 3.3 lend strong support to the hypothesis that the proposed MLE is consistent.



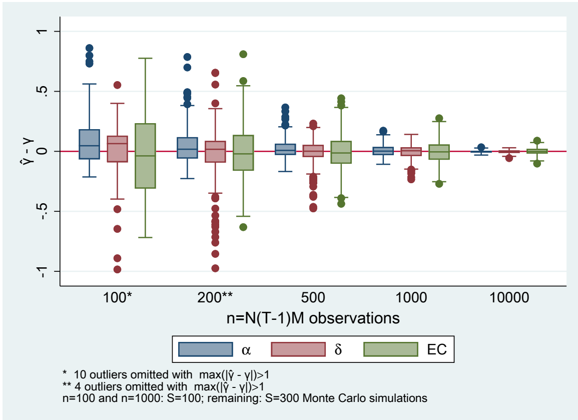
(a) firms $N \to \infty$



(b) periods $T \to \infty$



(c) markets $M \to \infty$



(d) $n = N(T-1)M \to \infty$

Figure 3.3.: Consistency: Box plots of the estimation errors $\hat{\gamma} - \gamma$ for different sample sizes. EC denotes the entry cost parameter θ^e.

Next, I determined the asymptotic distribution of the estimators. Figure 3.4 shows the QQ-plots of the re-scaled parameter estimates against the normal distribution. The figure clearly shows that the estimates follow a highly non-normal distribution for small sample sizes with only a few hundred firm-state observations. However, as the sample size increases to one thousand, or ten thousand observed firm-state transitions, the distribution of the re-scaled estimates begins to resembles that of the normal distribution. I also conducted formal tests for normality of the re-scaled parameter estimates, shown in the following table 3.1. The tests are actually based on the QQ-plots (Shapiro and Francia, 1972) and so it is unsurprising that their results are in line with the visual impression of the those plots. Table 3.1 shows the p-values of these tests; the null hypothesis is always a normal distribution of the re-scaled parameter estimates. The test for univariate normality of the entry cost parameter cannot reject the null hypothesis for all sample sizes. Rather surprisingly, the p-value is very large for small sample sizes, and tends to decrease as the sample size becomes larger, albeit non-monotonically. A plausible explanation is that the estimator is indeed normally distributed, so that the null hypothesis is true, and in consequence, the p-value is distributed uniformly on the interval $[0, 1]$. On the contrary, the small p-value for small sample sizes point to a strongly non-normal distribution of $\hat{\alpha}$ and $\hat{\delta}$. For the largest sample size, the null hypothesis cannot be rejected for all univariate components of the parameter vector. However, the Doornik-Hansen test for multivariate normality of the entire parameter vector $\hat{\boldsymbol{\gamma}}$ would still reject the null hypothesis at a conventional level of significance of 0.05. Still, it can be suspected, albeit not proven, that the MLE will come close to a multivariate normal distribution as the data size is increased even further.

3.4.2. Consistency of the hyper-parameter grid search

I now turn to the question whether the hyper-parameters $\boldsymbol{\psi}$ of the model – namely, the size of the strategic neighbourhood k, and the local interaction range d – can be estimated consistently by means of a grid search. Table 3.2 shows the distribution with which each of the grid points attained the maximum likelihood for a different number of observed time periods, with 100 Monte Carlo samples each.

Consider first panels (a) through to (c) of table 3.2. In table 3.2a, representing the small sample case with eleven observed time periods, a share of 0.43 of all estimates fell within the ϵ-box (see figure 3.5). The corresponding proportions for 51 and 101 observed time periods were 0.48 and 0.52, respectively. Since there are a total of seventeen grid points, and the ϵ-box consists of nine grid points, one would expect a share of $\frac{9}{17} \approx 0.53$ to fall in the ϵ-box even if the MLE was completely uninformative. The fact that the actual share of estimates to fall in the ϵ-box is smaller than this expected share suggests that the grid search procedure for the hyper-parameters is not very good, even in a comparatively large sample with 101 observed time periods. Yet, this could be due to a bias that vanishes as the sample size increases. Indeed, the share of estimates that fall into the ϵ-box increases as more time periods are observed. But an approximate t-test fails to confirm that this increase is statistically significant.[12] While this lack of significance could of course be due to the small power of the test as a result of the relatively small number of Monte Carlo simulations ($S = 100$), the combined evidence so far does not allow me to conclude that the grid search result improves as the number of observed data points increases. At last, consider the MSE statistics in table 3.2d. This table shows that the MSE for both $\hat{k}$ and $\hat{d}$ decreases as the number of observed time periods increases, but this decline is rather slow.

		consistency			p-values (H_0: normality)			
n	S	$MSE(\hat{\alpha})$	$MSE(\hat{\delta})$	$MSE(\hat{\theta}^x)$	$\hat{\alpha}$	$\hat{\delta}$	$\hat{\theta}^x$	$\hat{\boldsymbol{\gamma}}$
100	100	0.8671	0.7952	0.1547	<0.0001	<0.0001	0.8588	<0.0001
200	300	0.0253	0.0657	0.0551	<0.0001	<0.0001	0.6012	<0.0001
500	300	0.0064	0.0111	0.0239	<0.0001	<0.0001	0.8545	<0.0001
1,000	300	0.0025	0.0037	0.0101	0.0292	0.0001	0.5512	<0.0001
10,000	100	0.0002	0.0003	0.0010	0.1463	0.6654	0.2591	0.0370

Table 3.1.: The table shows the mean squared errors (MSE) and the p-values of distributional tests against the normal distribution (H_0: normality), for different data sizes n. S is the number of Monte Carlo simulations that are available for each sample size.

[12] Let $q_T^\epsilon \equiv \Pr\left(\hat{\boldsymbol{\psi}}_T \in \epsilon\text{-box}\right)$. (1) $H_0 : q_{11}^\epsilon \leq q_{51}^\epsilon$ yields $p = 0.239$ (2) $H_0 : q_{11}^\epsilon \leq q_{101}^\epsilon$ yields $p = 0.101$

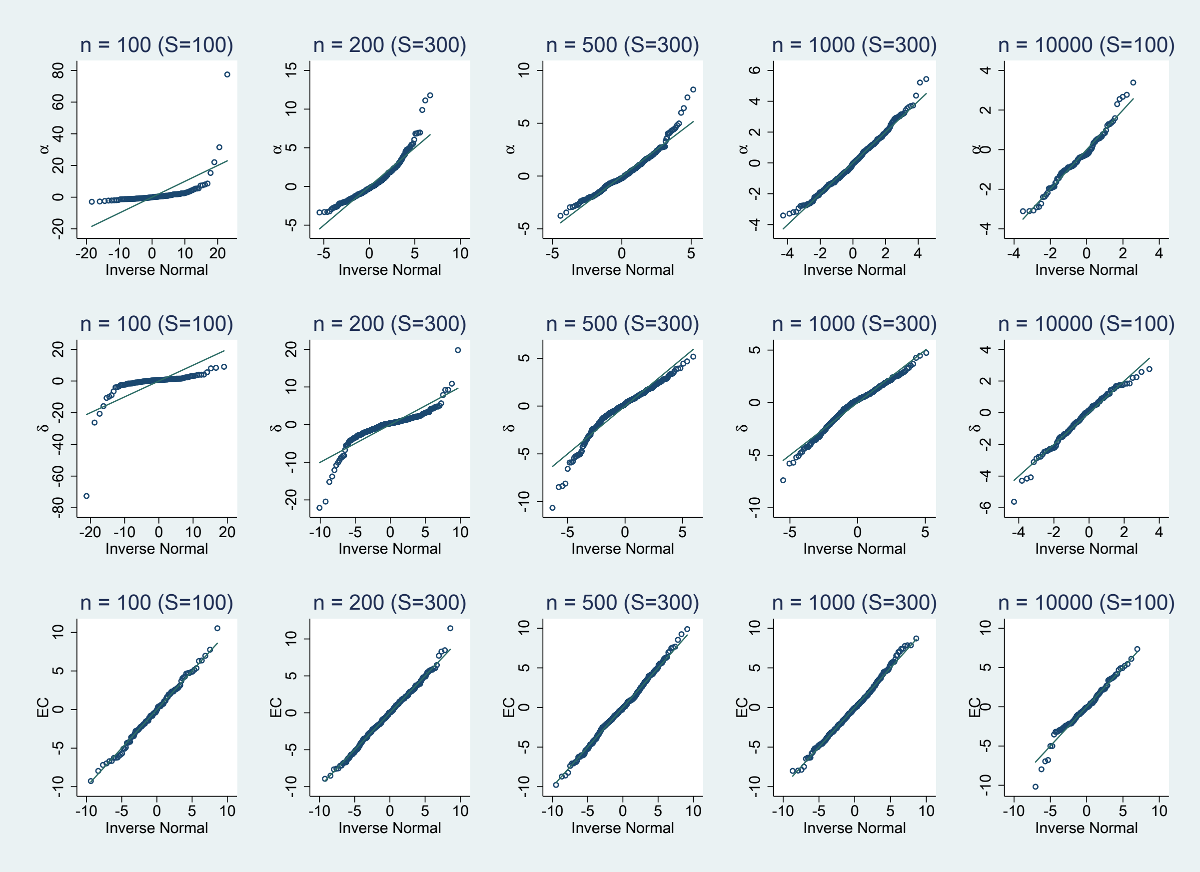


Figure 3.4.: QQ plots of the re-scaled parameter estimates $\sqrt{n}(\hat{\boldsymbol{\gamma}}_{k,s} - \boldsymbol{\gamma}_{k,0})$ against the normal distribution.

Surprisingly, the estimates of the regular parameters $\hat{\boldsymbol{\gamma}}$ are still rather good. This is shown in figure 3.5 which depicts box plots of the distribution of estimation errors for the regular parameters, depending on the sample size, and on whether the hyper-parameters were known, or estimated. The figure shows that the grid search introduces a slight bias in parameter estimates, and a lot more variance, for small sample sizes. However, as the sample size is increased, bias and variance decrease considerably. This is also reflected in table 3.3, which shows the corresponding estimates of the MSE. The last bloc of that table shows the ratio of the MSE as the hyper-parameters are known, or estimated. The table shows that the MSE of the regular parameters decreases as the number of observed time periods T is increased, even though the hyper-parameters are unknown and estimated via the grid search procedure. Furthermore, the ratio of the MSEs for $\hat{\alpha}$ and $\hat{\delta}$ decreases steadily. Interestingly, this is not true for the MSE of the entry cost parameter estimate $\hat{\theta}^e$. Taken together, figure 3.5 and 3.3 make it seem likely that the model's regular parameters can be estimated consistently even if the hyper-parameters are unknown and must be estimated via a grid search.

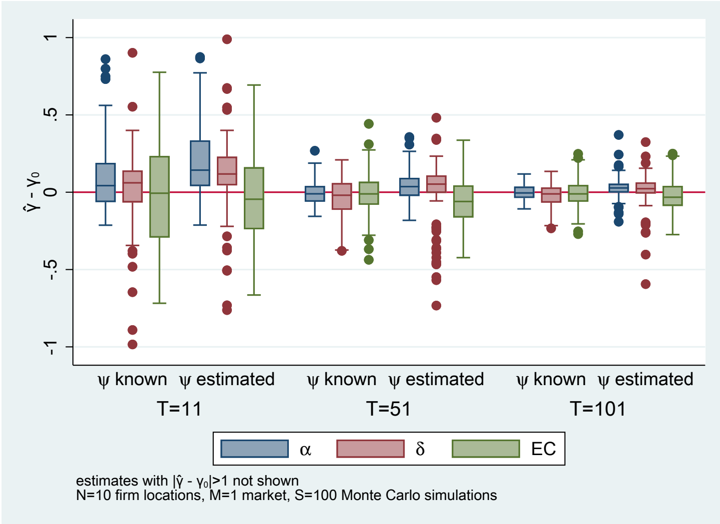


Figure 3.5.: Distribution of regular parameters' estimation errors $\hat{\gamma} - \gamma_0$ if the hyper-parameter are known, or estimated via a grid search.

3.4.3. Robustness with respect to unobserved entry locations

Figure 3.6 shows the distribution of estimation errors that obtains with, and without the presence of unobserved entry locations. The figure shows that the estimates of the profitability parameter α, and the local interaction parameter δ, are almost unaffected by the presence of unobserved entry locations. On the other hand, the presence of unobserved entry locations leads to a clear bias in estimating the entry cost parameter. A Kolmogorov-Smirnov test (with exact p-values) was used to test whether the distribution of estimation errors differs depending on whether unobserved entry locations are included in the estimation approach. This test did not reject the combined null hypothesis that the distribution functions for the estimation errors $\hat{\alpha}_s - \alpha_0$ and $\hat{\delta}_s - \delta_0$ differ in the two estimation scenarios ($p = 0.908$ and $p = 0.368$, respectively), but it did reject this hypothesis for the entry cost parameter ($p < 0.001$). I take note of the fact that this test has the equality of the two distributions as its null hypothesis, and so a failure to reject the null hypothesis may be due to a lack of power, rather than proving that the two distribution functions are statistically the same. Still, the failure to reject the null for the distribution functions of $\hat{\alpha}$ and $\hat{\delta}$, combined with the visual impression of figure 3.6 leads me to conclude that the parameters of the period return function can be robustly estimated in the presence of unobserved entry locations.

3.5. Conclusion

I have revisited the spatial dynamic entry model that was developed in chapter 2, and used Monte Carlo simulations to examine if its structural parameters can be estimated consistently. My results are threefold. First, my results confirm that model's regular parameters can be consistently estimated, and

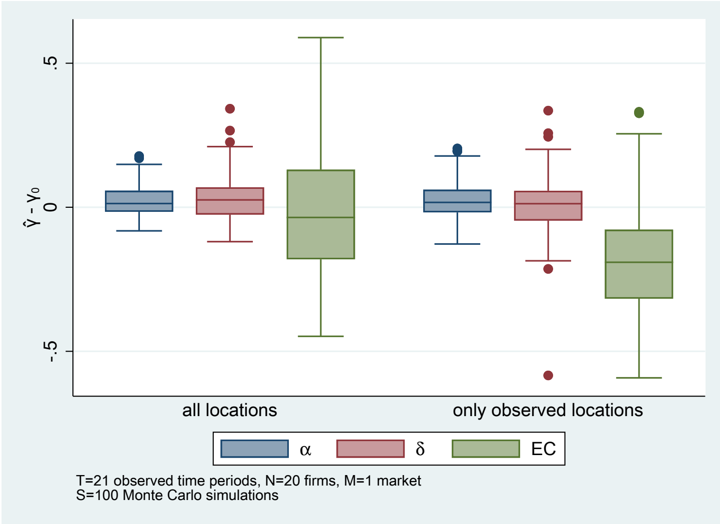


Figure 3.6.: Distribution of regular parameters' estimation errors $\hat{\gamma} - \gamma_0$ in the presence of unobserved entry locations.

that the estimates are asymptotically normally distributed, provided that the model's hyper-parameters – the spatial interaction radius, and the size of the strategic neighbourhood – are known. Second, if these hyper-parameters are unknown, the results in this paper imply that it is difficult to estimate them via a grid search approach. But despite this fact, it appears that the model's regular parameters can still be consistently estimated. Lastly, it is possible to obtain good estimates of the period return parameters even if some potential entry locations are unknown. However, in that case the entry cost parameter cannot be estimated consistently.

These results were derived by using a simplified version of the original model, and with a certain set of parameters. I expect that the conclusions in this paper can be upheld if the model's parameters are changed slightly, but I cannot prove that this is the case for all possible parameter constellations. In particular, there may be parameter constellations that generate market outcomes with insufficient variation to identify all parameters in realistically small samples.

Keeping in mind these caveats, my results have a number of interesting implications. First, the fact that the estimates appear to be normally distributed for larger sample sizes validates the use of asymptotic theory in order to derive standard errors in chapter 2. The fact the the model's parameters can still be estimated quite accurately without knowing the model's hyper-parameters, or in the presence of unobserved entry locations, lends support to the results obtained in chapter 2. On the other hand, it raises the question whether those parameters could not have been estimated in a substantially simpler framework, possibly neglecting dynamic forward-looking behaviour altogether. Furthermore, the fact that the model's hyper-parameters cannot be estimated well by means of a grid search make it appear unlikely that one could succeed in identifying the parameters of a more elaborate period return function without making substantial additional assumptions.

At last, the Monte Carlo simulations were only possible because I had access to potent computer infrastructure,[13] because the maximum likelihood estimator used a nested fixed point approach. This meant that I could not examine the estimator's behaviour in sample sizes that were as large as I wished them to be, or conducted as many Monte Carlo simulations as would actually be warranted to derive small-sample and asymptotic distributions. Future research should therefore focus on ways in which the spatial oblivious equilibrium concept developed in chapter 2 can be embedded in a more efficient (in terms of computer resources) estimation approach, such as a k-step estimator with a structural transition matrix.

[13] I used the MLS/WISO cluster within the bwHPC project, funded by the state of Baden-Württemberg.

	$\hat{k}$ \ $\hat{d}$	25	30	35	40	45
$\hat{k}$	3	0.13	–	0.08	–	0.13
	4	–	0.11	0.03	0.08	–
	5	0.09	0.01	0.02	0.03	0.04
	6	–	0.05	0.02	0.08	–
	7	0.03	–	0.02	–	0.05

(a) $T = 11$

	$\hat{k}$ \ $\hat{d}$	0.25	0.3	0.35	0.4	0.45
$\hat{k}$	3	0.06	–	0.09	–	0.13
	4	–	0.03	0.07	0.14	–
	5	0.05	0.03	0.08	0.03	0.04
	6	–	0.05	0.04	0.01	–
	7	0.06	–	0.04	–	0.05

(b) $T = 51$

	$\hat{k}$ \ $\hat{d}$	0.25	0.3	0.35	0.4	0.45
$\hat{k}$	3	0.09	–	0.11	–	0.11
	4	–	0.09	0.09	0.06	–
	5	0.01	0.05	0.05	0.07	0.07
	6	–	0.03	0.02	0.06	–
	7	0.05	–	0.02	–	0.02

(c) $T = 101$

T	$MSE(\hat{d})$	$MSE(\hat{k})$
11	56.00	2.13
51	46.25	2.06
101	44.00	1.95

(d) Mean squared estimation error

Table 3.2.: (a)–(c): Empirical distribution function of maximum likelihood estimates for the hyper-parameters k and d. The true values are $k_0 = 5$ and $d_0 = 0.35$. $N = 10$ firms and $M = 1$ market. $S = 100$ Monte Carlo samples in total. Cells denoted by – were not part of the search grid. (d): estimated mean squared error, based on panels (a) through to (c).

	MSE: ψ known (1)			MSE: ψ estd. (2)			ratio (2) / (1)		
T	$\hat{\alpha}$	$\hat{\delta}$	$\hat{\theta}^e$	$\hat{\alpha}$	$\hat{\delta}$	$\hat{\theta}^e$	$\hat{\alpha}$	$\hat{\delta}$	$\hat{\theta}^e$
11	0.856	0.264	0.156	3.737	4.470	0.154	4.4	16.9	1.0
51	0.006	0.018	0.021	0.013	0.086	0.023	2.3	4.8	1.1
101	0.003	0.006	0.009	0.006	0.013	0.012	2.2	2.1	1.4

Table 3.3.: The table shows the estimated mean squared errors (MSE) of the regular parameter estimates for different data sizes, and depending on whether the hyper-parameters $\boldsymbol{\psi} = (k, d)$ were known, or were estimates via a grid search.

4. What Happens when Separate and Unequal School Districts Merge?

joint work with Thilo Klein[1] and Josue Ortega[2]

4.1. Introduction

For students in many countries, the transition from primary to secondary school marks an important step towards adolescence that also affects their future educational and professional careers. The modalities of this transition vary between, and sometimes also within countries, but frequently involve an element of choice whereby students can express their preferences over a set of schools.[3] This set of alternative schools can be quite large and cover the entire country, or it can be limited to local school districts or other administrative boundaries. In the latter case, every district typically constitutes an independent assignment market. School district consolidation is the process by which previously independent assignment markets are merged so that students can now choose from a greater set of alternative schools. This phenomenon has taken place in the US for over a hundred years: the number of school districts has fallen from 125,000 in 1900 to 84,000 in 1950 to under 15,000 today (Brasington, 1999).[4] School district consolidations have also occurred in several other countries, e.g. in Germany (Riedel et al., 2010), Hungary (Bukodi et al., 2008), Sweden (Söderström and Uusitalo, 2010), and New Zealand (Waslander and Thrupp, 1995).

School district consolidation can be undertaken to reduce administrative costs, or to foster integration of racially and economically segregated areas. But in the case of the U.S., this consolidation of school districts is rarely a smooth process and is often met with reluctance by some of the independent districts that are to integrate (Berry and West, 2008). One of the many reasons for the reluctance of districts to merge is the concern that their students will have to attend worse schools after consolidation takes place (Fairman and Donis-Keller, 2012). This concern is not entirely unwarranted, as district consolidation not only leads to more choice, but also to more competition. Which effect dominates is unclear a priori and depends on many factors, not least on students' characteristics and preferences. We shed light on

[1]ZEW and Pforzheim University
[2]Queen's Management School, Queen's University, Belfast and formerly ZEW
[3]See matching-in-practice.eu, accessed on 19 September 2019
[4]Source: Institute of Education Sciences, U.S. Department of Education.

these effects of school district consolidation by means of a theoretical model, and an empirical analysis that is based on data from the Hungarian nationwide school assignment system.

In our theoretical model, we study district consolidation as the merger between disjoint Gale-Shapley many-to-one matching markets that are possibly different in terms of their size and their ratio between students and school seats. Students are assigned to schools using the student-optimal stable matching (SOSM) before and after consolidation takes place. Before district consolidation takes place, students can only attend schools within their own district.[5] We compute the expected welfare gains from consolidation for students in those markets (Proposition 2) and we show that district consolidation generates expected welfare gains for all students, particularly for those who belong to districts that are relatively small, or have a high ratio of local students per school seat.

These theoretical predictions are compared to empirical results that are obtained by using data from secondary school admissions in Hungary, and in particular, from its capital Budapest during 2015. We focus on Budapest because i) we have data of students' stated preferences over all schools in its 23 districts, as well as schools' priorities over all students from the 23 districts; ii) students are assigned using the student-optimal stable matching (SOSM) (Biró, 2008); iii) Hungary consolidated primary school districts in 2013 (Kertesi and Kézdi, 2013), and thus the analysis of the unconsolidated case is particularly meaningful; and iv) we have additional data on students' and schools' characteristics that allow us to tell which school features drive students' preferences, such as schools' previous results in mathematics and Hungarian, distance to the students' home address, and socio-economic status. Our empirical strategy is to compare the SOSM in the integrated market to the matching that results in a counterfactual disintegrated market. In order to compute the counterfactual matchings, we need to construct a complete set of preferences over all market participants – schools and students. To this end, our strategy is to estimate a parametric form of students' preferences over schools, and schools' priorities over students. Despite our data being remarkably detailed, we need to overcome two technical problems here.

The first problem that needs to be addressed is about estimating students' preferences: although in the SOSM it is a weakly dominant strategy for students to report their complete rank-order lists (ROLs) of schools truthfully, stated ROLs may differ from the real ones because students submit strategic ROLs by either omitting schools which they deem unattainable or by truncating their ROLs if they are confident to be assigned to more preferred schools. Both types of omissions have been consistently observed in the field (Chen and Pereyra, 2019) and in the lab (Castillo and Dianat, 2016); and both are particularly important for us because the average student in Budapest ranks only 4 schools, even when they are allowed and encouraged to rank all schools. The fact that students submit rather short preference lists is the reason why we need a parametric approach to construct "true" complete rank order list. But

[5]The SOSM is a stable matching such that there is no other stable matching in which a student is assigned to a more preferred school. It is consistently chosen in real-life school choice and college admissions in several regions, including Boston (Abdulkadiroğlu et al., 2014), Chile (Correa et al., 2019; Hastings et al., 2013), Hungary (Biró, 2008), Paris (Hiller and Tercieux, 2014) and Spain (Mora and Romero-Medina, 2001).

the fact that students may omit some of their top-ranked schools also renders standard approaches to estimates multinomial preferences inapplicable.

A second closely linked technical complication arises with estimating schools' priorities: Hungarian schools only report priorities over the set of students who actually apply to them and not over the entire set of students. In Hungary, and in many other countries, schools' priorities are based on tests, interviews, and previous grades with weights decided by each school (subject to basic governmental guidelines). Therefore, the admission criteria at each school contain important idiosyncratic components that are unobservable to us. Thus, even though Fack et al. (2019) have shown how to estimate students' preferences without assuming truth-telling behavior, we cannot directly apply their discrete choice methods which rely on observing complete schools' priorities over students (for example, when schools' priorities are based on a centralized exam). Yet, we draw on their insights and develop a method to consistently estimate students' preferences when their feasible choice sets are unknown, or latent.

