Estimation of Learning, Adoption and Diffusion over a Network

Nir Hak*

December 23, 2019

Click here for latest version.

Abstract

Firms often decide whether to adopt an innovation of uncertain value in markets where the outcomes of earlier adopters are observed. This paper introduces a flexible Bayesian model suitable for the analysis of social learning, competition, and diffusion in such environments. Agents in the model have (potentially misspecified) theories of how others' profits relate to their own, and use these to make their adoption decisions. When adopting, agents steal business from and inform others. I estimate the model exploiting a unique reform in Illinois that legalized slot machines, and empirically study how information and adoption diffuse through a network. This setting is well-suited for such analysis, since gambling data are publicly available, adoption is a discrete action, and the set of potential adopters (liquor license holders) is defined by law. I find that establishments that observe more adoption or higher neighbors' profits are more likely to adopt themselves, yet learning could improve since they do not use all the relevant information. Establishments have diffuse priors and they learn from more neighbors than they compete with. The direction and extent to which learning affects adoption are ex-ante ambiguous. In two counterfactual exercises I show that increasing information availability or learning substantially increases both adoption and total profits in the market.

*Department of Economics, Littauer Center, 1805 Cambridge Street, Cambridge MA 02138. Email: nirhak@g.harvard.edu. I am particularly indebted to my advisors Ariel Pakes, Robin Lee, and Ben Golub for invaluable guidance and support on this project. I also wish to thank Alexander MacKay, Krishna Dasaratha, Isaiah Andrews, Jon Roth, Sarah Armitage, Shengwu Li, Myrto Kalouptsidi, Yotam Shem-Tov, Tzachi Raz, Moya Chin, Pierre Dubois, Mark Shepard, Tianwang Liu, and Scott Kominers.

1 Introduction

Firms often have to make a decision whether to invest in a new product or technology. This investment could be in the adoption of a new product to sell, or in research and development. Firm typically have some uncertainty about the future outcome of such investment, and would prefer investments that yield higher profits with greater certainty. These decisions are not made in a vacuum—in many cases there are similar firms in the market that have already made similar decisions. Therefore, past investment decisions made by some firms could provide information to other firms. The information spillovers could affect their future decisions. A change in learning behavior would then affect the diffusion patterns of investment or adoption of the new product or technology. Adoption and diffusion of new technology have been extensively documented and studied back to at least Griliches (1957); the same is true of the effects of social learning on the actions taken by individuals or firms (Foster and Rosenzweig, 1995; Conley and Udry, 2010; Kellogg, 2011). While extensively studied in theoretical literature (e.g. by Board and Meyer-ter Vehn (2018); Sadler (2019)), there is limited empirical research on how social learning affects adoption and the its impact on market outcomes.

The extent to which agents learn depends on two factors: the availability of information, and the agents' interpretation of the available information. Learning occurs when firms believe that their potential profits are correlated with the profits of other firms. Therefore, when firms have informative observations of other firms' profits, they update their beliefs about own potential profits. The updating process is based on the firm's perceptions about this correlation in profits. This need not reflect the true correlation of profits. Modeling the possibility of misspecified beliefs is key to understanding firms' learning behavior, which subsequently affects their observed decisions.

This paper develops a model in which agents (firms) in a competitive market socially learn about the profitability of a new product and decide whether to adopt the product. An agent that adopts both competes with and informs others. To distinguish between the two effects I use the fact that realized profits are only affected by competition and not by learning. When deciding whether to adopt, agents are risk averse and fully Bayesian, and believe that their profits are spatially correlated with the profits of neighboring agents. The model allows the agents' priors to be misspecified—their beliefs about profits' spatial covariance could be different from the true spatial covariance. I use this property to distinguish between agents' perceptions and the truth, which is in turn used to quantify learning and its effects on decisions. Agents in the model, that have not yet adopted, observe the profits of agents that have already done so and dynamically form posteriors on their own potential profits. Given these beliefs and local competition, agents then decide whether to adopt. To the best of my knowledge, this is a first attempt to estimate the perceptions that explain the behavior of Bayesian firms using real-world market data.

I estimate this model with data from the slot machine industry in the state of Illinois. According to the available data, establishments slowly adopt slot machines over a few years. There are several properties that make this setting ideal for such an analysis. First, the set of potential adopters is defined by state law: only liquor license holders can install slot machines in their establishments. Secondly, the adoption decision is clearly defined, as adoption of slot machines requires an application to be approved by the Illinois Gaming Board (IGB). Finally, state regulation requires that adoption decisions and monthly gambling profits be publicly disclosed. Therefore, all market participants starting from the inception of this market are observed. Most importantly, the public information allows the agents in the model to observe others' profits, since the data can be accessed freely on the IGB's website. This is a rare environment where data about small retail businesses is available, shedding light on the decisions owners make and how they affect profits of other market participants. However, this type of learning is not limited to either small firms or profit observation; firms of any size could potentially learn from any outcome that is correlated with profits.[1]

I model agents' beliefs about profits as determine by information from and competition with others. To separate the effects estimation is carried broken into steps. In the first step, I estimate the competition levels, that is, what an establishment should expect in the way of business stealing when neighbors of different distances and types decide to adopt. I exploit the variation in the number of neighbors of an establishments that adopt and their distances to estimate how profits change with increased competition. This part of the estimation relies on the feature of the environment that establishment owners do not directly engage in price-setting behavior. I find that business stealing exists in the market and significantly affects establishments' profits. Moreover, the business stealing between establishments of the same type (e.g., two bars) is three times as large as between establishments of different types (e.g., a bar and a restaurant).

A selection problem arises when estimating the rest of the model. I observe only the profits of agents that decided to adopt, and the adoption decision depends on their beliefs when they choose to adopt. The beliefs are unobserved to the econometrician. In the second step of the estimation, I use variation in establishments' characteristics, and a mild assumption on the individuals' initial (time-zero) beliefs, to identify profit-relevant parameters. To pin down these parameters, I develop a new method that corrects for the selection problem us-

[1] Outcomes could include: sales, revenues, customers traffic, etc. For example, auto manufacturers that try to decide if the invest in electric engines, observing sales of competitors of different similarity.

ing a fixed-point algorithm. This algorithm imposes consistency of agents' beliefs with their adoption behavior. The final step is done using maximum likelihood estimation, maximizing the probability of observing the adoption patterns observed in the data. Steps 2 and 3 are intertwined: the fixed point in step 2 pins down the profit parameters given agents' prior beliefs, while step 3 pins down the parameters that define the agents' priors.

Using the results of the model, I analyze two counterfactual environments. In the first exercise, agents cannot learn from one another, simulating a baseline case of no information. In many markets firms are not obligated to reveal their profit data, which makes learning from others a harder task. This exercise demonstrates the learning effects on adoption and profits we can expect from policies that require firms to report profits or other outcome data. In the second counterfactual exercise, I examine how adoption and total profits would change if agents knew the true correlations between their profits and those of other firms. Compared to an environment without learning, this is the change that would expected if agents in a market are more sophisticated, or if the product is similar to existing products sold in the market and agents have better information about the correlations. In an environment in which agents are not sophisticated, this counterfactual measures the potential benefits from an intervention that informs the agents (such as newsletters or workshops that make profits in the area and their meaning more salient).

I find that the maximum distance at which firms compete with one another (the competition radius) is much smaller than their learning radius. Although firms compete only with firms in a radius of about 0.7 miles, they learn from neighbors much further away, in a radius of up to about 2 miles away. At the same time, the actual spatial correlation of firms' profits completely decays after about 13 miles. This means that agents ignore a lot of the available information and that learning could be improved. I also check how well agents' expected profits (in the period in which they choose to adopt) predict their realized profits. I find that, on average, agents are not far from being correct: for every extra dollar agents expect to earn, their realized profits are on average higher by about 85 cents.

Given the point estimates, I simulated two counterfactual environments. When agents do not observe their neighbors, and therefore cannot learn, the total adoption over the period of analysis drops by about 4.6%, while total market profits are expected to drop by 3.4%. The yearly net revenue of the slot machine market in Illinois is greater than $1 billion, which means that tens of millions of dollars of yearly revenue could be attributed to the government's decision to make the data publicly available. In the second simulation exercise, I checked how diffusion and profit patterns would change if agents' priors were correctly specified. If agents perfectly interpreted the information they observe about their neighbors' profits, total adoption would increase by about 3.1% compared to their baseline behavior, and total profits would increase by 3.6%.

Prior to estimating the model, I use features of the data to examine the adoption patterns. I showed that establishments that observe more adoption in their surrounding area or higher neighbors' profits are more likely to adopt themselves. Although these correlative findings are consistent with learning, they are also consistent with other stories. For example, adoption pattern is due to installation cost shock that leads to higher adoption in the area. To rule out this story and other alternative explanations that are consistent with heterogeneity in individual preferences and shocks that are spatially correlated, I also employ a semi-parametric method from Pakes and Porter (2016). I used this method to identify a lower bound on the level of learning using moment inequalities.

I model the decisions of the establishments as a simple binomial choice and use a simple revealed preferences idea: if agent i decides to adopt in some period t after not adopting in periods $s < t$, i's valuation of the choice "adopt" had to experience a more positive change than the choice "not adopt." The method's advantages in this setting stem from the richness of the unobserved heterogeneity that it allows. Namely, it allows for individual-choice fixed effects as well as serially correlated shocks and separately spatially correlated shocks to agents' utilities. The estimation results are that observing neighbors' adoption or higher profits serve as signals that affect adoption probability in the same way. In particular, comparing the period of adoption to earlier periods, observing extra neighbors' adoption has an effect equivalent to observing at least $2,700 in higher neighbors' yearly profits.

The structural model in this paper builds on ideas from the theoretical social learning and diffusion literatures. The social learning literature, an overview of which is in Golub and Sadler (2017), goes back to seminal papers of Bikhchandani et al. (1992); Banerjee (1992); Acemoglu et al. (2011); Bala and Goyal (1998); DeGroot (1974). This model set the foundations for many future learning models in which agents that are arranged in a network learn from one another. DeMarzo et al. (2003) and Molavi et al. (2018), for example, model agents that observe one signal at time $t = 0$ and learn about a fixed state of the world.[2] As in Harel et al. (2017), my model has agents that observe signals in every period and learn about a fixed state of the world. In all of the mentioned papers the state of the world is a scalar, while in my case the state is a vector (of everyone's profits) and each agent cares about one component of the vector (her own profit). The diffusion literature goes back to Bass (1969) and Bailey (1975), in which agents spontaneously adopt based on the fraction of adopters. Pastor-Satorras and Vespignani (2001) added a basic network structure that affects the probability of adoption. Conceptually, my paper is related to the environments in

[2] Frongillo et al. (2011) and Dasaratha et al. (2019) study agents that learn about a changing state of the world and observe a signal in every period.

Sadler (2019) and Board and Meyer-ter Vehn (2018) which analyze Bayesian agents that learn about the value of adoption from their neighbors and decide whether to adopt themselves.

Early papers on the diffusion of innovation took a reduced-form perspective. In Griliches (1957) and Coleman et al. (1966), diffusion of products is observed and the extent of it is estimated, while in Banerjee et al. (2013) a mechanical diffusion model is estimated using micro-level network data. The work by Bailey et al. (2019) and Kim et al. (2015) investigated how consumers' product adoption spreads, while Foster and Rosenzweig (1995) and Conley and Udry (2010) established that farmers learn from their neighbors. The estimation in this paper microfounds the learning process of the agents which influences their decisions, which in turn explains the diffusion process.

In the industrial organization literature, papers like Benkard (2000) and Doraszelski et al. (2018) structurally estimate how firms learn from own information and experience, though not from others. A paper that microfounds a diffusion process is Holmes (2011), which analyzes the diffusion of Walmart branches around the US. While both Holmes (2011) and my paper analyze the spatial introduction of a new store or type of product to the market, the mechanisms are very different. In my paper there are many separate one-time decisions made by individuals, while in his paper there is an underlying multi-period single-agent optimization problem.

Learning from others is examined in recent literature that investigates the decisions made by oil extraction firms. Methodologically the papers by Covert (2015) and Hodgson (2018) are closest to mine as they also use a Gaussian process to model a firm's beliefs about spatial correlation.[3] Both of those papers assume firms have perfect knowledge of the spatial correlation of oil reservoirs in the ground and perfectly learn from observing drilling outcomes. Covert (2015) uses the spatial correlation in the data as a Bayesian prior. Later, to rationalize his observations, he studies a heuristic model of agents who overweight their own signals. In this paper, I study a structural model with incorrect beliefs about correlation. I estimate the beliefs that Bayesian agents have that best explain their observed behavior. More fundamentally, the market and the questions in this paper are very different: this paper contributes to the understanding of the underlying processes and factors that affect adoption and diffusion of a technology or a product in a competitive market.

Outline. Section 2 describes the setting and the data. Section 3 provides descriptive evidence of spatial learning using non- and semi-parametric methods. Section 4 introduces the core empirical model of the paper. Section 5 discusses the identification and estimation

[3]In these two papers the outcomes from oil drilling or fracking in some spot is similar independent of the firm that chooses to do so, while in my paper the outcomes from adoption is unique for the establishments.

procedures. Section 6 presents and discusses the estimation results. In section 7, I simulate counterfactual environments. Section 8 discusses work in progress and future extensions. Finally, section 9 concludes.

2 Application: Slot Machines in Illinois

For many decades Nevada was the sole state allowing casino-like gambling. The opening of the first casino in 1978 in Atlantic City marked the beginning of a rapid expansion of the gambling industry in the US. In 2016, with 41 states that have casino-like gambling and more than $70 billion in yearly revenues,[4] gambling has become a major industry.

Casino gambling in Illinois started in September 1991, and there are 10 operating casinos in the state to date.[5] This industry has been a steady source of tax revenue to both the state and the municipalities, with more than $400 million of yearly tax revenue since 1999. In order to increase tax revenue, in July 2009 the state of Illinois enacted the Video Gaming Act that allows the installation of up to five video gaming terminals (slot machines) in retail establishments with pouring liquor licenses, veteran and fraternal establishments and Truck Stops. After lengthy legal and political battles, slot machines went live in September 2012. By the end of 2016 there were 24,840 active machines in 5,726 establishments. Slot machine gaming produced more than $1.1 billion of revenue in 2016, of which $330 million in tax payments went to the state and the local municipalities.[6]

An establishment owner that wishes to install slot machines is required to apply for a license with the IGB. The approval process takes usually one to two months, after which slot machines can be installed in the establishment by a terminal operator. Licensed by the IGB, a terminal operator is an entity that owns, installs, operates, and maintains slot machines in the establishments. By law, the post-tax revenue from the operation of the machines is split equally between the establishment's owner and the terminal operator.

Slot machine design has to comply with regulations defined in the Video Gaming Act. Slot machines provide low-stakes betting, with a maximum wager of $2 and a maximum jackpot of $500; lotteries also have to be i.i.d. with a minimum expected payout of 80% (i.e., the expected loss on a one-dollar bet is at most twenty cents). Moreover, all slot machines have to be connected to a centralized communication system at the IGB. The information

[4]This is based on the American Gaming Association's "State of the states" survey of the casino industry. These revenues include commercial and tribal casinos and video lottery terminals such as the ones described in this paper.

[5]Based on the IGB's 2018 annual report.

[6]States received 25% of total revenue in the form of taxes, while municipalities received 5%. Some municipalities also imposed application fees.

from the slot machines is automatically aggregated at the business level and advertised monthly on the IGB's website.

The slot machine market in Illinois holds several advantages for analyzing the adoption of a product in a new market. First, the set of potential adopters is well defined: only businesses that are allowed to install slot machines are those that hold a pouring liquor license.[7] Secondly, adoption is well defined: an establishment that applies for slot machines with the IGB is defined as an adopter. Moreover, the adoption decision is the only decision made by the establishment owner; after the adoption there are no pricing decisions made by the owners, and the operation of the machines is handled by the terminal operator. Finally, by regulation all of the data are released on a monthly basis and are publicly available for download, in the same website which establishment owners visit in order to apply for slot machine licenses.

Establishment owners may take into account factors other than gambling profits when deciding whether to adopt. A direct effect of installing slot machines is that they take up space which was used to make profits in some other way, most likely replacing tables or seats. Adoption could also have indirect effects: it could attract additional customers that wish to gamble, or keep customers for longer, which could lead to an increase in non-gambling profits. At the same time, customers who are deterred by gambling could be pushed away. Direct fixed costs are not a concern for owners, as the installation is at the expense of the terminal operator. However, in many cases establishment owners have some fixed costs, which could be due to improvements that they need or choose to make (such as an electricity upgrade or renovations).

2.1 Data and Descriptive Statistics

The analysis relies on data from three sources. The first is a panel dataset from the Illinois Department of Revenue containing a panel of liquor licenses in Illinois from 2008 to 2017, which defines the set of potential adopters. The dataset includes license numbers, addresses, effective and expiration dates, and the type of license (on- or off-premises consumption). The second source is Google Maps, a web-based map service that includes detailed information on businesses from which I gathered establishment characteristics. The third and main source is the IGB, from which I acquired multiple datasets:

- A panel dataset that contains monthly total gambling wagers and profits for all establishments that have ever adopted slot machines. The dataset is for the period

[7] As mentioned earlier, veteran and fraternal establishments as well as truck stops are also allowed to install slot machines. In practice, many veteran establishments have liquor licenses and are included in the analysis anyway, and the number of truck stops in the data is small.

	Mean	Median	S.D.
Bar profits	\$4,515	\$3,873	\$3,039
Others' profits	\$3,783	\$3,388	\$2,796
Establishments per municipality	11	4	19
Neighbors (3 miles)	45	32	42

Table 1: Summary Statistics
Summary statistics of the distribution of gambling profits of the establishments, split into bar-like types (bars, bar & grills, veterans establishments) and all other types (mostly restaurants). Neighbors are defined as nodes within a 3-miles radius of another node as used in section 3.

September 2012 through March 2017.

- A dataset that matches all liquor license numbers with gambling license numbers.
- A dataset that has application review dates for all applicants and the first day the machines went live.
- A dataset that has the municipalities that initially opted out (prohibited gambling), and if and when they opted in later.

Table 1 provides general summary statistics. Below are additional relevant institutional details and stylized facts:

Potential adopters and adoption. The set of potential adopters is defined based on the liquor-license dataset combined with the information on which municipalities opted out of the new regulation. The initial set of potential adopters includes only businesses that had on-premises consumption liquor licenses in September 2012, the month when machines were first allowed to be installed. The restriction on license type is due to the regulation allowing installation only in businesses with an on-premises consumption license. The restriction on the existence of a license in September 2012 is to keep the dataset homogeneous, as many other retail businesses (gas stations, laundromats, etc.) acquired a pouring liquor license for the express purpose of installing slot machines. The set of potential adopters in each period (month) was additionally restricted to account for municipalities opting in after the initial period, and for businesses that were shut down (those whose liquor license was not renewed) in later periods. In total there are 9,757 establishments in the dataset, of which 1,527 are in municipalities that initially prohibited slot machines but allowed them in later periods.

The adoption period of each establishment is defined by the month in which its application was reviewed by the board. This period is generally different from the first period of operation, as it usually takes one to two months from application review to first activation day. In the dataset a total of 4,286 establishments adopted within the time frame of the analysis. Figure 2.1 shows the pattern of adoption over time. About 75% of bar-like establishments adopted over the analysis period, while only 35% of the other businesses (mostly restaurants) adopted within that time frame.

Characteristics. The establishments in the data have different characteristics that are gathered from Google Maps. For each business, I gathered information about the price level of food and alcohol in the establishment[8], the number of reviews left on Google Maps by customers, and the average consumer rating of the business ("stars"). See figure 2.3 for details. For some of the businesses, some or all of the details are missing. I also observe the Census municipality level population data.

