Three Essays on Machine Learning and Applied Economics

by

Hyungjin Kim

B.A. in Economics and Philosophy, Sungkyunkwan University, 2015

M.A. in Economics, Sungkyunkwan University, 2017

Submitted to the Graduate Faculty of

the Dietrich School of Arts and Science in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2024

UNIVERSITY OF PITTSBURGH

DIETRICH SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Hyungjin Kim

It was defended on

March 28th 2024

and approved by

Arie Beresteanu, Department of Economics, University of Pittsburgh

Douglas Hanley, Associate Professor of Economics, University of Pittsburgh

Mahrad Sharifvaghefi, Assistant Professor of Economics, University of Pittsburgh

Robert A. Miller, Professor of Economics & Strategy, Carnegie Mellon University

Dissertation Director: Arie Beresteanu, Associate Professor of Economics,

University of Pittsburgh

Three Essays on Machine Learning and Applied Economics

Hyungjin Kim, PhD

University of Pittsburgh, 2024

This dissertation explores two main themes: the development of static games using machine learning, and the impact of firms' behaviors on economic and health outcomes. The first chapter focuses on the development of machine learning estimators for static games. The second chapter examines the effects of chain pharmacy entry on competition, market structure, and access to pharmacies in rural areas, utilizing the estimators developed in the first chapter. The third chapter investigates how mergers alter the post-merger market structure.

Table of Contents

Preface xiv

1.0 Machine Learning Estimators for Static Games 1

1.1 Introduction 1

1.1.1 Related Literature 3

1.2 Model Framework 4

1.2.1 Two Step Estimation 8

1.2.1.1 Step 1: Estimation of nuisance parameters 9

1.2.1.2 Step 2: Recovering the Structural Parameters 9

1.3 The DML-Static Game Estimator 12

1.3.1 The Definition of Neyman Orthogonal Moment Condition . 12

1.3.2 Orthogonal Moment Condition for Static Game: Two Players Example 13

1.3.3 Estimation Procedure 15

1.3.4 Asymptotic Analysis 16

1.3.5 Monte Carlo Simulation 18

1.3.5.1 The static entry/exit model and equilibrium 18

1.3.5.2 Data Generation 19

1.3.5.3 Estimation of the static entry/exit model 20

1.3.6 Simulation Results 21

1.4 Conclusion 22

1.5 Proofs 22

1.6 Tables 27

1.7 Figures 28
2.0 Rural Pharmacy Access and Competition 29
2.1 Introduction 29
2.1.1 Literature Review 32
2.2 Data and Background 33
2.2.1 Industry Background 33
2.2.2 Data 35
2.2.2.1 Market Definition of Geographic Level 36
2.2.2.2 Final Sample 37
2.2.3 Limited Pharmacy Access in Rural Towns 37
2.2.3.1 Summary Statistics 39
2.2.4 Market Structure 40
2.2.5 Reduced Form Evidence 42
2.2.5.1 Specification of Distance to Chain Pharmacies . . . 42
2.2.5.2 Effects on Market Structure 43
2.2.6 Preliminary Analysis on Strategic Interaction between Independent Pharmacies 45
2.3 Structural Analysis 46
2.3.1 Model Primitive 46
2.3.1.1 Equilibrium Concept 48
2.3.2 Discussion 48
2.3.2.1 Static versus Dynamic Framework 48
2.3.2.2 Profit Shifters for Identification 49
2.3.2.3 Isolated Market Assumption 50
2.3.3 Estimation Procedure: Two-Step Estimators 50
2.3.3.1 Time-Varying Unobserved Endogenous Variables . . 51

2.3.3.2 Existing Two-Step Estimators 52
2.3.3.3 Neyman Orthogonal Estimators 54
2.3.4 Estimation Results 55
2.3.5 Robustness Check 58
2.4 Counterfactual 59
2.4.1 Solution Method for the Static Game 59
2.4.2 Goodness of Fit 60
2.4.3 Chain Pharmacy Counterfactual: Fixing Entry of Chain Pharmacy afterward 2000 61
2.4.4 Policy Counterfactual: Providing Subsidy associated with Medicare in High Elderly Population Towns 61
2.5 Conclusion 63
2.6 Tables 64
2.7 Figures 70
3.0 Horizontal Merger and Post-Entry Market Structure: Evidence from Acquisition in the Retail Pharmacy Market 78
3.1 Introduction 78
3.2 The Retail Pharmacy Industry 82
3.2.1 Background of Merger between Walgreens and Rite Aid . . 83
3.3 Data 84
3.3.1 Geographic Illustration 85
3.3.2 Descriptive Evidence 86
3.4 Empirical Strategy 87
3.4.1 Estimating Dynamic Treatment Effects 87
3.4.2 Difference-in-Differences Estimators 89

3.4.3 Alternative Approach: Staggered Difference-in-Differences Estimators 90
3.4.4 Estimation Results from Staggered Difference-in-Differences Estimators 92
3.4.5 Identification 93
3.4.6 Heterogeneity Analysis 93
3.4.7 Robustness Check 94
3.5 Conclusion 95
3.6 Tables 97
3.7 Figures 102
Appendix A. Review of Discrete Static Games 110
A.1 Discrete Games Details 110
A.2 Discussion on Multiple Equilibria and Estimation Approach 112
A.3 Identification 114
Appendix B. Tables 121
Appendix C. Figures 150
Appendix D. Data Description 162
D.0.1 Construction of Data 162
D.0.2 Market level characteristics 163
D.0.3 Health related variables 163
D.1 Additional Discussions 164
D.1.1 Strategic Decision of Pharmacies 164
D.1.2 Pharmacy Profits and Machine Learning Approach 165
D.2 Estimation Details for Neyman Orthogonal Estimators 165
Bibliography 168

List of Tables

1 Simulation Results 27

2 Comparison of Estimator Specifications for Existing and Developed Orthogonal Estimators 64

3 Performance of Different Methods in First Stage CCP for Hold-Out Sample 65

4 Results from the Structural Model: High Elderly Population Town 66

5 Results from the Structural Model: Non-High Elderly Population Town 67

6 Pharmacy Accessibility under Subsidy Counterfactual Scenario (Year: 2019) 68

7 Expected Number of Stores under Chain Pharmacy Counterfactual (Year: 2019) 69

8 (Average) Summary Statistics by Market Type 97

9 Horizontal Merger and Pharmacy Stores 98

10 Horizontal Merger and Number of Pharmacy Stores 99

11 Horizontal Merger and Non-merged Pharmacy Stores 100

12 Horizontal Merger and Entries of Non-merged Pharmacy Stores . 101

13 Summary Statistics: High Elderly Population Township 121

14 Summary Statistics: Non-High Elderly Population Township . . 122

15 The New Entries in Chain Pharmacies with Different Distances and the Number of Independent Pharmacies. 123

16 Logit Regression on Independent Pharmacy's Entry. 124

17 Pharmacy Store Opening Costs (Example) 125
18 Top 10 Importance Features from Xg boosting 126
19 Past/Current Chain Pharmacies and the Number of Independent Pharmacies 127
20 Entries in Neighborhood (outside of township) and the Number of Independent Pharmacies. 128
21 (Selected) Descriptive Statistics: Non-High Elderly Population Township 129
22 (Selected) Descriptive Statistics: High Elderly Population Township 130
23 Confusion Matrix 131
24 Full Results for Bajari et al. (2010b) 132
25 Notation for Parameters and Data 133
26 First Stage Reduced Form CCP: Existing Method Bajari et al. (2010b) 134
27 Results from the Structural Model Robustness Check ($L = 4$) . . 135
28 Goodness of Fit: By Socio-Economic Characteristics 136
29 Expected Number of Stores under Counterfactual Scenario 2 (Year: 2019) 137
30 Comparison of Rite Aid Stores 138
31 Heterogeneity by Income Group: Horizontal Merger and Number of Pharmacy Stores 139
32 Heterogeneity by Income Group: Horizontal Merger and Number of Non-merged Pharmacy Stores 140
33 Heterogeneity by Income Group: Horizontal Merger and New Entries of Non-merged Pharmacy Stores 141

34 Robustness Check: Adding Adjacent Markets: Horizontal Merger and Number of Pharmacy Stores 142

35 Robustness Check of Adding Adjacent Markets: Horizontal Merger and Non-merged Pharmacy Stores 143

36 Robustness Check: Adding Adjacent Markets: Horizontal Merger and Entries of Non-merged Pharmacy Stores 144

37 Robustness Check: Alternative Demographic Controls: Horizontal Merger and Number of Pharmacy Stores 145

38 Robustness Check: Alternative Demographic Controls: Horizontal Merger and Non-merged Pharmacy Stores 146

39 Robustness Check: Alternative Demographic Controls: Horizontal Merger and Entries of Non-merged Pharmacy Stores 147

40 de Chaisemartin and Xavier D'Haultfoeuille's Staggered TWFE: Horizontal Merger and Pharmacy Stores 148

41 Description of Datasets 149

List of Figures

1 The distribution of the estimated structural parameters from simulation Sample Size: 2,000, Dim (s_0)=500 28
2 Trends in Limited Pharmacy Accessibility 70
3 Trends in Pharmacy Access by High-Elderly Population/Non-High Elderly Population Towns 71
4 An Example: Spatial Distribution in Independent/Chain Pharmacy 72
5 (Average) Number of Independent/Chain Pharmacies between 2000-2019 73
6 Event Study: The Effects of Chain Pharmacy Entry on Local Independent Pharmacies 74
7 Heterogeneity by Ratio of Elderly Population Towns: the Effects of Chain Pharmacy Entry on Local Independent Pharmacies . . 75
8 Algorithm Steps 76
9 Goodness of Fit: High elderly population towns 77
10 Walgreens and Rite Aid Merger Events Timeline 102
11 Geographic Distribution of Stores Location Before/After the Horizontal Merger 103
12 Number of Stores Before/After the Horizontal Merger 104
13 Number of Stores Before/After the Horizontal Merger 105
14 Geographic Distribution of Exit of Stores Before/After the Horizontal Merger 106

15 Event Study: The effects of the horizontal merger on the total number of pharmacies 107
16 Event Study: The effects of the horizontal merger on the new entries by non-merged firms 108
17 Horizontal Mergers and Pharmacy: Event Study 109
18 Timing of the Game 150
19 Industry Survey: Main Reason for Choosing Primary Pharmacy 151
20 Industry Survey: Main Reason for Switching Primary Pharmacy 152
21 Trends in Pharmacy Deserts: Alternative Definition 153
22 (Average) Number of Independent/Chain Pharmacies between 2000-2019 by Age Group 154
23 Distribution of Market Structure of Independently-Owned Pharmacy by Age Group 155
24 More Event Study: The effects of chain pharmacy entry on local independent pharmacy (Robustness Check) 156
25 ROC in First Stage Estimation of CCP: High Elderly Population Town 157
26 Change in Market Structures with Before/After the Merger Approval 158
27 Dynamic Effects of Horizontal Merger on Non-merging Pharmacies 159
28 Dynamic Effects of Horizontal Merger on Entries of Non-merging Pharmacies 160
29 Balancing Test of the Exit of Merged Firms' Treatment on Pre-Market Characteristics 161

Preface

I'm extremely grateful to my advisor, Arie Beresteanu, for his continuous support during my doctoral studies. Your commitment to teaching me not just what to think, but how to think critically and independently, has been instrumental in shaping my academic journey. I would never have obtained a Ph.D. on this path without you.

I would like to thank Douglas Hanley, Mahrad Sharifvaghefi, and Robert A. Miller for their support and valuable contributions to my research and for guiding me during my time at Pitt. I would also like to thank Anh Nguyen for providing thoughtful guidance and direction throughout my graduate education. To Hyunchul Kim and Minsoo Park, I want to express my gratitude for providing me with my initial opportunity in economic research almost a decade ago. Without this initial great experience, I couldn't have discovered how truly enjoyable research can be.

Finally, I am deeply indebted to my family and Seoyun Hong for joining my achievement and supporting me when I doubted myself. This dissertation is dedicated to you.

1.0 Machine Learning Estimators for Static Games

1.1 Introduction

Game-theoretic models of strategic interactions are widely used, including entry/exit (Bresnahan and Reiss, 1991), (Berry, 1992), product quality (Mazzeo, 2002), location (Seim, 2006), network (Jia, 2008), (Holmes, 2011), and promotional strategies (Jia, 2008). When analyzing firms' strategic interactions using data, it is crucial to consider the characteristics of local markets that can affect firm profits. Researchers face complex challenges when observing many covariates, ranging from traditional sources like rich market covariates and firm-level characteristics to modern unstructured high-dimensional datasets such as consumer reviews and product images. A key task for the researchers is to carefully select variables that are relevant to underlying payoffs and to choose controls that avoid omitted variable bias or multicollinearity. This process also requires prior knowledge of the correct functional form of firm profits and involves deciding on the transformations and interactions of the selected covariates. The variable selection and functional form specification, often done ad-hoc, present substantial challenges in applied research.

To tackle these challenges, researchers often turn to a data-driven approach by using machine learning (ML) methods with regularization, such as the Lasso estimator, which can handle high-dimensional market characteristics. However, ML methods face a trade-off between bias and variance: while they excel in prediction (reduce variance), 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. Similar to the causal inference framework, the model parameters of the discrete games of strategic interactions can be categorized into two groups - low-dimensional parameters of interest (e.g. strategic interaction effects) and high-dimensional nuisance parameters (the effects of many market covariates). While ML methods allow for high-dimensional nuisance parameters, regularization bias in the nuisance parameter estimates unavoidably will be transmitted into the primary parameter of interest. As a result, the low-dimensional structural parameters of interest would be badly biased.

In this paper, I provide a methodology for valid inference of underlying structural parameters of interest in the presence of high-dimensional nuisance parameters based on machine learning methods. I employ the framework in Chernozhukov et al. (2018b) to allow for the use of high-dimensional covariates in the model of strategic interaction with incomplete information. First, I use 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 components 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 and it performs well even in the presence of many irrelevant variables to the game model.

1.1.1 Related Literature

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. My 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). [1] 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.

1.2 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). For general purposes, I describe standard static game structures and I specifically tailor the discussion to pharmacies' entry and exit games in section 5.

I define a set of players, represented as $i \in \{1, ..., n\}$. Each player has two distinct choices, which I denote by $\mathcal{J} = 2$[2].

$$a_i = \begin{cases} 1 & \text{if Player } i \text{ chooses to being active.} \\ 0 & \text{if Player } i \text{ chooses being inactive.} \end{cases} \tag{1.2.1}$$

Let $\mathcal{A} = \{0, 1\}^n$ represent the Cartesian product of the choices made by all players in the same market with n player. I denote N as the total sample size, where n

[1] 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.

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

represents the number of players in each market. Without loss of generality, I assume that the number of players, n, is the same across different markets. 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 $\mathcal{A}$. Adhering to traditional game-theoretic 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 takes the $u(s, a_i, a_{-i}, \epsilon, \theta)$ where s denotes the relevant state variables and ϵ denotes the private information. The period utility function can be expressed as:

$$u_i(a_i, a_{-i}, s, \epsilon; \theta) = \Pi_i(a_i, a_{-i}, s; \theta) + \epsilon_i(a_i)$$

where the deterministic payoff function $\Pi_i(a_i, a_{-i}, s; \theta)$ is additive separable with respect to private information ϵ_i. I impose the standard assumptions on the model primitives:

Assumption 1.2.1 (Private 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 are not observed by analysts.*

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

Assumption 1.2.1 (a) highlights the conditional independence assumption, which specifies that ϵ_i is independent of ϵ_{-i} given s. This paper narrows its focus on settings characterized by incomplete information, as described in Assumption 1.2.1 (b). This suggests that the realized utility functions are private information to the respective players. 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 1.2.2 (Normalization of Outside Choice)**.**

$$\forall\ a_{-i} \in \mathcal{A}_{-i},\ \forall\ s,\ \Pi_i(a_i = 0, a_{-i}, s) = 0. \tag{1.2.2}$$

The assumption for normalization 1.2.2 is crucial for identifying payoffs when a player decides to be active.