To overcome these technical challenges in preference estimation, our empirical strategy builds on two identifying assumptions. Our first assumption is that the observed assignment is stable, which implies that a student's assigned school must be her top choice among her ex-post feasible schools (and vice versa for schools). The approach is similar to Fack et al. (2019) and Akyol and Krishna (2017). In their settings, ex-post feasible choice sets can be constructed because each student's priority at every school is observed. This is not the case in our setting, where students' and schools' feasible choice sets are latent and therefore need to be endogenized to point-identify parameters.[6] Our second identifying assumption is that students use undominated strategies, i.e. a school is ranked above another one if the former is preferred to the latter. The submitted ROLs then reveal the true partial preference order of students over schools (Haeringer and Klijn, 2009). The method is implemented as a Gibbs sampler that imposes bounds on the latent match valuations that are derived from stability and from the observed ROLs. This approach generalizes the matching estimator, proposed in Logan et al. (2008) and Menzel and Salz (2013) for the marriage market, from a one-to-one matching to a many-to-one matching setting, which is suitable for the school admissions problem studied in this paper. We test our proposed estimation method in Monte-Carlo simulations, and we find that it yields unbiased estimators for students' preferences and schools' priorities. Our estimator is available online.[7]

Our main finding is that the consolidated school market in Budapest is advantageous for the majority of students and yields large welfare gains when compared to a counter-factual situation in which students are only allowed to attend schools in their home districts. Throughout, we compare market assignments that obtain through the student-optimal stable matching. We can quantify these welfare gains as being equivalent to attending a school that is five kilometres closer to the students' home addresses. In other words, the average student would be willing to incur an additional travel distance of five kilometres

[6]In the case where schools are not strategically submitting priority lists, a two-step approach to this problem could be derived from He and Magnac (2019): First, estimate school priorities for all students using schools' observed ranking over applicants. Second, use the estimated priorities to construct personalized choice (or consideration) sets and apply the estimation strategy in Fack et al. (2019).

[7]The estimator is available in C++ and R at github.com/robertaue/stabest.

for being able to attend his assigned school in the consolidated market, rather than his counter-factual assigned school in her home district. We can empirically confirm our theoretical result which states that students who live in smaller districts or in districts with less school capacity benefit more from school district consolidation than the average student. Also, the median student incurs a welfare gain that is positive and almost as large as the average welfare gain. To explain these large utility gains, we devise a method to de-compose the total gains into a choice effect and a competition effect. We find that the large welfare gains are largely due to an enhanced choice set, and that the consolidated market does not lead to greatly increased competitive pressure. This can be explained by the institutional details of the school market in Hungary and in Budapest, which is characterized by a large nominal overcapacity of school seats relative to the number of students. In particular, we show that the gains from school district consolidation are much smaller if we adjust the schools' capacity so as to have just as many school seats as there are students in the aggregate.

The parametric specification of students' utility from choosing a school yields insights into what students value most about their school. We find that travel distance is a very important factor that determines students' choices, but students also prefer schools with a high average academic achievement, and those with a higher average socio-economic status. Unsurprisingly, our results imply that students dislike schools which hold additional oral entrance exams, all things else being equal. Moreover, we find that students have assortative preferences. For instance, students with a high socio-economic background have a stronger preference for schools with a high average socio-economic status than other students. The same holds for students who are particularly strong in Mathematics or in Hungarian language.

Our results have implications for the design of school choice markets that can be summarized as follows. Consolidated school choice markets generate expected welfare gains for students if there is enough aggregate capacity. In our empirical setting, significantly more than one half of all students gained from market consolidation which makes it seem possible to obtain majority support for consolidation projects.

Organization of the chapter This chapter proceeds as follows. Section 4.2 discusses the related literature. Section 4.3 presents the model and our theoretical results. Section 4.5 presents the estimation strategy. Section 4.4 introduces our data and the Hungarian school system. Section 4.6 presents our results, namely the welfare gains from district consolidation using both stated and estimated preferences for both students and schools. Section 4.7 concludes.

4.2. Related literature

Although there is a large literature in economics studying school district consolidation, the majority of it is unrelated to that of matching markets. This literature has four main findings: i) there is evidence of overall improvement in students' performance after district consolidation, yet these improvements are not uniformly distributed and there may be losses for specific groups of students (Berry, 2005; Berry and West, 2008; Cox, 2010; Leach et al., 2010);[8] ii) small and look-alike districts are more likely to merge (Brasington, 1999; Gordon and Knight, 2009); iii) although there is empirical evidence of increased fiscal efficiency due to district consolidation, most of the efficiency has already been achieved (Duncombe et al., 1995; Howley et al., 2011), and iv) district consolidation has diversified the racial composition of schools (Alsbury and Shaw, 2005; Siegel-Hawley et al., 2017).

Our paper is more closely related to the literature on two-sided matching, to which we contribute on two fronts. The first one is the theoretical study of consolidation of distinct Gale-Shapley matching markets. From this area, the closest papers to ours are Ortega (2018, 2019), which study the integration of different one-to-one disjoint matching markets; all of them balanced and of the same size. He shows that i) integration benefits more agents than those it harms, and ii) there are expected welfare gains from integration for all agents in random markets. We extend these results to the substantially more general setting of many-to-one matching markets in which each district has potentially different sizes and ratios between schools and students. Furthermore, we show that in any school choice problem there exists a way to partition of students and schools into districts such that district consolidation weakly harms every student when the SOSM is consistently chosen.

A related series of papers assume instead that the set of schools is disjoint and the pool of students is shared. This implies that some students may receive several admission offers whereas others may get none. Manjunath and Turhan (2016) and Turhan (2019) show that iterative matching procedures can lead to larger welfare gains and fewer incentives to misrepresent preferences when the initial partition of the society is coarser. Using a similar approach, Doğan and Yenmez (2017) show that students are weakly better off when all schools join a centralized clearinghouse, whereas Ekmekci and Yenmez (2019) show that no school has incentives to integrate. Hafalir et al. (2019) also studies district consolidations, but assumes that school districts are allowed to exchange students as long as each student becomes better off in the exchange. They identify conditions in which stable mechanisms satisfy individual rationality, diversity, and balancedness desiderata.

All the aforementioned papers assume there is a school choice system before and after consolidation occurs, but a few others assume instead that each school conducts its own admission system prior to consolidation (Chade et al., 2014; Che and Koh, 2016; Hafalir et al., 2018). Some empirical papers examine students' welfare after school choice is established (Baswana et al., 2019; Braun et al., 2010; Machado and Szerman, 2018), but to our knowledge none of those authors have studied district consolidation with school choice before and after the merge of districts occurs.

The second strand of the literature to which we contribute is the estimation of students' preferences and schools' priorities from observed data. There are several methods for preference estimation with more or less restrictive underlying assumptions. The most common identifying assumption is truth-

[8]There is also a well-established link relating larger school sizes with lower students' performances, which is not the focus of this paper.

telling, where under the SOSM, a student is truth-telling if she submits her k most preferred schools. Abdulkadiroğlu et al. (2017) and Che and Tercieux (2019), for example, follow this assumption in their analysis of the New York City high school match. However, truth-telling is only a weakly dominant strategy, even when schools can be listed at no cost. Commonly observed and rationalizable strategies that are inconsistent with truth-telling include skipping "infeasible" schools and truncating ROLs after "safe" schools. Therefore, other identifying assumptions have been explored in the literature. Fack et al. (2019) is the seminal reference for the estimation of students' preferences when their feasible choice sets are known to the researcher (and to the student).

A less restrictive identifying assumption is that students do not swap their true preference orderings over schools when submitting a ROL, and Fack et al. (2019) have used this assumption to estimate preferences in the Paris school choice context. This assumption is due to the fact that it is a strictly dominated strategy in the student-proposing Gale Shapley mechanism to rank school s' before school s if a student actually prefers school s over school s' (Haeringer and Klijn, 2009).

Another commonly used identifying assumption is stability of the observed matching, which implies that a student's assigned school must be the top choice among her ex-post feasible schools. Artemov et al. (2017) argue that stability is a more innocuous assumption than undominated strategies in that it permits inconsequential 'mistakes' (in the sense of playing dominated strategies). Also, asymptotic stability of the matching is guaranteed ex-post under fairly general conditions (Fack et al., 2019). The most pervasive problem for stablity-based inference in two-sided matching models is that stable matching games may possess multiple (stable) equilibria for a given set of preferences. In the absence of an equilibrium selection rule, models that rely on stability are therefore incomplete (Tamer, 2003). For the most part, the literature has therefore focussed on complete models (see Chiappori and Salanié, 2016; Fox, 2009, for surveys of the literature). One means to ensure uniqueness of the stable matching is to restrict the form of utility functions, mostly by assuming that preferences on both sides of the market are aligned to each other. For instance, Agarwal and Diamond (2014) show that preferences are non-parametrically identified in many-to-one matching markets with perfectly aligned preferences. This approach has been applied to capital and credit markets (Chen, 2013; Sørensen, 2007) and to the US medical match (Agarwal, 2015). A unique stable matching is also guaranteed where administrative admission rules guarantee a global ROL. In the school choice context, this has been applied for Paris (Fack et al., 2019), for college admissions in Mexico (Bucarey, 2018), Turkey (Akyol and Krishna, 2017), and Norway (Kirkebøen, 2012). If such assumptions are not met, then only the joint match surplus may be identifiable from observational data (Logan et al., 2008; Menzel, 2015; Menzel and Salz, 2013). Weldon (2016) also discusses the identification of preference parameters in school choice markets, and provides some Monte Carlo evidence on the convergence properties of stability-based estimators. He finds that estimation routines that rely exclusively on the observed matching being stable can be rather slow to converge as the student-school ratio becomes larger, a finding that we could confirm in our own simulations.

Our methodological contribution to the literature lies in developing a method to simultaneously estimate the parametric form of students' preferences and schools' priorities in such settings where only partial ROLs and the final assignment are known to the econometrician, but where preferences and priorities are not perfectly aligned. We follow the approach of Fack et al. (2019) in that we use the stability assumption in conjunction with the undominated strategies assumption. Their approach, however, is not directly applicable to our setting where the students' feasible choice sets are unobserved, and so we extend it to include latent feasible choice sets using a data augmentation approach.

4.3. Model

To study district consolidation from a theoretical perspective, we first introduce some notation. An extended school choice problem (ESCP) is a tuple $(T, S, \mathcal{D}, \succ, \rhd, q)$, where:

- T is a set of students.
- S is a set of schools. We refer to $\Omega = T \cup S$ as the society.
- q is the number of students that each school can accept.
- $\mathcal{D} := \{D_1, \ldots, D_r\}$ is a partition of $T \cup S$ into r subsets such that each of them has some students and some schools. T^{D_i} and S^{D_i} denote the set of students and schools in district D_i. A *population* P is the union of some (possibly all) districts.
- $\succ_t$ is the strict preference ordering of student t over all schools in S. We write $s \succ_t s'$ to denote that t prefers school s to school s' (and $s \succcurlyeq_t s'$ if either $s \succ_t s'$ or $s = s'$). We use $\succ := (\succ_t)_{t \in T}$ to denote the preference profile of all students.
- $\rhd_s$ is the strict priority structure of school s over all students in T. We use $t \rhd_s t'$ to represent that student t has a higher priority than student t' at school s. We use $\rhd := (\rhd_s)_{s \in S}$ to denote the priorities of all schools.

We assume that each district D_i has qn_i students, $n_i + k_i$ schools and $q(n_i + k_i)$ school seats, where k_i is a positive or negative integer that reflects the imbalance between the supply and demand for school seats in each district. If $k_i > 0$, the district is *underdemanded*; if $k_i < 0$ the district is *overdemanded*; if $k_i = 0$ then the district is *balanced* and each student is guaranteed a seat in his own district. We will assume that $K := \sum_i^r k_i \geq 0$, i.e. the society as a whole is either balanced or underdemanded and the size of its unbalance is K.[9] We also use $N := \sum_i^r n_i$.

The admission policy of each school s is given by a choice rule $\text{Ch}_s : 2^T \times \{q_s\} \mapsto 2^T$, which maps every nonempty subset $T' \subseteq T$ of students to a subset $\text{Ch}_s(T', q_s) \subseteq T'$ such that $|\text{Ch}_s(T', q_s)| \leq q_s$. We assume that for each school s, $\text{Ch}_s(\cdot, q_s)$ is responsive to the priority ranking $\rhd_s$, i.e. for each

[9] This assumption is satisfied in our data and is often satisfied in school choice markets.

$T' \subseteq T$, $\text{Ch}_s(T', q_s)$ is obtained by choosing the highest-priority students in T' until q_s students are chosen.

Given a population P with students T^P and schools S^P, a *matching* $\mu : T^P \cup S^P \mapsto T^P \cup S^P$ is a correspondence such that for each $(t, s) \in T^P \times S^P$, $\mu(t) \in S^P \cup \{t\}$, $\mu(s) \subseteq T^P$, $|\mu(s)| \leq q_s$ and $\mu(t) = s$ if and only if $t \in \mu(s)$. We write $\mu(t) = t$ if student t is unmatched under μ. A *matching scheme* σ is a function that specifies a matching for each district D_i, denoted by $\sigma(\cdot, D_i) : T^{D_i} \cup S^{D_i} \mapsto \mu : T^{D_i} \cup S^{D_i}$, as well as for the society as a whole, denoted by $\sigma(\cdot, \Omega) : T \cup S \mapsto T \cup S$. As no confusion shall arise, when referring to an arbitrary district, we will simply write $\sigma(\cdot, D)$. The matchings $\sigma(\cdot, D)$ and $\sigma(\cdot, \Omega)$ denote the assignment of students to schools before and after consolidation occurs, respectively.[10]

A matching $\mu : T^P \cup S^P \mapsto T^P \cup S^P$ is *stable* if $\nexists (t, s) \in T^P \times S^P$ such that i) $\mu(t) = t$ and $|\mu(s)| < q_s$, or ii) $s \succ_t \mu(t)$ and $t \rhd_s t' \in \mu(s)$. A matching scheme σ is stable if all its corresponding matchings $\sigma(\cdot, D)$ and $\sigma(\cdot, \Omega)$ are stable. The *students-optimal stable matching* (SOSM) is the stable matching that all students weakly prefer over any other stable matching and it is attained by the student-proposing deferred acceptance algorithm (Gale and Shapley, 1962; Roth and Sotomayor, 1992).

Welfare Effects of Consolidation To analyze students' welfare changes we quantify the gains from district consolidation in terms of the ranking of their assigned school. We focus on random ESCPs, in which the schools' priorities and students' preferences are generated uniformly at random. Random matching problems were first studied by Wilson (1972) and have been extensively studied ever since, and they facilitate an analytical treatment of market outcomes. Of course, the preferences in actual school choice problems are not random and so a failure of our theoretical predictions to hold empirically could also be due to the random market assumption not being met in practice. Yet, we argue that the random markets assumption is useful because it allows us to isolate the effects of a larger matching market, rather than being distracted by compositional details of the students and schools.

The *absolute rank* of a school s in the preference order of a student t (over all potential schools in the society) is defined by $\text{rk}_t(s) := |\{s' \in S : s' \succcurlyeq_t s\}|$. Given a matching μ, the *students' absolute average rank of schools* can be defined by

$$\text{rk}_T(\mu) \quad := \quad \frac{1}{|\overline{T}|} \sum_{t \in \overline{T}} \text{rk}_t(\mu(t))$$

where $\overline{T}$ is the set of students assigned to a school under matching μ. Then, the welfare gains from consolidation for students of district D_i are defined as

$$\gamma_T(\sigma_{\text{SOSM}}) = \text{rk}_T(\sigma_{\text{SOSM}}(\cdot, D_i)) - \text{rk}_T(\sigma_{\text{SOSM}}(\cdot, \Omega))$$

[10]Matching schemes are analogous to the concept of assignment schemes in cooperative game theory (Sprumont, 1990).

Proposition 2 approximates the students' welfare gains from consolidation as function of n_i and k_i, providing a set of interesting comparative statistics as a corollary.

Proposition 2. *In a random ESCP, the expected welfare gains from consolidation for students $\gamma_T(\sigma_{\text{SOSM}})$ can be approximated by*

$$\frac{N+K}{q}\left(\frac{\log(\frac{n_i+k_i}{k_i})}{n_i} - \frac{\log(\frac{N+K}{K})}{N}\right) \qquad \text{if } k_i \geq 0 \tag{4.1}$$

$$\frac{N+K}{q}\left(\frac{q(n_i+k_i)}{n_i\log(\frac{n_i}{k_i})} - \frac{\log(\frac{N+K}{K})}{N}\right) \qquad \text{if } k_i < 0 \tag{4.2}$$

We postpone the proof of Proposition 2 to appendix B.1. Expression (4.1) coincides with the one in Proposition 4 in Ortega (2018) when $k_i = K = 0$ and $q = 1$. The approximations presented have several testable implications, which we present below.

Corollary 1. *The gains from consolidation are positive for all districts, in particular:*

1. *If the whole society is underdemanded, students from overdemanded districts benefit more from consolidation than those from underdemanded districts.*
2. *A smaller size of the district size n_i leads to larger expected welfare gains from consolidation.*
3. *A smaller size of the global imbalance K leads to larger expected welfare gains from consolidation.*

Before moving to the empirical part of this paper, where we test our theoretical predictions, we provide some intuition for the comparative statistics. It is well-known that in a two-sided matching problem with different sizes, the agents in the short side choose whereas the agents in the large side get chosen, a phenomenon that increases as the imbalance between the two sides of the market grows (Ashlagi et al., 2017). Thus, if a local district is underdemanded, students get assigned to highly ranked schools before consolidation, which makes the gains from consolidation smaller. On the contrary, if students belong to an overdemanded district, they are assigned to a poorly ranked school before consolidation, which leads to large potential gains from consolidation (which indeed occur, since the whole society is underdemanded). This explains our first comparative statistic.

The two remaining comparative statistics have to do with the relationship between relative and absolute rankings. In small districts, even if students are assigned to some of their preferred schools within their district, it is unlikely that those schools are in the top of their preference list. Thus, in small districts there is large potential for welfare gains. Similarly, the larger the global imbalance K becomes, students are assigned to more preferred schools after consolidation takes place.

4.4. Data

This section describes the school admission system in Hungary, and the data that we use. Hungary has a nation-wide integrated school market. This means that every student can apply to any school in the entire country, and a centralized assignment mechanism is used to allocate students to schools. In this system, every student submits a rank order list (ROL) of arbitrary length ranking the school programmes that he would like to attend. In turn, each school programme ranks all the students that applied to it according to several criteria such as grades, additional exams and entrance interviews. The specific weighting of these criteria is decided upon by each school but must comply with specific regulations (e.g. the weight of the interview score cannot be more than 25%). School programmes submit a strict ranking of their more preferred students; the remaining students are simply deemed unacceptable and are not ranked against each other. The assignment of students to schools is conducted using the deferred acceptance student-proposing algorithm (Biró, 2008). This algorithm has been used since 2000 in a fully consolidated fashion, allowing students to apply and be assigned to any school in the entire country (see Biró (2012) for a detailed overview of its implementation).

For our empirical analysis, we use data from the national centralized matching of students to secondary schools in Hungary, the so-called KIFIR dataset,[11] along with student-level data from the national assessment of basic competencies (NABC), both from the year 2015. Our data encompasses the universe of all students in Hungary who apply to a secondary school programme in 2015 (at an age of 14, with some exceptions). Each secondary school offers general or specialised study programmes with different quotas that are known ex-ante by students. The reader is referred to B.2.4 for some details on these original data sources. Due to data protection arrangements, access to these data was restricted and our estimation routines were run by officials at the Hungarian ministry of education on their local computer.

We will restrict our attention to the greater Budapest area which comprises 23 well-defined districts, so as to obtain a realistic setting within which the (un)consolidation of school districts can be studied. Budapest lends itself to this type of analysis because it is a geographically relatively small market that is tightly integrated, and yet the market is large enough to permit a meaningful study of the unconsolidation into smaller and well-defined districts. Figure 4.1 shows the geographical area of Budapest with school district borders, and with arrows between districts that send their students to study to other districts. That figure shows that there is a considerable amount of inter-district movements, especially in the inner parts of the city.

Out of the 13,611 students from Budapest that we find in the NABC dataset, we are able to link 10,880 students to their corresponding application records in the KIFIR database. In order to attain comparable competitive conditions, we adjust the schools' capacities by removing any seats that were assigned to students not in our sample. In total, there are 881 school programmes of 246 schools that are located

[11] KIFIR stands for *Középiskolai Felvételi Információs Rendszer* which translates to "Information System on Secondary School Entrance Exams"

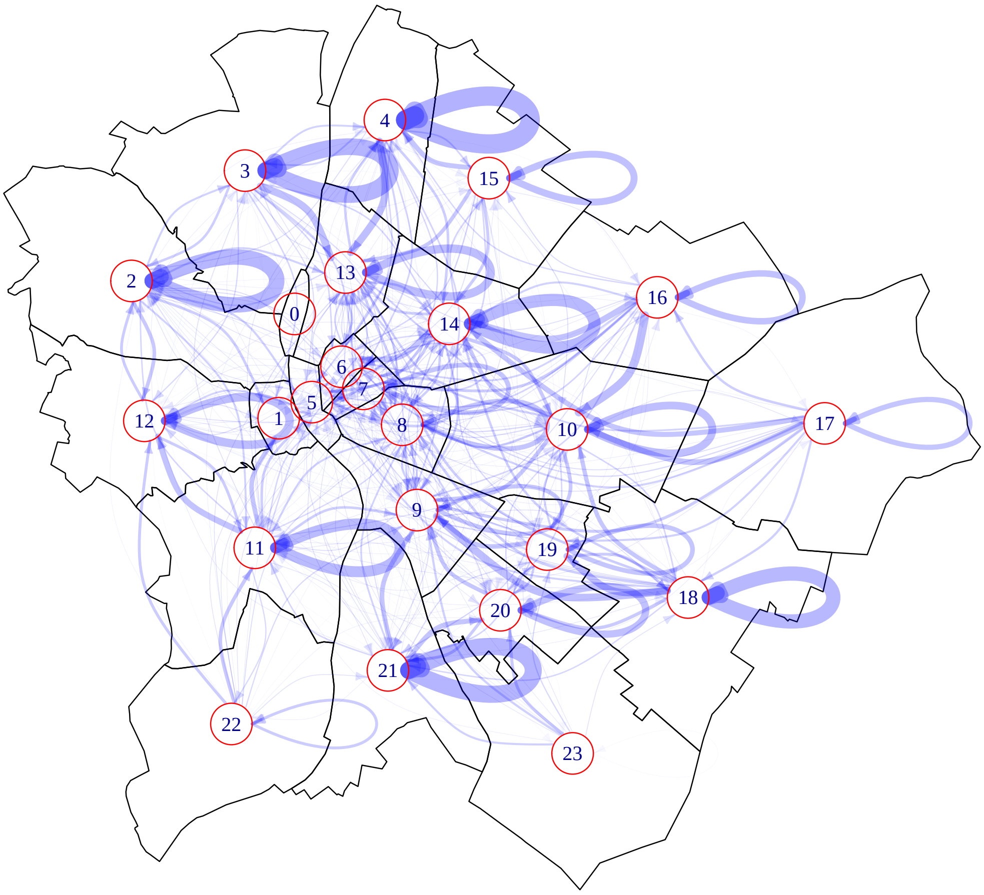


Figure 4.1.: The districts in Budapest, and flows of accepted students among those districts. Flows from one district A to another district B are bent to the left, when viewed from A. The width of the flow arrows from one district D to another district D' is proportional to the number of students who live in district D and who were accepted at a school in district D'.

in the city of Budapest. A school programme can constitute of a particular class in which students specialize on languages, or computer science, for instance. Thus, schools can offer multiple programs within the same age cohort. We aggregate school programmes at the school level in order to reduce the sample size and the associated computational burden which is not negligible in our context.[12] Combining the 246 schools with 10,880 students still leaves us with almost 2.7 million possible student-school combinations to be considered. We focus on three school types – four-year grammar schools, vocational secondary, and vocational schools – which the students apply to after having completed eight years of primary education. For all students in the sample, their location of residence is approximated by their zip code, and the Open Source Routing Machine (Luxen and Vetter, 2011) was used to compute travel distances from each of Hungary's zip code centroids to every known school location.

Table 4.1 shows student-level summary statistics of our data. Panel A shows that most students were born in 2002, and that there are as many girls as boys, as one would expect. The students' mean grade average in the previous school year is four (five is the highest grade in the Hungarian grading system). Their math, Hungarian, and SES scores from the NABC[13] were standardized by us since their absolute numbers have no meaning. The variable measuring students' socio-economic status (SES) is a composite measure that includes, amongst other variables, the number of books that the household has, or the level of parental education. This indicator was also standardized. Since the students' grade average, their math, and their Hungarian NABC scores are highly correlated, we created a composite measure that we call "ability" and which is constructed as the first principal component of these variables. Table 4.1 shows that the students from Budapest in our sample file applications to roughly four schools, on average.[14] Roughly seventy percent of the students apply to at least one school in their home district, and on average, students include only one school from their home district in their submitted rank order list. Panel B shows some attributes of students' first choice school, and panel C shows attributes of the students' actual assigned school. Panel C shows that the average match rank[15] is 1.46 with more than seventy percent of all students being assigned to their top choices. This is probably due to the fact that there is much excess capacity: the schools in the sample reportedly have vastly more seats than there are students (see below). This peculiar fact has been confirmed in conversation with officials from the Hungarian ministry of education on several occasions. The distribution of the number of programmes the students apply to, and of the actual match rank in the 2015 matching round, are shown in figure 4.2. This figure confirms that most students submit rather short ROLs, and the vast majority of students is assigned to their submitted top choice.