In addition to the aforementioned characteristics, I gathered data on each establishment's type, which is the definition of the business. There are 415 establishment types in the data (bar, restaurant, etc.), 1168 of which had no reported type. In the estimation, I used a total of ten types. The first eight types are the most prevalent ("biggest") types in the data. The remaining types represent the businesses that do not have type information in Google Maps, and the final type aggregates all the businesses of all other types. Therefore, each type's prevalence, average profits, and adoption rates presented in figure 2.2.

Theoretical and realized odds. By law, the expected value of each lottery has to be at least 0.8 of the wager and lotteries have to be i.i.d. In practice, the vast majority of businesses are far from this bound and, on average, the realized odds are about 0.92.

3 Reduced Form and Moment Inequalities

Learning occurs when agents observe other establishments and believe that their own outcomes are correlated with those of others. Therefore, observing higher than expected profits of others some nodes may choose to adopt. At the same time, adoption itself by others could be a positive signal about the profitability of slot machines in that area, the reaction that patrons have for slot machines[9], how installation changes the characteristics of the typical customer, etc.

The main contribution of this section is in establishing that there are patterns consistent with learning, and rule out few alternative explanations to learning. First, I show that the

[8] Represented by number of dollar signs. Cheap bars usually receive one dollar sign ("$"); high-end restaurants usually recieve three or four ($$$).

[9] In the model introduced in section 4 I assume that learning is only about gambling profits. In practice learning could be about different dimensions.

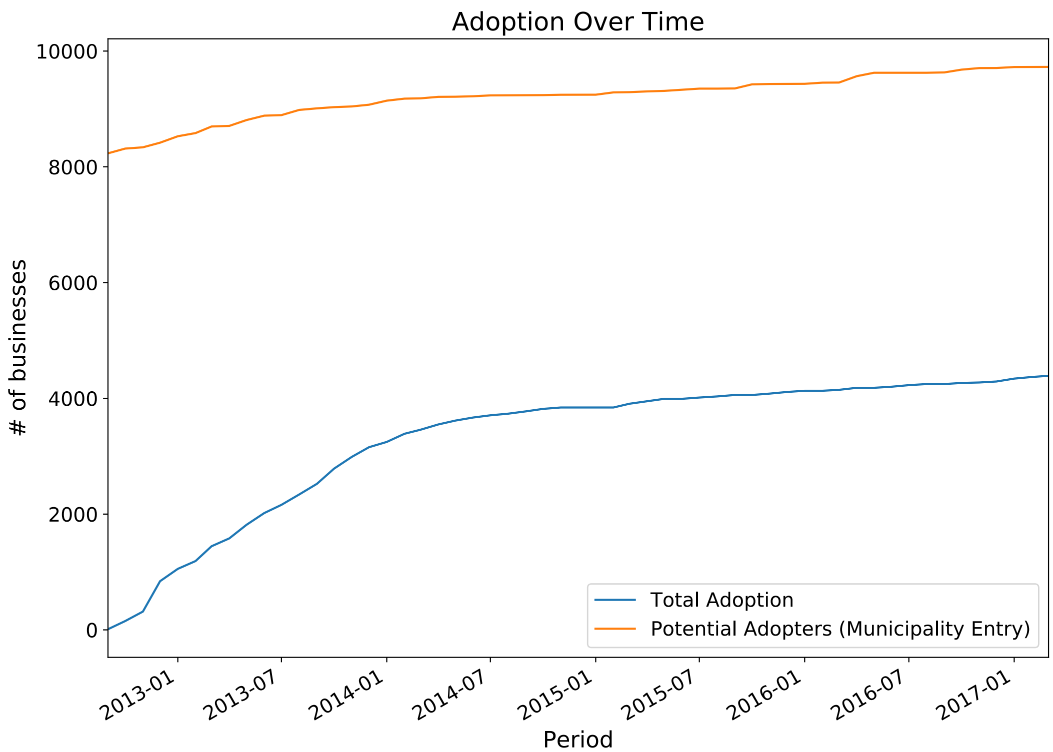


Figure 2.1: Potential Adoption and Adopters over Time
Vertical axis: number of businesses; horizontal axis: is the period. Orange: number of businesses that could adopt (potential set) over time as a result of municipalities opting in later. Blue: number of adopters up to and including the period.

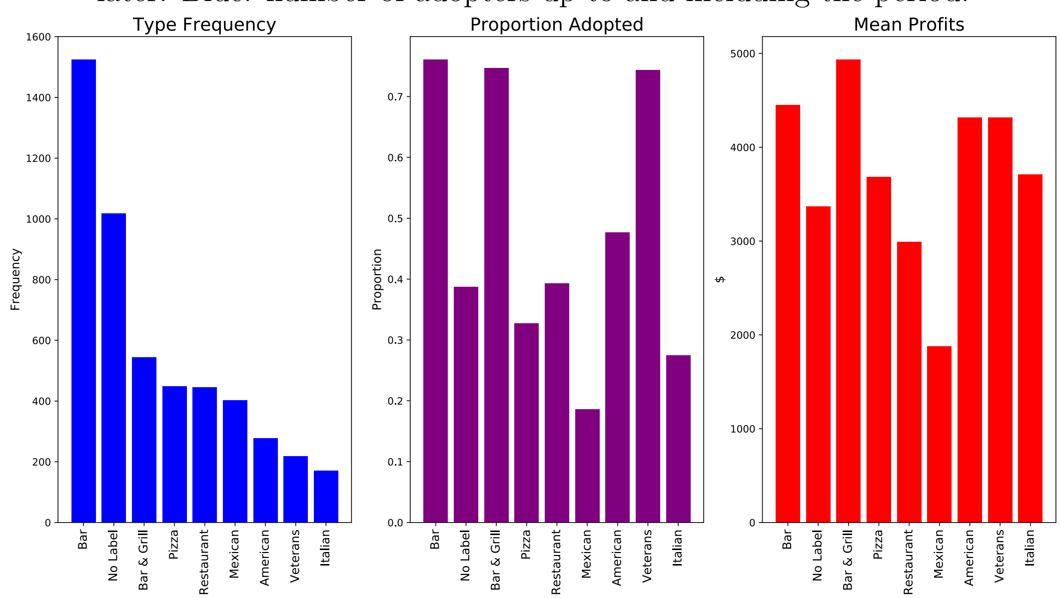


Figure 2.2: Statistics by Type
Left panel: frequency of establishments type. Middle panel: proportion of adopters of each type. Right panel: average profits of each type. The graphs are the 9 largest types.

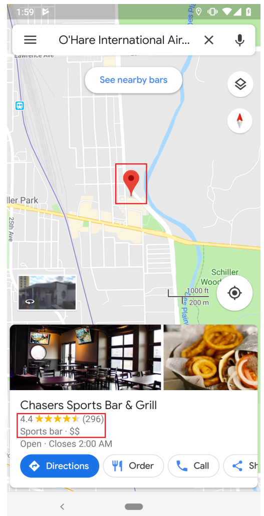


Figure 2.3: Google Maps Data
This figure represents the data gathered from Google Maps. Red squares: information used, which includes (in this example) the business type ("sports bar"), price level ("$$"), number of reviews (296), and business rating (4.4). The coordinates of the business were also extracted.

future adoption probability of nodes increases with both more adoption of neighbors and with higher neighbors' profits. I then turn to a semi-parametric estimation method of a simple binomial-choice model that is linear in observables. The estimation method is due to Pakes and Porter (2016) and relies on revealed preferences to identify the parameters. The method allows the errors to be serially and spatially correlated and allows for individual fixed effects.

3.1 Descriptive Evidence

This subsection provides descriptive evidence of information spillovers. Let i be a node, and let N_i be the set of nodes (or establishments) that lie within a 3-mile radius of i.[10] To check for information spillovers, I check whether i's probability of adoption changes with the number of i's neighbors that active in a period $A_{i,t} \subseteq N_i$. I also show how the adoption probability changes when the average profits in i's neighborhood change over time. To do so, I run the following simple regression:

$$\begin{aligned} a_{i,t+1} &= \underbrace{\beta_0 \Delta_{0,1} A_i + \beta_1 \Delta_{1,2} A_i + \beta_2 \Delta_{2,3} A_i + \beta_3 |A_{i,t-3}|}_{\text{Neighbors' adoption in recent periods}} \\ &+ \underbrace{\alpha_0 \Delta_{0,1} \overline{\pi}_i + \alpha_1 \Delta_{1,2} \overline{\pi}_i + \alpha_2 \overline{\pi}_{i,t-2}}_{\text{Neighbors' profits in recent periods}} + \underbrace{\gamma N_i}_{\#\text{ Neighbors}} + \epsilon_{i,t}, \end{aligned}$$

[10]The exercise was repeated with different radii and yielded similar results.

where $a_{i,t+1}$ is a binary variable that takes the value 1 if i adopted in period $t+1$ and 0 otherwise,[11] $\Delta_{s,s+1}A_i$ and $\Delta_{s,s+1}\overline{\pi}_i$ represent respectively the changes in the number of active neighbors of i between periods s and $s+1$, and $\overline{\pi}_{i,t}$ represents the average profits of i's neighbors in period t (in thousands of dollars).

As presented in table 2, observing more active neighbors that adopted in a recent periods is correlated with a higher probability of adoption. Column (1) suggests that the effect is substantial, with a baseline adoption probability of about 3%; observing any neighbor that adopted within the three previous periods increases the adoption probability by more than 20% (compared to the baseline). Observing higher average neighbors' profits in a period is also correlated with increased probability of adoption, though the magnitudes are smaller. The effect of changes in profits is small and disappears when period fixed effects are added (column (2)).

The regression results coincide with a story of information spillovers and resulting learning. Nevertheless, they do not rule out many alternative explanations, such as period cost shocks that induce higher adoption in some areas. In the next subsection I introduce a simple reduced form choice model and use a flexible method that puts minimal restrictions on the distribution of the errors in order to eliminate alternative stories.

3.2 Binomial-Choice Model

Let i be a node, with τ_i denoting the period in which i adopted slot machines. For every period $t \leq \tau_i$, each node had to make a decision whether to adopt or not; let $d \in \{0,1\}$ represent this decision, where 0 represents not adopting and 1 represents adopting. Since less than 1% of businesses removed slot machines after adoption, it is assumed that a node that adopts stays in forever. I check whether the number of neighbors of i that are active, denoted $|A^{\ell}_{i,t}|$, and their average profits, $\overline{\pi}^{\ell}_{i,t}$, had an effect on i's adoption decision.

Let $V_{d,i,t}$ denote i's perceived or expected value from choice d at period t. I assume first that i has additively separable preferences that take the form

$$V_{d,i,t} = d \cdot \left[\beta|A^{\ell}_{i,t-1}| + \alpha\overline{\pi}^{\ell}_{i,t-1}\right] + \lambda_{d,i} + \varepsilon_{d,i,t}.$$

The variable $\lambda_{d,i}$ is a fixed effect for each decision by individual i. It is important to note that the fixed effects are unrestricted so they could be spatially correlated.

The term $\varepsilon_{d,i,t}$ represents a decision and period shock, observed by i but not by the econometrician. The only distributional assumption is that $\boldsymbol{\varepsilon}_{i,\tau_i-k}|x_{i,\tau_i-k}, x_{i,\tau_i}, \boldsymbol{\lambda}_i \sim \boldsymbol{\varepsilon}_{i,\tau_i}|x_{i,\tau_i-k}, x_{i,\tau_i}, \boldsymbol{\lambda}_i$,

[11] Note that the regression panel is unbalanced, since if i adopts in period $t+1$ there are exactly $t+1$ observations of "group" i.

	$a_{i,t+1}$	
	(1)	(2)
$\Delta_{0,1}A_i$	0.007^{***}	0.003^{***}
	(0.001)	(0.001)
$\Delta_{1,2}A_i$	0.007^{***}	0.003^{***}
	(0.001)	(0.001)
$\Delta_{2,3}A_i$	0.006^{***}	0.002^{***}
	(0.001)	(0.001)
$A_{i,t-3}$	-0.0002^{***}	0.001^{***}
	(0.0001)	(0.0001)
$\Delta_{0,1}\overline{\pi}_i$	0.002^{***}	0.001
	(0.001)	(0.001)
$\Delta_{1,2}\overline{\pi}_i$	0.0003	-0.001
	(0.001)	(0.001)
$\overline{\pi}_{i,t-2}$	0.001^{**}	0.001^{**}
	(0.0003)	(0.0003)
$N_{i,t}$	-0.0001^{***}	-0.0003^{***}
	(0.00003)	(0.00003)
Constant	0.029^{***}	
	(0.002)	
Period Fixed Effect	NO	YES
Observations	182,102	182,102
R^2	0.010	0.022
Adjusted R^2	0.010	0.022
Residual Std. Error	0.174	0.173

Note: *p<0.1; **p<0.05; ***p<0.01

Table 2: Regression of Adoption on Neighborhood Statistics

where $\boldsymbol{\varepsilon}_{i,t} = \{\varepsilon_{0,i,t}, \varepsilon_{1,i,t}\}$, $\boldsymbol{\lambda}_i = \{\lambda_{0,i}, \lambda_{1,i}\}$, and $x_{i,t} = \left\{|A^{\ell}_{i,t-1}|, \overline{\pi}^{\ell}_{i,t-1}\right\}$. That is the errors have the same marginal distribution but can be freely serially correlated.[12]

Adding the assumption that agents are weakly rational is done for the purpose of imposing the following two inequalities:

$$V_{1,i,\tau_i} \geq V_{0,i,\tau_i} \text{ and } V_{1,i,\tau_i-k} \leq V_{0,i,\tau_i-k}.$$

That means that in the period in which i decided to adopt, her value from adoption was greater than her value from staying out, and that in every preceding period she had a positive value from staying out. Subtracting the inequalities yields the following inequality:

$$(V_{1,i,\tau_i} - V_{1,i,\tau_i-k}) - (V_{0,i,\tau_i} - V_{0,i,\tau_i-k}) \geq 0,$$

where taking expectations over this inequality yields the following inequality for every period k:

$$\frac{\beta}{\alpha}\left(|A^{\ell}_{i,\tau_i-1}| - |A^{\ell}_{i,\tau_i-1-k}|\right) + \overline{\pi}^{\ell}_{i,\tau_i-1} - \overline{\pi}^{\ell}_{i,\tau_i-1-k} + \Delta\varepsilon_{1,i} - \Delta\varepsilon_{0,i} \geq 0. \tag{3.1}$$

The coefficient, $\frac{\beta}{\alpha}$, puts a value on observing another neighbor that adopts. It defines an equivalence between seeing another neighbor of i that adopts and seeing that, on average, i's neighbors are making higher profits.

Estimation and interpretation To run this test, I define the set of neighbors that i could observe and learn from as nodes that lie within a 3-miles radius of i. (alternative radii were tested and yielded similar results.) I use the inequalities in (3.1) to define moments[13] for the estimation[14] of $\frac{\beta}{\alpha}$.

Since there is a large number of moments, I use the two-step method of Romano et al. (2014) that selects moments to use in calculating the bound on the parameters. The result of the estimation is that the lower bound on the 95% confidence set of $\frac{\beta}{\alpha}$ is greater than 45.3. This result means that when an agent observes another neighbor adopting, this information affects her probability of adoption in the same way as observing her neighbors' average profits increase by at least \$45.3 per month. Moreover, in the period of adoption a agent has on average 5 more neighbors than in periods she did not yet adopt. These facts mean that in the period i decides to adopt, the signals from others' adoption are equivalent, on average,

[12] For simplicity $\lambda_{d,i}, \varepsilon_{d,i,t}$ are additively separable, though in practice any function of the two, $f(\lambda_{d,i}, \varepsilon_{d,i,t})$, is allowed.

[13] I also use Mega Millions national-lotteries jackpots as an instrument, which doubles the number of moments.

[14] The estimation details, and a robustness check with a smaller number of moments, are presented in appendix A.

to i observing additional \$2,700 of her neighbors' yearly profits.

The flexibility of the model can account for spatial correlation and serial correlation, which combined withe the estimation results help to rule out some alternative explanations, such as the following:

- Adoption shocks: The profitability in an area is known ex ante, and adoption is due only to different cost shocks. This explanation is indeed consistent with observations of spatial correlation in adoption, though it is not consistent with the result that there is additional adoption due to increases in neighbors' profits.
- Imitation without learning: Businesses imitate each other as a result of changes in fashion (one business paints green, and then the one next door also paints green). This is another story that is consistent with spatial correlation, though not with the equivalence between observing higher neighbors' profits that lead to adoption.

Though this framework is flexible, it doesn't eliminate all alternative explanations. The literature on statistics of networks has shown that for every pattern of adoption over a network, there is a sequence of spatially and serially correlated shocks that would explain this pattern of adoption (Shalizi and Thomas, 2011). Explanations that are not eliminated include those in which shocks are spatially and serially correlated at the same time, such as traveling agents that drive between businesses over a period of months and persuade owners to install slot machines.

4 Structural Model

In the previous section, I established that there are patterns in the data consistent with learning: Owners that observe more agents that adopted or a higher profitability in their neighborhoods (as defined in Section 3) are more likely to adopt themselves. However, the estimates given there are only descriptive and cannot be used to make counterfactual inferences. Therefore, in this section I introduce a structural learning model that is used to study the extent to which learning affects adoption, and therefore how adoption patterns would change under different information structures or different learning behaviors. I now model the data generating process explicitly and use the estimates for counterfactual exercises.

Model overview. This is a model in which Bayesian agents decide in every period whether to adopt slot machines and install them in their establishment or not. The agents in the model believe that their profits are spatially correlated with the profits of their neighbors.

The beliefs that they have could be misspecified in one specific way: The true spatial correlation of the profits could be different from their beliefs about this correlation. Agents also compete with each other and are aware of the level of competition they would face when they make their adoption decisions.

In the rest of this section, I describe the environment and game. There is a set of municipalities, M, that are assumed to be observationally independent; therefore in this section I omit the municipality subscript. The estimation includes all municipalities and the municipality-specific notation is reintroduced in section 5.

I use small standard font letters represent scalars, small bold font letters represent vectors, and capital letters represent sets or matrices, which will be clarified explicitly or by context.

4.1 Description

Primitives. There is a set of agents $N = \{1, ..., n\}$. Each node $i \in N$ is associated with a deterministic vector of characteristics $\boldsymbol{x}_i$ and type θ_i. Each node's profit also has an unobservable persistent stochastic part to it, denoted by ξ_i.

The nodes are positioned on a plane; the distance between nodes i and j is denoted by d_{ij} where $d_{ii} = 0$.

Periods are discrete,

$$t \in \{0, 1, 2, ...\},$$

and at time $t = 0$ the vector of unobservables of all agents $\boldsymbol{\xi}$ is drawn from a joint normal distribution

$$\boldsymbol{\xi} \sim N\left(0, \Sigma_{\text{True}}\right),$$

where Σ_{True} is the true covariance matrix that the unobservables are drawn from.

The true correlation matrix, K_{True}, combined with the true scalar variance $\sigma^2_{\xi\text{True}}$, defines the covariance matrix; $\Sigma_{\text{True}} = \sigma^2_{\xi\text{True}} K_{\text{True}}$.

Information and observations. All agents have perfect knowledge of the characteristics and types of all agents and the full matrix of distances. Agents also have a common prior about the distribution from which the vector $\boldsymbol{\xi}$ is drawn, $N\left(0, \Sigma_0\right)$, where Σ_0 could be different from Σ_{True}.

