Double/Debiased Machine Learning for Static Games with Incomplete Information: An Application to Pharmacy Accessibility*

Hyung-Jin Kim†

Current draft: October 13th, 2023

Abstract

This paper develops estimation and inference methods for game-theoretic models with many covariates. The methods combine the double/debiased machine learning (DML) framework with static games with incomplete information structure as in Bajari et al. (2010b). I provide valid inferences for low-dimensional parameters of interest in the presence of high-dimensional nuisance parameters when using machine learning estimators.

An empirical application studies the growing issue of limited pharmacy access in rural Midwestern United States. The paper finds that the decline of independent pharmacies is associated with the new entry of chains which leads to more limited pharmacy access. The effects are more pronounced in elderly towns. Using the static games of independent pharmacies, this paper finds that the effect of the rival independent pharmacy is 1.5 times larger in Neyman orthogonal moment estimators compared to Bajari et al. (2010b)'s estimator. I further find that the predictive performance of machine learning plays an important role in this difference. The first counterfactual simulation studies the role of new chain pharmacies in the local market structure. In the second counterfactual experiment, this paper evaluates the effect of a subsidy program on improving limited pharmacy access, similar to the physician bonus program for Medicare-related services targeting areas with limited medical access.

Keywords: Static games; orthogonal moments; machine learning; high-dimensional data; limited pharmacy accessibility

*First draft: February 2023. I am indebted to my advisor, Arie Beresteanu, and committee members, Douglas Hanely and Mahrad Sharifvaghefi, for their guidance and support. I am grateful to David Agorastos, Japil Lee, Drake Coleman, Brett Hollenbeck, Andre Holt, Whitney Newey, and Anh Nguyen for their helpful comments and suggestions. I am also thankful to the conference and seminar participants at the 37th Pennsylvania Economic Association, North America Summer Meeting of the Econometric Society, 2023 (NASME 2023), International Association of Applied Econometrics Conference, 2023 (IAAE 2023), European Association for Research in Industrial Economics, 2023 (EARIE 2023), and the Econometrics seminar at the University of Pittsburgh. All errors in this paper are my own.

†Department of Economics, University of Pittsburgh, jin.kim@pitt.edu

1 INTRODUCTION

Static games involving strategic interactions are an active area of research, including entry/exit games (Bresnahan and Reiss 1991, Berry 1992), quality choices (Mazzeo, 2002), location choice models (Seim (2006)), multi-store choices (Jia 2008, Holmes 2011), networks (Nishida, 2015), pricing strategies (Ellickson and Misra, 2008), and stock recommendations submitted by analysts (Bajari et al., 2010b). To recover the underlying structural parameters of strategic interactions, Bajari et al. (2010b) is one of the most widely used two-step methods.[1]

Recently, there has been a growing interest in using high-dimensional data for empirical research in economics and solving real-world business problems. The demand for High-dimensional data rises due to more available big data with sparse features such as text, images, and the transformation of raw covariates such as interaction, polynomials, and splines. Researchers are motivated to include high-dimensional control variables to address omitted variable bias issues and allow for more flexible functional forms. However, while the aforementioned Bajari et al. (2010b) establishes desirable theoretical properties and performs well in practice using low-dimensional data, the method becomes infeasible in high-dimensional settings where the dimension of control variables p is relatively large compared to the sample size N.

In this context, researchers often turn to machine learning (ML) methods with regularization, such as the Lasso estimator, to enable estimation with high-dimensional covariates. However, Machine learning methods face a trade-off between bias and variance: while they excel in prediction, they can suffer from regularization bias. For example, the Lasso estimator is susceptible to regularization bias due to model selection errors such as selecting irrelevant covariates or not selecting relevant covariates. Consider the setting where the model parameters are categorized into two groups - low-dimensional parameters of interest (e.g. rivals effects) and high-dimensional nuisance parameters. Due to the described bias-variance trade-off, the use of machine learning methods to estimate high-dimensional nuisance parameters can enhance parameter prediction accuracy, but it may also transmit regularization bias in the nuisance parameter estimates into the primary parameter of interest.

The goal of this paper is to provide a valid inference of underlying structural parameters of interest in the presence of high-dimensional nuisance parameters estimated using machine learning methods. I employ the framework in Chernozhukov et al. (2018b) to enable the use of

[1] Bajari et al. (2010b)'s estimators take two steps: (1) estimating the rival's presence probabilities, represented as reduced-form parameters η_0, using non-parametric methods or simple logit models, and (2) recovering the low-dimensional structural parameters θ_0 based on first-stage estimates of beliefs regarding optimal choices made by other firms.

high-dimensional covariates in the model of strategic interaction with incomplete information in Bajari et al. (2010b). First, I employ machine learning methods to estimate the high-dimensional nuisance parameters including belief over competitors' choices and the effect of market characteristics. To prevent transmitting regularization bias in machine learning estimates of nuisance parameters to low-dimensional structural parameters of interest, I use two ingredients introduced in Chernozhukov et al. (2018b). I construct a moment function that satisfies the orthogonality condition which implies that the moment condition is locally insensitive to regularization bias in nuisance parameter estimates. In addition, I implement a cross-fitting algorithm to avoid imposing strong restrictions on the growth of entropy and model complexity. As long as convergence rates for ML estimators are faster than $N^{-1/4}$ and regularity conditions are satisfied, the proposed estimator achieves $\sqrt{N}$-consistency and asymptotic normality. Monte Carlo simulation evaluates the finite sample properties of developed estimators.

As an empirical application, I apply my framework to study the strategic entry decisions of independent pharmacies in rural Midwestern United States. First, I document that limited access to pharmacies in rural towns has increased over the past two decades, and this increased trend is associated with the exit of independent pharmacies. I further present data patterns that independent pharmacies choose to leave the market as more chain pharmacies enter. I use event study designs to study the effect of chain pharmacy entries on the behavior of independent pharmacies. After controlling for rich covariates, including socioeconomic characteristics, health-related variables, and market-year fixed effects, I find that new entry of chain pharmacies within 15 miles is associated with a significant decrease in the number of independent pharmacies in town, which leads to limited accessibility of pharmacy in town. The magnitudes of impact are larger in towns with a higher proportion of elderly residents who are particularly vulnerable to the issue of limited access to pharmacies.

Motivated by reduced form evidence, I estimate a structural model of the static games among independent pharmacies in the town. In this process, I use machine learning methods to recover rival's choice probabilities in the first stage and shed light on the benefit of using machine learning in estimation. There are two key findings. First, compared to classical methods in Bajari et al. (2010b), machine learning methods produce greatly improved prediction power in the estimation of beliefs over rivals' conditional choice probabilities (CCPs). Second, as a result, the structural parameter of rivals' interaction effects in the second stage is around 1.5 times larger than the estimate using Bajari et al. (2010b). This change in estimates can be primarily attributed to the ability of machine learning methods to more accurately capture beliefs about rivals' actions when compared to a simple conditional logit model.

Finally, I use the estimated model to simulate counterfactual scenarios aimed at improving pharmacy accessibility in elderly towns. The first scenario quantifies the role of new chain pharmacy entries on the market structure of local markets by looking at how many independent pharmacies would be active in the market if the number of chain pharmacies was fixed in the year 2000. In the absence of new chain entries after the year 2000, there would have been 10% more independent pharmacies, compared to the observed number in 2019. The first counterfactual simulation reveals that the new entries of chain pharmacies since 2000 can account for 40% of the variation in the closed independent pharmacies between 2000-2019.

The second scenario characterizes the equilibrium where the federal government provides a 10% subsidy from pharmacy sales associated with Medicare beneficiaries to pharmacies. This hypothetical subsidy program is inspired by the existing federal government's physician bonus program initiated in 2006 to enhance healthcare accessibility, by targeting limited medical access areas. The counterfactual analysis reveals that with this subsidy program, 16% of towns previously categorized as having limited pharmacy access would no longer fall into this category.

I chose retail pharmacy markets for the application of my methodology for three reasons. First, there is little knowledge of which market characteristics are relevant to the opening of pharmacies for econometricians. Instead of assuming that data-generating processes are known to econometricians, I incorporate a set of rich covariates from socioeconomic and health-related characteristics. This approach allows the data to detect which characteristics are relevant for the underlying payoffs of pharmacies. Secondly, given that the true functional forms of the underlying payoffs also remain unknown to econometricians, I moved away from the assumption of straightforward linear models. I instead utilize flexible functional forms, such as interaction terms, to capture the underlying payoffs of pharmacies. Finally, the pharmacy's most important strategic decision is entry/exit, which aligns with the discrete choice of games that I develop. Pharmacies compete in relatively tight geographical markets as consumers take the location of pharmacies into consideration when deciding where to shop. According to industry reports[2], consumers take the location (distance) into account for pharmacy choice decision, followed by the acceptance of health insurance and the quality of services. In Appendix Figure C.1, I observe that across different types of pharmacies (independent, chain, and, mass merchants), location consistently emerges as the most important factor determining consumers' pharmacy preferences. Admittedly, while health insurance and pricing do play roles, location remains the predominant factor of consideration.[3] Ap-

[2]2018 Pharmacy Satisfaction Pulse, Pharmacy Satisfaction Data from surveys

[3]Based on the anecdotal, I abstract away from decisions on other dimensions - prices, product variety,

pendix Figure C.2 reinforces this, underscoring location as the most important factor when consumers switch pharmacies.

Literature Review. Extensive literature exists on the estimation of structural parameters in the context of high-dimensional data. The literature focuses on the development of Neyman orthogonal moment functions to achieve $\sqrt{N}$-consistency and asymptotic normality of the estimator (Neyman 1959, Newey 1994, Belloni et al. 2016, Chernozhukov et al. 2018a, Chernozhukov et al. 2022, Ichimura and Newey 2022). Orthogonal moment functions have the property that the second-stage estimations of structural parameters are insensitive to first-stage local biases from machine learning methods. Coupled with sampling splitting, low-dimensional structural parameters of interests θ_0 follow $\sqrt{N}$-consistent and asymptotically normal, in the presence of high-dimensional data. The proposed orthogonal moment condition aligns with previous literature and has desirable asymptotic properties.

This paper also relates to deriving Neyman/orthogonal moment function for discrete choice game settings; two-stage methods (Bajari et al. 2010b, Chernozhukov et al. 2016, Nekipelov et al. 2022), dynamic games with value function approximation approach (Bajari et al. 2009, Adusumilli and Eckardt 2019), and partial identification (Semenova 2018). The previous literature in Bajari et al. (2009) and Bajari et al. (2010b) suggests that the influence function in discrete games corrects the player's own choice probabilities. In contrast, the orthogonal moments in this paper remove biases from the rival's choice probabilities because first-stage nuisance parameters include beliefs over the rival's choice probabilities. Semenova (2018) used partial identification for the dynamic discrete choice model whereas I provide point identification for the static game. My work differs from Adusumilli and Eckardt (2019) in that Adusumilli and Eckardt (2019) used the value function approximation for dynamic models based on Reinforcement Learning. Nekipelov et al. (2022) proposed correction terms for the two-player static game with the incomplete formation. My paper differs from theirs in several ways. First, as I allow firm-level shifters for the identification, it requires new correction terms, which differ from Nekipelov et al. (2022). [4] Second, this paper accommodates multiple players whereas Nekipelov et al. (2022) 's sketch includes two players' cases. Finally, while Nekipelov et al. (2022) requires the use of the loss function in the second stage, this paper uses the generalized method of moment.

The empirical application in this research also contributes to the growing literature on limited pharmacy accessibility, or "pharmacy desert" in public health literature (Amstislavski et al. 2012, Qato et al. 2014, Di Novi et al. 2020). I highlight three aspects. First, to the

health insurance in-network/out-of-cost, and qualities.

[4]Specifically, my correction terms includes conditional expectation of shifters x_i, x_{-i} and common controls x_0 to satisfy identification assumption in Tamer (2003) and Bajari et al. (2010b). In contrast, Nekipelov et al. (2022)' sketch uses only common controls x for both players.

best of my knowledge, this paper is the first to study the pharmacy desert through the lens of competition, by providing causal and structural estimates of the effect of chain pharmacy entry on local market structure. Second, I focus on rural towns only. Due to the fewer stores and smaller population in rural markets, these markets are most impacted by the closure of independent stores, increasing trends in limited pharmacy accessibility. Finally, I further document that towns with higher share of elderly population experienced rapid growth in limited accessibility - which is concerning because the elderly population faces higher transportation costs and limited mobility.

The rest of the paper is structured as follows: Section 2 provides a description of the conventional econometric models use in static games with incomplete information. The models are built upon the framework established in Bajari et al. (2010b) and Bajari et al. (2013). Section 3 characterizes the definition of Neyman orthogonal moments and their properties and also provides an asymptotic theory for proposed estimators. Section 3 also presents simulations to evaluate the finite sample properties. Section 4 illustrates data and background. Section 5 provides reduced form evidences and present the structural estimation results. Section 6 presents the counterfactual analysis. Section 7 concludes the paper. The appendix includes omitted proof in the main paper.

Notation. I use the symbol E to represent the expectation with respect to some probability measure that governs the law of the data and E_P to denote the expectation under a probability measure P.

2 MODEL FRAMEWORK

I focus on static games of incomplete information, closely following the framework presented in Bajari et al. (2010b) and Bajari et al. (2013). This context involves a defined set of players, represented as $i \in \{1, ..., N\}$. Each player has two distinct choices, which we denote by $\mathscr{J} = 2$. [5]

$$a_i = \begin{cases} 1 & \text{if Player } i \text{ chooses to being active.} \\ 0 & \text{if Player } i \text{ chooses being inactive.} \end{cases} \tag{2.1}$$

Let $\mathscr{A} = \{0, 1\}^N$ represent the Cartesian product of the choices made by all players in a market. In this framework, the objective of each player is to maximize their utility. I use $a = (a_1, \ldots, a_n)$ as a generic representation of $\mathscr{A}$. Adhering to traditional game-theoretic

[5] Without loss of generality, this framework can be extended to encompass multiple choices, represented by $|\mathscr{J}| > 2$, corresponding to a multinomial choice.

conventions, $-i$ denotes the rivals of player i, and $a_{-i} = (a_1, \ldots, a_{i-1}, a_{i+1}, \ldots, a_N)$ describes the strategy choices of all players excluding player i.

For an active player i, the flow utility is contingent upon the state variables $d_i \in \mathscr{D}_i$. The combined set of state variables for all players is represented by the vector $d = (d_1, \ldots, d_N) \in \mathscr{D}$. The Cartesian product of these state variables is denoted as $\mathscr{D} = \prod_i \mathscr{D}_i$. Additionally, I assume the state variable d is universally observed to both every player in the market and to econometricians.

In addition to the state variables, I incorporate private information, $\epsilon_i(a_i)$, drawn from distribution g_i, which is specific to each player's choices and directly influences their utilities. The private information for each player i is represented as:

$$\epsilon_i = (\epsilon_i(1), \epsilon_i(0)).$$

Here, $\epsilon_i(1)$ pertains to the situation where player i is active, while $\epsilon_i(0)$ corresponds to the player being inactive.

I make the assumption that while each player cannot observe the stochastic private information shock of their rivals, $\epsilon_{-i}(a_{-i})$, they are aware of the distribution g_{-i}. This type of information structure is prevalent in discrete choice scenarios.

Assumption 2.1 (Information)**.**

(a) *Private information is independently and identically distributed across both choices and players, drawn from a Type 1 Extreme Value Distribution,* g.

(b) *Each player privately observes their own* ϵ *which remains hidden from analysts.*

(c) *The state vector* d *is accessible to all players within the same market and is also discernible by analysts.*

Assumption 2.1 (a) highlights the conditional independence assumption, which specifies that ϵ_i is independent of ϵ_{-i} given d. This paper narrows its focus on settings characterized by incomplete information, as articulated in Assumption 2.1 (b). This suggests that the realized utility functions are private information to the respective players.

Given the private information ϵ and the observable state variable d, the utility of player i can be expressed in additively separable forms:

Assumption 2.2 (Additive Separability)**.**

$$u_i(a, d, \epsilon_i; \theta) = \Pi_i(a_i, a_{-i}, d; \theta) + \epsilon_i(a_i). \tag{2.2}$$

In the above equation, the payoff function $\Pi_i(a_i, a_{-i}, d; \theta)$ is additively separable with respect to private information. The payoff function depends on the agent's discrete choice a_i, the choices of other agents a_{-i}, and the state variable d. If no strategic interaction exists, implying that the flow utility is independent of other players' choices, the strategic model simplifies to a binary logit model. In this single-agent model, utility is solely a function of the individual's choice, relevant state variables, and private information.

In alignment with the standard discrete choice model, I normalize the utility associated with the outside option, which refers to the state of inactivity in the context of this game. This normalization is crucial for identifying payoffs when a player decides to be active. Formally, I assume the utility of the outside option to be zero.

Assumption 2.3 (Normalization of Outside Choice)**.**

$$\forall\ a_{-i} \in \mathscr{A}_{-i},\ \forall\ d,\ \Pi_i(a_i = 0, a_{-i}, d) = 0. \tag{2.3}$$

By setting the payoffs of inactivity to zero, I ensure that the payoff of being active is inherently measured relative to the inactive choice. Put simply, the payoff function reflects the differential between the utilities of being active and inactive. This assumption also highlights that the payoff derived from inactivity remains unaffected by the decisions of other players. Consequently, regardless of the actions taken by rival players a_{-i}, the payoff of the outside option remains at zero.

For a two-player game, I define the decision-choice rule as $a_i = \delta_i(d, \epsilon_i(a_i))$. Given that neither econometricians nor rivals observe ϵ_i, the decision rule is characterized by choice probabilities:

$$\sigma_i(a_i = 1|d) = \int \mathbf{1}\{\delta_i(d, \epsilon_i(a_i)) = a_i\} f(\epsilon_i) d\epsilon_i,$$

where $\mathbf{1}\{\delta(d, \epsilon_i(a_i)) = a_i\}$ stands as the indicator function that takes the value of 1 if player i chooses action 1, and 0 otherwise. It is pivotal to note that due to the private nature of information, the decision rules, specifically $\delta_i(d, \epsilon_i(a_i))$, remain independent of the private information of her rivals, ϵ_{-i}.

I also assume a simultaneous game setup wherein players make decisions simultaneously without observing the choices of their counterparts. In line with the discrete game literature, I incorporate the notion of belief information, ensuring that players hold correct beliefs about their rivals' choice probabilities. This view aligns with the Bayesian Nash equilibrium, where agents' beliefs about their rivals correspond with the actual conditional choice probabilities. However, this assumption might be strong in real-world scenarios. Xie (2022) proposed a more flexible approach by including unrestricted unknown functions in their model, suggesting an avenue for future research.

Assumption 2.4 (Correct Beliefs)**.**

In conjunction with the incomplete information structure, players form accurate beliefs about their rivals' choices.

Given private information assumptions 2.1 and the correct beliefs as per assumption 2.4, the choice-specific value function can be articulated in terms of the expected payoffs linked to rivals' choice probabilities, denoted as σ_{-i}.

$$\Pi_i(a_i = 1, d; \theta) = \sum_{a_{-i} \in A_{-i}} \sigma_{-i}(a_{-i}|d)\pi_i(a_i = 1, a_{-i}, d; \theta) \text{ for all } i = 1, ...n. \tag{2.4}$$

$$\text{where } \sigma_{-i}(a_{-i}|d) = \prod_{s \neq i} \sigma_s(a_s|d)$$

The given expression delineates the deterministic components of the expected payoff for player i, contingent on their chosen action a_i and the distribution of choice probabilities of their rivals, denoted by σ_{-i}. The extent of interaction stemming from rivals becomes particularly pronounced here: the utility that player i derives from opting to be active is intrinsically tied to the choices of her adversaries, a_{-i}.

Building on this, the optimal decisions of player i are encapsulated by the following:

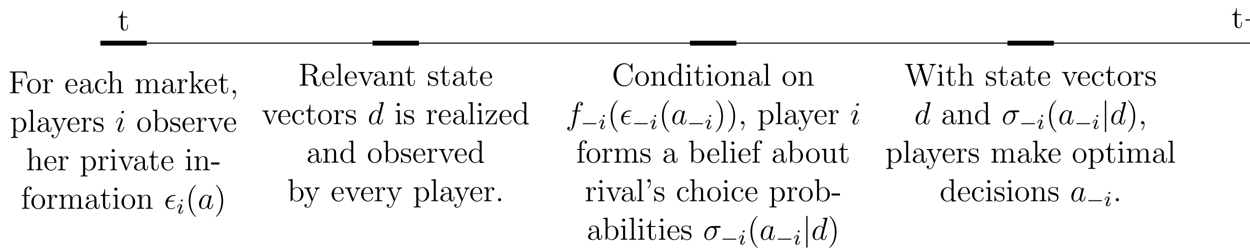
$$\sigma_i^*(a_i = 1|d) = \Pr\left[\underbrace{\Pi_i(a_i = 1, d)}_{=\sum_{a_{-i} \in A_{-i}} \sigma_{-i}(a_{-i}|d)\pi_i(a_i=1, a_{-i}, d)} + \epsilon_i(1) \geq \underbrace{\Pi_i(a_i = 0, d)}_{=0} + \epsilon_i(0)\right]$$

$$\sigma_{-i}^*(a_{-i} = 1|d) = \Pr\left[\underbrace{\Pi_{-i}(a_{-i} = 1, d)}_{=\sum_{a_{-i'} \neq -i \in A_{-i'}} \sigma_{-i'}(a_{-i'}|d)\pi_{-i}(a_{-i}=1, a_i, d)} + \epsilon_{-i}(1) \geq \underbrace{\Pi_{-i}(a_{-i} = 0, d)}_{=0} + \epsilon_{-i}(0)\right].$$

Agents choose to be active if and only if the sum of deterministic expected payoff and stochastic error components associated with being active is greater than the outside option and associated private information.

To provide a comprehensive overview, I've mapped out the sequence of events in the game, which is illustrated in Figure 2.1. At each time point t, in every market, players initially receive their private insights, symbolized by $\epsilon_i(a_i)$. Subsequently, all participants become aware of the relevant state vectors, d. Based on knowledge about their rivals' private choices, represented by $f_{-i}(a_{-i})$, each player i anticipates how their competitors might act, as signified by the choice probability $\sigma_{-i}(a_{-i}|d)$. Having gathered all this information, players then finalize their decisions, denoted a_i, and the game progresses to the succeeding period, $t+1$.

Figure 2.1: Timing of the Game



I now describe the mapping from choice-specific value functions to equilibrium choice probabilities. Under assumptions about correct belief, Type 1 Extreme Value distributions over private information, and normalized outside payoff, I can express equilibrium choice probabilities of choosing to be active as a system of equations. Taking a two-player game as an example, the choice probabilities, from both the econometrician's and the rival's perspectives, are articulated as:

$$\begin{aligned}\sigma_i(a_i=1|d) &= \Psi_i(\Pi_i(a_i,d)) := \frac{\exp(\pi_i(a_i=1,d))}{1+\exp(\pi_i(a_i=1,d))}\\ \sigma_i(a_{-i}=1|d) &= \Psi_{-i}(\Pi_{-i}(a_{-i},d)) := \frac{\exp(\pi_{-i}(a_{-i}=1,d))}{1+\exp(\pi_i(a_{-i}=1,d))}\end{aligned} \tag{2.5}$$

where equilibrium functions Ψ map the choice-specific value function into choice probabilities. By the correct belief assumption 2.4, the probability of being active is the equilibrium probability in that she makes her best responses after observing the state variable, which is consistent with Bayesian Nash equilibrium (BNE).

Building upon this foundation, I expand the scope to encompass games with n players. In this context, after fixing state variable d, $\sigma_i(a_i|d)$ will be the solution to the system of n equations:

$$\begin{aligned}\sigma_1(a_1=1|d) &= \frac{\exp\Big(\sum_{a_{-1}\in A_{-1}}\sigma_{-1}(a_{-1}|d)\pi_i(a_1=1,a_{-1},d)\Big)}{1+\exp\Big(\sum_{a_{-1}\in A_{-1}}\sigma_{-1}(a_{-1}|d)\pi_i(a_1=1,a_{-1},d)\Big)} \\ \sigma_2(a_2=1|d) &= \frac{\exp\Big(\sum_{a_{-2}\in A_{-2}}\sigma_{-2}(a_{-2}|d)\pi_i(a_2=1,a_{-2},d)\Big)}{1+\exp\Big(\sum_{a_{-2}\in A_{-2}}\sigma_{-2}(a_{-2}|d)\pi_i(a_2=1,a_{-2},d)\Big)}\end{aligned} \tag{2.6}$$

$$\vdots$$

$$\sigma_N(a_N = 1|d) = \frac{\exp\Big(\sum_{a_{-N}\in A_{-N}} \sigma_{-N}(a_{-N}|d)\pi_i(a_N = 1, a_{-N}, d)\Big)}{1 + \exp\Big(\sum_{a_{-N}\in A_{-N}} \sigma_{-N}(a_{-N}|d)\pi_i(a_N = 1, a_{-N}, d)\Big)}$$

I highlight the system of equations in 2.6 several issues. First, I presume the existence of a solution to equation (2.6), following McKelvey and Palfrey (1995)'s standard Brouwer's fixed-point argument. Secondly, the estimation of the equilibrium choice probabilities σ_i is complicated by the dependence of the rivals' choice probabilities σ_{-i} on the player i's choice probabilities.

There are primarily two approaches to address the aforementioned challenge. Building upon the nested fixed point (NXFP) method introduced by Rust (1987), Aguirregabiria and Mira (2002) devised an iterative algorithm tailored for dynamic games, which can be seamlessly adapted for static games. Specifically, within the inner loop, the algorithm iterates to the fixed point in 2.6, delineating the relationship between equilibrium choices, σ_i, and equilibrium beliefs for each player $i = 1, \ldots, N$. Subsequently, the outer loop employs each candidate parameter vector to compute a pseudo-likelihood, echoing the conventional maximum likelihood approach inherent to logistic regression. This iterative mechanism persists until convergence is attained. The NXFP method, sometimes referred to as the nested pseudo-likelihood approach, includes two primary limitations: the computational intensity arising from the dual-layered iteration and the assumption of a unique equilibrium in the model, which precludes the possibility of multiple equilibria[6].

The second approach, pioneered by Hotz and Miller (1993) and Bajari et al. (2010b) employs a two-step method. This method is computationally light and uses weaker assumptions about multiple equilibria compared to the NXFP algorithm. [7] In the first stage, I non-parametrically estimate conditional expectation $\sigma_{-i} = E[a_{-i} = 1|d]$ from observed choices and market characteristics d.[8] In the second stage, the econometrician estimates a single-agent random utility model. This model incorporates both market characteristics, d,

[6]For every prospective parameter vector, the algorithm necessitates the determination of a fixed point for equilibrium choices.

[7]For an in-depth discussion and comparison of these methodologies, I direct readers to Ellickson and Misra (2011). It's essential to note that I am not advocating for the superiority of the two-step methods over NXFP. My perspective stems from the ease with which one can integrate the findings of Newey (1994) and Chernozhukov et al. (2022). Given that the two-step approaches align with the classical semi-parametric estimation framework, it's feasible to apply the properties detailed in Chernozhukov et al. (2022). The incorporation of high-dimensional covariates based on NXFP methods is outside of the scope of this paper. For a relevant perspective, consider Dearing and Blevins (2019) and their exposition on zero Jacobian properties within the context of Efficient Pseudo-Likelihood.

[8]A formal introduction to the econometric principles underpinning the two-step methods will be presented in the subsequent section.

and the equilibrium beliefs, σ_{-i}, obtained from the first stage.

Given that multiple equilibria are prevalent in models with discrete games in the literature, I introduce an assumption regarding the selection of an equilibrium from the set of potential equilibria.

Assumption 2.5 (Equilibrium Selection)**.**

The data are generated by a single equilibrium from the set of possible multiple equilibria and observed equilibrium does not switch over different markets.

This assumption is relatively weaker compared to the uniqueness assumption, as it permits the existence of multiple equilibria in the model. As long as the equilibrium played in the data remains consistent across different markets or time periods, the initial stage of estimation accurately retrieves the choice probabilities of the underlying choice-specific value functions. Consequently, even if the obtained parameters might suggest other equilibria not played in the data, the estimates in the second stage remain consistent. Notably, this assumption is widely employed in two-stage estimation approaches, encompassing both static games (Bajari et al. 2010b, Ellickson and Misra 2011), and dynamic games (Aguirregabiria and Mira 2007, Bajari et al. 2007, Pesendorfer and Schmidt-Dengler 2008).

When coupled with the equilibrium selection assumption, the two-stage methods obviate the need for iterative model solving, effectively addressing the challenge posed by multiple equilibria in the estimation process. Additionally, researchers can derive a set of structural parameters without the need for repeated model solving, leading to a substantial reduction in computation time.

2.1 Identification

This section focuses on reviewing the identification results established in the literature, specifically in the works of Bajari et al. (2010b) and Bajari et al. (2010a). The purpose of revisiting Bajari et al. (2010b) is to highlight that the incorporation of high-dimensional state variables denoted as d does not alter the identification outcomes. Thus, the arguments developed in Bajari et al. (2010b) remain applicable even when dealing with high-dimensional covariates. To enhance readability and comprehension of the recovery process for underlying structural parameters θ, I present a restatement of the identification problems.