Table 4.2 shows the school-level summary statistics. School programmes in Budapest are very attractive so that many students from outside Budapest rank a school in Budapest as their top choice. Therefore, students from Budapest face strong competition in their "domestic" school market, and restricting

[12]We converted students' ROLs to the school level by keeping the most preferred school programme of every school.

[13]Where these scores were missing in our data, we imputed the missing values using predictive mean matching, as implemented in the package `mice` in R (van Buuren and Groothuis-Oudshoorn, 2011); see B.2.4.

[14]Actually, students apply for course programmes, many of which may be offered by the same school. Thus, the actual length of the students' rank order lists is larger than this.

[15]With 1 being the most preferred school.

	Mean	SD	Min	Max	N
Panel A. Student characteristics					
birth year	2,000.1	0.550	1,996	2,002	10,880
female	0.495	0.500	0	1	10,880
grade average	4.064	0.693	1.000	5.000	10,880
math score (NABC)*	0.000	1.000	−3.825	3.521	10,880
hungarian score (NABC)*	0.000	1.000	−4.186	3.176	10,880
ability†	1.472	1.398	-3.662	6.006	10,880
SES score*	0.000	1.000	−4.111	1.651	10,880
ROL length	4.093	1.800	1	24	10,880
applies to home district	0.680	0.466	0	1	10,880
ROL length within home district	1.054	0.965	0	7	10,880
Panel B. Attributes of first-choice school					
distance (km)	7.100	4.630	0.105	36.645	10,880
ave. math score (enrolled students)	0.320	0.716	−1.971	1.754	10,880
ave. hungarian score (enrolled students)	0.352	0.699	−2.006	1.686	10,880
ave. SES score (enrolled students)	0.090	0.582	−1.886	1.212	10,880
Panel C. Attributes of assigned school					
match rank	1.476	0.924	1.000	11.000	9,783
matched to first choice	0.711	0.453	0.000	1.000	9,783
distance (km)	7.061	4.653	0.105	36.645	9,783
assigned to home district	0.297	0.457	0.000	1.000	9,783
ave. math score (enrolled students)	0.195	0.686	−1.971	1.754	9,783
ave. hungarian score (enrolled students)	0.230	0.669	−2.006	1.686	9,783
ave. SES score (enrolled students)	−0.012	0.571	−1.886	1.212	9,783

Variables indicated with an asterisk are z-normalized. The 2015 Hungarian and math test scores are taken by the students as part of the admissions process. † ability is the first principal component of the joint distribution of students' grades, their math, and their hungarian scores. Socioeconomic status is a composite measure which includes, amongst other variables, the number of books that the household has, or the level of parental education.

Table 4.1.: Secondary School Applicants in Budapest: Summary Statistics.

Statistic	N	Mean	St. Dev.	Min	Max
capacity	246	137.098	96.306	6	502
adjusted capacity	246	116.447	90.586	6	498
applications	246	411.199	456.929	7	2,392
ROL1 applications	246	44.228	44.254	0	251
acceptable applications	246	130.638	124.433	0	698
assigned students	246	39.768	31.499	0	157
ave. match rank	242	47.229	34.011	2.250	187.298
entrance interview	246	0.439	0.497	0	1
enrolled students' average					
math	246	−0.130	0.778	−1.971	1.754
Hungarian	246	−0.084	0.747	−2.006	1.686
SES	246	−0.185	0.643	−1.886	1.212
assigned students' average					
math	246	−0.248	0.670	−2.355	1.643
Hungarian	246	−0.253	0.694	−2.332	1.476
SES	246	−0.135	0.638	−1.789	1.282

Table 4.2.: Summary statistics of secondary schools in Budapest.

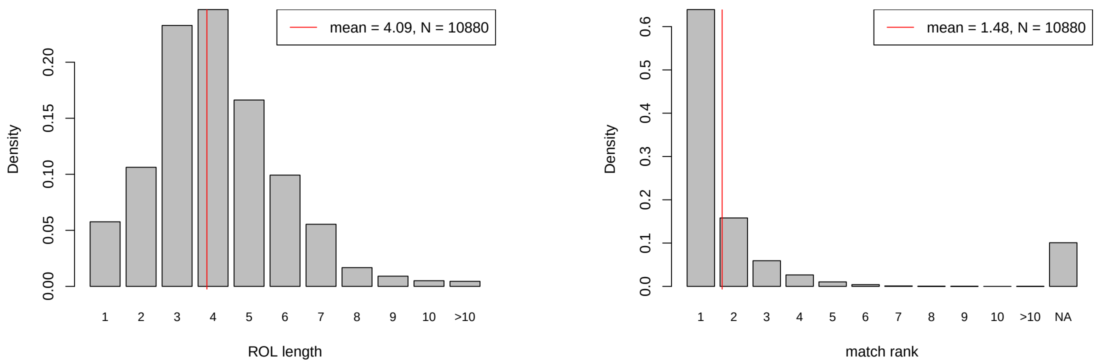


Figure 4.2.: Distribution of the length of students' ROLs, and of their realized match rank, in the 2015 KIFIR database for students in our sample. These numbers refer to rank order lists that were aggregated at the level of the schools (see text for explanation).

the attention to students from Budapest will likely lead to a much more relaxed assignment problem. In order to circumvent this problem, we subtracted the number of admitted students from outside Budapest from the schools' capacity so as to maintain the original "tightness" of the market – this is the adjusted capacity that is used throughout our analysis. The average school receives over four hundred applications, of which only 130 are deemed "acceptable". In the end, about forty students are assigned to each school on average. The comparably small number of acceptable applications could indicate that it is quite costly for schools to rank all their applicants consistently, and so they focus on only ranking those students which are most likely to be admitted to the school. Note that our estimation approach assumes that schools submit their priority lists truthfully, *i.e.* that every student who is labelled "unacceptable" really ranks lower than any other applicant. This assumption could be violated if schools strategically choose to omit very high achieving students, because they feel that these students are more likely to be admitted to a more prestigious school, and thus want to avoid the workload of prioritizing these students. But we think that this is probably a minor problem, and that schools are overall truth-telling. We also collected data on whether a school holds an additional entrance interview[16] and we found that about forty percent of all schools do so. The table also summarizes the school-level averages of admitted and currently enrolled students. The standard deviation of these school-level averages is more than two thirds of the total variance across students, which is normalized to one. Thus, there is evidence for a substantial amount of sorting by ability and socio-economic status.

4.5. Empirical strategy

Our empirical strategy to estimate the gains from district consolidation in a school choice market can be summarized as follows: we compute the SOSM in an unconsolidated, district-level school market

[16]This information was manually collected from the website felvizsga.eu which provides information about admission procedures at different Hungarian schools. Last accessed on 11 November 2019.

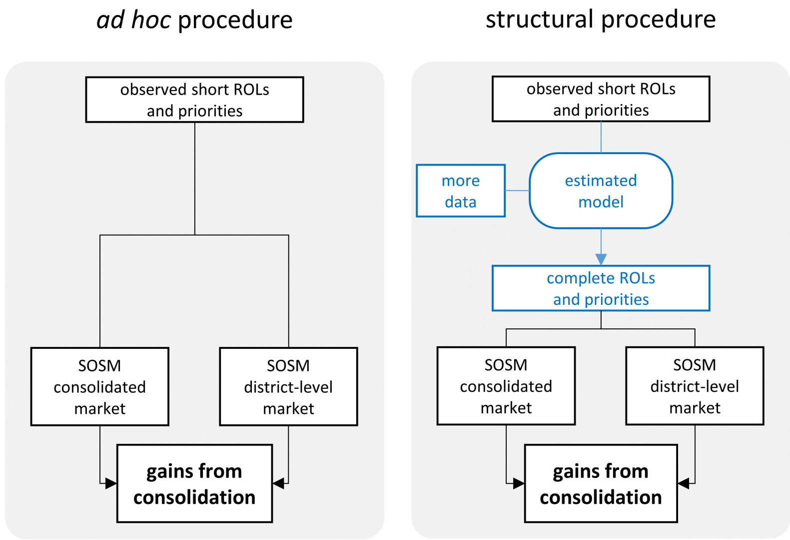


Figure 4.3.: Our empirical strategy

and compare it to the SOSM in the consolidated, city-wide school market. In a first pass, we use the submitted rank order lists to obtain an *ad hoc* measure of the consolidation gains. This approach has some shortcomings due to the fact that the submitted rank order lists are incomplete, as will be outlined below. To circumvent these shortcomings, we develop a procedure to estimate the complete preference order of all market participants. This allows us to compute a more complete SOSM in the unconsolidated market, and also to compare utility outcomes. Figure 4.3 summarizes our strategy at a glance.

In section 4.3 we have shown theoretically that one can expect overall welfare gains from school district consolidation, but that the magnitude of these gains may depend on the specific market characteristics. We test these predictions using student-level administrative data from the Hungarian school assignment system KIFIR.[17] The KIFIR dataset contains the stated preferences of students over all schools that are included in their submitted rank order lists, and the respective rankings of schools over their applicants. These submitted rank order lists allow us to perform an *ad hoc* qualitative assessment of the consolidation gains in terms of foregone rank order items.

However, using the short submitted rank order lists two has shortcomings. The first problem is related to the computation of the matching in an unconsolidated district-level school market. As table 4.1 shows, over thirty percent of all students have not included any school from their home district in their submitted rank order lists, and on average, students included only a single school from their home district in their submitted rank order list. This is probably due to the fact that the school market in Budapest has been consolidated for a long time. As a result, many students would remain unmatched in a counter-factual, disintegrated school market. Moreover, it seems reasonable to assume that students would adjust their submitted rank order lists if the school market was to be disintegrated. Thus, the SOSM in a disintegrated school market cannot be well described by using the submitted short rank

[17]See section 4.4 for details on the Hungarian school choice system, and of the data.

order lists from the consolidated school market. Second, it is unclear how a change in a student's match rank translates to utility gains or losses, because the former is an ordinal concept, whereas the latter is a cardinal concept. Also, the cardinal concept of utility is more appropriate to compute aggregate welfare measures. To overcome these problems, we present a data augmentation approach to back out the "true" complete preference ordering from the submitted rank order lists. Our method is based on the discrete choice framework (Train, 2009) and we use it to compute the different SOSM allocations, and to evaluate their welfare consequences. The method is outlined in more detail below.

4.5.1. Preference estimation: methodology

We observe a school choice market with a set of students (T) and a set of schools (S). We write students' utilities over the set of schools $U_t(s)$, and schools' valuations over the set of students $V_s(t)$ as

$$U_t(s) \quad = \quad U_{t0} + \mathbf{X}_{ts}\beta + \epsilon_{ts} \tag{4.3}$$

$$V_s(t) \quad = \quad V_{s0} + \mathbf{W}_{st}\gamma + \eta_{st} \tag{4.4}$$

where $\mathbf{X}_{ts}$ and $\mathbf{W}_{st}$ are observed characteristics that are specific to the school-student match st. $\mathbf{X}_{ts}$ could, for instance, include a school fixed effect or the travel distance from t to s. The terms U_{t0} and V_{s0} are the outside utilities of not being matched to any student or school. These are assumed to be zero, so that the latent utilities represent the net utility of being matched. The match valuations $U_t(s)$ and $V_s(t)$ are treated as latent variables that are to be estimated along with the structural parameters β and γ. Throughout, we will denote by $\mathbf{U}_t$ the vector of student i's utilities over the entire set of schools, and by $\mathbf{V}_s$ school s's valuations over the entire set of students. We make use of the common indexing notation whereby the elements of some vector $\mathbf{Z}$ that do not refer to the student-school pair ts are denoted by $\mathbf{Z}_{-ts}$, i.e. $\mathbf{U}_{-ts}$ denotes the entire set of utility numbers but for $U_t(s)$. We further assume that the structural error terms ϵ_{ts} and η_{st} are independent across alternatives, and normally distributed with unit variance. While one could in principle allow for more general correlation structures, it is customary (and necessary) in the discrete choice literature to put some structure on the error terms in order to ensure identification (Train, 2009). We also think that including a sufficiently rich set of controls and co-variates allows us to model the dependencies across alternatives in a more transparent manner than if we had left the co-variance structure completely unspecified.

Some more notation will be convenient below. The econometrician observes students' submitted partial rank order lists over schools, $\mathbf{rk}$, and schools' submitted partial priority orderings over students, $\mathbf{pr}$. Following the notation of Fack et al., we denote the observed rank order list of student t as $L_t = (s_t^1, s_t^2, \ldots, s_t^{K_t})$, where $s_t^k \in \mathcal{S}$ is some school. Denote the rank that student t assigns to school s as $rk_t(s)$, with $1 \leq rk_t(s) \leq K_t$ if $s \in L_t$ and $rk_t(s) = \emptyset$ else. The observed rank order lists $\mathbf{rk}$ encompass all individually observed rankings $rk_t(s)$. Similarly, denote the set of students who apply to school s as L_s, and let the the priority number that school s assigns to student t be $pr_s(t)$. Priority numbers are like ranks, in that they take discrete values, and a lower priority number means higher priority. Schools are required to prioritize all students who apply to them, but they may rank some students as "unacceptable". We say that $pr_s(t) = +\infty$ if student t is unacceptable to school s, and $pr_s(t) = \emptyset$ if student t did not apply at school s. Thus, $pr_s(t) \in \{1, 2, \dots, |L_s|, \infty, \emptyset\}$.

Given the specification of the error terms and the observed rankings, equations (4.3) and (4.4) can be regarded as representing two distinct rank-ordered probit models (Train, 2009, p.181). However, the complications outlined in the introductory part of this section imply that an estimation as such is unlikely to succeed in obtaining the true preference parameters. Because schools only rank students who apply to them, and geographical distance is not an admission criterion, we cannot follow the approach of Burgess et al. (2015) to construct the feasible choice set of each student in order to identify her true preferences. For the same reason, the construction of the stability-based estimator that is proposed in Fack et al. (2019) cannot by applied. Still, we follow the main ideas outlined in their paper in that we use a combination of identifying assumptions to identify the model parameters. These will be described in turn. We chose a Bayesian data augmentation approach, owing to its flexibility, and because it allows us to directly estimate the latent variables $\mathbf{U}$ and $\mathbf{V}$ which are our prime objects of interest for the purpose assessing the gains of integration. Similar approaches have been used by Logan et al. (2008) and Menzel and Salz (2013) in the context of one-to-one matching markets. Following Lancaster (2004, p.238), who describes a data augmentation approach for an ordered multinomial probit model, we simulate draws from the posterior density of the structural preference parameters $p(\beta, \gamma|data)$ by considering the component conditionals $p(\mathbf{U}|\beta, \gamma, \mathbf{V}, data)$, $p(\mathbf{V}|\beta, \gamma, \mathbf{U}, data)$, $p(\beta|\gamma, \mathbf{U}, \mathbf{V}, data)$ and $p(\gamma|\beta, \mathbf{U}, \mathbf{V}, data)$. We assume a vague prior for the structural preference parameters γ and β. Details of the conditional posterior distributions are spelled out in B.2.2. Our *data* comprises of the co-variates $\mathbf{X}$ and $\mathbf{W}$, of the assignment μ and of the submitted rank order and priority lists. In general, the Gibbs algorithm to sample for the posterior density can be described as follows:

1. for all t, s: draw $U_t(s)$ from $p(U_t(s)|\beta, \gamma, \mathbf{U}_{-ts}, \mathbf{V}, data) = N(\mathbf{X}_{is}\beta, 1)$, truncated to $[\underline{U}_t(s), \overline{U}_t(s)]$
2. for all s, t: draw $V_s(t)$ from $p(V_s(t)|\beta, \gamma, \mathbf{V}_{-st}, \mathbf{U}, data) = N(\mathbf{W}_{st}\gamma, 1)$, truncated to $[\underline{V}_s(t), \overline{V}_s(t)]$
3. draw β from $p(\beta|\gamma, \mathbf{U}, \mathbf{V}, data) = N\left(b, (\mathbf{X}'\mathbf{X})^{-1}\right)$, with $b = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{U}$
4. draw γ from $p(\gamma|\beta, \mathbf{U}, \mathbf{V}, data) = N\left(g, (\mathbf{W}'\mathbf{W})^{-1}\right)$, with $g = (\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'\mathbf{V}$
5. repeat steps 1–4 N times

Key to our estimation methodology are the truncation intervals for $U_t(s)$ and $V_s(t)$. These intervals are functions of the data and of the latent variables in the model, and they are specific to the particular set of identifying restrictions that is used. The bounds of these intervals could be very tight, or they could encompass the entire real line. We describe the various kinds of identifying restrictions below, and outline how they can be used to construct these truncation intervals; a detailed derivation of the truncation intervals is deferred to B.2.1.

Weak truth-telling (WTT) Weak truth-telling requires that the student truthfully submits his or her top-K_t choices, and that any unranked alternative is valued less than any ranked alternative. Formally, this implies that $U_i(s) \geq U_i(s')$ if (but not only if) $rk_t(s) < rk_t(s')$ or $s' \notin L_t$. That is, any unranked school is assumed to be less preferable than any ranked school. A similar reasoning can be applied to schools' priorities over students, with the difference that a school s cannot rank a student t unless t applies to s. However, a school can label a student as "unacceptable" which implies that all students labelled in this manner are valued less than any other ranked student. So we can bound $V_s(t) \geq V_s(t')$ if $s \in L_t \cap L_{t'}$ and $pr_s(t) < pr_s(t')$ or $pr_s(t') = +\infty$. Taken together, these bounds pin down the truncation intervals and the component conditionals in steps 1 and 2 above.

Undominated Strategies (UNDOM) The assumption of undominated strategies is similar to that of weak truth-telling, but is restricted to the submitted rank order lists. That is, we can bound $U_t(s) \geq U_t(s')$ if $s, s' \in L_t$ and $rk_t(s) < rk_t(s')$. The bounds for the school's valuation over students are the same as in the weak truth-telling case because a school cannot decide to not rank a student; it must at least decide whether the student is acceptable or not. Undominated strategies is thus a weaker, but also more general, condition than weak truth-telling in the sense that the latter implies the former, but not vice versa.

Stability If we assume that the matching of students to schools is *stable* in the sense outlined in section 4.3, a different set of bounds can be applied to the latent valuations. Denote the observed matching as μ such that $\mu(t) = s$ and $i \in \mu(s)$ if student t is assigned to school s. Stability implies that there is no pair of a student t and a school s such that $V_s(t) > \min_{t' \in \mu(s)} V_s(t')$ (so there is no school s that would like to see student t enrolled rather than one of its currently enrolled students) and $U_t(s) > U_t(\mu(t))$ (no student t would prefer being enrolled at s rather than at his current school). This condition implies that we can bound the realization of $U_t(s)$ conditional on the matching μ, and on the match valuations $\mathbf{U}_{-ts}$ and $\mathbf{V}_{-ts}$. Analogous bounds can be placed on $V_s(t)$ with straightforward extensions for cases where schools are not operating at full capacity. These bounds are spelled out in appendix B.2.1 in greater detail. This identifying assumption can be used on its own, or in conjunction with the assumption of undominated strategies.

4.5.2. Identification

Fack et al. (2019) provide an illuminating discussion of the merits of different estimation procedures in the Paris school choice context where students' priorities at all schools are observed by the econometrician, and we draw on their insights below. They argue that the identifying restriction *stability* alone allows for point-identification in large markets as in the Paris setting,[18] but can also be used in

[18]Weldon (2016, p.158) studies identification of preference parameters using stability-based estimators in a large number of small independent matching markets, and concludes that identification depends strongly on the precise parameter configurations of the matching agents.

conjunction with *UNDOM*. While we characterize our estimation approach in the same terms as they do, our setting differs from theirs in that the students' relative rankings at various schools is only incompletely observed. Our preferred identifying assumption is the combination of undominated strategies and stability because it allows point identification, and it guarantees that the observed matching μ is stable under the estimated latent match valuations. The stability property is also convenient because it allows us to replicate the observed matching by computing the SOSM based on priority and preference lists that are computed from the estimated latent match valuations.

The usual conditions on identification in additive random utility models apply, and preference parameters are identified up to the variance of the unobserved random utility component which we restrict to unity. In these models, only utility differences are identified, and so we can identify only up to $J - 1$ alternative specific constants in a choice situation with J alternatives, with one constant being normalized to zero. Moreover, the effect of the decision makers' characteristics are only identified as interactions with alternative-variant characteristics. Furthermore, since only utility differences matter, only the differences of the error terms are identified. This is handled implicitly in our data augmentation approach, by drawing the errors subject to lower and upper bounds that are implied by the observed rank order lists. Lastly, parameters are only identified if there is sufficient heterogeneity in the observed choices: If everyone were to choose the same option, then any parameter which leads to this option being assigned a utility of plus infinity could rationalize what is observed in the data (Cameron and Trivedi, 2005; Train, 2009).

Preference parameters under the identifying restriction of weak truth-telling can in principle be identified by means of a rank ordered model where the choice set encompasses the entire set of schools.[19] However, because students may omit some of their most preferred schools if chances of admission are small, this assumption is often violated and parameter estimates are biased in such a model (Fack et al., 2019). To see this, consider some very popular school s^+ to which chances of admission are so small that most students, although they would rank it first, never actually include it in their submitted ROL. But then, the probability that school s^+ is the most preferred option differs from the probability that it is ranked first, and so the likelihood is misspecified. This may not be a problem at all if the researcher was merely concerned with describing the actual application behaviour of students in an existing school choice problem, but it becomes a problem if one is to study the effects of changing the rules of an existing allocation mechanism. In that case, it seems reasonable to assume that students' true underlying preferences would remain unchanged, but that the changed admission rules would lead them to alter their applications behaviour. Therefore, an analysis that is based on student's true preferences would retain its validity in a counter-factual allocation mechanism, while an analysis based on reported preferences would not be applicable.

The alternative, and weaker, identifying assumption of undominated strategies merely makes a statement about how likely it is for an individual student to prefer school s over school s', given the student's

[19] Variants of this are the rank ordered logit model (Beggs et al., 1981) or a rank ordered probit model (Yao and Böckenholt, 1999). Whereas the rank ordered logit model has analytically tractable expressions for the likelihood, the rank ordered probit model has not, and thus requires simulation or Bayesian estimation techniques.

and the schools' observable characteristics. This probability can be identified non-parametrically from the observed ROLs, conditional on s and s' being part of the submitted ROL, even if some top choices, or some very unattractive alternatives, were omitted due to strategic reasoning. If we assume that the student's decision to include both s *and* s' in her ROL is independent of whether she ranks s or s' higher, then these conditional non-parametric estimates can be matched to the unconditional model-implied probabilities, and hence the model is completely specified. Therefore, the coefficients on alternative-varying covariates can in principle be identified by their relative contribution to the probability that a particular choice s is ranked before an alternative s'. Of course, the usual limitations that apply in multinomial choice models also apply here; for example, preference parameters are only identified up to the scale of the error variance. In this regard we deviate from Fack et al. (2019, p.1507) who argue that an econometric model based on undominated strategies is incomplete in the sense of Tamer (2003), because "the assumption [...] does not predict a unique ROL for the student". From this they conclude that this assumption alone does not permit point identification of preference parameters.

If, in addition, one is willing to make the assumption that the observed matching is stable with respect to the decision makers' true preferences, this stability assumption can serve as an additional source of identification. To illustrate this, consider some school s^- which is so unpopular that only few students have included it in their ROLs. Because of this, the probability that this school is preferred to some other school s' is only poorly identified, and this could lead to large uncertainties in the parameter estimates. But if school s^- has some vacant seats, the stability of the observed matching implies that no other student prefers this school over her currently assigned school. In general, the stability assumption imposes additional bounds on a student's latent match valuation if some school has vacant seats and if the student is matched to another school; or if a school's latent valuation of this student is larger than the least valued student who is currently assigned to that school. Similar considerations apply for the bounds on schools' valuations over students. So, the stability assumption places additional identifying restrictions on the distributions of latent errors and structural parameters.

Measurement error As was described in section 4.4, we do not exactly observe the students' characteristics which the school can condition their admission choices on. Instead, we must rely on supplementary information from the NABC, and we also make use of imputed data because it is important to have a complete set of students for our empirical approach. Thus, the measurement error in our explanatory variables is likely to attenuate our parameter estimates towards zero. Therefore, we may overestimate the contribution of the unobserved idiosyncratic preference and priority shocks to the formation of students' preferences and schools' priorities.

4.5.3. Monte-Carlo evidence

Complementary to the above discussion on identification, we present Monte Carlo evidence below to show that our method works as intended. Specifically, we compare various estimation approaches that are based on different identifying assumptions as laid out above, and we show that a combination of stability and undominated strategies allows us to obtain unbiased parameter estimates with a reasonably small variance.