Given the utility function, I define the decision-choice rule as $a_i = \delta_i(s, \epsilon_i(a_i))$. As neither econometricians nor rivals observe ϵ_i, the decision rule is characterized by choice probabilities:

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

where $\mathbf{1}\{\delta(s, \epsilon_i(a_i)) = a_i\}$ stands as the indicator function that takes the value of 1 if player i chooses action 1, and 0 otherwise. f denotes the distribution of private information. Due to the private information assumption, the decision rules of $\delta_i(s, \epsilon_i(a_i))$, remain independent of the private information of her rivals, ϵ_{-i}.

Given private information assumptions 1.2.1, the choice-specific value function Π_i can be expressed as the expected payoffs linked to rivals' choice probabilities, denoted as σ_{-i}.

$$\Pi_i(a_i = 1, s; \theta) = \sum_{a_{-i} \in A_{-i}} \sigma_{-i}(a_{-i}|s) \pi_i(a_i = 1, a_{-i}, s; \theta) \text{ for all } i = 1, ...n. \tag{1.2.3}$$

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

where π denotes the payoff which depends on the rival's action, and common state variable s. The choice-specific value function Π_i describes 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 upon this foundation, I expand the scope to encompass games with n players. In this context, after fixing state variable s, $\sigma_i(a_i|s)$ will be the solution to the system of n equations:

$$\begin{aligned}
\sigma_1(a_1 = 1|s;\theta) &= \frac{\exp\Big(\sum_{a_{-1}\in A_{-1}} \sigma_{-1}(a_{-1}|s;\theta)\pi_i(a_1 = 1, a_{-1}, s;\theta)\Big)}{1+\exp\Big(\sum_{a_{-1}\in A_{-1}} \sigma_{-1}(a_{-1}|s;\theta)\pi_i(a_1 = 1, a_{-1}, s;\theta)\Big)} \qquad (1.2.4)\\
\sigma_2(a_2 = 1|s;\theta) &= \frac{\exp\Big(\sum_{a_{-2}\in A_{-2}} \sigma_{-2}(a_{-2}|s;\theta)\pi_i(a_2 = 1, a_{-2}, s;\theta)\Big)}{1+\exp\Big(\sum_{a_{-2}\in A_{-2}} \sigma_{-2}(a_{-2}|s;\theta)\pi_i(a_2 = 1, a_{-2}, s;\theta)\Big)}\\
&\vdots\\
\sigma_n(a_n = 1|s;\theta) &= \frac{\exp\Big(\sum_{a_{-n}\in A_{-n}} \sigma_{-n}(a_{-n}|s;\theta)\pi_i(a_n = 1, a_{-n}, s;\theta)\Big)}{1+\exp\Big(\sum_{a_{-n}\in A_{-n}} \sigma_{-n}(a_{-n}|s;\theta)\pi_i(a_n = 1, a_{-n}, s;\theta)\Big)}
\end{aligned}$$

I highlight the system of equations in 1.2.4, which represents the best response function between players for several points. First, I presume the existence of a solution to equation (1.2.4), 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. Here, I use the standard equilibrium concept of the Bayesian Nash Equilibrium. For the issue of multiple equilibria and identification challenges, Appendix A.2 provides detailed discussions.

1.2.1 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 using 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. 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.

With the exclusion identification assumption in the Appendix section A.3, relevant state variables are decomposed into $s = (s_x, s_i, s_{-i})$ where s_x denote the common market characteristics for players, s_i denotes the player i specific shifter, and s_{-i} denotes rivals' shifter. The conditional value function assumes the following form:

$$\pi_i(a_i = 1, s_x, s_i, s_{-i}; \theta_\gamma, \beta) = \sum_{a_{-i} \in A_{-i}} \sigma_{-i}(a_{-i}|s_x, s_i, s_{-i})\theta_\gamma + s_i\beta_\kappa + s_x'\beta_s \tag{1.2.5}$$

Here, θ_γ represents the interaction effects from rivals, β_κ shows the effect of player-specific productivity shock for player i, and β_s encompasses the effect of common market characteristics. I denote $\beta = (\beta_\kappa, \beta_s)$.

1.2.1.1 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:

$$\gamma_{-i} = E[a_{-i}|s_x, s_{-i}, s_i] \text{ for all } i = 1, ..., n.$$

Nonparametric 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 be consistent estimates of σ_{-i}.

1.2.1.2 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_γ and β from equation (1.2.5). To accomplish this, I follow Bajari et al. (2010b)'s semi-parametric models. Coupled with the Type 1 Extreme Value Distribution, I assume that the moment condition, based on the first-order condition of the logit likelihood satisfies the following condition

$$E[m(w; \theta_{\gamma 0}, \beta_0)] = 0 \tag{1.2.6}$$

where w denotes the data $w_i = (a_i, s_i, s_x)$ for $i = 1, ..., n$. Given the sample of observations $w_i = (w_1, w_2, ..., w_n)$, I define $(\theta_{\gamma 0}, \beta_0)$ as the solution to

$$\frac{1}{N}\sum_{i=1}^{N} m(w_i; \theta_\gamma, \beta) = \frac{1}{N}\sum_{i=1}^{N} [z_i(a_i - \Lambda(s_x, s_i, \hat{\gamma}_{-i}; \theta_\gamma, \beta))] = 0 \tag{1.2.7}$$

where $z_i = (z_1, ..., z_N)$ denotes the correctly specified variables $z_i = (a_i, s_i, s_x)$ and Λ denotes the logit-link function. With low-dimensional vector s_x, 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).

Next, I consider the setting where common market characteristics s_x include high-dimensional covariates, that is, the dimension of s_x 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 other parameters (γ_{-i}, β) as nuisance parameters. Therefore, $\theta = \theta_\gamma$ and $\eta = (\gamma_{-i}, \beta)$ in the below and in section 4. However, I incorporate a subset of β as the parameter of interest instead of the nuisance parameters, in the empirical application in section 6. The developed methodology in section 4 can be trivially extended to this case.

In this high-dimensional setting, the conventional Generalized Method of Moments (GMM) becomes infeasible due to ill-conditioned properties encountered when calculating the Jacobian for a high-dimensional matrix. To avoid this bottleneck, I allow for machine learning methods with regularization to enable estimation. For example, I could employ a Logistic Lasso estimator to estimate first-stage choice probability $\hat{\gamma}_{-i}$ as well as covariates effects $\hat{\beta}$ in the second stage. This imposes the regularization in that it requires the sparsity assumption such as

$$\frac{p^2 \log^2(dim(s_x) \vee N)}{N} \to 0$$

to enable estimation with high-dimensional covariates, where p denotes the (true) low-dimensional relevant market characteristics and $dim(s_x)$ refers to the dimension of a pool of market covariates s_x.

However, when machine learning estimates are used to estimate nuisance parameter $\eta = (\gamma_{-i}, \beta)$, the parameters of interest θ fail to achieve $\sqrt{N}$-consistency, as documented in Chernozhukov et al. (2022). To illustrate, estimating θ associated with machine learning estimators of $\hat{\eta}$ exhibits a slower convergence rate than $\sqrt{N}$. Also, the regularization biases are transmitted into the second stage moment in the equation (1.2.7), making moment functions sensitive to these regularization biases. This effect can be shown that the directional (Gateaux) derivative with respect to the nuisance parameters η is non-zero:

$$\partial_{\gamma_{-i}} E[m(w_i;\theta,\eta)][\gamma_{-i} - \gamma_{-i0}] = E\left[z_i \cdot \Lambda' \cdot \left(-\theta_\gamma \sum_{a_{-i}} \prod_{s \neq i,i} (1-\gamma_s)\right) \cdot (\gamma_{-i} - \gamma_{-i0})\right]$$

$$\neq 0 \text{ for } -i = 1, \ldots, i-1, i+1, \ldots, n.$$

$$\partial_\beta E[m(w_i;\theta,\eta)][\beta - \beta_0] = E\left[(z_i \Lambda' \cdot \theta_x) \cdot (\beta - \beta_0)\right] \neq 0$$

where the directional derivative is defined in section 4. This implies that the moment condition in 1.2.7 is not robust to the local misspecification of the first-stage nuisance parameters. Therefore, the first-order biases in the nuisance parameter could affect the $\sqrt{N}$-consistency of the target parameter. Consequently, regularization and overfitting biases resulting from the use of machine learning estimators may lead to inconsistency in the parameters of interest θ.

To overcome these limitations, in the next section, I introduce so-called Neyman orthogonal moment functions that are insensitive to local misspecification from first-stage nuisance estimators η, based on the original moment function in equation (1.2.7). These methods can provide robust estimations of parameters of interest in the presence of high-dimensional nuisance parameters. Structural parameters of interest based on the developed method also achieve $\sqrt{N}$-consistency and asymptotic normality.

1.3 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 4.1 formally introduces the definition of Neyman orthogonal moment condition, while section 4.2 provides new moment conditions derived from the GMM equation (1.2.7). In section 4.3, the estimation procedure combined with the cross-fitting algorithm is outlined. The asymptotic properties of the proposed estimator are analyzed in section 4.4, and the estimator is extended to cover games with multiple players in section 4.5. Finally, section 4.6 evaluates the finite sample properties of developed estimators through a series of Monte Carlo simulations.

1.3.1 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:

$$E[\psi(W; \theta_0, \eta_0)] = 0. \tag{1.3.1}$$

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

$$\partial_\eta E_P[\psi(W; \theta_0, \eta_0)][\eta - \eta_0] := D_0[\eta - \eta_0], \ \eta \in \mathcal{T} \tag{1.3.2}$$

for convenience.

Definition 1. *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 (1.3.1) 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*

$$\partial_\eta E_P[\psi(W, \theta_0, \eta_0)][\eta - \eta_0] = 0 \text{ for all } \eta \in \mathcal{T}_N. \tag{1.3.3}$$

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

1.3.2 Orthogonal Moment Condition for Static Game: Two Players Example

For illustrative purposes, I focus on two-player games, and multi-player settings are given in the Appendix A.2. I construct the moment function that satisfies the orthogonality condition defined in section 4.1 by adding the "bias correction terms" to the original moment function (1.2.7). 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 the orthogonal moment function.

Let $f_i \equiv \sqrt{\Lambda(\cdot)(1-\Lambda(\cdot))}$ where Λ denotes the choice probabilities of being active, induced by the original moment function. The construction of the Neyman orthogonal moment function is based on the linear projection of z_i on $x_i = (s_i, s_x)$ with weighting f_i, similar to Belloni et al. (2016).

$$f_i z_i = f_i s_i' \mu + u_i, \;\; E[f_i s_i u_i] = 0 \tag{1.3.4}$$

Then, the orthogonal moment function satisfies $E[\psi(w; \theta_0, \eta_0)] = 0$ where

$$\begin{aligned} \psi(w_i; \theta, \eta) &= m(w_i; \theta, \eta) + \phi(w_i; \theta, \alpha, \eta) \\ m(w_i; \theta, \eta) &= (z_i - x_i'\mu)\left[a_i - \Lambda(\gamma_{-i}, \theta_\gamma, \beta)\right] = \mu_z \left[a_i - \Lambda(\gamma_{-i}, \theta_\gamma, \beta)\right] \\ \phi(w_i; \theta, \alpha, \eta) &= -E[\mu_z \Lambda(\cdot)(1-\Lambda(\cdot))\theta_\gamma | s_i, s_{-i}, s_x](a_{-i} - \gamma_{-i}) = \alpha(a_{-i} - \gamma_{-i}) \end{aligned} \tag{1.3.5}$$

where $\mu_z = (f_i z_i - f_i x_i'\mu)/f_i = z_i - x_i'\mu$ and $\alpha = E[\mu_z \Lambda(\cdot)(1-\Lambda(\cdot))\theta_\gamma | s_i, s_{-i}, s_x]$. This moment function defined above satisfies the orthogonality condition.

Theorem 1.3.1. *The moment function (1.3.5) obeys the Neyman orthogonality condition.*

Theorem 1.3.1 states that the moment condition $E[\psi(w_i; \theta, \eta)] = 0$ identifies the true parameter and is insensitive to misspecification of η in the neighborhood of η_0. The proof of Theorem 1.3.1 can be found in the Appendix.

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

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 D.2.3. Note that the estimators of nuisance parameters are required to have convergence rates that 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 (1.3.5) 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 $\hat{\eta}_k$ using I_k^c.
   a. Obtain $\hat{\gamma}_{-ik}$ using the ML Classifier of a_{-i} on s_{-i}, s_i and s_x.
   b. Obtain $\hat{\beta}_k$ using the Logit Lasso estimator of a_i on $\hat{\gamma}_{-i}, s_i$ and s_x.
   c. Compute $\hat{\theta}_{\gamma k}$ from original moment function 1.2.7.
   d. Compute the conditional densities $\hat{f}_k$.
   e. Estimate $\hat{\mu}_{zk}$ using the Lasso estimator of $\hat{f}_k z_i$ on $\hat{f}_k x_i$.
   f. Collect $\hat{\eta}_k = (\hat{\gamma}_{-ik}, \hat{\beta}_k, \hat{\mu}_{zk})$.
3. Construct the estimator $\hat{\theta}_\gamma$ as the solution to

$$\frac{1}{K}\sum_{k=1}^{K} L_{n,k}(\theta_\gamma, \hat{\eta}_k) = 0$$

where $L_{n,k}(\theta_\gamma) = \dfrac{\{E_{n,k}\left[\psi(\theta_\gamma, \hat{\eta}_k)\right]\}^2}{E_{n,k}\left[\psi(\theta_\gamma, \hat{\eta}_k)^2\right]}$ 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 $\psi(w_i; \theta, \eta) = \mu_z\left[a_i - \Lambda(\theta_\gamma, \beta_d, \beta_x)\right] - \alpha[a_{-i} - \gamma_{-i}]$ where $\alpha = E[\mu_z \Lambda(\cdot)(1 - \Lambda(\cdot))\theta_\gamma | s_i, s_{-i}, s_x]$.

1.3.4 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 1.3.1 (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 1.3.2 (Regularity Condition)**.**

i) $W_i = (A_i, s_{-i}, s_i, s_x)$ *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) $E[\{Y_{-i} - \hat{\gamma}(s_{-i}, s_i, s_x)\}^2 | s_{-i}, s_i, s_x]$ *and* $\hat{\alpha}$ *are bounded.*

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

Theorem 1.3.2. *Suppose that assumptions B.1. and B.2. in Online Appendix B hold. For* $V = M^{-1} E[\psi_0(W)\psi_0(W)'] M^{-1}$*, the DML static game estimators constructed in orthogonal moment conditions in 1.3.5 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 1.3.2 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 B.2 Online Appendix. 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 1.3.2 can be found in the Online Appendix A.

1.3.5 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.

1.3.5.1 The static entry/exit model and equilibrium

I simplify the model described in section 3 into two players, normalizing the payoff to be inactive to zero. The payoff of player i is a function of common market characteristics s_x for both players i and $-i$, rival's choice probabilities σ_{-i}, and a player-specific variable d_i, expressed as follows:

$$\pi_i(a_i = 1|\gamma_{-i}, s_i, s_x; \theta_\gamma, \beta) = \sigma_{-i}(a_{-i} = 1|s_i, s_{-i}, s_x)\theta_\gamma + \beta_\kappa s_i + s_x'\beta_x.$$

Under Type 1 Extreme Value distribution assumptions, conditional choice probability can be expressed in terms of relevant state variables and choice probabilities σ. I can solve equation (1.2.4) via a fixed-point algorithm as it contains two unknowns and two equations for two-players games. 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.