Definition 1 (Identification)**.** *Deterministic payoff components $\pi(a_i, a_{-i}, d)$ are identified if different deterministic payoff components $\sigma_i(a_i = 1|d) \neq \tilde{\sigma}_i(a_i = 1|d)$ yield alternative equilibrium probabilities $\pi(a_i, a_{-i}, s) \neq \tilde{\pi}(a_i, a_{-i}, d)$.*

The identification condition requires that different payoffs should generate different equilibrium choice probabilities. A necessary condition implies that without further assumptions about exclusion restrictions, the identification of the underlying model cannot be achieved. Manski (1993) called this issue a reflection problem associated with social interaction. To further illustrate this issue, consider the following illustrative examples featuring two players, denoted as $(i = 1, 2)$, engaging in binary choices. Their respective choice-specific value functions can be expressed as:

$$\begin{aligned}
&\underbrace{\Pi_1(a_1 = 1|d)}_{\text{unknown}} = \sigma_2(a_2 = 1|d)\underbrace{\pi_1(a_1 = 1, a_2 = 1|d)}_{\text{unknown}} + (1 - \sigma_2(a_2 = 1|d))\underbrace{\pi_1(a_1 = 1, a_2 = 0|d)}_{\text{unknown}} \\
&\underbrace{\Pi_1(a_1 = 0|d)}_{\text{known}} = 0 \\
&\underbrace{\Pi_2(a_2 = 1|d)}_{\text{unknown}} = \sigma_1(a_1 = 1|d)\underbrace{\pi_2(a_1 = 1, a_2 = 1|d)}_{\text{unknown}} + (1 - \sigma_1(a_1 = 1|d))\underbrace{\pi_1(a_1 = 0, a_2 = 1|d)}_{\text{unknown}} \\
&\underbrace{\Pi_2(a_2 = 0|d)}_{\text{known}} = 0
\end{aligned} \tag{2.7}$$

After fixing d, the left-hand side of equation (2.7) comprises two unknown components: the deterministic utilities $\Pi_1(a_1 = 1, d)$ and $\Pi_2(a_2 = 1, d)$. In accordance with the assumption 2.3, the expected payoff of remaining inactive, $\Pi_1(a_1 = 0, d)$ and $\Pi_2(a_2 = 0, d)$, is known to the econometrician. Conversely, on the right-hand side of the equation, there exist four unknowns: $\pi_1(a_1 = 1, a_2 = 1, d)$, $\pi_1(a_1 = 1, a_2 = 0, d)$, $\pi_2(a_1 = 1, a_2 = 1, d)$, and $\pi_2(a_1 = 0, a_2 = 1, d)$. This results in an under-identified scenario.[9]

The utilization of exclusion restrictions is a common strategy for disentangling the system of equations in (2.7) to satisfy identification condition.[10] The exclusion restriction requires that the pertinent state variable d can be split into two components: one that is universal across all players within the same market, referred to as d_0, and player-specific shocks denoted as d_i for each player $i = 1, ..., N$. Notably, player-specific shocks do not directly impact the payoffs of player $-i$, but they do influence the rival's payoffs indirectly through their effects on the rival's endogenous choices.

Assumption 2.6 (Exclusion Restriction)**.**

$$\pi_i(a_i, a_{-i}, d) = \pi(a_i, a_{-i}, d_0, d_i).$$

[9]The recovery of choice probabilities σ_i relies on first-stage reduced form choice probabilities.

[10]For more comprehensive discussions, see Bajari et al. (2010b), Bajari et al. (2010b).

Given the imposition of exclusion restrictions, let $d = (d_0, d_1, d_2)$, where d_0 represents common state variables for players 1 and 2 in the same markets, d_1 is player 1's specific state variable, and d_2 is player 2's specific state variable. When d_0 is held constant, it can be omitted for simpler notation. For the exposition, I further assume that each shifter takes binary values: 'H' denotes High, and 'L' denotes Low. To streamline the discussion, let $\Pi_i(a_i = 1|d_1 = H, d_2 = H) := \Pi_i(a_i = 1|H, H)$ and $\sigma_i(a_i = 1, a_{-i} = 1|d_1 = H, d_2 = H) := \sigma_i(a_i = 1, a_{-i} = 1|H, H)$ for brevity.

$$\begin{aligned}
\Pi_1(a_1 = 1|H, H) &= \sigma_2(a_2 = 1|H, H) \underbrace{\pi_1(a_1 = 1, a_2 = 1|d_1 = H, d_2 = H)}_{=\pi_1(a_1=1,a_2=1|d_1=H)} \\
&\quad + (1 - \sigma_2(a_2 = 1|H, H)) \underbrace{\pi_1(a_1 = 1, a_2 = 0|d_1 = H, d_2 = H)}_{=\pi_1(a_1=1,a_2=0|d_1=H)} \\
\Pi_1(a_1 = 1|H, L) &= \sigma_2(a_2 = 1|H, L) \underbrace{\pi_1(a_1 = 1, a_2 = 1|H, L)}_{=\pi_1(a_1=1,a_2=1|d_1=H)} \\
&\quad + (1 - \sigma_2(a_2 = 1|H, L)) \underbrace{\pi_1(a_1 = 1, a_2 = 0|H, L)}_{=\pi_1(a_1=1,a_2=0|d_1=1)} \\
\Pi_1(a_1 = 1|L, H) &= \sigma_2(a_2 = 1|L, H) \underbrace{\pi_1(a_1 = 1, a_2 = 1|L, H)}_{=\pi_1(a_1=1,a_2=1|d_1=L)} \\
&\quad + (1 - \sigma_2(a_2 = 1|L, H)) \underbrace{\pi_1(a_1 = 1, a_2 = 0|L, H)}_{=\pi_1(a_1=1,a_2=0|d_1=L)} \\
\Pi_1(a_1 = 1|L, L) &= \sigma_2(a_2 = 1|L, L)\pi_1(a_1 = 1, a_2 = 1|L, L) \\
&\quad + (1 - \sigma_2(a_2 = 1|L, L)) \underbrace{\pi_1(a_1 = 1, a_2 = 0|L, L)}_{=\pi_1(a_1=1,a_2=0|d_1=L)} \qquad (2.8)
\end{aligned}$$

The exclusion restriction implies that $\pi_1(a_1 = 1, a_2 = 1|d_1 = H, d_2 = H) = \pi_1(a_1 = 1, a_2 = 1|d_1 = H, d_2 = L)$, which leads to the following equation:

$$\begin{aligned}
\Pi_1(a_1 = 1|H, H) &= \sigma_2(a_2 = 1|H, H)\pi_1(a_1 = 1, a_2 = 1|d_1 = H) \\
&\quad + (1 - \sigma_2(a_2 = 1|H, H))\pi_1(a_1 = 1, a_2 = 0|d_1 = H) \\
\Pi_1(a_1 = 1|H, L) &= \sigma_2(a_2 = 1|H, L)\pi_1(a_1 = 1, a_2 = 1|d_1 = H) \\
&\quad + (1 - \sigma_2(a_2 = 1|H, L))\pi_1(a_1 = 1, a_2 = 0|d_1 = H) \\
\Pi_1(a_1 = 1|L, H) &= \sigma_2(a_2 = 1|L, H)\pi_1(a_1 = 1, a_2 = 1|d_1 = L) \\
&\quad + (1 - \sigma_2(a_2 = 1|H, L))\pi_1(a_1 = 1, a_2 = 0|d_1 = L) \\
\Pi_1(a_1 = 1|L, L) &= \sigma_2(a_2 = 1|L, L)\pi_1(a_1 = 1, a_2 = 1|d_1 = L) \\
&\quad + (1 - \sigma_2(a_2 = 1|L, L))\pi_1(a_1 = 1, a_2 = 0|d_1 = L) \qquad (2.9)
\end{aligned}$$

The left-hand side of the system of equations in 2.9 involves $dim(d_1) \times dim(d_2) = 2 \times 2 = 4$ unknowns. In contrast, the right-hand side of the equation encompasses four unknowns $(\pi_1(a_1 = 1, a_2 = 1|d_1 = H), \pi_1(a_1 = 1, a_2 = 0|d_1 = H), \pi_1(a_1 = 1, a_2 = 1|d_1 = L), \pi_1(a_1 = 1, a_2 = 0|d_1 = L))$. This implies that the equation 2.9 is identified. A similar argument can be applied to demonstrate the identification of payoffs for player 2 based on equation 2.10.

$$
\begin{aligned}
\Pi_2(a_1 = 1|H, H) &= \sigma_1(a_1 = 1|H, H)\pi_2(a_1 = 1, a_2 = 1|d_2 = H) \\
&\quad + (1 - \sigma_1(a_1 = 1|H, H))\pi_2(a_1 = 1, a_2 = 0|d_2 = H) \\
\Pi_2(a_1 = 1|H, L) &= \sigma_1(a_1 = 1|H, L)\pi_2(a_1 = 1, a_2 = 1|d_2 = H) \\
&\quad + (1 - \sigma_1(a_1 = 1|H, L))\pi_2(a_1 = 1, a_2 = 0|d_2 = H) \\
\Pi_2(a_1 = 1|L, H) &= \sigma_1(a_1 = 1|L, H)\pi_2(a_1 = 1, a_2 = 1|d_2 = L) \\
&\quad + (1 - \sigma_1(a_1 = 1|H, L))\pi_2(a_1 = 1, a_2 = 0|d_2 = L) \\
\Pi_2(a_1 = 1|L, L) &= \sigma_1(a_1 = 1|L, L)\pi_2(a_1 = 1, a_2 = 1|d_2 = L) \\
&\quad + (1 - \sigma_1(a_2 = 1|L, L))\pi_2(a_1 = 1, a_2 = 0|d_2 = L) \qquad (2.10)
\end{aligned}
$$

More generally, the choice-specific value functions as expressed in equation (2.4) could lead to the following expression:

$$
\underbrace{\Pi_i(a_i = 1, d)}_{\text{unknown}} = \sum_{a_{-i} \in A_{-i}} \underbrace{\sigma_{-i}(a_{-i}|d)}_{\text{known}} \underbrace{\pi_i(a_i = 1, a_{-i}, d)}_{\text{unknown}} \text{ for all } i = 1, ..., N. \qquad (2.11)
$$

This equation implies that, once d_0 is fixed, the left-hand side of the equations contains N unknowns $(\Pi_1, .., \Pi_N)$, whereas the right-hand side equations encompass $N \times 2^{N-1}$ unknowns $(\pi_i(a_i = 1, a_{-i}, d))$.[11] Consequently, without the introduction of exclusion restrictions, the system of equations in (2.11) cannot be identified. Next, armed with the assumption of exclusion restrictions as given in Assumption 2.6, I can reformulate the choice-specific value function as follows:

$$
\underbrace{\Pi_i(a_i = 1, d_i, d_{-i})}_{\text{unknown}} = \sum_{a_{-i} \in A_{-i}} \underbrace{\sigma_{-i}(a_{-i}|d_i, d_{-i})}_{\text{known}} \underbrace{\pi_i(a_i = 1, a_{-i}, d_i)}_{\text{unknown}} \text{ for all } i = 1, ..., N. \qquad (2.12)
$$

Evidently, the number of unknowns (free parameters) on the left-hand side has reduced from $\pi_i(a_i = 1, a_{-i}|d_i, d_{-i})$ to $\pi_i(a_i = 1, a_{-i}, d_i)$. This induces more variations on the right-hand side than the number of unknowns on the left-hand side flow utilities $\Pi_i(a_i = 1, d_i, d_{-i})$,

[11] The challenge posed by the curse of dimensions is evident, as players are required to formulate beliefs about all possible choices for their rivals.

which is the over-identified case.[12] It follows the necessary condition that the supports of beliefs $\sigma_{-i}(a_{-i} = 1|d_i, d_{-i})$ need to be sufficiently rich with 2^{N-1} points. It is evident that the incorporation of high-dimensional common market characteristics d_0 does not impact the identification conditions, as long as the rank conditions for d_{-i} given d_i are satisfied. The crucial determinants of identification are the number of players, choices, and variations $d_{-i}|d_i$ involved in the system, rather than the dimensionality of the common market characteristics d_0.

Remark 1 (Identification)**.**

1. *Suppose that Assumption 2.3 and Assumption 2.6 are satisfied. As long as there are 2^{N-1} points in the support of conditional distribution $d_{-i}|d_i$ related to $\sigma_{-i}(a_{-i}|d_i, d_{-i})$, then the necessary condition holds.*

2. *Allowing high dimensional market characteristics d_0 does not change necessary conditions for identification.*

A common example of exclusion restrictions is player-specific productivity shocks (Ericson and Pakes (1995)). Other instances encompass factors like the distance of a store to its distribution center, as explored in studies such as (Holmes, 2011) and (Jia, 2008). The underlying concept behind imposing exclusion restrictions is as follows: by imposing a restriction on the distance to the distribution center, the distance between player $-i$ and the distribution center of player $-i$, denoted as d_{-i}, will directly impact player $-i$'s entry probability. Conversely, player i's entry probability is indirectly influenced by d_{-i} through the choices made by the rival. The variation in the distance between player $-i$ and the distribution center of player $-i$ provides more equations in comparison to the number of unknowns on the left-hand side of the equation (2.12).

2.2 Two Step Estimation

The two-step method described here is an approach widely used for estimating structural parameters in static games with discrete choices. This method involves two key steps: in the first step, nuisance parameters like the choice probabilities of rivals are estimated through techniques like non-parametric methods, machine learning methods, or simple conditional logit models under parametric assumptions. In the second step, the estimated nuisance parameters are leveraged to formulate a method of moment conditions, upon which the Generalized Method of Moments (GMM) is employed to estimate the parameters of interest.

[12]In practice, testing for over-identification can be performed as suggested in Bajari et al. (2013).

The benefit of this approach lies in its flexibility and computational efficiency, facilitating the incorporation of high-dimensional market characteristics and the utilization of contemporary machine-learning techniques.

I additionally assume the linearity of flow payoffs, another common assumption in the empirical literature. The advantage of this assumption lies in the property that the choice-specific value function $\Pi_i(a_i = 1, d; \theta)$ will have a linear dependence on the expected payoffs. For instance, this takes the functional form:

$$\Pi_i(a_i = 1, d; \theta) = \sum_{a_{-i} \in A_{-i}} \sigma_{-i}(a_{-i}|d)\theta_\gamma + d_i'\beta_1 + d_0'\beta_2 \tag{2.13}$$

Here, θ_γ represents the interaction effects from rivals, β_1 shows the effect of player-specific productivity shock for player i, and β_2 encompasses the effect of common market characteristics. I denote $\beta = (\beta_1, \beta_2)$.

Step 1: Estimation of nuisance parameters

An analyst observes data on choices $A = \{0, 1\}^N$ for all N players and relevant state variables d. In the first stage, one can construct and estimate the conditional choice probability of being active for each player $-i$. Formally, the first stage reduced form choice probability can be expressed as:

$$\gamma_{-i} = E[a_{-it}|d_{-i}, d_i, d_0] \text{ for all } -i = 1, ..., N.$$

Non-parametric methods such as kernel or series estimation, or conditional logit can be employed to estimate the nuisance parameters in the first stage. It is important to note that this requires a unique equilibrium in the data so that the first stage estimates $\hat{\gamma}_{-i}$ can serve as consistent estimates of σ_{-i}.

Step 2: Recovering the Structural Parameters

Given the correctly specified first stage $\hat{\gamma}_{-i}$, the next step is to recover the underlying structural parameters of interest θ and β from equation 2.13. To accomplish this, I follow Bajari et al. (2010b)'s semi-parametric models. Coupled with the Type 1 Extreme Value Distribution, an econometrician can construct moment condition using a properly specified set of variables z_i:

$$\arg\min_{\theta,\beta} \; m(w_i; \theta, \beta) = z_i(a_i - \sigma(\hat{\gamma}_{-i}, \beta)). \tag{2.14}$$

With low-dimensional vector d_0, all the parameters in the model, the interaction effect θ_γ and covariate effects β, can be recovered with $\sqrt{N}$-consistency and asymptotic normality as established in Bajari et al. (2010b) and Newey and McFadden (1994). Now I consider the setting where common market characteristics d_0 encompass high-dimensional covariates, that is, the dimension of d_0 being comparable to or potentially larger than the sample size n. The goal is to develop an inference for the parameter of interest θ in the presence of high-dimensional nuisance parameters η. For illustration purposes, I set the interaction effect θ_γ as the main parameter of interest and parameters associated with d_0, (γ_{-i}, β), as nuisance parameters. However, I incorporate a subset of β as the parameter of interest in the empirical application and the methodology in Section 3 can be trivially extended to this case.

In this high-dimensional setting, the conventional GMM method becomes infeasible, so I could implement machine learning methods with regularization to enable estimation. For example, I could employ a Logistic Lasso estimator to estimate first-stage choice probability $\hat{\gamma}_{-i}$. This imposes the regularization in that it requires the sparsity assumption such as

$$\frac{s^2 \log^2(dim(d_0) \vee N)}{N} \to 0$$

to enable estimation with high-dimensional covariates, where $dim(d_0)$ refers to the dimension of d_0.

However, when machine learning estimates are used to estimate nuisance parameter η, the parameters of interest θ do not necessarily achieve $\sqrt{N}$-consistency, as documented in Chernozhukov et al. (2022). This is because using the estimator $\hat{\eta}$ to estimate $\hat{\theta}$ incurs the first-order bias, but machine learning estimators usually exhibit slower convergence rate than $\sqrt{N}$. Formally, this can be examined by looking at the directional (Gateaux) derivative with respect to the nuisance parameters η being non-zero:

$$\begin{gathered}
\partial_{\gamma_{-i}} E[m(w_i;\theta,\eta)][\gamma_{-i} - \gamma_{-i0}] = E\left[z_i \cdot \Lambda' \cdot \left(-\theta_\gamma \sum_{a_{-i}} \prod_{s \neq i,i} (1-\gamma_s)\right) \cdot (\gamma_{-i} - \gamma_{-i0})\right] \neq 0 \\
\text{for } -i = 1, \ldots, i-1, i+1, \ldots, n. \\
\partial_\beta E[m(w_i;\theta,\eta)][\beta - \beta_0] = E\left[(z_i \Lambda' \cdot \beta) \cdot (\beta - \beta_0)\right] \neq 0
\end{gathered}$$

where the directional derivative is defined in Section 3. This implies that the moment condition in (2.14) is not robust to the local misspecification of the first-stage nuisance parameters. Therefore, the first-order biases of the nuisance parameter would affect the target parameter, which are regularization and overfitting biases from using machine learning estimators.

To overcome these limitations, in the next section, I introduce so-called Neyman orthogonal moments that are insensitive to local misspecification from first-stage nuisance estimators η, based on the original moment in equation (2.14). These methods can provide robust and efficient estimations of parameters of interest in the presence of high-dimensional nuisance parameters. This method also achieves $\sqrt{N}$-consistency and asymptotic normality.

3 The DML-STATIC GAME ESTIMATOR

This section presents the DML-static game estimator, which is based on the works of Bajari et al. (2010b), Belloni et al. (2016), and Chernozhukov et al. (2022). Section 3.1 formally introduces the definition of Neyman orthogonal moment condition, while Section 3.2 provides new moment conditions derived from the GMM equation (2.14). In Section 3.3, the estimation procedure combined with the cross-fitting algorithm is outlined. The asymptotic properties of the proposed estimator are analyzed in Section 3.4, and the estimator is extended to cover games with multiple players in Section 3.5. Finally, Section 3.6 evaluates the finite sample properties of developed estimators through a series of Monte Carlo simulations.

3.1 The Definition of Neyman Orthogonal Moment Condition

This section presents the concept of Neyman orthogonal moment condition following the framework in Chernozhukov et al. (2018a). I introduce the definition in my context for clarity. Let $\theta \subset R^{dim(\theta)}$ be the structural parameters of interest and $\eta \in \mathcal{T}$ be the infinite-dimensional nuisance parameter where $\mathcal{T}$ is a convex subset of some normed vector space with norm denoted by $||\cdot||_{\mathcal{T}}$. Under true values θ_0 and η_0, the following moment function is assumed to satisfy:

$$E[\psi(W;\theta_0,\eta_0)] = 0. \tag{3.1}$$

Following Chernozhukov et al. (2018a) and Van der Vaart (2000) section 20.2, I define the directional (Gateaux) derivative map $D_\tau : \tilde{\mathcal{T}} \to R^{dim\theta}$ as

$$D_\tau[\eta-\eta_0] := \partial\tau\{E_P[\psi(W,\theta_0,\eta_0+\tau(\eta-\eta_0)]|\},\ \eta \in \mathcal{T}$$

for all $\tau \in [0,1)$ and I assume its existence. The derivative at $\tau = 0$ is denoted as

$$\partial_\eta E_P[\psi(W;\theta_0,\eta_0)][\eta-\eta_0] := D_0[\eta-\eta_0],\ \eta \in \mathcal{T} \tag{3.2}$$

for convenience.

Definition 2. *The moment function $\psi(W,\theta,\eta)$ obeys Neyman orthogonality condition at (θ_0,η_0) with respective to the nuisance parameter realization set $\mathcal{T}_N \subset \mathcal{T}$ if equation (3.1) holds and the Gateaux derivative $D_\tau[\eta-\eta_0]$ exists for all $\tau \in [0,1)$ and $\eta \in \mathcal{T}$, and the orthogonality condition holds, that is*

$$\partial_\eta E_P[\psi(W,\theta_0,\eta_0)][\eta-\eta_0] = 0 \text{ for all } \eta \in \mathcal{T}_N. \tag{3.3}$$

For the rest of the paper, I will refer to the moment function that satisfies the Neyman orthogonality condition as the orthogonal moment function.

3.2 Orthogonal Moment Condition for Static Game: Two Players Example

For illustrative purposes, the exposition initially focuses on two-player games, adhering to the assumptions (2.1) -(2.6). I construct the moment function that satisfies the orthogonality condition defined in Section 3.1 by adding the "bias correction terms" to the original moment function (2.14). This makes the new moment function insensitive to the first-stage bias from the nuisance parameter estimate $\hat{\gamma}$. Specifically, the bias correction term with respect to nuisance parameter β follows the optimal instrument approach in Belloni et al. (2016), and the bias correction term with respect to γ_{-i} follows the approach in Chernozhukov et al. (2022). Additionally, there is a new nuisance parameter μ_z generated in the process of constructing orthogonal moment function.

Let $f_i \equiv \sqrt{\sigma(\cdot)(1-\sigma(\cdot))}$ where σ denotes the choice probabilities from original moment function. The construction of the Neyman orthogonal moment function is based on the linear projection of z_i on $x_i = (d_i, d_0)$ with weighting f_i, similar to Belloni et al. (2016).

$$f_i z_i = f_i x_i' \mu + u_i, \;\; E[f_i x_i u_i] = 0 \tag{3.4}$$

Then, the orthogonal moment function is

$$\begin{aligned} \psi(w_i;\theta,\eta) &= m(w_i;\theta_0,\eta_0) + \phi(w_i;\theta_0,\alpha_0,\eta_0) \\ m(w_i;\theta_0,\eta_0) &= (z_i - x_i'\mu)\left[a_i - \sigma(\gamma_{-i},\theta_\gamma,\beta)\right] = \mu_z\left[a_i - \sigma(\gamma_{-i},\theta_\gamma,\beta)\right] \\ \phi(w_i;\theta_0,\alpha_0,\eta_0) &= -E[\mu_z\sigma(\cdot)(1-\sigma(\cdot))\theta_\gamma|d_i,d_0](a_{-i}-\gamma_{-i}) = \alpha_0(a_{-i}-\gamma_{-i}) \end{aligned} \tag{3.5}$$

where $\mu_z = (f_i z_i - f_i x_i'\mu)/f_i = z_i - x_i'\mu$ and $\alpha = E[\mu_z\sigma(\cdot)(1-\sigma(\cdot))\theta_\gamma|d_i,d_0]$. This moment function satisfies the orthogonality condition defined above.

Theorem 3.1. *The moment function (3.5) obeys the Neyman orthogonality condition.*

Theorem 3.1 states that the moment condition $E_P[\psi(w_i;\theta,\eta)] = 0$ identifies the true parameter and is insensitive to misspecification of η in the neighborhood of η_0. The proof of Theorem 3.1 can be found in the Appendix.

3.3 Estimation Procedure

I present the estimation procedure utilizing the cross-fitting algorithm proposed by Chernozhukov et al. (2018a) combined with the two-step estimation method in Bajari et al. (2010b).

Let K denote a positive integer and take a K-fold random partition $I_1, ..., I_K$ of observation indices $\{1, \ldots, N\}$. For simplicity, let each fold I_k have an equal size with $n = N/K$. Define the auxiliary sample $I_k^c = \{1, ..., N\}/I_k$ for each $k \in \{1, \ldots, K\}$.

Step 1. Estimation of nuisance parameters using ML

For each $k \in \{1, \ldots, K\}$, estimate the set of nuisance parameters $\hat{\eta} = (\hat{\gamma}_{-i}, \hat{\beta}, \hat{\mu}_z)$ only using observations not in the group k as

$$\hat{\eta}_k = \hat{\eta}\Big((W_i)_{i\in I_k^c}\Big).$$

For choice probabilities γ_{-i}, econometricians can use modern machine learners such as Logit Lasso, Random Forests Classifiers, or Neural Network Classifiers. For β, I use Logistic Lasso following Belloni et al. (2016). To learn μ, I use Lasso based on the equations 3.4. Note that the estimators of nuisance parameters are required to have convergence rates are faster than $N^{-1/4}$.

Step 2. Recovering structural parameters θ

Using the estimated nuisance parameter estimates $\hat{\eta}_k$, I evaluate the moment condition in equation 3.5 on the sample I_k. I obtain the final estimator $\hat{\theta}$ by aggregating the objective functions for each $k \in \{1, \ldots, K\}$. The formal estimation algorithm is summarized below.

Algorithm

1. Take a K-fold random partition $(I_k)_{k=1}^K$ with same size $n = N/K$. For each $k \in \{1, \ldots, K\}$, define I_k^c as the complement of I_k.

2. For each $k \in \{1, \ldots, K\}$, construct an ML estimator $\hat{\eta}_k = \hat{\eta}((W_i)_{i\in I_k^c})$.

(a) Obtain $\hat{\gamma}_{-ik}$ using ML Classifier of a_{-i} on d_{-i}, d_i and d_0.

(b) Obtain $\hat{\beta}_k$ using Logit Lasso estimator of a_i on $\hat{\gamma}_{-i}, d_i$ and d_0.

(c) Compute $\hat{\theta}_{\gamma k}$ from original moment function 2.14.

(d) Compute the conditional densities $\hat{f}_k$.

(e) Estimate $\hat{\mu}_{zk}$ from the Lasso estimator of $\hat{f}_k z_i$ on $\hat{f}_k d_0$.

(f) Collect $\hat{\eta}_k = (\hat{\gamma}_{-ik}, \hat{\beta}_k, \hat{\mu}_{zk})$.

3. Construct the estimator $\hat{\theta}_\gamma$ as

$$\hat{\theta}_\gamma \in \arg\min_\theta \frac{1}{K} \sum_{k=1}^{K} L_{n,k}(\theta_\gamma)$$

where $L_{n,k}(\theta_\gamma) = \dfrac{\{E_{n,k}[\psi(\theta_\gamma, \hat{\eta}_k)]\}^2}{E_{n,k}[\psi(\theta_\gamma, \hat{\eta}_k)^2]}$ and $E_{n,k}$ is the empirical expectation over I_k, that is, $E_{n,k}[\psi(w)] = n^{-1} \sum_{i \in I_k} \psi(w_i)$. The moment function used in the objective function is $\psi(w_i; \theta, \eta) = \mu_z \left[a_i - \Lambda(\gamma_{-i}\theta_\gamma, d_i\hat{\beta_{d_i}}, d_0\hat{\beta})\right] - \alpha[a_{-i} - \gamma_{-i}]$ where $\alpha = E[\mu_z \Lambda(\cdot)(1 - \Lambda(\cdot))\theta_\gamma | d_i, d_0]$.

3.4 Asymptotic Analysis

In this section, I provide an asymptotic theory for DML static games with two players. The analysis for multiple players can be provided by using the same arguments. I closely follow the assumptions and proofs in Chernozhukov et al. (2022).

Assumption 3.1 (Convergence Rates)**.**

For each $\ell = 1, ..., L$,

i) $||\hat{\gamma}_{-ih\ell} - \gamma_{-ih0}|| = O_p(n^{-1/4})$

ii) $||\hat{\beta}_\ell - \beta_0|| = O_p(n^{-1/4})$

iii) $||\hat{\mu}_\ell - \mu_0|| = O_p(n^{-1/4})$

Assumption 3.2 (Regularity Condition)**.**

i) $W_i = (A_i, D_{-i}, D_i, D_{0i})$ are bounded.

ii) M is twice differentiable with uniformly bounded derivatives bounded from zero.

where $M \equiv \dfrac{\partial m(w,\gamma,\beta;\theta)}{\partial\theta}$

iii) $E[\{Y_{-i} - \hat{\gamma}(D_{-i}, D_i, D_0)\}^2 | D_{-i}, D_i, D_0]$ *and* $\hat{\alpha}$ *are bounded.*

iv) $E[m(W,\gamma_0,\theta_0)^2] < \infty$ *and* $\int ||m(w,\hat{\gamma}_\ell,\theta_0) - m(w,\gamma_0,\theta_0)||^2 F_0(dw) \xrightarrow{p} 0$

Theorem 3.2. *Suppose that assumption 2.1-2.6, and assumption 3.2 holds. For* $V = M^{-1}E[\psi_0(W)\psi_0(W)']M^{-1}$, *the DML static game estimators constructed in orthogonal moment conditions in 3.5 obeys*

$$\sqrt{N}(\hat{\theta} - \theta_0) \to N(0, V).$$

Also, the variance estimator $\hat{V}$ *is consistent where*

$$\hat{V} = \Big(\frac{1}{K}\sum_k^K E[M]\Big)^{-1} \frac{1}{K}\sum_k^K E[\psi^2(w,\hat{\theta},\hat{\eta}_k)] \Big(\frac{1}{K}\sum_k^K E[M]\Big)^{-1}.$$

Theorem 3.2 establishes the asymptotic normality of the proposed DML estimator for static games. The theorem shows that the proposed DML method is $\sqrt{N}$-consistent and asymptotically normal. Additionally, even in the worst-case scenario, as long as the nuisance parameters converge at a rate faster than $N^{-1/4}$, many machine learning methods satisfy the convergence rates of the nuisance parameters specified in Assumption 3.1. For instance, conditions for Lasso/Logit Lasso are provided in Belloni et al. (2012), faster rates (shallow trees) for Random Forest in Syrgkanis and Zampetakis (2020), and faster rates based on critical radius in neural networks in Chernozhukov et al. (2021). The proof of Theorem 3.2 can be found in the Appendix.