The data generating process of our Monte Carlo study is borrowed from Fack et al. (2019),[20] but with slight adjustments. We consider markets with $T \in \{100, 200, 500\}$ students and six schools with a total capacity of $0.95 \cdot T$ seats, so there is slight excess demand.Students' utility over schools is given by

$$U_t(s) = \delta_s - d_{ts} + 3 \cdot (a_t \cdot \bar{a}_s) + \epsilon_{ts}$$

where δ_s is a school fixed effect, d_{ts} is the distance from student t to school s, a_t is the students' grade and $\bar{a}_s$ is the average grade of all students at school s (or put differently, the schools' academic quality). Hence, the true preference parameter in the data generating process is a vector $\beta_0 = (1, -1, 3)'$. ϵ_{ts} follows a standard normal distribution. For the exposition, we assume that δ_s is known to the econometrician and therefore enters the estimation as an additional co-variate. The schools' valuation over students (which translates into the students' priorities) is given by

$$V_s(t) = a_t + \eta_{st}$$

where η_{is} is also standard normally distributed. Here, the true priority parameter γ_0 is a scalar equal to one. We subsume all preference and priority parameters as $\theta_0 = (\beta_0', \gamma_0)'$. In the market, students choose their optimal application portfolio, given their equilibrium beliefs about admission probabilities, and a small application cost. This leads some students to skip seemingly unattainable top choices, or to truncate their ROL at the bottom. As a result, the submitted ROLs are likely to violate the assumption of WTT. Based on the simulated submitted ROLs, students and school seats are matched according to the SOSM. We refer the reader to the online appendix of Fack et al. for further details. Our major departure from their approach is the assumption that a student's relative ranking at a school is unknown to the econometrician. Instead, the econometrician only observes the relative rankings of students who applied at school s. Also, normally distributed errors are used on both sides of the market instead of the type-I extreme value distributed errors used by Fack et al..

For our Monte Carlo study, we simulated one hundred independent realizations of these markets. In the simulated markets with two hundred students, a share of 0.69 of the submitted rank order lists satisfied WTT across all simulations.[21] For every sample k, we estimated students' preferences over schools ($\hat{\beta}_k$), and schools' priorities over students ($\hat{\gamma}_k$) using the data augmentation approach described above. The following different sets of identifying assumptions were used to compute the truncation intervals based on the strategically submitted ROLs:

1. weak truth-telling (WTT)

[20] Their data generating process is described, and the code is made available, in their online appendix.

[21] See section 4.5.1. In the market with one hundred students, this share was 0.72, and in the market with five hundred students, it was 0.64.

2. stability
3. undominated strategies
4. stability + undominated strategies

As a benchmark, we estimated the model under the assumption of undominated strategies based on true and complete ROLs.[22] We let the Gibbs sampler run for 20,000 iterations, with a burn-in period of 10,000 iterations. To reduce the parameter estimates' serial correlation, we used only every fifth sample, and discarded the rest.

Figure 4.4 shows box plots[23] of the estimation errors ($\hat{\theta}_k - \theta_0$) across the one hundred realized data sets, for different estimation approaches. Table 4.3 shows the corresponding mean squared error and bias statistics.[24] The first three panels of figure 4.4 depict the distribution of the estimation errors of students' preference parameters ($\hat{\beta}_k - \beta_0$). As expected, the benchmark case where the complete ROLs are known on both sides allows us to identify the parameters very precisely. Furthermore, the estimates for student preferences that are derived under the assumption of weak truth-telling are biased. This too is to be expected because the assumption of weak truth-telling does not hold in the data generating process. When the estimation is conducted using only the stability assumption, the results are very noisy, and also biased. Under the stability assumption, the best estimation results are those for the coefficient on travel distances d_{ts}, but worse results are obtained for the schools' quality δ_s and for the interaction parameter. This is in line with the previous literature on stability based estimators of preferences in small two-sided matching markets. That literature has reached a consensus that the preference parameters are only identified under certain assumptions on the observable characteristics (Weldon, 2016, pp.158-168) or certain preference structures such as perfectly aligned preferences (Agarwal and Diamond, 2014), and may not be identified at all in other circumstances. Note that this is not necessarily at odds with Fack et al. (2019) who argue that a stability based estimator can be used to point-identify preference parameters, for their stability-based estimator is based on the assumption that students' feasible choice sets are known, whereas we assume that this is not the case. The estimates that are derived under undominated strategies are much more precise, but also appear to suffer from a slight bias, which could be a result of the small sample size. Finally, when we combine stability and undominated strategies, our estimates are virtually indistinguishable from the benchmark estimates that are derived using the true and complete ROLs. Interestingly, estimates for the schools' priority function are quite good in all estimation approaches, although the priority lists are only incompletely observed. This insight could lend support to alternative two-step estimators where the schools' priority structure is estimated first, and students' preferences are estimated in a second step, as in He and Magnac (2019).

To confirm that the combination of stability and undominated strategies is indeed able to correct the estimation bias due to strategic reporting, we computed the share of submitted ROLs satisfying WTT

[22]With completely observed ROLs, this is equivalent to the assumption of WTT.
[23]All box plots in this paper are drawn according to the "basic box plot" tyle as in McGill et al. (1978).
[24]B.2.3 presents the same results for $T = 100$ and $T = 500$ students.

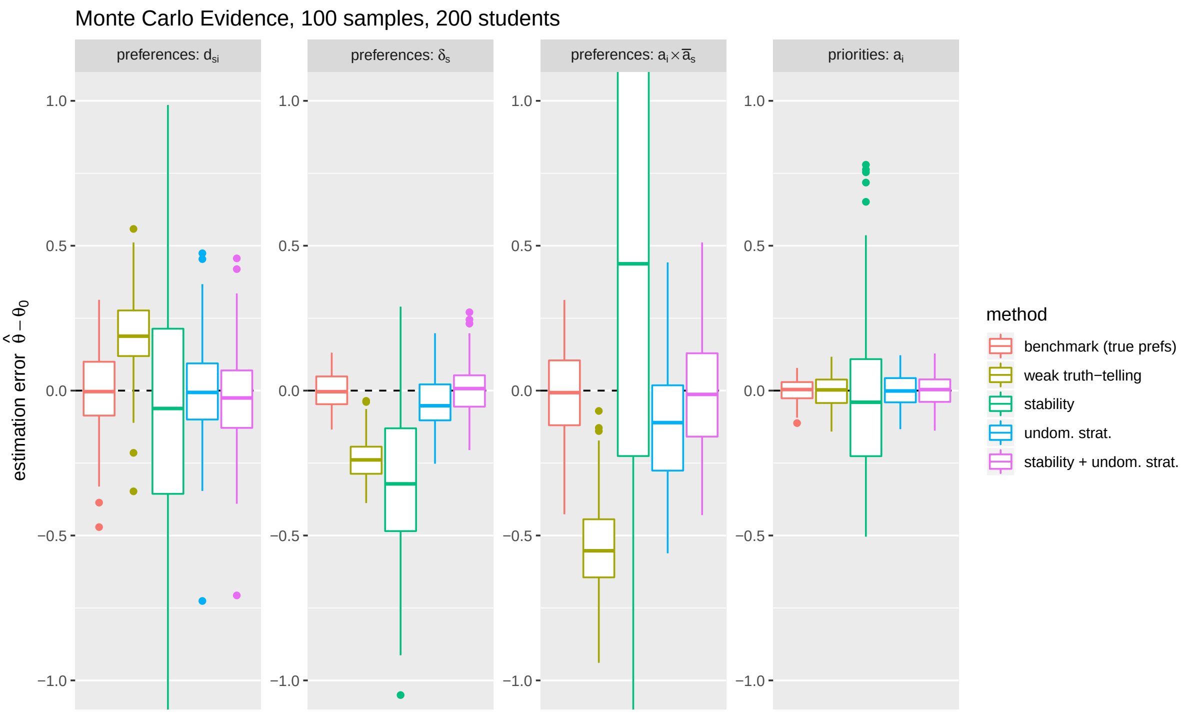


Figure 4.4.: Box plots of the distributions of estimation errors across one hundred simulated markets (six schools with 190 seats and 200 students).

	preferences			priorities
method	d_{is}	δ_s	$a_i \cdot \bar{a}_s$	a_i
benchmark (true prefs.)	0.0187	0.0038	0.0227	0.0016
weak truth−telling	0.0598	0.0581	0.3243	0.0032
stability	0.2903	0.1597	4.8612	0.0788
undominated strategies	0.0338	0.0103	0.0539	0.0030
stability + undom. strat.	0.0323	0.0088	0.0448	0.0030

(a) Mean squared error (MSE)

	preferences			priorities
method	d_{is}	δ_s	$a_i \cdot \bar{a}_s$	a_i
benchmark (true prefs.)	-0.0066	-0.0023	-0.0027	-0.0009
weak truth−telling	0.1937	-0.2302	-0.5425	0.0004
stability	-0.1273	-0.3132	0.9949	-0.0204
undominated strategies	0.0055	-0.0421	-0.1179	0.0001
stability + undom. strat.	-0.0219	0.0134	-0.0183	0.0026

(b) Bias

Table 4.3.: MSE and bias statistics for various estimation methods based on the Monte Carlo simulation described above.

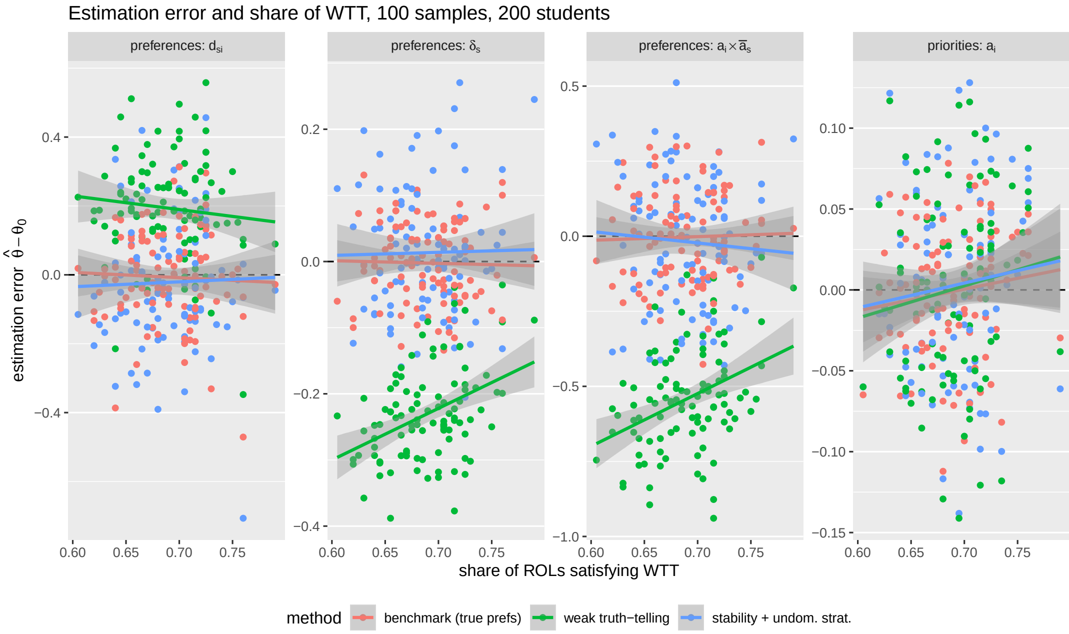


Figure 4.5.: Dependence of the estimation error in different specifications on the share of submitted ROLs that satisfy the WTT assumption. Every dot represents one parameter estimate in one sample market. One hundred simulated markets, six schools with 190 seats, and 200 students.

in each sample market, and we plotted this share against the parameter estimate in that sample. This is done in figure 4.5. Each dot in that figure represents one parameter estimate in one single simulated market, the lines represent the least square estimates, and the shaded areas are the 0.95 confidence intervals around the least square predictions. Table 4.4 shows the corresponding regression coefficients from separate linear regressions of the estimation error on the share of ROLs satisfying WTT, by estimation approach and parameter. Significance is indicated by asterisks. The leftmost three panels of that figure show that the estimation error for students' utility parameters under the WTT assumption decreases in absolute terms as the share of submitted ROLs satisfying WTT increases (green line). On the other hand, the benchmark estimates and the estimates under stability and undominated strategies are not dependent on the share of ROLs that satisfy WTT. For schools' priority parameters, there is no significant relation between either of the estimates and the WTT share, although the point estimates are weakly positive. We conclude from this graph that the proposed estimation approach that relies on a combination of undominated strategies and stability is robust to the strategic submission of preference lists.

method	preferences d_{is}	δ_s	$a_i \cdot \bar{a}_s$	priorities a_i
benchmark (true prefs.)	-0.155	-0.040	0.122	0.134
weak truth–telling	-0.398	0.780***	1.756***	0.199
stability	0.488	0.143	-10.336*	1.670**
undominated strategies	0.157	0.083	-0.307	0.181
stability + undom. strat.	0.143	0.047	-0.381	0.153

p-values indicated by $^{*} < 0.1$; $^{**} < 0.05$; $^{***} < 0.01$

Table 4.4.: Robustness of estimation procedures to violations of the WTT assumption. The table shows the coefficients from separate linear regressions of the estimation error on the share of ROLs satisfying WTT, by estimation approach and parameter. For an estimation approach to be robust to violations of the WTT assumption, the estimation error should not depend on the share of ROLs satisfying WTT.

4.6. Empirical results

This section reports our estimates of the gains from consolidation. First, we present results that are based on the actual submitted preference lists. Next, we present our estimates of students' preferences that are used to construct complete preference lists. These complete preference lists are used to estimate the consolidation gains, circumventing the restrictions that are imposed by the first approach. See figure 4.3 for a brief depiction of our empirical strategy.

4.6.1. Gains from consolidation: using reported preferences

We first approach the problem of estimating the gains from consolidation from a purely descriptive standpoint. To this end, we take the students' submitted rank order lists (ROLs) as given, and re-compute the SOSM under different district consolidation scenarios.[25] As a benchmark outcome, we use the matching in the consolidated market comprising all districts in Budapest. This matching is denoted by μ_{BP} and it is almost identical to the actual matching observed in the KIFIR dataset. This matching is compared to the matching that obtains in a district-level school market (μ_d). For every student, we compare the match rank obtained in the district-level market to the match rank in the benchmark scenario. This difference in match ranks is used as a measure for the consolidation gains. There are two major complications: first, a considerable number of students do not include any school from their home district in their submitted rank order list, and second, some individual school districts actually lack capacity to accommodate all domestic students, despite the fact that there is much excess school capacity in the aggregate. These problems lead to a large number of students not being matched in the counter-factual matching. We assume that these unmatched students would prefer being matched rather than being unmatched, and that the option of being unmatched is as good as the school that they ranked last. In doing so, we obtain a lower bound for the consolidation gains. Because district

[25]For all purposes, we made use of the implementation of the SOSM that is provided as part of the R package *matching-Markets*, available on cran.r-project.org/package=matchingMarkets.

number 23 has only one single school, it does not even offer one school for every track (gymnazium, secondary or vocational). Therefore, we merge this district to its neighbouring district number 20. We show some summary statistics of the district-level and consolidated matches in table 4.5 below.

district markets:	
# matched	6554
share top choice match	0.78
ave. match distance [km]	3.49
consolidated market:	
# matched	10494
share top choice match	0.43
share matched in home district	0.30
ave. match distance [km]	7.10

Table 4.5.: Match statistics of the district-wise and consolidated student-school matching, using reported partial preferences lists.

Table 4.6 contains a detailed account of the consolidation gains per district. That table shows that the vast majority of students is strictly better off in the consolidated market, either because they are assigned to a more preferred school in the consolidated market, or because they are unmatched in the unconsolidated market. In fact, there is not a single district in which more students would prefer the unconsolidated market over the integrated market in Budapest. Motivated by the general insights of corollary 1, figure 4.6 shows how the share of students who strictly gain from consolidation varies along two key dimensions: district size (left panel) and excess capacity (right panel). Figure 4.6a shows that the share of consolidation winners is practically unrelated to district size, but is above fifty percent throughout. The share seems to be negatively correlated with the excess capacity in a district, as shown in Figure 4.6b. To test whether these relationships are significant, we computed a linear regression of the winners' shares per district on the size and relative excess capacity per district. Column (1) in table 4.7 shows that the relationship with a district's size is insignificant, albeit estimated to be negative. The coefficient for a district's capacity is negative and significantly different from zero.

Despite the fact that the share of winners is above fifty percent in all districts, it is by no means clear that district consolidation would also be politically feasible *ex ante*. Our majority share measure is composed of those who strictly gain from consolidation *ex post*. As Fernandez and Rodrik (1991) have noted, *ex ante* uncertainty about the identity of those who gain and those who loose due to a reform induces a bias towards the status quo in majority votes. This bias can effectively prevent the implementation of a reform even when it would be supported by a majority *ex post*. This would be especially true for those districts where the majority share of winners is not so large.

Next, we examine how our theoretical predictions about the distribution of quantitative rank order gains relates to our empirical results. Corollary 1 states that the expected gains from consolidation are larger for smaller markets, and for markets with less capacity. Figure 4.7a shows that there is practically no correlation between district size, as measured by the number of students per district, and

[26] The relative excess capacity in district i is computed as $(N_i - K_i)/N_i$, where N_i and K_i are the number of students, and the total number of school seats in district i, respectively.

district	seats	students	excess seats	−	0	+	unmatched
1	338	95	243	3	9	26	50
2	1191	634	557	36	241	190	148
3	928	743	185	32	263	227	213
4	865	746	119	32	319	241	151
5	625	217	408	5	50	32	122
6	1243	172	1071	14	24	51	77
7	1312	212	1100	14	73	48	70
8	2524	290	2234	11	79	77	119
9	2116	275	1841	19	73	98	77
10	2012	591	1421	45	120	224	194
11	1025	713	312	13	181	169	347
12	956	359	597	17	142	108	90
13	3290	449	2841	44	148	152	100
14	2893	796	2097	52	189	247	291
15	701	454	247	11	99	120	219
16	770	659	111	1	96	162	397
17	147	628	-481	0	40	107	481
18	503	873	-370	17	177	245	432
19	773	444	329	13	68	120	237
20	1643	573	1070	31	157	189	189
21	2518	641	1877	14	258	204	157
22	273	316	-43	7	51	92	165
Total	28646	10880	17766	431	2857	3129	4326

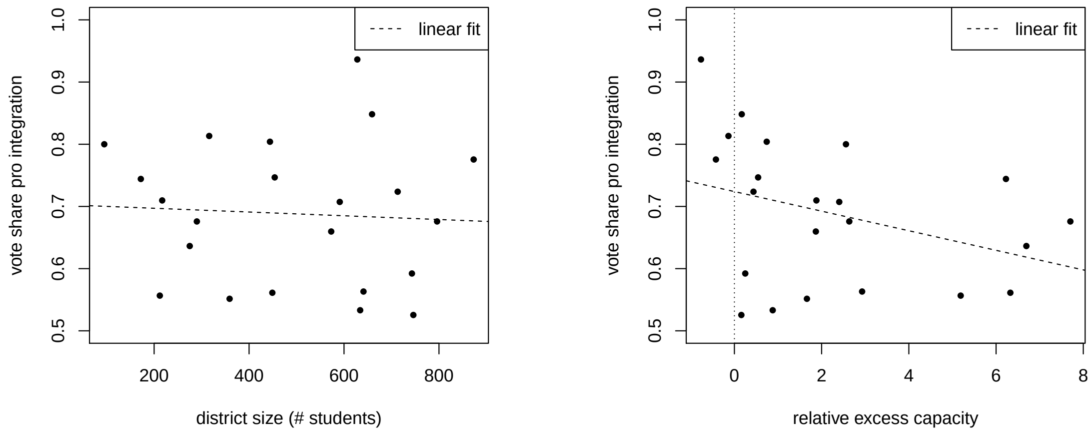
Table 4.6.: Losers (−) and winners (+) from integrating districts in Budapest. seats: number of seats after removing seats given to students from outside Budapest; students: number of students; excess seats: seats − students; −,0,+: number of losers, indifferences and winners from consolidation; unmatched: number of unmatched students. District 23 was merged with district 20 (see text for explanation).

the average rank gains from consolidation. Moreover, panel 4.7b shows that there is a strong negative partial correlation between the average rank order gains, and the districts' excess capacity. Column (2) in table 4.7 contains the estimated coefficients and standard errors from a regression of average rank order gains per district on the size and capacity per district. The table shows that the coefficient for district size is rather small, and also insignificant, whereas the the coefficient for district-level capacity is significantly negative. Therefore, we find robust empirical support for the first part of Corollary 1, but we cannot statistically confirm the validity of the second part.

4.6.2. Preference estimation results

We now turn to the key building block of our structural approach to computing the gains from consolidation. In order to derive the complete preference ordering over schools and students, we estimate a general model of students' preferences and schools' priorities that was described in detail in section 4.5.1. See section 4.4 for an in depth discussion of the data sources.

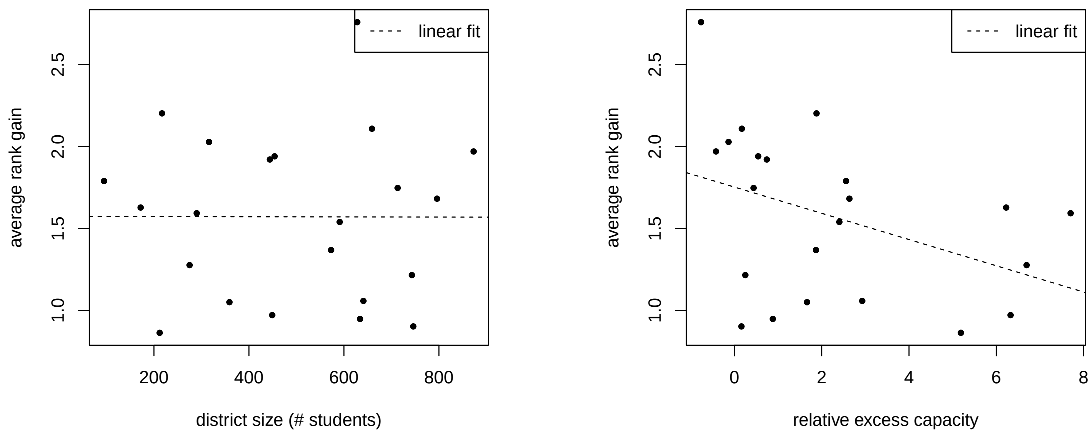
We assume that students' preferences over schools depend on the geographical distance and on the squared distance, between a student's place of residence and the schools' location. To proxy for the



(a) Share of winners and district sizes

(b) Share of winners and excess capacity[26]

Figure 4.6.: Majority support for an integrated market in Budapest, using stated preference lists. One observation denotes one district.



(a) Average rank gains and district sizes

(b) Average rank gains and excess capacity[26]

Figure 4.7.: Rank order gains from an integrated market in Budapest, using stated preference lists. One observation denotes one district.

	Dependent variable:	
	consolidation winners' share	average rank gain
district size (# students)	−0.0002	−0.0008
	(0.0001)	(0.0005)
relative excess capacity	−0.0256**	−0.1191**
	(0.0111)	(0.0478)
Observations	22	22
Note:	*p<0.1; **p<0.05; ***p<0.01	
	Standard errors in parentheses, intercept not shown.	

Table 4.7.: Formal tests for the relationship of consolidation gains to key district statistics.

schools' academic quality, we computed the average of the mean NABC scores in math and Hungarian of students currently enrolled at that school. Also, we computed the average SES score of those students. Finally, we included the interaction terms of the students' math, Hungarian, and socio-economic scores with their respective school-level means in order to test whether there is evidence for assortative matching, similar to what Fack et al. (2019) find. To account for any unobserved heterogeneity across schools, we include school dummies, as we have a rather small set of observable school-level characteristics.[27] We assume that schools select their students based on their gender, math and Hungarian NABC scores, and the SES score. The NABC scores are a proxy for the outcome of a nationwide assessment center, which we do not observe. We estimated a separate set of coefficients for each tier of the Hungarian school system. Our Gibbs sampler was initialized with zero values for all parameters and valuations. Because the estimation procedure is rather time consuming, we let it run for only ten thousand iterations and discarded the first five thousand iterations. To reduce the serial correlation, only every tenths estimate of the remaining five thousand iterations was used so that the posterior means are averaged across five hundred iterations. By visual inspection, we confirmed that the coefficient estimates had converged to their stationary posterior distribution after about two thousand iterations.

The posterior means of the parameter estimates for two different identifying assumptions that were discussed in section 4.5.1 – weak truth-telling (WTT), and stability in combination with undominated strategies – are shown in table 4.8 below and will be discussed in turn. Notice that our Bayesian estimation approach allows us to directly sample from the posterior parameter distribution, so that we do not need to rely on asymptotic results as in conventional estimation approaches. That is why table 4.8 does not include asymptotic p-values but instead shows the 95% confidence intervals of the posterior distribution.

[27] Because we are essentially estimating a discrete choice model over the set of schools, the preference specification cannot include an intercept, as this would not be identified. For the same reason, the first school dummy was omitted lest an intercept is introduced by means of a linear combination of school dummies. In the empirical specification, it turned out that some multicollinearity problems arose even when excluding one school dummy, possibly due to numerical inaccuracies or the presence of interactions. Thus, some more school dummies had to be excluded. To this end, we chose the following approach: In a first step, all fixed effects for schools numbered 2 through to 246 were used to generate a design matrix $\mathbf{X}$ for the problem at hand. In step k, we checked whether the matrix $\mathbf{X}'\mathbf{X}$ had full rank. If not, we dropped one school fixed effect and continued with step $k + 1$, else we stopped. This procedure resulted in a set of fixed effects for the schools numbered 2 through to 243.