1.3.5.2 Data Generation

Using the best response function 1.2.4, 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 $\theta_\gamma = -1.5$, $\beta_\kappa = 2$.[3] For common market characteristics, I set $s_x'\beta_s = x_{s1}\beta_{s1} + x_{s2}\beta_{s2}$ where $\beta_{s1} = 0.8$ and $\beta_{s2} = 1.4$. 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 $s_x = (s_{x1}, s_{x2})$ and player-specific shifter s_i, s_{-i} from a uniform distribution with mean zero and variance one. [4]
2. I draw probabilities from a uniform distribution on $[0,1]$ and player i chooses to

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

[4] 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.

be active ($a_i = 1$) if the draw is less than or equal to the probability of being active, $\sigma_i(a_i = 1|s_i, s_{-i}, s_x)$.

When implementing the Plug-in estimator and Orthogonal estimator, I additionally include many covariates s_0 with $dim(s_0) = 500$ in common market characteristics. s_0 is drawn from a standard normal distribution, then further normalized to mean zero and variance one.

1.3.5.3 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 s_x as common market characteristic and shifter (s_i, s_{-i}). 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 (s_x, s_0) as common market characteristic and shifter (s_i, s_{-i}) and imposing regularization to Bajari et al. (2010b). I estimate first-stage conditional choice probabilities using the Logit Lasso estimator[5] and obtain $\hat{\theta}_\gamma$ adopting regularization to GMM estimation using equation (1.2.7).

Orthogonal estimator: Estimator of $\hat{\theta}_\gamma$ using (s_x, s_0) as common market characteristic and shifter (s_i, s_{-i}) and using orthogonal moment condition. I estimate first-stage conditional choice probabilities using the Logit Lasso estimator and use

[5] 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.

the orthogonal moment condition in the second stage, given in section 4.3. I also employ a cross-fitting algorithm with $K = 5$ as described in section 4.3.

1.3.6 Simulation Results

Table 1 summarizes the Monte Carlo Simulation results. The histogram of the simulation result with $(N, p) = (2000, 500)$ is illustrated in Figure 1. 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.

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.

1.4 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 valid 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 extend DML with dynamic games, utilizing zero-jabocian properties in efficient pseudo-likelihood (E-NPL) methods Dearing and Blevins (2019).

1.5 Proofs

Proofs of Theorem 1.3.2.

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

Assumption 1.5.1 (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 1.5.2 (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 1.5.3 (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 1.5.4.

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 1.5.5 (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, $\int |\partial g_j(w, \hat{\gamma}, \theta_0)/\theta_k - \partial g_j(w, \gamma_0, \theta_0)/\partial_k| F_0(dw) \xrightarrow{p} 0$

Proof of Assumption 1.5.1:

Assumption 1.5.1 part (i) is implied by Assumption 1.3.2. 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 1.3.2

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

Assumptions 1.5.1 part (ii) holds by the triangle inequality.

For part (iii), again, by the assumption 1.3.2

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

Assumptions 1.5.1 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 1. If Assumption 1.3.2 holds, then,

$$\hat{\theta}_\ell = \theta_0 + O_p(n^{-s_1}).$$

Proof of Lemma 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 2. If Assumption 1.3.2 holds, then

$$|\hat{\alpha}_{k\ell}-\alpha_{h0}| = O_p(n^{-s_1})$$

Proof of Lemma 2:

$$|\hat{\alpha}_{h\ell}-\alpha_{h0}|\leq C|\hat{\theta}_\gamma-\theta_0|+|\hat{\gamma}_{h\ell}-\gamma_{h0}|+|\hat{\mu}_\ell-\mu_0|\leq O_p(n^{-s_1}).(\because \text{triangle inequality}).$$

Proof of Assumption 1.5.2: 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 2 and assumption 1.3.1,

$$\sqrt{n}\int ||\hat{\Delta}_\ell||F_0(dw) \leq \sum_h \sqrt{n}||\hat{\alpha}_{h\ell} - \alpha_{h0}||\ ||\hat{\gamma}_{h\ell} - \gamma_{h0}||$$

$$= O_p(\sqrt{n}n^{-s_1}\ n^{-s_1}) = o_p(1)\ (\because -\frac{1}{2} < -2s_1 < -1) \text{ by the assumption 1.3.1 .}$$

For part (ii),

$$\int ||\hat{\Delta}||^2 F_0(dw) \leq \sum_h ||\hat{\alpha}_{h\ell} - \alpha_{h0}||^2||\hat{\gamma}_{h\ell} - \gamma_{h0}||^2 = O_p(n^{-2s_1}) = o_p(1)$$

It follows that triangle inequality.

Proof of Assumption 1.5.3: The first condition follows by assumption 1.3.1 for Lasso and other machine learning methods. For the second condition, taking Taylor approximation for nuisance parameters $\hat{\gamma} = (\hat{\gamma}_{-ik}, \hat{\mu}, \hat{\beta})$:

$$\begin{aligned}
\bar{\psi}(w, \hat{\gamma}, \alpha_0, \theta_0) :=& \int \hat{\mu}\{a_i - \Lambda(\hat{\beta}, \theta, \hat{\gamma}_{-i})\} + \int \alpha_k[a_{-i} - \hat{\gamma}_{-i}]F_0(dw),\\
=& \int (\hat{\mu} - \mu_0)[y - \hat{G}]F_0(dw) + \int \hat{\mu}\hat{G}'(\hat{\beta} - \beta_0)F_0(dw)\\
+& \int \hat{\mu}\hat{G}'(\hat{\gamma}_{-i} - \gamma_{-i0})F_0(dw)\\
\leq& O_p(n^{-2s_1}) + O_p(n^{-2s_1}) + O_p(n^{-2s_1}) = O_p(n^{-2s_1}) = C||\hat{\gamma} - \gamma_0||^2.
\end{aligned}$$

Proof of Assumption 1.5.4: 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 1.5.3.

Proof of Assumption 1.5.5: 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.*

1.6 Tables

Table 1: Simulation Results

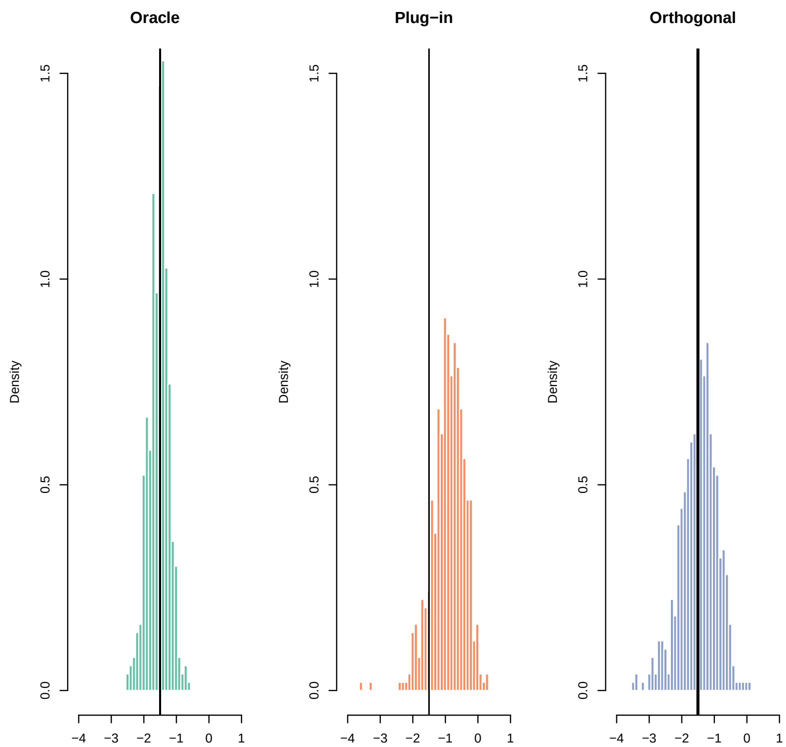
	Oracle estimator				Plug-in estimator				Orthogonal estimator			
	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
$(N, dim(s_0))$	Mean bias	Bias(%)	CP	RMSE	Mean bias	Bias(%)	CP	RMSE	Mean bias	Bias(%)	CP	RMSE
(1000,500)	-1.595	-6.323	0.960	0.522	-1.013	32.5	0.736	1.025	-1.529	-1.901	0.880	0.936
	(0.410)				(0.600)				(0.615)			
(1500,500)	-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)			
(2000,500)	-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 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.

1.7 Figures

Figure 1: The distribution of the estimated structural parameters from simulation

Sample Size: 2,000, Dim (s_0)=500



Notes: The estimated effect of rival coefficients is based on 500 simulations. The true value of the rival effects is $\theta_\gamma = -1.5$. The Oracle method employs low-dimensional (three-dimensional) relevant covariates, which serve as infeasible estimators under the assumption that econometricians possess knowledge of the true identity of controls. 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 and cross-fitting algorithm.

2.0 Rural Pharmacy Access and Competition

2.1 Introduction

Pharmacies have played pivotal roles in community healthcare, offering services beyond medication dispensing, such as immunizations, chronic disease care, and substance use treatment. Easy access to nearby pharmacies is essential for ongoing medication availability and following prescriptions per doctor's note. The literature documents that the absence of nearby pharmacies and travel difficulties can greatly reduce how well individuals, especially older ones, adhere to their healthcare providers' recommended behaviors. (Amstislavski et al. 2012; Qato et al. 2014; Di Novi et al. 2020). Moreover, approximately 65.1% of the elderly population in the U.S. takes multiple medications (Young et al., 2021). Therefore, understanding barriers to pharmacy access and related health disparities is essential for comprehending and improving health policies.

This paper provides the first empirical evidence that rural towns have experienced an increase in limited pharmacy access[1] over the past two decades in the Midwestern United States, a trend closely linked with the exit of independent pharmacies. Often, in these rural towns, independent pharmacies serve as the sole providers, and their closure results in restricted access to local pharmacy services. I document that the entries of chain pharmacies in urban towns induce the exit of independent pharmacies in nearby rural towns. To understand the role of chain pharmacy entries in shaping rural pharmacy market structure, I employ event studies using staggered two-way

[1]Limited pharmacy access, or "pharmacy desert" refers to when there is no pharmacy in the town, following Qato et al. (2014). Section 2.3 describes detailed background and discussion.

fixed effects (de Chaisemartin and D'Haultfoeuille 2023; Callaway and Sant'Anna 2021;Borusyak et al. 2021; Sun and Abraham 2021) with panel dataset spanning 2000-2019. I find that new entry of chain pharmacies within 15 miles has led to a significant decrease in the number of independent pharmacies, limiting the accessibility of pharmacies in rural towns. The magnitude of the impact of new chain pharmacy entries is larger in towns with a higher proportion of elderly residents who are particularly vulnerable to the issue.

I use the existing structural games of Bajari et al. (2010b) to decompose the competition effects of chain pharmacies and another independent pharmacy on the underlying profits of pharmacies in rural towns. I model binary entry/exit games among independent pharmacies, capturing strategic interactions with rival independent pharmacies and the competitive impact of chain pharmacies.[2] In this context, researchers face the challenge of selecting many market covariates to include in the model, to avoid issues such as omitted variable bias and multicollinearity. This process also requires knowledge of the correct functional form for pharmacy profits and the appropriate transformation/interaction of selected market covariates. The selection of variables and specification of functional forms, often carried out on an ad-hoc basis, present significant challenges in practical research.[3] To address these challenges, researchers frequently adopt a data-driven approach, using machine learning (ML) methods like the Lasso estimator, which selects relevant market covariates associated with a store's profits. However, ML methods suffer from regularization bias; for instance, the Lasso estimator may have model selection errors such as selecting irrelevant covariates or not selecting relevant covariates. This regularization bias from

[2]Appendix E.1 explains the rationale behind choosing entry and exit as forms of strategic interactions.

[3]Appendix E.2 provides detailed discussion.

ML methods can be transmitted into the low-dimensional parameters of competition effects, resulting in our primary parameter of interest being biased.

I compare the results from using existing methods with those from the developed methodology, which uses ML methods for estimating high-dimensional nuisance parameters including beliefs about competitors' choices and market covariates effects. My developed estimator offers two key benefits in recovering profits for independent pharmacies. First, it allows for more systematic variable selection and flexible functional form specification, avoiding ad-hoc choices. Second, in contrast to the existing methods Bajari et al. (2010b), using ML methods significantly enhances the prediction accuracy in estimating beliefs about rivals' conditional choice probabilities (CCPs). While I do not observe the true coefficient, the estimated structural parameter for the rival's effects on the profits in the second stage is approximately 50% larger than existing estimates. This change in estimates can be primarily attributed to the ability of ML 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 towns with a high elderly population ratio. The absence of chain entries scenario quantifies the role of new chains' entries on the market structure. To do this, I simulate the equilibrium and how many independent pharmacies would be active in the market if the number of chain pharmacies was fixed in the year 2000. The absence of new chain entries scenario 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.

Next, the subsidy counterfactual characterizes the equilibrium where the federal government provides a 10% subsidy for pharmacy sales associated with Medicare beneficiaries to pharmacies in towns with a high elderly population ratio. This hypothetical subsidy program is inspired by the existing federal government's physician bonus program initiated in 2006 to enhance healthcare accessibility, targeting areas with limited medical access. 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.

My results suggest that limited pharmacy access might have a heterogeneous impact across demographic groups and socioeconomic statuses. The elderly population, especially those with limited mobility and relatively higher transportation costs, may face challenges in correctly managing multiple medications. On the other hand, groups with better transportation access and younger demographics may benefit from the entry of chain pharmacies, which often offer competitive pricing and higher-quality services. These findings shed light on the broader discussion of health disparities and the need for targeted public policies to improve pharmacy access in rural towns with a high portion of the aging population.

2.1.1 Literature Review

The primary contribution of this paper is to provide the first evidence of how chain pharmacy entries impact the rural market structure and access to pharmacies. Previous public health literature has focused on the effect of restricted pharmacy access on negative health outcomes and examined related health equity issues (Amstislavski et al. 2012, Qato et al. 2014, Di Novi et al. 2020). Buchmueller et al. (2006) examined the effect of restricted hospital access on deaths from heart attacks and unintentional injuries. However, there is scarce literature focusing on the mechanism leading to the rise in limited pharmacy access. I contribute to the literature by providing empirical evidence of mechanisms: increased entries by chain pharmacies in urban towns are associated with the exit of independent pharmacies in nearby rural towns, leading to limited pharmacy access in rural towns. I further document that towns with a high elderly population experience more rapid growth in limited accessibility - a concerning trend given their higher transportation costs and limited mobility - and the effect of new chain pharmacies on independent pharmacies is larger in those towns. To the best of my knowledge, these empirical findings are the first in the literature.

Secondly, this paper relates to the effects of firm entry on economic outcomes, market structure, spatial competition, and associated policy questions (Jia 2008, Ellickson and Grieco 2013, Grieco 2014, Caoui et al. 2022). While previous literature has focused on the grocery industry and associated food deserts (Chenarides et al. 2021, Lopez et al. 2023), I contribute to the understanding of the market structure of pharmacy industries and their role in shaping limited pharmacy access in rural U.S.

2.2 Data and Background

In this section, I provide the background of the pharmacy industry, data sources, limited pharmacy access in rural towns, and descriptive statistics

2.2.1 Industry Background

Pharmacies industry has begun with 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.

However, the industry landscape began to shift in 1970 when chain pharmacies (e.g., Walgreens, CVS, Rite Aid) and mass merchandise pharmacies (e.g., Walmart, Sam's Club, Target) 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 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. Chains, mass merchandise, and supermarket pharmacies (e.g., Kroger, Publix) often offer competitive prices and might offer differentiated products in terms of better health insurance coverage, by leveraging their bulk purchasing power, substantial bargaining power against health insurance companies, and vertical relation with health insurance companies. For example, Walmart launched a $4 generic prescription program in 2006.

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%. [4] Although the mail order market share steadily rose to 15% in 2008, then their market share had reverted to 13.7% in 2018. I abstract away mail-order in the analysis because it takes a relatively smaller portion of the market share.