3.5 Extension: Multiple Players

In this section, I extend the two-player game to a multiple-player game. Let S denote the total number of players in the market. A significant difference is that now players form beliefs about all the other players, resulting in a much larger number of possible combinations: $2^{N-1} - 1$. For clarity, I provide the original moment condition for multiple players, based on Bajari et al. (2010b):

$$m(w_i;\theta,\eta) = z_i\{a_i - \Lambda(\pi_i)\} \tag{3.6}$$

$$\text{where } \pi_i = \theta_\gamma \sum_{a_{-i}} \prod_{s\neq i}^S \gamma_s + \beta_1 d_i + \beta_2 d_0.$$

As the number of choice probabilities to estimate in the first step increases, the number of nuisance parameters increases as well. Therefore, I need to develop a moment function that satisfies the orthogonality condition with respect to the new nuisance parameters γ_s. The corresponding moment function is

$$\begin{aligned}\psi(w_i;\theta,\eta) &= m(w_i;\theta_0,\eta_0) + \phi(w_i;\theta_0,\alpha_0,\eta_0) \\ m(w_i;\theta_0,\eta_0) &= (z_i - x_i'\mu) = \mu_z\left[a_i - \Lambda(\gamma_{-i},\theta_\gamma,\beta)\right] \\ \phi(w_i,\theta_0,\alpha_0,\eta_0) &= -\sum_{a_{-i}}\prod_{s\neq i,-i}\underbrace{\alpha_{-i}}_{=\mu_z\Lambda(\cdot)(1-\Lambda(\cdot))(1-\gamma_s)\theta_\gamma}[a_{-i}-\gamma_{-i}].\end{aligned} \tag{3.7}$$

The new moment functions introduced in this paper differ from the influence functions for discrete games proposed in the literature Bajari et al. (2010b) and Bajari et al. (2009). In particular, the bias correction terms of the new moment functions incorporate the rival's choice probability, as opposed to the player's own beliefs in the previous literature. This distinction arises as the moment conditions include nuisance parameters over the rivals, not the player's own choices. With an increase in the number of players, the complexity of beliefs over rivals also increases exponentially, leading to a corresponding exponential increase in the correction terms. In the following, this paper formally derives the near orthogonal moments for multiple players, building upon the results for two-player games in Theorem 3.1.

Theorem 3.3. *The moment function (3.7) obeys the Neyman orthogonality condition.*

The proof of Theorem 3.3 can be found in Appendix.

3.6 Monte Carlo Simulation

I conduct a series of Monte Carlo experiments to evaluate the finite sample properties of the proposed method. I design the experiments to be analogous to the static entry/exit model with incomplete information as in Bajari et al. (2010b). Then, I report the results to compare the performance of the debiased estimator to the plug-in estimator in the high-dimensional setting.

The static entry/exit model and equilibrium

I simplify the model described in Section 2 into two players, normalizing the payoff to be inactive to zero. The payoff of player i is a function of common market characteristics $d_0 = (d_1, d_2)$ for both players i and $-i$, rival's choice probabilities σ_{-i}, and a player-specific variable d_i, expressed as follows:

$$\pi_i(a_i = 1|\gamma_{-i}, d_i, d_0; \theta_\gamma, \delta, \beta) = \sigma_{-i}(a_{-i} = 1|d_i, d_{-i}, d_0)\theta_\gamma + \beta_1 d_i + d_0'\beta_2.$$

Under Type 1 Extreme Value distribution assumptions, conditional choice probability can be expressed in terms of relevant state variables and choice probabilities σ:

$$\begin{aligned} \sigma_i(a_i = 1|\sigma_{-i}, d_i, d_{-i}, d_0) &= \frac{\exp\left(\sigma_{-i}(a_{-i} = 1|d_i, d_{-i}, d_0)\theta_\gamma + \beta_1 d_i + d_0'\beta_2\right)}{1 + \exp\left(\sigma_{-i}(a_{-i} = 1|d_i, d_{-i}, d_0)\theta_\gamma + \beta_1 d_i + d_0'\beta_2\right)} \\ \sigma_{-i}(a_{-i} = 1|\sigma_i, d_i, d_{-i}, d_0) &= \frac{\exp(\sigma_i(a_i = 1|d_{-i}, d_i, d_0)\theta_\gamma + \beta_1 d_i + d_0'\beta_2)}{1 + \exp(\sigma_i(a_i = 1|d_{-i}, d_i, d_0)\theta_\gamma + \beta_1 d_i + d_0'\beta_2)}. \end{aligned} \tag{3.8}$$

Consistent with Bayesian Nash equilibrium, player i's best response function σ_i depends on her rival's best response (σ_{-i}), and vice versa.

I can solve equation (3.8) via a fixed-point algorithm as it contains two unknowns and two equations. The algorithm converges when the difference in the choice probabilities of being active between the $(k+1)$th and kth iterations is smaller than a predetermined tolerance level ϵ for both i and $-i$. Throughout the convergence process, I did not encounter issues related to multiple equilibria.

Data Generation

Using the best response function 3.8, I simulate market-firm level data for a decision-maker who lives for ten periods and makes decisions on whether to be active or inactive in each period. I set $\theta_\gamma = -1.5$, $\beta_1 = 2$.[13] For common market characteristics, I set $d_0'\beta_2 = d_1\beta_{21} + d_2\beta_{22}$ where $\beta_{21} = 0.8$ and $\beta_{22} = 1.4$. To achieve varied finite samples, I simulate datasets for a number of markets 50, 75, and 100 with two firms and allow ten periods of time, resulting in 1,000, 1,500, and 2,000 total observations, respectively. I denote the total observation size as N. I generate data from the model following these steps:

1. I independently draw common market characteristics for players in the same market $d_0 = (d_1, d_2)$ and player-specific shifter d_i, d_{-i} from a uniform distribution with mean zero and variance one.[14]

2. I draw probabilities from a uniform distribution on $[0, 1]$ and player i chooses to be active ($a_i = 1$) if the draw is less than or equal to the probability of being active,

[13] I follow the specifications of the parameter from Arcidiacono and Miller (2011)'s game with modifications to avoid multiple equilibria.

[14] For this task, I draw 20 grid points from standard normal distributions using Python *np.random.seed(0)*, and scaled by mean 0 and variance 1.

$\sigma_i(a_i = 1|d_i, d_{-i}, d_0)$.

When implementing the Plug-in estimator and Orthogonal estimator, I additionally include many covariates d_x with $dim(d_x) = p = 500$ in common market characteristics. d_x is drawn from a standard normal distribution further normalized to mean zero and variance one.

Estimation of the static entry/exit model

The first stage of estimation requires the estimation of nuisance parameters $\hat{\eta} = (\hat{\gamma}_{-i}, \hat{\beta})$. In the second stage, the structural parameter θ_γ is recovered. I compare the performance of three different estimators: Oracle estimator, Plug-in estimator, and Orthogonal estimator.

Oracle estimator: Estimator of $\hat{\theta}_\gamma$ based on Bajari et al. (2010b) using only d_0 as common market characteristic. This estimator assumes the knowledge of the true identity of common market characteristics. I use a logit estimator for first-stage conditional choice probabilities.

Plug-in estimator: Estimator of $\hat{\theta}_\gamma$ using (d_0, d_x) as common market characteristic and imposing regularization to Bajari et al. (2010b). I estimate first-stage conditional choice probabilities using the Logit Lasso estimator[15] and obtain $\hat{\theta}_\gamma$ adopting regularization to GMM estimation using 2.14.

Orthogonal estimator: Estimator of $\hat{\theta}_\gamma$ using (d_0, d_x) as common market characteristic and using orthogonal moment condition. I estimate first-stage conditional choice probabilities using the Logit Lasso estimator and use orthogonal moment condition in the second stage. I employ a cross-fitting algorithm with $K = 5$ as described in Section 3.3.

3.7 Simulation Results

Table 3.1 summarizes the Monte Carlo Simulation results. The histogram of the simulation result with $(N, p) = (2000, 500)$ is illustrated in Figure C.4. The true parameter value is -1.5. These true parameters are partially from the literature Arcidiacono and Miller (2011) and I do not find multiple equilibria issues and corner solution problems. The mean bias, percentage of bias relative to the true parameter value, 95% coverage probability, and RMSE of three estimators are reported respectively in columns 2-5, 6-9, and 10-13.

When using the Oracle estimator, the estimates are well-centered around the true value and show coverage probability close to 95%. This is as expected since the true model is known to the econometrician. The results using the Oracle estimator provide a benchmark when evaluating the performance of Plug-in and Orthogonal estimators.

[15] I use the penalty term recommended by Belloni et al. (2016), $\lambda = c\sqrt{n}\Phi^{-1}(1 - \gamma/\{2pn\})$ and *hdm* package in R.

When the Plug-in estimator is used, rival effects are severely biased upward due to regularization bias in machine learning estimators. Column 7 reports the biases in percentage terms and shows that the magnitude of biases is not negligible. Similarly, the coverage probability reported in column 8 is far below the nominal level of 95%, indicating that the Plug-in estimator has invalid inferential properties. The RMSE is also much larger compared to the Oracle estimator.

When using the orthogonal estimator, the estimates are centered around the true values, comparable to the result using the Oracle estimator and having a smaller bias compared to the Plug-in estimator. The coverage probability and RMSE show better performance than the Plug-in estimator but worse than the Oracle estimator.

Table 3.1: Simulation Results

	Oracle estimator				Plug-in estimator				Orthogonal estimator			
(N, p)	Mean bias	Bias(%)	CP	RMSE	Mean bias	Bias(%)	CP	RMSE	Mean bias	Bias(%)	CP	RMSE
(1000,500)	-1.595 (0.410)	-6.323	0.960	0.522	-1.013 (0.600)	32.5	0.736	1.025	-1.529 (0.615)	-1.901	0.880	0.936
(1500,500)	-1.562 (0.329)	-4.151	0.956	0.388	-0.913 (0.464)	39.143	0.668	0.872	-1.456 (0.472)	2.947	0.872	0.627
(2000,500)	-1.545 (0.289)	-3.007	0.958	0.307	-0.905 (0.402)	-39.691	0.608	0.775	-1.477 (0.407)	1.526	0.858	0.566

Notes: Mean and Standard Deviation for 500 simulations. Column (1) represents the simulation scenario specifying the number of observations (N) and the dimension of market characteristics (p). Columns (2)-(5) used the Oracle estimator and columns (6)-(9) used the Plug-in estimator, and Columns (10)-(13) used developed Orthogonal estimator. For each estimator, the mean bias, the percentage of bias, 95% coverage probability, and root mean square error (RMSE) are reported.

4 Data and Background

In this section, I provide the background of independent pharmacies, describe data sources, illustrate limited pharmacy accessibility, and provide related descriptive statistics.

4.1 Industry Background

Independent pharmacies have traditionally been characterized as independently owned-stores or single-owner establishments. Independent pharmacies offer a wide range of services, functioning as community hubs where individuals can get their prescriptions filled, seek advice on minor ailments, and purchase over-the-counter medications and other everyday items.

However, the landscape began to shift in 1970 when chain pharmacies, mass-merchandised, and supermarket-based pharmacies began to challenge the dominance of independent pharmacies. Walgreens, founded in 1901 in Chicago; CVS Pharmacy, founded in 1963; and Rite

Aid, founded in 1962, all embarked on expansion sprees by opening new stores or acquiring smaller chains. In the mass-merchandised pharmacy market, Walmart launched Walmart Pharmacy in 1978 in Rogers, Arkansas, and has since grown to over 5,000 stores nationwide, making it one of the largest pharmacy chains in the United States. Target opened its first pharmacy in 1996, in Minneapolis, Minnesota, and has since expanded to over 1,600 stores nationwide. In 2015, Target sold its pharmacy business to CVS Health, the second-largest pharmacy chain in the United States. Due to their bulk purchasing power and significant resources, chain, mass merchandise, and supermarket pharmacies often offer competitive prices. They do this by leveraging economies of scale and substantial bargaining power against health insurance companies. For example, Walmart offers a $4 generic prescription program.

By 1999, the market share in prescription sales for chain pharmacies reached 40.3%. Independent pharmacies trailed at 25.6%, with mass merchandisers at 10.1%, supermarket pharmacies at 11.00%, and mail orders at 13.0% [16]. Although the mail order market share steadily rose by 15% in 2008, by 2018, their market share had reverted to 13.7%. I abstracted away mail-order in the analysis because it takes a smaller portion of the market share.

In the 2000s, there was the continuing expansion of both merchandise-based pharmacies (e.g., Walmart, Sam's Club, Target) and supermarket-based pharmacies (Kroger, Publix), particularly after 2005, which made less room for independent pharmacies. By 2019, there were 22,773 chain pharmacies, 21,683 independent pharmacies, 8,427 supermarket-based pharmacies, and 8,597 mass merchant-based pharmacies [17].

4.2 Data

I combine data from multiple sources to construct the final dataset. In this dataset, geographic units, referred to as "towns," are defined as markets.[18] Each entry contains details about the pharmacy's market entry/exit decisions and observable characteristics of both the stores and the market.

My primary dataset is sourced from the Data Axle Historical Business Database, which chronicles the operations of business establishments, including pharmacies, in the United States from 1997 to 2021, updated annually. This dataset has been employed in recent studies such as Dearing and Blevins (2019), Koh (2023), and Lepoev (2023). Since Data Axle includes the addresses of each pharmacy store, I can assign these addresses to townships

[16]Source: https://www.kff.org/wp-content/uploads/2000/06/3019-prescription-drug-trends-a-chartbook.pdf

[17]Source: 2020 NATIONAL COMMUNITY PHARMACISTS ASSOCIATION DIGET

[18]I explain the reasoning behind defining the geographic market at the town level in the subsequent section.

using Python's Geopandas. Furthermore, the panel data structure enables me to define entry and exit every year.

Additionally, I obtain market-level data on demographic characteristics from the Census and the American Community Survey (ACS) at the township (county subdivision) level. This data offers rich market characteristics as well as consumer demographics. It allows me to study how market characteristics and consumer demographics affect independent pharmacies' decisions to enter or exit a market. I also obtain health-related characteristics from the Current Population Studies (CPS) and ZIP Code Business Patterns dataset. Appendix D provides more detailed information on data construction. Given that the nearest census data is available starting from the year 2000, my analysis covers the period from 2000 to 2019.[19].

Market Definition of Geographic Level

I define a geographic market based on townships (county subdivisions) that had pharmacies at any point between 1997 and 2021. I chose the township as the geographic unit for the following reasons:

1) As reflected in the survey results shown in Figures C.1 and C.2, consumers consider the location of a pharmacy to be one of the most important factors. Consumers generally prefer a pharmacy closer to their neighborhood, which aligns with the current market definition.[20]

2) My market definition follows earlier healthcare studies that used towns Schaumans and Verboven (2008) and zip codes Lepoev (2023).[21]

3) From an econometric analysis perspective, it is advantageous that the market-level characteristics (e.g., population) from the Census, which I will describe in the next section, align with the township-level geographic market. This means researchers can easily merge township-level market characteristics into the township-level dataset.

4) Market definition of town level could be suitable as I provide reduced-form evidence in section 5.2, which suggests that new entries of independent pharmacies outside of a township have a minimal effect on pharmacies within that township. The widely used isolated market assumption by Bresnahan and Reiss (1991) is likely valid in my settings.

[19]I have excluded data from early 2020 onwards from the analysis because of the onset of the pandemic, as the market equilibrium might differ significantly from the pre-pandemic period.

[20]While my focus is granular township level, one might be concerned about the possibility that consumers visit pharmacies while commuting to work. However, OFT (2003) reported that only 6% of patients visit their pharmacy during their commute, further confirming the local nature of competitive interactions.

[21]Admittedly, there are other ways to define a market, such as pre-specified regions like census tracts or cluster analysis (k-means clustering) as used in Ellickson and Misra (2008). However, I chose the township level because it is a pre-specified region that relatively follows the rectangular styles shown in the subsequent figure 4.3.

Final Sample

Given the focus on rural areas, for a township to be included in the dataset as a rural geographic market, it must: (i) ensure its geographic boundaries do not overlap with the urbanized areas defined by the Census, as described in Appendix D.1, (ii) have a population of more than 100 people, (iii) have had at least one pharmacy in operation between 1997 and 2001, (iv) not have had more than two independent pharmacies operating simultaneously between 2000 and 2019[22], (v) not have had more than seven chain pharmacies within a 15-mile radius.

The first two criteria ensure that the sample is limited to rural areas. Restrictions (iii) and (iv), which impose limits on the number of stores in a township, could potentially introduce an endogenous sample selection issue. However, these restrictions are necessary to maintain computational feasibility and to exclude townships that are close to urban clusters. Additionally, I control for outliers in (v) as 95.3% of my final samples include at most seven chain stores. This is because townships with more than seven chain pharmacies are likely to be fundamentally different from typical rural townships.

4.3 Low Accessibility in Pharmacy

In this section, I document recent trends in towns with limited accessibility to pharmacy. Here, consumers must travel several miles to obtain prescriptions. These areas are sometimes referred to as “limited access to pharmacy area” or “pharmacy deserts”. The term ”pharmacy desert” is inspired by the concept of a ”food desert”, an area where residents struggle to find healthy foods due to a lack of nearby supermarkets or affordable food stores. Similarly, a ”pharmacy desert” is an area without easy access to a pharmacy, making it difficult for residents to obtain their medications.

I describe areas with “limited access to pharmacy” as those townships without any pharmacies. This definition is based on the approach taken by Qato et al. (2014), who used census tracts as a geographical reference, which are similar to townships. Elderly individuals are especially vulnerable to these challenges due to mobility issues and high transportation costs. As a result, my focus is on the Midwest rural areas, where the aging population is a growing concern, as highlighted by (Mather et al., 2015).

Limited accessibility to pharmacy can lead to adverse health outcomes, such as increased emergency department visits and hospitalizations.The consequences of patients not taking their medications as prescribed, known as “non-adherence”, have been studied in Di Novi

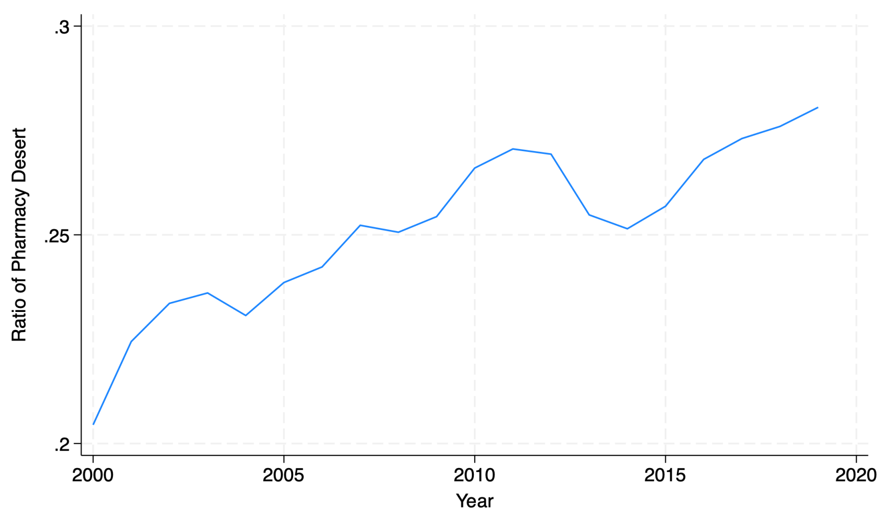
[22] In the rural townships of my sample, 99.30% of samples have at most two pharmacies operating concurrently.

et al. (2020). For a comprehensive discussion on the widespread issues of limited pharmacy access and the resulting negative health impacts, see Di Novi et al. (2020). [23]

Trends in Limited Accessibility in Pharmacy

Figure 4.1 illustrates the escalating trends in limited access to pharmacy within the Midwest using my final sample of 802 townships. The percentage of towns with limited access to pharmacy stores has surged from 20.44% to 28% (a 37% increase). In Figure C.3, I present alternative definitions of limited pharmacy access, considering population weights and a 5-mile distance. These alternative specifications reveal qualitatively similar trends.

Figure 4.1: Trends in Limited Pharmacy Accessbility



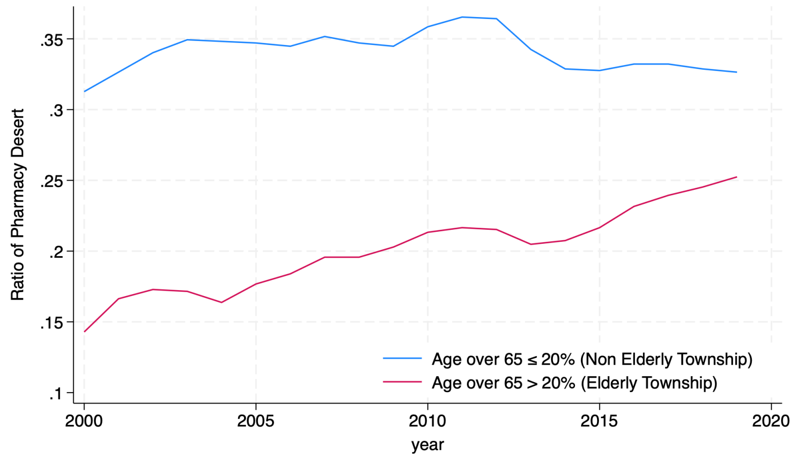
Notes: The data points represent a three-year moving average based on the limited pharmacy access, taken from a final sample of 802 townships. A township is designated as a limited pharmacy access (indicator value of 1) if there are no independent or chain pharmacies within its boundaries.

I also find heterogeneity in limited access to pharmacy by elderly population share. Figure 4.2 illustrates that non-elderly townships maintained relatively stable figures, while elderly townships-those where more than 20% of the population is aged 65 or older a marked increase in limited pharmacy access, rising from 14.29% in 2000 to 25.44% in 2019. Given that elderly townships are particularly vulnerable to pharmacy access challenges, this raises public concern over easy prescription access in elderly towns.

To shed light on these trends, Table 4.1 provides insights into the types of pharmacies that have exited townships, leaving them without any available pharmacies. The table re-

[23]Non-adherence happens when patients don't follow their medication instructions. It's especially common among those taking many different drugs, like older adults who often have multiple health conditions. Not taking medicine correctly can increase death risks and lead to more use of other health services, like hospital stays or emergency room visits. This behavior can waste resources and harm the patient's health.

Figure 4.2: Trends in Pharmacy Deserts by Elderly/Non-Elderly Township



Notes: The data points are based on a three-year moving average, showcasing limited pharmacy access from a sample of 802 townships. A township receives an indicator value of 1 for limited pharmacy access if it has no independent or chain pharmacies within its boundaries. Townships with over 20% of their population aged 65 or older are classified as "elderly", while those with less than 20% are termed "non-elderly".

veals that a significant 86.82% of the stores exiting the market were independently owned pharmacies. In other words, independently owned pharmacies played a crucial role in providing prescription accessibility in rural areas. Had these independently owned pharmacies managed to sustain their operations, the surge in pharmacy deserts would likely have been less pronounced.

Table 4.1: Exits of Last Pharmacies in Rural Townships Resulting in Limited Pharmacy Accessibility

	The number of chain pharmacies		Total
The number of independent pharmacies	0	1	
0	0	54	54
1	362	2	369
2	7	0	7
Total	369	56	425
Percentage	86.82%	13.18%	100%

Notes: This table details the distribution of independent and chain pharmacies from the previous year. Towns without these pharmacies in the subsequent year are classified as having limited pharmacy accessibility due to these exits.

Summary Statistics

Table 4.2 and 4.3 provide descriptive summary statistics of my final sample, which comprises 291 non-elderly townships and 511 elderly ones for twenty years, with a total 16,040

Table 4.2: (Selected) Descriptive Statistics: Non Elderly Township

		Panel A. Year 2000-2009					*Panel B. Year 2010-2019*				
Variable	Frequency	Mean	S.D	Median	Min	Max	Mean	S.D	Median	Min	Max
Township-level variables											
Pop.[a]	Decennial	2077	1578	1775	114	14388	2087	1592	1826	123	14738
Income per Capita[b]	Decennial	16410	3012	16455	8360	35705	20912	4450	20807	10306	42282
Prop. Age over 65[c]	Decennial	0.159	0.03	0.17	0.05	0.20	0.168	0.04	0.17	0.05	0.37
Prop. Female	Decennial	0.507	0.02	0.51	0.39	0.62	0.502	0.02	0.50	0.31	0.55
Prop. Black	Decennial	0.011	0.05	0.00	0.00	0.50	0.011	0.05	0.00	0.00	0.52
Prop. Vehicle = 0	Decennial	0.071	0.06	0.06	0.00	0.54	0.065	0.07	0.05	0.00	0.72
Pharmacy Desert[d]	Annual	0.342	0.47	0.00	0.00	1.00	0.341	0.47	0.00	0.00	1.00
Ind. Pharmacies (Town)[e]	Annual	0.671	0.55	1.00	0.00	2.00	0.646	0.58	1.00	0.00	2.00
Chain Pharmacies (15 miles)[f]	Annual	0.746	1.12	0.00	0.00	7.00	1.382	1.73	1.00	0.00	7.00
County-level characteristics											
Physician Offices	Annual	9.425	11.46	5.00	1.00	96.00	9.155	10.66	5.00	1.00	83.00
State-level characteristics											
Prop. Insurance Age 18-64[g]	Annual	0.871	0.02	0.87	0.83	0.93	0.873	0.04	0.87	0.79	0.97
Prop. Insurance Age over 65[g]	Annual	0.992	0.01	0.99	0.97	1.00	0.991	0.01	0.99	0.96	1.00
Ind. Pharmacies characteristics	Annual										
Employee	Annual	6.693	9.938	6	0	400	6.591	4.166	6	0	50
Years in business	Annual	5.787	3.364	6	0	12	10.201	6.971	11	0	22
N			2,910					2,910			

Notes: "Non Elderly township" is defined as townships with an age over 65 population ratio lower than 20% in the year 2000. A comprehensive list of descriptive statistics for the final dataset can be found in Appendix A. "Decennial" implies that the census is conducted every ten years. "Annual" indicates that updates are made on a yearly basis. [a] "Pop." refers to the total population of each township. [b] "Income per Capita" represents the median income of each township. [c] "Prop." stands for the proportion of a specific demographic group within the population. [d] "Pharmacy deserts" is a binary variable taking the value 1 if there are no available pharmacies within the township. [e] "Ind. Pharmacy" denotes the average number of independent pharmacies within the township. [f] "Chain Pharmacy" denotes the average number of chain pharmacies within a 15-mile radius of the centroid of township. [g] "Prop. Insurance" refers to the ratio of the population within each age groups enrolled in health insurance.

market-level observations(802 towns * 20 years)[24]. Elderly townships typically have a smaller population than non-elderly townships, which typically results in lower market demand. As chain pharmacies prefer to enter markets with higher demand, elderly townships have more independent pharmacies on average. This highlights that independent pharmacies play an important role in providing prescriptions in rural towns.

While demographics remain relatively stable over time for both groups of townships, the pharmacy industry has undergone significant changes. Over the past two decades, the number of chain pharmacies within a 15-mile radius has doubled for both groups of townships. In contrast, the number of independent pharmacies in elderly townships declined, a trend not observed in non-elderly townships, which aligns with previous findings.

In the subsequent section, I study the prevalence of limited accessibility to pharmacy in the Midwest United States. I focus on understanding the mechanisms driving the increasing trend in pharmacy deserts through the lens of competition, with a particular emphasis on the entry of chain pharmacies.

[24]For a full list of variables, Appendix D.2 and D.2 provide descriptive statistics of my final sample. I show key selected variables for brevity in main Table 4.2 and 4.3.