First, consider the results of the college selection equation (top panel) across the two identifying assumptions. These results are qualitatively similar to each other: students dislike schools that are further away from them, but the marginal disutility of travelling is *decreasing* because the squared distance term is positive. Students also value academic quality and prefer schools with a higher average SES score, but they dislike the presence of an oral entrance exam. The coefficient for the presence of an entrance exam is much smaller (*i.e.* more negative) in the WTT result: this is an indication that students strategically omit highly competitive schools which hold an entrance exam, so that the WTT estimates of the oral interview are biased downwards, whereas our stability based estimator corrects for this bias. This result confirms how important it is correct for biases due to strategic reporting when estimating students' preferences. The interaction terms are all positive, which suggests that there is sorting on both academic ability and on socioeconomic background. Both estimation approaches yield results that are qualitatively quite similar. Note that the variance of the interacted variables is much larger than that of the school-level variables, so that the interaction terms' contribution towards explaining student preferences is actually quite large.

The results of the student selection equation (bottom panel) show that students' math and Hungarian scores are important variables that schools condition their choices on. Somewhat surprisingly, the female coefficient is negative in the stability + undom. specification, whereas it is very small in the WTT specification. The large negative estimated coefficients for the female indicator is due to the stability requirement: In the data, female students have higher Hungarian scores than male students.[28] At the same time, the Hungarian score is also a key determinant of the schools' priority decision. But in the aggregate, roughly as many female students as male students are admitted to each school, and so the negative female coefficient is needed to ensure that not too many female students form instabilities with school seats occupied by male students.[29] Hence, we think that the negative female coefficient merely reflects the schools' desire to have a balanced gender composition, but it does not indicate discrimination of female students per se. Also, all schools except for vocational schools appear to select on the students' socioeconomic status although the coefficient is rather small compared to the Hungarian score. Yet, in combination with the students' taste for schools with a higher average socio-economic status, and the tendency of students with higher socio-economic backgrounds to prefer schools with a higher average socio-economic status, these results may be indicative of social sorting patterns that could be interesting in their own right.

Contructing complete preference lists In order to obtain complete preference lists for the entire market, we use the estimated coefficients of the student and school selection equations as represented in table 4.8 and combine them with one set of draws from the distribution of error terms that respect

[28] See table B.4 in the appendix.

[29] A quick way to check if this explanation is correct would be to re-estimate the model without the stability bounds. However, we were as of now unable to re-do the analysis due to difficult remote data access conditions.

	stability + undom.		WTT	
Student's selection of schools	$\bar{\beta}$	*95% CI*	$\bar{\beta}$	*95% CI*
distance (km)	-0.148	[-0.152;-0.144]	-0.339	[-0.341;-0.336]
distance (km^2)	0.002	[0.002; 0.003]	0.007	[0.007; 0.007]
academic quality	0.750	[0.681; 0.818]	1.487	[1.458; 1.515]
ave. SES	1.520	[1.411; 1.650]	0.462	[0.418; 0.509]
oral entrance exam	-1.457	[-1.698;-1.240]	-4.436	[-4.587;-4.288]
math × ave. math	0.183	[0.166; 0.196]	0.185	[0.175; 0.195]
hungarian × ave. Hungarian	0.222	[0.205; 0.237]	0.303	[0.293; 0.315]
SES × ave. SES	0.294	[0.279; 0.308]	0.356	[0.347; 0.368]
Schools' selection of students	$\bar{\gamma}$	*95% CI*	$\bar{\gamma}$	*95% CI*
gymnazium				
female	-0.930	[-0.947;-0.909]	-0.013	[-0.040; 0.014]
math score	0.049	[0.033; 0.066]	0.194	[0.171; 0.218]
Hungarian score	0.394	[0.376; 0.413]	0.224	[0.199; 0.249]
SES score	0.038	[0.024; 0.053]	0.096	[0.076; 0.116]
secondary school				
female	-0.439	[-0.481;-0.401]	0.124	[0.089; 0.159]
math score	0.184	[0.163; 0.205]	0.236	[0.208; 0.265]
Hungarian score	0.287	[0.262; 0.315]	0.231	[0.203; 0.259]
SES score	0.053	[0.032; 0.072]	0.103	[0.082; 0.123]
vocational school				
female	0.094	[0.043; 0.158]	0.051	[-0.031; 0.131]
math score	0.101	[0.063; 0.136]	0.078	[0.025; 0.129]
Hungarian score	0.189	[0.152; 0.226]	0.144	[0.089; 0.200]
SES score	0.011	[-0.023; 0.044]	0.015	[-0.020; 0.051]

Table 4.8.: Posterior means of preference and priority parameters under two different identifying assumptions (see section 4.5.1). Fixed effects for schools numbered 2 through to 243 were included in students' preference equation, and are not reported here. Academic quality is the average of the school-level averages for the Hungarian and math scores. Confidence intervals from the posterior parameter distribution of the Gibbs sampler.

the upper and lower bounds derived from stability and imposed by submitted preference lists. Thus, the estimated utility for student i visiting school s is

$$\hat{U}_t(s) = \mathbf{X}_{ts}\bar{\beta} + \hat{\epsilon}_{ts}$$

where $\hat{\epsilon}_{ts}$ is one particular realization of the latent error distribution such that $\hat{U}_t(s)$ respects the bounds that are imposed by the identifying assumptions. This estimated latent utility comes straight from the Gibbs sampler. Schools' latent match utilities are constructed analogously. These estimates of the latent valuations can then be used to construct, for each market participant, a complete preference ordering of the other market side. Note however, that every such set of valuations is only one particular draw from an infinite manifold of possible realizations. Currently, we only use a single realization of the valuations, and we believe that the large market size validates this approach.

4.6.3. Gains from consolidation: using estimated preferences

We now repeat the analysis of section 4.6.1 above, but using the complete rank order lists described above. Again, we compare the outcome of a consolidated city-wide match to the district-level matching scheme. Instead of the rank order gains, we computed the average gains in latent utility. For a student t, this is defined as the utility difference between visiting the assigned school in the consolidated market, $\mu_{BP}(t)$, and the assigned school in the district level market, $\mu_d(t)$:

$$\Delta U_t \equiv \hat{U}_t(\mu_{BP}(t)) - \hat{U}_t(\mu_d(t))$$

Utility is a unitless quantity which is hard to interpret per se, but our utility specification allows us to express these gains in terms of travel distances:

$$\Delta U_t^{km} \approx \frac{\Delta U_t}{\left|\frac{\partial \hat{U}_t(\mu_{BP}(t))}{\partial d_{t\mu_{BP}(t)}}\right|},$$

with $d_{t\mu_{BP}(t)}$ being the travel distance between student t's zip code of residence and her assigned school in the consolidated market.[30] Because students dislike utility, we use the absolute value in the denominator, so that $\Delta U_t^{km} > 0$ corresponds to a positive welfare gain due to market consolidation. Therefore, ΔU_t^{km} is a measure of the additional travel time that a student would be willing to incur in order to visit the school in the consolidated market, rather than the assigned school in the district level school market. As before, we merge district 23 to its neighbouring district number 20. As a robustness

[30]Because distance travelled enters the utility specification quadratically, it matters in principle whether the partial derivative is evaluated at the district level matching μ_d, or at the integrated matching μ_{BP}. However, the estimated quadratic term is very small (see table 4.8), which allows us to use the following approximation:

$$\left|\frac{\partial \hat{U}_t(\mu_{BP}(t))}{\partial d_{t\mu_{BP}(t)}}\right| = \left|-0.148 + 2(0.002 d_{t\mu_{BP}(t)})\right| \approx 0.148.$$

Hence, one utility unit is approximately worth seven kilometres of avoided travel distance.

district markets:	
# matched	9,986
share top choice match	0.83
ave. match distance [km]	3.55
integrated market:	
# matched	10,880
share top choice match	0.66
share matched in home district	0.29
ave. match distance [km]	7.14

(a) Assignment statistics of the district-wise and integrated student-school matching.

	Mean	SD	Min	Median	Max	N
total gains						
in latent utility units	0.819	0.916	-1.895	0.600	5.799	9,986
in equivalent kilometres	5.532	6.187	-12.805	4.054	39.180	9,986
decomposition						
choice effect I	0.750	0.798	0.000	0.548	5.010	10,880
competition effect I	0.103	0.590	-3.655	0.000	5.352	9,986
choice effect II	0.865	0.899	0.000	0.663	5.799	9,986
competition effect II	-0.040	0.295	-3.000	0.000	2.000	10,880

(b) Various measures of consolidation gains, expressed in latent utility changes.

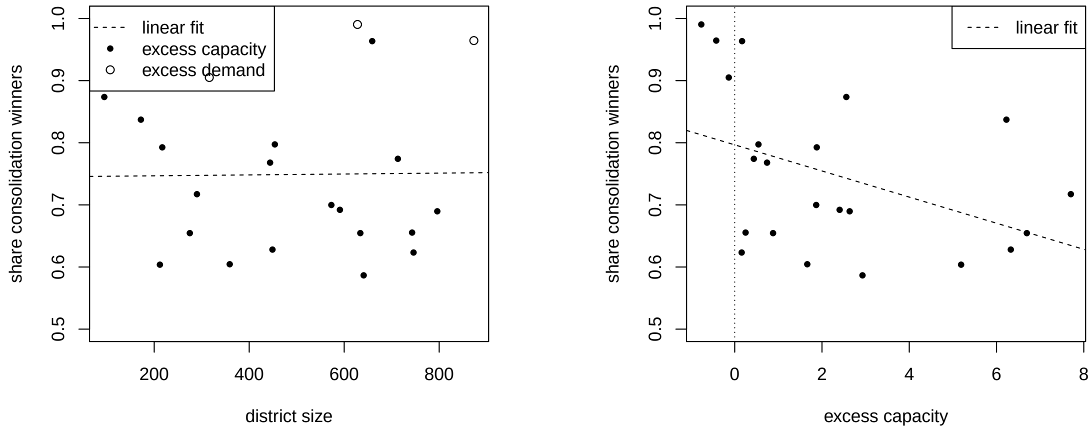
Table 4.9.: Gains from consolidation using inferred complete preferences lists

check, we also conducted the same analysis with artificially balanced markets where the number of school seats was equal to the number of students in every district. Those results are reported in B.2.6.

Table 4.9a shows some summary statistics of the resulting district-level and consolidated, city-wide matchings. The table shows that some students remain unmatched in the district-level matching. This is because the school market in Budapest has been an integrated one for a long time already, so some districts do not have enough school seats to accommodate all students of their own district. In the consolidated market, all students are matched because there is enough capacity in the the aggregate, and because preference lists are complete.

The first row in table 4.9b shows summary statistics of the consolidation gains ΔU_t. Because not all students are matched in the unconsolidated market, those gains cannot be computed for all students. The average gains are positive, but some students also lose due to market consolidation. However, the median is positive so that the majority of all students gain. The second row of that table shows the utility gains, converted to distance units ΔU_t^{km}. It shows that the average student's gains are equivalent to saving more than five kilometres in travel distances, even though students actually incur longer travel distances in the consolidated market, as table 4.9a shows. Accordingly, the utility gains greatly outweigh the additional travel distances that are incurred in the consolidated market.

As in section 4.6.1, we now ask whether market consolidation can be decided upon unanimously if every district had one vote, and if those votes were bound to reflect the majority view in those districts. It is assumed that students who are unmatched in the district-level matching prefer the consolidated matching. Of course, this is an ex post perspective, as was already discussed in section 4.6.1. Figure



(a) Share of winners and district sizes

(b) Share of winners and excess capacity[26]

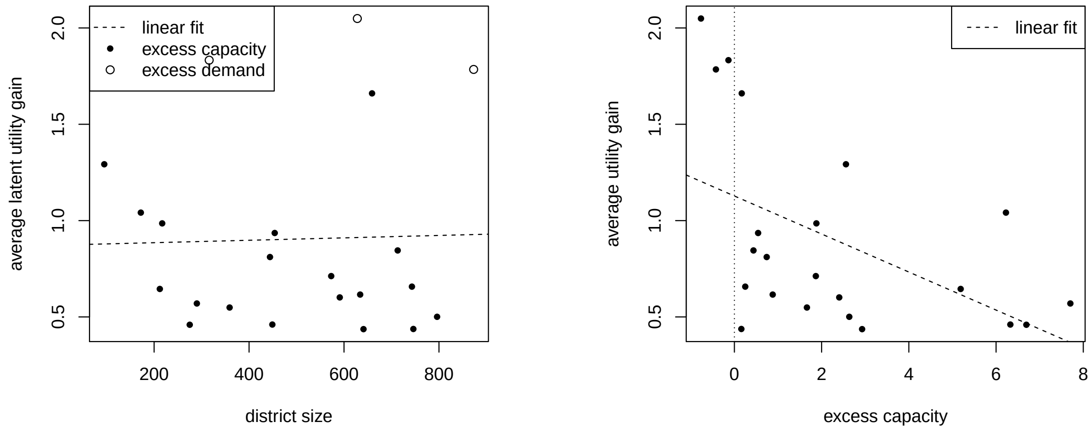
Figure 4.8.: Majority support for an integrated market in Budapest, using inferred complete preference lists. One observation denotes one district.

4.8 shows that a majority of all students in every district strictly prefers the consolidated market over the disintegrated market. The left panel of that figure shows that there is no correlation between the majority shares and the district sizes, and the right panel of that figure shows a strong negative correlation between the majority shares and the relative excess capacities, by district.

Next, we relate the average consolidation gains in latent utility units per district to two key district characteristics, size and capacity. Figure 4.9 shows that there is a weakly positive correlation between the average utility gains and district size, and a negative correlation between average gains and district-level excess capacity. A test based on a regression of district-level average gains on district characteristics is reported in table 4.10 and shows that both the district size (as measured in hundreds of students) as well as the district capacity (as measured by relative excess capacity) have a negative partial effect

	Dependent variable:	
	share consol. winners	ave. latent util. gain
district size (100 students)	−0.0190	−0.0850
	(0.0135)	(0.0499)
relative excess capacity	-0.0306^{**}	-0.1417^{***}
	(0.0120)	(0.0442)
Observations	22	22
Note:		*p<0.1; **p<0.05; ***p<0.01
		Standard errors in parentheses, intercept not shown.

Table 4.10.: Formal tests for the relationship of consolidation gains to key district statistics, using complete rank order lists. One observation corresponds to one district.



(a) Average utility gains and district sizes

(b) Average rank gains and excess capacity[26]

Figure 4.9.: Average latent utility gains (ΔU_t) of an integrated market in Budapest, using inferred complete preference lists. One observation denotes average statistics in one district.

on the average gains in latent utility, but only the marginal effect of district capacity is significantly different from zero. Qualitatively, these results are in line with parts one and two of Corollary 1 in section 4.3. But the graphical results as well as the lack of significance for the effect of district size show that these postulated relationships are quite noisy. This can be explained by the fact that the theoretical results were derived under the stark assumption of random preferences on both sides of the market. But the previous subsection has just revealed the opposite, namely that preferences systematically depend on market observables. It is therefore quite understandable that the district level results exhibit a considerable amount of variability that cannot be explained by theory alone.

Decomposition of the utility gains As we write in the theoretical section, district consolidation has two effects on students' welfare: first, it leads to more choice, which is unambiguously good, and second, it may increase or decrease competition. Increased competition means that it becomes more difficult for a given student to be admitted to his or her favourite schools. Whether competition increases or decreases depends on many factors: If the schools in some sub-market are very attractive, or if this market is not as tight as the aggregate market (from the students' perspectives), then district consolidation will lead to more competition, so that domestic students may be hurt. The composition of choice and competition effects may help to explain the large utility gains from consolidation that we find. In order to explain these gains, we isolate the effects of choice and competition in a decomposition exercise. The idea is to keep an individual student t fixed, and assign her to the most preferred feasible school, given that all other students are either restrained to attend only local schools, or may attend any school in the integrated market. The competition effect is then the change of student t's welfare as all other students' choice sets are enlarged to include the entire integrated market. Similarly, the

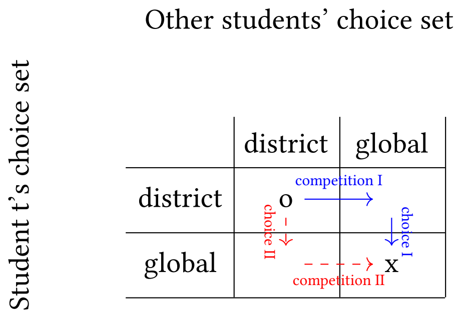


Figure 4.10.: Decomposition of the gains from market consolidation into choice and competition effects

choice effect is the change of that student t's welfare as her choice set is expanded to include all schools, keeping the other students' choice sets constant. This is repeated for all students, and the results are aggregated. The idea is illustrated in figure 4.10, and more details on the procedure can be found in B.2.5. As this figure shows, there are always two ways to measure either the choice, or the competition effect. We shall refer to the resulting statistics as type-I and type-II effects.

Table 4.9b shows summary statistics of the choice and competition effects that are calculated in both ways. In general, the sum of the competition and choice effects of either type should be equal to the total welfare effect of consolidation. However, because not all students are assigned to a school in the district level matching (c.f. table 4.9a), the type-I competition effect and the type-I choice effect cannot be computed for all students. But this affects only very few students and so the average choice gains and the average competition effects approximately add up to the total gains. The results show that the choice effects account for the vast share of total welfare gains, while the average competition effects are much smaller in magnitude, and vary in sign. Whereas the average type-II competition effect is small and negative, the type-I competition effect is small and positive. Therefore, it must remain an open question whether competition is stronger in the consolidated market, or in the district-level markets.[31] The fact that the competition effects are so small in magnitude is probably related to the fact that the Hungarian school market is characterized by much excess capacity, as was already discussed elsewhere in this paper. And so an integrated market leads to large welfare gains due to increased choice, but increases the competitive pressure by only a small amount.

In order to further explain the gains from market consolidation, we related the student-level gains, and the competition and choice effects that were computed above, on student- and district-level observables. Table 4.11 shows the results of this linear regression analysis. The coefficients describe a "consolidation premium" that can be ascribed to various observable student characteristics. The results for the type-I

[31] At first glance, it may seem counter-intuitive that competition could be weaker in the consolidated, aggregate market. But this can be explained by the fact that the school districts are very different. A few districts have a large number of school seats that far exceeds the number of their domestic students (see table 4.6). While the market tightness increases for students in those districts as all districts are integrated, the aggregate market tightness may decrease as a result. Therefore, the majority of students may experience more favourable competitive competitions in the aggregate market.

and type-II decomposition are similar, and so we discuss only results related to the type-I choice and competition effects. The first column of this table shows that students with a higher socio-economic status (SES) benefit *relatively* more from district consolidation. The italicised adverb is important because all students benefit on average, but some students benefit more than others. However, the effect is estimated imprecisely and so one cannot reject the null hypothesis of there being no effect at all. The second and third columns reveal that this is due to the fact that students with a higher SES benefit more from increased choice, but benefit less from more favourable competitive conditions in the consolidated market.[32] Again, these effects are insignificant although the effects in the fourth and fifth column that are related to the type-II effects would indeed be significantly different from zero. A similar, but exacerbated pattern can be observed for students with higher academic ability. High-ability students benefit more from district consolidation than average students, and they benefit comparatively more from an enhanced choice set, and less from more relaxed competitive conditions in the aggregate. These effects are statistically significant. Students in larger districts, or in those districts with a lot of excess capacity, benefit significantly less than other students. This is consistent with the predictions of corollary 1 and with the district-level findings reported in table 4.10. Contrary to what we find in that table, the negative effect of district size on the consolidation gains is now estimated to be significantly different from zero.

The results imply that there is a consolidation premium for high-ability students, and possibly for students from a higher socio-economic background. As table 4.1 shows, the explanatory variable SES is standardized and has unit variance, whereas the variance of "ability" is about 1.5. Because the estimated coefficient in table 4.11 is also larger for "ability" than for SES, it follows that an increase in student ability by one standard deviation increases the consolidation premium by about $1.5 \times 0.014 \approx 0.021$ utility units, whereas an increase of the socio-economic status indicator by one standard deviation increases the consolidation premium by only 0.009. So besides being insignificant, the estimated effect of a higher socio-economics status on consolidation gains is also much less relevant. Thus it appears that the highly selective consolidated Hungarian school system benefits high-ability students more than those from higher socio-economic background, if the latter benefit at all. But of course, there are some caveats to this conclusion. First, the variables measuring SES and student ability are highly correlated ($r = 0.47$) and so there will be a large overlap of high-SES and high-ability students among those who benefit a lot from market consolidation. Second, the overall effects are rather small compared to the total variance of the consolidation gains, which is close to one (see table 4.9b). On that account, the systematic factors driving the consolidation gains are rather small, and idiosyncratic factors seem to be the most important determinants (but note our comment on measurement error in section 4.5.2).

[32] Recall that the type-I competition effects are positive on average.

	Dependent variable: latent utility gains				
		type-I decomposition		type-II decomposition	
	total	choice	competition	choice	competition
	(1)	(2)	(3)	(4)	(5)
socio-economic status SES	0.0085	0.0141	-0.0067	0.0187**	-0.0097***
	(0.0097)	(0.0086)	(0.0062)	(0.0095)	(0.0034)
ability	0.0143**	0.0443***	-0.0238***	0.0267***	-0.0139***
	(0.0070)	(0.0062)	(0.0045)	(0.0069)	(0.0025)
district size (100 students)	-0.1408***	-0.2106***	0.0692***	-0.1443***	0.0038
	(0.0172)	(0.0160)	(0.0110)	(0.0169)	(0.0063)
relative excess capacity	-0.3101***	-0.0457**	-0.2720***	-0.2958***	-0.0116
	(0.0246)	(0.0221)	(0.0157)	(0.0242)	(0.0087)
school type: gymnazium	-0.1308***	-0.2487***	0.0785***	-0.1544***	0.0368***
	(0.0352)	(0.0311)	(0.0225)	(0.0346)	(0.0123)
school type: secondary	-0.0288	-0.0670**	0.0134	-0.0425	0.0188
	(0.0331)	(0.0292)	(0.0211)	(0.0326)	(0.0115)
Constant	2.2938***	1.9240***	0.4160***	2.2997***	-0.0215
	(0.0767)	(0.0691)	(0.0490)	(0.0755)	(0.0272)
district FE	Yes	Yes	Yes	Yes	Yes
Observations	9,986	10,880	9,986	9,986	10,880
Note:					*p$<$0.1; **p$<$0.05; ***p$<$0.01 Standard errors in parentheses.

Table 4.11.: Explaining gains from consolidation with students observables. The table shows regression coefficients of students' gains on student observables. 'ability' is a composite variable (see table 4.1 and section 4.4). Variables 'district size' and 'relative excess capacity' refer to the students' home districts; the school type refers to the school type of the assigned school in the integrated market.

4.7. Conclusion

We have analysed the effects of school market consolidation theoretically under the random markets assumption and empirically by means of a structural preference model. The theoretical predictions have shown that market consolidation leads to substantial welfare gains for students, and that students who live in smaller markets, or in markets with fewer available school seats, are expected to have larger welfare gains. Our empirical results confirm that the average student greatly benefits from having a consolidated school market, and that more than half of all students are better off in the consolidated school market. We find that the gains from consolidation are larger in school districts which have very little capacity compared to the number of students, and in smaller districts. By and large, these results are independent of whether students' stated preferences are used, or whether an inferred complete preference ranking is used. Moreover, our results indicate that high-ability students benefit more from market consolidation than do other students. It is important to note that our empirical results are derived in a school market having much excess capacity, and do not necessarily carry over to markets which have just enough capacity for all students.

As a by-product, we have established a method to consistently estimate students' preferences in school markets with school-specific admission criteria unknown to the researcher. Our estimation approach avoids a bias that is otherwise introduced by students' strategic reporting of their preferences, and we showed by means of a Monte Carlo study that this method works as intended. We find – perhaps unsurprisingly – that students favour nearby schools which have a high academic and social reputation, but dislike having to write dedicated school-specific entrance exams. We also find that there is evidence for sorting according to academic ability, and social status. Schools appear to base their admissions mostly on the students' abilities in Hungarian, with math scores and socio-economic background being less important.

We have computed the consolidation gains under the assumption that the students' and schools' characteristics remain fixed throughout, while only the admission system is changed. Thus, our results should be interpreted as measuring the isolated, or partial effect of the admission system on students' welfare. We think that we can accurately describe and measure this partial effect, and that it is a valuable statistic in itself that informs the debate on the merits of centralized assignment mechanisms. But of course, there are other effects that could be taken into account. Recall that the status quo, and the starting point of our analysis, is the completely consolidated school market in Budapest, so that the gains from consolidation are more accurately described as hypothetical losses from market disintegration. But if that school market were to be disintegrated, then both students and schools could probably react in unforeseeable ways, and this could attenuate the losses of disintegration and, conversely, reduce the gains from consolidation. For instance, schools could increase their capacity, but they could also increase the diversity of their educational profile in response to the changed environment. Also, the unobservable component to schools' attractiveness that we subsume in a fixed effect for each school could change as a result, so that the students' preference orderings may actually change, thus leading to a different counter-factual assignment. It could appear to the reader that one could estimate these second order consequences of district consolidation by means of an iterative procedure whereby the schools' average academic qualities, and students' preferences, are updated in turns until a "steady state" is reached. But in our opinion, such a mechanistic steady state analysis is unlikely to mirror the multitude of individual and institutional responses, and would thus be rather speculative. Therefore, we refrained from this approach, focusing on what we can measure, and not on what we cannot measure.