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

In the 2000s, there was the continuing expansion of both merchandise-based pharmacies and supermarket-based pharmacies, 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 in the United States.[5]

2.2.2 Data

I combine data from multiple sources to construct the final dataset, in which geographic units of county subdivisions (that I refer to as 'towns') defined by the Census are markets.[6]

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, and is annually updated. Since Data Axle includes the addresses of each pharmacy store, I can assign these addresses to townships using Python's Geopandas. Furthermore, the panel data structure enables me to define entry and exit every year. This dataset has been used in recent studies such as Dearing and Blevins (2019) and Koh (2023).

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 Survey (CPS) and

[5]Source: 2020 NATIONAL COMMUNITY PHARMACISTS ASSOCIATION DIGET

[6]I explain detailed reasons for defining the geographic market at the town level in the subsequent section.

ZIP Code Business Patterns dataset. Appendix D provides more detailed information on data construction. Given that the nearest Census data is available from 2000 onward, my analysis covers the period 2000-2019.[7]

2.2.2.1 Market Definition of Geographic Level

I define a geographic market based on townships that had pharmacies at any point between 1997 and 2021.[8] I choose the township (county subdivisions) as the geographic unit for the following reasons: First, as reflected in the survey results shown in Appendix 19 and Appendix 20, 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. [9] Second, my market definition follows earlier healthcare studies that used towns Schaumans and Verboven (2008).[10] Third, 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. Fourth, the market definition of town level could be suitable as I provide reduced-form evidence in section 5.2, which suggests that

[7]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.

[8]If any townships had no pharmacy at any point between 1997 and 2021, these were automatically dropped from subsequent analysis in logit or other types of discrete choice models.

[9]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.

[10]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.

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.

2.2.2.2 Final Sample

For a township to be included in the dataset, it must: (i) not overlap with the Census-defined urbanized areas, 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; (v) not have had more than seven chain pharmacies within a 15-mile radius.[11]

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.

2.2.3 Limited Pharmacy Access in Rural Towns

In this section, I document recent trends in the number of towns with limited access to pharmacies. These areas are sometimes referred to as “limited access to

[11] 99.30% of the rural townships in my sample have at most two pharmacies operating simultaneously.

pharmacy areas" or "pharmacy deserts." The term "pharmacy desert" is inspired by the concept of a "food desert" in the literature, 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. In such towns, consumers must travel several miles to obtain prescriptions.

I describe areas with "limited access to pharmacy" as those townships without any pharmacies, following the approach in Qato et al. (2014). Qato et al. (2014) use census tracts as a geographical reference, which is 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 pharmacy access can cause 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 documented in Di Novi et al. (2020). For a comprehensive discussion on the impact of limited pharmacy access on negative health outcomes and increased health care costs, I refer to Di Novi et al. (2020). [12]

Figure 2 illustrates the escalating trends in the number of towns with limited access to pharmacy in 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 Appendix 21, I present alternative definitions of lim-

[12]Non-adherence issues occur 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.

ited pharmacy access, considering population weights and a 5-mile distance. These alternative specifications show qualitatively similar trends.

I also find substantial heterogeneity in limited access to the pharmacy by the elderly population share. Figure 3 illustrates that non-high elderly population townships maintained relatively stable figures, while high elderly population townships - those where more than 20% of the population is aged 65 or older - marked increase in limited pharmacy access, rising from 14.29% in 2000 to 25.44% in 2019. Given that high elderly population townships are particularly vulnerable to pharmacy access challenges, it could arouse health/public policy concerns over prescription access in high elderly population towns.

2.2.3.1 Summary Statistics

Appendix 13 and 14 provide descriptive summary statistics of my final sample, which comprises 291 non-high elderly population townships and 511 high elderly population towns for twenty years, with a total of 16,040 market-level observations (802 towns * 20 years).[13] High elderly population townships typically have a smaller population than non-high elderly population townships, which typically results in lower market demand. As chain pharmacies prefer to enter markets with higher demand, high elderly population 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

[13]For a full list of variables, Appendix 13 and 14 provide descriptive statistics of my final sample. I show key selected variables for brevity in main Table 21 and 22.

for both groups of townships. In contrast, the number of independent pharmacies in high elderly-population townships declined, a trend not observed in non-high elderly-population townships.

In the subsequent section, I study the mechanisms driving the increasing trend in pharmacy deserts through the lens of competition, with a particular emphasis on the entry of chain pharmacies.

2.2.4 Market Structure

The entry of chain pharmacies has significantly 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 include several formats: standalone retail pharmacies (e.g., Walgreens), supermarket-based pharmacies (e.g., Kroger Pharmacy), and merchandised-based pharmacies (e.g., Walmart 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 focuses on the "Superior 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, 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.

Figure 5 demonstrates the overall negative correlation between independent and chain pharmacies by showing the average number of stores within the same area over the period from 2000 to 2019. Within towns, the average number of independent pharmacies decreased by 0.18, while the average number of chain pharmacies increased by 0.13 units, resulting in a net decrease of 0.05 units. In accordance with the anecdotal evidence presented in Figure 4, the average number of chain pharmacies located up to 15 miles outside of towns increased by 0.38 units. This suggests that new chain pharmacies tend to be established in relatively urbanized areas or nearby shopping malls that are distant from rural towns.

In Appendix 22, I also examine changes in the number of independent/chain pharmacies in high elderly population townships and non-high elderly population 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 heterogeneously impact independent pharmacies by township demographics.

Finally, Appendix 23 shows changes in the market structure of independent pharmacies by towns. Specifically, it examines the distribution of townships unserved, monopolies, and duopolies among independent pharmacies. For both high elderly-population townships and non-high elderly-population townships, monopolies are decreasing, while unserved areas are increasing. The changes are greater in high elderly population townships, which is aligned with findings in Appendix 22.

2.2.5 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.

2.2.5.1 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 radius from the centroid of towns. Appendix 15 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 negligible as they are not statistically significant.

2.2.5.2 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:

$$Y_{mt} = \sum_{\tau} \delta_{\tau} \text{Entry}_{m,t-\tau} + \beta X_{mt} + \lambda_m + \alpha_t + \gamma_{st} + \epsilon_{mt} \tag{2.2.1}$$

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. [14]

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 potential 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 overcome these concerns, I adopt generalized event study frameworks that detect possible pre-trends as well as are 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

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

different dates). Figure 6 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. In contrast, post-treatment shows that the entry of chain pharmacies leads to a decreased number of independent pharmacies in the township. In Appendix 24, 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.[15]

Next, I examine the heterogeneity by market demographics of age distribution, high elderly population townships in Figure 7a and non-high elderly population townships in Figure 7b. Consistent with earlier findings, the effects of new chain entries are quite large in high elderly population townships. The impact of new chain pharmacies on the number of independent pharmacies is smaller in non-high elderly population townships and their dynamic effects are negligible after three years of event. This suggests that the elderly population is more price-sensitive, as chain pharmacies typically offer competitive prices.[16] I interpret that these competition effects might be significantly different between high elderly population towns and non-high elderly population towns, so in the structural analysis in the section 5, I separately recover parameters of interest in the two sets of different aged-population township types;

[15] This empirical illustration and the results should be interpreted cautiously, as the current framework does not capture exit events of chain pharmacies. Unlike policy treatment effects, chain pharmacies often enter and exit the market quickly if it is less profitable. To the best of my knowledge, I have not found a methodology suitable for my setting, which allows for continuous and multiple treatments, as well as (multiple time) switchers. I also caution that the entry effects of chain pharmacy are immediate, unlike dynamic treatment effects like Caoui et al. (2022). One potential channel is for chain pharmacies to enter the market by acquiring independent pharmacies. However, data shows that only around 10% of closed independent pharmacies are transformed into new chain pharmacies, which mitigates this concern.

[16] Given their bargaining power with insurance companies, national chains might offer lower prices, and they can also obtain more discounts through bulk contracts.

high elderly population towns and non-high elderly population towns.

2.2.6 Preliminary Analysis on Strategic Interaction between Independent Pharmacies

To demonstrate how entry and exit patterns change with the endogenous competition with the rival's independent store, I provide reduced form evidence of strategic interactions using simple logit regressions. An independent firm in market m makes a binary decision of each firm a_{imt} where $a_{imt} = 0$ if firm i is active in market m period t and $a_{imt} = 1$ if i being inactive. As these regressions do not take into account the simultaneous entry of rival firms, the results do not reveal causality, but correlation.

Appendix 16 presents the results of a logit model on entry, controlling for the presence of chain pharmacies within a 15-mile radius, demographic variables, and health-related characteristics. Additionally, I have included a binary variable indicating whether states adopted the Medicaid expansion policy after 2014. Firstly, the presence of rival stores in the same town is strongly and negatively correlated with entry decisions.

This suggests that the presence of a rival may significantly decrease the latent payoff of being active in the market. Since difference-in-differences estimators, as discussed in Section 2.5, are unable to capture endogenous rival effects, this motivates the use of a structural model in which I can separate the competition effects stemming from other independent pharmacies and chain pharmacies.[17] Secondly, the number of chain pharmacies is negatively correlated with entry decisions, though this correlation is weaker than the impact of rivals. This implies that chain pharmacies might offer distinct types of services (e.g., higher quality, better in-network premi-

[17]More specifically, in the DID model, the dependent variable represents the number of independent pharmacies so I cannot capture competition effects between independent pharmacies.

ums), positioning them more as secondary competitors. Interestingly, my findings align with those of Grieco (2014), which found that the effect of chain supermarkets is less pronounced than that of independent groceries.

Games played between independent pharmacies are endogenous in that independent pharmacies' optimal choices are the best response to other beliefs over the probability of rival's choice probabilities. To correctly recover underlying payoffs with strategic interactions, it requires pharmacies' entry and exit game, detailed in the next section.

2.3 Structural Analysis

In this section, motivated by reduced-form analysis in section 2.6, I present a structural model of independent pharmacies and estimation strategies, and then report the estimation results of the model.

2.3.1 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, in my data, 99.30% of towns contain at most two operating independent pharmacies.

I assume that the information structure of the games between independent pharmacies is characterized by incomplete information, as detailed in Assumption 1.2.1. Each player observes her own private information but cannot observe her rival's. Instead, she knows the distribution of her rival's private information.

I further assume that, from the perspective of independent pharmacies, the entry of chain pharmacies is given. This assumption is reasonable for rural independent pharmacy markets, as the decision for chain pharmacies to enter these areas is likely driven by broader regional demographics, network structures, and the locations of their distribution centers. The competition from small independent stores is less of a concern for the national chain pharmacies. Therefore, the model is greatly simplified by treating decisions made by national chains as given for local independent stores. This approach aligns with methodologies used in other studies examining strategic interactions between local stores (Ackerberg and Gowrisankaran 2006, Grieco 2014). [18] Based on the empirical evidence in Table 15, the analysis focuses on the number of chain pharmacies within a 15-mile radius of town centers.

The store i's choice-specific value function u when active in the market depends on the beliefs over a rival's conditional choice probabilities (CCP) of actions a_{-i} in market m at period t:

$$\begin{aligned}&u_{imt}(a_{imt}=1, s_{mt}, \epsilon_{imt}(1); \theta) = \pi_i(a_{imt}=1, s_{mt}; \theta) + \epsilon_{imt}(1).\\&= \sigma_{-i}(a_{-imt}=1|s_{mt})\pi_i(a_{imt}=1, a_{-imt}, s_{mt}; \theta)\\&+ (1-\sigma_{-i}(a_{-imt}=1|s_{mt}))\pi_i(a_{imt}=1, a_{-imt}, s_{mt}; \theta) + \epsilon_{imt}(1).\end{aligned} \tag{2.3.1}$$

where π_i is the expected payoffs and $\sigma_{-i}(a_{-imt}=1|s_{mt})$ is the probability of rival's being active, conditional on observable market characteristics s_{mt}. θ denotes the set of structural parameters affecting the pharmacy's per-period payoff, and $\epsilon_{imt}(1)$

[18] An alternative approach involves a comprehensive model that accommodates the endogenous entry of both independent and chain pharmacies. For merchandise-based pharmacies such as Walmart, entry decisions are influenced by factors like existing merchandise department stores, distribution centers, and network effects with nearby stores. Incorporating these elements, however, requires complex methodologies similar to those in Holmes (2011) or Jia (2008). Nonetheless, these models have their constraints, particularly in ignoring the strategic interactions among chain pharmacies. This paper, however, primarily investigates the impact of chain pharmacy entry on rural markets rather than the expansion strategies of these chains. Thus, a detailed structural model encompassing the endogenous entry of chain pharmacies is outside this paper's scope.

denotes being active-specific private information for pharmacy i. I further assume that the value of being inactive is normalized to zero.

2.3.1.1 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 coupled with 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. As econometricians cannot observe private information, the optimal strategy can be expressed as choice probabilities:

$$\sigma_i^*(a_{imt} = 1 | s_{mt}; \theta) = \frac{\exp(\pi_i(a_{imt} = 1, s_{mt}; \theta))}{1 + \exp(\pi_i(a_{imt} = 1, s_{mt}; \theta)} \tag{2.3.2}$$

2.3.2 Discussion

Before presenting the estimation procedure and the results, I detail the assumptions underlying the structural model.

2.3.2.1 Static versus Dynamic Framework

I model the discrete choice of an independent pharmacy as a static game for the following reasons. First, a dynamic model might be more appropriate if the entry costs, including fixed sunk costs, of entering the industry are substantial. A dynamic model, capable of distinguishing sunk costs from fixed costs, accommodates forward-looking behaviors observed in industries like cement (Ryan (2012)) and hardware (Igami and Uetake (2020)). However, anecdotal evidence from pharmacy industry reports, as shown in Appendix 17, suggests that the sunk costs of opening independent pharmacies are relatively small compared to their yearly gross profits, [19] which 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)). Second, Appendix 19 shows a regression analysis of past entries of chain pharmacies on the number of independent pharmacies in township m during period t. The results indicate that controlling for the number of chain pharmacies in the same year, the effect of chain pharmacies in the past year on current independent pharmacies' payoff is negligible. Therefore, based on this observation, I model that a player's payoff depends on current state variables.

2.3.2.2 Profit Shifters for Identification

To achieve identification, following the approach discussed in Bajari et al. (2010b), I adopt firm-specific variables previously utilized in the existing literature. [20] For instance, Grieco (2014) employed the existence of a two-year-old firm as a shifter variable for independent grocery stores. As the lagged variable of entry might be endogenous to the current rival's payoff, I use the employment size from three years prior as an alternative variable. This exclusion restriction is considered valid if an operating store's revenue increases with its number of employees, yet remains unrelated to the profits of rival stores. The underlying assumption here is that a store's profit is influenced by its own employment levels, whereas the employment levels of

[19] 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%.

[20] Ideally, distances to headquarters or distribution centers are extensively used in IO literature (e.g., Chen 2014, Xie (2022).

rival stores affect a store's profit solely through the rivals' decisions to be an active market. [21] In section 5.4, I will provide further discussion on the suggested shifter for independent pharmacies.

2.3.2.3 Isolated Market Assumption

Following the approach pioneered by Bresnahan and Reiss (1991), the literature on strategic interaction typically focuses on isolated markets with a relatively small number of firms. The assumption might be invalid if independent pharmacies outside the township significantly influence the entry or exit decisions of pharmacies within the township. To address this concern, I observe that the average distance between town boundaries is 14 miles, suggesting sufficient separation between markets. This finding implies that most towns in my final sample approximate isolated markets. Furthermore, Appendix 20 displays a regression analysis of new independent pharmacy entries outside the township against the number within the town, factoring in market characteristics, town-fixed effects, year-fixed effects, and market-year interactions. Appendix 20 provides suggestive evidence that new entries of independent pharmacies in nearby towns (outside the township) may have minimal impact on those within the township.