Table 4.3: (Selected) Descriptive Statistics: Elderly Township

		Panel A. Year 2000-2009					*Panel B. Year 2010-2019*				
Variable	Frequency	Mean	S.D	Median	Min	Max	Mean	S.D	Median	Min	Max
Township-level variables											
Pop.[a]	Decennial	1434	725	1307	153	4859	1385	732	1227	117	4745
Avg. Income[b]	Decennial	16807	2206	16721	10022	27227	21345	3417	21202	12156	37437
Prop. Age over 65[c]	Decennial	0.262	0.05	0.25	0.20	0.48	0.246	0.05	0.24	0.10	0.49
Prop. Female	Decennial	0.528	0.02	0.53	0.37	0.59	0.518	0.02	0.52	0.29	0.62
Prop. Black	Decennial	0.002	0.01	0.00	0.00	0.08	0.004	0.01	0.00	0.00	0.17
Prop. Vehicle = 0	Decennial	0.076	0.03	0.07	0.00	0.22	0.062	0.04	0.06	0.00	0.25
Pharmacy Desert[d]	Annual	0.178	0.38	0.00	0.00	1.00	0.224	0.42	0.00	0.00	1.00
Ind. Pharmacies (Town)[e]	Annual	0.841	0.52	1.00	0.00	2.00	0.715	0.58	1.00	0.00	2.00
Chain Pharmacies (15 miles)[f]	Annual	0.357	0.78	0.00	0.00	6.00	0.676	1.16	0.00	0.00	7.00
County-level characteristics											
Physician Offices	Annual	5.699	8.90	3.00	1.00	80.00	6.622	10.07	3.00	1.00	80.00
State-level variables											
Prop. Insurance Age 18-64[g]	Annual	0.878	0.02	0.88	0.83	0.93	0.879	0.04	0.88	0.79	0.97
Prop. Insurance Age over 65[g]	Annual	0.993	0.01	0.99	0.97	1.00	0.991	0.01	0.99	0.96	1.00
Ind. Pharmacies characteristics											
Employee	Annual	5.904	4.074	5	0	71	6.170	3.958	5	0	71
Years in business	Annual	6.081	3.324	6	0	12	12.390	6.711	14	0	22
N			5,110					5,110			

Notes: "Non-elderly township" is defined as townships with an age over 65 population ratio lower than 20% in the year 2000. A comprehensive list of descriptive statistics for the final dataset can be found in Appendix A. "Decennial" implies that the census is conducted every ten years. "Annual" indicates that updates are made yearly. [a] "Pop." refers to the total population of each township. [b] "Income per Capita" represents the median income of each township. [c] "Prop." stands for the proportion of a specific demographic group within the population. [d] "Pharmacy deserts" is a binary variable taking the value 1 if there are no available pharmacies within the township. [e] "Ind. Pharmacy" denotes the average number of independent pharmacies within the township. [f] "Chain Pharmacy" denotes the average number of chain pharmacies within a 15-mile radius of the centroid of a township. [g] "Prop. Insurance" refers to the ratio of the population within each age group enrolled in health insurance.

4.4 Market Structure

The entry of chain pharmacies has profoundly transformed the landscape of the retail pharmacy market, introducing a new competitive format and challenges for independently owned pharmacies. Following the classification by Grieco (2014), independently owned pharmacies are defined as either single stores or those sharing a parent company with fewer than three stores. Chain pharmacies encompass two distinct formats: standalone retail pharmacies (e.g., Walgreens) and supermarket-based pharmacies (e.g., Pharmacy).

As an illustrative case, I present a snapshot of changes in the pharmacy market environment due to the entry of new chain pharmacies over time. Figure 4.3 focuses on Super Township in Kansas, which is shaded in gray. In this figure, each red circle denotes independent pharmacies, and each blue star denotes chain pharmacies. Each boundary delineates a township, averaging around 29 square miles in size and 5.5 miles in width, in line with the typical dimensions of townships in the current dataset. In 2000, there was one independent pharmacy in the town, accompanied by one chain pharmacy within a 15-mile radius from the centroid of town. By 2009, the market experienced more chain pharmacy entry, with a total of three chain pharmacies actively operating. By 2019, more chain pharmacies had entered, bringing the total to six within the 15-mile radius. Due to intensified competition from these chain pharmacies, the independent pharmacies in Superior Township shut down. After the independent pharmacy left markets in 2019, the town was classified as a "limited access to pharmacy" area. Based on Figure 4.3, I summarize the following observations:

1. Chain pharmacies are more abundantly and densely situated in high-demand areas, such as shopping malls.

2. The new entry of chain pharmacies is associated with the exit of independent pharmacies.

3. The decline of independent pharmacies is associated with the more prevalent limited pharmacy accessibility at the town level.

To see the overall patterns, Figure 4.4 illustrates the negative correlation between independently owned pharmacies and chain pharmacies by documenting the average number of stores in the final same over the period 2000-2019. Within towns, the average number of independently owned pharmacies decreased by 0.18 whereas the average number of chain pharmacies increased by 0.13 units, in total 0.05 units decreased. In line with anecdotal evidence presented in Figure 4.3, the average number of chain pharmacies outside of town to 15 miles increased by 0.38 units. It indicates that new chain pharmacy entries in distant urban areas or shopping mall towns.

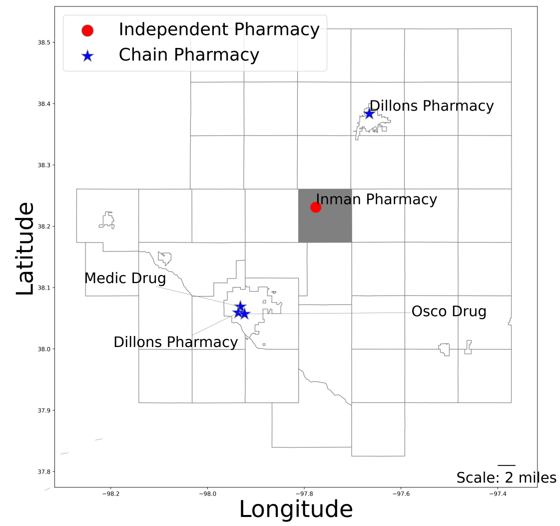
In Appendix Figure C.5, I also examine changes in the number of independent/chain pharmacies by age group; elderly townships and non-elderly townships. In non-elderly townships, independent pharmacies exited the market less frequently than in elderly townships, despite the more pronounced increase in chain pharmacies outside of townships. This empirical finding suggests that the entry of new chain pharmacies might impact independent pharmacies differently across age groups.

Finally, Figure C.6 shows changes in the market structure of independent pharmacies by age group. Specifically, it examines the distribution of townships unserved, monopolies, and duopolies among independent pharmacies. For both elderly townships and non-elderly townships, monopolies are decreasing, while unserved areas are increasing. The changes are greater in elderly townships, which is aligned with Figure C.5.

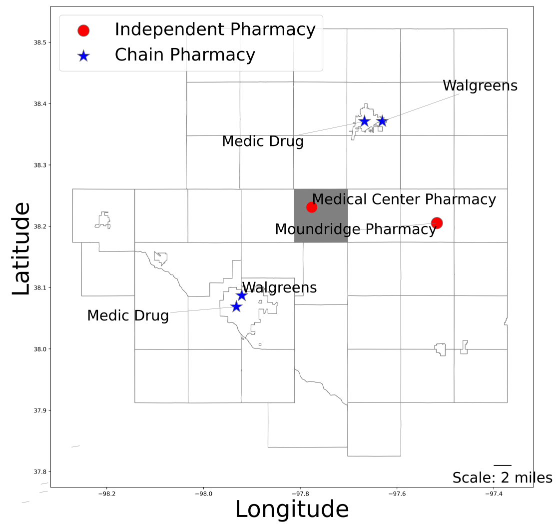
4.5 Reduced Form Evidence

In this section, I present evidence on the impact of chain pharmacies on local independently-owned pharmacies. The goal is to evaluate whether or not the new entry of chain pharmacies is associated with a decrease in the number of local independent pharmacies. I use the final dataset between 2000 and 2019 for the analysis.

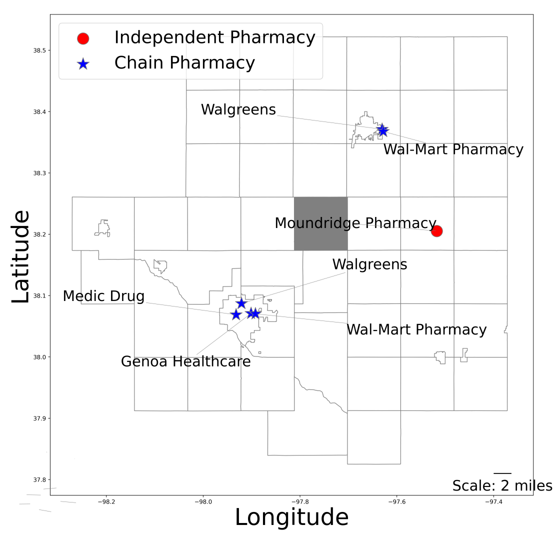
Figure 4.3: An Example: Spatial Distribution in Independent/Chain Pharmacy



(a) Year: 2000
# of Chain within 15 miles: 1



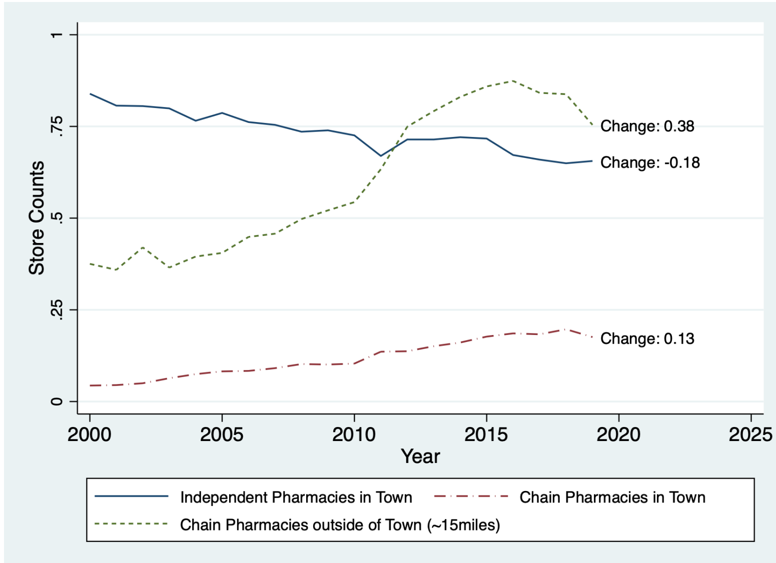
(b) Year: 2010
# of Chain within 15 miles: 3



(c) Year: 2019
# of Chain within 15 miles: 5

Note: The samples in this study are drawn from Superior Township in Kansas and their neighborhood, covering the years 2000 to 2019. In the visual representation, Superior Township is highlighted in grey. Independent pharmacies are marked with red circles, while chain pharmacies are indicated by blue stars. The vertical labels represent latitude, and the horizontal labels denote longitude. This figure highlights the following: 1. Chain pharmacies are more abundant and densely situated in high-demand areas, such as shopping malls. 2. The decline of independent pharmacies has contributed to the growth of pharmacy deserts in the US.

Figure 4.4: (Average) Number of Independent/Chain Pharmacies between 2000-2019



To provide a descriptive overview of the impact of chain pharmacies on independent pharmacies, I document the distribution of active local independent pharmacies relative to the presence of chain pharmacies within a 15-mile radius. Table 4.4 shows that towns with nearby chain pharmacies have a 44.06% rate of un-served markets, in contrast to 25.60% in towns without chain pharmacies in the vicinity. Furthermore, there is a notable reduction in monopolies and duopolies run by independent pharmacies in these areas. Also, towns with nearby chain pharmacies have a higher rate of unserved markets and fewer monopolies and duopolies run by independent pharmacies. This suggests that chain pharmacies made less room for independent pharmacies.

Table 4.4: Distribution (of the Number) of Active Independent Pharmacies in Town 2000-2019 by the Number of Chain Pharmacies (%)

	Number of chain pharmacies within 15 miles		Total
Active independent pharmacies	0	1+	Overall
0	25.60	44.06	32.40
1	67.77	51.38	61.73
2	6.63	4.56	5.87
N	12,822	7,228	16,040

In Table 4.5, the distribution of active independent pharmacies is segmented by the age profile of the townships (elderly or non-elderly) and further bifurcated by the proximity of chain pharmacies. In non-elderly townships, the influence of chain pharmacies on independent pharmacy distribution is subtle. When no chain pharmacies are nearby, 37.69% of the areas do not have an independent pharmacy, 57.41% have one, and a mere 4.90% have two or more. In contrast, with the presence of at least one chain pharmacy, the figures change slightly to 39.94%, 55.75%, and 4.31%, respectively. This indicates that in non-elderly townships, the proliferation of chain pharmacies has a small impact on the number of independent pharmacies. On the other hand, the dynamics in elderly townships show a noticeable difference. Absent any chain pharmacies, 20.43% of the areas lack an independent pharmacy, with a significant 72.20% having just one, and 7.37% with two or more. However, with the introduction of a chain pharmacy, the distribution changes: 47.74% of the areas have no independent pharmacy, 47.49% have a single store, and 4.77% have more than one. This difference emphasizes the pronounced influence of chain pharmacies on the prevalence of independent pharmacies in elderly townships.

Table 4.5: Distribution (of the Number) of Active Independent Pharmacies in Town 2000-2019 by Age Group (%)

	Num. independent pharmacies		
	0	1	2+
Non Elderly Township			
No Chain	37.69	57.41	4.90
Chain ≥ 1	39.94	55.75	4.31
Elderly Township $\geq$ 1,500			
No Chain	20.43	72.20	7.37
Chain ≥ 1	47.74	47.49	4.77

Specification of Distance to Chain Pharmacies

To inform whether the new entry of a chain pharmacy within a certain radius is associated with competition in the independent pharmacy in the town, I regressed the number of independent pharmacies on the new entry of chains with different mile radii from the centroid of towns. Table 4.6 provides suggestive evidence that considering the new entry of a chain within 15 miles may be suitable for modeling independent pharmacy entry/exit. Outside of 15 miles, the effects are not statistically significant.

Effects on Market Structure

Next, I conduct an event study to present the effects of chain pharmacy entry over the years before and after their introduction. In this regression, I estimate:

$$Y_{mt} = \sum_{\tau} \delta_{\tau} \text{Entry}_{m,t-\tau} + \beta X_{mt} + \lambda_m + \alpha_t + \gamma_{st} + \epsilon_{mt} \tag{4.1}$$

Table 4.6: The New Entries in Chain Pharmacies with Different Distances and the Number of Independent Pharmacies.

	(1) Independent Stores	(2) Independent Stores	(3) Independent Stores	(4) Independent Stores
I(Chain Entry=1, 0-5 miles)	-0.446***			
	(0.0402)			
I(Chain Entry=1, 5-10 miles)		-0.0569*		
		(0.0255)		
I(Chain Entry=1, 10-15 miles)			-0.0274+	
			(0.0159)	
I(Chain Entry=1, 15-20 miles)				-0.00700
				(0.00977)
Township FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Market × Year FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	16040	16040	16040	16040
Mean of Dep. Variable	0.735	0.735	0.735	0.735
Adjusted R^2	0.573	0.537	0.537	0.537

Note: Estimates are from fixed effects regressions of the new entry of chain pharmacies within different distances on the number of independent pharmacies in township m and year t. Column (1) denotes the entry of chain pharmacies within 5 miles, Column (2) denotes the entry of chain pharmacies between 5 and 10 miles, Column (3) denotes the entry of chain pharmacies between 10 and 15 miles and Column (4) denotes the entry of chain pharmacies between 15 and 20 miles. Significance levels are denoted by + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$.

where $Entry_{m,t}$ denotes a dummy variable for whether a chain store has entered location m by period t. The outcome of interest variables Y_{mt} denotes the number of independent pharmacies at township m in period (year) t. I control for township-level demographics X_{mt}, unobserved township-level fixed effects λ_m, and yearly time fixed effects α_t. To control time-varying unobserved heterogeneity, I incorporate market-year fixed effects γ_{st} where s denotes the state level. I focus only on binary specification, meaning that $Entry_{mt}$ takes the value 1 if chain stores enter and 0 otherwise, instead of the number of entries. [25]

As the entry of chain pharmacy is heterogeneous across townships, this boils down to staggered Difference-in-Difference with two-way fixed effects (TWFE) designs (e.g. Goodman-Bacon (2021), Callaway and Sant'Anna (2021)). I address two issues: (i) heterogeneous treatment effects in the presence of different timing of treatment, which can induce bias in coefficients due to the use of different timing groups (early versus late-treated) as controls, and (ii) pre-treatment effects. To do this, I run event studies that detect possible pre-trends as well as robust to heterogeneous treatment timing.

My preferred TWFE models are those by de Chaisemartin and D'Haultfoeuille (2023) because de Chaisemartin and D'Haultfoeuille (2023)'s approach can accommodate one-shot treatment with heterogeneous treatment periods (e.g. Hurricane in different dates). Figure 4.5 shows that both the standard TWFE and the de Chaisemartin and D'Haultfoeuille (2023) methods indicate an absence of statistically significant effects in terms of pre-trends. However, post-treatment shows that the entry of chain pharmacies is associated with a decreased number of independent pharmacies in the township. In Appendix Figure C.7, I also provide the results using alternative weights on heterogeneity-robust estimators by Borusyak et al. (2021), Callaway and Sant'Anna (2021), and Sun and Abraham (2021). These results show that alternative ways of constructing weights for event study are robust to my preferred TWFE design.[26]

4.6 Preliminary Analysis

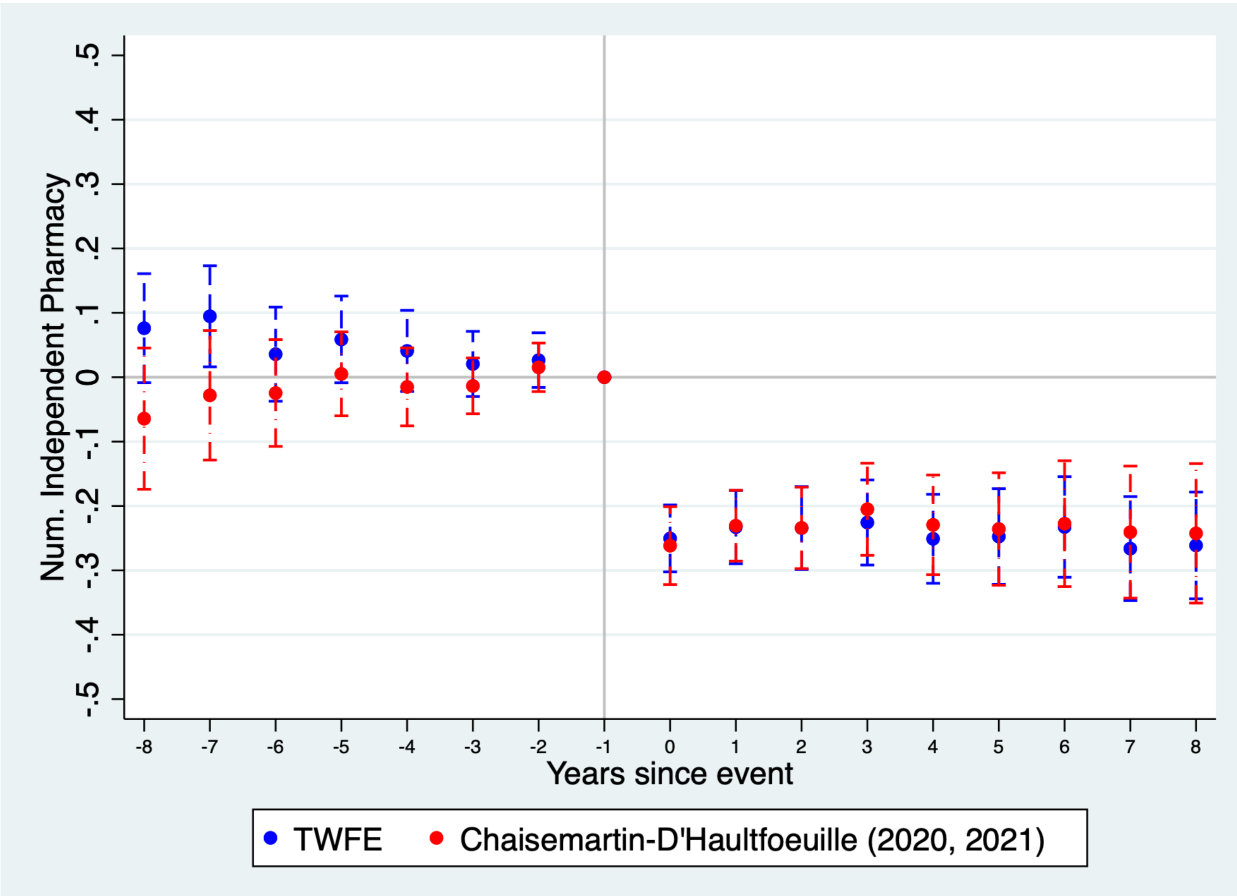
To demonstrate how entry and exit patterns change with the rival's independent store, I present the data patterns using simple logit regressions.

An independent firm in market m makes a binary decision of each firm $a_{imt} = (0, 1)$ where $a_{imt} = 0$ if firm i being active in market m period t and $a_{imt} = 1$ if i being inactive. I also

[25] In my final sample, when an entry event occurs, 88% of entries were the entry of one chain pharmacy.

[26] See Appendix for similar exercises for subsamples: non-elderly townships in Figure C.8 and elderly townships in Figure C.9. Consistent with earlier findings, the effects of new chain entries are larger in elderly townships. Non-elderly townships have somewhat mixed results, so in the structural analysis in the next section, I separately recover parameters of interest in the two distinct township types: non-elderly and elderly.

Figure 4.5: Event Study: The effects of chain pharmacy entry on local independent pharmacy



Note: Coefficient plots from event-study difference-in-differences analyses that regress the number of independent pharmacies in a township on year fixed effects, county fixed effects, control variables, and market× year fixed effects. The sample consists of 802 townships between 2000 and 2019. The omitted baseline period is $t = -1$, which is the last pre-treatment period. Standard errors are clustered at the county level and error bars represent 95 confidence intervals.

include a binary variable of whether states adopted Medicaid expansion policy afterwards 2006. As these regressions do not take into account the simultaneous entry of rival firms, the results do not reveal causality, but correlation.

Table 4.7: Logit Regression on Independent Pharmacy's Entry.

	(1)	(2)
I(Rival Store=1)	-3.273***	-3.293***
	(0.176)	(0.178)
Chain Pharmacies within 15 mi	-0.543***	-0.526***
	(0.0713)	(0.0727)
Pharmacy's Employee Size	1.951***	1.965***
	(0.222)	(0.224)
Rival's Employee Size	-1.222***	-1.210***
	(0.259)	(0.259)
Total Pop.	0.972***	0.966***
	(0.221)	(0.221)
Income Per Capita	0.0418	0.262
	(0.290)	(0.464)
Physician Offices	-0.00292	0.0295
	(0.133)	(0.134)
Prop. Age over 65	5.772**	5.734**
	(1.930)	(1.926)
Prop. Female	4.086	4.056
	(4.021)	(4.052)
Prop. Black	-7.822	-7.379
	(6.323)	(6.414)
Prop. - High School Graduates	-0.975	-0.859
	(1.358)	(1.367)
Prop. Unemployment	-1.586	-1.220
	(1.961)	(2.064)
Prop. Vehicle = 0	5.855**	5.869**
	(1.798)	(1.837)
Medicaid Expansion	0.218**	0.0781
	(0.0822)	(0.112)
Prop. Insurance Age over 65	-1.483	-5.553
	(3.096)	(3.439)
County FE	Yes	Yes
Year FE	No	Yes
Observations	32,040	32,040
Mean of Dep. Variable	0.367	0.367
Adjusted R^2	0.348	0.350

Notes: Binary Logit estimates of stay-in/out in township m and year t. These Results do not control for the endogeneity of decisions between small independent pharmacy stores. Standard errors are clustered at the county level. Significance levels are denoted by + p<0.10, * p<0.05, ** p<0.05.*** p<0.01.

Table 4.7 reports the result of a logit model on entry, controlling for chain pharmacies within 15 miles, demographic variables, and health-related characteristics. First, the presence of rival stores in the same town is strongly negatively correlated with entry decisions. That is, the presence of a rival might substantially lower the latent payoff of being active, which explains why it is important to capture the primary competitor - the rival. Second, the number of chain pharmacies is negatively correlated with entry decisions, but weaker than the competition effects from the rival. It means that chain pharmacies might offer different types of services (e.g., higher qualities, better in-network-premium) so chain pharmacies are more like a secondary competitor. Interestingly, my findings are similar to Grieco (2014) in that the magnitude of the effect of chain supermarkets is smaller than that of independent groceries.

5 Structural Analysis

In this section, motivated by reduced-form analysis, I present a structural model of independent pharmacies and report the estimation results of the model.

5.1 Model Primitive

I model the entry decision of an independently owned pharmacy as a discrete-time, simultaneous-move game. Each year, every store decides whether to be active or inactive in the market. I focus on duopoly markets because 99.30% of my samples contain at most two operating independent pharmacies.

I assume that the information structure of the games between independent pharmacies follows incomplete information, as in Assumption 2.1. Each player observes her own private information while she cannot observe that of her rival. Instead, she knows the distribution of her rival's private information. Allowing players to have private information significantly reduces the computational costs of estimation.

The model also assumes that the entry of chain pharmacies is given from the perspective of independent pharmacies. This is a reasonable assumption for rural independent pharmacy market settings because the decision for chain pharmacies to enter is more likely driven by the broader demographics of larger regions, their network structures, and where their distribution centers are located. The competition from small independent stores is less of a concern for the national chain pharmacies. Therefore, the model is greatly simplified by assuming that decisions made by national chains are given to local independent stores. These approaches are used in other studies examining the strategic interaction between local stores. (Ackerberg and Gowrisankaran 2006, Grieco 2014). [27] Based on the empirical evidence in Table 4.6, the analysis focuses on the number of chain pharmacies within a 15-mile radius of town centers.

The store's choice-specific value function when active in the market depends on the beliefs over a rival's binary choices a_{-i} in market m at period t:

$$u_{imt}(a_{imt}=1, d_{mt}, \epsilon_{imt}(1); \theta) = \prod_i (a_{imt}=1, d_{mt}; \theta) + \epsilon_{imt}(1).$$
$$= \sigma_{-i}(a_{-imt}=1|d_{mt})\pi_i(a_{imt}=1, a_{-imt}, d_{mt}; \theta) + (1-\sigma_{-i}(a_{-imt}=1|d_{mt}))\pi_i(a_{imt}=1, a_{-imt}, d_{mt}; \theta) + \epsilon_{imt}(1). \tag{5.1}$$

where $\sigma_{-i}(a_{-imt}=1|d_{mt})$ is the probability of rival's being active, conditional on observable market characteristics d_{mt}. θ denotes the set of structural parameters affecting the pharmacy's per-period payoff, and $\epsilon_{imt}(1)$ denotes being active-specific private information for pharmacy i. I further assume that the value of being inactive is normalized to zero.

Equilibrium Concept

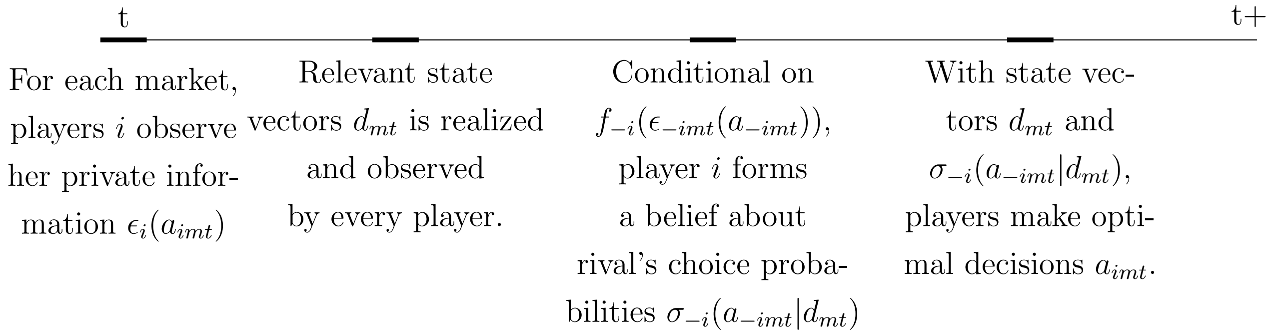
I focus on the Bayesian Nash equilibrium, where a store's choices are the best responses conditional on its belief about the rival. Under the rational expectation assumption in Section 2, each store forms correct beliefs about the rival and observes its own private information, but not its rival's. Pharmacy stores observe the distribution of the rival's private information, which I assume follows a Type 1 Extreme Value Distribution. Each firm's strategy is a function of the probability of a rival's entry, the observed state variable, and its private, choice-specific shocks. Because econometricians cannot observe private information, the optimal strategy can be expressed as choice probabilities:

$$\sigma_i^*(a_{imt}=1|d_{mt}, \sigma_{-i}^*) = \frac{\exp(\prod_i (a_{imt}=1, d_{mt}))}{1+\exp(\prod_i (a_{imt}=1, d_{mt})} \tag{5.2}$$

Given the equilibrium above, each player's decision is illustrated in Figure 5.1.

[27] Alternatively, one can consider a comprehensive model of the endogenous entry of independent pharmacies and chain pharmacies. The decision of each merchandise-based pharmacy (e.g., Walmart, Costco, and Sam's Club) depends on pre-existing merchandise department stores, distribution centers, and network effects with nearby department stores. However, considering such factors requires a very complicated method like Holmes (2011) or Jia (2008). These models have clear limitations: they cannot capture strategic interaction with other chain pharmacies due to the way they are developed. Moreover, from the perspective of giant chains like Walmart, their internal network structure would likely take precedence over considerations regarding a small, local independent pharmacy. In addition, the decision-making of these big-box stores might be fundamentally different from that of rural independent stores. Given the size of the stores, a dynamic setting might be more suitable, which would require a separate model for the national chains. More importantly, the goal of this paper is to understand the effect of the entry of chain pharmacies on the landscape of the rural market, not to understand the expansion of chain pharmacies. For these reasons, it is beyond the scope of this paper to consider the endogenous entry model of chain pharmacies with a properly defined structural model

Figure 5.1: Timing of the Game



5.2 Discussion

Before presenting the estimation procedure and discussing the results of the structural estimation, I provide a detailed discussion of the assumptions for the structural model.