Our results contribute to the growing literature on school market consolidation, and its effects on student welfare. If the aggregate school market has excess capacity, then a consolidated school market probably leads to large welfare gains that benefit substantially more than half of all students. Intuitively, students greatly benefit from an expanded choice set, while the competitive pressure does not increase by very much. On the other hand, our supplementary analysis in B.2.6 shows that if the school market as a whole is about balanced, with just enough capacity to accommodate all students, then a consolidated market does not increase students' welfare very much and the median student neither gains nor loses due to market consolidation. The reason is that the benefits of an expanded choice set in the consolidated market are largely offset by increased competitive pressure. High-ability students benefit most from school market consolidation, which is presumably due to a rather competitive assignment system that allows those students to attend the best schools in an increased choice set. Students with a high socio-economic background also benefit relatively more, but less so than high-ability students.

The analysis that we presented here is concerned with the consolidation of formerly independent deferred acceptance school markets into one large deferred acceptance school market. And so our results are not applicable to an introduction of a school choice system where no such system has been in place before. Also, the individual welfare effects may of course run counter to the objectives of a social planner who would also be concerned about inequality and segregation. We do not study these topics, but our approach to estimating student's preferences could in principle be useful to analyse the interplay of choice and institutions that determines these distributional aspects more closely.

Our results also have implications for the theoretical research on matching markets. As our estimates show, students' and schools' preferences and priorities over each other have an observable and an unobservable idiosyncratic component. Thus, theoretical results that rely on uniformly and randomly generated preferences may be an inappropriate tool to describe a real-world matching market, and future theoretical research should incorporate or explore this aspect in greater detail.

B. Appendix to chapter 4

B.1. Proofs

The proof of proposition 2 is as follows:

Proof. Combining two existing results, we can show that in random ESCPs

$$\mathrm{rk}_T(\sigma_{\text{SOSM}}(\cdot,\Omega)) \approx \frac{N+K}{qN}\log\left(\frac{N+K}{K}\right)+1$$

We obtain the expression above by combining two known properties of matching markets: i) each many-to-one matching market with responsive preferences has a corresponding one-to-one matching market (lemma 5.6 in Roth and Sotomayor, 1992), and ii) the students' absolute average rank of schools in random one-to-one matching markets can be approximated by $\frac{N+K}{N}\log(\frac{N+K}{K})$ (theorem 2 in Ashlagi et al., 2017).[1] This approximation maps remarkably well the simulation for many-to-one markets in Table 4 in Ashlagi et al. (2017). For example with $N = 198, K = 2, q = 5$, their simulations give a rank of 1.9 whereas the approximation gives 1.93. We emphasize that our approximation only works for relatively small values of q; when q is large instead then there is a large probability that each agent will be assigned to his most desired school, and thus $\mathrm{rk}_T(\sigma_{\text{SOSM}}(\cdot,\Omega)) \approx 1$.

To compare the gains from consolidation, we only need to approximate $\mathrm{rk}_T(\sigma_{\text{SOSM}}(\cdot,D))$. To do this, we define the *relative rank* of a school s in the preference order of a student $t \in T^{D_i}$ (over potential schools in within his own district) as $\hat{\mathrm{rk}}_t(s) := \left|\{s' \in S^{D_i} : s' \succcurlyeq_t s\}\right|$. Given a matching μ, the *students' relative average rank of schools* is defined by

$$\hat{\mathrm{rk}}_T(\mu) \quad := \quad \frac{1}{|\overline{T}|}\sum_{t\in\overline{T}}\hat{\mathrm{rk}}_t(\mu(t))$$

where $\overline{T}$ is the set of students assigned to a school under matching μ.

[1] Ashlagi et al. (2017) prove that for any stable matching, the following inequalities hold with high probability: $(1-\epsilon)\,\frac{N+K}{N}\log(\frac{N+K}{K}) \le \mathrm{rk}_T(\mu) \le (1+\epsilon)\,\frac{N+K}{N}\log(\frac{N+K}{K})$.

In a district with qn_i students, $q(k_i+n_i)$ school seats and with $k_i > 0$, we can approximate the students' relative average rank of schools (using the same tools as before) as

$$\hat{\text{rk}}_T(\sigma_{\text{SOSM}}(\cdot, D)) \approx \frac{n_i + k_i}{qn_i} \log\left(\frac{n_i + k_i}{k_i}\right) + 1 \tag{B.1}$$

whereas in a district with $k_i < 0$, the approximation becomes

$$\hat{\text{rk}}_T(\sigma_{\text{SOSM}}(\cdot, D)) \approx \frac{n_i + k_i}{1 + \frac{n_i}{n_i+k_i} \log\left(\frac{n_i}{k_i}\right)} \tag{B.2}$$

The final step in the proof closely follows the proof of Proposition 3 in Ortega (2018). To relate the students' relative average rank of schools before consolidation to the absolute ranking, suppose that a school is ranked h among all schools in its district. A random school from another district could be better ranked than school 1, between schools 1 and 2, ..., between schools $h-1$ and h, ..., between schools n_i+k_i-1 and n_i+k_i, or after school n_i+k_i. Therefore, a random school from another district is in any of those gaps with probability $1/(n_i+k_i+1)$ and thus has $h/(n_i+k_i+1)$ chances of being more highly ranked than our original school with the relative rank h. There are $N+K-n_i-k_i$ schools from other districts. On average, $\frac{h(N+K-n_i-k_i)}{n_i+k_i+1}$ schools will be ranked better than it. Furthermore, there were already h schools in its own district better ranked than it. This implies that his expected ranking is $h + \frac{h(N+K-n_i-k_i)}{n_i+k_i+1} \approx \frac{h(N+K)}{n_i+k_i}$. Substituting h for (B.1) and (B.2), respectively, we obtain students' relative average rank of schools before consolidation. After some algebra any by eliminating the constants (which are irrelevant in large markets), it follows that

$$\gamma_T(\sigma_{\text{SOSM}}) \approx \frac{N+K}{q}\left(\frac{\log(\frac{n_i+k_i}{k_i})}{n_i} - \frac{\log(\frac{N+K}{K})}{N}\right)$$

if $k_i \geq 0$, and

$$\gamma_T(\sigma_{\text{SOSM}}) \approx \frac{N+K}{q}\left(\frac{q(n_i+k_i)}{n_i \log(\frac{n_i}{k_i})} - \frac{\log(\frac{N+K}{K})}{N}\right)$$

if $k_i < 0$. □

B.2. Supplementary material

B.2.1. Explicit computation of the bounds on latent valuations

The estimation procedure relies on imposing upper and lower bounds on the latent valuations. This section describes explicitly how these bounds can be computed at every step of the estimation procedure, under various identifying restrictions. For convenience, we repeat the notation that is used to describe students' and schools' ordinal preferences and priorities here.

We denote the observed rank order list of student i of length L_t as $\mathcal{L}_t = (s_t^1, s_t^2, \ldots, s_t^{L_i})$, where $s_t^k \in S$. Denote the rank that student t assigns to school s as $rk_t(s)$, with $1 \leq rk_t(s) \leq L_t$ if $s \in \mathcal{L}_t$ and $rk_t(s) = \emptyset$ else. Collect all observed ranks into $\mathbf{rk} = \{rk_t(s)\}_{t \in T, s \in S}$. The preference orderings induced by these observed ranks are a subset of students' unobserved strict preference ordering $\succ = \{\succ_t\}_{t \in T}$, i.e. $rk_t(s) < rk_t(s') \Rightarrow s \succ_t s'$ but not vice versa, because students may find it optimal to not rank all schools if the application procedure is costly. This is the „skipping at the top" and „truncation at the bottom" problem that was discussed in the main text and that precludes the application of standard revealed preference arguments to estimate a reduced-form model of students' preferences.

Similarly, denote the set of students who apply to school s as $\mathcal{L}_s$, and let the the priority number that school s assigns to student t be $pr_s(t)$. Priority numbers are like ranks, in that they take discrete values and a lower priority number means higher priority. Schools are required to prioritize all students who apply to them, but they may rank some students as "unacceptable". We say that $pr_s(t) = +\infty$ if student t is unacceptable to school s, and $pr_s(t) = \emptyset$ if student t did not apply at school s. Furthermore, denote the set of ranked students that are acceptable to school s as $\ell_s = \{t \in \mathcal{L}_s : pr_s(t) < \infty\}$ and define the largest priority number of any school s as $\overline{pr}_s = \max\{pr_s(t) : t \in \mathcal{L}_s\} \in \{|\ell_s|, \infty\}$. Thus, $pr_s(t) \in \{1, 2, \ldots, |\ell_s|, \infty, \emptyset\}$. The set of all observed priority rankings is given by $\mathbf{pr} = \{pr_s(t)\}_{t \in T, s \in S}$. Again, the priority structure induced by pr_s is a subset of the unobserved true priority ordering $\rhd = \{\rhd_s\}_{s \in S}$.

Because the bounds depend on the observed ranks and priorities, but also on the latent valuations of students and schools, they must be computed anew in every iteration of the Gibbs sampler. More concretely, the vector of latent utilities at the current iteration step k is constructed as

$$\mathbf{U}_{ij}^{(k)} = \begin{cases} \mathbf{U}_{ij}^{(k)} & \text{if the pair } ij \text{ has been visited in iteration } k \\ \mathbf{U}_{ij}^{(k-1)} & \text{else.} \end{cases}$$

An analogue updating scheme is used to construct the vector of latent valuations $\mathbf{V}$. This Gauss-Seidel style updating scheme ensures that, at any point in the iteration scheme, the upper and lower bounds are satisfied for the entire vector of latent utilities and valuations, but it comes at a higher computational burden. The alternative would be to compute upper and lower bounds once in every iteration k, using only the last estimates of the latent utilities $\mathbf{U}_{ij}^{(k-1)}$. In what follows, we will omit the index of the current iteration round k, and assume that any reference to $\mathbf{U}_{is} = U_i(s)$ is made with respect to the most recent available estimate of $U_i(s)$, either from iteration k or from iteration $k-1$.

Lastly, we will in the following exposition use the order $>$ on the set of ranks, or priorities. Since either a rank $rk_t(s)$ or a priority $pr_s(t)$ can take the value $\emptyset$, it is necessary to define the behaviour of this operator with respect to $\emptyset$: we will assume that the statement $a > \emptyset$ is false for all values of a, whereas $a \geq \emptyset$ is true if, and only if, $a = \emptyset$. Also, as a convention, the minimum of an empty set returns ∞ and the maximum of an empty set returns $-\infty$.

Weak truth-telling (WTT)

Having clarified the notation, we now turn to describe how upper and lower bounds implied by the weak truth-telling assumption (WTT) are constructed. WTT posits that, on the side of the students, any unranked alternative school $s : rk_t(s) = \emptyset$ is worse than any ranked alternative s' with $rk_t(s') \neq \emptyset$. Given latent valuations $\mathbf{U}_{-it}$, and observed ranks $\mathbf{rk}$, the upper and lower bounds for utility $U_t(s)$ can be expressed as follows:

$$\overline{U}_t(s) = \begin{cases} +\infty & rk_t(s) = 1 \\ \min_{s' \in \mathcal{L}_t}\{U_t(s') : rk_t(s') < rk_t(s)\} & rk_i(s) > 1 \\ \min_{s' \in \mathcal{L}_t}\{U_t(s')\} & rk_i(s) = \emptyset \end{cases}$$

$$\underline{U}_t(s) = \begin{cases} \max_{s' \in \mathcal{L}_t}\{U_t(s') : rk_t(s') > rk_t(s)\} & rk_t(s) < L_t \\ \max_{s' \notin \mathcal{L}_t}\{U_t(s')\} & rk_t(s) = L_t < |S| \\ -\infty & rk_t(s) = \emptyset \wedge rk_t(s) = |S| \end{cases}$$

In our setting, schools only get to see those students who apply to them and hence, $pr_s(t) = \emptyset$ does not imply that the school s considers student t worse than any or all of their ranked students $t' \in \mathcal{L}_s$ that showed up their application list. Therefore, WTT does not allow us to infer anything about the upper and lower valuation bounds for those students that did not apply at school s. Schools are required to prioritize all students that apply to them, but if school s deems student $t \in \mathcal{L}_s$ unacceptable, it assigns $pr_s(t) = \infty$ to that student, which implies that this student t is less preferred than any other ranked student $t' \in \mathcal{L}_t : pr_s(t') < \infty$. This, however, does not allow us to infer anything about how school s prioritizes student t relative to other students that are equally unacceptable. Hence, the upper bounds for school s's valuation of student t, $V_s(t)$, conditional on $\mathbf{V}_{-st}$ and observed priorities $\mathbf{pr}$ are given by

$$\overline{V}_s(t) = \begin{cases} +\infty & pr_s(t) \in \{1, \emptyset\} \\ \min_{t' \in \mathcal{L}_s}\{V_s(t') : pr_s(t') < pr_s(t)\} & 1 < pr_s(t) \leq \overline{pr}_s \end{cases}$$

and the lower bounds by

$$\underline{V}_s(t) = \begin{cases} -\infty & pr_t(s) \in \{\overline{pr}_s, \emptyset\} \\ \max_{t' \in \mathcal{L}_s}\{V_s(t') : pr_s(t') > pr_s(t)\} & 1 \leq pr_s(t) < \overline{pr}_s \end{cases}$$

Undominated Strategies (UNDOM)

Under undominated strategies (UNDOM), unranked alternatives are not assumed to be worse thank ranked alternatives, from the students' perspective. Therefore, UNDOM imposes fewer restrictions than WTT. Given latent valuations $\mathbf{U}_{-it}$, and observed ranks $\mathbf{rk}$, the upper and lower bounds for utility $U_t(s)$ can be expressed as follows:

$$\overline{U}_t(s) = \begin{cases} +\infty & rk_t(s) \in \{1, \emptyset\} \\ \min_{s' \in \mathcal{L}_t}\{U_t(s') : rk_t(s') < rk_t(s)\} & rk_i(s) > 1 \end{cases}$$
$$\underline{U}_t(s) = \begin{cases} -\infty & rk_t(s) \in \{L_t, \emptyset\} \\ \max_{s' \in \mathcal{L}_t}\{U_t(s') : rk_t(s') > rk_t(s)\} & rk_t(s) < L_t \end{cases}$$

Because schools cannot choose to intentionally not rank a student who applies there, the upper and lower bounds under UNDOM are exactly the same that were derived under WTT.

Stability

Finally, consider an observed matching μ where $\mu(s)$ denotes the set of all students that are assigned to school s, and $\mu(t)$ denotes the assigned school of student t (a student can only be assigned to one school at once). If student t is unassigned, $\mu(t) = t$. Every school can accommodate at most q_s students, so we define the convenience function

$$\chi(s) = \mathbf{1}\left(|\mu(s)| = q_s\right)$$

that indicates whether a school is at full capacity or not. Further, define the *feasible set* of student t as the set of schools that do not classify student t as unacceptable or have not ranked student t, and that either have some vacant seats, or would favour student t over one of their currently admitted students:

$$\mathcal{F}_t = \left\{ s \in S : (pr_s(t) < \infty \vee pr_s(t) = \emptyset) \wedge \left(\neg\chi(s) \vee V_s(t) > \min_{t' \in \mu(s)} V_s(t') \right) \right\}.$$

This feasible set of student t is unobserved (it is a latent set) because it depends on the latent valuations $\mathbf{V}$.

We now outline conditions on the valuations and utilities that, if satisfied, guarantee that the observed matching μ is stable. Logan et al. (2008) have used similar conditions to estimate the parameters of a one-to-one marriage market model, and we adapt their setting to a many-to-one matching market. Before we proceed, we introduce the following assumption:

Assumption 1 (Non-wastefulness)**.** *The matching μ is non-wasteful: all schools operate at full capacity ($|\mu(s)| = q_s$) or no student is unmatched ($\mu(t) \neq t$).*

This assumption is convenient in order to ensure that one can always find utilities and valuations that are consistent with a stable matching and it is also the approach that was taken by Sørensen (2007, p.2732). Without this assumption, it would be necessary to specify outside options for agents, which would complicate the analysis, but pose no substantial challenges to it. Conditional on the latent set $\mathcal{F}_t$, stability requires that student t's utility for any school in this latent set be less than that of her currently assigned school. Therefore, the upper bound for a student t's valuation of school s is given by

$$\overline{U}_t(s) = \begin{cases} U_t(\mu(t)) & \mu(t) \notin \{s, t\} \wedge s \in \mathcal{F}_t \\ +\infty & \text{else} \end{cases}$$

Similarly, the lower bounds are given by

$$\underline{U}_t(s) = \begin{cases} \max_{s' \in \mathcal{F}_t \setminus \{s\}} \{U_t(s')\} & \mu(t) = s \\ -\infty & \text{else} \end{cases}$$

Note that we assume that all schools are acceptable to the student. This implies that if student t is unmatched ($\mu(t) = t$), then we cannot bound her utility for any school, be it in her feasible set or not. Instead, stability requires that her feasible set be empty. This places bounds on the schools' valuations for student t which will be described shortly.

We define school s's feasible set as the set of students who are acceptable to school s, and who would prefer going to school s rather than to their current school, or are unassigned under the matching μ. We chose to include only students that are acceptable to school s in this set because it simplifies the notation below. Thus, the feasible set is given by

$$\mathcal{F}_s = \{t \in T : pr_s(t) < \infty \wedge (U_t(s) > U_t(\mu(s)) \vee \mu(t) = t)\}.$$

Again, this is a latent set that depends on the latent student utilities $\mathbf{U}$. Then, upper and lower bounds of school s's valuation of student t can be constructed if school s is at full capacity, i.e. if $\chi(s)$ is true:

$$\overline{V}_s(t) = \begin{cases} \min_{t' \in \mu(s)} \{V_s(t')\} & \chi(s) \wedge t \notin \mu(s) \wedge t \in \mathcal{F}_s \\ +\infty & \text{else.} \end{cases}$$

Similarly, the lower bounds are given by

$$\underline{V}_s(t) = \begin{cases} \max_{t' \in \mathcal{F}_s \setminus \mu(s)} \{V_s(t')\} & \chi(s) \wedge t \in \mu(s) \\ -\infty & \text{else.} \end{cases}$$

In general, the upper and lower bounds on utilities and valuations are interdependent, and are not unique.

Combination of UNDOM and Stability

The combination of the two assumptions that students and schools play undominated strategies, and that the assignment is stable, allows us to tighten the bounds. For instance, let $[\underline{U}_t^{rk}(s), \overline{U}_t^{rk}(s)]$ be the bound that is imposed by the assumption of undominated strategies on the valuation $U_t(s)$, and let $[\underline{U}_t^{\mu}(s), \overline{U}_t^{\mu}(s)]$ be the bounds that follow from the requirement that the observed matching μ be stable. An obvious way to combine these two bounds is to simply set

$$\underline{U}_t(s) = \max\left\{\underline{U}_t^{rk}(s), \underline{U}_t^{\mu}(s)\right\}$$
$$\overline{U}_t(s) = \min\left\{\overline{U}_t^{rk}(s), \overline{U}_t^{\mu}(s)\right\}$$

and for $V_t(s)$ in an analogous manner. Now, the question is whether so truncation intervals that are constructed in this way are non-empty, *i.e.* whether $\underline{U}_t(s) \leq \overline{U}_t(s)$. We will show that, for any given stable matching μ, observed priorites $\mathbf{pr}$ and preference ranks $\mathbf{rk}$, there is at least one set of preferences $\mathbf{U}$ and valuations $\mathbf{V}$ such that the assumptions UNDOM and stability are satisfied:

Lemma 3. *Consider any given non-wasteful stable matching μ that is derived from the observed partial rankings $\mathbf{rk}$ and priority structures $\mathbf{pr}$. Then, there exists a complete preference structure $\succ$ and priority ordering $\rhd$ such that*

1. *$\succ$ and $\rhd$ are consistent with $\mathbf{rk}$ and $\mathbf{pr}$, respectively and*
2. *μ is stable under $\succ$ and $\rhd$.*

Thus, the set of utilities $\mathbf{U}$ and valuations $\mathbf{V}$ that satisfies the bounds imposed by UNDOM *and* stability *is non-empty for any observed matching μ.*

Proof. The first point is obvious: fix an arbitrary set of utility numbers $\{U_t(s) : s \in \mathcal{L}_t\}$ and valuation numbers $\{V_s(t) : t \in \mathcal{L}_s\}$ that respect the ordering implied by the observed ranks $\mathbf{rk}$ and priorities $\mathbf{pr}$; there will always be such numbers. For the second point, note that we can equivalently express students' preferences and schools' priorities in terms of their partial rank and priority order lists, or in terms of their utilities and valuations. Since the observed matching μ is stable under the former, it must also be stable under the latter representation and so, any set of utility and priority numbers that respects the bounds imposed by *UNDOM* also satisfies the bounds that are imposed by *stability*. Next, we need to show that there are always utility and valuation numbers for the remaining non-ranked pairs such that there are no *blocking pairs*. Consider any such pair t, s such that $s \notin \mathcal{L}_t$. Under the student-proposing deferred acceptance mechanism, no student can be assigned to a school that she did not include in her stated rank order list $\mathbf{rk}_t$, and hence $s \neq \mu(t)$. Then there are four remaining cases to consider:

Case 1 Student t is not unmatched, and school s is at full capacity, i.e. $\mu(t) \neq t$ and $|\mu(s)| = q_s$. Stability is satisfied if $U_t(s) < U_t(\mu(t))$ or $V_s(t) < \min_{t' \in \mu(s)} V_s(t')$, or both.

Case 2 Student t is not unmatched, and school s has spare capacity. Stability is satisfied for all $U_t(s) < U_t(\mu(t))$ and $V_s(t) \in \mathbb{R}$.

Case 3 Student t is unmatched, and school s is at full capacity. Stability is satisfied for all $V_s(t) < \min_{t' \in \mu(s)} V_s(t')$ and $U_t(s) \in \mathbb{R}$.

Case 4 Student t is unmatched, and school s has spare capacity. This case is ruled out under the assumption that μ is non-wasteful.

Hence, if the matching μ is non-wasteful, it will always be possible to find utilities and valuations that respect both the partially observed rank and priority structures, and stability properties. □

However, we observe in our dataset that roughly ten percent of all students are not assigned to a school in the first matching round (*c.f.* table 4.1) so that the allocation is not non-wasteful in the sense outlined above, and the last case of the proof does not go through.[2] This could appear to be a problem for our estimation approach, because the existence of an unmatched student t and a school that has spare capacity s necessarily leads to instability in our estimation approach. The solution would be to endogenously determine "latent" unacceptable students, to exclude such students from the sample, or to artificially label them as being "unacceptable", neither approach of which is very attractive. Instead, we note that if there exists a student t who is unmatched, and a school s with spare capacity, it must either be that t did not apply to s, in which case the bounds on the latent utility and on the latent valuation are $\pm\infty$, or that student t did rank school s, but school s ranked student t as unacceptable, in which case the valuation and utility bounds are well defined. Only the former case represents a case of true instability, whereas the latter case is well covered by our estimation approach. Most importantly, if such a case of true instability should occur, it will not affect the parameter estimates in either direction, because the utilities and valuations are not restricted and simply add some white noise to the parameter updates.

[2]In the Hungarian school choice system, the main matching round is followed by a subsequent round in which any unmatched students are assigned to the closest feasible school.

B.2.2. Posterior distributions

The Bayesian estimator uses the data augmentation approach (proposed by Albert and Chib, 1993) that treats the latent valuation variables as nuisance parameters. This section describes the components conditionals of the Gibbs sampler that is used to sample from the posterior distribution of the parameters of interest β and γ, $p(\beta, \gamma|data)$ where the $data$ are observed co-variates, and possibly rank and priority structures or matching information.

Conditional distribution of utilities and valuations

Recall that it is assumed that $\epsilon_{ts}, \eta_s t \sim N(0, 1)$, as is customary and necessary in the discrete choice literature. Then, the component conditionals for the unobserved latent utilities and valuations are given by

$$p(U_t(s)|\beta, \gamma, \mathbf{U}_{-ts}, \mathbf{V}, data) \propto \exp\left\{\frac{-(U_t(s) - \mathbf{X}_{ts}\beta)^2}{2}\right\} \mathbf{1}(U_t(s) \in [\underline{U}_t(s)\overline{U}_t(s)])$$

$$p(V_s(t)|\beta, \gamma, \mathbf{V}_{-st}, \mathbf{U}, data) \propto \exp\left\{\frac{-(V_s(t) - \mathbf{W}_{st}\gamma)^2}{2}\right\} \mathbf{1}(V_s(t) \in [\underline{V}_s(t), \overline{V}_s(t)])$$

Note that, although the error terms are uncorrelated and independent across alternatives, the utilities are not because their truncation intervals are endogenously determined. For example, if we observe a student's ranking across three different schools A, B, and C such that $rk_t(A) < rk_t(B) < rk_t(C)$, this implies that $U_t(A) > U_t(B) > U_t(C)$. Therefore, the distribution of utilities across schools is not iid normal, but rather a multivariate normal distribution subject to a system of linear inequality constraints (Train, 2009, p.181). Commonly known techniques for sampling from these distributions with linear constraints are rather slow when the number of alternatives is very large, as is the case in our setting with thousands of students, and hundreds of schools.[3] Instead, we embed the sampling from this intractable distribution into our Gibbs sampler. However, we found that this procedure is rather slow to converge, and also exhibits very strong serial correlation so that a sufficiently large number of Gibbs samples must be drawn.