At every period t there is a set $A_t \subseteq N$ which includes active adopters up to and including period t. The true gambling profits $\pi^g_{i,t}$ for agent $i \in A_t$ are

$$\pi^g_{i,t} = \xi_i + \boldsymbol{x}_i\boldsymbol{\beta} - C_i(A_t) + z_t + \epsilon^g_{j,t},$$

where $C_i(\cdot)$ is a function that represents the level of competition faced by i when a set A_t has already adopted, $\boldsymbol{\beta}$ is a vector of parameters, z_t is a period fixed effect, and

$$\epsilon^g_{i,t} \sim N\left(0, \sigma^2_\epsilon\right)$$

is an i.i.d. profit shock. The parameters $\boldsymbol{\beta}$, C_i, σ^2_ϵ, and z_t are known to the agents.

In period t the gambling profits $\pi^g_{i,t}$ of each agent that adopted up to and including that period, $i \in A_t$, are revealed and publicly observed. Additionally, agents in the model have perfect recall, hence they remember all observations, and since information is public, the information at period t is

$$\mathcal{I}_t = (\boldsymbol{\pi}^g_{A_s})^t_{s=0}.$$

The notation $\boldsymbol{\pi}^g_{A_s}$ denotes the vector of profits in period s of all agents that adopted up to and including period s.

Utility and adoption. When the game starts, the set of adopters is empty. Then, at every period $t \in \{0, 1, 2, ...\}$, each agent i that has not yet adopted ($i \notin A_t$) forms her estimate of $\xi_i|\mathcal{I}_t$.

In every period, each agent draws some period-specific i.i.d. adoption cost shock $\eta_{i,t}$ from some known common distribution. She observes this period-independent adoption cost shock and chooses to adopt if and only if her discounted expected utility from adoption given her information is greater than 0:

$$E\left[u\left(\pi^g_{i,t} + \Delta\pi^{core}_{i,t} + \eta_{i,t}\right)|\mathcal{I}_{t-1}\right] > 0,$$

Where $\Delta\pi^{\text{core}}_{i,t}$ represents the change in non-gambling profits as result of adoption (e.g. sales of food and beverages)[15]. Since all available information is public, it is sufficient to condition on the public information set $\mathcal{I}_{t-1}$ to obtain the best estimate of one's expected profits from adoption. When agent j adopts, her expectations about her own profits are known to all other agents in the municipality (since they also have perfect knowledge of her characteristics). Therefore, the value of η_{j,τ_j} in j's adoption period τ_j provides no information to others and does not affect posterior beliefs about profits, that is,

$$E\left[\pi^g_{i,s\geq t}|\mathcal{I}_t\right] = E\left[\pi^g_{i,s\geq t}|\mathcal{I}_t, \boldsymbol{\eta}_{A_t}\right], \tag{4.1}$$

where $\boldsymbol{\eta}_{A_t}$ represents the entire history of the adoption shocks of other agents, including when they adopted.

[15]The term is further discussed in subsection 4.4

If different agents had different (private) information at the time of adoption (or in the case that the adoption shocks $\eta_{i,t}$ were not i.i.d.) the equality in (4.1) would generally not hold.

4.2 Interpretation

In every period, agents that did not yet adopt decide whether to adopt. Adoption is assumed to be irreversible since less than 1% of adopters remove the slot machines (though many establishments shutdown entirely). This means that after adoption establishments make no decisions and their only role is in the information their profits provide to potential adopters.

One of the assumptions is that agents' information in every period is the full history of profits in their municipality. This assumption is made due to the structure in which all gambling profits are published online by the IGB. As seen in figure 4.1, profits can be observed in three ways: all establishments statewide, all establishments in a municipality, or a specific establishment. The first and last options provide too much or very little information and therefore less useful. Viewing profits in the municipality level would usually result an amount of information that is significant, and that could be parsed and therefore establishments are assumed to observe profits in their municipality.

Agents in the model form the best estimate of their profit in every period based on a common prior and the public information available in that period. The assumption that the mean of $\boldsymbol{\xi}$ is zero represents the idea that agents update their beliefs based on some deviation from their expectations they have for an establishment's profits given its characteristics.

The assumptions on common prior and public information generate an environment in which an agent's decision to adopt is "no surprise" to other agents. When j decides to adopt, since i perfectly knows j's characteristics and has the same information that is available to j, the adoption decision itself doesn't add any new information to i about her own unobservable ξ_i or about her expected profits. Only when i observes j's period profits does i update her beliefs about her own profits.

Finally, at this point the utility function can be fairly flexible. The utility function itself doesn't have a direct effect on the way that agents learn from the information they observe, it can affect the availability of information though.

4.3 Learning

The only object (other than profit i.i.d. shocks) the agents are uncertain about is their vector of unobservables, $\boldsymbol{\xi}$. This vector is drawn from a joint normal distribution. When agents observe some information about components of this vector, that is, other nodes' period

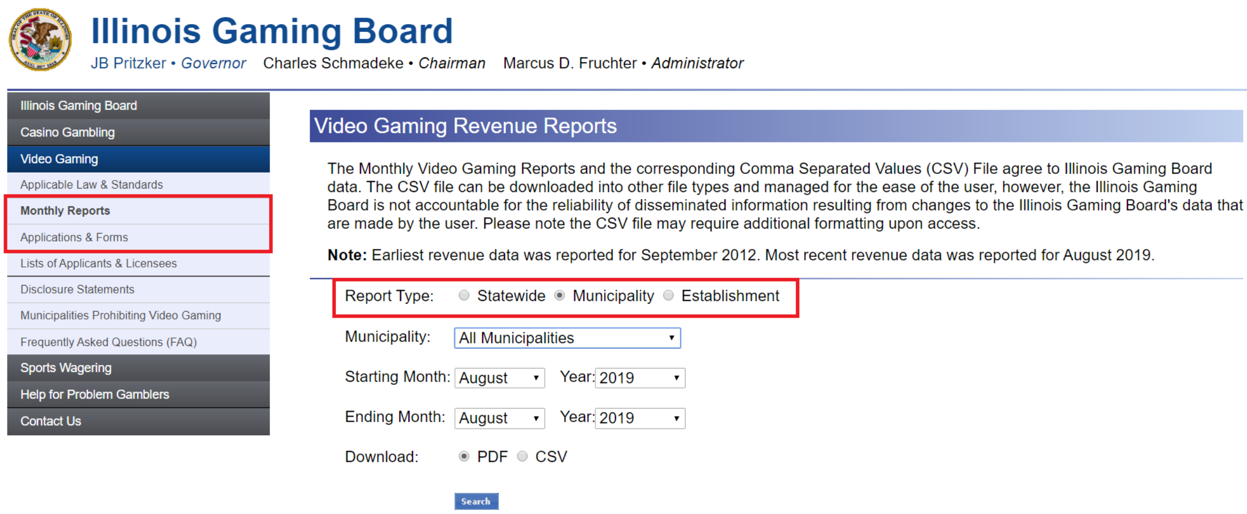


Figure 4.1: Illinois Gaming Board Website
Snapshot of the Illinois Gaming Board web page from which data on slot machines' profits you can be generated and downloaded. Users can choose whether to observe data on either the whole state, a specific municipality or a single establishment. The menu on the left has a link to the page which contains the forms that establishments need to submit in order to apply for slot machines.

profits, they can update their beliefs about their own unobservable. This in turn results in a posterior distribution of $\boldsymbol{\xi}$. Note that since the prior is common and the information is public, the posterior is also common.

Let j be an agent that adopted in some period τ_j before t. Since all other agents know the correct $\boldsymbol{\beta}$ and competition faced by j's neighbors, when they observe her profit at period t they can extract the following:

$$\xi_{j,t} \equiv \xi_j + \epsilon^g_{j,t} = \pi^g_{j,t} - \boldsymbol{x}_j \boldsymbol{\beta} + C_j(A_t) - z_t. \tag{4.2}$$

It is important to note that

$$\xi_{j,t} \sim N\left(\xi_j, \sigma^2_\epsilon\right).$$

This means that a noisy signal around the true unobservable can be extracted from every observation of per period profits of agents that adopted; recall that σ_ϵ is common knowledge.

Let $\omega_{j,t} = t - \tau_j - 1$ be the number of observations of j's profits that are available at time t, let P^ω_t denote the diagonal precision matrix of all observations[16] up to and including period t, and let $\bar{\xi}_{j,t}$ denote the average of the signals about the unobservable for agent j; this average is distributed around the truth, $\bar{\xi}_{j,t} \sim \left(\xi_j, \frac{\omega_{j,t}}{\sigma^2_\epsilon}\right)$.

[16]Nodes that have not yet adopted don't have information hence the precision of these nodes' "signals" is 0.

The updated distribution of $\boldsymbol{\xi}$'s for all of the agents in the municipality is a multivariate normal with posterior covariance matrix equal to

$$\Sigma_t = \left(\Sigma_0^{-1} + P_t^{\omega}\right)^{-1}, \tag{4.3}$$

where posterior means equal to

$$\boldsymbol{\mu}_t = \left(\Sigma_0^{-1} + P_t^{\omega}\right)^{-1} P_t^{\omega} \bar{\boldsymbol{\xi}}_t. \tag{4.4}$$

These define the best estimate of ξ for each agent in period t, $\mu_{i,t} \equiv E\left[\xi_i | \mathcal{I}_t\right]$, $\sigma_{i,t}^2 \equiv Var\left[\xi_i | \mathcal{I}_t\right]$. On the equilibrium path of play, agents use these posterior distributions as inputs to their utility functions, and these in turn define which agents adopt in each period.

4.4 Additional Parametric Assumptions

Additional parametric and functional form assumptions need to be made in order to estimate the model. In this part I impose additional structure on several objects of interest, including the competition function $C_j(\cdot)$, the prior distribution covariance Σ_0, the change in profits of the core business from adoption $\Delta\pi_{i,t}^{core}$, the utility function $u(\cdot)$, and the distribution of the adoption cost shocks. I also specify the utility function of the agents; the specification in this paper implies that agents are not forward looking and so don't take into account potential adoption of neighbors when adopting. In section 8 and appendix C I discuss current work that alters these specifications in order to allow agents with forward-looking best responses.

Competition function. The competition function takes the following functional form:

$$C_i(A_t) = c_s \sum_{j \in A_{i,t}^C, \theta_i = \theta_j} s_{ij,t}^C + c_d \sum_{j \in A_{i,t}^C, \theta_i \neq \theta_j} s_{ij,t}^C.$$

The term $s_{ij,t}^C = 1 - \frac{d_{ij}}{r^C}$ represents the proximity of i to j for some physical distance d_{ij} between the two nodes and some threshold r^C beyond which the competitive proximity is 0. This functional form takes into account that a closer neighbors steal more business from each-other. This property is similar to how competition is modeled in Seim (2006), the main difference is that she used a step function. The two separate sums in the function, multiplied by the different coefficients, allows for different levels of business stealing between adopters that are of the same type (c_s) as i and these that are of different type (c_d).

Prior. At time $t = 0$, agents' prior beliefs are a vector of unobservables $\boldsymbol{\xi}$ drawn randomly from a joint normal distribution:

$$\begin{pmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_n \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} & \cdots & \Sigma_{1n} \\ \Sigma_{21} & \Sigma_{22} & & \vdots \\ \vdots & & \ddots & \vdots \\ \Sigma_{n1} & \cdots & \cdots & \Sigma_{nn} \end{pmatrix}\right).$$

All diagonal terms have the same variance $\Sigma_{ii} = \sigma_\xi^2$, while the off-diagonal terms represent the spatial covariances between the ξ's, where

$$\Sigma_{ij} = \kappa\left(d_{ij}|\rho, r^\ell\right)\sqrt{\Sigma_{ii}\Sigma_{jj}}.$$

The function $\kappa(\cdot)$ is a kernel function that defines the correlation between the unobservables of two location based on their distance:

$$\kappa\left(d_{ij}|\rho, r^\ell\right) = \rho\left(\left(1 - \frac{d_{ij}}{r^\ell}\right)_+\right)^2,$$

where the correlation between nodes for which $d_{ij} \geq r^\ell$ is equal to zero.[17] It is important to note that the parameters in κ and the variance σ_ξ^2 are not necessarily the correct ones; they are merely the beliefs that the agents have about the level of correlation.

The parameter $\rho \in [0, 1]$ represents the maximum correlation with a distance-zero neighbor. In some settings, such as Covert (2015); Hodgson (2018) that analyzes distribution of oil in the ground, it is reasonable to assume perfect correlation when distance is 0, $\kappa(0) = 1$. This would translate to assuming that $\rho = 1$. Since in this paper there is an inherent heterogeneity in establishments' outcomes I allow ρ to be flexible. Therefore, two businesses that are in the same location might have different levels of success when adopting slot machines.

Core profit difference. The value $\Delta\pi_{i,t}^{core}$ represents the change in profits that establishments would see in their "core" business if they choose to adopt (e.g., in sales of food or alcohol). This term also takes into account the fixed cost due to alteration of their business layout. Changes in profits would be determined by the extent to which slot machines complement or substitute other products. Slot machines mechanically substitute part of the business since they take up space in the bar that had an alternative use, they could further

[17] This kernel function guarantees positive definiteness of any covariance matrix (see Williams and Rasmussen (2006)).

influence the number and type of customers which would affect the sales of food and alcohol and most likely depend on the type of the establishment. Additional considerations of the owners include their personal preferences regarding gambling, the potential change in atmosphere, or their uncertainty about these outcomes. I also take into account the fact that businesses that do not adopt for longer could on average differ not only on the information they have but also on the cost side. Therefore the assumed functional form is:

$$\Delta\pi_{i,t}^{core} = \sum_{\theta} \gamma_\theta \mathbb{1}_{\theta=\theta_i} + \gamma_t t,$$

where γ_θ is the type fixed effect, $\mathbb{1}_{\theta=\theta_i}$ is an indicator function for the type, and γ_t is a time trend coefficient.

Utility function. I assume that agents may be risk averse when they face their adoption decision. In every period, agents have beliefs about their outcome based on the updating process described in subsection 4.3.

Agents are expected utility maximizers, with a CARA utility function of the form

$$u(x) = -e^{-\lambda x},\ \lambda \geq 0.$$

Since all the terms here represent differences from agents' outcomes if they did not adopt, the outside option is to get 0 with certainty. In addition, since in the case of adoption all agents' realized period profits are drawn from a normal distribution, a known result[18] is that if $x \sim N\left(\mu_x, \sigma_x^2\right)$, then

$$EU(x) = -e^{-\lambda\left(\mu_x - \frac{\lambda}{2}\sigma_x\right)},$$

therefore agents that maximize utility adopt if and only if $\mu_x - \frac{\lambda}{2}\sigma_x \geq 0$.

Period adoption shocks. $\eta_{i,t}$ is assumed to be an EV(1) logit shock with variance $\frac{\pi^2}{6\alpha}$ for some parameter α.

5 Estimation & Identification

To estimate the parameters of the model, I use a procedure which finds the parameters that match best the adoption behavior observed in the data. The estimation procedure has three

[18]See derivation for example Sargent (1987).

main steps: (i) estimation of the competition function, (ii) estimation of $\boldsymbol{\beta}$ that is nested in step (iii) that estimates of Σ_0, and the parameters of the utility function. I later use the estimates of $\boldsymbol{\beta}$ to estimate the parameters of Σ_{True}. This section starts with describing the variation and identifying assumptions, followed by the likelihood function maximization and the stages of the estimation.

5.1 Variation and Identification

The estimation procedure uses three types of variation in the data in order to identify the parameters:

- competition (business stealing) is measured using changes in nodes' profits with variation in the set of active neighbors over time;
- firms' baseline profit parameters ($\boldsymbol{\beta}$) are measured using changes in baseline firms' profits with variation in firms' characteristics and types;
- priors and adoption decision parameters are measured using the changes in likelihood of adoption given variation in information (neighbors profits) different establishments observe.

Using these sources of variation, there are three model features that are used for identification:

1. The period profit shocks are i.i.d. normally distributed: $\epsilon^g_{i,t} \sim N\left(0, \sigma^2_\epsilon\right)$.
2. The period adoption cost shocks, $\eta_{i,t}$, which are i.i.d. logit shocks.
3. The expected ex-ante values of all of the unobservables are zero: $E[\xi_i] = 0$ for all i.

While the first two identifying assumptions are standard, the third is used to ensure that knowing the agents' beliefs in the period of adoption is sufficient to correct for bias due to selection based on unobserved beliefs. This is an initial condition that ensures that the expected value of the unobservable of each of the first adopters was 0.

5.2 Likelihood Function and Maximization

Under the assumptions given in section 4, agent i adopts in period t if and only if

$$\tilde{v}_{i,t} \equiv \underbrace{(\mu_{i,t} + \boldsymbol{x}_j\boldsymbol{\beta} - C_j(A_t) + z_t)}_{E[\pi^g_{i,t}]} - \frac{\lambda}{2}\sigma_{i,t} + \underbrace{\gamma_{\theta_i} + \gamma_t t}_{\pi^{core}_{i,t}} + \eta_{i,t} \geq 0.$$

While $\eta_{i,t}$ is observed by the agents, it is not observed by the econometrician, which yields an implied probability of adoption given by

$$p_{i,t} = \frac{\exp\left(v_{i,t}\right)}{1+\exp\left(v_{i,t}\right)},$$

where $v_{i,t} = \tilde{v}_{i,t} - \eta_{i,t}$. Hence the likelihood of observing adoption based on the parameters for municipality m is

$$\mathscr{L}_m\left(\Psi\right) = \prod_{i\in N_m, t\leq\tau_i}\left[\mathbb{1}_{a_{i,t}=1}p_{i,t} + \mathbb{1}_{a_{i,t}=0}\left(1-p_{i,t}\right)\right],$$

where $\Psi \equiv \{\boldsymbol{\beta}, c_s, c_d, \sigma_\xi, \sigma_\epsilon, \rho, r^\ell, \boldsymbol{\gamma}, \lambda\}$ is the set of parameters that enter the estimation. The estimated parameters would be

$$\Psi^* \quad = \quad \arg\max_{\Psi}\prod_{m\in M}\mathscr{L}_m\left(\Psi\right) \tag{5.1}$$

5.3 Step 1: Competition

The model assumes that the only action that an agent takes is deciding whether to adopt in every period in which they are not yet in the market. After adoption, agents stay idle and see a stream of profits. I use this assumption and the variation of a node's profits over time to measure the effect of business stealing.

Since both ξ_i and $\boldsymbol{x}_i$ are fixed over time, in the first step of the estimation I can recover the coefficients for the same type, c_s, and for other types, c_d, by running a simple linear regression,

$$\pi^g_{i,t} = \underbrace{f_i}_{\xi_i+\boldsymbol{x}_i\boldsymbol{\beta}} + c_s\sum_{j\in A^C_{i,t},\theta_i=\theta_j} s^C_{ij,t} + c_d\sum_{j\in A^C_{i,t},\theta_i\neq\theta_j} s^C_{ij,t} + z_t + \epsilon^g_{i,t},$$

where f_i and z_t represent establishment and period fixed effects.

The estimates of the establishment fixed effects, f_i, play a crucial role in the rest of the estimation—they are the baseline establishment profits, net of competition, and used in subsection 5.4 to estimate $\boldsymbol{\beta}$. In addition, the standard deviation of $\epsilon^g_{i,t}$ will be used as the estimate of σ_ϵ. I assume that the standard deviation, σ_ϵ, of the period i.i.d. profit shocks, $\epsilon^g_{i,t}$, is known to all agents and is used as the observations' signal variance.

Estimating r^C. In subsection 4.4 I introduced the parameter r^C that defines the maximal distance: the distance beyond which competition is set to 0. This radius parameter is estimated offline and is taken as exogenous in most of the estimation process. In order to estimate this parameter, I count the number of adopters in bands ("donuts") of some radius around an agent that has already adopted. I then regress the profitability of the agents on these counts, including time and node fixed effects, in a fashion similar to the regression described earlier in this section.