Static versus Dynamic

I model the discrete choice of an independent pharmacy as a static game for the following reasons. First, if the entry costs (including fixed sunk costs) of entering the industry are substantial, it might be suitable to consider a dynamic model. As the dynamic model can differentiate between sunk costs and fixed costs, it can accommodate forward-looking behaviors observed in manufacturing (e.g., the cement industry Ryan (2012)) and the hardware industry (Igami and Uetake (2020)). However, anecdotal evidence from pharmacy industry reports suggests that the sunk costs of opening independent pharmacy stores are relatively small compared to the yearly gross profits from running stores, as shown in Appendix Table C.6[28]. This implies that sunk costs may not be substantial. The static approach to studying pharmacies has been utilized in other studies (Aradillas-López and Gandhi (2016)). Next, Appendix Table C.7 documents the regression of past entries of chain pharmacies on the number of independent pharmacies in township m at period t. After controlling for the number of chain pharmacies in the same year, the behavior of independent pharmacies does not change, suggesting that past chain entry may be negligible. Based on this observation, I assume that a player's payoff depends only on current state variables, not on state variables from past periods.

[28]Source: Elabed et al. (2016). The industry report indicates that the dollar metrics of entry's sunk costs constitute a relatively smaller portion of yearly profits. Specifically, the components of sunk costs of entry are approximately \$107,000, and yearly gross profits are around \$748,000, thus the ratio of entry's sunk costs to gross profits is around 14.3%.

Identification

To achieve identification, as discussed in Bajari et al. (2010b), I adopt firm-specific variables that have been utilized in existing works. For example, Grieco (2014) used the existence of a two-year-old firm as a shifter for independent grocery stores. To alleviate concerns regarding endogenous entry, I instead use employment size from three years ago. The suggested exclusion restriction is valid if the revenue of an operating store increases with respect to the number of employees but is unrelated to the profits of rival stores. The underlying assumption is that a store's profit can be affected by its own employment, while the employment of rival stores will impact a store's profit only through the rival's choice of remaining in the market.[29]

Assumption: Isolated Market

Pioneered by Bresnahan and Reiss (1991), the literature on strategic interaction typically focuses on isolated markets with few firms. If independent pharmacies outside of the township influence the entry/exit of independent pharmacies within the township, then the assumption is likely to be invalid. To address this concern, I observe the following. The average distance between town boundaries is 14 miles which ensures far enough distance between markets. This evidence suggests that most towns in my final sample are closer to being isolated markets. Additionally, Table 5.1 presents the regression of new independent pharmacy entry outside of the township on the number of independent pharmacies within the town, including market characteristics, town fixed effects, year fixed effects, and market $\times$ year fixed effects. Table 5.1 provides suggestive evidence that the new entry of independent pharmacies in nearby towns (outside of the township) might have little impact on independent pharmacies within the township.

5.3 Estimation Method: Two-Step Estimators

I use a two-step estimator to recover underlying structural parameters of interest. Specifically, in the first stage, I obtain reduced-form estimates of beliefs over the rival's CCP from the data. In the second stage, I use the rival's CCP, observed market characteristics, chain pharmacy effects, and firm-specific shifters to construct moment conditions. Finally, I minimize these moment conditions over a set of candidate structural parameters. Additionally, I estimate the model separately using the sample of elderly and non-elderly towns, as I

[29] Following this shifter, I decompose notation $d_{mt} = d_{imt}, d_{-imt}, d_{0mt}$, where d_{imt} denotes store i specific shifter, d_{-imt} denotes store $-i$ specific shifter, and d_{0mt} denotes the common market characteristics.

Table 5.1: Entries in Neighborhood (outside of township) and the Number of Independent Pharmacies.

	(1) Independent Pharmacies
(Independent Pharmacy Entry Outside of Town Boundary=1)	-0.0131
	(0.0198)
Township FE	Yes
Year FE	Yes
Market × Year FE	Yes
Controls	Yes
Observations	16,040
Mean of Dep. Variable	0.735
Adjusted R^2	0.534

Note: Estimates are from fixed effects regressions of the new entry of independent pharmacies outside of township (within 10 miles) on the number of independent pharmacies in township m and year t. Standard errors are clustered at the town level. Significance levels are denoted by + p<0.10, * p<0.05, ** p<0.05.*** p<0.01.

find substantial heterogeneity in market dynamics in the previous section. This approach is widely used in the industrial organization literature. (e.g., Ellickson and Misra (2008)).

Time-Varying Unobserved Endogenous Variables: I attempt to address the issues discussed in Berry and Compiani (2023) for both the estimators of Bajari et al. (2010b) and Orthogonal estimators. Berry and Compiani (2023) demonstrated that the first-stage estimation of choice probabilities should not be adversely affected by the presence of unobserved and time-varying endogenous state variables. To address this issue, I employ three distinct strategies. First, I include county-fixed effects to control for time-invariant, unobserved, market-specific shocks. 90 percent of my final sample experienced a population change of less than 250 people between the 2000s and the 2010s, which means that demographics are quite stable over the years. This implies that the county-fixed effects capture much of the unobserved heterogeneity. Second, if one is willing to agree that stores make optimal hiring decisions spontaneously in response to changes in market characteristics, labor employment might capture much of the time-varying, unobserved market characteristics. Third, I also include state Medicaid expansion, a policy-relevant variable, as suggested by Berry and Compiani (2023). In the logit regression result of the independent pharmacy's entry in Table 4.7, the expansion of Medicaid coverage has a positive and significant effect on entry, while this effect becomes insignificant when including year-fixed effects as in column (2). This suggests that both year-fixed effects and county-fixed effects absorb much of the unobservable market characteristics.

Estimation method: For comparison, I present the result using two different estimators: Bajari et al. (2010b)'s estimators with a few pre-selected variables and developed orthogonal estimators using flexible ML methods with many covariates. The main differences in estimation algorithms between the two estimators are 1) the set of control variables used, 2) the use of the Machine Learning estimator, and 3) the use of a cross-fitting algorithm. The details of the estimation algorithm are as below.

5.3.1 Bajari et al. (2010b)'s approach

First Stage Nuisance Parameter γ_{-i} Estimation: The goal is to recover reduced-form beliefs over the rival's equilibrium CCP from the data. The reduced-form estimates of CCP take the form of conditional expectation:

$$\hat{\gamma}_{-imt} = E[a_{-imt}|c_{mt}, d_{-imt}, d_{imt}, d_{0mt}^{pre}, y_t, \text{county}_f] \tag{5.3}$$

where a_{-imt} denotes the rival's binary choice, c_{mt} represents the number of chains within 15 miles, d_{-imt} indicates the rival's shifter, which is the number of the rival's employees, d_{imt} denotes player i's shifter, which is the number of employees, d_{0mt}^{pre} denotes common market characteristics, y_t represents the year fixed effects, and county_f represents the county fixed effects.[30] For d_{0mt}^{pre}, I assume that only a relatively small number of pre-selected market characteristics are relevant for independent pharmacies' payoffs, as in the previous empirical IO/health literature. I employ a simple logit model to estimate the conditional expectation of equation 5.3.

After recovering the equilibrium strategies from the data, the goal of the second stage is to estimate the structural parameters of interest. First, I construct standard empirical models of entry and market structure to describe the moment conditions. Since I lack data on prices and quantities, I model the profit functions in a reduced-form manner, following the conventional approach in static entry literature (e.g., Berry (1992), Seim (2006)). The average period profit per store in market m in period t is characterized as follows:

$$\begin{aligned}&\prod_i(a_{imt} = 1, \hat{\gamma}_{-imt}, c_{mt}, d_{imt}, d_{0mt}^{pre}, y_t, \text{county}_f; \theta) = \\ &\hat{\gamma}_{-imt}\theta_\gamma + c_{mt}\theta_c + d_{imt}\beta_1 + d_{0mt}^{pre} \cdot \beta + \alpha_t y_t + \alpha_{\text{county}}\text{county}_f\end{aligned} \tag{5.4}$$

where $\hat{\gamma}_{-imt}$ represents beliefs about the rival's CCP.

[30] I choose county-level fixed effects, as town-level fixed effects are too granular; some estimates did not converge due to too many fixed effects in a simple conditional logit model

Second Stage Structural Parameter Estimation: I construct the logit-likelihood function, as developed by Bajari et al. (2010b) and Bajari et al. (2013). The logit-likelihood function depends on the payoff function given in equation (5.4) and equilibrium function (5.2), and estimates a set of parameters:

$$\text{argmin}_{\theta_\gamma,\theta_c,\beta_1,\beta,\alpha_t,\alpha_{\text{county}}} \ln\mathscr{L} = \sum_t \sum_m \sum_i \; a_{imt}\ln(\sigma_{imt}) + (1-a_{imt})\ln(1-\sigma_{imt}) \quad (5.5)$$

where

$$\sigma_{imt} = \Lambda(\hat{\gamma}_{-imt}\theta_\gamma + c_{mt}\theta_c + d_{imt}\beta_1 + d^{pre}_{0mt}\cdot\beta + \alpha_t y_t + \alpha_{\text{county}}\text{county}_f)$$

Here, Λ represents the standard logit link function under Type 1 Extreme Value Distribution. The consistency and asymptotic normality of the estimator have been established by Bajari et al. (2010b). I account for correlation in error terms by taking the clustered standard error at the county level.

5.3.2 Neyman Orthogonal Estimators

Neyman Orthogonal estimators facilitate the use of flexible ML methods, which perform well in prediction, and also relax assumptions regarding the data-generating process. Consequently, I do not pre-select socio-economic variables, utilizing instead the pool of variables d^{pool}_{0mt} as shown in the summary statistics Table D.2 and Table D.2.

Nuisance Parameter Estimation

Estimation of γ_{-i}: I implement ML classifiers to obtain conditional expectations using a richer pool of demographics:

$$\hat{\gamma}_{-imt} = E[a_{-imt}|c_{mt}, d_{-imt}, d_{imt}, d^{pool}_{0mt}, y_t, \text{county}_f] \quad (5.6)$$

Estimation of (β_1, β): Given $\hat{\gamma}_{-imt}$ in hand, I estimate the nuisance parameters β_1 and β. I accommodate flexible interactions between these richer independent variables $d^{interaction}_{omt}$. Due to the nature of high-dimensional settings, I construct the following Logit Lasso specification:

$$(\hat{\theta}\gamma, \hat{\theta}c, \hat{\beta}_1, \hat{\beta}, \hat{\alpha}_t, \alpha_{\text{county}}) \quad \in \text{argmin}\theta\gamma, \theta_c, \beta_1, \beta, \alpha_t, \alpha_{\text{county}} \left[E_n[\Lambda_i(\theta_\gamma, \theta_c, \beta_1, \beta, \alpha_t, \alpha_{\text{county})}] + \frac{\lambda_1}{n}||(\theta_\gamma, \theta_c, \beta_1, \beta)||_1 \right] \quad (5.7)$$

where λ_1 denotes the ℓ_1 penalty terms, trained by the 5-fold cross-fitting algorithm in the R package 'cv.glmnet'.

Estimation of μ_γ, μ_c: While ML methods perform well in prediction, regularization can induce biases in the structural parameters of interest. To obtain unbiased estimates of these parameters, I employ the cross-fitting algorithm described in Section 3. Furthermore, I construct and estimate the correction terms μ_γ and μ_c, following Section 3.3, associated with the rival independent pharmacy effects θ_γ and the number of chain pharmacy effects θ_c.

Structural Parameter Estimation: Given cross-fitted nuisance parameters $\hat{\eta} = (\hat{\gamma}_{-i}, \hat{\beta}_1, \hat{\beta}, \hat{\mu}_\gamma, \hat{\mu}_c)$, I construct orthogonal moment functions to obtain unbiased estimates of θ_γ, θ_c:

$$\text{argmin}_{\theta_\gamma, \theta_c} \psi(w_i; \theta, \eta) = E\left[\underbrace{(\mu_\gamma, \mu_c)[a_{imt} - \sigma_{imt}]}_{=m(w_{imt};\theta,\eta)} - (\alpha_\gamma, \alpha_c)[a_{-i} - \gamma_{-i}]\right] \tag{5.8}$$

$$\text{where } \sigma_{imt} = \Lambda(\hat{\gamma}_{-imt}\theta_\gamma + c_{mt}\theta_c + d_{imt}\hat{\beta}_1 + d_{0mt}^{interaction} \cdot \hat{\beta})$$

$$\alpha_\gamma = E[\mu_\gamma \Lambda(\cdot)(1 - \Lambda(\cdot))\theta_\gamma | d_{imt-3}, c_{mt-3}, d_{0mt-3}^{interaction}]$$

$$\alpha_c = E[\mu_c \Lambda(\cdot)(1 - \Lambda(\cdot))\theta_\gamma | c_{mt-3}, d_{0mt}]$$

I further define $M \equiv \frac{\partial m(w,\gamma,\beta,\eta)}{\partial \theta}$. The variance-covariance matrix has the following form:

$$\hat{V} = \left(\frac{1}{K}\sum_k^K E[M]\right)^{-1} \frac{1}{K}\sum_k^K E[\psi^2(w, \hat{\theta}, \hat{\eta_k}] \left(\frac{1}{K}\sum_k^K E[M]\right)^{-1}$$

I also employ county-level clusters to allow the correlation of error terms within the county level. For summary purposes, Table 5.2 presents the specifications of the classical methods and the orthogonal estimators developed in each estimation step. Additionally, Table 5.3 provides a summary of the notation for parameters to be estimated and the data.

5.4 Estimation Results

First Stage Estimation Results

Table 5.4 reports the estimation results of reduced-form CCP γ_{-i} based on Bajari et al. (2010b)'s estimator. The suggested shifter, employment size in pharmacy shows expected signs. The number of employees in a pharmacy is positively correlated to the probability of staying in the market, and the rival's number of employees is negatively associated, meaning that the number of employees might represent a proxy for sales and higher quality provision

Table 5.2: Summary of Specifications for Bajari et al. (2010b)'s Estimators and Orthogonal Estimators

Estimators	Stages	Parameters	Covariates	Dim (Covariates)	County & Year FE	Method
Bajari et al. (2010b)						
	First Stage	γ_{-imt}	$c_{mt}, d_{imt}, d_{-imt}, d_{0mt}^{pre}$	13	Yes	Simple Logit
	Second Stage	$\theta_\gamma, \theta_c, \theta_\gamma, \beta_1, \beta$	$\hat{\gamma}_{-i}, c_{mt}, d_{imt}, d_{0mt}^{pre}$	13	Yes	Logit Likelihood
Orthogonal Estimators						
	First Stage	γ_{-i}	$c_{mt}, d_{imt}, d_{-imt}, d_{0mt}^{pool}$	35	Yes	XG Boosting
	First Stage	β_1, β	$\hat{\gamma}_{-i}, c_{mt}, d_{imt}, d_{0mt}^{pool}, d_{0mt}^{interaction}$	563	Yes	Logit-Lasso
	First Stage	μ_γ	$c_{mt-3}, d_{imt-3}, d_{0mt-3}^{pool}, d_{0mt-3}^{interaction}$	563	Yes	Lasso
	First Stage	μ_c	$c_{mt-3}, d_{0mt}^{pool}, d_{0mt}^{interaction}$	563	Yes	Lasso
	Second Stage	θ_γ, θ_c	$\hat{\gamma}_{-i}, c_{mt}\hat{\beta}, \hat{\mu}_\gamma, \hat{\mu}_c$	2	No	Neyman Orthogonal Moment

Notes: pre denotes the pre-selected socio-economic variables. *pool* denotes the pool of richer socio-economic variables. *interaction* denotes the interaction term between the d_0^{pool}.

Table 5.3: Notation for Parameters and Data

Parameters	Description
γ_{-i}	Beliefs over rival CCP
θ_γ	The effect of rival independent pharmacy
θ_c	The effect of the number of chain pharmacies within 15 miles
β_1	The effect of store-specific shifter: Employees Size
β	The effect of common market characteristics
Data	**Description**
a_{imt}	Binary action of being active ($a_{imt} = 1$) or being inactive ($a_{imt} = 0$
c_{mt}	The number of chain pharmacy within 15 miles
d_{imt}	Player i specific shifter: Employees Size
d_{-imt}	Player $-i$ specific shifter: Employees Size
d_{0mt}	Common market characteristics
d_{0mt}^{pre}	Pre-selected Common market characteristics
d_{0mt}^{pool}	Pool of richer Common market characteristics
$d_{0mt}^{interaction}$	Interaction of d_{0mt}^{pool}

by the pharmacy. As expected, the impact of the total population appears to positively affect latent profits, as the total population can be a good proxy for the market size. The share of the population over 65 also appears to positively affect latent profits because the population over 65 may have higher demands for prescription drugs. Consistent with reduced-form evidence, chain pharmacies are larger in elderly towns than in non-elderly towns.

Next, I present the estimation results of predicting a rival's CCP using various ML methods. To shed light on the choice of ML methods, Table 5.5 summarizes the findings of applying various procedures and reports the out-of-sample (hold-out sample) accuracy level. Not surprisingly, XG Boosting outperforms ordinary logit and other ML methods based on linear models. Given this empirical pattern, I employ XG Boosting in my first-stage estimates of CCP. I use shallow trees (max depth: 3) to adhere to theoretical rates $O(n^{-1/4})$ and use the "off-the-shelf" *xgboost* package in R. Luo et al. (2016) discuss the theoretical limits for ℓ_2 boosting models.

For comparison, in Appendix Figure C.10, I report the prediction performance in terms of the area under the curve (AUC)[31] in first-stage reduced-form CCP estimates in elderly towns for Bajari et al. (2010b)'s style simple logit and the XG Boosting method. Compared to logit methods, XG Boosting improves the AUC by more than 25%, effectively predicting the rival's stay-in decision. In Appendix Table C.8, I also present confusion matrices for both elderly towns and non-elderly towns. The accuracy in both cases exceeds 0.95, an outstanding performance given the complex nature of the games.

Table 5.6 showcases the primary factors driving outcomes, as identified by the XGBoost model, in both elderly and non-elderly townships. For both town categories, employment numbers by store and rival stores stand out as top determinants. In elderly townships, the presence of chain pharmacies emerges as a notable contributor, while in non-elderly townships, demographic attributes such as the female population and vehicle ownership gain prominence. These differences underscore the heterogeneous socio-economic dynamics at play in each town type.

Second Stage Estimation Results

Table 5.7 and Table 5.8 respectively report the point estimates from the observed sample along with standard errors for elderly and non-elderly towns, respectively. I account for correlation in error terms by taking the clustered standard error at the county level.

[31] AUC denotes the area under the ROC (Receiver Operating Characteristic) curve, where ROC represents the true positive rate against the false positive rate (FPR). The AUC provides an aggregate measure of model performance across all possible classification thresholds. An AUC of 1 indicates a perfect model, while an AUC of 0.5 signifies a model that is no better than random guessing. An AUC below 0.5 indicates the model is performing worse than random guessing.

Table 5.4: First Stage Reduced Form CCP: Classical Method Bajari et al. (2010b)

	(1) Elderly Town	(2) Non-Elderly Town
Chain Pharmacies within 15 mi	-0.421***	-0.145**
	(0.0702)	(0.0537)
Pharmacy's Employee Size	3.396***	3.172***
	(0.390)	(0.509)
Rival's Employee Size	-3.080***	-2.028***
	(0.392)	(0.488)
Total Pop.	0.568***	-0.0480
	(0.155)	(0.188)
Income Per Capita	0.208	-0.527
	(0.304)	(0.475)
Physician Offices	0.0921	0.0101
	(0.0863)	(0.139)
Prop. Age over 65	-0.0954	5.279*
	(1.311)	(2.587)
Prop. Female	-0.382	8.551
	(2.920)	(5.324)
Prop. Black	-4.145	-1.491
	(5.639)	(6.552)
Prop. - High School Graduates	-0.107	-0.943
	(0.812)	(1.424)
Prop. Unemployement	-1.429	-0.109
	(1.309)	(1.545)
Prop. Vehicle = 0	4.491**	1.962+
	(1.389)	(1.145)
Medicaid Expansion	0.0164	0.0462
	(0.0771)	(0.0984)
Prop. Insurance Age over 65	-0.141	-8.509*
	(2.117)	(3.384)
County FE	Yes	Yes
Year FE	Yes	Yes
Observations	20,400	11,640
Mean of Dep. Variable	0.388	0.329
Adjusted R^2	0.165	0.172

Notes: Binary Logit estimates of stay-in/out in township m and year t. Standard errors are clustered at the county level. Significance levels are denoted by + p<0.10, * p<0.05, ** p<0.05.*** p<0.01.

Table 5.5: Performance of Different Methods in First Stage CCP for Hold-Out Sample

Method	County FE & Year FE	Interaction Terms	Prediction Performance (AUC)
Ordinary Logit	Yes	Yes	0.7533
Ridge	Yes	Yes	0.7608
Lasso	Yes	Yes	0.7861
Elastic Net	Yes	Yes	0.7852
XG Boosting	Yes	No	0.9539
Random Forest	Yes	No	0.8320
Support Vector Machine	Yes	No	0.6730

Notes: AUC denotes the area under the ROC(Receiver Operating Characteristic) curve, where ROC denotes the value of the true positive rate against the false positive rate (FPR). The AUC gives an aggregate measure of the model's performance across all possible classification thresholds. An AUC of 1 indicates a perfect model. An AUC of 0.5 indicates a model that is no better than random guessing. If the AUC is less than 0.5, it means the model is performing worse than random guessing.

Table 5.6: Top 10 Importance Features from Xg boosting

(a) Elderly Township

	Feature	Gain
1	#. Employment	0.764
2	#. Rival's Employment	0.138
3	#. Chain Pharmacy	0.019
4	Log(Total Household)	0.010
5	Log(Total Pop.)	0.005
6	Unemployment Rates (%)	0.005
7	Rental Ratio (%)	0.005
8	Female (%)	0.004
9	Age over 65 (%)	0.003
10	Commuting: Walk (%)	0.003

(b) Non-Elderly Township

	Feature	Gain
1	#. Employment	0.738
2	#. Rival's Employment	0.106
3	Female (%)	0.016
4	Vehicle=1 (%)	0.013
5	Rental Ratio (%)	0.010
6	Log(Total Pop.)	0.008
7	Log(Total Household) (%)	0.007
8	Black (%)	0.006
9	Log(Income Per Capita)	0.005
10	High School Graduate (%)	0.005

Notes: Results from Xg Boosting over within sample. I separately estimate in the elderly township/non-elderly township. #. denote the number and % denotes the share of demographic groups out of the total population in the towns.

Table 5.7: Results from the Structural Model: Elderly Town

Parameters	Variables	Bajari et al. (2010b)	Orthogonal Estimators
θ_γ	Rival independent pharmacies	-5.420	-8.055
		(0.685)	(0.495)
θ_c	No. of chain pharmacies	-1.065	-1.138
	(within 15 miles)	(0.085)	(0.057)
Observations		20,400	20,400
Pre-selection of Socio-Economic Controls		Yes	No
Interaction between Socio-Economic Controls		No	Yes
Dimension of Controls		13	563
Counties FE		Yes	Yes
Year FE		Yes	Yes

Notes: I use Lasso estimates for Plug-in estimators with cross-validated penalty parameters from 5-fold cross-fitting Standard errors clustered at the county level. Standard errors are in parenthesis.

Table 5.8: Results from the Structural Model: Non-Elderly Town

Parameters	Variables	Bajari et al. (2010b)	Orthogonal Estimators
θ_γ	Rival independent pharmacies	-4.000	-6.648
		(0.449)	(0.470)
θ_c	No. of chain pharmacies	-0.269	-0.258
	(within 15 miles)	(0.085)	(0.015)
Observations		11,640	11,640
Pre-selection of Socio-Economic Controls		Yes	No
Interaction between Socio-Economic Controls		No	Yes
Dimension of Controls		13	563
Counties FE		Yes	Yes
Year FE		Yes	Yes

Notes: I use Lasso estimates for Plug-in estimators with cross-validated penalty parameters from 5-fold cross-fitting Standard errors are clustered at the county level. Standard errors are in parenthesis.

Interpreting parameters could be challenging as in most discrete-choice models; nonetheless, several empirical findings emerge. The results suggest that the effect of the local rival independent pharmacy appears to be stronger, which aligns with Grieco (2014). I also find that the flexible ML method captures this interaction effect well, which is 1.5 times larger than that identified by Bajari et al. (2010b). Importantly, this observation is consistent for both elderly towns and non-elderly towns, as presented in Table 5.8. Additionally, the estimated results suggest that the number of chain pharmacies has a relatively weak impact on the value of profits for independent pharmacies in the market. Chain effects are less pronounced in non-elderly towns, which aligns with reduced-form evidence. I interpret this as an indication of the elderly population being more price-sensitive, given that chain pharmacies may offer competitive prices.[32]

[32] Several channels exist for this theory: As the national chain has bargaining power over the insurance company, they might offer lower prices. In addition, they can get more discounts through the bulk contracts.

5.5 Robustness Check

I have conducted a series of robustness checks that are not reported in the main text. Since the number of cross-fitting folds, K, does not have a rule of thumb, I experimented with an alternative $K = 4$, which is also widely utilized in the debiased machine learning literature. In Appendix Table C.11, I demonstrate that utilizing a different number of cross-fitting folds yields quite similar results.

To explore the possibility that the choice of hyper-tuning parameters in XG Boosting might not adequately capture beliefs over the rival's choice, I attempted alternative hyper-tuning with cross-validation methods. The results are qualitatively similar.

6 Counterfactual

The structural parameters I have estimated, combined with the underlying structural model, enable me to perform counterfactual experiments. The counterfactual analysis simulates the entry behavior of independent pharmacies to characterize new equilibrium outcomes under different scenarios. As elderly towns have experienced rapid increases in limited access to pharmacies, I focus on elderly towns in my counterfactual scenarios.

6.1 Solution Method for the Static Game

To conduct counterfactuals in different scenarios, I first solve for the equilibrium of the model based on Equation 5.2. I employ a nested fixed-point algorithm, which solves the following system of equations:

$$\sigma_i(a_i = 1|d) = \frac{e^{\sigma_i(a_{-i}=1|d)\theta_\gamma + d_c\theta_c + d_i\theta_{d_i} + d_0\cdot\beta}}{1 + e^{\sigma_i(a_{-i}=1|d)\theta_\gamma + d_c\theta_c + d_i\theta_{d_i} + d_0\cdot\beta}} \tag{6.1}$$

$$\sigma_{-i}(a_{-i} = 1|d) = \frac{e^{\sigma_i(a_i=1|d)\theta_\gamma + d_c\theta_c + d_i\theta_{d_i} + d_0\cdot\beta}}{1 + e^{\sigma_i(a_i=1|d)\theta_\gamma + d_c\theta_c + d_i\theta_{d_i} + d_0\cdot\beta}} \tag{6.2}$$

Here, Equation 6.1 denotes the conditional choice probability (CCP) of player i, and Equation 6.2 denotes the CCP of player $-i$. Given the two unknowns (σ_i and σ_{-i}) and the two equations (Equations 6.1 and 6.2), I apply an iterative method. The iteration continues until the difference between the kth iteration and the $(k+1)$th iteration is less than a tolerance level, $\epsilon = 0.00001$.[33]

[33] I initiate the process with the estimated values of γ_{-i}. As long as I use observed choice probabilities as input, I did not encounter multiple equilibria.

6.2 Goodness of Fit

Figure 6.1 shows the overall predicted and observed number of independent pharmacies between 2000 and 2019. Overall, the simulated outcome captures the downward-sloping trend in the observed number of stores in elderly towns, but the simulated outcome slightly overpredicts after 2016. I also report the average predicted number of stores and the observed number of stores for each town, conditional on various socioeconomic characteristics, in Appendix Table C.12. In line with the overall trends, the predicted averages for stores closely resemble the observed counts.

6.3 Counterfactual 1

In the first counterfactual scenario, I use the counterfactual Bayesian Nash Equilibrium to simulate a situation where chain pharmacies are restricted from expanding starting afterwards 2000. The primary aim of this simulation is to quantify the extent to which independent pharmacies stay in the market in 2019 with the absence of new entry of chain pharmacies after 2000. Table 6.1 reports the results of the counterfactual which highlights two things: 1) The counterfactual predictions suggest that, in the absence of chain pharmacy expansion, the average number of stores in total markets would increase. Counterfactual experiments indicate that, without the expansion of chain pharmacies, there would be a rise in the average number of total market stores. Specifically, the expected store count would see an uptick by 10.40%. Notably, between 2000 and 2019, the count of independent pharmacies dropped by 26.6%. The entry of new chain pharmacies accounted for 40% of this variation.

I also examine the variations in these changes across different market types. M I observe the following: towns with larger populations, higher proportions of elderly residents, and a greater percentage of households without a vehicle would have had more independent stores. This suggests that in towns with an elderly population and limited transportation options, pharmacy accessibility might have been enhanced if chain pharmacies hadn't entered the market after 2000.