Conditional distribution of utility and valuation parameters

We assume a vague prior for the structural parameters β and γ which, together with the assumption that the error terms have unit variance, implies that the posteriors of β and γ follow a normal distribution (Lancaster, 2004, p.120). Also, we note that the scale and the location of the utilities and valuations are not identified, as in any discrete choice model. Our assumption that the idiosyncratic errors have unit variance pins down the scale of utility, and the assumption that these errors are zero in expectation pins

[3]The function `rtmvnorm2` in the R package tmvtnorm (`https://cran.r-project.org/package=tmvtnorm`, version 1.4-10) does provide such a a method

down the location of utilities. Hence the component conditional distribution of the utility parameter is given by

$$p(\beta|\gamma, \mathbf{U}, \mathbf{V}, data) = p(\beta|\mathbf{U}, data) = N\left(b, (\mathbf{X}'\mathbf{X})^{-1}\right)$$

for $b = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{U}$, and similarly, the conditional component for the priority parameter γ reads

$$p(\gamma|\beta, \mathbf{U}, \mathbf{V}, data) = p(\gamma|\mathbf{V}, data) = N\left(g, (\mathbf{W}'\mathbf{W})^{-1}\right)$$

for $g = (\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'\mathbf{V}$.

B.2.3. More Monte Carlo results

100 students

In a smaller market with only one hundred students, the stability-based estimator performs very poorly compared to any other estimation strategy:

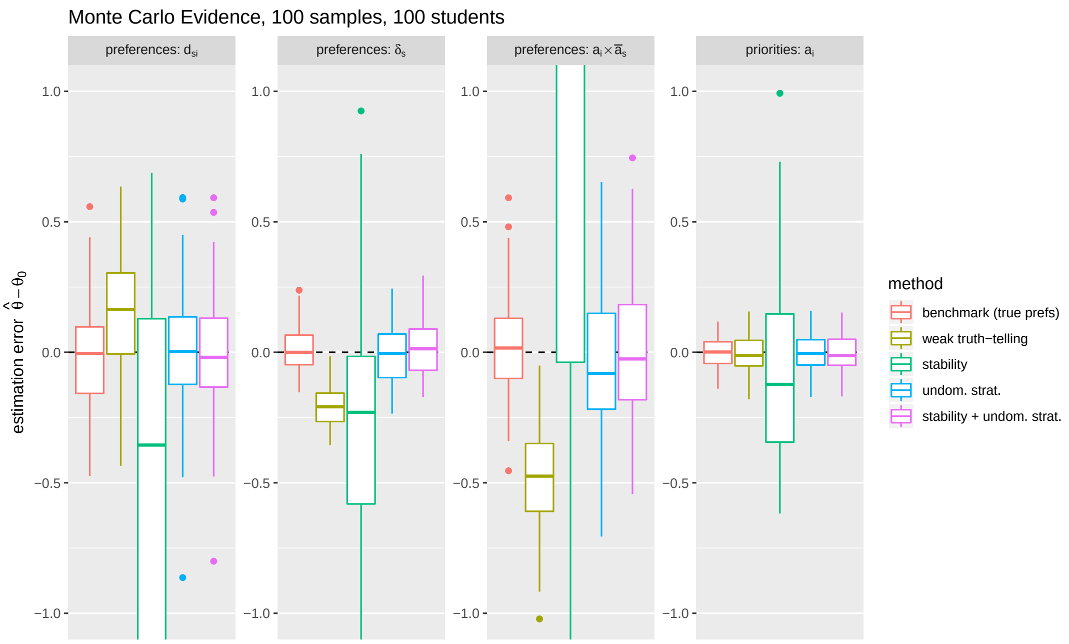


Figure B.1.: Distributions of estimation errors across one hundred simulated markets (six schools with 95 seats and 100 students).

(a) Mean squared error (MSE)

method	preferences d_{is}	δ_s	$a_i \cdot \bar{a}_s$	priorities a_i
benchmark (true prefs.)	0.0393	0.0065	0.0391	0.0028
weak truth−telling	0.0643	0.0496	0.2832	0.0048
stability	4.7436	0.3016	28.3876	0.1974
undominated strategies	0.0547	0.0114	0.0801	0.0049
stability + undom. strat.	0.0517	0.0107	0.0798	0.0047

(b) Bias

method	preferences d_{is}	δ_s	$a_i \cdot \bar{a}_s$	priorities a_i
benchmark (true prefs.)	-0.0144	0.0091	0.0223	0.0001
weak truth−telling	0.1498	-0.2091	-0.4897	-0.0050
stability	-1.0446	-0.2945	3.0821	-0.0283
undominated strategies	-0.0016	-0.0073	-0.0323	0.0003
stability + undom. strat.	-0.0213	0.0146	0.0093	0.0011

Table B.1.: MSE and bias statistics for one hundred simulated markets (six schools with 95 seats and 100 students).

500 students

The variance of the estimates improves considerably in larger markets, as figure B.2 below shows. However, the stability based estimator still produces estimates that are rather imprecise, and also biased.

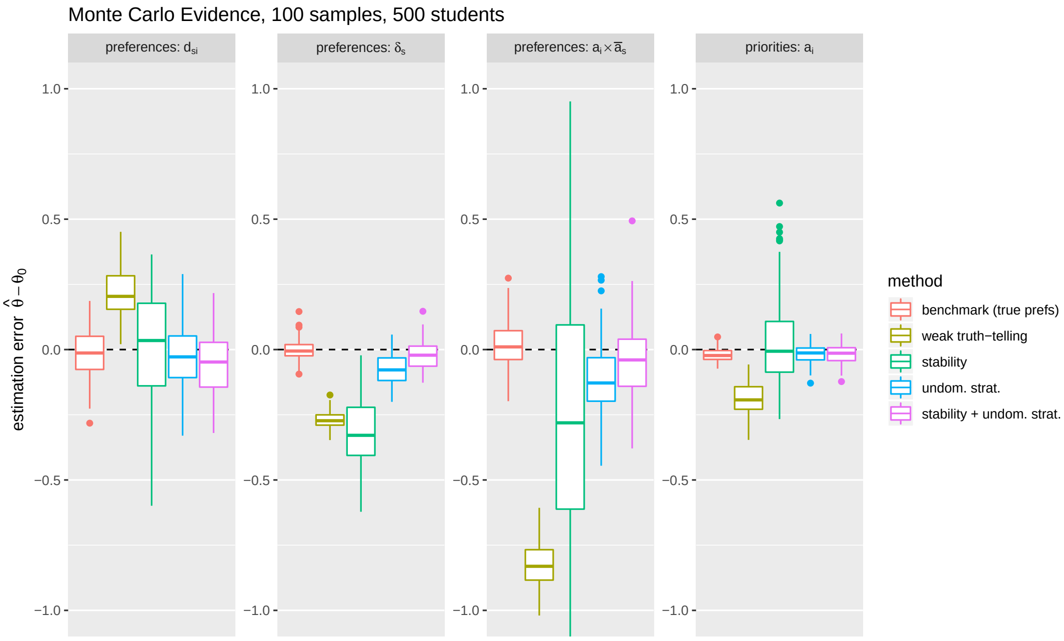


Figure B.2.: Distributions of estimation errors across one hundred simulated markets (six schools with 475 seats and 500 students).

(a) Mean squared error (MSE)

method	preferences			priorities
	d_{is}	δ_s	$a_i \cdot \bar{a}_s$	a_i
benchmark (true prefs.)	0.0083	0.0016	0.0093	0.0010
weak truth–telling	0.0549	0.0743	0.6863	0.0398
stability	0.0446	0.1207	0.3576	0.0328
undominated strategies	0.0158	0.0086	0.0343	0.0015
stability + undom. strat.	0.0163	0.0035	0.0217	0.0016

(b) Bias

method	preferences			priorities
	d_{is}	δ_s	$a_i \cdot \bar{a}_s$	a_i
benchmark (true prefs.)	-0.0129	-0.0019	0.0152	-0.0208
weak truth–telling	0.2179	-0.2704	-0.8242	-0.1892
stability	0.0141	-0.3243	-0.2801	0.0242
undominated strategies	-0.0207	-0.0726	-0.1198	-0.0181
stability + undom. strat.	-0.0515	-0.0238	-0.0404	-0.0172

Table B.2.: MSE and bias statistics for one hundred simulated markets (six schools with 475 seats and 500 students).

B.2.4. Data sources and construction

Table B.3 shows summary statistics of the student-level NABC data. Most students are fifteen years old at the time of the NABC test (in 2015). The NABC scores in Hungarian and mathematics are the results of a standardized test procedure. The socio-economic status (SES) is a composite measure that is based on responses given by students in an accompanying survey, so that this variable has more missing data. Also, the grade average is based on student's own responses and may thus be biased. Therefore, we use the NABC scores as a proxy for student's academic ability. The table also shows that, unsurprisingly, students from Budapest perform better in the NABC test, and have a higher socio-economic status.

statistic	mean	SD	min	max	N
Entire country					88967
Birth year	2000.1	0.58216	1996	2002	88959
Female	0.494	0.5	0	1	88967
Last grade average	3.9837	0.7668	1	5	60843
NABC score Hungarian	1559.9	202.36	820.97	2199.2	82237
NABC score math	1612.1	196.5	907.81	2307.3	82176
Socioeconomic status (csh)	-0.0226	1.01	-3.15	1.88	64971
Budapest					13611
Birth year	2000.1	0.55451	1996	2002	13609
Female	0.497	0.5	0	1	13611
Last grade average	4.1929	0.65506	1	5	8392
NABC score Hungarian	1634.9	190.44	820.97	2199.2	12480
NABC score math	1685.7	190.24	947.4	2307.3	12467
Socioeconomic status (csh)	0.616	0.88	-3.15	1.88	9029

Table B.3.: Summary statistics of the original NABC (2015) data

Table B.4 shows that there are significant differences in test outcomes and between male and female students. Female students perform much better in Hungarian on average (almost one third of a standard deviation), whereas male students perform better in math on average (one tenth of a standard deviation). Also, female students obtain a slightly better SES index (five percent of the standard deviation) but notice that the SES index is based on self reporting, so it could be due to different reporting behaviour. In all cases, the differences in means are significant at the one percent level.

		mean			t-test		
statistic	N	all	male	female	diff.	t-stat.	p-value
NABC score Hungarian	12,480	1,635	1,583	1,636	-52.94	-14.379	$<$0.001
NABC score math	12,467	1,686	1,670	1,651	19.26	5.283	$<$0.001
Socioeconomic status (csh)	9,029	0.616	0.426	0.473	-0.047	-2.697	0.007

Table B.4.: Gender differences in test outcomes. Raw NABC (2015) data; all students from Budapest. Two-sample t-test with equal variance.

Table B.5 shows key statistics of the nation-wide matching scheme. The data comprises almost four hundred thousand applications from almost ninety thousand students to over six thousand school programs. Each record corresponds to the application of a student to a school and contains the ranks $rk_t(s)$ and $rk_s(t)$, an indicator whether the school finds the student acceptable, and a match indicator. On average, each student applies to 4.5 school programs, or to 2.8 different schools. Almost 95% of all students are assigned to a school, of which three quarters are eventually assigned to their top choice program.[4] We link this data to a school survey in order to obtain the precise location of each school, and the school's district.

# students	88, 401
# school programs	6, 181
# schools (OMid-telephely-tipus)	1, 793
# student-school applications	395, 222
length of submitted ROL (school programs)	4.471
length of submitted ROL (schools)	3.002
# assigned	83, 482
.. share top choice	0.759
.. average match rank	1.486

Table B.5.: Summary statistics of the original application data (KIFIR)

The analysis was conducted for us by the Hungarian ministry of education who used a confidential concordance table to link records from the KIFIR and NABC datasets. The following table B.6 shows that a large share of the students from Budapest can be linked, which leaves us with a sizeable sample of students. We further restricted the sample to include only those students who filed an application for at least one school from Budapest (last row).

# students in KIFIR	88, 401
# students in NABC	88, 967
.. of which Budapest	13, 611
# students in KIFIR and linked to NABC	80, 385
.. of which Budapest	10, 962
.. of which in sample	10, 880

Table B.6.: Merging the KIFIR and the NABC datasets.

As table B.3 shows, the NABC scores and, in particular, the SES are missing for a quite substantial share of our sample. Because the computation of the student-optimal stable matching depends on the composition of the student sample, we were reluctant to drop records with missing data, as this would have left us with rather few complete records. Instead, we opted for a data imputation approach and used the R package `mice` to construct a complete dataset. Missing variables were imputed using predictive mean matching, were missing values are replaced by actual values from other records that resemble the incomplete record, conditional on other observed characteristics. As predictors, we used an extended set of variables that included also some results from the 2017 NABC round (where available), and further student level variables that are not shown here. This procedure is repeated a few times, until the

[4]The admission system ensures that any students who are unmatched at the end of the main matching round are assigned to the nearest school which still has free capacity.

imputed values converge in expectation. It is recommended that researchers construct multiple imputed datasets to assess the robustness of their analysis with respect to these imputations, but due to the substantial computational burden of our estimation procedures, this was infeasible in our context. The following table B.7 shows details of the imputation procedure. It can be seen that the imputed mean of the variables referring to academic ability is lower than in the original data. Our imputation procedure naturally introduces measurement error into the data, which, in a classical regression framework, should lead to estimated coefficients that biased towards zero. We expect that this is also true for our estimation procedure which essentially is a data augmentation approach with a linear regression. Nevertheless, it is our opinion that the drawbacks of using an imputed data set are greatly outweighed by the benefit of having a comprehensive set of students for the estimation procedure (which relies on stability considerations, and thus, on the entirety of the student population) and for the counter-factual matches (which are more directly dependent on the entire student population).

statistic	N	mean	sd	N.imp	mean.imp	sd.imp	pval
Birth year	10,879	2,000.06	0.55	1	2,001		0.09
Sex (1=female,2=male)	10,880	1.50	0.50	0			
Last grade average	6,598	4.12	0.68	4282	3.97	0.70	0
NABC score Hungarian	9,934	1,659.63	183.87	946	1,612.10	192.57	0
NABC score math	9,948	1,607.88	186.60	932	1,569.42	189.95	0
Socioeconomic status (SES)	7,097	0.45	0.87	3783	0.41	0.88	0.02

Table B.7.: Results of the imputation procedure, using predictive mean matching and ten iterations. The p-value is computed for a two-sided t-test with unequal variances.

In order to ease the interpretation of estimated preference parameters, we decided to standardize the NABC scores and the SES index to having a mean of zero, and unit standard deviation. This is shown in table 4.1 in the main text.

B.2.5. Decomposing the gains from consolidation: details

This section presents in detail how we construct the decomposition of the students' consolidation gains into a choice effect and a competition effect. In doing so, we make use of the large market approximation to matching markets (Azevedo and Leshno, 2016) by which school-specific cutoff scores play the role of prices that balance the supply of, and the demand for school seats. The cutoff score at school s under the matching μ is the lowest valuation among all students who where admitted to that school under μ, or

$$c_s(\mu) = \min_{t \in \mu(s)} V_s(t)$$

We assume that the school market consists of relatively few schools and a large number of students so that the addition (or deletion) of a single student has practically no effect on a schools' cutoff score, in line with the framework Azevedo and Leshno (2016). In order to decompose the total consolidation gains, we compute the school-level cutoff scores under the district wise matching μ_d and under the integrated matching μ_{BP}. The effect of increased choice, keeping everything else constant, can then be computed as the difference between student t being matched to her most preferred feasible school in her own district, and globally, using either the district-level or the city-wide cutoffs. Let the feasible set of student t under the cutoffs $\{c_s(\mu)\}_{s \in S}$ be

$$\mathcal{F}_t^{\mu} = \{s \in S : V_s(t) \geq c_s(\mu)\}$$

and denote the set of schools in district d as S^d. Then, the choice gain of student t can either be expressed as

$$\Delta^{ch-I} U_t = \max_{s \in \mathcal{F}_t^{\mu_{BP}}} U_t(s) - \max_{s \in \mathcal{F}_t^{\mu_{BP}} \cap S^d} U_t(s)$$

or

$$\Delta^{ch-II} U_t = \max_{s \in \mathcal{F}_t^{\mu_d}} U_t(s) - \max_{s \in \mathcal{F}_t^{\mu_d} \cap S^d} U_t(s)$$

as is illustrated in figure 4.10. The only difference between $\Delta^{ch-I} U_t$ and $\Delta^{ch-II} U_t$ is the usage of a different baseline scenario to compute the cutoffs – the global cutoffs $\{c_s(\mu_{BP})\}$ for $\Delta^{ch-I} U_t$ and the local cutoffs for $\Delta^{ch-II} U_t$. It is easy to see that the choice gains will always be weakly positive by construction. It can also happen that a student is not assigned in one of the counter-factual scenarios. In our empirical application, the choice gains $\Delta^{ch-I} U_t$ are missing for about one quarter of all students, because their set of feasible schools within their home district is empty under the global cutoff scores. In a similar manner, one can compute the change in student t's welfare as the market is opened up to external competition. We call this change a competition gain, but it is not a priori clear whether students actually gain or lose from competition. The competition gain can be computed either as

$$\Delta^{co-I} U_t = \max_{s \in \mathcal{F}_t^{\mu_{BP}} \cap S^d} U_t(s) - \max_{s \in \mathcal{F}_t^{\mu_d} \cap S^d} U_t(s)$$

or as

$$\Delta^{co-II}U_t = \max_{s \in \mathcal{F}_t^{\mu_{BP}}} U_t(s) - \max_{s \in \mathcal{F}_t^{\mu_d}} U_t(s)$$

Now, Δ_t^{co-I} differs from $\Delta^{co-II}U_t$ in that student t's choice set is restricted to feasible schools within her home district d in the former, but not in the latter. It is easy to see that the sum of $\Delta^{ch-I}U_t$ and $\Delta^{co-I}U_t$ is identical to the sum of $\Delta^{ch-II}U_t$ and $\Delta^{co-II}U_t$ unless some type-I choice gains are missing. Also, the sum of the choice and competition gains are equal to the total welfare gains.

B.2.6. Balanced markets

The Hungarian school market is characterized by a great amount of nominal excess capacity. To see whether, and if so, how, this affects the conclusions drawn in the main text, we repeated the analysis in section 4.6.3 with an artificially balanced market. This was achieved by scaling the schools' capacities proportionally (up to the integer constraint) within each district until the total number of seats equals the total number of students. In doing so, we guarantee that every student is matched to *some* school. Of course, this is a highly artificial setting, but it serves as a useful comparison benchmark against which the results from the main text can be viewed.

Table B.8a shows the match statistics for the balanced markets. The consolidation gains were computed analogue to the main text. The first row of table B.8b shows that the consolidation gains are now very small compared to the large gains achieved in the unbalanced markets, and the median student neither gains nor looses due to district consolidation. The de-composition into choice and competition effects, also shown in table B.8b, shows why this is the case: The choice, and the competition effects now have about equal magnitudes and opposite signs, and so they cancel each other.[5] Interestingly, the competition effects are now strictly negative.[6]

Figure B.3a shows that there is a weakly negative relationship between ex post majority support for market consolidation and district size. A linear regression analysis (not shown here) confirms this, but does not find a significant effect ($p = 0.139$). The important difference to results from the main text, which were derived with the original amount of excess capacity, is that not all districts have a majority of consolidation winners. Figure B.3b shows that there is a weakly negative correlation between average latent utility gains and district size, similar to figure 4.9a. But again, this negative effect is insignificant in a linear model ($p = 0.448$). Thus, we cannot confirm the prediction of Corollary 1 in this case. Because all district-level school markets were exactly balanced in this exercise, it is not possible to determine how the excess capacity affects the gains from consolidation.

We also estimated a linear regression of the students' total gain and their choice and competition gains on student and district observables, the results of which are shown in table B.9. Contrary to table 4.11, students with a higher SES gain less than the average student, but the coefficient is equally insignificant. Furthermore, high-ability students have significantly larger consolidation gains. Interestingly, the estimated effect of a student's home district size is now significantly positive, contrary to the correlation

[5] The choice and competition effects of type-I could not be computed for one quarter of the students because the school market is now balanced, and thus very tight. This leads to the situation where many students have no feasible school in their home district, given the consolidated school-level cutoff. This problem does not arise with the choice- and competition effects of type-II.

[6] This is a rather peculiar results, and it is worth some discussion. Recall that the competition gains are computed by comparing the students' feasible choice sets under different scenarios, and that those are in turn based on the schools' admission cutoffs (see section B.2.5). With balanced markets, it turns out that the school level cutoffs are empirically larger than the district-level cutoffs. This holds true for all but one school. The fact that there is one exception leads to the conclusion that this is an empirical phenomenon that arises in a large market, but that it is not a strict implication of the way we constructed the feasible choice sets per se. The larger cutoffs in the integrated market results in smaller feasible choice sets, and so the competition effects are negative in our sample.

district markets	
# matched	10,880
share top choice match	0.57
ave. match distance [km]	3.53
consolidated market	
# matched	10,880
share top choice match	0.54
share matched in home district	0.18
ave. match distance [km]	8.99

(a) Assignment statistics of the district-wise and integrated student-school matching.

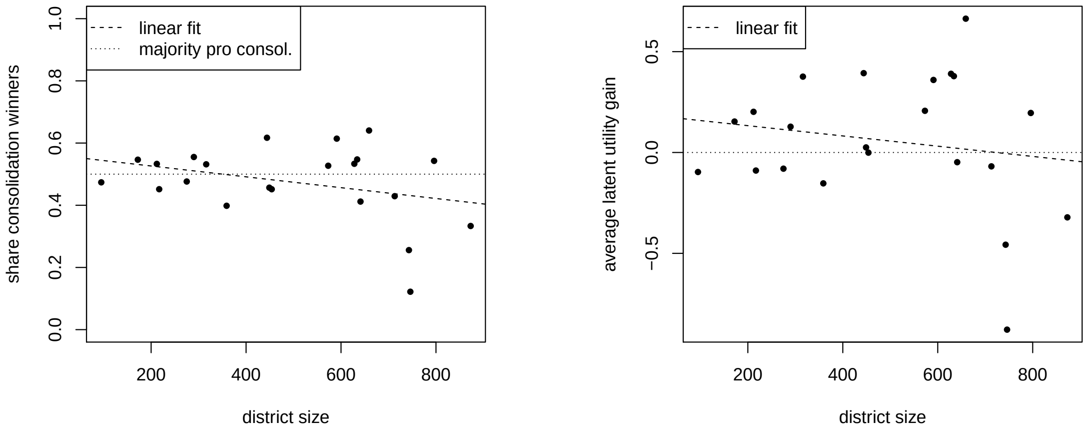
	Mean	SD	Min	Median	Max	N
total gains						
in latent utility units	0.033	1.131	-7.000	0.000	5.000	10,880
in equivalent kilometres	0.223	7.639	-47.297	0.000	33.784	10,880
decomposition						
choice effect I	1.190	1.203	0.000	0.929	7.445	7,536
competition effect I	-1.055	1.289	-8.060	-0.566	0.000	7,536
choice effect II	1.050	0.981	0.000	0.888	6.725	10,880
competition effect II	-1.017	0.818	-6.791	-0.924	0.000	10,880

(b) Various measures of consolidation gains, expressed in latent utility changes.

Table B.8.: Gains from market consolidation using inferred complete preferences lists and artificially balanced markets: summary statistics

in figure B.3b. However, when the district FEs are not included (results not shown here), the effect is significantly negative.

In conclusion, this appendix shows that the results from the main text do not necessarily carry over to situations where the aggregate school market has less excess capacity or is even balanced. With artificially balanced markets, the median student neither gains nor losses due to market consolidation, and the share of students who gain is below fifty percent in many districts. This is at odds with the theoretical predictions in chapter 4.3 where we showed that a smaller aggregate market imbalance should lead to *larger* expected welfare gains (Corollary 1), and it could be due to the fact that those theoretical results were derived under the stark assumption of uniform and random preferences. Therefore, it seems imperative for theoretical researchers to extend the set of possible preference structures that are accommodated by their models.



(a) Share of winners and district sizes

(b) Average latent utility gains and district sizes

Figure B.3.: Majority support for, and average latent utility gains of market consolidation, using inferred complete preference lists and balanced markets.

	Dependent variable: latent utility gains				
		type-I decomposition		type-II decomposition	
	total (1)	choice (2)	competition (3)	choice (4)	competition (5)
socio-economic status SES	-0.0099 -0.0121	0.0585*** -0.0164	-0.0716*** -0.017	0.0494*** -0.0106	-0.0594*** -0.0091
ability	0.0839*** -0.0088	0.0404*** -0.0123	0.0235* -0.0127	0.0066 -0.0077	0.0772*** -0.0066
district size (100 students)	0.2044*** -0.0559	-0.1431 -0.1533	0.2173 -0.1587	0.0203 -0.0492	0.1841*** -0.042
school type: gymnazium	0.0993*** -0.0295	0.0854** -0.0424	-0.0089 -0.0438	-0.0193 -0.0259	0.1186*** -0.0222
school type: secondary	-0.5061*** -0.1579	1.4227*** -0.4762	-1.4857*** -0.4928	1.3117*** -0.1388	-1.8178*** -0.1186
Constant	-0.5061*** -0.1579	1.4227*** -0.4762	-1.4857*** -0.4928	1.3117*** -0.1388	-1.8178*** -0.1186
district FE	Yes	Yes	Yes	Yes	Yes
Observations	10,880	7,536	7,536	10,880	10,880
Note:			*p<0.1; **p<0.05; ***p<0.01 Standard errors in parentheses.		