I choose the radius r^C as the value after which I observe that the correlation decays to 0, as described in section 6.

5.4 Step 2: Estimating $\boldsymbol{\beta}$

5.4.1 Why not a simple regression? - Accounting for selection

The next step of the estimation recovers the coefficients in $\boldsymbol{\beta}$ by projecting f_i on the characteristics space $\boldsymbol{x}_i$ and type fixed effects[19]. The problem: since agents self select to adopt, a linear regression of f_i on characteristics generally result in biased parameters. Namely, from the model we know that

$$f_i = \boldsymbol{x}_i \boldsymbol{\beta} + \xi_i, \tag{5.2}$$

and adoption depends on characteristics and observations of others' profits. Therefore, in general, $E\left[\xi_i | \boldsymbol{x}_i, i\text{'s adoption}\right] \neq 0$, and linear regression would lead a bias in $\boldsymbol{\beta}$. We solve this issue by using the agents' beliefs at adoption as an estimator for ξ_i.

The unobservable of each agent that adopted can be written in the following way $\xi_i = \mu_{i,\tau_i} + \epsilon_i$ where μ_{i,τ_i} is i's expectations of her unobservable in the period she decided to adopt, τ_i; ϵ_i is i's prediction error. Her error has two components, the first is a result of her misspecified prior ($\Sigma_0 \neq \Sigma_{\text{True}}$), denoted ϵ_i^p , and the second is due to the random draw of ξ_i, denoted ϵ_i^ξ. Therefore we can rewrite 5.2:

$$f_i = \boldsymbol{x}_i \boldsymbol{\beta} + \mu_{i,\tau_i} + \epsilon_i^p + \epsilon_i^\xi.$$

As mentioned in subsection 4.1, I assume that all agents in a municipality have the same information, i.e. there is no private information. This means that i updates her beliefs about her own profits base only the observed profits of others. Additionally, the real ξ_i assumed to be independent of $\boldsymbol{x}_i$, and therefore $E\left[\epsilon_i^\xi | \boldsymbol{x}_i, \mathcal{I}_{\tau_i}\right] = 0$.

Lemma 1. *Under the model assumptions, the mean of ϵ_i^p conditional on the observations at adoption is zero:*

$$E\left[\epsilon_i^p | \mathcal{I}_{\tau_i}\right] = 0. \tag{5.3}$$

[19] Type fixed effects account for ξ's being drawn from a distribution with mean different than 0.

Proof. See appendix D. □

This lemma implies that even though agents have an incorrect prior, the expectation their prediction error is zero given their observations.

Therefore, $E\left[\epsilon_i|\boldsymbol{x}_i, \mathcal{I}_{\tau_i}\right] = 0$, if μ_{i,τ_i} is known, $\boldsymbol{\beta}$ can be estimated with the following regression:

$$F_i - \mu_{i,\tau_i} = \boldsymbol{x}_i\boldsymbol{\beta} + \epsilon_i. \tag{5.4}$$

5.4.2 Estimation Procedure

Although all agents in the market know the values of the coefficients $\boldsymbol{\beta}$, this parameter vector is unknown to the econometrician, and therefore pose a difficulty in determining the values of μ_{i,τ_i} for all adopters. Without knowing μ_{i,τ_i} for all i that have ever adopted, it is impossible to determine the values of $\boldsymbol{\beta}$, while without knowing $\boldsymbol{\beta}$ it is impossible to determine μ_{i,τ_i} of all i. Therefore, what is needed is a method that estimates the two jointly and ensures that they are internally consistent. The equilibrium solution holds for a given Σ_0. In subsection 5.5 I search over the parameters that define Σ_0.

A candidate solution has two components: a vector $\boldsymbol{\mu_\tau}^c$ of length equal to the number of agents ever adopted, representing the expectations of the unobservable each establishment had at adoption, and a vector of parameters$\boldsymbol{\beta}^c$. At equilibrium:

1. The beliefs at adoption, $\boldsymbol{\mu}_\tau^c$ yield $\boldsymbol{\beta}^*$ (using the regression (5.4)).

2. The parameters $\boldsymbol{\beta}^c$ can be used to calculate the realized beliefs at adoption, $\boldsymbol{\mu}_\tau^*$, using the updating process in subsection 4.3 (given a prior Σ_0).

3. In equilibrium $\boldsymbol{\mu}_\tau^c = \boldsymbol{\mu}_\tau^*$, $\boldsymbol{\beta}^c = \boldsymbol{\beta}^*$.

The parameters $\boldsymbol{\beta}^*$ are the ones for which the parameters are consistent with agents' beliefs.

In order to find such equilibrium, an iterative algorithm is used. Starting with candidate beliefs, $\boldsymbol{\mu}_\tau^c$, the candidate parameters, $\boldsymbol{\beta}^c$, are calculated. Then, $\boldsymbol{\beta}^c$ is used to calculate the realized beliefs at adoption, denoted $\boldsymbol{\mu}_\tau^{c+1}$. Therefore, in equilibrium condition 3, the part about $\boldsymbol{\beta}^c$ holds by definition. If $\boldsymbol{\mu}_\tau^c = \boldsymbol{\mu}_\tau^{c+1}$ then candidate is an equilibrium solution. If not, $\boldsymbol{\mu}_\tau^{c+1}$ is the new candidate. I repeat this iterative process until convergence.[20] A detailed description of the algorithm is described in appendix E.1; appendix E.2 uses simulated data to show how this algorithm recovers the correct parameters when the data is selected.

[20] Computational convergence, below some specified tolerance.

5.5 Matching Adoption Decision

So far I showed how I estimated some of the model parameters. First, the competition function $C_i(A_t)$ was estimated, using only post-adoption data on the nodes' profits. Then $\boldsymbol{\beta}$ was estimated with a method that accounts for the selection problem that arises due to the fact that agents' adoption is correlated with their expectations of profits. This second step of the estimation used a fixed guess of the parameters of the kernel correlation function ρ, r^ℓ, and the standard deviation parameter σ_ξ. The first two steps were agnostic as to the reasons for agents' adoption, as long as the adoption decision itself did not provide information about other nodes' expected profits—and that, on average, the expected profits were correct.

The third and last step of the estimation provides some functional form assumptions on the value of adoption. I then match these calculated values of adoption and match them with the observed adoption in the data, pinning down parameters ρ, r^ℓ, and σ_ξ, and the other parameters in the value function.

5.5.1 MLE: Matching on adoption

This part directly follows the functional form and the maximization from subsection 5.2. As a reminder, the probability that agent i adopts at period t is

$$p_{i,t} = \frac{\exp\left(v_{i,t}\right)}{1 + \exp\left(v_{i,t}\right)},$$

with

$$v_{i,t} = \left(\mu_{i,t} + \boldsymbol{x}_j \boldsymbol{\beta} - C_j(A_t) + F_t\right) - \frac{\lambda}{2}\sigma_{i,t} + \gamma_{\theta_i} + \gamma_t t. \tag{5.5}$$

I maximize the likelihood function to obtain

$$\left(\rho^*, r^{\ell *}, \lambda, \boldsymbol{\gamma}^*\right) = \arg\max_{\rho, r^\ell, \gamma} \prod_{i \in N_m, t \leq \tau_i} \left[\mathbb{1}_{a_{i,t}=1} p_{i,t} + \mathbb{1}_{a_{i,t}=0}\left(1 - p_{i,t}\right)\right],$$

by searching over the values of ρ, r^ℓ and σ_ξ nonlinearly. For every guess, I use the values of $\mu_{i,t}$ and $\sigma_{i,t}$ estimated in the previous steps as inputs, running a simple logit regression to recover the parameters γ_π, λ, $\boldsymbol{\gamma}_\theta$, and γ_t.

5.6 Estimating true parameters

When recovering the true covariance parameters between agents' unobservables ($\sigma_{\xi \text{True}}$, ρ_{True} and r^ℓ_{True}) only estimates of a subset of the populations' unobservables are available, which may lead to a selection problem. To solve this issue I rely on the fact that when some agent

i adopts in period t, using the true prior parameters and current data would result in the best estimate of the joint distribution of i's and previous adopters' unobservables (given the information available at period t, regardless of the reason that i chose to adopt).

To estimate the true covariance parameters I first recover the $\xi_{i,t}$s of all adopters using the parameter estimates from this section (see equation 4.2). The true correlation parameters maximize the likelihood of the realized unobservables taking into account the available information at each agent's adoption period. Let i be an agent that adopts in period τ_i, for a guess of true parameters I calculate the Bayesian posterior covariance of i's unobservable with unobservables of all adopter up to the period, A_{τ_i} (given observations up to the period). This posterior for each i is used to construct a covariance matrix of the joint distribution of unobservables of all adopters at the period in which each adopted. I construct such covariance matrix for each municipality and use the matrices to calculate the likelihood of observing the realized unobservables. I choose $\sigma_{\xi \text{True}}$, ρ_{True} and r^{ℓ}_{True} that maximize this likelihood.

6 Estimation Results

In this section I describe the parameter estimates and the economic interpretation of these parameters. I start by describing them in the same order in which they were estimated in section 5. In subsection 6.2 I analyze the implications of the model regarding the way agents learn, and I describe additional estimates of the true spatial covariance parameters described in subsection 4.1. This section ends with analysis of the model fit. Note that the confidence intervals and standard errors were conservatively calculated using bootstrap at the municipality level (except for the competition estimates that are presented in table 4, which were produced in a simple linear regression). Note that the parameters in steps 1 and 2 of the estimation (described in subsections 5.3 and 5.4) are estimated using all of the available data. The final step of the estimation (subsection 5.5) uses a subset of the establishments. To prevent biases[21] in adoption decisions, only establishments in municipalities that allowed adoption from the inception of the market are included. In appendix B, I show that the results are robust to using all the nodes.

[21] There is a concern that nodes in municipalities that initially prohibited adoption would have greater information in the period of adoption.

6.1 Parameter Estimates

Competition. I start with estimation of r^C. To determine its value, I regress the gambling period profits of establishments that have already adopted on the numbers of their neighbors that have already adopted. The area around a node is divided into bands, and the node's neighbors are divided into groups based on these bands. Table 3 shows the results of this regression under two different band systems, and with different combinations of fixed effects. While all the specifications have node fixed effects, some specifications also have period fixed effects and county time trends. I can see that business stealing (i.e., competition) disappears after about 0.7 miles, and therefore r^C would take this value.

Given that value of r^C, the values of $c_s = 229.86$ and $c_d = 77.03$ are reported in column (1) of table 4. These values represent the business stealing that an additional next-door neighbor imposes when it decide to adopt. The former value represents the business stealing imposed by a node of the same type, while the latter value represents the business stealing that occurs when an agent of a different type adopts. This means that if i and j are neighbors of the same type, j's adoption has three times the effect than if the type of j is different that of i. Columns (2) and (3) show that the values of these coefficients are robust to the different fixed effects that are included in the estimation.

β and selection correction. The estimates of β are reported in table 5. The "with correction" column represents the estimates of β from the model, controlling for the beliefs of adopting agents at the period of adoption and using the procedure described in subsection 5.4. In general, establishments should ex ante expect a few thousands of dollars in monthly profits from installation of slot machines, while, bar-like establishments (bars, bar and grills, and veteran organizations) generally have higher profits. However, missing information, such as no reported price level or a lack of a reported rating of the business, significantly lowers the expected profits from gambling. The result is that there are situations in which businesses should expect low monthly profits (for example, a Mexican restaurant that has no reported price level and has never been rated could expect profits that are lower than \$1,000).

The "without correction" column represents the estimates of $\boldsymbol{\beta}$ that would have been produced without accounting for selection at adoption. The selection problem is due to the fact that the profits observed in the data are only those of agents that chose to adopt based on their beliefs. The data in that column are reported for purpose of comparison to the true parameters, and the rightmost column in the table is the percentage change in the parameters as a result of the selection correction procedure. I can see that baseline profits of all types decreased, with the maximum decrease reaching more than 8%, while the changes in other parameters went in both directions and reached up to 13.6%.

	$\pi^g_{i,t}$					
	(1)	(2)	(3)	(4)	(5)	(6)
$\|A_i\|$ 0–0.1 miles	-154.915^{***}	-117.950^{***}	-137.968^{***}			
	(49.634)	(44.739)	(45.996)			
$\|A_i\|$ 0.1–0.2 miles	-92.227^{*}	-48.752	-60.746			
	(51.351)	(46.859)	(47.565)			
$\|A_i\|$ 0.2–0.3 miles	-81.233	-83.176	-96.993			
	(64.402)	(59.505)	(60.814)			
$\|A_i\|$ 0.3–0.4 miles	73.363	78.222	66.208			
	(68.219)	(66.748)	(66.877)			
$\|A_i\|$ 0.4–0.5 miles	-45.320	-77.858	-84.894			
	(57.838)	(54.628)	(55.398)			
$\|A_i\|$ 0–0.5 miles				-68.884^{***}	-55.769^{***}	-68.489^{***}
				(22.806)	(20.544)	(21.440)
$\|A_i\|$ 0.5–0.6 miles	-51.208	-51.136	-64.845	-44.593	-48.302	-61.095
	(56.958)	(53.072)	(54.298)	(56.927)	(53.069)	(54.217)
$\|A_i\|$ 0.6–0.7 miles	-36.105	-72.108	-83.596^{*}	-36.544	-72.115	-83.297^{*}
	(49.759)	(47.302)	(47.910)	(49.411)	(47.177)	(47.743)
$\|A_i\|$ 0.7–0.8 miles	14.337	-17.421	-32.109			
	(57.407)	(54.363)	(55.190)			
$\|A_i\|$ 0.7–0.75 miles				49.904	4.070	-14.133
				(85.913)	(83.074)	(83.827)
$\|A_i\|$ 0.75–0.8 miles				-16.818	-34.378	-44.420
				(82.056)	(76.401)	(77.417)
$\|A_i\|$ 0.8–0.9 miles	63.559	12.424	2.324	68.564	15.846	6.284
	(49.577)	(47.068)	(47.854)	(49.743)	(47.238)	(47.995)
$\|A_i\|$ 0.9–1 miles	48.020	36.187	18.177	54.653	37.469	20.198
	(57.817)	(50.227)	(51.215)	(57.645)	(50.075)	(50.965)
Period Fixed Effects	Yes	No	Yes	Yes	No	Yes
Node Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
County Time Trend	No	Yes	Yes	No	Yes	Yes
Observations	153,623	153,623	153,623	153,623	153,623	153,623
R^2	0.826	0.827	0.833	0.826	0.827	0.833
Adjusted R^2	0.821	0.822	0.828	0.821	0.822	0.828
Residual Std. Error	1,559.823	1,555.256	1,527.369	1,560.231	1,555.555	1,527.696
Note:						*p<0.1; **p<0.05; ***p<0.01

Table 3: Period Profits on Number of Adopters in Bands
Regression of profits of existing adopters on the number of their neighbors that adopted in different radius bands.

		$\pi^g_{i,t}$	
	(1)	(2)	(3)
Competition: Other Types	-77.034^{**}	-55.508^{*}	-77.125^{**}
	(36.600)	(33.252)	(35.559)
Competition: Same Type	-229.856^{***}	-228.268^{***}	-255.202^{***}
	(76.850)	(70.381)	(71.789)
Period Fixed Effects	Yes	No	Yes
Node Fixed Effects	Yes	Yes	Yes
County Time Trend	No	Yes	Yes
Observations	153,623	153,623	153,623
R^2	0.826	0.827	0.833
Adjusted R^2	0.821	0.822	0.828
Residual Std. Error	1,560.089	1,555.419	1,527.568
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 4: Competition: Period profits on total neighbors proximity
Regression of profits of existing adopters on total proximity of neighboring adopters as defined in subsection 5.3.

The correction procedure shifts the parameters in both directions, although most of the parameters decrease there are some parameter values that go up. In order to understand the total impact of the selection correction procedure we need to understand its total effect on the ex-ante expected profits of the establishments ($x_i\beta$). In figure 6.1 I plot the distributions of the ex-ante expected profits of all establishments, with and without correction. The mean value of $x_i\beta$ goes down by about 5%, and it is apparent that the the values of $x_i\beta$ shift to the left. Running a Kolmogorov Smirnov test confirms that the distributions are significantly different.

Belief parameters. The estimates of the updating parameters are presented in table 6. The correlation estimates ρ_0 and r_0^ℓ indicate that agents believe that the maximum correlation with a zero-distance neighbor is about 0.2, and that the correlation with their neighbors decays after a little more than 2 miles, about three times the radius of the competition they face. Even though all of the model's parameters are identified, in practice there is a problem with the estimation of $\sigma_{\xi 0}$. The values of $\sigma_{\xi 0}$ are too large to be computationally estimable. When $\sigma_{\xi 0}$ is large, changing its value has negligible effect on the rest of the parameters. Therefore, in the estimation, I calibrate $\sigma_{\xi 0}$ to a very large number. At the limit the values

	With Correction	Without Correction	Δ
No Price Dummy	-1926.676***	-1850.816***	-4.1%
	(301.833)	(271.89)	
Price Level	-909.399***	-837.577***	-8.6%
	(206.904)	(186.30)	
No Rating Dummy	-1397.593**	-1588.13**	+12%
	(654.334)	(566.21)	
Rating	-210.012	-233.801**	-9.9%
	(131.837)	(118.6)	
Population/$\|N\|$	0.025***	0.022***	+13.6%
	(0.0086)	(0.003)	
American Restaurant	6621.499***	6779.421***	-2.23%
	(903.338)	(667.12)	
Bar	6641.801***	6814.336***	-3.25%
	(905.276)	(577.1)	
Bar and Grill	7035.818***	7184.183***	-2.1%
	(917.15)	(600.9)	
Italian Restaurant	5792.791***	6184.658***	-6.6%
	(1392)	(759.446)	
Mexican Restaurant	4113.553***	4483.172***	-8.3%
	(1073.3)	(690.72)	
Missing Type	6431.563***	6702.122***	-4%
	(989.97)	(647.273)	
Pizza Restaurant	5935.446***	6170.37***	-3.9%
	(1075.7)	(653.29)	
Restaurant	5402.674***	5552.862***	-2.8%
	(962.256)	(632.89)	
Veterans Organization	6686.412***	7018.335***	-4.8%
	(1094.04)	(657.9)	
Small Types	6424.562***	6664.113***	-3.7%
	(968.8)	(596.38)	
Observations	4286	4286	
R2	0.065	0.058	
Adjusted R2	0.062	0.055	
Residual Std. Error	3134.452	3173.386	
F Statistic	21.065***	18.739***	
Note:		*p<0.1; **p<0.05; ***p<0.01	

Table 5: β without and with selection correction
Estimates of the vector of parameters $\boldsymbol{\beta}$. Column (1) has the model estimates. Column (2) has the estimates without selection correction. Column (3) has the percentage difference between the two. Note: standard errors in (2) do not account for error from the first stage of the estimation therefore smaller.