6.4 Counterfactual 2: Providing Subsidy associated with Medicare in Elderly Town

The second scenario investigates the potential outcomes of providing subsidies to independent pharmacies in elderly towns. For context, I reference the Health Professional Shortage Area Physician Bonus Program (HPSAPB), which provides a 10% subsidy on Medicare-covered services to physicians in a designated HPSAPB region. Analogously, I explore the possibility

Figure 6.1: Goodness of Fit: Elderly Town

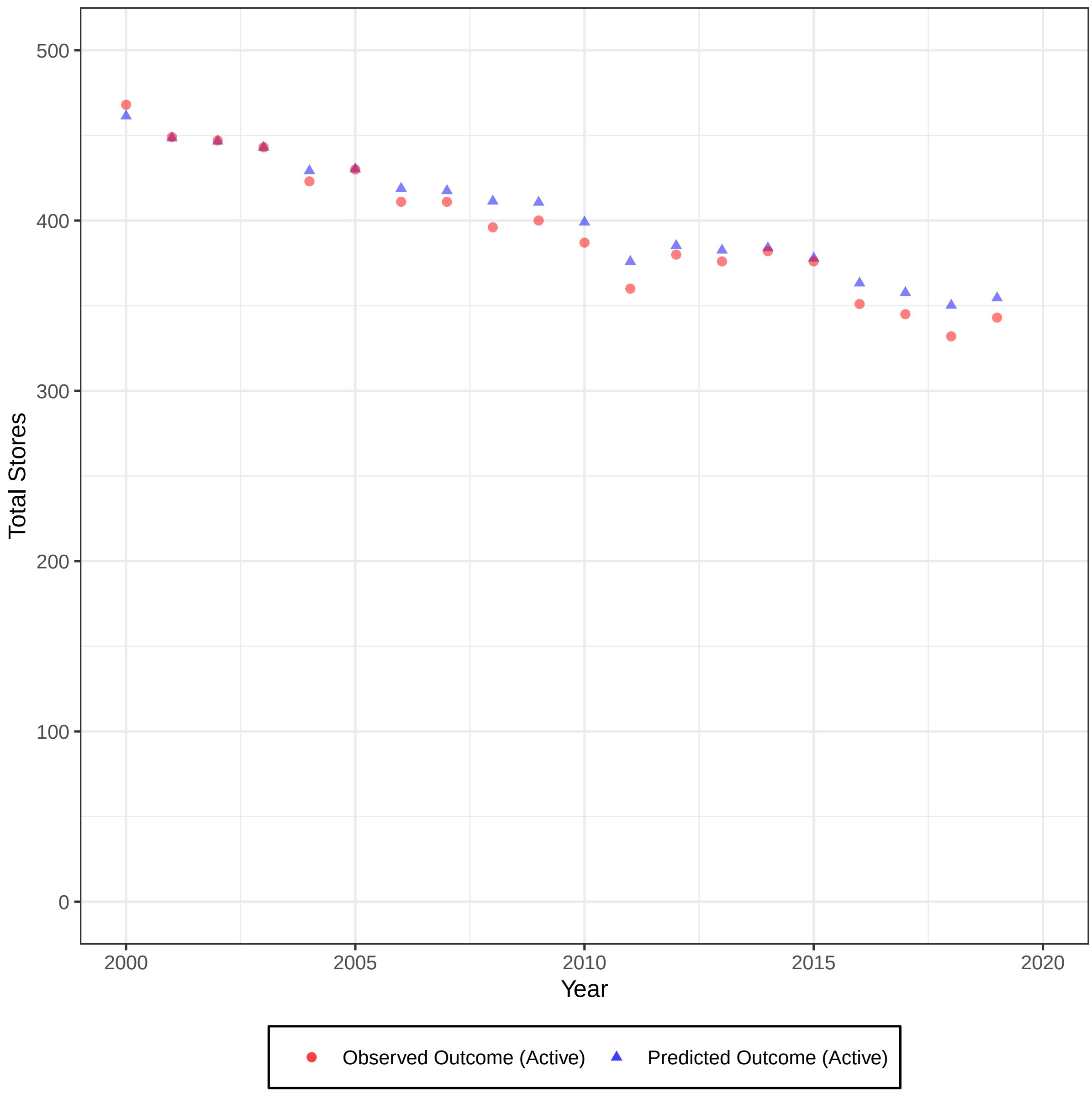


Table 6.1: Expected Number of Stores under Counterfactual Scenario 1 (Year: 2019)

	(Average) Independent Pharmacy Counts			
	Predicted	CF S1	△	△%
Total Markets	0.672	0.742	0.070	10.42
Total Population				
Below median (1,226)	0.588	0.592	0.004	0.68
Above median (1,226)	0.780	0.890	0.110	14.10
Prop. Vehicle=0				
Below median (0.055)	0.668	0.702	0.034	5.09
Above median (0.055)	0.690	0.780	0.090	13.04
Prop. under Poverty Line				
Below median (0.12)	0.640	0.722	0.082	12.81
Above median (0.12)	0.736	0.760	0.024	3.26
Share of Age over 65				
Below median (0.24)	0.682	0.726	0.044	6.45
Above median (0.24)	0.686	0.756	0.070	10.21
Presence of Chain Pharmacy in 2000				
No chain pharmacy within 15 miles	0.700	0.740	0.040	5.71
Chain pharmacy present within 15 miles	0.614	0.750	0.136	22.15
Minority Group				
Below 10%	0.682	0.738	0.056	8.21
Above 10%	0.732	0.800	0.068	9.29

Notes:

of increasing access to pharmacies within elderly towns by providing pharmacists with a 10% subsidy for prescriptions associated with Medicare beneficiaries. To simulate this policy, I factor the subsidies into my estimated latent profits by calibrating the revenue share from Medicare.[34] Subsequently, I calculate the counterfactual CCP using these adjusted profits for independently owned pharmacies.[35]

Table 6.2 show how the expected number of independent pharmacy in 2019 is predicted to change in the counterfactual (CF)- relative to predicted market equilibrium by total market and socio-demographic characteristics respectively. I find that, in the hypothetical world in which independent stores in elderly town gets a 10% subsidy associated with Medicare beneficiaries, markets would have on average 20% more independent pharmacy than the observed number of pharmacies.

To further illustrate how pharmacy accessibility within town would have been improved, I compare the predicted pharmacy accessibility and pharmacy accessibility under scenario 2 in Table 6.3. On average, elderly towns in 2019 with limited pharmacy accessibility would decrease by 5.7% or change in rates by 16.71 %. Interestingly, the effects are largest in locations where the share of minority groups is above 10%, which implies that minority groups will get the most benefits from this suggested subsidy program.

7 Conclusion

This paper successfully addresses the challenge of combining static discrete games with double machine learning (DML) in the context of high-dimensional data. By introducing DML static game estimators, researchers can now obtain reliable inferences, even when dealing with high-dimensional nuisance parameters estimated using machine learning techniques. The results highlight the robustness of the proposed DML static game estimator, which exhibits $\sqrt{N}$-consistency and asymptotic normality. Simulation studies demonstrate the proposed estimators' effectiveness in the unbiased estimation of structural parameters and the validity of inferences.

There are several avenues for future research. One possibility is to examine the relaxation of assumptions about incomplete information structures (e.g., Grieco 2014), building upon the DML framework. Another possibility is to extend DML with nested pseudo-likelihood

[34]Specifically, I reference the average gross markup for independent pharmacies, which stands at 22 percent according to the 2020 National Community Pharmacists Association (NCPA) Digest. This margin typically fluctuates between 22-24 percent annually. Further, I discovered that 30 percent of sales come from Medicare Part D. Taking these factors into account, subsidies result in a 13.5 percent increase in latent profits (0.135 = 1/0.22 (gross mark-up rate) * 0.3 (sales share of Medicare) * 0.1 (subsidy rate).

[35]This framework operates within a partial equilibrium context. It does not consider potential reactions from chain pharmacies or their eligibility for this subsidy program.

Table 6.2: Expected Number of Stores under Counterfactual Scenario 2 (Year: 2019)

	(Average) Independent Pharmacy Counts			
	Predicted	CF S2	△	△%
Total Markets	0.672	0.820	0.148	22.02
Total Population				
Below median (1,226)	0.588	0.686	0.098	16.67
Above median (1,226)	0.780	0.952	0.172	22.05
Prop. Vehicle=0				
Below median (0.055)	0.668	0.812	0.144	21.56
Above median (0.055)	0.690	0.828	0.138	20.00
Prop. under Poverty Line				
Below median (0.12)	0.640	0.796	0.156	24.38
Above median (0.12)	0.736	0.844	0.108	14.67
Share of Age over 65				
Below median (0.24)	0.682	0.796	0.114	16.72
Above median (0.24)	0.686	0.844	0.158	23.03
Presence of Chain Pharmacy in 2000				
No chain pharmacy within 15 miles	0.700	0.842	0.142	20.29
Chain pharmacy present within 15 miles	0.614	0.718	0.104	16.93
Minority Group				
Below 10%	0.682	0.82	0.138	20.23
Above 10%	0.732	0.834	0.102	13.93

Notes:

Table 6.3: Pharmacy Accessibility under Counterfactual Scenario 2 (Year: 2019)

	Limited Pharmacy Accessibility (%)			
	Predicted	CF S2	△	△%
Total Markets	34.1	28.4	-5.7	-16.71
Total Population				
Below median (1,226)	40.0	35.3	-4.7	-11.76
Above median (1,226)	28.2	21.6	-6	-23.61
Prop. Vehicle=0				
Below median (0.055)	35.7	31	-4.7	-13.19
Above median (0.055)	32.5	25.9	-6.6	-20.48
Prop. under Poverty Line				
Below median (0.12)	38	31.7	-6.28	-13.45
Above median (0.12)	30.2	25.1	-5.1	-16.88
Share of Age over 65				
Below median (0.24)	33.7	29.0	-4.7	-13.95
Above median (0.24)	34.5	27.8	-6.6	-19.32
Presence of Chain Pharmacy in 2000				
No chain pharmacy within 15 miles	32.9	27.1	-5.8	-17.65
Chain pharmacy present within 15 miles	39.6	34.3	-5.2	-13.16
Minority Group				
Below 10%	35.6	30	-5.6	-15.98
Above 10%	16.6	10	-6.6	-40.00

Notes:

(NPL) methods, following the work of Aguirregabiria and Mira (2002) and Aguirregabiria and Mira (2007). These topics are left open for future research.

References

Ackerberg, Daniel A and Gautam Gowrisankaran, "Quantifying equilibrium network externalities in the ACH banking industry," *The RAND journal of economics*, 2006, *37* (3), 738–761.

Adusumilli, Karun and Dita Eckardt, "Temporal-Difference estimation of dynamic discrete choice models," *arXiv preprint arXiv:1912.09509*, 2019.

Aguirregabiria, Victor and Pedro Mira, "Swapping the nested fixed point algorithm: A class of estimators for discrete Markov decision models," *Econometrica*, 2002, *70* (4), 1519–1543.

_ **and** _ , "Sequential estimation of dynamic discrete games," *Econometrica*, 2007, *75* (1), 1–53.

Amstislavski, Philippe, Ariel Matthews, Sarah Sheffield, Andrew R Maroko, and Jeremy Weedon, "Medication deserts: survey of neighborhood disparities in availability of prescription medications," *International journal of health geographics*, 2012, *11* (1), 1–13.

Aradillas-López, Andres and Amit Gandhi, "Estimation of games with ordered actions: An application to chain-store entry," *Quantitative Economics*, 2016, *7* (3), 727–780.

Arcidiacono, Peter and Robert A Miller, "Conditional choice probability estimation of dynamic discrete choice models with unobserved heterogeneity," *Econometrica*, 2011, *79* (6), 1823–1867.

Bajari, Patrick, C Lanier Benkard, and Jonathan Levin, "Estimating dynamic models of imperfect competition," *Econometrica*, 2007, *75* (5), 1331–1370.

_ **, Han Hong, and Denis Nekipelov**, "Game theory and econometrics: A survey of some recent research," in "Advances in economics and econometrics, 10th world congress," Vol. 3 2013, pp. 3–52.

_ **,** _ **, and Stephen P Ryan**, "Identification and estimation of a discrete game of complete information," *Econometrica*, 2010, *78* (5), 1529–1568.

_ **,** _ **, John Krainer, and Denis Nekipelov**, "Estimating static models of strategic interactions," *Journal of Business & Economic Statistics*, 2010, *28* (4), 469–482.

_ , **Victor Chernozhukov, Han Hong, and Denis Nekipelov**, "Nonparametric and semiparametric analysis of a dynamic discrete game," *Unpublished manuscript, Dep. Econ., Univ. Minn., Minneapolis*, 2009.

Belloni, Alexandre, Daniel Chen, Victor Chernozhukov, and Christian Hansen, "Sparse models and methods for optimal instruments with an application to eminent domain," *Econometrica*, 2012, *80* (6), 2369–2429.

_ , **Victor Chernozhukov, and Ying Wei**, "Post-selection inference for generalized linear models with many controls," *Journal of Business & Economic Statistics*, 2016, *34* (4), 606–619.

Berry, Steven T, "Estimation of a Model of Entry in the Airline Industry," *Econometrica: Journal of the Econometric Society*, 1992, pp. 889–917.

_ **and Giovanni Compiani**, "An instrumental variable approach to dynamic models," *The Review of Economic Studies*, 2023, *90* (4), 1724–1758.

Borusyak, Kirill, Xavier Jaravel, and Jann Spiess, "Revisiting event study designs: Robust and efficient estimation," *arXiv preprint arXiv:2108.12419*, 2021.

Bresnahan, Timothy F and Peter C Reiss, "Entry and competition in concentrated markets," *Journal of political economy*, 1991, *99* (5), 977–1009.

Callaway, Brantly and Pedro HC Sant'Anna, "Difference-in-differences with multiple time periods," *Journal of econometrics*, 2021, *225* (2), 200–230.

Chernozhukov, Victor, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins, "Double/debiased machine learning for treatment and structural parameters," *The Econometrics Journal*, 01 2018, *21* (1), C1–C68.

_ , **Juan Carlos Escanciano, Hidehiko Ichimura, Whitney K Newey, and James M Robins**, "Locally Robust Semiparametric Estimation. arXiv e-prints, page," *arXiv preprint arXiv:1608.00033*, 2016.

_ , _ , _ , _ , **and** _ , "Locally robust semiparametric estimation," *Econometrica*, 2022, *90* (4), 1501–1535.

_ , **Whitney K Newey, and Rahul Singh**, "Learning L2-continuous regression functionals via regularized riesz representers," *arXiv preprint arXiv:1809.05224*, 2018, *8.*

— , — , **Victor Quintas-Martinez, and Vasilis Syrgkanis**, "Automatic debiased machine learning via neural nets for generalized linear regression," *arXiv preprint arXiv:2104.14737*, 2021.

de Chaisemartin, ClÃ©ment and Xavier D'Haultfoeuille, "Difference-in-Differences Estimators of Intertemporal Treatment Effects," 2023.

Dearing, Adam and Jason R Blevins, "Efficient and convergent sequential pseudo-likelihood estimation of dynamic discrete games," *arXiv preprint arXiv:1912.10488*, 2019.

der Vaart, Aad W Van, *Asymptotic statistics*, Vol. 3, Cambridge university press, 2000.

Elabed, Azam, Albert Wertheimer, and Sabri Ibrahim, "Opening A New Independent Pharmacy 101," *INNOVATIONS in pharmacy*, 2016, *7* (1).

Ellickson, Paul B and Sanjog Misra, "Supermarket pricing strategies," *Marketing science*, 2008, *27* (5), 811–828.

— **and** — , "Structural workshop paper-Estimating discrete games," *Marketing Science*, 2011, *30* (6), 997–1010.

Ericson, Richard and Ariel Pakes, "Markov-perfect industry dynamics: A framework for empirical work," *The Review of economic studies*, 1995, *62* (1), 53–82.

Goodman-Bacon, Andrew, "Difference-in-differences with variation in treatment timing," *Journal of Econometrics*, 2021, *225* (2), 254–277.

Grieco, Paul LE, "Discrete games with flexible information structures: An application to local grocery markets," *The RAND Journal of Economics*, 2014, *45* (2), 303–340.

Holmes, Thomas J, "The diffusion of Wal-Mart and economies of density," *Econometrica*, 2011, *79* (1), 253–302.

Hotz, V Joseph and Robert A Miller, "Conditional choice probabilities and the estimation of dynamic models," *The Review of Economic Studies*, 1993, *60* (3), 497–529.

Ichimura, Hidehiko and Whitney K Newey, "The influence function of semiparametric estimators," *Quantitative Economics*, 2022, *13* (1), 29–61.

Igami, Mitsuru and Kosuke Uetake, "Mergers, innovation, and entry-exit dynamics: Consolidation of the hard disk drive industry, 1996–2016," *The Review of Economic Studies*, 2020, *87* (6), 2672–2702.

Jia, Panle, "What happens when Wal-Mart comes to town: An empirical analysis of the discount retailing industry," *Econometrica*, 2008, *76* (6), 1263–1316.

Koh, Paul S, "Stable outcomes and information in games: An empirical framework," *Journal of Econometrics*, 2023, *237* (1), 105499.

Lepoev, Strahil, "Welfare Effects of Partial Acquisition in the Retail Pharmacy Market - the Case of Rite Aid and Walgreens," *Unpublished manuscript, Dep. Econ., Johns Hopkins University*, 2023.

Luo, Ye, Martin Spindler, and Jannis Kück, "High-Dimensional L_2 Boosting: Rate of Convergence," *arXiv preprint arXiv:1602.08927*, 2016.

Manski, Charles F, "Identification of endogenous social effects: The reflection problem," *The review of economic studies*, 1993, *60* (3), 531–542.

Mather, Mark, Linda A. Jacobsen, and Kelvin M. Pollard, "Aging in the United States," *Population Bulletin*, 2015, *70* (2).

Mazzeo, Michael J, "Product choice and oligopoly market structure," *RAND Journal of Economics*, 2002, pp. 221–242.

McKelvey, Richard D and Thomas R Palfrey, "Quantal response equilibria for normal form games," *Games and economic behavior*, 1995, *10* (1), 6–38.

Nekipelov, Denis, Vira Semenova, and Vasilis Syrgkanis, "Regularised orthogonal machine learning for nonlinear semiparametric models," *The Econometrics Journal*, 2022, *25* (1), 233–255.

Newey, Whitney K, "The asymptotic variance of semiparametric estimators," *Econometrica: Journal of the Econometric Society*, 1994, pp. 1349–1382.

_ **and Daniel McFadden**, "Large sample estimation and hypothesis testing," *Handbook of econometrics*, 1994, *4*, 2111–2245.

Neyman, Jerzy, "Optimal asymptotic tests of composite hypotheses," *Probability and statsitics*, 1959, pp. 213–234.

Nishida, Mitsukuni, "Estimating a model of strategic network choice: The convenience-store industry in Okinawa," *Marketing Science*, 2015, *34* (1), 20–38.

Novi, Cinzia Di, Lucia Leporatti, and Marcello Montefiori, "Older patients and geographic barriers to pharmacy access: When nonadherence translates to an increased use of other components of health care," *Health economics*, 2020, *29*, 97–109.

OFT, "The Control of Entry Regulations and Retail Pharmacy Services in the UK.," Technical Report OFT609, Report of Fair Trading 2003.

Pesendorfer, Martin and Philipp Schmidt-Dengler, "Asymptotic least squares estimators for dynamic games," *The Review of Economic Studies*, 2008, *75* (3), 901–928.

Qato, Dima M, Martha L Daviglus, Jocelyn Wilder, Todd Lee, Danya Qato, and Bruce Lambert, "'Pharmacy deserts' are prevalent in Chicago's predominantly minority communities, raising medication access concerns," *Health Affairs*, 2014, *33* (11), 1958–1965.

Rust, John, "Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher," *Econometrica: Journal of the Econometric Society*, 1987, pp. 999–1033.

Ryan, Stephen P, "The costs of environmental regulation in a concentrated industry," *Econometrica*, 2012, *80* (3), 1019–1061.

Schaumans, Catherine and Frank Verboven, "Entry and regulation: evidence from health care professions," *The Rand journal of economics*, 2008, *39* (4), 949–972.

Seim, Katja, "An empirical model of firm entry with endogenous product-type choices," *The RAND Journal of Economics*, 2006, *37* (3), 619–640.

Semenova, Vira, "Machine learning for dynamic discrete choice," *arXiv preprint arXiv:1808.02569*, 2018.

Sun, Liyang and Sarah Abraham, "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects," *Journal of Econometrics*, 2021, *225* (2), 175–199.

Syrgkanis, Vasilis and Manolis Zampetakis, "Estimation and inference with trees and forests in high dimensions," in "Conference on learning theory" PMLR 2020, pp. 3453–3454.

Tamer, Elie, "Incomplete simultaneous discrete response model with multiple equilibria," *The Review of Economic Studies*, 2003, *70* (1), 147–165.

Xie, Erhao, "Inference in Games Without Equilibrium Restriction: An Application to Restaurant Competition in Opening Hours," *Journal of Business & Economic Statistics*, 2022, *40* (4), 1803–1816.

A Appnedix: Proof

Proof of Lemma ??.

$$E[\psi(w_i;\theta_0,\eta_0)] = E\Big[\underbrace{(z_i - E[z_i|x])\,[a_i - \Lambda(\gamma_{-i}\theta, X\beta)]}_{\text{Part A}} - \underbrace{\alpha[a_{-i} - \gamma_{-i}]}_{\text{Part B}}\Big]$$

Part A is equal to zero by the law of iterated expectations.

$$\begin{aligned}
&E[z_i[a_i - \Lambda(\gamma_{-i}\theta, X\beta)]] - E[E[z_i|x](a_i - \Lambda(\gamma_{-i}\theta, X\beta))] \\
&= 0 + E[E[z_i|x](a_i - \Lambda(\gamma_{-i}\theta, X\beta))] \ (\because \text{original moment is equal to zero.}) \\
&= E[E[z_i|x](E[a_i|d_i, d_{-i}, x] - \Lambda(\gamma_{-i}\theta X\beta))] \\
&= 0 \ (\because \Lambda = Pr(a_i = 1|d_i, d_{-i}, x))
\end{aligned}$$

Also, Part B equals zero by the law of iterated expectations.

$$E[\alpha[a_{-i} - \gamma_{-i}(d_{-i}, X)] = E[E[\alpha|d_{-i}, X][E[a_{-i}|d_{-i}, X] - E[a_{-i}|d_{-i}, X]] = 0$$

The conclusion follows as the sum of parts A and B equals zero. $Q.E.D.$

Proof of Theorem 3.1.

$$\begin{aligned}
\frac{\partial\psi(w_i,\theta_0;\gamma_{-i0},\beta,\mu_{z0})}{\partial\beta} &= E\left[-\mu_z\Lambda' x\cdot(\beta-\beta_0)\right] = -E[u_i f_i x_i\cdot(\beta-\beta_0)] = 0 \ (\because \text{equation } 3.4) \\
\frac{\partial\psi(w_i,\theta_0;\gamma_{-i},\beta_0,\mu_{z0})}{\partial\gamma_{-i}} &= E[(-\mu_z\Lambda'\theta_\gamma + \alpha)\cdot(\gamma_{-i}-\gamma_{-i0})] \\
&= E[(-\mu_z\Lambda'\theta_\gamma + \mu_z\Lambda'\theta_\gamma, \mu_{z0})\cdot(\gamma_{-i}-\gamma_{-i0})] = 0. \\
\frac{\partial\psi(w_i,\theta_0;\gamma_{-i},\beta_0,\mu_z)}{\partial\mu_z} &= E[a_i - \Lambda(\theta_\gamma\gamma_{-i0}, X\beta_0)\cdot(\mu_z-\mu_{z0})] = 0.
\end{aligned}$$

$Q.E.D.$

Proofs of Theorem 3.2.

For clarity, I re-state the Assumption made in Chernozhukov et al. (2022) and verify that these assumptions are satisfied.

Assumption A.1 (Mean-Square Consistency)**.**

$E[||\psi(W, \theta_0, \gamma_0, \alpha_0)||^2] < \infty$ and

$$(i) \int ||m(w, \hat{\gamma}_\ell, \theta_0) - m(w, \gamma_0, \theta_0)||^2 F_0(dw) \xrightarrow{p} 0,$$
$$(ii) \int ||\phi(w, \hat{\gamma}_\ell, \alpha_0, \theta_0) - \phi(w, \gamma_0, \alpha_0, \theta_0)||^2 F_0(dw) \xrightarrow{p} 0,$$
$$(iii) \int ||\phi(w, \gamma_0, \hat{\alpha}_\ell, \tilde{\theta}_\ell) - \phi(w, \gamma_0, \alpha_0, \theta_0)||^2 F_0(dw) \xrightarrow{p} 0.$$

To give mild mean-square consistency conditions for $\hat{\gamma}_\ell$ and $(\hat{\alpha}_\ell, \tilde{\theta}_\ell)$ separately. I denote

$$\hat{\Delta}_\ell(w) := \phi(w, \hat{\gamma}_\ell, \hat{\alpha}_\ell, \tilde{\theta}_\ell) - \phi(w, \gamma_0, \hat{\alpha}_\ell, \tilde{\theta}_\ell) - \phi(w, \hat{\gamma}_\ell, \alpha_0, \theta_0) + \phi(w, \gamma_0, \alpha_0, \theta_0).$$

Assumption A.2 (Convergence Rate for Interaction Remainder)**.**

For each $\ell = 1, ..., L$,

$$\sqrt{n} \int \hat{\Delta}_\ell(w) F_0(dw) \xrightarrow{p} 0.$$

and

$$\int ||\hat{\Delta}_\ell(w)||^2 F_0(dw) \xrightarrow{p} 0.$$

Assumption A.3 (Convergence Rates for γ)**.**

For each $\ell = 1, ..., L$,

$||\hat{\gamma}_\ell - \gamma_0|| = o_p(n^{-1/4})$ *and* $||\bar{\psi}(\gamma, \alpha_0, \theta_0)|| \leq C||\gamma - \gamma_0||^2$ *for all* γ *with* $||\gamma - \gamma_0||$ *small enough.*

Assumption A.4.

For each $\ell = 1, ...L$,

$$\int ||m(w, \hat{\gamma}_\ell, \tilde{\theta}_\ell) - m(w, \hat{\gamma}_\ell, \theta_0)||^2 F_0(dw) \xrightarrow{p} 0 \text{ and } \int ||\hat{\Delta}_\ell||^2 F_0(dw) \xrightarrow{p} 0$$

Assumption A.5 (Convergence of the Jacobian)**.** *G exsits and there is a neighborhood* $\mathcal{N}$ *of* θ_0 *and* $||\cdot||$ *such that (i) for each* ℓ, $||\hat{\gamma}_{h\ell} - \gamma_0|| \xrightarrow{p} 0$; *(ii) for all* $||\hat{\gamma}_\ell - \gamma_{h0}||$ *small enough,* $m(W, \gamma, \theta)$ *is differentiable in* θ *on* $\mathcal{N}$ *with probability approaching 1 and there are* $C > 0$ *and* $d(W, \gamma)$ *such that, for* $\theta \in \mathcal{N}$ *and* $||\hat{\gamma} - \gamma_0||$ *small enough,*

$$\left|\left|\frac{\partial m(W, \hat{\gamma}, \hat{\theta})}{\partial \theta} - \frac{\partial m(W, \hat{\gamma}, \theta_0)}{\partial \theta}\right|\right| \leq d(W, \gamma)||\hat{\theta} - \theta_0||^{1/C}; \; E[d(W, \hat{\gamma})] < C$$

(iii) For each $\ell = 1, ...L, j$ *and* k, $\int |\partial g_j(w, \hat{\gamma}, \theta_0)/\theta_k - \partial g_j(w, \gamma_0, \theta_0)/\partial_k| F_0(dw) \xrightarrow{p} 0$

B Proof

Proof of Assumption A.1:

Assumption A.1 part (i) is implied by Assumption 3.2. For $k = 1, .., K$, let $\phi_k(w, \gamma_k, \alpha_k) = \alpha_k[y_{-ik} - \gamma_k]$ and

$$\phi(w, \gamma, \alpha, \theta) = \sum_{k=1}^{K} \phi_k(w, \gamma_k, \alpha_k, \theta).$$

For part (ii), by the assumption 3.2

$$\int ||\phi(w, \hat{\gamma}_\ell, \alpha, \theta_0) - \phi(w, \gamma_0, \alpha_0, \theta_0)||^2 F_0(dw) = \int ||\alpha_0^2[\hat{\gamma}_{h\ell} - \gamma_{h0}]^2||F_0(dw)$$
$$\leq C||\hat{\gamma}_{k\ell} - \gamma_{h0}||^2 \xrightarrow{p} 0.$$

Assumptions A.1 part (ii) holds by the triangle inequality.

For part (iii), again, by the assumption 3.2

$$\int ||\phi(w, \gamma_0, \hat{\alpha}, \tilde{\theta}) - \phi(w, \gamma_0, \alpha_0, \theta_0)||^2 F_0(dw) = \int ||(\hat{\alpha}_{h\ell} - \alpha_{h0})^2[a_{-i} - \gamma_{h0}]^2||F_0(dw)$$
$$\leq C||\alpha_{h\ell} - \alpha_{h0}||^2 \xrightarrow{p} 0.$$

Assumptions A.1 part (iii) holds by the triangle inequality.

The following lemma gives a convergence rate for the preliminary naive plug-in estimator of $\hat{\alpha}$

Lemma B.1. If Assumption 3.2 holds, then,

$$\hat{\theta}_\ell = \theta_0 + O_p(n^{-d_1}).$$

Proof of Lemma B.1: Similar to Chernozhukov et al. (2022), the convergence rate for quasi maximum likelihood would be slower by the convergence rate for nuisance parameters γ.

Next, the following lemma gives a convergence rate for the unknown function $\hat{\alpha}$.

Lemma B.2. If Assumption 3.2 holds, then

$$|\hat{\alpha}_{k\ell} - \alpha_{h0}| = O_p(n^{-d_1})$$

Proof of Lemma B.2:

$$|\hat{\alpha}_{h\ell} - \alpha_{h0}| \leq C|\hat{\theta}_\gamma - \theta_0| + |\hat{\gamma}_{h\ell} - \gamma_{h0}| + |\hat{\mu}_\ell - \mu_0| \leq O_p(n^{-d_1}).(\because \text{triangle inequality}).$$

Proof of Assumption A.2: For part (i), I observe that

$$\hat{\Delta}_\ell = \sum_h (\hat{\alpha}_{h\ell} - \alpha_{h0})(\hat{\gamma}_{h\ell} - \gamma_{h0})$$

Then, by the conclusion of Lemma B.2 and assumption 3.1,

$$\begin{aligned}
&\sqrt{n}\int ||\hat{\Delta}_\ell|| F_0(dw) \leq \sum_h \sqrt{n}||\hat{\alpha}_{h\ell} - \alpha_{h0}||\ ||\hat{\gamma}_{h\ell} - \gamma_{h0}|| \\
&= O_p(\sqrt{n}n^{-d_1}\ n^{-d_1}) = o_p(1)\ (\because -\frac{1}{2} < -2d_1 < -1) \text{ by the assumption ??}.
\end{aligned}$$

For part (ii),

$$\int ||\hat{\Delta}||^2 F_0(dw) \leq \sum_h ||\hat{\alpha}_{h\ell} - \alpha_{h0}||^2 ||\hat{\gamma}_{h\ell} - \gamma_{h0}||^2 = O_p(n^{-2d_1}) = o_p(1)$$

It follows that triangle inequality.