2.3.3 Estimation Procedure: 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(conditional choice probabilities) from the data. In the second stage,

[21] Following this shifter, I decompose notation $s_{mt} = s_{imt}, s_{-imt}, s_{xmt}$, where s_{imt} denotes store i's specific shifter, s_{-imt} denotes store $-i$'s specific shifter, and s_{xmt} denotes the common market characteristics, which could be extended into many covariates.

I use the rival's CCP, observed market characteristics, chain pharmacy effects, and firm-specific shifters to construct moment conditions. Finally, these moment conditions are minimized over a set of candidate structural parameters. Additionally, I separately estimate the model for samples of high elderly population and non-high elderly population towns, acknowledging the substantial heterogeneity in market dynamics found in section 2.5.[22] It relaxes the assumption that only one equilibrium be played across all towns. Instead, I assume that a unique equilibrium be played in each town with a high elderly population and towns with a low elderly population ratio.

2.3.3.1 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) demonstrate that the first-stage estimation of choice probabilities should not be affected by the presence of unobserved and time-varying endogenous state variables. To address this issue, I employ three 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 assume 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 include state Medicaid expansion, a policy-relevant variable, as suggested instrument

[22] This approach is widely used in the industrial organization literature. (e.g., Ellickson and Misra (2008)).

variables by Berry and Compiani (2023). In the logit regression result of the independent pharmacy's entry in Table 16, 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 purposes, I employ two distinct estimators: the existing estimators as described by (Bajari et al., 2010b) which utilize pre-selected variables, and newly developed orthogonal estimators that leverage flexible machine learning (ML) methods with rich covariates. The primary distinctions between these two estimators lie in 1) the set of control variables used, 2) the application of the machine learning approach, and 3) the implementation of a cross-fitting algorithm. Detailed procedures for the existing estimator are provided in section 5.3.1, and for the orthogonal estimators in section 5.3.2.

2.3.3.2 Existing Two-Step Estimators

First-stage nuisance parameter γ_{-i} estimation: The goal of the first stage 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:

$$\hat{\gamma}_{-imt} = E[a_{-imt}|c_{mt}, s_{-imt}, s_{imt}, s^{pre}_{xmt}, y_t, \text{county}_f] \tag{2.3.3}$$

where a_{-imt} denotes the rival's binary choice, c_{mt} represents the number of chains within 15 miles, s_{-imt} indicates the rival's shifter, which is the number of the rival's employees, s_{imt} denotes player i's shifter, which is the number of employees, s^{pre}_{xmt} denotes common market characteristics, y_t denotes the year fixed effects, and county_f

represents the county fixed effects. [23] For s_{xmt}^{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, and their summary statistics can be found in Appendix 21 for towns with a low elderly population and Appendix 22 for towns with a high elderly population, respectively. I employ a simple logit model to estimate the conditional expectation of equation (2.3.3).

Second Stage Structural Parameter Estimation: After recovering the equilibrium strategies from the data, the goal of the second stage is to estimate the structural parameters of interest. First, I model the profit functions in a reduced-form manner, following the conventional approach in static entry literature (e.g., Berry (1992), Seim (2006)), then describe the moment conditions to identify the relevant structural parameters. The average period profit per store in market m in period t is characterized as follows:

$$\begin{aligned} &\pi_{imt}(a_{imt} = 1, \hat{\gamma}_{-imt}, c_{mt}, s_{imt}, s_{xmt}^{pre}, y_t, \text{county}_f; \theta) = \\ &\hat{\gamma}_{-imt}\theta_\gamma + c_{mt}\theta_c + s_{imt}\beta_e + s_{xmt}^{pre} \cdot \beta_x + \alpha_t y_t + \alpha_{\text{county}}\text{county}_f \end{aligned} \tag{2.3.4}$$

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

As developed by Bajari et al. (2010b) and Bajari et al. (2013), I construct the logit-likelihood function, which depends on the profit function given in equation (2.3.4) and equilibrium function (2.3.2), and estimates a set of parameters:

$$\underset{\theta_\gamma,\theta_c,\beta_e,\beta_x,\alpha_t,\alpha_{\text{county}}}{\arg\min} \ \ln\mathcal{L} = \sum_t \sum_m \sum_i \ a_{imt}\ln\left(\Lambda\left(\pi_{imt}\right)\right) + (1 - a_{imt})\ln\left(1 - \Lambda\left(\pi_{imt}\right)\right) \tag{2.3.5}$$

[23] 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

Here, Λ represents the standard logit link function under Type 1 Extreme Value Distribution and π denotes the expected profit per period from equation (2.3.4). The consistency and asymptotic normality of the estimator are established by Bajari et al. (2010b). I account for correlation in error terms by taking the clustered standard error at the county level.

2.3.3.3 Neyman Orthogonal Estimators

Neyman Orthogonal estimators facilitate the use of flexible Machine Learning (ML) methods, which allow data-driven selection for covariates and flexible functional forms. Consequently, I do not pre-select socio-economic variables and instead utilize the pool of variables s_{xmt}^{pool}, as shown in the summary statistics in Appendix 13 and Appendix 14 and their flexible interactions terms $s_{xmt}^{interaciton}$. Here, I provide an overview of the estimation steps, and Appendix Section F describes additional details of the algorithm steps, as well as the functional forms of standard errors in the second-stage structural parameters.

Step 1. Estimation of nuisance parameters using ML: Let K denote a positive integer and take a K-fold random partition $I_1, ..., I_K$ of observation indices $\{1, \ldots, N\}$. I use random sample splitting with K=5 folds across the year cluster.[24] I also define the auxiliary sample $I_k^c = \{1, ..., N\}/I_k$ for each $k \in \{1, \ldots, K\}$. For each k, I use machine learning methods to estimate the set of nuisance parameters $\hat{\eta} = (\hat{\gamma}_{-i}, \hat{\beta}_e, \hat{\beta}_x, \hat{\mu}_\gamma, \hat{\mu}_c)$ only using observations not in the group k as $\hat{\eta}_k = \hat{\eta}\Big((W_i)_{i \in I_k^c}\Big)$.[25]

[24] I attempted to use the multi-way clustering method proposed by Chiang et al. (2022), but I found that county fixed effects are crucial for capturing time-invariant market characteristics when estimating β. Since county fixed effects learned in the training set cannot be applied to the test set (due to cross-fitting), I have chosen to use the year as the clustering level for the cross-fitting algorithm. The advantage of this approach is that it allows for the use of county-fixed effects to predict nuisance parameters in the test set.

[25] (μ_γ, μ_c) are additional nuisance parameters not in the original method. They are generated in

Step 2. Recovering structural parameters $(\theta_\gamma, \theta_c)$: Using the nuisance parameter estimates $\hat{\eta}_k$, I evaluate the original moment condition in equation (2.3.6) on the sample I_k. Finally, I obtain the Neyman Orthogonal estimator $\hat{\theta} = (\hat{\theta}_\gamma, \hat{\theta}_c)$ by aggregating the orthogonal moment functions for each $k \in \{1, \ldots, K\}$. Figure 8 illustrates the estimation algorithm.

$$\begin{aligned}
\psi(w_{imt}; \theta, \eta) &= m(w_{imt}; \theta, \eta) + \phi(w_{imt}; \theta, \alpha, \eta) \\
m(w_{imt}; \theta, \eta) &= (\mu_\gamma, \mu_c)\left[a_{imt} - \Lambda(\gamma_{-imt}, \theta_\gamma, \theta_c, \beta_e, \beta_x)\right] \\
\phi(w_{imt}; \theta, \alpha, \eta) &= -(\alpha_\gamma, \alpha_c)(a_{-imt} - \gamma_{-imt}) \\
\alpha_\gamma &= E[\mu_\gamma \Lambda(\cdot)(1 - \Lambda(\cdot))\theta_\gamma | c_{mt}, s_{imt}, s_{-imt}, s_{xmt}] \\
\alpha_c &= E[\mu_c \Lambda(\cdot)(1 - \Lambda(\cdot))\theta_\gamma | c_{mt-3}, s_{xmt}]
\end{aligned} \tag{2.3.6}$$

As a summary, Table 2 presents the specifications of the existing methods and the orthogonal estimators developed in each estimation step. Additionally, Appendix 25 reviews the notation for parameters to be estimated and the data.

2.3.4 Estimation Results

First Stage Estimation Results: Appendix 26 reports the estimation results of reduced-form CCP γ_{-i} based on existing estimators, Bajari et al. (2010b)'s estimator. Estimates suggest that player-specific shifters, and employment size in pharmacy show 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 good proxy for sales and higher quality provision by the pharmacy. As

the process of constructing the moment function to satisfy the orthogonality property.

expected, the impact of the total population appears to positively affect latent profits, as the total population might capture well for the overall size of the market. 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, the effect of chain pharmacies on profits is larger in high-elderly-population towns than in non-high elderly-population towns.

Next, I present the estimation results of predicting a rival's CCP using various ML methods. To shed light on the performance of different ML methods, Table 3 summarizes the findings of applying various procedures and reports the out-of-sample (hold-out sample) accuracy level. XG Boosting outperforms ordinary logit and other ML methods that are based on linear models.[26] This suggests that the flexible features of the tree-based model perform well in predicting the rival's CCP. Given this empirical pattern, I employ XG Boosting in my first-stage estimates of CCP.[27] Luo et al. (2016) provide and discuss the theoretical limits for ℓ_2 boosting models.

To shed light on the performance of XG Boosting, Appendix 18 presents the primary factors driving outcomes, as identified by the XG Boost model, in both the

[26]For comparison, in Appendix 25, I report the prediction performance in terms of the area under the curve (AUC). 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. This is demonstrated in the first-stage reduced-form CCP estimates for towns with high elderly populations, comparing Bajari et al. (2010b)'s simple logit method 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 23, I also present confusion matrices for both high-elderly-population towns and non-high-elderly-population towns. The accuracy in both cases exceeds 0.95, an outstanding performance given the complex nature of the games.

[27]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.

high elderly population and the non-high elderly population townships. For both town categories, employment numbers by store and rival stores stand out as top determinants. In high elderly population townships, the presence of chain pharmacies emerges as a notable contributor, while in non-high elderly population 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: Tables 4 and 5 present structural parameters with standard errors from the observed sample for towns with high and non-high elderly populations Standard errors account for correlations at the clustered county level. Column (1) in each table follows a baseline specification assuming known and linear market characteristics. Column (2), however, differs in both tables by incorporating: i) flexible functional forms for the rival's CCP in nuisance parameter estimation, ii) data-driven market characteristic selection, and iii) interaction effects in richer market covariates.

First, I compare the estimates from the existing estimators in column (1) with my orthogonal estimators in column (2) across towns with high elderly populations and non-high elderly populations. The orthogonal estimators in column (2) of Table 4 present that the rival effect from another independent pharmacy is 1.5 times larger than the existing estimators (Bajari et al., 2010b) in column (1) and their differences are statistically significant from Wald test[28]. This difference stems from the first-stage accuracy, where the beliefs about the rival's CCP in the first-stage ML of column (2) are 30% more accurate than in the simple logit of column (1).

[28]The Wald test, using estimated standard errors and coefficients, indicates the coefficients are statistically different at the 0.01 significance level for both high elderly population towns sample and non-high elderly population towns sample.

Furthermore, XG boosting effectively addresses non-linear complexities in the rival's entry and exit strategies and data-driven selection of market characteristics.

Second, I compare the strategic interactions from another independent pharmacy and the competitive effects of chain pharmacies. In towns with both high and non-high elderly populations, the effect of another local rival independent pharmacy on profit is larger than that of a chain pharmacy, which aligns with the findings of Grieco (2014). The difference between the two estimated parameters, derived from different moment conditions, is statistically significant from the Wald test. This is expected, as another rival independent pharmacy is more substitutable from the consumer's perspective than a chain pharmacy.

Third, I compare the estimates associated with competition between towns with high elderly populations and those with non-high elderly populations. The effect of another rival independent pharmacy on the value of running a store is more pronounced in towns with high elderly populations, aligning with Grieco (2014)'s findings. This is likely due to the smaller populations in these towns, where local markets often cannot support two independent pharmacies, especially in rural areas. Similarly, chain pharmacies have a smaller impact on latent profits in towns with non-high elderly populations, reflecting a pattern consistent with Grieco (2014) and previous empirical findings. Consequently, competition tends to be more intense in towns with high elderly populations due to smaller market sizes.

2.3.5 Robustness Check

I have conducted a series of robustness checks that are not reported in the main paper. 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 double/debiased machine learning literature. In Appendix 27, I demonstrate that utilizing a different number of cross-fitting folds yields quite similar results.

To examine whether the hyper-tuning parameters in XG Boosting might fail to adequately capture beliefs about the rival's choices, I try alternative hyper-tuning with cross-validation methods for the hyper-tuning parameters in XG Boosting. The results are qualitatively similar.

2.4 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 high elderly population towns have experienced rapid increases in limited access to pharmacies, I focus on high elderly population towns in my counterfactual scenarios.

2.4.1 Solution Method for the Static Game

To conduct counterfactuals in different scenarios, I first solve for the equilibrium of the model based on Equation (2.3.2). I employ a nested fixed-point algorithm, which solves the following system of equations:

$$\sigma_i(a_{imt}=1|c_{mt},s_{imt},s_{xmt}) = \frac{e^{\sigma_i(a_{-imt}=1|c_{mt},s_{imt},s_{xmt})\theta_\gamma+c_{mt}\theta_c+s_i\beta_\kappa+s_{xmt}\cdot\beta_x}}{1+e^{\sigma_i(a_{-imt}=1|c_{mt},s_{imt},s_{xmt})\theta_\gamma+c_{mt}\theta_c+s_i\beta_\kappa+s_{xmt}\cdot\beta_x}} \tag{2.4.1}$$

$$\sigma_{-i}(a_{-imt}=1|c_{mt},s_{-imt},s_{xmt}) = \frac{e^{\sigma_i(a_{imt}=1|c_{mt},s_{imt},s_{xmt})\theta_\gamma+c_{mt}\theta_c+s_{imt}\beta_\kappa+s_{xmt}\cdot\beta_x}}{1+e^{\sigma_i(a_{imt}=1|c_{mt},s_{imt},s_{xmt})\theta_\gamma+c_{mt}\theta_c+s_{imt}\beta_\kappa+s_{xmt}\cdot\beta_x}} \tag{2.4.2}$$

Here, Equation (2.4.1) denotes the conditional choice probability (CCP) of player i, and Equation (2.4.2) denotes the CCP of player $-i$. Given the two unknowns (σ_i and σ_{-i}) and the two equations (2.4.1) and (2.4.2), I use 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, $\epsilon = 0.00001$.[29]

2.4.2 Goodness of Fit

Figure 9 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 high elderly population towns, but the simulated outcome slightly over-predicts 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 28. In line with the overall trends, the predicted averages for stores closely resemble the observed counts.

[29] 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.

2.4.3 Chain Pharmacy Counterfactual: Fixing Entry of Chain Pharmacy afterward 2000

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 7 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.

2.4.4 Policy Counterfactual: Providing Subsidy associated with Medicare in High Elderly Population Towns

The second scenario investigates the potential outcomes of providing subsidies to independent pharmacies in high elderly population 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 of increasing access to pharmacies within high elderly population 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.[30] Subsequently, I calculate the counterfactual CCP using these adjusted profits for independently owned pharmacies. [31] 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.

Appendix 29 shows how the expected number of independent pharmacies 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 highly elderly population towns get a 10% subsidy associated with Medicare beneficiaries, markets would have on average 20% more independent pharmacies 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. On average, high elderly population 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

[30]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 and 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).

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

suggested subsidy program.