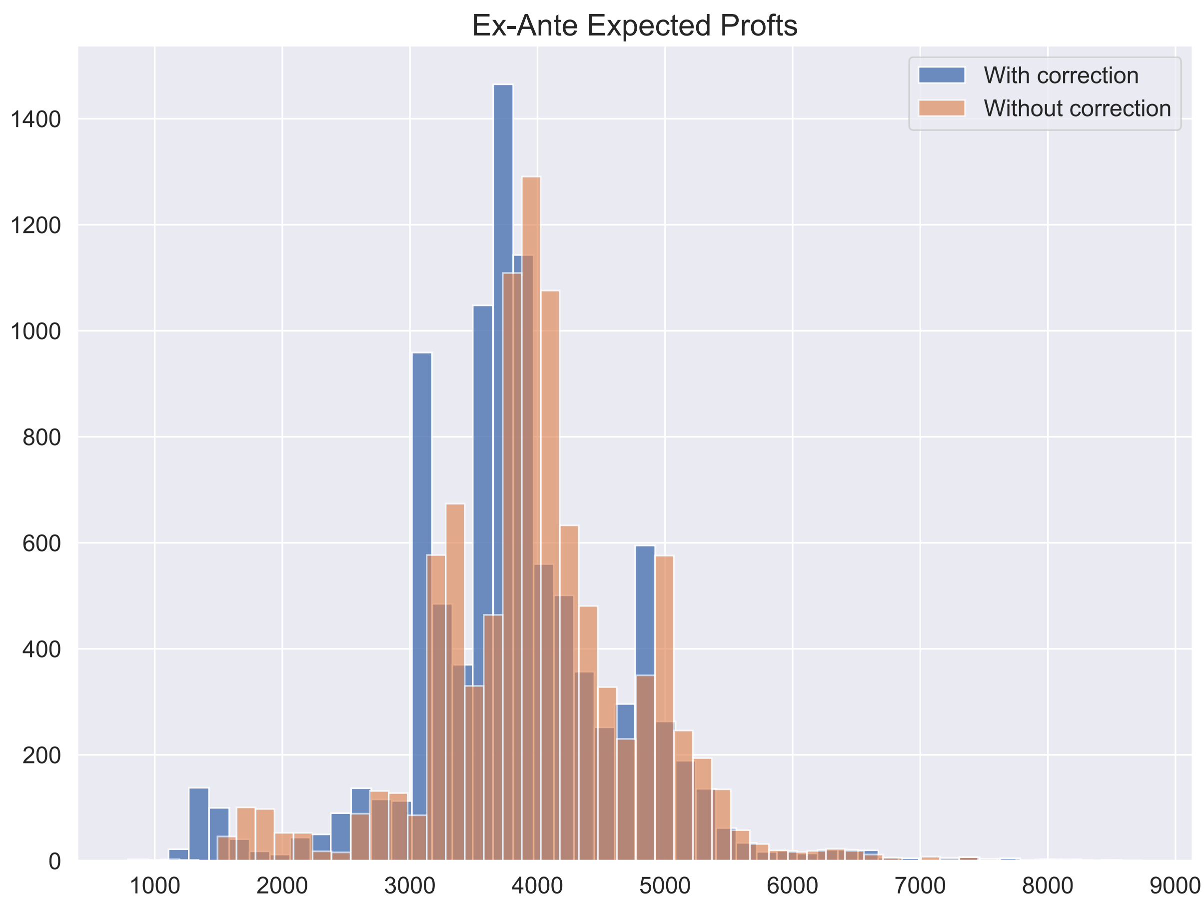


Figure 6.1: Distribution of $x_i\beta$
Orange: distribution of ex-ante expected profits **without** correcting for selection. Blue: distribution of ex-ante expected profits correcting for selection.

of $\sigma_{\xi 0}$ and λ are not separately identified, though the product of the two is. This part is formally stated and proved in appendix F. A very high $\sigma_{\xi 0}$ suggests that agents have a diffuse prior and therefore put very low or no weight on it; they behave as if they were frequentists. Therefore, the moment they observe a signal about another node's profits, they immediately update based only on the correlation they believe they have with that node.

The estimated parameters are presented in table 7. The value of $-\frac{\lambda}{2}\sigma_{\xi 0}$ represents the initial influence that the uncertainty and risk aversion have on the value of the certainty equivalent of choosing to adopt slot machines. The other parameters are components of the change in core profits that firms would see if they decided to adopt, $\Delta\pi_{i,t}^{\text{Core}}$. One interpretation of the value of γ_t is that, on average, adoption is more costly and the slot machines are more of a substitute for establishments that choose not to adopt for more periods. It turns out that, on average, slot machines complement the core business for bars that adopted up to period 52 providing positive core value. On the other hand, a restaurant that adopts at around period 20 or later expects on average a loss to their core business.

6.2 Agents' Beliefs at Adoption and True Spatial Correlation

In the model I made the uncommon assumption that agents' perceptions about the spatial correlation may be different from the real spatial correlation of their profits. This kind of learning behavior could lead to a prediction that is less accurate than it could be, or to a prediction that is potentially incorrect. In this subsection I first find how well the agents predict. I then present the parameters of the true spatial correlation in the data and see how the agents' beliefs compare to the truth.

Agents' unobservable predictions. Using the estimates of $\boldsymbol{\beta}$ and the other parameters that were recovered in the estimation process, I also recover estimates of the true unobservables of all agents that have ever adopted. Given the estimates of the model, I also know the beliefs that each agent had at the period in which they adopted, μ_{i,τ_i}. In table 8 and figure 6.2 I see that the agents beliefs about their unobservable (albeit noisy) are predictive of their true profits at adoption: An agent that believes that its profits will be higher by a dollar have, in expectation, profits that are higher by about 85 cents. Moreover, the hypothesis that the coefficient of μ_{i,τ_i} is different from 1 cannot be rejected (although only barely). Therefore, we cannot reject that, at adoption, agents' expectations about their profits are correct on average.

Spatial covarinace: beliefs vs. truth. Using the estimates of the unobservables that were recovered in the estimation process, I can estimate the true spatial correlation param-

Parameter	Estimate	Confidence Interval
ρ_0	0.198	$[0.0755, 0.99]$
r_0^ℓ	2.09	$[1.4, 18.12]$
σ_ϵ	1537.9	$[1477, 1596.5]$

Table 6: Beliefs Parameters
Estimates and confidence intervals parameters that are the agents' priors and which they use when updating their beliefs about their unobservable.

Parameter	Value	Confidence Interval
$-\frac{\lambda}{2}\sigma_{\xi 0}$	-33435.4	$[-476228, -3357]$
γ_t	-310.7	$[-1947, -234]$
γ_{bar}	16101.6	$[-91240, 129565]$
$\gamma_{\text{bar\&grill}}$	15636	$[-89516, 128992]$
γ_{Veteran}	15044.1	$[-99454, 128732]$
$\gamma_{\text{MissingType}}$	11139.1	$[-133919, 125997]$
γ_{American}	10900.3	$[-130355, 125421]$
$\gamma_{\text{restaurant}}$	10627.2	$[-140070, 125611]$
$\gamma_{\text{OtherTypes}}$	9189.32	$[-143517, 124171]$
$\gamma_{\text{PizzaPlace}}$	7951.1	$[-151335, 123014]$
γ_{Italian}	7085.1	$[-161208, 122060]$
γ_{Mexican}	5699.1	$[-178417, 122127]$

Table 7: Adoption Decision Parameters
Estimates and confidence intervals of the parameters that represent initial risk ($-\frac{\lambda}{2}\sigma_{\xi 0}$) and that define the change in core profits ($\Delta\pi_{i,t}^{\text{Core}}$).

Figure 6.2: Realized Unobservable and Beliefs at Adoption
Blue: the regression line and confidence interval of the regression in table 8. Green: grouping at $500 bins of establishment's beliefs at adoption, the averages of the best estimates (realization) of establishments' unobservables and confidence intervals are plotted. Orange: 45 degree line.

	Realized ξ	
μ_{τ_i}	0.858***	0.847***
	(0.077)	(0.08)
const		25.001
		(49.634)
Observations	4287	4286
R2	0.028	0.026
Adjusted R2	0.028	0.025
Residual Std. Error	3128.083	3128.355
F Statistic	125.064***	112.986***
Note:	*p<0.1; **p<0.05; ***p<0.01	

Table 8: Realized Unobservable and Beliefs at Adoption
Regression of the best estimate of adopters' unobservables on the beliefs those adopters had at the period in which they decided to adopt

eters that the ξ's are drawn from. The estimates of these parameters are $\sigma_{\xi True} = 3128.16$, $\rho_{True} = 0.155$, and $r^{\ell}_{True} = 13.5$. First, I notice that the true standard deviation of the distribution of the unobservables is about twice the signal standard deviation σ_ϵ, which implies that agents should weight their prior when observing signals and updating their beliefs. Moreover, agents believe their correlation with very close neighbors is higher than the actual correlation ($\rho_{True} = 0.155 < 0.198 = \rho_o$) therefore they overweight them. However, they ignore valuable information—they take into account neighbors in about a 2-mile radius, while the spatial correlation decays after about 13.5 miles. Figure 6.3 shows the difference between the beliefs agents have about the spatial correlation with neighbors of different distances (blue) and the actual spatial correlation with neighbors of different distances (orange). The smaller radius could be a result of different beliefs that the agents could have, though there are alternative scenarios which will result in the same pattern. For example, if not all agents observe profits through the website, and instead talk to their neighbors or just observe the changes in traffic in neighboring businesses, it will shrink the estimated radius parameters.

6.3 Adoption Behavior and Model Fit

The core question of this model and the pattern that this paper set out to explain is what caused agents to delay adoption of an existing product for many months and sometimes even a few years. The channel that I explore is that agents observe adoption by others and update their beliefs about both the profitability of the product and the riskiness of the adoption. Even if agents have some expectations about the profitability of the product, they might be averse to adoption because they are unsure of how this action would affect their business.

In fact, in the estimates and in the data, I an indication that a decrease in agents' uncertainty, given their expected profits, has positive impact on adoption. One way to see this result is in figure 6.4, where I analyze the set of bar-like establishments that waited at least fifteen periods before they eventually decided to adopt. These nodes had ample time and most likely multiple observations, when they decided to adopt. I show (in blue) that in the 15 periods prior adoption, agents initially saw an increase in their expected profits from adoption (changes in $\mu_{i,t}$). Around 10 periods before adoption this change in expected value halts and remains approximately constant until the period in which agents adopt. Even though the expected values do not change, the certainty equivalent ($\mu_{i,t} - \frac{\lambda}{2}\sigma_{i,t}$, depicted in green) continues to rise until the period of adoption, since over time agents observe more signals from more neighbors. In the 15 periods before adoption, establishment owners see an average increase in the value of adoption of around \$900. Analysis of the levels of profit

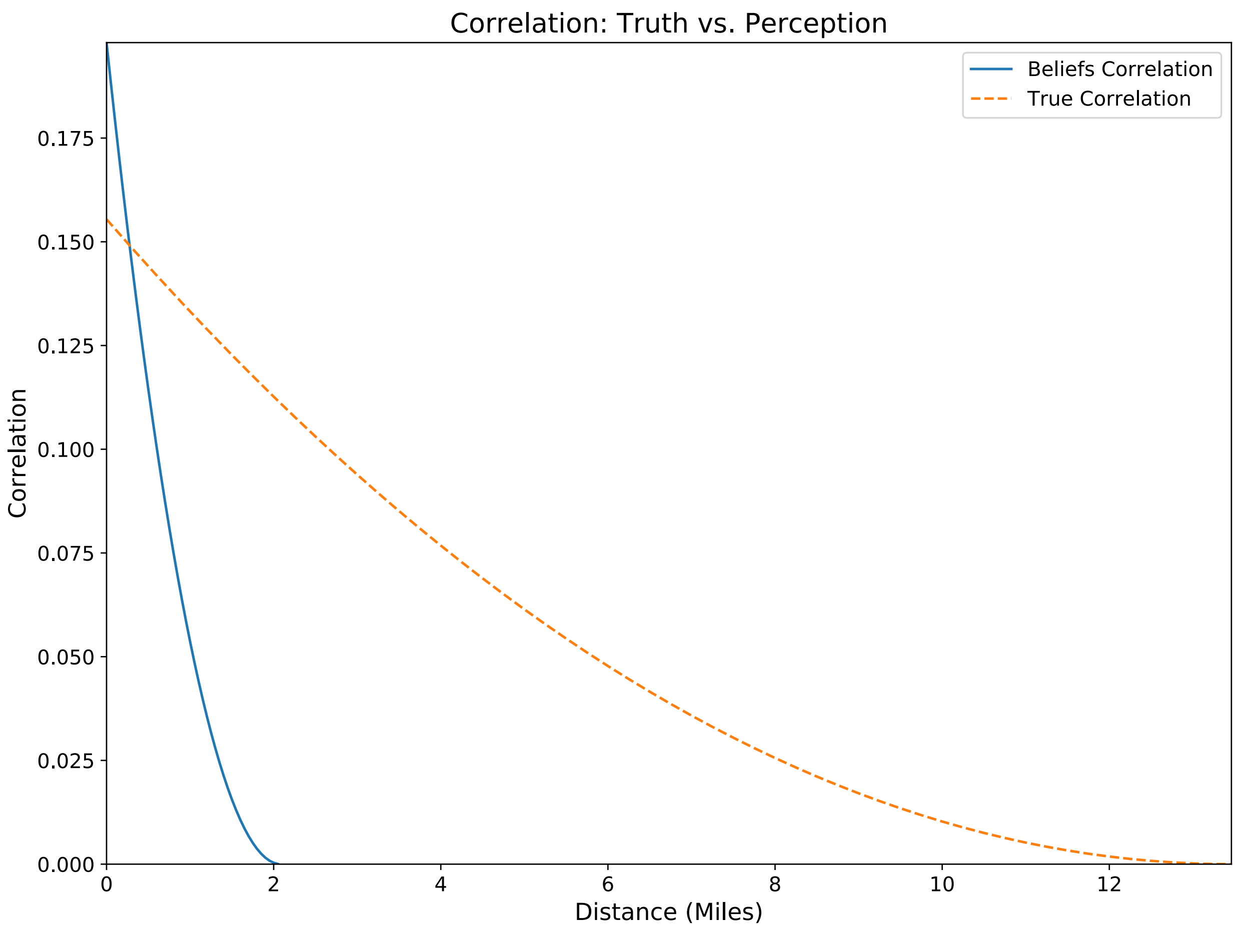


Figure 6.3: Spatial Correlation: Beliefs vs. Truth
The vertical axis represents how much spatial correlation there is between two nodes, with distance on the horizontal axis. The blue curve represents the correlation agents believe they have with their neighbors. The orange curve represents the true spatial correlation in the data.

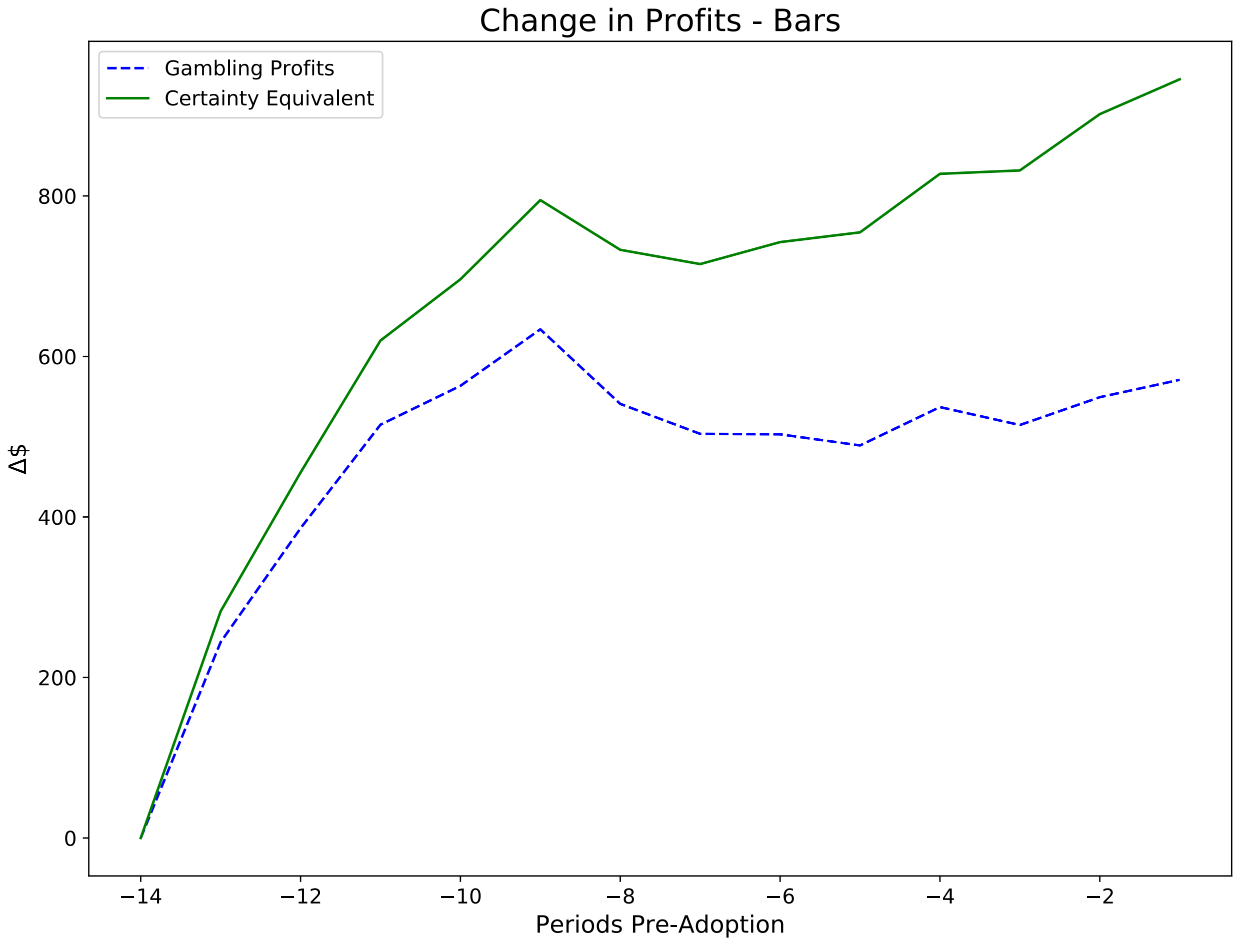
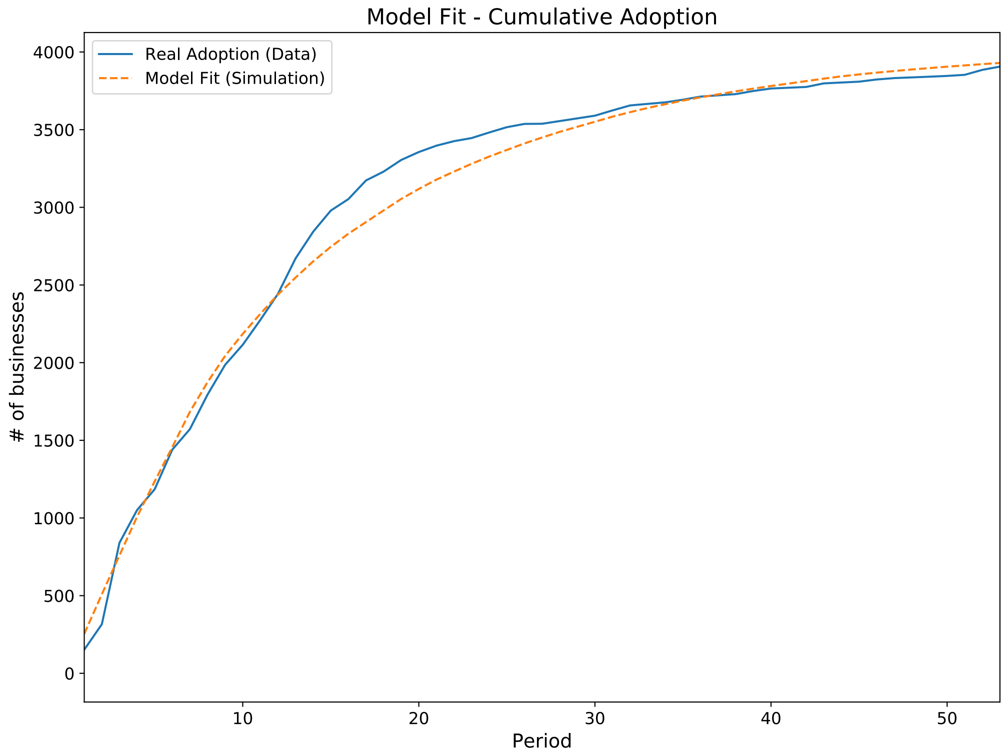