Table B.9.: Explaining gains from consolidation with students observables (balanced markets). The table shows regression coefficients of students' gains on student observables. The school type refers to the school type of the assigned school in the integrated market.

Bibliography

ABDA (2016). Die Apotheke – Zahlen, Daten, Fakten 2016. Technical report, ABDA - Bundesvereinigung Deutscher Apothekerverbände e. V. Retrieved from abda.de on 2 June 2020.

Abdulkadiroğlu, A., Agarwal, N., and Pathak, P. A. (2017). The welfare effects of coordinated assignment: Evidence from the New York City high school match. *American Economic Review*, 107(12):3635–89.

Abdulkadiroğlu, A., Angrist, J., and Pathak, P. (2014). The elite illusion: Achievement effects at Boston and New York exam schools. *Econometrica*, 82(1):137–196.

Agarwal, N. (2015). An empirical model of the medical match. *American Economic Review*, 105(7):1939–78.

Agarwal, N. and Diamond, W. (2014). Identification and estimation in two-sided matching markets. Cowles Foundation Discussion Papers 1905, Cowles Foundation for Research in Economics, Yale University.

Aguirregabiria, V. and Magesan, A. (2019). Identification and Estimation of Dynamic Games When Players' Beliefs Are Not in Equilibrium. *The Review of Economic Studies*, 87(2):582–625.

Aguirregabiria, V. and Mira, P. (2002). Swapping the nested fixed point algorithm: A class of estimators for discrete markov decision models. *Econometrica*, 70(4):1519–1543.

Aguirregabiria, V. and Mira, P. (2007). Sequential estimation of dynamic discrete games. *Econometrica*, 75(1):1–53.

Aguirregabiria, V. and Mira, P. (2010). Dynamic discrete choice structural models: A survey. *Journal of Econometrics*, 156(1):38–67.

Aguirregabiria, V. and Vicentini, G. (2016). Dynamic spatial competition between multi-store retailers. *The Journal of Industrial Economics*, 64(4):710–754.

Akyol, P. and Krishna, K. (2017). Preferences, selection, and value added: A structural approach. *European Economic Review*, 91:89–117.

Albert, J. H. and Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, 88(422):669–679.

Allen, T. and Arkolakis, C. (2014). Trade and the topography of the spatial economy. *The Quarterly Journal of Economics*, 129(3):1085.

Alsbury, T. and Shaw, N. (2005). Policy implications for social justice in school district consolidation. *Leadership and Policy in Schools*, 4(2):105–126.

Anderson, S. P., de Palma, A., and Thisse, J.-F. (1992). *Discrete choice theory of product differentiation*. MIT Press.

Anderson, S. P., Goeree, J. K., and Ramer, R. (1997). Location, location, location. *Journal of Economic Theory*, 77(1):102–127.

Ansari, A., Economides, N., and Steckel, J. (1998). The max-min-min principle of product differentiation. *Journal of Regional Science*, 38(2):207–230.

Artemov, G., Che, Y.-K., and He, Y. (2017). Strategic 'mistakes': Implications for market design research. Working paper. Retrieved from sites.google.com/site/yinghuahe on 3 June 2020.

Ashlagi, I., Kanoria, Y., and Leshno, J. (2017). Unbalanced random matching markets: The stark effect of competition. *Journal of Political Economy*, 125(1):69–98.

Azevedo, E. M. and Leshno, J. D. (2016). A supply and demand framework for two-sided matching markets. *Journal of Political Economy*, 124(5):1235–1268.

Bajari, P., Benkard, C. L., and Levin, J. (2007). Estimating dynamic models of imperfect competition. *Econometrica*, 75(5):1331–1370.

Baswana, S., Chakrabarti, P. P., Chandran, S., Kanoria, Y., and Patange, U. (2019). Centralized admissions for engineering colleges in india. In *Proceedings of the 2019 ACM Conference on Economics and Computation*, pages 323–324. Association for Computing Machinery, New York.

Beggs, S., Cardell, S., and Hausman, J. (1981). Assessing the potential demand for electric cars. *Journal of Econometrics*, 17(1):1–19.

Ben-Akiva, M., de Palma, A., and Thisse, J.-F. (1989). Spatial competition with differentiated products. *Regional Science and Urban Economics*, 19(1):5–19.

Berry, C. R. (2005). School district consolidation and student outcomes: Does size matter? In Howell, W. G., editor, *Besieged: School Boards and the Future of Education Politics*, pages 56–80, Washington, DC. Brookings Institution Press.

Berry, C. R. and West, M. R. (2008). Growing pains: The school consolidation movement and student outcomes. *The Journal of Law, Economics, & Organization*, 26(1):1–29.

Berry, S. and Reiss, P. (2007). Empirical models of entry and market structure. In Armstrong, M. and Porter, R., editors, *Handbook of Industrial Organization*, volume 3, pages 1845–1886. Elsevier.

Berry, S. T. (1994). Estimating discrete-choice models of product differentiation. *The RAND Journal of Economics*, 25(2):242–262.

Bester, H., de Palma, A., Leininger, W., Thomas, J., and von Thadden, E.-L. (1996). A noncooperative analysis of Hotelling's location game. *Games and Economic Behavior*, 12(2):165–186.

Biró, P. (2008). Student admissions in Hungary as Gale and Shapley envisaged. Technical Report TR-2008-291, Department of Computing Science, University of Glasgow. Retrieved from researchgate.net/profile/Peter_Biro4 on 2 June 2020.

Biró, P. (2012). Matching practices for secondary schools – Hungary. Technical report, MiP Country Profile 6. Retrieved from matching-in-practice.eu on 2 June 2020.

Biscaia, R. and Mota, I. (2013). Models of spatial competition: A critical review. *Papers in Regional Science*, 92(4):851–871.

Brasington, D. M. (1999). Joint provision of public goods: the consolidation of school districts. *Journal of Public Economics*, 73(3):373 – 393.

Braun, S., Dwenger, N., and Kübler, D. (2010). Telling the truth may not pay off: An empirical study of centralized university admissions in Germany. *The BE Journal of Economic Analysis & Policy*, 10(1).

Brekke, K. R., Nuscheler, R., and Rune Straume, O. (2006). Quality and location choices under price regulation. *Journal of Economics & Management Strategy*, 15(1):207–227.

Brekke, K. R., Siciliani, L., and Straume, O. R. (2011). Hospital competition and quality with regulated prices. *The Scandinavian Journal of Economics*, 113(2):444–469.

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

Bucarey, A. (2018). Who pays for free college? Crowding out on campus. Working paper, MIT Department of Economics. Retrieved from economics.mit.edu/files/14234 on 2 June 2020.

Bukodi, E., Róbert, P., and Altorjai, S. (2008). The hungarian educational system and the implementation of the isced-97. *The International Standard Classification of Education*, pages 200–215.

Bundesamt für Kartographie und Geodäsie, editor (2018). *Digitales Landschaftsmodell 1:250 000.* © GeoBasis-DE / BKG.

Burgess, S., Greaves, E., Vignoles, A., and Wilson, D. (2015). What parents want: School preferences and school choice. *The Economic Journal*, 125(587):1262–1289.

BZ (2017). Fulminanter Börsenstart für die Zur-Rose-Aktie. *Börsen-Zeitung*, (128):18. Retrieved from wiso-net.de/document/BOEZ__2017128092 on 2 June 2020 (login required).

Cameron, A. C. and Trivedi, P. K. (2005). *Microeconometrics. Methods and Applications.* Cambridge University Press.

Castillo, M. and Dianat, A. (2016). Truncation strategies in two-sided matching markets: Theory and experiment. *Games and Economic Behavior*, 98:180–196.

Chade, H., Lewis, G., and Smith, L. (2014). Student portfolios and the college admissions problem. *Review of Economic Studies*, 81(3):971–1002.

Che, Y.-K. and Koh, Y. (2016). Decentralized college admissions. *Journal of Political Economy*, 124(5):1295–1338.

Che, Y.-K. and Tercieux, O. (2019). Efficiency and stability in large matching markets. *Journal of Political Economy*, 127(5):2301–2342.

Chen, J. (2013). Estimation of the loan spread equation with endogenous bank-firm matching. *Advances in Econometrics*, 3:251–289.

Chen, L. and Pereyra, J. S. (2019). Self-selection in school choice. *Games and Economic Behavior*, 117:59–81.

Chiappori, P.-A. and Salanié, B. (2016). The econometrics of matching models. *Journal of Economic Literature*, 54(3):832–61.

Chisholm, D. C. and Norman, G. (2012). Spatial competition and market share: an application to motion pictures. *Journal of Cultural Economics*, 36(3):207–225.

Christaller, W. (1933). *Die zentralen Orte in Süddeutschland.* Fischer, Jena.

Coenen, M., Haucap, J., Herr, A., and Kuchinke, B. A. (2011). Wettbewerbspotenziale im deutschen Apothekenmarkt. *ORDO*, 62(1):205–230.

Correa, J., Epstein, R., Escobar, J., Rios, I., Bahamondes, B., Bonet, C., Epstein, N., Aramayo, N., Castillo, M., Cristi, A., et al. (2019). School choice in Chile. In *Proceedings of the 2019 ACM Conference on Economics and Computation*, pages 325–343. Association for Computing Machinery, New York.

Cox, B. (2010). A decade of results: a case for school district consolidation? *Education*, 131(1):83–92.

Crawford, G. S. (2012). Endogenous product choice: A progress report. *International Journal of Industrial Organization*, 30(3):315–320.

d'Aspremont, C., Gabszewicz, J. J., and Thisse, J.-F. (1979). On Hotelling's "Stability in Competition". *Econometrica*, 47(5):1145–1150.

Davis, P. (2006). Spatial competition in retail markets: movie theaters. *The RAND Journal of Economics*, 37(4):964–982.

de Palma, A., Ginsburgh, V., Papageorgiou, Y. Y., and Thisse, J.-F. (1985). The principle of minimum differentiation holds under sufficient heterogeneity. *Econometrica*, 53(4):767–781.

Destatis (2017). Unternehmen, Beschäftigte, Umsatz und weitere betriebs- und volkswirtschaftliche Kennzahlen im Handel: Deutschland, Jahre, Wirtschaftszweige, Umsatzgrößenklassen. Tabelle 45341-0001. *Genesis-Online Datenbank*.

Deutscher Apotheker Verlag (2001-2016). Bundesapothekenregister.

Diggle, P. J. (2014). *Statistical Analysis of Spatial Point Patterns*. CRC Press.

Doornik, J. A. and Hansen, H. (2008). An omnibus test for univariate and multivariate normality. *Oxford Bulletin of Economics and Statistics*, 70(s1):927–939.

Doraszelski, U. and Escobar, J. F. (2010). A theory of regular markov perfect equilibria in dynamic stochastic games: Genericity, stability, and purification. *Theoretical Economics*, 5(3):369–402.

Doraszelski, U. and Judd, K. L. (2012). Avoiding the curse of dimensionality in dynamic stochastic games. *Quantitative Economics*, 3(1):53–93.

Doraszelski, U. and Pakes, A. (2007). A framework for applied dynamic analysis in IO. volume 3 of *Handbook of Industrial Organization*, pages 1887–1966. Elsevier.

Doraszelski, U. and Satterthwaite, M. (2010). Computable markov-perfect industry dynamics. *The RAND Journal of Economics*, 41(2):215–243.

Doğan, B. and Yenmez, B. (2017). Unified enrollment in school choice: How to improve student assignment in Chicago. Working paper. Retrieved from ssrn.com/abstract=2999373 on 4 June 2020.

Duncombe, W., Miner, J., and Ruggiero, J. (1995). Potential cost savings from school district consolidation: A case study of New York. *Economics of Education Review*, 14(3):265–284.

Eaton, B. C. and Lipsey, R. G. (1975). The principle of minimum differentiation reconsidered: Some new developments in the theory of spatial competition. *The Review of Economic Studies*, 42(1):27.

Economides, N. (1986). Nash equilibrium in duopoly with products defined by two characteristics. *The RAND Journal of Economics*, 17(3):431–439.

Ekmekci, M. and Yenmez, B. (2019). Common enrollment in school choice. *Theoretical Economics*, 14(4):1237–1270.

Ericson, R. and Pakes, A. (1995). Markov-perfect industry dynamics: A framework for empirical work. *The Review of Economic Studies*, 62(1):53–82.

Fack, G., Grenet, J., and He, Y. (2019). Beyond truth-telling: Preference estimation with centralized school choice and college admissions. *American Economic Review*, 109(4):1486–1529.

Fairman, J. C. and Donis-Keller, C. (2012). School district reorganization in maine: Lessons learned for policy and process. *Maine Policy Review*, 21(2):24–40.

Fernandez, R. and Rodrik, D. (1991). Resistance to reform: Status quo bias in the presence of individual-specific uncertainty. *The American Economic Review*, 81(5):1146–1155.

Fischer, K. (2011). Central places: The theories of von Thünen, Christaller, and Lösch. In Eiselt, H. A. and Marianov, V., editors, *Foundations of Location Analysis*, pages 471–505, Heidelberg. Springer.

Fox, J. T. (2009). Structural empirical work using matching models. In Durlauf, S. N. and Blume, L. E., editors, *The New Palgrave Dictionary of Economics*. Palgrave Macmillan.

Gale, D. and Shapley, L. S. (1962). College admissions and the stability of marriage. *The American Mathematical Monthly*, 69(1):9–15.

Ghemawat, P. and Nalebuff, B. (1985). Exit. *The RAND Journal of Economics*, 16(2):184–194.

Gordon, N. and Knight, B. (2009). A spatial merger estimator with an application to school district consolidation. *Journal of Public Economics*, 93(5-6):752–765.

Haeringer, G. and Klijn, F. (2009). Constrained school choice. *Journal of Economic theory*, 144(5):1921–1947.

Hafalir, I., Kojima, F., and Yenmez, B. (2019). Interdistrict school choice: A theory of student assignment. Working paper. Retrieved from ssrn.com/abstract=3307731 on 2 June 2020.

Hafalir, I. E., Hakimov, R., Kübler, D., and Kurino, M. (2018). College admissions with entrance exams: Centralized versus decentralized. *Journal of Economic Theory*, 176:886–934.

Hastings, J. S., Neilson, C. A., and Zimmerman, S. D. (2013). Are some degrees worth more than others? Evidence from college admission cutoffs in Chile. Working Paper 19241, National Bureau of Economic Research.

He, Y. and Magnac, T. (2019). Application costs and congestion in matching markets. Working Paper 17-870, Toulouse School of Economics. Retrieved from tse-fr.eu/publications on 2 June 2020.

Hiller, V. and Tercieux, O. (2014). Choix d'écoles en France. *Revue économique*, 65(3):619–656.

Ho, K. and Ishii, J. (2011). Location and competition in retail banking. *International Journal of Industrial Organization*, 29(5):537–546.

Hopenhayn, H. A. (1992). Entry, exit, and firm dynamics in long run equilibrium. *Econometrica*, 60(5):1127–1150.

Horvath, D. (2010). *Die regulierte Apothekenversorgung in Deutschland*. Dissertation, Universität Duisburg-Essen.

Hotelling, H. (1929). Stability in competition. *The Economic Journal*, 39(153):41–57.

Hotz, V. J. and Miller, R. A. (1993). Conditional choice probabilities and the estimation of dynamic models. *The Review of Economic Studies*, 60(3):497.

Howley, C., Johnson, J., and Petrie, J. (2011). Consolidation of schools and districts: What the research says and what it means. *National Education Policy Center*.

IQVIA (2019). The global use of medicine in 2019 and outlook to 2023. Forecasts and areas to watch. Technical report, IQVIA Institute for Human Data Science. Retrieved from iqvia.com/insights on 2 June 2020.

Irmen, A. and Thisse, J.-F. (1998). Competition in multi-characteristics spaces: Hotelling was almost right. *Journal of Economic Theory*, 78(1):76–102.

Kertesi, G. and Kézdi, G. (2013). School segregation, school choice, and educational policies in 100 Hungarian towns. Working paper, Institute of Economics, Hungarian Academy of Sciences. Retrieved from econ.core.hu/kiadvany/bwp on 4 June 2020.

Kirkebøen, L. (2012). Preferences for lifetime earnings, earnings risk and nonpecuniary attributes in choice of higher education. Discussion Paper 725, Statistics Norway. Retrieved from hdl.handle.net/10419/192707 on 4 June 2020.

Kulke, E. (2013). *Wirtschaftsgeographie*. UTB, Stuttgart.

Lancaster, T. (2004). *An Introduction to Modern Bayesian Econometrics*. Blackwell.

Leach, J., Payne, A. A., and Chan, S. (2010). The effects of school board consolidation and financing on student performance. *Economics of Education Review*, 29(6):1034–1046.

Logan, J. A., Hoff, P. D., and Newton, M. A. (2008). Two-sided estimation of mate preferences for similarities in age, education, and religion. *Journal of the American Statistical Association*, 103(482):559–569.

Lösch, A. (1940). *Die räumliche Ordnung der Wirtschaft*. Fischer, Jena.

Lucas, R. E. (1976). Econometric policy evaluation: A critique. In *Carnegie-Rochester conference series on public policy*, volume 1, pages 19–46.

Luxen, D. and Vetter, C. (2011). Real-time routing with OpenStreetMap data. In *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pages 513–516. Association for Computing Machinery, New York.

Machado, C. and Szerman, C. (2018). Centralized admissions and the student-college match. Working paper. Retrieved from ssrn.com/abstract=2844131 on 4 June 2020.

Manjunath, V. and Turhan, B. (2016). Two school systems, one district: What to do when a unified admissions process is impossible. *Games and Economic Behavior*, 95:25–40.

Maskin, E. and Tirole, J. (1988). A theory of dynamic oligopoly, II: Price competition, kinked demand curves, and Edgeworth cycles. *Econometrica*, 56(3):571–599.

McGill, R., Tukey, J. W., and Larsen, W. A. (1978). Variations of box plots. *The American Statistician*, 32(1):12–16.

Menzel, K. (2015). Large matching markets as two-sided demand systems. *Econometrica*, 83(3):897–941.

Menzel, K. and Salz, T. (2013). Robust decisions with incomplete structural models of strategic interactions. Working paper, Department of Economics, New York University. Retrieved from economics.sas.upenn.edu on 4 June 2020.

Mora, R. and Romero-Medina, A. (2001). Understanding preference formation in a matching market. Working paper, Departamento de Economía, Universidad Carlos III de Madrid. Retrieved from hdl.handle.net/10016/262.

Netz, J. S. and Taylor, B. A. (2002). Maximum or minimum differentiation? Location patterns of retail outlets. *The Review of Economics and Statistics*, 84(1):162–175.

OpenStreetMap contributors (2017). Open Street Map data extracts. Retrieved from download.geofabrik.de in 2017.

Ortega, J. (2018). Social integration in two-sided matching markets. *Journal of Mathematical Economics*, 78:119–126.

Ortega, J. (2019). The losses from integration in matching markets can be large. *Economics Letters*, 174:48–51.

Ottino-Loffler, B., Stonedahl, F., Veetil, V. P., and Wilensky, U. (2017). Spatial competition with interacting agents. Working paper. Retrieved from ssrn.com/abstract=2697263 on 4 June 2020.

Pakes, A. and McGuire, P. (1994). Computing Markov-perfect nash equilibria: Numerical implications of a dynamic differentiated product model. *The RAND Journal of Economics*, 25(4):555–589.

Pakes, A. and McGuire, P. (2001). Stochastic algorithms, symmetric markov perfect equilibrium, and the 'curse' of dimensionality. *Econometrica*, 69(5):1261–1281.

Pakes, A., Ostrovsky, M., and Berry, S. (2007). Simple estimators for the parameters of discrete dynamic games (with entry/exit examples). *The RAND Journal of Economics*, 38(2):373–399.

Panteli, D., Arickx, F., Cleemput, I., Dedetn, G., Eckhardt, H., Fogarty, E., Gerkens, S., Henschke, C., Hislop, J., Jommi, C., Kaitelidou, D., Kawalec, P., Keskimäki, I., Kroneman, M., Bastida, J. L., Barros, P. P., Ramsberg, J., Schneider, P., Spillane, S., Vogler, S., Vuorenkoski, L., Kildemoes, H. W., Wouters, O., and Busse, R. (2016). Pharmaceutical regulation in 15 European countries. *Health Systems in Transition*, 18(5). Retrieved from eprints.lse.ac.uk/68290 on 4 June 2020.

Riedel, A., Schneider, K., Schuchart, C., and Weishaupt, H. (2010). School choice in German primary schools. How binding are school districts? *Journal for Educational Research Online*, 2(1):94–120.

Roth, A. and Sotomayor, M. (1992). *Two-sided matching: A study in game-theoretic modeling and analysis*. Cambridge University Press.

Rushton, G. (1972). Map transformations of point patterns: Central place patterns in areas of variable population density. *Papers in Regional Science*, 28(1):111–132.

Rust, J. (1987). Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher. *Econometrica*, 55(5):999–1033.

Schaumans, C. and Verboven, F. (2008). Entry and regulation: evidence from health care professions. *The RAND Journal of Economics*, 39(4):949–972.

Seim, K. (2006). An empirical model of firm entry with endogenous product-type choices. *The RAND Journal of Economics*, 37(3):619–640.

Shapiro, S. S. and Francia, R. S. (1972). An approximate analysis of variance test for normality. *Journal of the American Statistical Association*, 67(337):215–216.

Siegel-Hawley, G., Bridges, K., and Shields, T. J. (2017). Solidifying segregation or promoting diversity? School closure and rezoning in an urban district. *Educational Administration Quarterly*, 53(1):107–141.

Smithies, A. (1941). Optimum location in spatial competition. *Journal of Political Economy*, 49(3):423–439.

Söderström, M. and Uusitalo, R. (2010). School choice and segregation: Evidence from an admission reform. *Scandinavian Journal of Economics*, 112(1):55–76.

Sørensen, M. (2007). How smart is smart money? A two-sided matching model of venture capital. *The Journal of Finance*, 62(6):2725–2762.

Sprumont, Y. (1990). Population monotonic allocation schemes for cooperative games with transferable utility. *Games and Economic Behavior*, 2(4):378–394.

Stata Corp., editor (2017). *Stata 15 Base Reference Manual*. Stata Press.

Statistische Ämter des Bundes und der Länder (2018). Regionaldatenbank Deutschland. Retrieved from regionalstatistik.de on 4 June 2020.

Tamer, E. (2003). Incomplete simultaneous discrete response model with multiple equilibria. *The Review of Economic Studies*, 70(1):147–165.

Thomadsen, R. (2007). Product positioning and competition: The role of location in the fast food industry. *Marketing Science*, 26(6):792–804.

Train, K. E. (2009). *Discrete Choice Methods with Simulation.* Cambridge University Press, second edition.

Turhan, B. (2019). Welfare and incentives in partitioned school choice markets. *Games and Economic Behavior*, 113:199–208.

van Buuren, S. and Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software, Articles*, 45(3):1–67.

Vogel, J. (2008). Spatial competition with heterogeneous firms. *Journal of Political Economy*, 116(3):423–466.

Waslander, S. and Thrupp, M. (1995). Choice, competition and segregation: An empirical analysis of a New Zealand secondary school market, 1990-93. *Journal of Education Policy*, 10(1):1–26.

Weber, A. and Pick, G. (1909). *Reine Theorie des Standorts.* Mohr, Tübingen.

Weintraub, G. Y., Benkard, C. L., and Van Roy, B. (2008). Markov perfect industry dynamics with many firms. *Econometrica*, 76(6):1375–1411.

Weintraub, G. Y., Benkard, C. L., and Van Roy, B. (2010). Computational methods for oblivious equilibrium. *Operations Research*, 58(4-part-2):1247–1265.

Weldon, M. (2016). *School choice, competition and ethnic segregation in Lancashire: Evidence from structural models of two-sided matching*. Dissertation, Lancaster University.

Wilson, L. (1972). An analysis of the stable marriage assignment algorithm. *BIT Numerical Mathematics*, 12(4):569–575.

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and panel data.* MIT Press, Cambridge, Massachusetts, second edition.

Yao, G. and Böckenholt, U. (1999). Bayesian estimation of Thurstonian ranking models based on the Gibbs sampler. *British Journal of Mathematical and Statistical Psychology*, 52(1):79–92.

Zensus (2011). Bevölkerung im 100 Meter-Gitter. Retrieved from zensus2011.de on 16 July 2017.

Ehrenwörtliche Erklärung

Ich versichere hiermit, dass ich die Dissertation selbstständig und ohne Benutzung anderer als der angegebenen Quellen und Hilfsmittel angefertigt und die den benutzten Quellen wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe. Diese Arbeit hat in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen.

Ich bin damit einverstanden, dass diese Arbeit zum Zwecke des Plagiatsabgleiches in elektronischer Form verarbeitet wird.

Mannheim, den 11. Juni 2020 Jan Robert Aue