2.5 Conclusion

This study finds that machine learning methods can be used to improve the estimation of structural parameters in models of strategic interaction among independent pharmacies. The absence of new entries by chain pharmacies' counterfactual simulations suggests that new chain pharmacy entries can account for 40% of the variation in the closed independent pharmacies between 2000 and 2019. The policy counterfactual also suggests that a 10% subsidy from pharmacy sales associated with Medicare beneficiaries to independent pharmacies could improve limited pharmacy access in 16% of towns.

2.6 Tables

Table 2: Comparison of Estimator Specifications for Existing and Developed Orthogonal Estimators

		Existing Estimators	Orthogonal Estimators
First Stage	Data	Pre-selected market covariates	Many market covariates
	Estimators	Logit	ML Methods (XG Boosting)
Second Stage	Data	Pre-selected market covariates	Many market covariates + their interactions
	Estimators	Logit	Logit Lasso

Notes: Existing estimators are based on the approach described in Bajari et al. (2010b). I use pre-selected market covariates, as described in Appendix 21 and Appendix 22. For my developed orthogonal estimators, which employ a data-driven approach to variable selection, I utilize a pool of market characteristics described in Appendix 13 and Appendix 14. For the first stage ML methods, I use Ridge, Lasso, Elastic Net, XG Boosting, Random Forest, and Support Vector Machine. I compare their out-of-sample performance in section 5.4.

Table 3: Performance of Different Methods in First Stage CCP for Hold-Out Sample

Method	County FE & Year FE	Interaction Terms	AUC Score
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 (Area under the curve) 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. If the AUC is less than 0.5, it means the model is performing worse than random guessing. I use high elderly population towns samples for this analysis. Non-high elderly population towns produce very similar results.

Table 4: Results from the Structural Model: High Elderly Population Town

		(1)	(2)
Parameters	Variables	Existing Estimators	Orthogonal Estimators
θ_γ	Rival independent pharmacy	-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-selected market characteristics		Yes	No
Interaction between market characteristics		No	Yes
Dimension of Controls		13	563
Counties FE		Yes	Yes
Year FE		Yes	Yes

Notes: Samples include towns with a high elderly population in the years 2000-2019. In column (1), I use existing estimators based on the approach described in Bajari et al. (2010b). I use pre-selected market covariates, as described in Appendix 22. In column (2), I use my developed orthogonal estimators, which employ a data-driven approach to variable selection, I utilize a pool of market characteristics described in Appendix 13. I further use sample splitting and moment conditions based on equation (2.3.6) to remove biases from ML in the first stage of nuisance parameters estimation. Standard errors are clustered at the county level. Significance levels are denoted by + p<0.10, * p<0.05, ** p<0.01, *** p<0.001.

Table 5: Results from the Structural Model: Non-High Elderly Population Town

		(1)	(2)
Parameters	Variables	Existing Estimators	Orthogonal Estimators
θ_γ	Rival independent pharmacy	-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-selected market characteristics		Yes	No
Interaction between market characteristics		No	Yes
Dimension of Controls		13	563
Counties FE		Yes	Yes
Year FE		Yes	Yes

Notes: Samples include towns with a non-high elderly population in the years 2000-2019. In column (1), I use existing estimators based on the approach described in Bajari et al. (2010b). I use pre-selected market covariates, as described in Appendix 22. In column (2), I use my developed orthogonal estimators, which employ a data-driven approach to variable selection, I utilize a pool of market characteristics described in Appendix 13. I further use sample splitting and moment conditions based on equation (2.3.6) to remove biases from ML in the first stage of nuisance parameters estimation. Standard errors are clustered at the county level. Significance levels are denoted by + p<0.10, * p<0.05, ** p<0.01, *** p<0.001.

Table 6: Pharmacy Accessibility under Subsidy Counterfactual Scenario (Year: 2019)

	Limited Pharmacy Access in Towns (%)			
	Predicted	CF S2	$\triangle$	$\triangle$%
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: Sample includes towns with a high ratio of elderly populations in the year 2019. Based on the structural parameters from Table 4, I re-solve the equilibrium by using a nested fixed point algorithm in equations (2.4.1) and (2.4.2), by proving 10% subsidy to independent pharmacies.

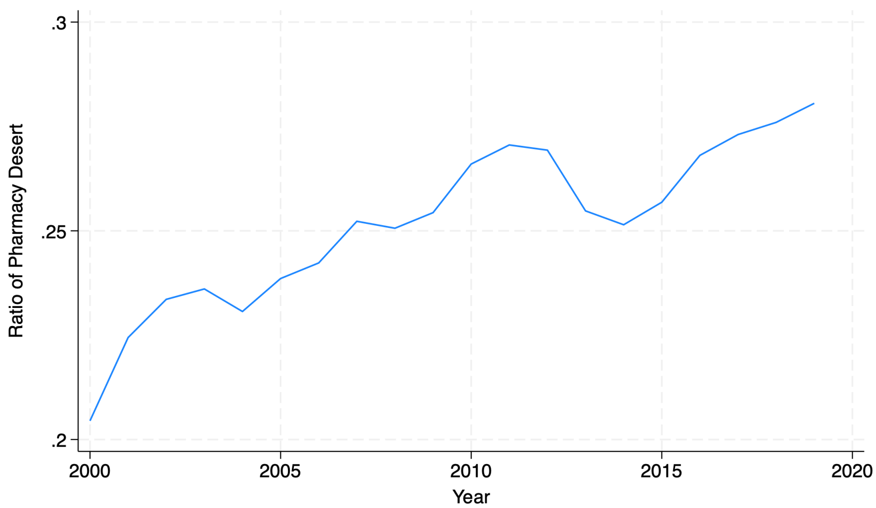
Table 7: Expected Number of Stores under Chain Pharmacy Counterfactual (Year: 2019)

	(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: Towns with high ratio of elderly populations are used. Based on the structural parameters from Table 4, I re-solve the equilibrium by using a nested fixed point algorithm in equations (2.4.1) and (2.4.2), fixing the number of chain pharmacies in 2000.

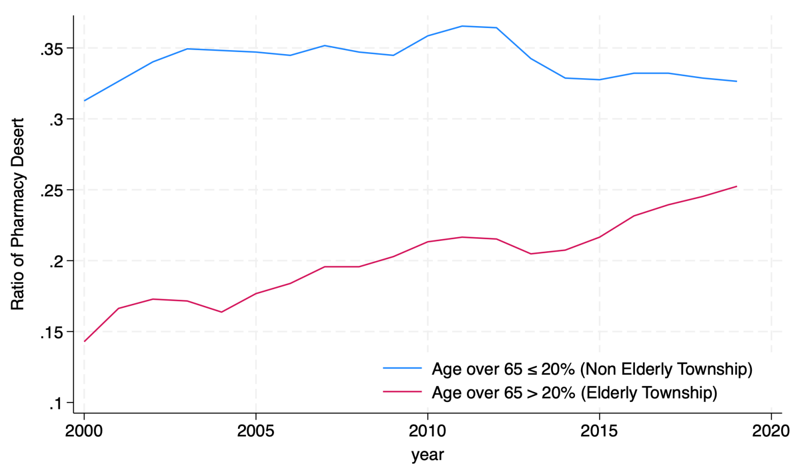
2.7 Figures

Figure 2: Trends in Limited Pharmacy Accessibility



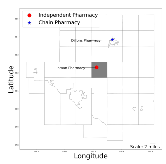
Notes: The data points represent a three-year moving average based on the limited pharmacy access, taken from a final sample of 802 townships. A township is designated as a limited pharmacy access (indicator value of 1) if there are no independent or chain pharmacies within its boundaries.

Figure 3: Trends in Pharmacy Access by High-Elderly Population/Non-High Elderly Population Towns

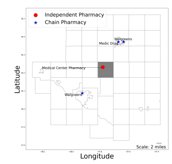


Notes: The data points are based on a three-year moving average, illustrating limited pharmacy access from a sample of 802 townships. A township takes an indicator value of 1 for limited pharmacy access if it has no independent or chain pharmacies within its town boundaries. Townships with over 20% of their population aged 65 or older in the year 2000 are classified as "elderly", while those with less than 20% in the year 2000 are defined as "non-elderly."

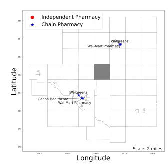
Figure 4: An Example: Spatial Distribution in Independent/Chain Pharmacy



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

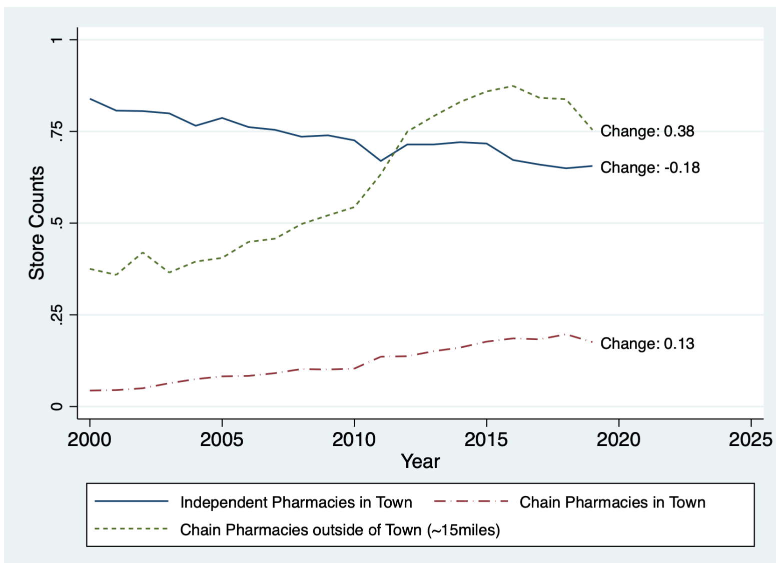
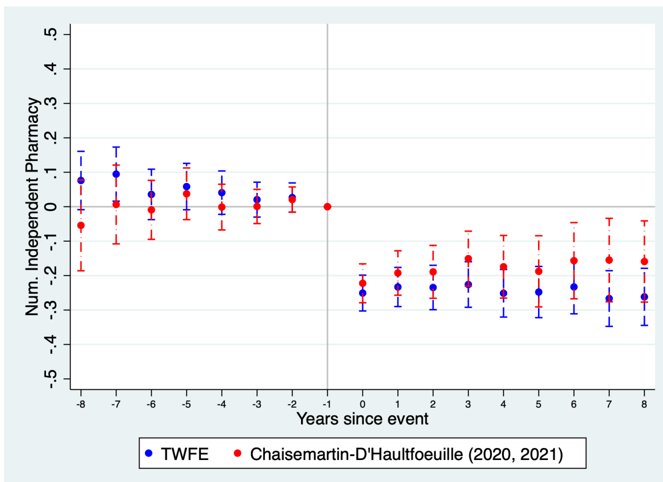
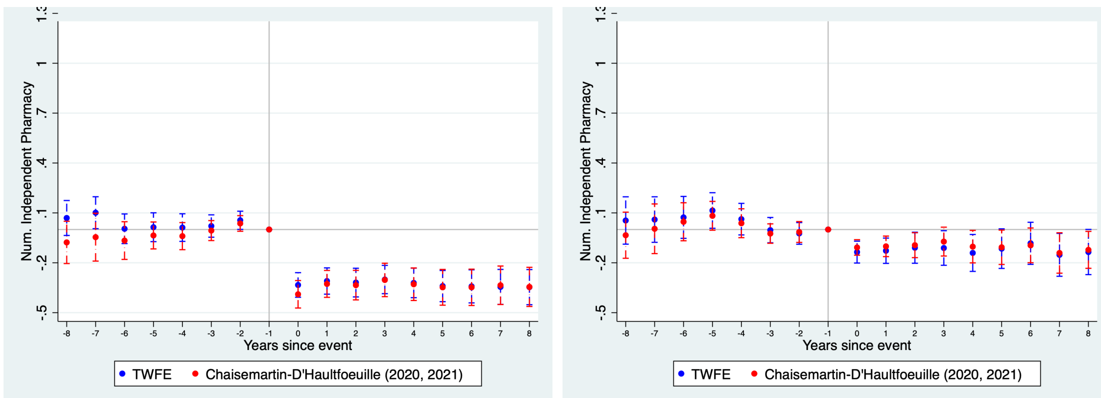


Figure 6: Event Study: The Effects of Chain Pharmacy Entry on Local Independent Pharmacies



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$\times$ 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 7: Heterogeneity by Ratio of Elderly Population Towns: the Effects of Chain Pharmacy Entry on Local Independent Pharmacies



(a) High Elderly Population Townships (b) Non-High Elderly Population Township

Note: This figure presents coefficient plots from event-study difference-in-differences analyses, which regress the number of independent pharmacies in a township on year-fixed effects, town fixed effects, control variables, and market× year-fixed effects. Figure 7a includes data from 291 "high elderly population towns", defined as townships with an over-65 population ratio higher than 20% in the year 2000. Figure 7b includes data from 511 "non-high elderly population towns", defined as townships with an over-65 population ratio lower than 20% in the year 2000. The baseline period, omitted in this analysis, is $t = -1$, representing the last pre-treatment period. Standard errors are clustered at the town level, and error bars represent 95% confidence intervals.

Figure 8: Algorithm Steps

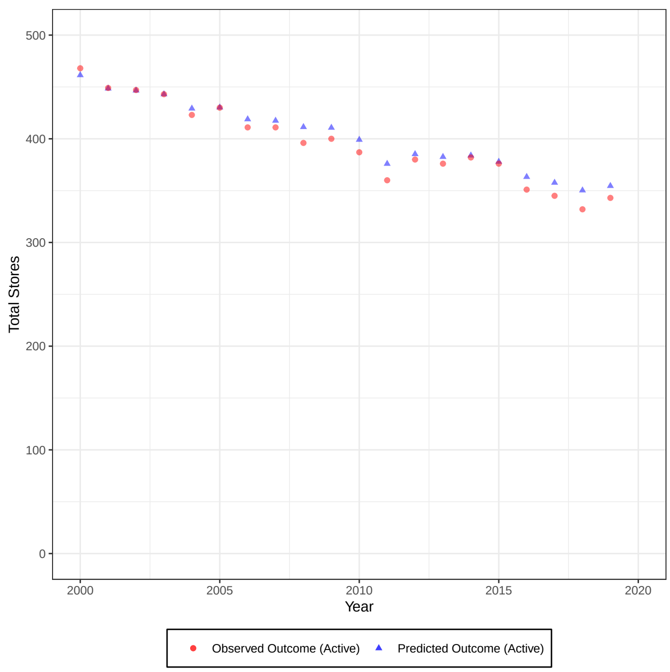
Take a K-fold random partition $(I_k)_{k=1}^{K}$ with the same size $n = N/K$. For each $k \in \{1, \ldots, K\}$, define I_k^c as the complement of I_k.

1. For each $k \in \{1, \ldots, K\}$, construct an ML estimator $\hat{\eta}_k$ using the subsample I_k^c.
   a. Obtain $\hat{\gamma}_{-imtk}$ using ML Classifiers of a_{-imt} on s_{-imt}, s_{imt} and s_{xmt}.
   b. Obtain $\hat{\beta}_k = (\hat{\beta}_{ek},\ \hat{\beta}_{xk})$ using Logit Lasso estimator of a_{imt} on $\hat{\gamma}_{-imtk}, s_{imt}$ and s_{xmt}.
   c. Compute $\hat{\theta}_k = (\hat{\theta}_{\gamma k}, \hat{\theta}_{ck})$ from the original moment function (2.3.5).
   d. Compute the conditional densities $\hat{f}_k$.
   e. Estimate $\hat{\mu}_k = (\hat{\mu}_{\gamma k}, \hat{\mu}_{ck})$ from the Lasso estimator of $\hat{f}_k z_{imt}$ on $\hat{f}_k x_{imt}$.
   f. Collect $\hat{\eta}_k = (\hat{\gamma}_{-imtk}, \hat{\beta}_k, \hat{\mu}_{zk})$.
2. Construct the estimator $(\hat{\theta}_\gamma, \hat{\theta}_c)$ as the solution to

$$\frac{1}{K}\sum_{k=1}^{K} L_{n,k}(\theta, \hat{\eta}_k) = 0$$

where $L_{n,k}(\theta) = \{E_{n,k}\left[\psi(\theta, \hat{\eta}_k)\right]\}^2$ and $E_{n,k}$ is the empirical expectation over I_k, that is, $E_{n,k}[\psi(w)] = n^{-1}\sum_{imt \in I_k} \psi(w_{imt})$. The moment function used in the objective function is defined in equation (2.3.6).