Proof of Assumption A.3: The first condition follows by assumption 3.1 for Lasso and other machine learning methods. For the second condition, taking Taylor approximation for nuisance parameters $\hat{\gamma} = (\hat{\gamma}_{-ik}, \hat{\mu}, \hat{\beta})$:

$$\begin{aligned}
\bar{\psi}(w, \hat{\gamma}, \alpha_0, \theta_0) &:= \int \hat{\mu}\{a_i - M(X\hat{\beta}, \theta\hat{\gamma}_{-i})\} + \int \alpha_k[a_{-i} - \hat{\gamma}_{-i}]F_0(dw), \\
&= \int (\hat{\mu} - \mu_0)[y - \hat{G}]F_0(dw) + \int \hat{\mu}\hat{G}'(\hat{\beta} - \beta_0)F_0(dw) + \int \hat{\mu}\hat{G}'(\hat{\gamma}_{-i} - \gamma_{-i0})F_0(dw) \\
&\leq O_p(n^{-2d_1}) + O_p(n^{-2d_1}) + O_p(n^{-2d_1}) = O_p(n^{-2d_1}) = C||\hat{\gamma} - \gamma_0||^2.
\end{aligned}$$

Proof of Assumption A.4: Similar to Chernozhukov et al. (2022), the first condition of Assumption 4 can be deduced from the convergence of the probabilities towards 1 and the uniform boundedness of the $\hat{\gamma}_{h\ell}$. The second condition follows by the proof of part ii) in Assumption A.3.

Proof of Assumption A.5: Following Chernozhukov et al. (2022), Assumption 5 can be deduced from the previously given boundedness properties. Finally, the conclusion follows by Theorem 9 in Chernozhukov et al. (2022). *Q.E.D.*

Proof of Lemma ??.

$$E[\psi(w_i;\theta_0,\eta_0)] = E\Big[\underbrace{(z_i - E[z_i|x_i])\,[a_i - M(\gamma_{-i}\theta, X\beta)]}_{\text{Part A}} - \underbrace{\sum_{a_{-i}}\prod_{s\neq i,-i}\alpha_{-i}[a_{-i}-\gamma_{-i}]}_{\text{Part B}}\Big]$$

Part A is equal to zero by the law of iterated expectations.

$$\begin{aligned}
&E[z_i[a_i - \Lambda(\gamma_{-i}\theta, X\beta)]] - E[E[z_i|x](a_i - \Lambda(\gamma_{-i}\theta, X\beta))]\\
&= 0 + E[E[z_i|x](a_i - \Lambda(\gamma_{-i}\theta, X\beta))] \text{ } (\because \text{original moment is equal to zero.})\\
&= E[E[z_i|x](E[a_i|d_i, d_{-i}, x] - \Lambda(\gamma_{-i}\theta X\beta))]\\
&= 0 \text{ } (\because \Lambda = Pr(a_i = 1|d_i, d_{-i}, x))
\end{aligned}$$

Also, Part B equals zero by the law of iterated expectations.

$$E[\prod_{s\neq i,-i}\alpha_{-i}[a_{-i}-\gamma_{-i}(d_{-i},X)] = E[E[\prod_{s\neq i,-i}\alpha_{-i}|d_{-i},X][E[a_{-i}|d_{-i},X] - E[a_{-i}|d_{-i},X]] = 0$$

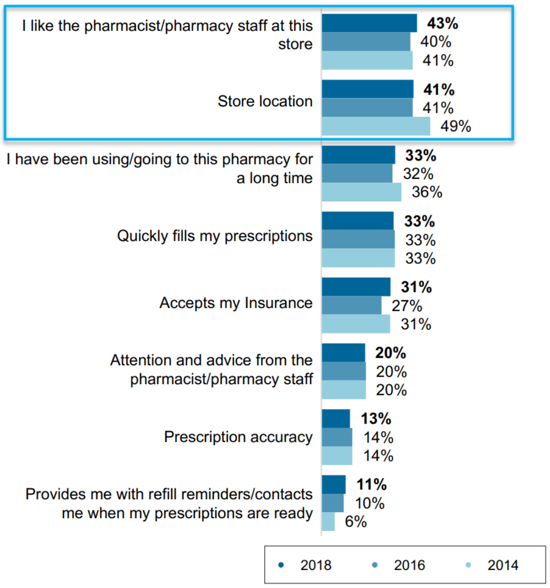
Proof of Theorem 3.3.

$$\begin{aligned}
\frac{\partial\psi(w_i,\theta_0;\gamma_{-i0},\beta,\mu_{z0})}{\partial\beta} &= E\left[-\mu_z\Lambda' x\cdot(\beta-\beta_0)\right] = -E[u_i f_i x_i\cdot(\beta-\beta_0)] = 0 \text{ } (\because \text{equation 3.4})\\
\frac{\partial\psi(w_i,\theta_0;\gamma_{-i},\beta_0,\mu_{z0})}{\partial\gamma_{-i}} &= E[(-\sum_{a_{-i}}\prod_{s\neq i,-i}(1-\gamma_s) - \alpha)\cdot(\gamma_{-i}-\gamma_{-i0})]\\
&= E[(-\sum_{a_{-i}}\prod_{s\neq i,-i}(1-\gamma_s) + \sum_{a_{-i}}\prod_{s\neq i,-i}(1-\gamma_s))\cdot(\gamma_{-i}-\gamma_{-i0})] = 0.\\
\frac{\partial\psi(w_i,\theta_0;\gamma_{-i},\beta_0,\mu_z)}{\partial\mu_z} &= E[a_i - \Lambda(\theta_\gamma\gamma_{-i0}x\beta_0)\cdot(\mu_z-\mu_{z0})] = 0.
\end{aligned}$$

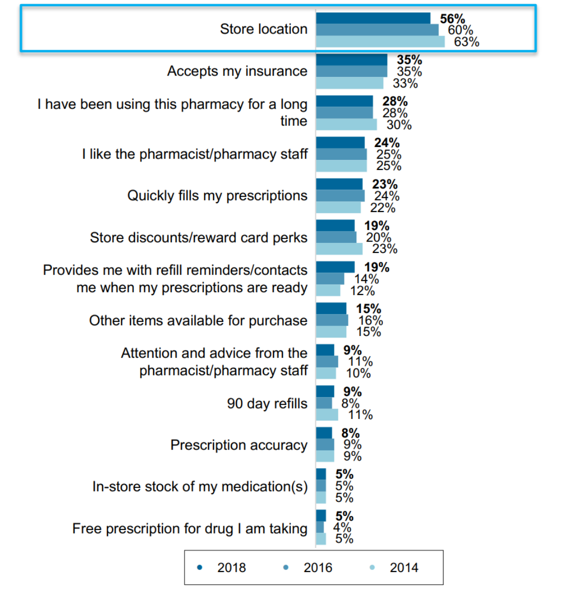
Q.E.D.

C Appendix: Supplemental Figure and Tables

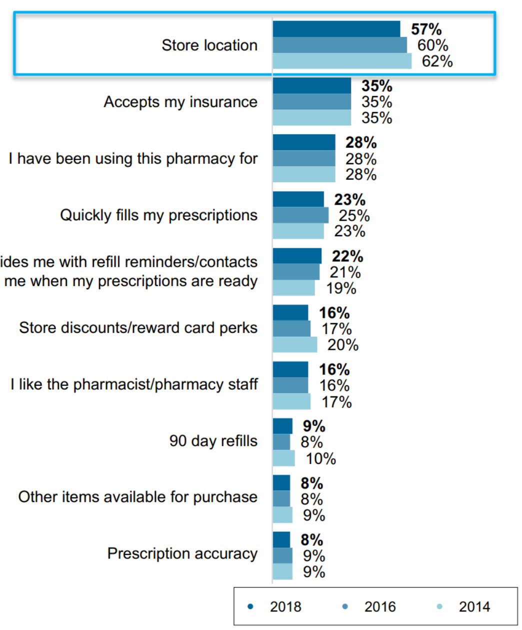
Figure C.1: Industry Survey: Main Reason for Choosing Primary Pharmacy



(a) Independent Pharmacy



(b) Supermarket Pharmacy



(c) Chain Pharmacy

Sources: Pharmacy Satisfaction Data Summary Report, 2018 Boehringer Ingelheim Pharmaceuticals, Inc

Figure C.2: Industry Survey: Main Reason for Switching Primary Pharmacy

	Total 2018 (n=2,244)	Total 2016 (n=2,035)
Close to my house/work	**33%**	34%
Health insurance requires me to use this pharmacy	**21%**	19%
I moved	**18%**	20%
90 day refills	**18%**	19%
Like the pharmacy staff	**18%**	13%
Refill reminders/contacts me when Rx ready	**17%**	14%
Has other items available for purchase	**12%**	12%
Offers store discounts/reward card perks	**10%**	9%
Other	**16%**	19%

Sources: Pharmacy Satisfaction Data Summary Report, 2018 Boehringer Ingelheim Pharmaceuticals, Inc

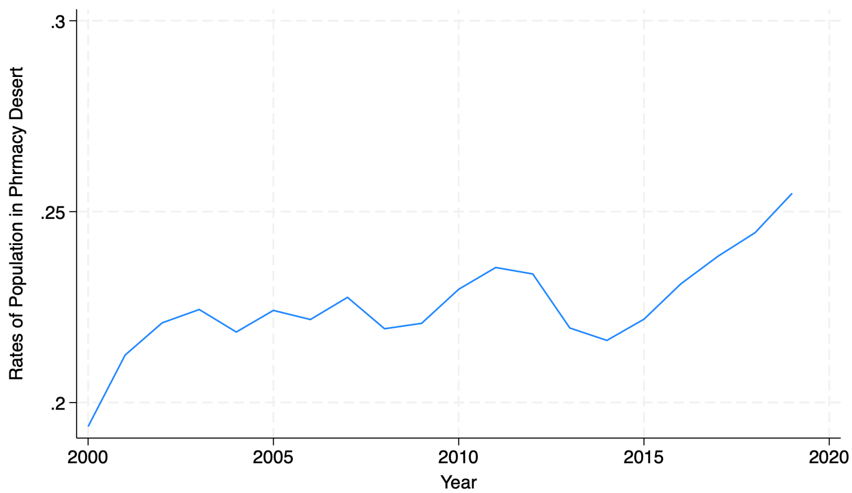
Tables and Figures

Table C.1: Simulation Results
(Sample Size: 1,000, Dimension of d_0: 500)

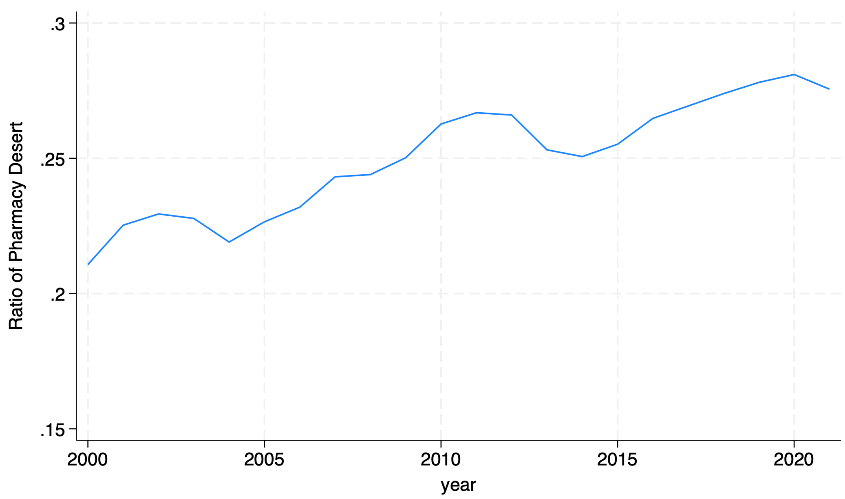
		Oracle				Naive				Orthogonal			
	DGP (1)	Estimates (2)	Bias(%) (3)	CI (4)	RMSE (5)	Estimates (6)	Bias(%) (7)	CI (8)	RMSE (9)	Estimates (10)	Bias(%) (11)	CI (12)	RMSE (13)
θ_r	-1.5	-1.595	-6.323	0.960	0.522	-1.013	32.5	0.736	1.025	-1.529	-1.901	0.880	0.936
		(0.410)				(0.600)				(0.615)			

Notes: Mean and Standard Deviation for 500 simulations. Column (1) represents the true value for the rival effect under the model. Columns (2)-(5) used Oracle and columns (6)-(9) used Naive plug-in estimators. Columns (10)-(13) used developed Neyman orthogonal moments. Column (2), (6), and (10) shows the mean and (median) standard errors for the estimated parameters. CI denotes the probability of a 95 percent confidence interval based on standard errors from 500 simulations. RMSE denotes the root mean square error between estimated parameters and true parameters.

Figure C.3: Trends in Pharmacy Deserts: Alternative Definition



(a) Population Weight-Average



(b) Within 5 miles

Notes: The figures depict trends in pharmacy deserts using alternative definitions. The units of observation are based on a three-year moving average of the pharmacy desert indicator for a final sample of 802 townships. In the left panel (a), the pharmacy desert indicator takes a value of 1 if townships have at least one independent or chain pharmacy, and it's weighted by the township's population. In the right panel (b), the pharmacy desert indicator takes a value of 1 if there are no both independent and chain pharmacies within a 5-mile radius.

Table C.2: Simulation Results
(Sample Size: 1,500, Dimension of d_0: 500)

		Oracle				Naive				Orthogonal			
	DGP (1)	Estimates (2)	Bias(%) (3)	CI (4)	RMSE (5)	Estimates (6)	Bias(%) (7)	CI (8)	RMSE (9)	Estimates (10)	Bias(%) (11)	CI (12)	RMSE (13)
θ_r	-1.5	-1.562	-4.151	0.956	0.388	-0.913	39.143	0.668	0.872	-1.456	2.947	0.872	0.627
		(0.329)				(0.464)				(0.472)			

Notes: Mean and Standard Deviation for 500 simulations. Column (1) represents the true value for the rival effect under the model. Columns (2)-(5) used Oracle and columns (6)-(9) used Naive plug-in estimators. Columns (10)-(13) used developed Neyman orthogonal moments. Column (2), (6), and (10) shows the mean and (median) standard errors for the estimated parameters. CI denotes the probability of a 95 percent confidence interval based on standard errors from 500 simulations. RMSE denotes the root mean square error between estimated parameters and true parameters.

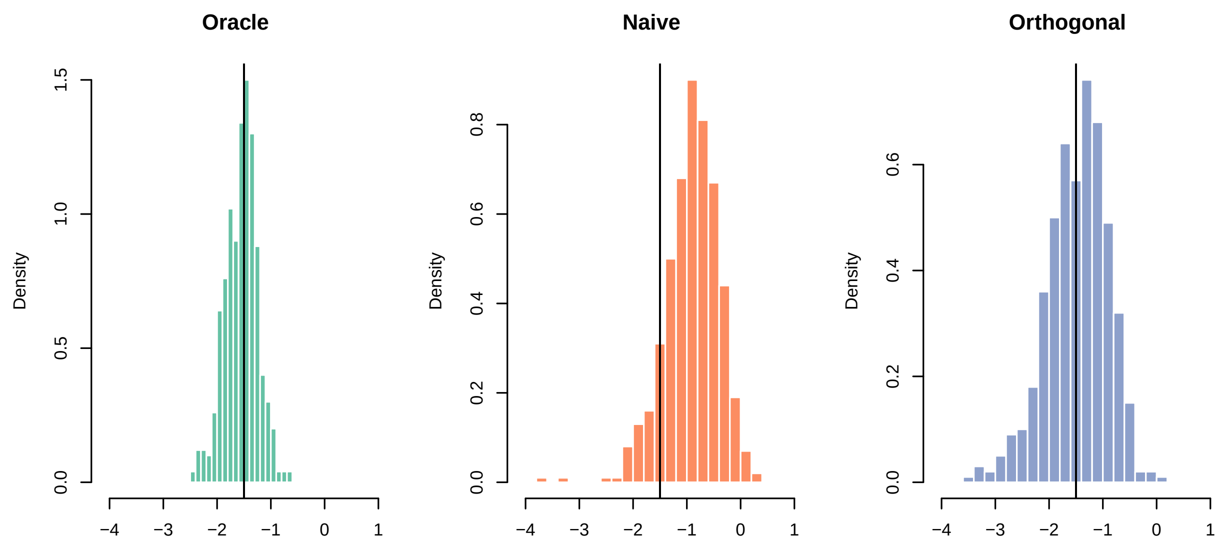
Table C.3: Simulation Results
(Sample Size: 2,000, Dimension of d_0: 500)

		Oracle				Naive				Orthogonal			
	DGP (1)	Estimates (2)	Bias(%) (3)	CI (4)	RMSE (5)	Estimates (6)	Bias(%) (7)	CI (8)	RMSE (9)	Estimates (10)	Bias(%) (11)	CI (12)	RMSE (13)
θ_r	-1.5	-1.545	-3.007	0.958	0.307	-0.905	-39.691	0.608	0.775	-1.477	1.526	0.858	0.566
		(0.289)				(0.402)				(0.407)			

Notes: Mean and Standard Deviation for 500 simulations. Column (1) represents the true value for the rival effect under the model. Columns (2)-(5) used Oracle and columns (6)-(9) used Naive plug-in estimators. Columns (10)-(13) used developed Neyman orthogonal moments. Column (2), (6), and (10) shows the mean and (median) standard errors for the estimated parameters. CI denotes the probability of a 95 percent confidence interval based on standard errors from 500 simulations. RMSE denotes the root mean square error between estimated parameters and true parameters.

Figure C.4: The distribution of the estimated structural parameters from simulation

Sample Size: 2,000, Dimension of d_0: 500



Notes: The estimated effect of rival coefficients is based on 500 simulations. The true value of the rival effects is $\theta_0 = -1.5$. The oracle method uses low-dimensional relevant covariates, while the naive plug-in method uses high-dimensional covariates without correcting for biases. The orthogonal method also uses high-dimensional covariates like the naive plug-in method, but it corrects for biases using the proposed Neyman orthogonal method.

Table C.4: Summary Statistics: Non-Elderly Township

Variable	Frequency	*Panel A. Year 2000-2009*					*Panel B. Year 2010-2019*				
		Mean	S.D	Median	Min	Max	Mean	S.D	Median	Min	Max
Township-level variables											
Pop.[a]	Decennial	2079	1578	1775	114	14388	2087	1592	1826	123	14738
Income per Capita[b]	Decennial	16410	3012	16455	8360	35705	20912	4450	20807	10306	42282
Prop. Age in 6-17[c]	Decennial	0.205	0.03	0.20	0.11	0.32	0.182	0.03	0.18	0.08	0.30
Prop. Age 18-65[c]	Decennial	0.570	0.03	0.57	0.47	0.70	0.583	0.03	0.58	0.48	0.74
Prop. Age over 65	Decennial	0.159	0.03	0.17	0.05	0.20	0.168	0.04	0.17	0.05	0.37
Prop. Female	Decennial	0.507	0.02	0.51	0.39	0.62	0.502	0.02	0.50	0.31	0.55
Prop. White	Decennial	0.920	0.16	0.98	0.03	1.00	0.907	0.17	0.97	0.03	1.00
Prop. Black	Decennial	0.011	0.05	0.00	0.00	0.50	0.011	0.05	0.00	0.00	0.52
Prop. Native	Decennial	0.044	0.15	0.00	0.00	0.96	0.047	0.16	0.00	0.00	0.96
Prop. Asian	Decennial	0.014	0.03	0.00	0.00	0.23	0.019	0.04	0.01	0.00	0.31
Avg. Household Size	Decennial	799	606	687	49	6062	820	619	707	58	6193
Prop. Education 9-12 years	Decennial	0.121	0.05	0.11	0.03	0.30	0.091	0.05	0.09	0.00	0.29
Prop. - High School Graduates	Decennial	0.399	0.07	0.40	0.19	0.66	0.403	0.08	0.40	0.19	0.59
Prop. Education - Some college	Decennial	0.204	0.05	0.20	0.03	0.40	0.211	0.06	0.21	0.00	0.44
Prop. Bachelor	Decennial	0.151	0.05	0.15	0.04	0.32	0.189	0.07	0.18	0.04	0.71
Prop. Graduate	Decennial	0.040	0.02	0.04	0.00	0.16	0.046	0.03	0.04	0.00	0.27
Prop. Unemployment	Decennial	0.055	0.04	0.04	0.00	0.36	0.083	0.06	0.07	0.00	0.46
Prop. Commuting to Work - Vehicle	Decennial	0.887	0.07	0.90	0.15	0.97	0.883	0.09	0.90	0.00	1.00
Prop. Commuting to Work - Public transportation	Decennial	0.003	0.01	0.00	0.00	0.06	0.004	0.01	0.00	0.00	0.08
Prop. Commuting to Work - Taxi	Decennial	0.000	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.03	
Prop. Commuting to Work - Walk	Decennial	0.045	0.04	0.04	0.00	0.30	0.045	0.05	0.03	0.00	0.37
Prop. in Commuting to Work - Motorcycle/Bicycle	Decennial	0.013	0.04	0.01	0.00	0.62	0.016	0.04	0.01	0.00	0.61
Prop. Poverty	Decennial	0.121	0.07	0.10	0.01	0.40	0.140	0.08	0.13	0.00	0.39
Prop. Housing Vacancy	Decennial	0.129	0.11	0.09	0.02	0.62	0.158	0.12	0.12	0.03	0.76
Prop. in Rent	Decennial	0.250	0.09	0.24	0.03	0.57	0.260	0.10	0.25	0.06	0.59
Prop. Vehicle = 0	Decennial	0.071	0.06	0.06	0.00	0.54	0.065	0.07	0.05	0.00	0.72
Prop. Vehicle = 1	Decennial	0.309	0.08	0.31	0.04	0.50	0.290	0.09	0.30	0.00	0.51
Prop. Vehicle = 2	Decennial	0.391	0.06	0.39	0.07	0.65	0.381	0.08	0.39	0.06	0.69
Prop. Vehicle = 3	Decennial	0.159	0.05	0.15	0.02	0.34	0.174	0.07	0.17	0.03	0.56
Prop. Vehicle = 4	Decennial	0.048	0.03	0.04	0.00	0.30	0.061	0.04	0.05	0.00	0.31
Pharmacy Desert[d]	Annual	0.342	0.47	0.00	0.00	1.00	0.341	0.47	0.00	0.00	1.00
Ind. Pharmacies (Town)[e]	Annual	0.671	0.55	1.00	0.00	2.00	0.646	0.58	1.00	0.00	2.00
Chain Pharmacies (15 miles)[f]	Annual	0.746	1.12	0.00	0.00	7.00	1.382	1.73	1.00	0.00	7.00
County-level characteristics											
Physician Offices	Annual	9.425	11.46	5.00	1.00	96.00	9.155	10.66	5.00	1.00	83.00
State-level characteristics											
Prop. Insurance Age 18-64[g]	Annual	0.871	0.02	0.87	0.83	0.93	0.873	0.04	0.87	0.79	0.97
Prop. Insurance Age over 65[g]	Annual	0.992	0.01	0.99	0.97	1.00	0.991	0.01	0.99	0.96	1.00
N			2910					2910			

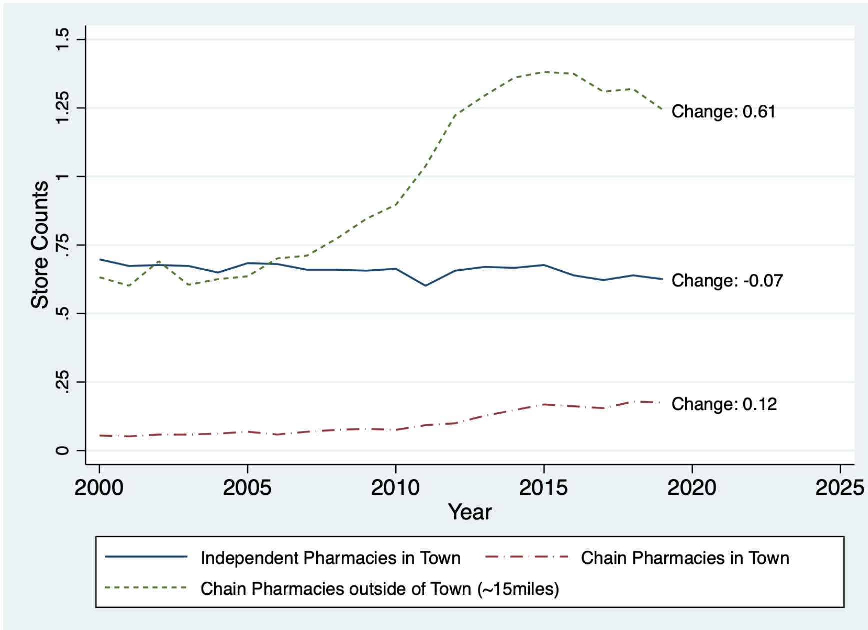
Notes: "Non Elderly township" is defined as townships with an age over 65 population ratio lower than 20% in the year 2000. "Decennial" implies that the census is conducted every ten years. "Annual" indicates that updates are made on a yearly basis. [a] "Pop." refers to the total population of each township. [b] "Income per Capita" represents the median income of each township. [c] "Prop." stands for the proportion of a specific demographic group within the population. [d] "Pharmacy deserts" is a binary variable taking the value 1 if there are no available pharmacies within the township. [e] "Ind. Pharmacy" denotes the average number of independent pharmacies within the township. [f] "Chain Pharmacy" denotes the average number of chain pharmacies within a 15-mile radius of the centroid of township. [g] "Prop. Insurance" refers to the ratio of the population within each age groups enrolled in health insurance.