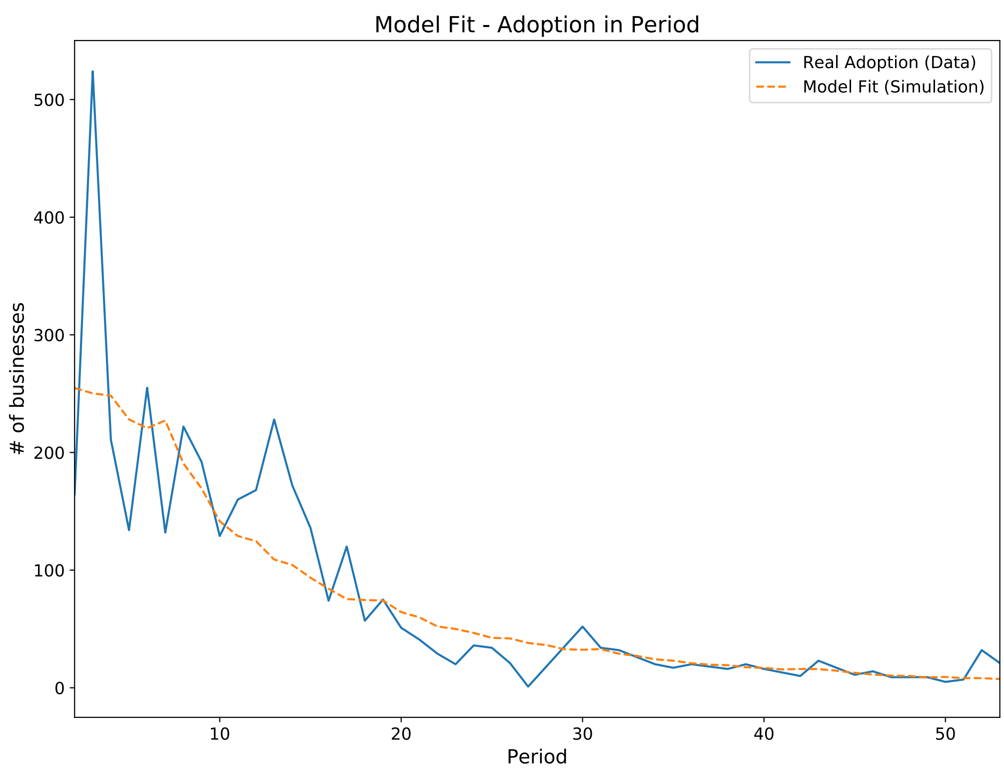
Figure 6.4: Pre-adoption Expectation and Certainty Equivalent.

and of restaurants and all other businesses, before adoption, is presented in appendix G.

Model fit. To check the model fit, I use the estimated true spatial correlation parameters to simulate unobservables for all agents. Then I forward simulate adoption, and therefore the implied observations all agents have and the competition that they face in every period. I repeated the simulation 100 times and average the adoption level in every period. As seen in figure 6.5, the fit of the model is good, the curve of predicted adoption follows the actual adoption pattern fairly closely. The maximal deviation in the number of cumulative predicted adopters compared to the actual number of adopters is 218, which is about 5.5% of the total number of adopters. In the final period available in the data set, the total deviation is only about 1.2%. In



(a)



(b)

Figure 6.5: Model Fit
In both panels the blue curve indicates the number of adopters over time in the data; The orange curve provides the simulated adoption pattern. Panel (a) is has cumulative adoption while (b) has in-period adoptions.

7 Counterfactual Analysis

In this section I examine two environments. In the first (no learning), agents have no information about their neighbors' profits and therefore cannot learn about their own profitability from their neighbors' profits. In the second environment (true correlations), I assume that agents have perfect knowledge of the true spatial correlation between everyone's unobservables, and therefore use the available information properly.[22] In these exercises I simulate the market's diffusion pattern, as well as how total market profits would change, and compare them to a baseline in which agents learn according to their estimated beliefs (model fit).

The no-learning exercise simulates an environment that is similar to most markets with non-public firms. In most cases private firms do not reveal their periodic profits, which creates an environment that is not conducive to learning. Using this exercise, I can test for how adoption would change if policy makers were to obligate firms to reveal their profits on new products. This is an extreme case in which firms have no information at all; in most cases firms would probably observe some noisy signal about the success of some product (e.g., changes in consumer traffic, price levels, or advertising). The quality of the signal would vary substantially and could be confounded by many factors and by the levels of success of other dimensions of the business which they observe.

The true-correlations exercise simulates an environment in which agents are more knowledgeable. In this exercise, I check how adoption patterns and total profits would change if agents had perfect information and knew exactly how to interpret it. The greater sophistication of agents in this environment could result from advertisement campaign or workshops that better inform them about the profitability levels in the area or how to interpret the information that is available to them. (This scenario would be of greater relevance if the assumption that all establishments are aware of everyone's profits is incorrect.) One way to interpret this exercise is as the effects information spillovers would have in markets with more sophisticated sellers, or for new products that sellers are more familiar with (for example, in this case, a new beer).

Theoretically, the direction of the change in adoption patterns under the two exercises is not ex ante clear. The effect of decreasing uncertainty from observing more and better information would generally make adoption more likely, though it would greatly depend on the observations themselves. Observing neighbors' profits that are disappointing would deter potential adopters.

The simulated diffusion patterns under the three simulations are presented in figure 7.1. I can see that providing more information to the agents leads to more and earlier adoption.

[22]Note that for technical reasons I maintain the assumption of infinite prior variance.

Compared to no learning, adoption under the model fit exercise increases by 4.8%, while the total profits in the market over that period increase by only 3.5%. In the true correlations exercise, adoption increases by another 3.1%, for an overall effect of 8%, and total profits increase by additional 3.6%, for an overall effect of 7.2%.

The results of the simulation suggest that making gambling profits public on IGB's website resulted in an yearly revenue increase in the order of tens of millions of dollars. For example, in 2016 about $38 million of net revenue can be attributed to learning, resulting in $11.5 million in taxes. While adding information increases both adoptions rates and profits, the two are not increasing at the same rate. Comparing the relative increase between the two levels of learning in the simulation, we see that the change in profits between true correlation and model fit are higher than the change between model fit and no learning, while the opposite is true for the change in adoption. This finding would suggest that additional information leads not only to more adoption but leads to more efficient adoption by more profitable establishments. If there is some fixed cost of adoption, this would increase the effects of information on efficiency by even more.

8 Extensions and Robustness in Progress

Dynamics - Forward looking agents. The agents in the current model are not taking into account how potential future adoption by firms which would affect their profits or the level of information available to them when they make their adoption decision in every period. It is reasonable to believe that when agents decide whether to adopt, they take into account the fact that others around them might also adopt. Future adoption of neighbors could affect their decision in a few ways. First, expecting adoption by neighbors would lead to lower expected profits and might deter adoption, though this expectation might also lead to early adoption if one thinks they would deter others. Secondly, there is an option value in waiting: If one is not sure whether adoption is worthwhile, they can free ride the information from neighbors. This effect might lead to delayed adoption.

In order to determine the magnitude of these dynamic effects, the beliefs of the agents about adoption by establishments around them have to be modeled. Modeling these beliefs poses additional problems. First, the state space is very large, since agents need to have beliefs about the probability of observing each combination of neighbors that adopted in every period; thus number of states for each agent grows exponentially with the number of neighbors. Secondly, agents could be strategic about their behavior. Finally, agents' beliefs affect their probability of adoption, which in turn affects the observations of others; therefore, agents' beliefs are a function of their behavior. Agents' beliefs and actions should

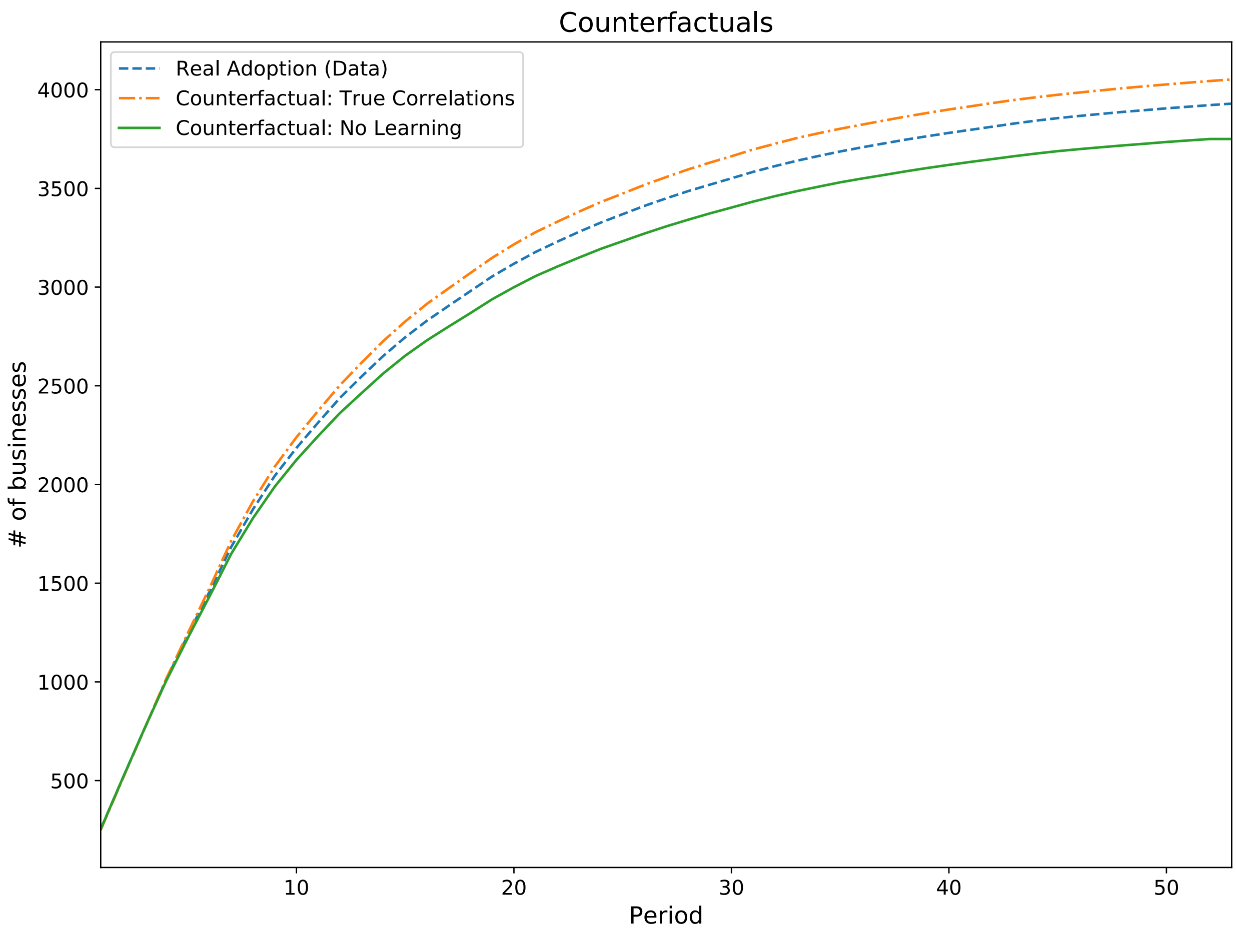


Figure 7.1: Counterfactual Estimation
In the figure are simulated adoption patterns over time. Blue: using the estimated beliefs of the agents (the model's fit). Green: if agents cannot learn from their neighbors. Orange: if agents have perfect knowledge of the correlation with their neighbors.

be consistent with each other.

In an ongoing extension I suggest a model that deals with those issues. To deal with the first and second issues, I reduce the state space and the dimensionality of the agents' beliefs, by implementing the method suggested by Krusell and Smith (1998) in a way similar to Lee (2013). To solve the final problem, I use the normality of the signals, and the fact that adoption decisions are based on a threshold rule of the agents, to determine the value of adoption in every period by backward induction and find parameters that imply that agents are consistent. Additional details are provided in appendix C.

Learning from Different Types. When estimating the model, the spatial competition, that is, the level of business stealing, is allowed to depend on the types of the establishments. Indeed, I found that the competition between establishments of the same type is fiercer. At the same time, learning differently from different types is restricted in the model, forcing agents to learn in the same way from all types.

In the preliminary analysis, I allowed for agents' potential profits to have different levels of correlation with the same and different types by adding a parameter that attenuates the learning from agents of different types. The result of the estimation was that agents learn equally from neighbors of the same type as from those different types (no attenuation). However, additional investigation should be made into these results to see how robust they are to the definitions and the number of types.

9 Conclusion

This paper provided evidence that the information firms gain from observing adoption and outcomes of others has a substantial effect on their decision to adopt products in a new market. I estimated a model which separates agents' perceptions about spatial correlation of outcomes from the true correlation. This property is especially relevant in markets of a new product where agents have no prior information and where an assumption that they have perfect knowledge of the distribution of profits would make little economic sense. Using this model, I also estimated the spatial competition between agents.

Counterfactual exercises showed that additional information leads to more adoption and higher market profits. It also showed that agents undervalue the informativeness of others' profits, and therefore do not use all the information available to them. Moreover, if agents used all of the available information, not only adoption and profits but also the profit per adopter would increase, suggesting that firms would adopt earlier or more efficiently.

This paper also developed a framework to correct for the selection problem in adoption when the actions of firms are based on their beliefs which in turn depend on the actions of other firms. The method relies on consistency or equilibrium-like notions: the parameters and primitives of the model lead to certain beliefs on the part of agents when they adopt, and these beliefs in turn imply certain values of parameters of the model, when the model estimates are a fixed point of this mapping.

While this paper has agents that compete and learn from one another based on their physical distances, the methods introduced can be implemented in many competitive environments where there is a measure of similarity between firms and social learning is key. One well suited example would be the car manufacturers with "distances" defined based on some similarity measure that decide whether to invest in electric engines, learning about the profitability from success of competitors of different levels of similarity.

References

Acemoglu, D., M. A. Dahleh, I. Lobel, and A. Ozdaglar (2011): "Bayesian learning in social networks," *The Review of Economic Studies*, 78, 1201–1236.

Bailey, M., D. M. Johnston, T. Kuchler, J. Stroebel, and A. Wong (2019): "Peer Effects in Product Adoption," NBER Working Paper No. w25843. Available at SSRN: https://ssrn.com/abstract=3390991.

Bailey, N. T. (1975): *The mathematical theory of infectious diseases and its applications*, Hafner Press/ MacMillian Pub. Co.

Bala, V. and S. Goyal (1998): "Learning from neighbours," *The review of economic studies*, 65, 595–621.

Banerjee, A., A. G. Chandrasekhar, E. Duflo, and M. O. Jackson (2013): "The diffusion of microfinance," *Science*, 341, 1236498.

Banerjee, A. V. (1992): "A simple model of herd behavior," *The quarterly journal of economics*, 107, 797–817.

Bass, F. M. (1969): "A new product growth for model consumer durables," *Management science*, 15, 215–227.

Benkard, C. L. (2000): "Learning and forgetting: The dynamics of aircraft production," *American Economic Review*, 90, 1034–1054.

Bikhchandani, S., D. Hirshleifer, and I. Welch (1992): "A theory of fads, fashion, custom, and cultural change as informational cascades," *Journal of political Economy*, 100, 992–1026.

Board, S. and M. Meyer-ter Vehn (2018): "Learning Dynamics in Social Networks," Tech. rep., Working Paper.

Coleman, J. S., E. Katz, and H. Menzel (1966): *Medical innovation: A diffusion study*, Bobbs-Merrill Co.

Conley, T. G. and C. R. Udry (2010): "Learning about a new technology: Pineapple in Ghana," *American economic review*, 100, 35–69.

Covert, T. R. (2015): "Experiential and social learning in firms: the case of hydraulic fracturing in the Bakken Shale," *Available at SSRN 2481321*.

Dasaratha, K., B. Golub, and N. Hak (2019): “Learning from Neighbors About a Changing State,” *Working Paper*.

DeGroot, M. H. (1974): “Reaching a consensus,” *Journal of the American Statistical Association*, 69, 118–121.

DeMarzo, P. M., D. Vayanos, and J. Zwiebel (2003): “Persuasion bias, social influence, and unidimensional opinions,” *The Quarterly journal of economics*, 118, 909–968.

Doraszelski, U., G. Lewis, and A. Pakes (2018): “Just starting out: Learning and equilibrium in a new market,” *American Economic Review*, 108, 565–615.

Foster, A. D. and M. R. Rosenzweig (1995): “Learning by doing and learning from others: Human capital and technical change in agriculture,” *Journal of political Economy*, 103, 1176–1209.

Frongillo, R. M., G. Schoenebeck, and O. Tamuz (2011): “Social learning in a changing world,” in *International Workshop on Internet and Network Economics*, Springer, 146–157.

Golub, B. and E. Sadler (2017): “Learning in social networks,” *Available at SSRN 2919146*.

Griliches, Z. (1957): “Hybrid corn: An exploration in the economics of technological change,” *Econometrica, Journal of the Econometric Society*, 501–522.

Harel, M., E. Mossel, P. Strack, O. Tamuz, et al. (2017): “Groupthink and the Failure of Information Aggregation in Large Groups,” *arXiv preprint arXiv:1412.7172*.

Hodgson, C. (2018): “Information Externalities, Free Riding, and Optimal Exploration in the UK Oil Industry,” *Working Paper*.

Holmes, T. J. (2011): “The diffusion of Wal-Mart and economies of density,” *Econometrica*, 79, 253–302.

Kellogg, R. (2011): “Learning by drilling: Interfirm learning and relationship persistence in the Texas oilpatch,” *The Quarterly Journal of Economics*, 126, 1961–2004.

Kim, D. A., A. R. Hwong, D. Stafford, D. A. Hughes, A. J. O’Malley, J. H. Fowler, and N. A. Christakis (2015): “Social network targeting to maximise population behaviour change: a cluster randomised controlled trial,” *The Lancet*, 386, 145–153.

Krusell, P. and A. A. Smith, Jr (1998): "Income and wealth heterogeneity in the macroeconomy," *Journal of political Economy*, 106, 867–896.

Lee, R. S. (2013): "Vertical integration and exclusivity in platform and two-sided markets," *American Economic Review*, 103, 2960–3000.

Molavi, P., A. Tahbaz-Salehi, and A. Jadbabaie (2018): "A Theory of Non-Bayesian Social Learning," *Econometrica*, 86, 445–490.

Pakes, A. and J. Porter (2016): "Moment inequalities for multinomial choice with fixed effects," Tech. rep., National Bureau of Economic Research.

Pastor-Satorras, R. and A. Vespignani (2001): "Epidemic spreading in scale-free networks," *Physical review letters*, 86, 3200.

Romano, J. P., A. M. Shaikh, and M. Wolf (2014): "A practical two-step method for testing moment inequalities," *Econometrica*, 82, 1979–2002.

Sadler, E. (2019): "Diffusion games," *Available at SSRN 2624865*.

Sargent, T. J. (1987): *Macroeconomic Theory: 2nd (second) Edition*, Emerald Group Publishing Limited.

Seim, K. (2006): "An empirical model of firm entry with endogenous product-type choices," *The RAND Journal of Economics*, 37, 619–640.