Figure 9: Goodness of Fit: High elderly population towns



Notes: Towns with a high ratio of elderly populations are used. Based on the structural parameters from Table 4, I re-solve the equilibrium by using a nested fixed point algorithm in equations (2.4.1) and (2.4.2). Then, I compare the observed number of independent pharmacies with the predicted number of independent pharmacies from the equilibrium choice probabilities, σ^*.

3.0 Horizontal Merger and Post-Entry Market Structure: Evidence from Acquisition in the Retail Pharmacy Market

*"**Guideline 4. Mergers Should Not Eliminate a Potential Entrant in a Concentrated Market**: Mergers can substantially lessen competition by eliminating a potential entrant. For instance, a merger can eliminate the possibility that entry or expansion by one or both firms would have resulted in new or increased competition in the market in the future. A merger can also eliminate current competitive pressure exerted on other market participants by the mere perception that one of the firms might enter."*

Source: 2023 Horizontal Guidelines by U.S. Department of Justice and the Federal Trade Commission

3.1 Introduction

Healthcare industries have undergone substantial consolidation, including hospitals, physicians, health insurers, and pharmacies through horizontal mergers and acquisitions. For example, there were more than 200 hospital mergers between 1998 and 2010 (Gaynor (2011)). In addition, vertical mergers between health insurance companies and pharmacies have accelerated the consolidation in the overall healthcare industry.

With the increasing prevalence of market consolidation in the U.S., the effects of horizontal mergers on competition and market dynamics remain heated debates for regulators and economists. While extensive research has been conducted on the immediate price effects and welfare implications of such mergers, the subsequent patterns of market entry behavior have been largely overlooked. This paper aims to address this gap by providing the first causal evidence of how horizontal mergers between dominant firms affect post-entry market behaviors, specifically analyzing the pharmacy sector's response to the Walgreens-Rite Aid merger in 2018.

Horizontal mergers, particularly among leading market players, raise both public and antitrust concerns. A key question is whether the merger lessens competition, ultimately harming consumer welfare. The Horizontal Merger Guidelines by the Department of Justice and the Federal Trade Commission emphasize the importance of assessing post-merger entry, which should ideally be timely, likely, and sufficient to deter any negative impacts on competition. Guided by regulators, this study examines the ex-post analysis of a high-profile merger between two top-ranking pharmacies - Walgreens and Rite Aid.

To understand the post-merger behaviors of pharmacies, I use a comprehensive dataset from 2010 to 2021, which includes the partial acquisition between Walgreens and Rite Aid in 2018. I examine the merger's effect on the total number of operational stores, a proxy for competition, and further investigate the possibility of market entry by non-merging competitors, a key determinant of a healthy competitive market. Using a difference-in-differences estimation framework, I find that merger approval decreases the number of operational stores by an average of 0.4 units post-merger, representing a 19% reduction in store count. This suggests that horizontal mergers might decrease competition, with no new market entries by non-merging competitors observed. I find no empirical evidence that the horizontal merger induces new market entries by non-merging competitors. These findings challenge the argument presented by merging firms that any reduction in competition resulting from a merger would be offset by new entries.

I also explore alternative methodologies for defining treatment. Specifically, in towns that have either a Walgreens or Rite Aid one year prior to the merger's approval, I designate those as 'treated towns' if one of these pharmacies closes its stores post-approval. This approach allows for a nuanced analysis of the merger's impact on local pharmacy availability. Employing a staggered Difference-in-Differences approach, as outlined by Callaway and Sant'Anna (2021), the main findings are consistent in that the merger decreases the number of pharmacies by 0.6 units or equivalently 27%. The reported magnitude is somewhat larger than observed in preferred specifications, as exit behaviors decrease the total number of pharmacies. Additionally, these results are robust, whether the comparison involves defining the untreated group as a never-treated group or a not-yet-treated group Further analysis of heterogeneity across income groups reveals that towns with lower incomes experienced a higher rate of pharmacy closures. This suggests new policy implications: mergers could be more carefully approved based on the income levels of the towns affected. The results are further supported by findings that only the highest-income towns experienced an increase in the number of non-merging pharmacies

This paper contributes to the growing body of literature on consolidation, particularly in healthcare, as exemplified by Gaynor et al. (2015), and examines the impact of mergers on price changes in the healthcare industry, following Tenn (2011), Haas-Wilson and Garmon (2011), and Thompson (2011). It also relates to empirical studies on the price effects and welfare changes following horizontal mergers in other retail industries, as seen in Weinberg (2008), Miller and Weinberg (2017), Miller et al. (2017), and Mansley et al. (2023). Collard-Wexler (2014) and Igami and Uetake (2020) use dynamic structural models to simulate entry behaviors before and after mergers. As the dynamic structural models of Collard-Wexler (2014) and Igami and

Uetake (2020) require numerous assumptions and substantial computational costs, this paper's approach to causal inference offers insights into the impact of mergers on market structure and competition, akin to the principle of Occam's Razor. To my knowledge, this is the first paper to offer causal estimates of the effects of horizontal mergers on post-entry behaviors. Considering that Walgreens and Rite Aid may offer differentiated products, such as a wider range of generic drugs and superior in-network health insurance options, the entry behaviors following these mergers could provide new insights into consumer welfare. For instance, approving a merger might limit the choices available to consumers in certain geographic areas, potentially leading to a decrease in consumer welfare.

This study also contributes to the theoretical literature on market entry and competition. The literature indicates that the presence and degree of sunk costs and entry barriers influence the timing of market entries following a horizontal merger (Werden and Froeb (1998); Cabral (2003); Marino and Zábojník (2006); and Davidson and Mukherjee (2007)). My empirical investigation sheds light on these theoretical predictions within the context of the pharmacy sector, which is hypothesized to have relatively low sunk costs associated with opening stores Kim (2023). The findings offer suggestive evidence regarding the duration and permanence of the impact of horizontal mergers on market structure. This paper indicates that even in industries with relatively low barriers to entry, the effects on the number of stores can be prolonged.

3.2 The Retail Pharmacy Industry

Approximately 60,000 pharmacies are operating in the United States. Of those, about 35-40% belong to one of the three leading pharmacy chains: Walgreens, CVS, and Rite Aid. Walgreens was founded in 1901 in Chicago, CVS Pharmacy in 1963, and Rite Aid in 1962. All three have expanded by opening new stores or acquiring smaller chains.

These national chains have rapidly increased their market shares through acquisitions. For example, in 2012, Walgreens acquired a Mid-South drugstore chain operating under the banners of USA Drug, Super D Drug, May's Drug, Med-X, and Drug Warehouse. In 2006, CVS acquired Eckerd Drug Stores, which had over 2,500 stores, and in 2015, CVS purchased Target's pharmacy and clinic businesses for approximately $1.9 billion, adding over 1,600 pharmacies to its network. Rite Aid began expanding rapidly in the 1980s and, in 1996, acquired Thrifty PayLess with over 1,300 stores. In 2007, Rite Aid acquired Genovese Drug Stores with over 1,200 stores[1].

The consolidation of small pharmacy chains in North Carolina exemplifies this trend. In 2014, CVS acquired Navarro Discount Pharmacy, which had 17 stores in the state, and Walgreens acquired Kerr's 76 retail drugstores. In 2015, CVS acquired Target's pharmacy and clinic businesses, which included 13 stores in North Carolina, and Walgreens bought Duane Reade with 10 stores in the state.

In the 2000s, there was continued expansion of both merchandise-based pharmacies (e.g., Walmart, Sam's Club, Target) and supermarket-based pharmacies (e.g., Kroger, Publix), particularly after 2005, which reduced the market share of indepen-

[1] Source: https://www.usatoday.com/story/money/2015/10/27/walgreens-rite-aid/74684642/ and https://fortune.com/2017/06/29/walgreens-rite-aid-merger-ftc/

dent 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 in the United States.[2]

3.2.1 Background of Merger between Walgreens and Rite Aid

In October 2015, Walgreens announced its plan to acquire Rite Aid for $6.8 billion, a move that would have created the largest pharmacy chain in the United States with over 14,000 stores. However, the merger faced immediate opposition from antitrust regulators concerned about the potential for reduced competition and increased consumer prices. Media coverage echoed these concerns, highlighting fears of a monopolistic retail pharmacy market. Public opinion further emphasized worries about diminished consumer choice, given that Walgreens and Rite Aid are among the nation's leading pharmacy chains.

The Federal Trade Commission (FTC) initiated investigations into the merger in November 2015. In April 2016, the FTC filed a lawsuit to block the merger, arguing that it would reduce competition in 1,900 local markets where Walgreens and Rite Aid were two of the few pharmacy chains operating. The FTC also argued that the merger would give Walgreens too much power over drug pricing.

The trial began in June 2016 and lasted for several weeks. The FTC presented evidence from economists and industry experts who testified that the merger would lead to higher prices and reduced choice for consumers. Walgreens presented evidence from its experts who testified that the merger would benefit consumers by leading to lower prices and more innovation.

In January 2017, Judges of the United States District Court for the District of

[2]Source: 2020 NATIONAL COMMUNITY PHARMACISTS ASSOCIATION DIGET

Columbia agreed with the FTC and blocked the merger. The judges found that the merger would significantly reduce competition in the retail pharmacy market. Walgreens and Rite Aid then negotiated a new agreement in June 2017, under which Walgreens would partially acquire 2,186 Rite Aid stores for $5.2 billion, instead of a full takeover. After further negotiation with the FTC, the Commission closed its investigation of a revised transaction under which Walgreens would acquire some Rite Aid stores, while Rite Aid would retain most of its network. Figure 10 provides a summary of the timeline of the partial acquisition. Despite the approval, consumer groups and public opinion continued to express concerns about potential anti-competitive effects.

3.3 Data

My data sources are fourfold: Data Axle, which tracks pharmacy openings and closings; public records from the FTC's investigation into the mergers of Walgreens and Rite Aid; demographic information from the 2010 to 2021 American Community Surveys at the census tract level; and the 2010 US Census Tract Shapefile.

Using Data Axle's information on business establishments, I identify the location and status (open or closed) of each pharmacy. After mapping each pharmacy's location using longitude and latitude, I group them into census tract markets. I then count the pharmacies, separating those affiliated with Walgreens or Rite Aid from those belonging to other competitors. I define a market based on the census tract, similar to the sizes found in towns, as noted by Schaumans and Verboven (2008). For this analysis, I only consider markets that had at least one pharmacy operating at any time from 2010 to 2021 and had a census tract with a population of at least one thousand.

To check the quality of the data, I look at the number of stores reported in North Carolina and compare it to the number in my final data set, as shown in the Appendix Table 30. The reported number of stores from my data is very close to the reported numbers in Rite Aid's financial reports.

3.3.1 Geographic Illustration

In this section, I illustrate how the horizontal merger led to the consolidation of the pharmacy industry in North Carolina. First, Figure 11 shows the locations of pharmacies before and after the merger. To compare the geographic distribution of stores, I captured snapshots of each store's location in 2017 and 2021. I observed the following: 1. Pharmacy stores are more prevalent in census tracts with higher population densities. 2. Following the merger approval, Walgreens and Rite Aid closed 187 stores, reducing the number from 510 in 2017 to 353 in 2021. This was implemented by Walgreens administrative levels to avoid cannibalization and save costs. As a result of the partial merger, a few Rite Aid stores remained under Rite Aid ownership.

To further investigate the pattern of exits, Figure 14 displays the markets where Walgreens and Rite Aid operated during my sample period from 2010 to 2021. In cases where Walgreens and Rite Aid locations were nearby, the merged entity (Walgreens Boots Alliance) tended to close Rite Aid stores. Although both chains generally prefer locations in areas of higher population density due to greater demand, the decisions to close stores appear to be geographically dispersed across the state. This distribution mitigates concerns regarding endogenous treatment. Section 3.4.2 provides additional discussion on this alternative definition of treatment.

Appendix Figure 26 illustrates the changes in the market structure before and after the merger approval in 2018. I observe that there is no significant change in the percentage of monopolies, duopolies, or oligopolies in markets not treated by the merger. As Walgreens and/or Rite Aid reduced their store count by more than 30%, the overall ratio of monopolies to oligopolies decreased, while the number of unserved areas increased. This suggests that Walgreens may have been closing unprofitable stores following a detailed evaluation of each store's performance.

3.3.2 Descriptive Evidence

I examine the impact of horizontal mergers on market structure changes using descriptive statistics. First, I assess the effectiveness of the horizontal merger between Walgreens and Rite Aid on the number of pharmacies, as shown in Figure 12a. There was a consistent upward trend due to increased entries from supermarket-based pharmacies and big-box stores like Walmart. The store count for Walgreens and Rite Aid remained relatively stable but experienced a notable decline following the merger approval, which aligns with empirical evidence from Figure 11. Figure 12b further outlines the average number of stores within treated and untreated groups. I classify a census tract as 'treated' if one of the merging entities (either Walgreens or Rite Aid) was active there in the year 2017-the year before the merger's approval.[3] Overall, Figure 12 indicates that the horizontal merger led to the closure of stores by the merged entity in certain locations.

To further illustrate the changes in post-entry behaviors, Figure 13 examines the behavior of non-merged competitors. Figure 13a presents the number of active stores operated by non-merged firms, and I find no evidence to suggest that the

[3] I use the alternative definition of 'treated' group in Section 3.4.2 refers to instances where the merging entity closed one of its stores.

horizontal merger has facilitated new entries by these firms. Figure 13b shows the number of new entries by non-merged firms, indicating a change in new entries in the year 2020. To examine whether these trends are pronounced after controlling for observable characteristics, I will introduce dynamic event studies in Section 4.

3.4 Empirical Strategy

In this section, I estimate the effects of horizontal mergers on market structures (number of pharmacies, number of non-merging identities' pharmacies), and whether mergers induce new entries by non-merger identities. I also show that the findings are robust to other potential concerns.

Table 8 presents the mean and standard deviations of market characteristics of total markets, markets with the presence of Walgreens/Rite Aid in the year 2017, and markets with the absence of Walgreens/Rite Aid in the year 2017. There are differences in the average levels of some of the variables(e.g., total population, population density) across towns. Furthermore, given that closing decisions of stores by Walgreens and Rite Aid are endogenous, I will closely illustrate the possibility of pre-trends well as balance tests, in subsequent sections.

3.4.1 Estimating Dynamic Treatment Effects

To estimate changes in treatment effects over time, including potential pre-trends, I conduct an event study to illustrate the effects of horizontal mergers in the years leading up to and following the merger's approval in 2017, which took effect in 2018.

In this regression, I estimate the following equation:

$$Y_{mt} = \sum_{\tau} \delta_\tau \text{Event}_{m,t-\tau} + \beta X_{mt} + \lambda_m + \alpha_t + \gamma_{ct} + \varepsilon_{mt} \tag{3.4.1}$$

Here, $Event_{m,t}$ is a dummy variable indicating whether the merged entity was active in 2018 at location m in period t. The dependent variable Y_{mt} represents the total number of retail pharmacies, including new entries by non-merged entities, in census tract m during year t. I control for census tract-level demographics X_{mt}, unobserved census tract-level fixed effects λ_m, and annual time fixed effects α_t. To account for time-varying unobserved heterogeneity, I include county-year fixed effects γ_{ct}, where c denotes the county and t the year. The analysis takes a binary specification, such that $Event_{mt}$ takes the value of 1 if the merged firms were active in 2018. [4]

Figure 15 shows that the standard TWFE (two-way fixed effect) indicates an absence of statistically significant effects in terms of pre-trends. However, the negative coefficients in the post-merger years suggest that the merger led to changes in the market with Walgreens and Rite Aid closing stores, as stated anecdotally. The event study supports the empirical findings that the horizontal merger had a negative impact on the number of pharmacies, which is consistent with the narrative that Walgreens and Rite Aid closed many stores following the merger.