Table C.5: Summary Statistics: Elderly Township

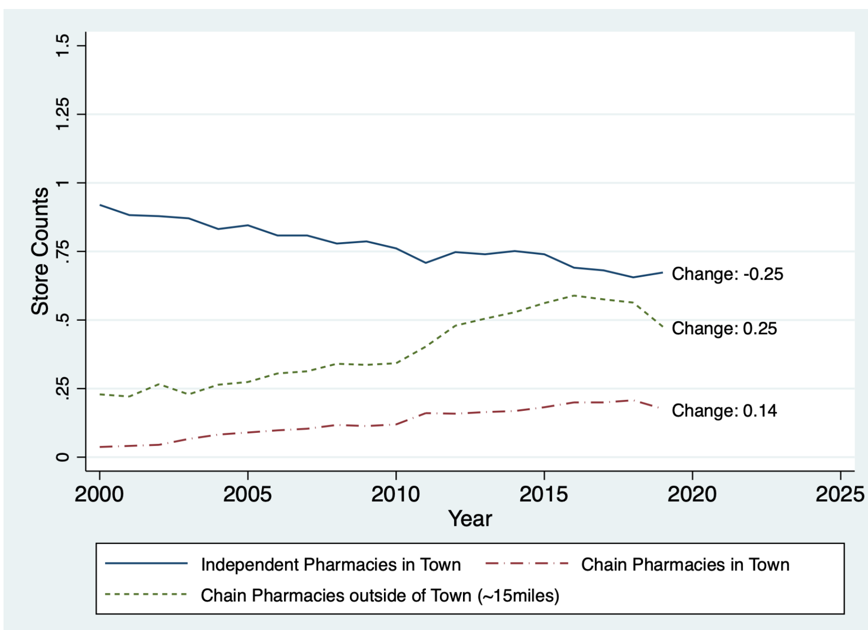
		Panel A. Year 2000-2009					*Panel B. Year 2010-2019*				
Variable	Frequency	Mean	S.D	Median	Min	Max	Mean	S.D	Median	Min	Max
Township-level variables											
Pop.[a]	Decennial	1434	725	1307	153	4859	1385	732	1227	117	4745
Income per Capita [b]	Decennial	16808	2206	16721	10022	27227	21345	3417	21202	12156	37437
Prop. Age in 6-17[c]	Decennial	0.176	0.02	0.18	0.09	0.24	0.160	0.03	0.16	0.08	0.25
Prop. Age 18-65[c]	Decennial	0.509	0.04	0.51	0.37	0.65	0.535	0.04	0.54	0.39	0.76
Prop. Age over 65	Decennial	0.262	0.05	0.25	0.20	0.48	0.246	0.05	0.24	0.10	0.49
Prop. Female	Decennial	0.528	0.02	0.53	0.37	0.59	0.518	0.02	0.52	0.29	0.62
Prop. White	Decennial	0.972	0.04	0.98	0.55	1.00	0.959	0.06	0.97	0.41	1.00
Prop. Black	Decennial	0.002	0.01	0.00	0.00	0.08	0.004	0.01	0.00	0.00	0.17
Prop. Native	Decennial	0.010	0.03	0.00	0.00	0.43	0.013	0.04	0.00	0.00	0.53
Prop. Asian	Decennial	0.008	0.02	0.00	0.00	0.16	0.012	0.02	0.01	0.00	0.25
Avg. Household Size	Decennial	605	302	560	74	2189	597	306	544	49	2079
Prop. Education 9-12 years	Decennial	0.101	0.03	0.10	0.04	0.22	0.076	0.04	0.07	0.00	0.23
Prop. High School Graduates	Decennial	0.379	0.06	0.38	0.17	0.57	0.389	0.07	0.39	0.13	0.59
Prop. Some college	Decennial	0.211	0.04	0.21	0.12	0.35	0.220	0.05	0.22	0.10	0.53
Prop. Bachelor	Decennial	0.163	0.04	0.16	0.04	0.31	0.204	0.06	0.20	0.03	0.37
Prop. Graduates	Decennial	0.040	0.02	0.04	0.00	0.13	0.044	0.02	0.04	0.00	0.17
Prop. Unemployment	Decennial	0.045	0.03	0.04	0.00	0.23	0.060	0.04	0.05	0.00	0.37
Prop. Commuting to Work - Vehicle	Decennial	0.867	0.05	0.88	0.60	0.97	0.868	0.07	0.88	0.47	1.00
Prop. Commuting to Work - Public transportation	Decennial	0.002	0.00	0.00	0.00	0.04	0.004	0.02	0.00	0.00	0.24
Prop. Commuting to Work - Taxi	Decennial	0.000	0.00	0.00	0.00	0.01	0.000	0.00	0.00	0.00	0.03
Prop. Commuting to Work - Walk	Decennial	0.074	0.04	0.07	0.00	0.35	0.067	0.05	0.06	0.00	0.32
Prop. Commuting to Work - Motobicycle	Decennial	0.010	0.01	0.01	0.00	0.12	0.015	0.02	0.01	0.00	0.10
Prop. Poverty	Decennial	0.101	0.04	0.09	0.03	0.32	0.131	0.07	0.12	0.00	0.47
Prop. Housing Vacancy	Decennial	0.133	0.11	0.10	0.02	0.73	0.162	0.12	0.13	0.03	0.76
Prop. in Rent	Decennial	0.249	0.07	0.24	0.04	0.47	0.271	0.07	0.26	0.06	0.53
Prop. Vehicle = 0	Decennial	0.076	0.03	0.07	0.00	0.22	0.062	0.04	0.06	0.00	0.25
Prop. Vehicle = 1	Decennial	0.349	0.05	0.35	0.10	0.53	0.325	0.07	0.33	0.00	0.57
Prop. Vehicle = 2	Decennial	0.382	0.05	0.38	0.22	0.50	0.380	0.06	0.38	0.19	0.67
Prop. Vehicle = 3	Decennial	0.138	0.04	0.14	0.02	0.29	0.161	0.06	0.15	0.02	0.42
Prop. Vehicle = 4	Decennial	0.040	0.02	0.04	0.00	0.26	0.049	0.03	0.04	0.00	0.18
Pharmacy Desert[d]	Annual	0.178	0.38	0.00	0.00	1.00	0.224	0.42	0.00	0.00	1.00
Ind. Pharmacies (Town)[e]	Annual	0.841	0.52	1.00	0.00	2.00	0.715	0.58	1.00	0.00	2.00
Chain Pharmacies (15 miles)[f]	Annual	0.357	0.78	0.00	0.00	6.00	0.676	1.16	0.00	0.00	7.00
County-level characteristics											
Physician Offices	Annual	5.699	8.90	3.00	1.00	80.00	6.622	10.07	3.00	1.00	80.00
State-level characteristics											
Prop. Insurance Age 18-64[g]	Annual	0.878	0.02	0.88	0.83	0.93	0.879	0.04	0.88	0.79	0.97
Prop. Insurance Age over 65[g]	Annual	0.993	0.01	0.99	0.97	1.00	0.991	0.01	0.99	0.96	1.00
N			5110					5110			

Notes: "Elderly township" is defined as townships with an age over 65 population ratio higher than 20% in the year 2000. "Decennial" implies that the census is conducted every ten years. "Annual" indicates that updates are made on a yearly basis. [a] "Pop." refers to the total population of each township. [b] "Income per Capita" represents the median income of each township. [c] "Prop." stands for the proportion of a specific demographic group within the population. [d] "Pharmacy deserts" is a binary variable taking the value 1 if there are no available pharmacies within the township. [e] "Ind. Pharmacy" denotes the average number of independent pharmacies within the township. [f] "Chain Pharmacy" denotes the average number of chain pharmacies within a 15-mile radius of the centroid of the township. [g] "Prop. Insurance" refers to the ratio of the population within each age group enrolled in health insurance.

Figure C.5: (Average) Number of Independent/Chain Pharmacies between 2000-2019 by Age Group



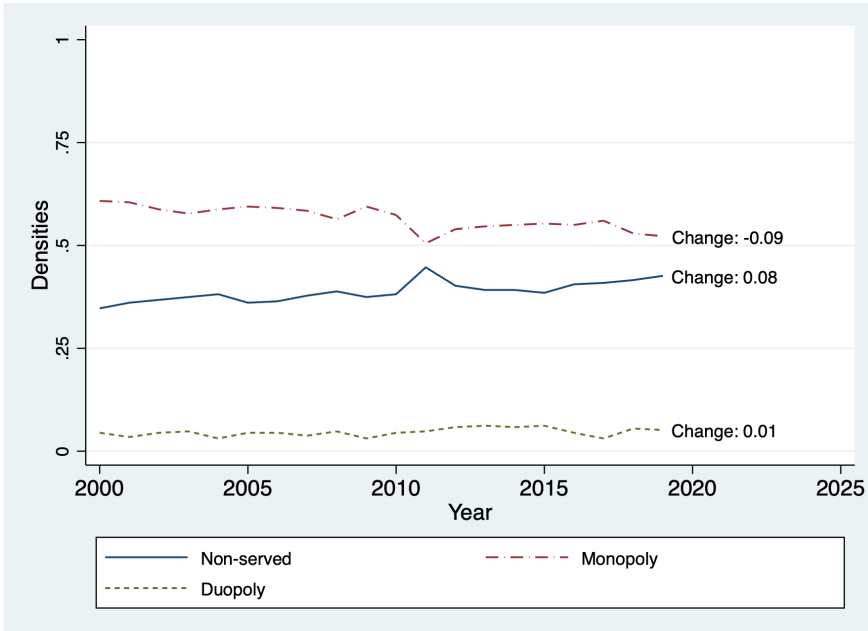
(a) Non Elderly Township



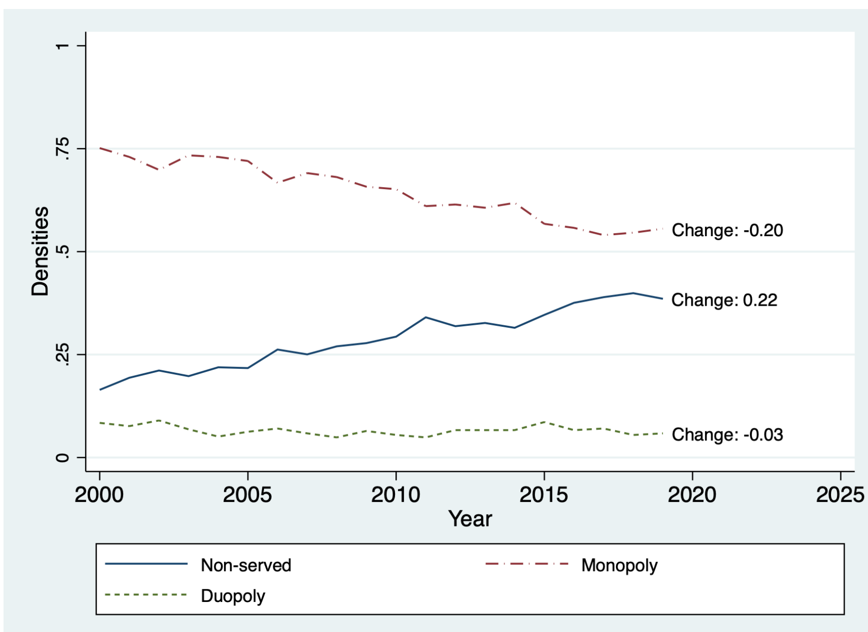
(b) Elderly Township

Notes:

Figure C.6: Distribution of Market Structure of Independently-Owned Pharmacy by Age Group



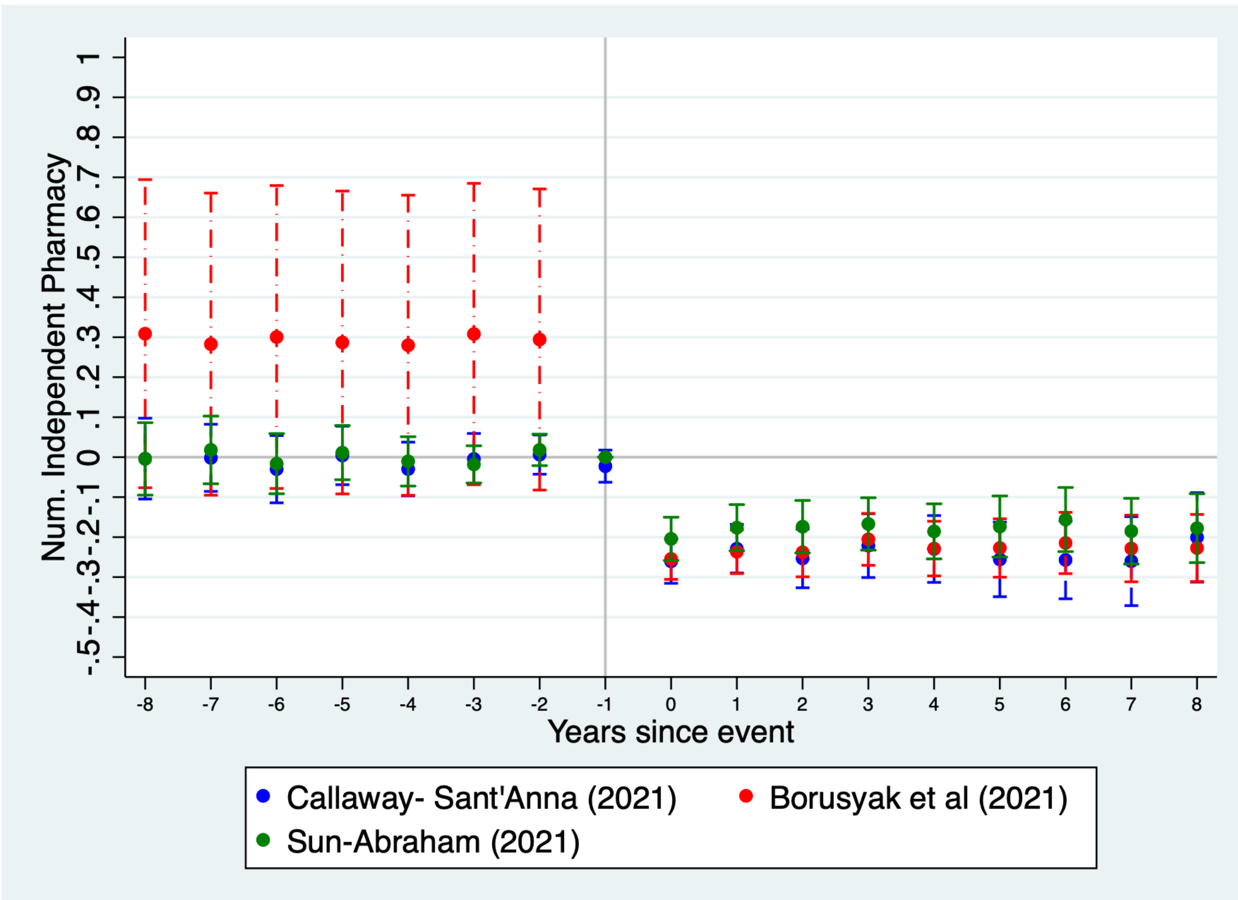
(a) Non Elderly Township



(b) Elderly Township

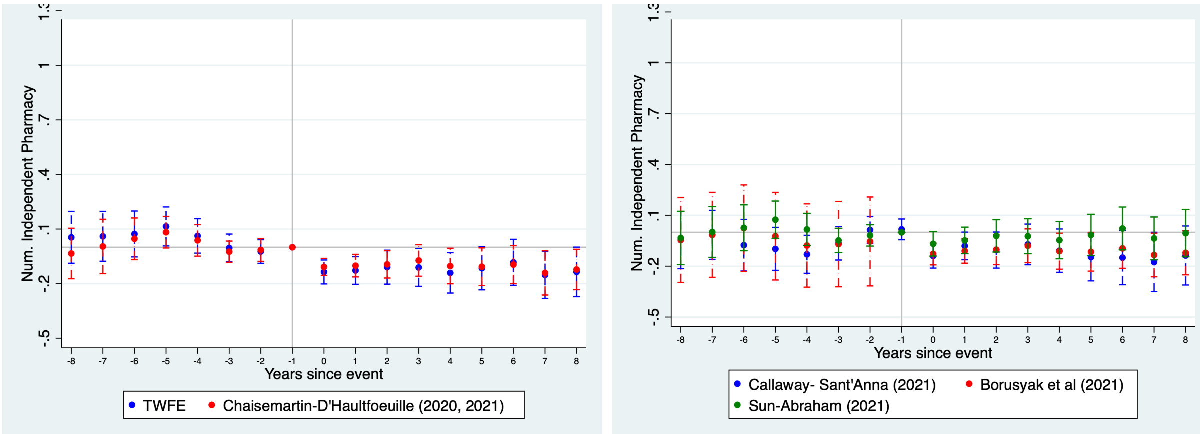
Notes:

Figure C.7: More Event Study: The effects of chain pharmacy entry on local independent pharmacy (Full Samples)



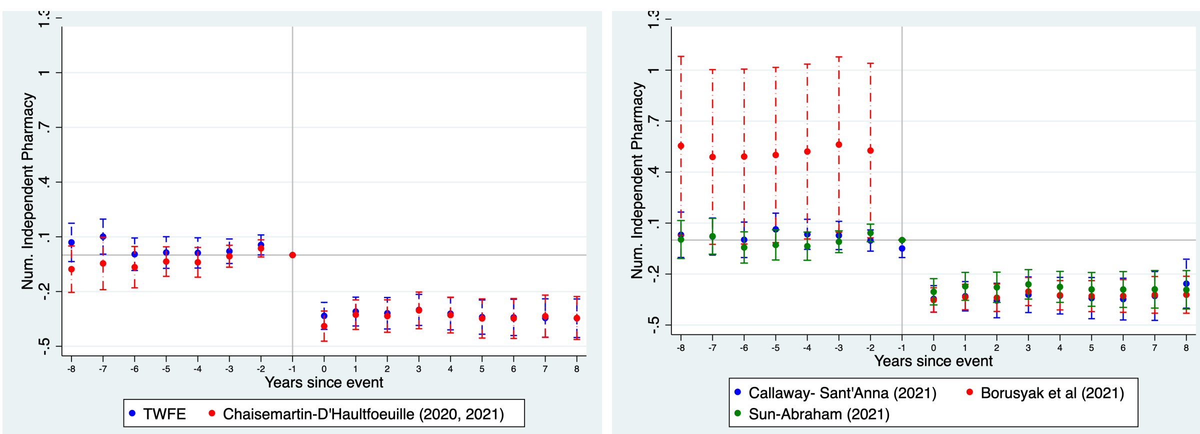
Note: Coefficient plots from event-study difference-in-differences analyses that regress the number of independent pharmacies in a township on year fixed effects, county fixed effects, control variables, and market× year fixed effects. The sample consists of 802 townships between 2000 and 2019. The omitted baseline period is $t = -1$, which is the last pre-treatment period. Standard errors are clustered at the county level and error bars represent 95 confidence intervals.

Figure C.8: Event Study: The effects of chain pharmacy entry on local independent pharmacy (Samples: Non Elderly Township)



Note: Coefficient plots from event-study difference-in-differences analyses that regress the number of independent pharmacies in a township on year fixed effects, county fixed effects, control variables, and market× year fixed effects. The sample consists of 291 townships between 2000 and 2019. The omitted baseline period is $t = -1$, which is the last pre-treatment period. Standard errors are clustered at the county level and error bars represent 95 confidence intervals.

Figure C.9: Event Study: The effects of chain pharmacy entry on local independent pharmacy (Samples: Elderly Township)



Note: Coefficient plots from event-study difference-in-differences analyses that regress the number of independent pharmacies in a township on year fixed effects, county fixed effects, control variables, and market× year fixed effects. The sample consists of 511 townships between 2000 and 2019. The omitted baseline period is $t = -1$, which is the last pre-treatment period. Standard errors are clustered at the county level and error bars represent 95 confidence intervals.

Table C.6: Pharmacy Store Opening Costs (Example)

Item Description		Expected Cost($)
Building		
Permits	Construction, including electrical, plumbing, architect drawing/building, plumbing & electrical permits, cost of building material and supplies	2,000
Construction	Bathroom refresh, drywall, electrical, plumbing, pharmacy and clinic sink, paint	25,000
Pharmacy/clinic outfit	Cabinetry, countertops, shelving, storage, medication fridge	20,000
Controlled medsafe	Purchase and bolted to floor	5,000
Shelving	Store perimeter wall	10,000-15,000
Signage	For exterior (marketing) and interior (location of products) pharmacy drop off/pick up, outside boxed sign, in-store signage	5,000
Inventory		
Furniture	Waiting area	2,000
Pharmacy supplies	Vials, labeling, stationery, compounding supplies, paper	2,000
Electronic		
Electronic items	Computers, cash register, phone system, TV, fax machine dispensing system, phones/fax/printer/cash register, ATM machines	15,000
Cable services	Comcast internet, phone (3 lines), TV services connection	500
Other		
Insurance	Building, workers comp, Professional Liability	250
Security	Gates for pharmacy, blinds for clinic, remote alarm, camera system	15,000
Advertising/Printing	Multi-language promo material, business cards, leaflets, patient education, newspaper advertisements (American, Chinese/Vietnamese papers), calendars/mugs/etc	
Total Costs		107,750-112,750

Source: Elabed et al. (2016)

Table C.7: Past/Current Chain Pharmacies and the Number of Independent Pharmacies

	(1) #. Independent Pharmacy Within Town
I(Entry of Chain =1 at t	-0.103***
	(0.0118)
I(Entry of Chain =1 at $t-1$	-0.0134
	(0.00942)
Township FE	Yes
Year FE	Yes
Market × Year FE	Yes
Controls	Yes
Observations	16,040
Mean of Dep. Variable	0.735
Adjusted R^2	0.547

Note Estimates are from fixed effects regression of the new entry of independent pharmacies outside of township but within 10 miles on the number of independent pharmacies in township m and year t. Significance levels are denoted by + p<0.10, * p<0.05, ** p<0.05.*** p<0.01.

Table C.8: Confusion Matrix

(a) Elderly Town

	Actual	
Predicted	Stay Out	Stay In
Stay Out	12,412 (0.608)	1,130 (0.055)
Stay In	78 (0.004)	6,780 (0.332)
Total N: 20,440	12,490 (0.6111)	7,950 (0.389)

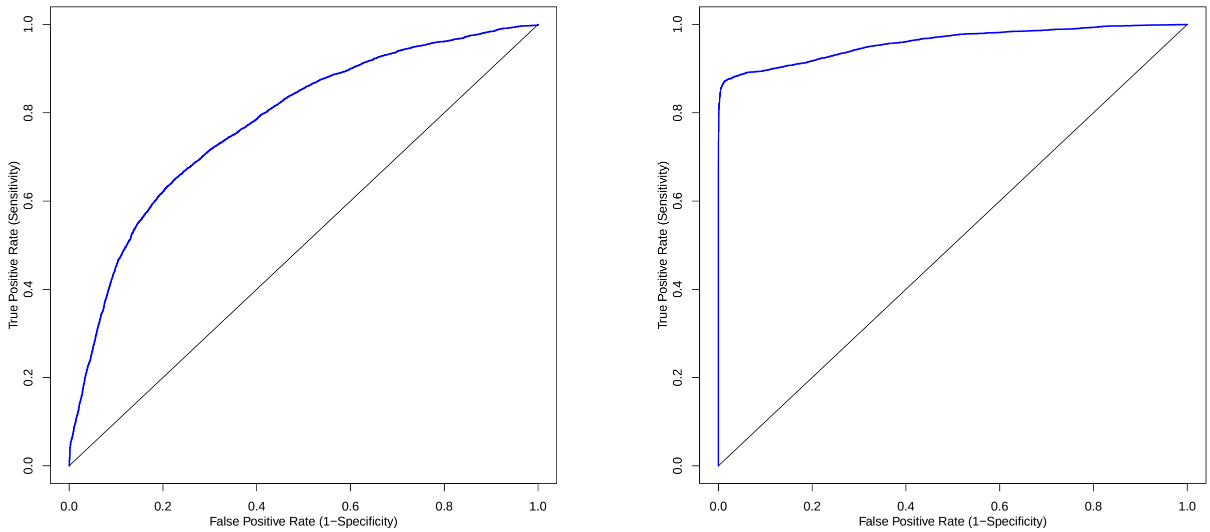
(b) Non-Elderly Town

	Actual	
Predicted	Stay Out	Stay In
Stay Out	7,773 (0.668)	741 (0.064)
Stay In	34 (0.003)	3,092 (0.266)
Total N: 11,640	7,807 (0.671)	3,833 (0.329)

Table C.9: Full Results for Bajari et al. (2010b)

	(1) Elderly Town	(2) Non-Elderly Town
Rival independent pharmacies	-5.420***	-4.000***
	(0.499)	(0.685)
Chain Pharmacies within 15 mi	-0.882***	-0.269***
	(0.0848)	(0.0604)
Store's Employment	0.592***	1.865***
	(0.165)	(0.318)
Total Pop.	1.229***	-0.0878
	(0.180)	(0.194)
Income Per Capita	0.384	-0.941*
	(0.316)	(0.477)
Physician Offices	0.205*	0.0160
	(0.0875)	(0.139)
Prop. Age over 65	-0.262	9.843***
	(1.348)	(2.725)
Prop. Female	-1.125	15.96**
	(3.011)	(5.802)
Prop. Black	-8.333	-2.918
	(5.695)	(6.875)
Prop. - High School Graduates	-0.180	-1.664
	(0.831)	(1.487)
Prop. Unemployement	-3.004*	-0.236
	(1.354)	(1.570)
Prop. Vehicle = 0	9.540***	3.520**
	(1.486)	(1.243)
Medicaid Expansion	0.0324	0.0804
	(0.0775)	(0.0989)
Prop. Insurance Age over 65	-0.519	-15.24***
	(2.170)	(3.711)
County FE	Yes	Yes
Year FE	Yes	Yes
Observations	20400	11640
Mean of Dep. Variable	0.388	0.329
Adjusted R^2	0.179	0.180

Figure C.10: ROC in First Stage Estimation of CCP: Elderly Town



(a) Logit
AUC: 0.7766

(b) XG Boosting
AUC: 0.9592

Notes: ROC denotes receiver operating characteristic curve. AUC denotes the Area under the ROC Curve.

Table C.10: Results from the Structural Model: Elderly Town

Parameters	Variables	Bajari et al. (2010b)	Plug-in	Orthogonal Moments
θ_γ	Rival independent pharmacies	-5.420	-7.972	-8.055
		(0.685)		(0.495)
θ_c	No. of chain pharmacies	-1.065	-1.065	-1.138
	(within 15 miles)	(0.085)		(0.057)
Observations		20,400	20,400	20,400
Socio-Economic Interaction		No	Yes	Yes
Dimension of Controls		13	563	563
Counties FE		Yes	Yes	Yes
Year FE		Yes	Yes	Yes

Notes: I use Lasso estimates for Plug-in estimators with cross validated penalty parameters from 5 fold cross fitting Standard errors are clustered at the county level. Standard errors are in parenthesis.

Table C.11: Results from the Structural Model
Robustness Check ($L = 4$)

Parameters	Variables	Orthogonal Moments	Orthogonal Moments
θ_γ	Rival independent pharmacies	-8.830	-6.794
		(0.355)	(0.563)
θ_c	No. of chain pharmacies	-1.477	-0.226
	(within 15 miles)	(0.035)	(0.018)
Observations		20,400	11,640
Socio-Economic Interaction		Yes	Yes
Dimension of Controls		563	563
Counties FE		Yes	Yes
Year FE		Yes	Yes

Notes: I use Lasso estimates for Plug-in estimators with cross validated penalty parameters from 5 fold cross fitting Standard errors are clustered at the county level. Standard errors are in parenthesis.

Table C.12: Goodness of Fit: By Socio-Economic Characteristics

	(Average) Independent Pharmacy Counts	
	Observed	Predicted
Total Markets	0.684	0.672
Total Population		
Below median (1,226)	0.612	0.588
Above median (1,226)	0.732	0.780
Prop. Vehicle=0		
Below median (0.055)	0.632	0.678
Above median (0.055)	0.714	0.690
Prop. under Poverty Line		
Below median (0.12)	0.620	0.632
Above median (0.12)	0.726	0.738
Share of Age over 65		
Below median (0.24)	0.654	0.682
Above median (0.24)	0.690	0.686
Presence of Chain Pharmacy in 2000		
No chain pharmacy within 15 miles	0.814	0.788
Chain pharmacy present within 15 miles	0.440	0.512
Minority Group		
Below 10%	0.670	0.682
Above 10%	0.700	0.732

Notes:

D Appendix: Data Appendix

Table D.1: Description of Datasets

Dataset Source	Description
Pharmacy Entry/Exit Data	
Data Axle Historical Business Database	This proprietary dataset, accessible via https://www.dataaxleusa.com/lp/data-axle/ and the Carnegie Library of Pittsburgh, is provided by Data Axle - data analytics marketing firm. The dataset encompasses 361 million digitized records of historical and contemporary business establishments from 1997-2021. I collected panel histories of pharmacies and mapped their addresses to township IDs using census shapefiles below.
Geographic Information System	
2000/ 2010 US Township (county subdivision) Shapefiles	These shapefiles, available at https://www.census.gov/cgi-bin/geo/shapefiles/index.php[1], outline each township's boundaries. It allows me to geocode the addresses of pharmacies and assign township IDs from the Census data.[2]
2010 Rural-Urban Commuting Area (RUCA) Codes	Sourced from https://www.ers.usda.gov/data-products/rural-urban-commuting-area-codes/documentation/, these codes define the census classifications for rural areas. I keep pertaining to rural townships.
2000-2010 Township Crosswalk	Available at https://www.census.gov/geographies/reference-files/time-series/geo/relationship-files.2010.html#list-tab-1709067297, this file details the relationships between 2010 Census county subdivisions and their 2000 Census counterparts.
Health-related variables	
CBP (County Business Patterns)	Sourced from https://www.census.gov/data/developers/data-sets/cbp-nonemp-zbp/cbp-api.html, CBP presents data on county-level business establishments, categorized by North American Industry Classification System (NAICS) codes. For this research, I extracted data on physician offices in each county using the physician's code (NAICS code: 621111).
Health Insurance Coverage	Available at https://cps.ipums.org/cps/index.shtml, the Annual Social & Economic Supplement of the Current Population Survey provides data on health insurance enrollment rates at the year-state level, grouped by age groups 6-17, 18-64, and above 65.

[1] On the website, I selected "Year 2010" followed by the "County Subdivisions (township)" layer type, enabling the download of shapefiles for both 2000 and 2010.

[2] In this study, the 2010 shapefiles were utilized for township IDs.

D.1 Construction of Data

To construct my final dataset, I levarage data from various sources, including pharmacy establishment datasets, market level characteristics, and health-related variables.

To begin with, I created a panel dataset of pharmacies in the Midwest U.S., organized by township and year. This allowed me to track the openings and closings of both independent and chain pharmacies over time. The Data Axle database provides data on pharmacies from 1997 to 2021.

Next, I defined the market boundaries for each pharmacy using the 2010 U.S. Census townships. I used the 2010 census boundaries for consistency, even though they have changed slightly over time.

To focus on rural areas, I used the census's definition of rural territories based on the RUCA. I used Python's geopandas tool to identify rural townships that don't overlap with urban census tracts derived from RUCA. I then used the 2000-2010 township crosswalk dataset to maintain the townships in line with the 2010 shapefiles.

Once I had a definitive list of rural townships, I used geopandas for geocoding. I used Yahoo Bing's reverse geocoding feature to translate pharmacy addresses into their corresponding longitude and latitude coordinates. I then aligned each independent pharmacy with its corresponding township ID.

To find out how many chain pharmacies are within a certain distance of each township, I first found the center of each township. Then, I drew circles around each township with radius of 5 to 30 miles. I counted the number of different types of chain pharmacies in each circle to get an exact count of the chain pharmacies within the specified distances. After the data-cleaning process, for each township, I have independent pharmacies with number of chains within 5-30 miles.

D.2 Market level characteristics

I also collect market-level data on a pool of demographic characteristics from the Census and ACS at the township level. This data allows me to estimate the latent profits of independent stores, as it proxies for both the demand for prescriptions and the costs of operating stores. Note that the decennial census was released in 2000 and 2010 during my sample periods, so most market-level characteristics are decennial. I list a full list of the demographic variables' geographic units and their frequencies in Table and Table .

D.3 Health related variables

To account for potential time-varying in prescription demands, I incorporate the number of physicians per county per year and health insurance enrollment rates per state per year, drawing data from the Annual Social & Economic Supplement of the Current Population Survey program. In addition, I include "Medicaid Expansion" dummy variable[36]. This variable is assigned a value of 1 in a given year if the state expanded Medicaid coverage to nearly all adults with incomes up to 138% of the Federal Poverty Level ($20,120 for an individual in 2023).

[36] Source: available at https://www.kff.org/medicaid/issue-brief/status-of-state-medicaid-expansion-decisions-interactive-map/