Shalizi, C. R. and A. C. Thomas (2011): "Homophily and contagion are generically confounded in observational social network studies," *Sociological methods & research*, 40, 211–239.

Williams, C. K. and C. E. Rasmussen (2006): *Gaussian processes for machine learning*, vol. 2, MIT press Cambridge, MA.

A Semi-Parametric Estimation Details

This appendix provides details of the estimation described in section 3.

The moments used in the estimation are achieved by taking the expectation over inequality (3.1):

$$E_i \left[\underbrace{\frac{\beta}{\alpha} \left(|A^{\ell}_{i,\tau_i-1}| - |A^{\ell}_{i,\tau_i-1-k}| \right) + \overline{\pi}^{\ell}_{i,\tau_i-1} - \overline{\pi}^{\ell}_{i,\tau_i-1-k}}_{\Delta V_{i,k}} \right] \geq 0.$$

These expectations can be estimated by averages of the observations for each k. Additional moments can be created by multiplying each element $\Delta V_{i,k}$ by a weakly positive number h_{i,τ_i-k} that depends on the characteristics of i. These moments are

$$\frac{1}{|N_k|} \sum_{i \in N_k} [\Delta V_{i,k} h_{i,\tau_i-k}] \geq 0,$$

where $N_k = \{i | \tau_i - k \geq 0\}$ is the set of all agents that waited at least k periods before adopting. The values of h_{i,τ_i-k} that I use are defined as

$$h_{i,\tau_i-k} = \begin{cases} m_{\tau_i-k}, & m_{\tau_i-k} > \bar{m} \\ 0, & m_{\tau_i-k} \leq \bar{m}, \end{cases}$$

where m_t is the size of the jackpot of the first Mega Millions lottery in period t and $\bar{m}$ is the average of m_t. This means that I multiply values in periods where the jackpot was high by the value of the jackpot, and the values in other periods by 0. Since the data set has 55 periods, with two instruments: the jackpots instrument and constant vector, there are 108 moments that can be used to give a lower bound to $\frac{\beta}{\alpha}$.[23]

Since there is a large number of moments, with many potentially far from binding, I use the two-step method in Romano et al. (2014) that selects moments with which to calculate the bound on the parameters. As mentioned in the main text, the result of the estimation is that the lower bound on the 95% confidence set of $\frac{\beta}{\alpha}$ is greater than 45.3. This means, that when an agent observes another neighbor adopting, this information affects her probability of adopting in the same way as observing her neighbors' average profits increase by at least \$45.3 per month. Given that this represents a lower bound, and that on average an agent observes adoption of about 5 neighbors between her adoption period and any preceding period, this would translate to an equivalent of having observed an increase of more than

[23]Since only adoption is observed and in practice there is no exit, only a lower bound can be calculated.

$2,700 in yearly profits.

One issue is that in the last few periods there are only a few adopters, which means that for high k there are not many observations. Therefore, one might be concerned that the covariance matrix of the moments is not accurately estimated and hence the results might be biased. To alleviate this concern, I ran the same exercise limited to periods with at least 2,500 observations, which translated to $k \in \{1, ..., 8\}$. In this setup the results were similar, with $\frac{\beta}{\alpha} > 63.6$, and average adoption of 3.3 nodes between an agent's adoption period and any preceding period.

B Robustness: All Municipalities

This section provides estimates for the model parameters without removing nodes in municipalities that initially prohibited adoption, as described in the first paragraph of section 6. Since competition is estimated offline, the estimates are the same and therefore not reported for this exercise.

The reported estimates are similar in sign and magnitude to the ones in the main text, where municipalities that initially prohibited adoption were excluded from the second part of the analysis.

Table 11: Adoption Decision Parameters

Parameter	Value	Confidence Interval
$-\frac{\lambda}{2}\sigma_0$	-23000.93353660123	$[-107257, -3619]$
γ_t	-312.8563745	$[-2658, -204]$
γ_{American}	-3410.525342	$[-206999, 80317]$
γ_{bar}	2852.874509	$[-206999, 80317]$
$\gamma_{\text{bar\&grill}}$	2396.961355	$[-141317, 83563]$
γ_{Italian}	-7826.736723	$[-143176, 83225]$
γ_{Mexican}	-9803.907215	$[-255907, 76750]$
$\gamma_{\text{MissingType}}$	-3498.456434	$[-287930, 76448]$
$\gamma_{\text{PizzaPlace}}$	-6724.532465	$[-210664, 79801]$
$\gamma_{\text{restaurant}}$	-4364.550335	$[-241638, 77288]$
γ_{Veteran}	1028.813781	$[-224755, 79380]$
$\gamma_{\text{OtherTypes}}$	-5723.088948	$[-158106, 82362]$

Table 9: $\boldsymbol{\beta}$ Estimates

Parameter	Value	SE
pop/N_m	0.02531138	0.00973083(***)
Rating (0-5)	-206.9918717	152.5865952
No Rating Reported	-1369.642018	746.0134164(*)
Price Level (1,2,3,4)	-922.6888963	220.8321112(***)
No Price Level Reported	-1940.59901	315.8133797(***)
β_{American}	6611.321771	1132.563941(***)
β_{bar}	6617.8685	1091.665016(***)
$\beta_{\text{bar\&grill}}$	7016.111362	1094.747158(***)
β_{Italian}	5734.092723	1613.743986(***)
β_{Mexican}	4071.444886	1332.811259(***)
$\beta_{\text{MissingType}}$	6392.465508	1195.043283(***)
$\beta_{\text{PizzaPlace}}$	5899.182716	1331.286606(***)
$\beta_{\text{restaurant}}$	5386.172002	1170.31366(***)
β_{Veteran}	6639.290589	1272.325301(***)
$\beta_{\text{OtherTypes}}$	6390.839606	1169.508027(***)

Table 10: Beliefs Parameters

Parameter	Value	Confidence Interval
ρ	0.253098	$[0.0814733, 0.99999]$
r^{ℓ}	1.93054	$[1.206, 38.845]$
σ_{ϵ}	1537.88	$[1490.57, 1583.25]$
σ_0	Large	Large

C Dynamics: Forward-Looking Agents

This part models only the value function that agents use when deciding when to adopt. Agents learn from observations, as described in the main text.

This model is an infinite-horizon model in which agents believe that if they do not adopt until the last period they will never adopt, though if they adopt at some period t they will receive their expected profit in perpetuity; therefore, the expected value of adoption is

$$\begin{aligned}
V_{i,t}(\eta_{i,t}, \Omega_{i,t}) = \\
\max\Big\{\delta(1-\delta)^{-1}\tilde{\mu}_{i,t} - E\left[\textstyle\sum_{s=1}^{\infty}\delta^s C_{i,t+s}|\Omega_{i,t}, a_{i,t}=1\right] - FC + \eta_{i,t}, \\
\delta E\left[V_{i,t+1}(\eta_{i,t+1}, \Omega_{i,t+1})|\Omega_{i,t}\right]\Big\},
\end{aligned}$$

where the first term in the max function is the expected profit from adoption and the second term is the expected profit from staying out, with $\tilde{\mu}_{i,t} = \mu_{i,t} + X_i\beta$ and $\eta_{i,t}$ being and i.i.d. logit shock that represents the adoption cost.

Therefore, let $EV(\Omega_{i,t})$ represent the expected value function given the state. Then

$$\begin{aligned} EV_i(\Omega_{i,t}) &\equiv \int_{\eta_{i,t}} V_{i,t}(\eta_{i,t}, \Omega_{i,t})dP(\eta_{i,t}) \\ &= \ln\Big(\exp\Big(\delta(1-\delta)^{-1}\tilde{\mu}_{i,t} - E\left[\textstyle\sum_{s=1}^{\infty}\delta^s C_{i,t+s}|\Omega_{i,t}, a_{i,t}=1\right] - FC\Big) \\ &\qquad\qquad + \exp\left(\delta E\left[EV_i(\Omega_{i,t+1})|\Omega_{i,t}\right]\right)\Big). \end{aligned}$$

The agents have some beliefs, or a misspecified model, about the definition of the state of the world and how this state of the world evolves. There is a discrete state space, where each node i in period t is in one of the states. This state of the world is a vector of size 3, where the first coordinate is the fraction of competitors that adopted $\left(f^c_{i,t}\right)$, the second coordinate represents the fraction of neighbors that could learn from those that adopted $\left(f^l_{i,t}\right)$, and the third coordinate is the "era" the period is in. The first two coordinates partition the set of neighbors of node i into p^c and p^l , and there are 9 "eras" of length 6 months in order to reduce the size of the state space; therefore, the size of the state space is $p^c \times p^l \times 9$. Explicitly, the assumption here is that

$$\tilde{\Omega}_{i,t} = \left(f^C_{i,t}, f^l_{i,t}, a_{i,t}, e(t)\right),$$

where $e(t) = \left\lceil \frac{t}{9} \right\rceil$. A reasonable partition of the other two would be dividing each into 4 parts, making the state space of size 144. Node i believes that the state of the world can transition to any other state of the world with some probability, according to some Markov matrix of transition probabilities:

$$P_{\theta(i)}\left(\tilde{\Omega}_{i,t+1}|\tilde{\Omega}_{i,t}\right),$$

which also depends on the type $\theta(i)$ of the node.

Agent i also has some primitive characteristics, such as θ_i, N_i^C, N_i^l are, respectively, the set of neighbors that are competitors of i and the set that i could learn from; $\overline{C}_i$ which represents the agent's maximal potential competition; and $\underline{\Sigma}_i$, which is the minimal potential variance that i could observe. These two bounds are achieved when all of i's competitors and other neighbors adopt.

For a given parameter ρ, and a given state of the world $\tilde{\Omega}_{i,t}$, there is a distribution of

$\tilde{\Omega}_{i,t+k}$ for every $k > 0$. Measures of the competition and variance at a future state are

$$\tilde{C}_i^{t+k} = C_i^t + \left(\overline{C}_i - C_i^t\right) \bar{f}_{i,t,k}^C$$

and

$$\tilde{\Sigma}_{ii}^{t,t+k} = \frac{\Sigma_{ii}^t \underline{\Sigma}_{ii}}{\bar{f}_{i,t,k}^l \Sigma_{ii}^t + \left(1 - \bar{f}_{i,t,k}^l\right) \underline{\Sigma}_{ii}},$$

where $\bar{f}_{i,t,k}^C$ and $\bar{f}_{i,t,k}^l$ are the expected fraction of competition and learning neighbors that adopted between the two periods:

$$\bar{f}_{i,t,k}^C = \frac{f_{i,t+k}^C - f_{i,t}^C}{1 - f_{i,t}^C}; \;\; \bar{f}_{i,t,k}^l = \frac{f_{i,t+k}^l - f_{i,t}^l}{1 - f_{i,t}^l}.$$

For these two measures, I assume that i believes that both competition and information arrive uniformly.

In period final decision period T, i's decision problem is fairly simple, which is adopt if and only if

$$EV_{i,1}^t(\Omega_T) \equiv \underbrace{\beta(1-\beta)^{-1}\left[\tilde{\mu}_{i,T} - E\left[C_{i,T+1}|\Omega_{i,T}, a_{i,T} = 1\right]\right] - FC}_{\tilde{\mu}_{\Omega_T}} + \eta_{i,T} > 0 \equiv EV_0^t(\Omega_T).$$

Additionally, from Lemma 2 we know that $\text{Var}_t(\mu_i^{t+k}) = \Sigma_{ii}^t - \Sigma_{ii}^{t+k}$, and since agents are Bayesian, $E\left[\tilde{\mu}_{i,T}|\tilde{\mu}_{i,t}\right] = \tilde{\mu}_{i,t}$. Therefore in any period t, and for every possible state in periods $t' > t$, we know that

$$\tilde{\mu}_{i,\Omega_{t'}}^t \sim N\left(\beta(1-\beta)^{-1}\tilde{\mu}_{i,t} - \sum_{s=t'+1}^{\infty} \beta^{s-t'} E\left[C_{i,s}|\Omega_{t'}\right] - FC, \underbrace{\Sigma_{ii}^t - \tilde{\Sigma}_{ii}^{t,t'}}_{\sigma_{\Omega_{t'}}^t}\right)$$

Since if i stays out at period T she will get 0, we can use the truncated normal formula to get that

$$EV_i^t(\Omega_{i,T}) = E\left[\tilde{\mu}_{i,\Omega_T}^t\right] + \frac{\sigma_{i,\Omega_T}^t \phi\left(\frac{-\tilde{\mu}_{i,\Omega_T}^t}{\sigma_{i,\Omega_T}^t}\right)}{1 - \Phi\left(\frac{-\tilde{\mu}_{i,\Omega_T}^t}{\sigma_{i,\Omega_T}^t}\right)}.$$

Using that result, we can calculate the value of staying out at period $T-1$,

$$EV_{i,0}^t(\Omega_{i,T-1}) = \beta \sum_{\Omega_T} Pr\left(\Omega_T|\Omega_{i,T-1}\right) \cdot EV_i^t(\Omega_T),$$

and the expected value of adopting,

$$EV^t_{i,1}(\Omega_{T-1}) = \beta(1-\beta)^{-1}\tilde{\mu}_{i,T-1} - \sum_{s=T}^{\infty}\beta^{s+1-T}E\left[C_{i,s}|\Omega_{i,T-1}, a_{i,t'}=1\right] - FC + \eta_{i,T-1}$$

Therefore, the expected value in period $T-1$ is

$$EV^t_i(\Omega_{T-1}) = E\left[\tilde{\mu}^t_{i,\Omega_{T-1}}\right] + \frac{\sigma^t_{\Omega_T}\phi\left(\frac{EV^t_{i,0}(\Omega_{i,T-1})-\tilde{\mu}^t_{i,\Omega_{T-1}}}{\sigma^t_{\Omega_{i,T-1}}}\right)}{1-\Phi\left(\frac{EV^t_{i,0}(\Omega_{i,T-1})-\tilde{\mu}^t_{\Omega_{i,T-1}}}{\sigma^t_{\Omega_{i,T-1}}}\right)}.$$

Continuing in this fashion, using backwards induction, we can calculate the expected value in every period, arriving at $EV^t_{i,1}(\Omega_t)$ and $EV^t_{i,0}(\Omega_t)$.

In equilibrium;

1. in each period t, agent i chooses to adopt if and only if $EV^t_{i,1}(\Omega_t, \eta_{i,t}) \geq EV^t_{i,0}(\Omega_t)$, given the beliefs regarding transition probabilities $P_{\theta(i)}\left(\tilde{\Omega}_{i,t+1}|\tilde{\Omega}_{i,t}\right)$.

2. The beliefs about transition probabilities are consistent with the realized values of $EV^t_{i,d}(\Omega_{T-1}, \eta_{i,t})$, which are consistent with actions taken.

Now since there is an expected value in every period t for every agent i and decision $d \in \{0,1\}$, the moments to match are that the expected adoption in each period equals the actual adoption:

$$m_t \equiv \left|\sum_{i\notin A_{t-1}} \frac{\exp\left(EV^t_{i,1}(\Omega_t,\eta_{i,t})\right)}{\exp\left(EV^t_{i,1}(\Omega_t,\eta_{i,t})\right)+\exp\left(EV^t_{i,0}(\Omega_t)\right)} - (A_t - A_{t-1})\right|,$$

and we choose ρ, β, FC that satisfy

$$\arg\min_{\rho,\beta,FC}\sum_t m_t.$$

Lemma 2. *Let*

$$\begin{pmatrix} x \\ \mathbf{y} \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_x \\ \mu_{\mathbf{y}} \end{pmatrix}, \begin{pmatrix} \Sigma_x & \Sigma_{x\mathbf{y}} \\ \Sigma'_{x\mathbf{y}} & \Sigma_{\mathbf{yy}} \end{pmatrix}\right),$$

where x is a scalar and $\mathbf{y}$ is a vector. Then

$$Var(E(x|\mathbf{y})) = \Sigma_x - Var(x|\mathbf{y}).$$

Proof. See appendix D. □

The Evolution of Variance

The belief of agent i in period t is that

$$\xi_{i,t} \sim N(\mu_i^t, \Sigma_{ii}^t).$$

Denote by x an independent normal signal that combines all of the information that i could still potentially receive. In other words, an independent signal that would reduce her variance to the variance shed had in the case she had perfect knowledge of her neighbors' unobservables, $\underline{\Sigma}_{ii}$. Formally, using a known result for optimal combination of private normal signals

$$Var\left(\frac{\Sigma_x \xi_{i,t} + \Sigma_{ii}^t x}{\Sigma_x + \Sigma_{ii}^t}\right) = \underline{\Sigma}_{ii}$$

and isolating Σ_x, we get

$$\Sigma_x = \frac{\underline{\Sigma}_{ii}\Sigma_{ii}^t}{\Sigma_{ii}^t - \underline{\Sigma}_{ii}}.$$

If we assume that x is an average of many "smaller" independent signals with identical variance, and that i observes only a fraction $\bar{f}_{i,t,k}^l$ of these signals, the variance of the equivalent signal, x_f, would be

$$\Sigma_{x_f} = \frac{\Sigma_x}{f} = \frac{\underline{\Sigma}_{ii}\Sigma_{ii}^t}{f(\Sigma_{ii}^t - \underline{\Sigma}_{ii})}$$

Combined optimally,

$$\tilde{\Sigma}_{ii}^{t,t+k} = Var\left(\frac{\Sigma_{x_f}\xi_{i,t} + \Sigma_{ii}^t x}{\Sigma_{x_f} + \Sigma_{ii}^t}\right) = \frac{\Sigma_{ii}^t \underline{\Sigma}_{ii}}{\bar{f}_{i,t,k}^l \Sigma_{ii}^t + \left(1 - \bar{f}_{i,t,k}^l\right)\underline{\Sigma}_{ii}}.$$

D Proofs

D.1 Lemma 1

Proof. Let $A_t = \{1, ..., m-1\}$ represent the set of agents that adopted up to and including period t, where the indices of the agents represent the order in which they adopted. For simplicity of notation, we assume that only one agent adopted at a time, and that they adopt in their index period (i.e., i adopted in period i), though it is easily generalized.

The expectation of agent m's beliefs, in period m is

$$\mu_{m,m} = \underbrace{\hat{\Sigma}_{m,A_m}\left(\hat{\Sigma}_{A_m} + \sigma_\epsilon D_{A_m,m}\right)^{-1}}_{\hat{w}_m}\bar{\boldsymbol{\xi}}_{A_m,m},$$

where $\hat{\Sigma}_{A_m}$ is the prior beliefs about the covariance matrix of the subset of agents A_m, and $\hat{\Sigma}_{m,A_m}$ is the prior beliefs of the correlations m with the existing adopters. D_{m,A_m} is a diagonal matrix in which each diagonal element is the inverse of the number of observations of profits made by the relevant agent in period m, and $\bar{\boldsymbol{\xi}}_{A_m,m}$ is the average of these observations. The expression that is multiplied by $\bar{\boldsymbol{\xi}}_{A_m,m}$ is a vector of the weights $\hat{w}_m$ that m puts on her signals when she updates her own beliefs. Similarly, I denote by w_m the weights she would've put on these beliefs if she had the correct prior.

The error that is due to her incorrect beliefs is exactly

$$\epsilon_m^{Beliefs} = (w_m - \hat{w}_m)\,\bar{\boldsymbol{\xi}}_{A_m,m} = \sum_{i=1}^{m}\left(w_m^i - \hat{w}_m^i\right)\bar{\xi}_{i,m},$$

where $\hat{w}_m^i$ represents the weights that m puts on i, and $\bar{\xi}_{i,m}$ represents the average of observations made by i in period m.

The rest of the proof uses induction. By assumption, the error of person 1 is exactly zero, since her beliefs under the two priors are 0. The error of the second adopter is

$$\epsilon_2^{Beliefs} = \left(w_2^1 - \hat{w}_2^1\right)\bar{\xi}_{1,2}$$

We know that ex ante $E\left[\xi_1\right] = 0$; therefore, $E\left[\bar{\xi}_{1,2}\right] = 0$ and $E\left[\epsilon_2^{Beliefs}\right] = 0$. By definition, we can write $\xi_2 = \hat{\mu}_2 + \epsilon_2^{Beliefs} + \epsilon_2^{\xi}$, hence

$$\xi_2 = \hat{w}_2^1\bar{\xi}_{1,2} + \left(w_2^1 - \hat{w}_2^1\right)\bar{\xi}_{1,2} + \epsilon_2^{\xi}$$

and $E\left[\xi_2\right] = 0$. Therefore, for agent 3 we have

$$\epsilon_3^{Beliefs} = \left(\hat{w}_3^1 - w_3^1\right)\bar{\xi}_{1,3} + \left(\hat{w}_3^2 - w_3^2\right)\bar{\xi}_{2,3},$$

and hence $E\left[\epsilon_3^{Beliefs}\right] = 0$. This procedure can be repeated for any m. $\square$

D.2 Lemma 2

Proof. A known result is that if $\mathbf{y}$ is observed, then

$$x|\mathbf{y} \sim N(\mu_{x|\mathbf{y}}, \Sigma_{x|\mathbf{y}}),$$

with

$$E(x|\mathbf{y}) = \mu_x + \Sigma_{x\mathbf{y}}\Sigma_{\mathbf{yy}}^{-1}(\mathbf{y} - \mu_{\mathbf{y}})$$

$$\text{Var}(\text{x}|\mathbf{y}) = \Sigma_x - \Sigma_{x\mathbf{y}}\Sigma_{\mathbf{yy}}^{-1}\Sigma'_{x\mathbf{y}}.$$

If it is known that $\mathbf{y}$ is to be revealed but its value is not yet known, then

$$\begin{aligned}\text{Var}(E(x|\mathbf{y})) = \text{Var}(\mu_x + \Sigma_{x\mathbf{y}}\Sigma_{\mathbf{yy}}^{-1}(\mathbf{y} - \mu_{\mathbf{y}}))&\\ = \text{Var}(\Sigma_{x\mathbf{y}}\Sigma_{\mathbf{yy}}^{-1}\mathbf{y}) = \Sigma_{x\mathbf{y}}\Sigma_{\mathbf{yy}}^{-1}\Sigma'_{x\mathbf{y}} &= \Sigma_x - \text{Var}(x|\mathbf{y}).\end{aligned}$$

□

E Computation of β and Simulation

E.1 Computation Of $\boldsymbol{\beta}$

Here I describe the details of the algorithm that gives an unbiased estimate of $\boldsymbol{\beta}$. This is given a guess of ρ and r^ℓ, θ, meaning that one knows or assumes the parameters of the kernel function κ. Knowing the function κ means that for every municipality there is one constant correlation matrix K.

The main idea of this algorithm is to iteratively update nodes' prior beliefs, and beliefs about their observables in the period in which they adopt until we find the beliefs that will satisfy the equilibrium described in the main text.

In the first step, I initialize the vector of beliefs at adoption for every agent that entered, $\boldsymbol{\mu}_\tau^0$, to a vector of zeros.

Given this initialization, we describe the process of each iteration c:

1. Given $\boldsymbol{\mu}_\tau^c$, a regression of $\boldsymbol{F} - \boldsymbol{\mu}_\tau^c$ of X is used to do the following:
   (a) generate $\boldsymbol{\beta}^c$;
   (b) generate the observations $\xi_{j,t}^c = \pi_{j,t}^g - \boldsymbol{x}_j\boldsymbol{\beta}^c + C_j(A_t)$ that are used in the updating process.
2. Using Σ_0 and the objects calculated in 1, the beliefs of every agent in every period calculated, in particular $\boldsymbol{\mu}_\tau^{c+1}$.
3. If
$$\frac{\sum_i(\mu_{\tau_i}^{c+1} - \mu_{\tau_i}^c)^2}{N} < \epsilon^{\text{tolerance}} \equiv 10^{-12},$$
then stop. Else, repeat for $c + 1$.

E.2 Simulation

In order to show that the algorithm above actually works on some data, I generated data using some initial parameters, and tried to recover the parameters using the procedure described above. I repeated the process for 1,020 simulation draws.

In each iteration of the simulation, there is one municipality of size $N = 800$, and each node's longitude and latitude are each drawn from $N(0,\ 4)$. Using a correlation parameter $\rho = 0.7$, and each node is randomly assigned into one of two types with equal probabilities, type 1 has $\mu_{\theta=1} = 800$ and $\sigma_{\theta=1} = 2000$, and type 2 has $\mu_{\theta=2} = 700$ and $\sigma_{\theta=2} = 1000$. Signal variance is $\sigma_\epsilon = 20$. Each node is also assigned an independent $X_i \sim N(40,\ 225)$ and the parameter that is multiplied by X is defined to be $\beta = 20$. These define the true $F_i = X_i\beta + \xi_i$.

Afterwards, an adoption procedure was simulated in which there is a constant fixed cost of adoption equal to $FC = 55\bar{F}$. Simulating adoption logit shock, each node adopts with probability

$$p_{i,t} = \frac{\exp(x_i\beta + \mu_{i,t})}{\exp(x_i\beta + \mu_{i,t}) + \exp(FC)}.$$

For each node that adopts, there is a realization of an observation $F_i + \epsilon_{i,t}$, and the neighbors of node i update according to the updating process in subsection 4.3. These determine $\mu_{i,t}$ and hence the probability of adoption.

With the simulated data, I used the procedure described above in order to estimate the parameters of β and also to see how close the estimates of $\boldsymbol{\mu}_\theta$ and $\boldsymbol{\sigma}_\theta$ are to the actual realized ones in every iteration. I also added an estimate done by a linear regression as a benchmark comparison. Estimation using this procedure seems to be consistent, while the regression is biased, as seen in the figure Figure E.1 on page 59.

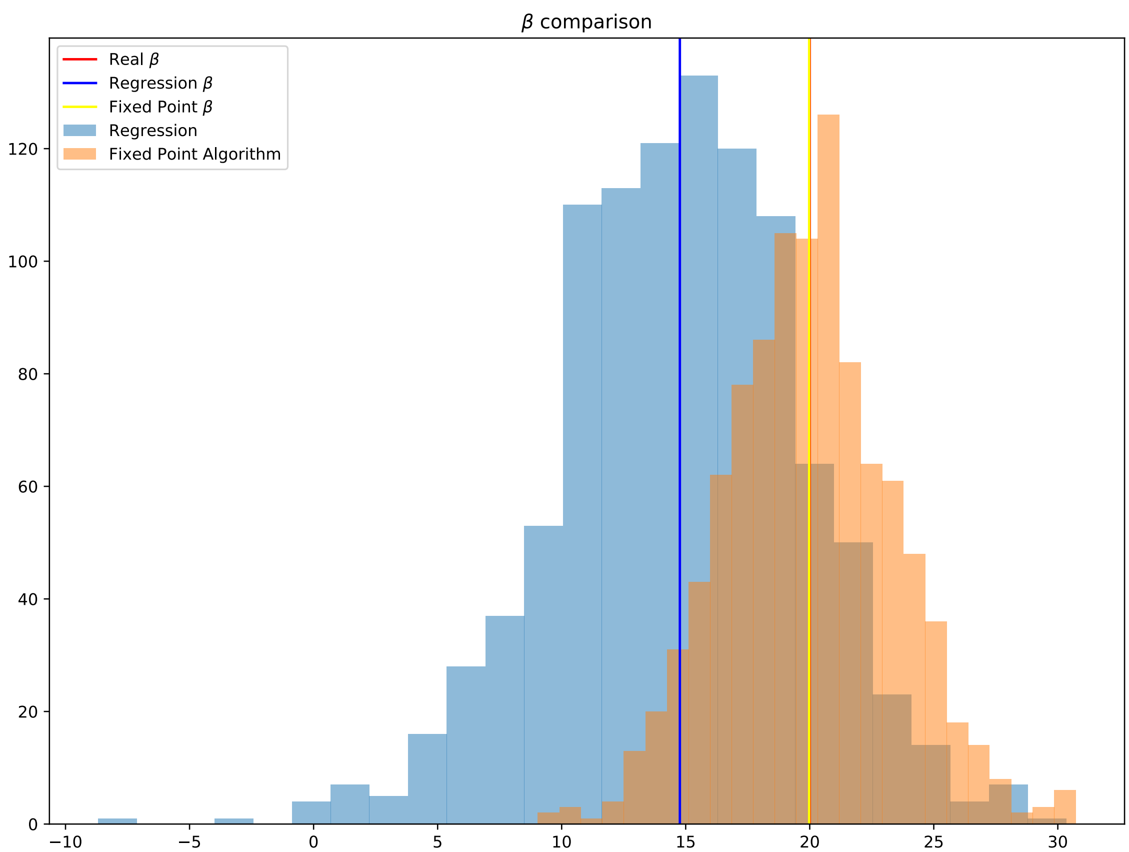
F Estimation of $-\frac{\lambda}{2}\sigma_{\xi 0}$

In section 6 I presented the estimation results for the parameter $-\frac{\lambda}{2}\sigma_{\xi 0}$ that represents the total effect of risk aversion and uncertainty on the certainty equivalent of adoption. I do so since in the data the value of $\sigma_{\xi 0}$ is too large numerically estimate in a meaningful way, and that when $\sigma_{\xi 0} \to \infty$ the parameters λ and $\sigma_{\xi 0}$ are not separately identified.

Lemma 3. *Under the assumptions of the model in section 4. If $\sigma_{\xi 0} \to \infty$, then the parameters λ and $\sigma_{\xi 0}$ are not separately identified.*

Proof. The value of the variance of agents i at period t that observed the profits of the agents in A_t is:

Figure E.1: Comparing β Estimates to Truth



Red line: (which almost coincides with the yellow line) the true parameter; Yellow: parameter estimated with the fixed point procedure; Blue: regression estimator. Orange and blue histograms: distributions of estimators for the fixed point procedure and regression respectively.

$$\sigma_{i,t}^2 = \sigma_{\xi 0}^2 - \Sigma_{i,A_t}\Sigma_{A_t}^{-1}\Sigma_{A_t,i},$$

where Σ_{i,A_t} and $\Sigma_{A_t,i}$ are row and column vectors of the covariance of i with all current adopters, while Σ_{A_t} is the covariance of the ξ's of all current adopters. These can be further separated to the matching correlation vectors ρ_{i,A_t} and $\rho_{A_t,i}$, and the correlation matrix K_{A_t}:

$$\sigma_{i,t}^2 = \sigma_{\xi 0}^2 - \sigma_{\xi 0}^2\rho_{i,A_t}\left(\sigma_{\xi 0}^2 K_{A_t} + \sigma_\epsilon^2 \mathrm{diag}(\boldsymbol{\omega}_{A_t})\right)^{-1}\sigma_{\xi 0}^2\rho_{A_t,i}.$$

At the limit, as $\sigma_{\xi 0}^2 \to \infty$ the value of $\sigma_{i,t}^2$ is of course also going to infinity. With that said, in the model, the value λ also depends on values of $\sigma_{i,t}^2$. Therefore, since $\sigma_\epsilon \ll \sigma_{\xi 0}$:

$$\begin{aligned}\lim_{\sigma_{\xi 0}^2\to\infty}\lambda\sigma_{i,t} &= \lambda\sqrt{\sigma_{\xi 0}^2 - \sigma_{\xi 0}^2\rho_{i,A_t}\left(\sigma_{\xi 0}^2 K_{A_t}\right)^{-1}\sigma_{\xi 0}^2\rho_{A_t,i}} \\ &= \lambda\sigma_{\xi 0}\sqrt{\left(1-\rho_{i,A_t}K_{A_t}^{-1}\rho_{A_t,i}\right)}\end{aligned}$$

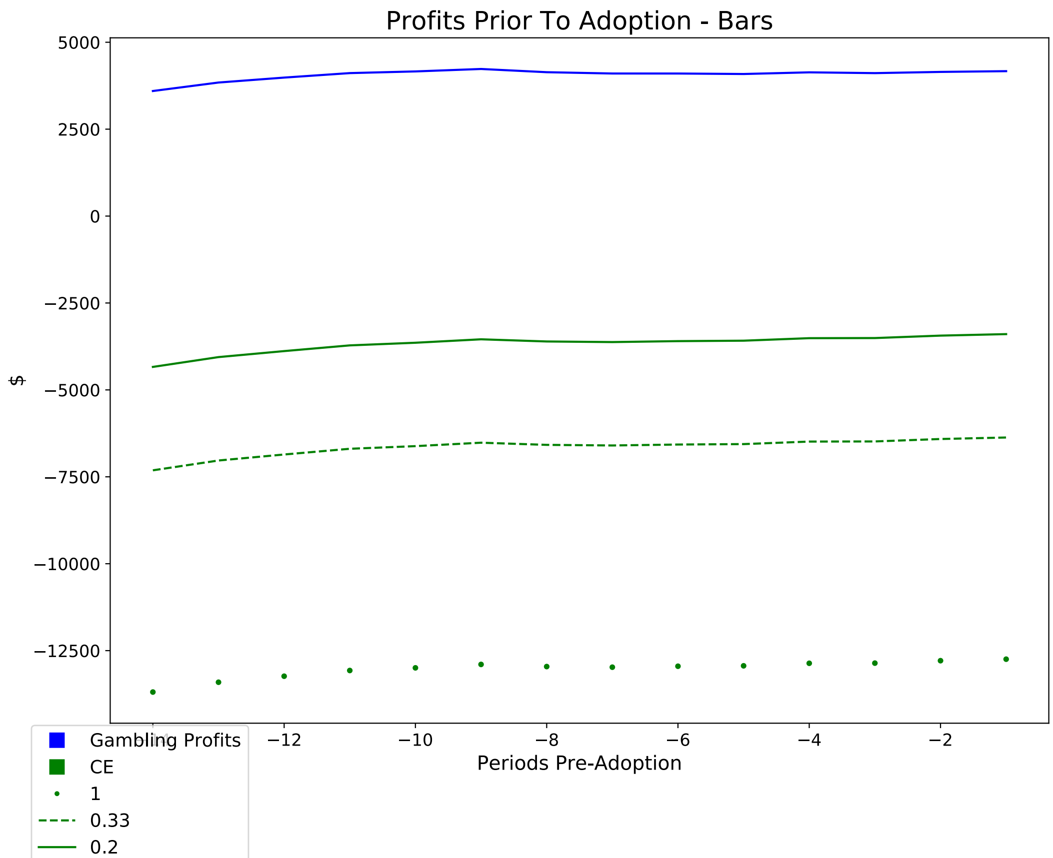
and since at the limit $\sigma_{i,t}$ is a linear function of $\sigma_{\xi 0}$, λ and $\sigma_{\xi 0}$ are not separately identified.

□

G Pre-adoption analysis

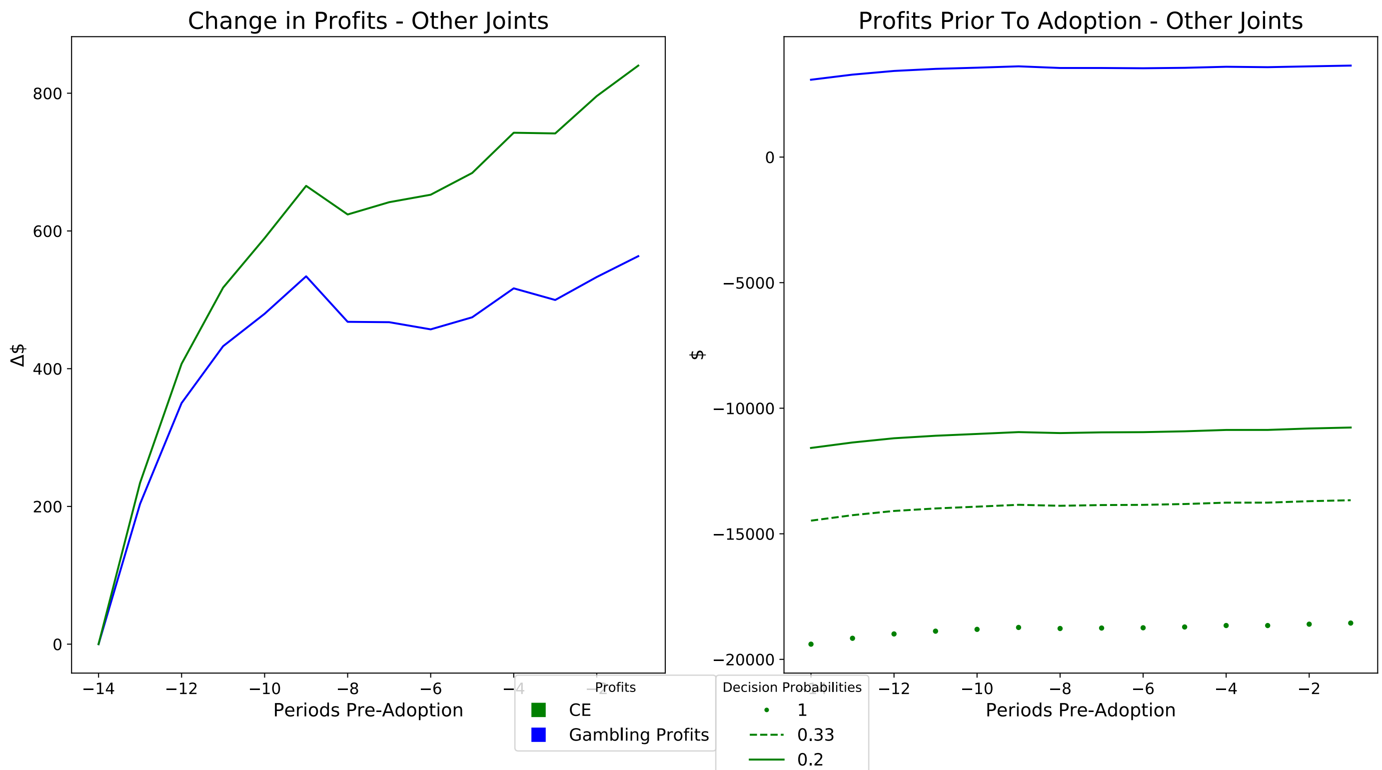
In subsection 6.3 I analyzed how the beliefs of bar-like establishments evolved in the period preceding the period of adoption. Figure 6.4 shows that their expected profits and certainty equivalent increased by about \$900 before they adopted. In figure G.1 we see the levels of beliefs that bars owners have in the period before adoption. Bar owners believe that they will make in expectations more than \$3500 when they decide to adopt, though the risk makes the owners' certainty equivalent of the lottery at less than negative \$12,000. However, if we believe that not all bar owners make an active decision whether to adopt every month, the certainty equivalent of choosing adoption increases significantly: If each agent considers adoption with probability 1/3, the model estimation results in certainty equivalents of more than negative \$7,500. If agents make an adoption decision in expectation once every five months, the certainty equivalent is greater than negative \$5,000.

Figure G.1: Beliefs Levels before Adoption: bars



Bar owners' beliefs before adoption. In blue we see their expected gambling profits before adoption. The green dots show the change in their certainty equivalent. The dashed and solid lines are the certainty equivalents in models where agents decide whether to adopt in every period with probability 1/3 and 1/5 respectively.

Figure G.2: Beliefs Change and Levels before Adoption: Other businesses



The analysis in this figure is the same analysis as in figures 6.4 and G.1 but applied to all non-bar-like businesses.