To evaluate how these industry changes have led to new entries by competitors, I defined the dependent variables as 'the number of competitors' stores' and 'the number of new competitors' entries, respectively. For both event studies in Figure 16, I found no pre-trends or post-merger effects, suggesting that the closure of Walgreens and Rite Aid stores did not lead to new entrants.

[4] As I assumed a one-shot event for this analysis, it does not require a staggered two-way fixed effects design.

3.4.2 Difference-in-Differences Estimators

To study the connection between horizontal mergers between dominant firms and post-market competition, I exploit variations in the timing and locations of pharmacy store closures in North Carolina between 2010 and 2021. The market (census tract) is considered to be treated if the number of stores by merged firms decreases after the merger was approved in 2018.

As I did not find pretends between treated and control groups from the event studies, I estimate the effect of horizontal mergers on the post-market structure using the standard two-way fixed effects:

$$y_{mt} = +\beta D(t > T_m) + \beta X_{mt} + \alpha_m + \gamma_t + \theta_{ct} + \epsilon_{mt}, \tag{3.4.2}$$

where y_{mt} is the outcome of interest; the total number of stores, the total number of non-merged pharmacies, entries by non-merged pharmacies, T_m denotes the horizontal merger being effective census tract with Walgreens and Rite Aid stores in the previous year, $D(\cdot)$ denotes the indicator function, which takes one if T_m takes the value one and otherwise zero. I also use two-way fixed effects by controlling time-invariant market fixed effects α_m, and time-fixed effects γ_t.

I begin by estimating the difference-in-differences (DD) framework outlined in Equation 3.4.2. In my baseline specification, Column (1) of Table 9 indicates a 0.4 unit decrease in the total number of pharmacies, suggesting that the horizontal merger is associated with a significant decrease in the number of pharmacies in the treated census tracts. Next, I examine the reactions of competitors by analyzing the total number of non-merging pharmacies and their entries, presented in Columns (2) and (3) of Table 9, respectively. I interpret this as follows: 1) The pharmacy market may already be saturated, indicating that there is insufficient demand for new pharmacies to enter the market even after the closures. 2) Alternatively, there could be high barriers to entry, such as established brand loyalties and significant product differentiation, which make it challenging for new competitors to penetrate the market.

3.4.3 Alternative Approach: Staggered Difference-in-Differences Estimators

As competitors are unlikely to have incentives to enter the market unless there are store closures by the merged entity- either Walgreens or Rite Aid, I employ an alternative definition of treatment. I designate a census tract as treated if the number of stores from either Rite Aid or Walgreens decreases in that census tract following the merger approval. Conversely, census tracts without such a decrease are defined as untreated. The primary distinction from the previous definition of treatment is that non-merged firms now have a clear incentive to enter the market when one of the merged identities closes the store, providing incentives for potential entrants.

As the closure of pharmacy stores by the merged entity has been heterogeneous across the years and market, I adopt 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.

Given the possibility that markets untreated by horizontal mergers might fundamentally differ due to the presence of Rite Aid or Walgreens stores before the merger approval, I experiment with constructing sub-samples. These sub-samples are based on census tracts where a Rite Aid or Walgreens store was present before the horizontal merger approval. Furthermore, I conduct experiments with different untreated groups. This involves distinguishing between groups that were never treated and those not yet treated, to account for staggered treatment.

Admittedly, the decision to entry or exit decision by pharmacies could be endogenous. In particular, it is possible that a firm's plans to enter or exit a market are determined, in part, by expectations about future market cost and demand factors that also affect post-entry or post-exit prices.

Figure 17 shows that Callaway and Sant'Anna (2021)' estimator indicates an absence of statistically significant effects in terms of pre-trends and robust to different untreated groups and never-treated/not-yet-treated groups. Post-treatment effects show that the entry of chain pharmacies is associated with a decreased number of independent pharmacies in the census tract. Compared to the preferred event study framework shown in Figure 6, the magnitudes are somewhat larger. This increase is attributed to the construction of treatment effects, which involve the exit of either Walgreens or Rite Aid.

To further examine post-entry behaviors by non-merged competitors, I implement a similar staggered TWFE approach to detect pre-trends in the number of non-merged firms and the number of entries by non-merged firms, as illustrated in Appendix 27 and 29. Similarly, I find no evidence of pre-trends similar to the event study as in Figure 13b.

3.4.4 Estimation Results from Staggered Difference-in-Differences Estimators

To investigate whether the horizontal merger leads to new entries from non-merged firms, I employ a staggered DID specification. Here, 'treated' is defined as those instances where either Walgreens or Rite Aid closed stores after the merger took effect. Table 10 demonstrates that this alternative definition of treatment yields effects qualitatively similar to those previous findings. However, the impact of the horizontal merger on the total number of pharmacies is statistically different and larger from that in Table 9, likely because the treatment now focuses on the exit behaviors of Walgreens or Rite Aid, resulting in more pronounced effects. Column (2) considers 'not-yet-treated' as the untreated group and shows results that are quite similar. To address concerns that untreated groups could differ significantly depending on whether Walgreens or Rite Aid were present before the merger approval, I use sub-samples with the presence of either Walgreens or Rite Aid prior year to the merger approval. The findings in Columns (3) and (4) confirm that the results remain robust.

Next, I document the reaction of the rivals' pharmacies and whether the exit of the incumbent merged identity changes the market structure. As a proxy, I look at two outcomes: the total number of rival pharmacies to capture both entry and exit behaviors of rival pharmacies, and new entries by non-merged pharmacies. Appendix Table 11 and Appendix Table 12 show that results are again consistent with the previous findings. For the full sample, regardless of the choices of constructing an untreated group as never treated or not yet treated, the exit of merged firms does not change the market environments. I caution that if non-merging firms perfectly know when Walmart or Rite Aid closed associated with mergers, then the timing of treatment could be an issue. In Appendix Table 29, after adding market-level fixed effects in sub-graphs (c), I fail to detect significant pre-trends and provide empirical evidence for the exogeneity of closures by Walgreens/Rite Aids, using a balance test.

3.4.5 Identification

The identification assumption in the differences-in-differences analysis is that the entry/exit decisions by non-merging pharmacies exposed to treatment would have followed the same trend as the store choices not exposed to the merger between Walgreens and Rite Aid. In Figure 12, I show that the number of non-merging pharmacy stores that were and were not competing with Walgreens/Rite Aid did not diverge from one another before the horizontal mergers. Additionally, I find that treated and untreated markets followed the same trend before the horizontal mergers, suggesting that untreated markets are a good control group for treated markets, which provides empirical justification for the identification assumption (see Figure 16 and Figure 17). While the sample selection is always a concern with differences-in-differences research designs, this is less concerning for my setting since the approval of the horizontal merger was unexpected by both the public media and rival firms. Furthermore, anecdotal evidence reveals that the timing of the closure of Walgreens or Rite Aid was nearly random and also unknown to rival firms, implying that the store decisions of rival pharmacies were not affected by the expectation of being "treated".

3.4.6 Heterogeneity Analysis

Which market characteristics driving these changes in market structure? As Walgreens or Rite Aid might want to keep their stores in urban areas, or higher-income markets, competitors might respond heterogeneously to these changes. To investigate this, I quantify the closures of Walgreens or Rite Aid associated with Mergers by different income quantiles. Following the Staggered TWFE framework, Table 31 represents that lower per captia groups might experience a more intense decrease in the total number of pharmacies. In Table 32, the effects of horizontal mergers on the number of non-merged firms are only statistically significant in higher per capita income groups, which implies that merger approval by low-income groups might require more careful investigation by regulatory authorities.[5]

3.4.7 Robustness Check

In this section, I address additional potential concerns and confirm that previous empirical findings remain valid under further robustness checks.

Alternative Treatment: Expand into Adjacent Markets: Since Walgreens and Rite Aid's store closures may affect adjacent markets, I redefine the treatment area to encompass adjacent census tracts. This modification ensures that the analysis considers the potential spillover effects of the mergers into neighboring areas. In Appendix Tables 35, 36, and 34, I demonstrate that the impact of horizontal mergers on the total number of pharmacies is somewhat moderated due to the dynamics of spatial competition. Furthermore, I corroborate the previous findings with untreated groups with never-treated or not-yet-treated, and subgroups analysis.

Alternative TWFE: I also use an alternative event-study estimator (de Chaisemartin and D'Haultfoeuille, 2023), which is robust to heterogeneous treatment timing, and the outcome may be affected by treatment lags. Estimates from this approach can be found in Appendix Table 40. The post-treatment estimates are highly

[5]I could not find different entries effects by income groups in Table 33.

similar to those from Callaway and Sant'Anna (2021)'s method, corroborating the robustness of staggered TWFE estimators in my settings.

Alternative demographic variables: Building on the work of Igami and Yang (2016), population density could serve as an alternative proxy for market demands-prescription. I demonstrate the robustness of my findings to alternative demographic controls, including population dentisites, in Appendix Tables 39, 37, and 38.

3.5 Conclusion

This study offers insights into the dynamics of competition and market entry following horizontal mergers, with a focus on the significant merger between Walgreens and Rite Aid in 2018. The key findings suggest that such mergers, especially among dominant firms, do not always encourage market entry by new entrants or non-merging competitors, contradicting claims often made by the merging entities. Employing the preferred staggered TWFE method by Callaway and Sant'Anna (2021), I observed a 24% decrease in the total number of stores post-merger, indicating a potential reduction in competition.

Heterogeneity analysis reveals that the effects of the merger on pharmacy numbers are more pronounced in lower-income towns compared to higher-income towns. This implies that regulatory agencies could consider the income levels of towns when approving mergers. The findings are consistent with observations that only the highest-income towns saw an increase in the number of non-merging pharmacies. Given that the Federal Trade Commission approved a partial acquisition between Walgreens and Rite Aid, more detailed scrutiny of spatial competition across different market types could protect consumer surplus and ensure the availability of more pharmacies in nearby areas.

3.6 Tables

Table 8: (Average) Summary Statistics by Market Type

Market Characteristics	Market Type		
	Total	Without Walgreens/Rite Aid (2017)	With Walgreens/Rite Aid (2017)
Total Population	4263.14	4181.94	4086.22
	(1478.27)	(1495.36)	(1403.93)
Population Density	1500.49	1445.51	1725.41
	(1399.55)	(1436.22)	(1212.91)
Mean Per Capita Income	28194.38	27773.2	29917.38
	(13880.97)	(13503.68)	(15210.28)
Share of Population $\geq$ Age over 65	0.153	0.151	0.159
	(0.069)	(0.071)	(0.061)
Share of Population Minority	0.330	0.329	0.332
	(0.235)	(0.237)	(0.224)
Share of Vacancy Housing Units	0.131	0.134	0.116
	(0.101)	(0.106)	(0.073)
Total Observations	11,424	9,180	2,244

Notes: The unit of observation for demographic variable is the census-tract-year. Standard deviations are in parentheses.

Table 9: Horizontal Merger and Pharmacy Stores

Outcome:	Pharmacies	Non-merging Pharmacies	Non-merging Pharmacy Entries
	(1)	(2)	(3)
Horizontal Merger	-0.394***	0.0410	0.0170
	(0.0292)	(0.0282)	(0.0192)
Census Tract FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Controls	Yes	Yes	Yes
Observations	11413	11413	11413
Outcome mean	1.520	1.198	0.143
Adjusted R^2	0.802	0.756	0.0448

Notes: Estimates are from difference-in-differences regressions of the number of stores in census tract c in year t on an indicator called "horizontal merger," which equals one for a census tract in the years following the activity of either Walgreens or Rite Aid and zero otherwise. Column (1) includes the total number of pharmacies. Column (2) includes the total number of pharmacies from non-merging firms. Column (3) includes the total number of new entries by non-merging pharmacies. Standard errors are clustered at the census tract level. Significance levels are denoted by + p<0.10, * p<0.05, ** p<0.01, *** p<0.001.

Table 10: Horizontal Merger and Number of Pharmacy Stores

Outcome:	Total Pharmacies			
	(1)	(2)	(3)	(4)
Closure of Merged Pharmacy	-0.682***	-0.680***	-0.836***	-0.800***
	(0.071)	(0.071)	(0.136)	(0.121)
Census Tract FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	11,408	11,413	2,244	2,244
Outcome mean	1.520	1.520	2.37	2.37
Sample	Full sample	Full sample	Sub-sample	Sub-sample
Untreated Group	Never-treated	Not-yet-treated	Never-treated	Not-yet-treated

Notes: Estimates are from staggered TWFE Callaway and Sant'Anna (2021) regressions of the number of stores in census tract c in year t on an indicator called "closure of merged pharmacy," which equals one for a census tract in the years following the closure of either Walgreens or Rite Aid after the merger approval and zero otherwise. The outcome of interest is the total number of pharmacies. Column (1) includes a full sample with never treated as an untreated group. Column (2) includes a full sample with not-yet-treated as an untreated group. Column (3) includes census tract with the presence of either Walgreens or Rite Aid prior year to merger approval with never treated as an untreated group. Column (4) census tract with the presence of either Walgreens or Rite Aid prior year to merger approval with not-yet-treated as an untreated group. Standard errors are clustered at the census tract level. Significance levels are denoted by + p<0.10, * p<0.05, ** p<0.01, *** p<0.001.

Table 11: Horizontal Merger and Non-merged Pharmacy Stores

Outcome:	Total Number of Non-merging Pharmacies			
	(1)	(2)	(3)	(4)
Closure of Merged Pharmacy	0.076	0.078	-0.066	-0.045
	(0.057)	(0.057)	(0.153)	(0.133)
Census Tract FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	11,408	11,413	2,244	2,244
Outcome mean	1.198	1.198	1.329	1.329
Sample	Full sample	Full sample	Sub-sample	Sub-sample
Untreated Group	Never-treated	Not-yet-treated	Never-treated	Not-yet-treated

Notes: Estimates are from staggered TWFE Callaway and Sant'Anna (2021) regressions of the number of stores in census tract c in year t on an indicator called "closure of merged pharmacy," which equals one for a census tract in the years following the closure of either Walgreens or Rite Aid after the merger approval and zero otherwise. The outcome of interest is the total number of pharmacies. Column (1) includes a full sample with never treated as an untreated group. Column (2) includes a full sample with not-yet-treated as an untreated group. Column (3) includes census tract with the presence of either Walgreens or Rite Aid prior year to merger approval with never treated as an untreated group. Column (4) census tract with the presence of either Walgreens or Rite Aid prior year to merger approval with not-yet-treated as an untreated group. Standard errors are clustered at the census tract level. Significance levels are denoted by + $p<0.10$, * $p<0.05$, ** $p<0.01$, *** $p<0.001$.

Table 12: Horizontal Merger and Entries of Non-merged Pharmacy Stores

Outcome:	Total Number of Entries of Non-merging Pharmacies			
	(1)	(2)	(3)	(4)
Closure of Merged Pharmacy	0.021	0.022	-0.066	-0.030
	(0.039)	(0.0039)	(0.058)	(0.063)
Census Tract FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	11,408	11,413	2,244	2,244
Outcome mean	0.143	0.143	0.171	0.171
Sample	Full sample	Full sample	Sub-sample	Sub-sample
Untreated Group	Never-treated	Not-yet-treated	Never-treated	Not-yet-treated

Notes: Estimates are from staggered TWFE Callaway and Sant'Anna (2021) regressions of the number of stores in census tract c in year t on an indicator called "closure of merged pharmacy," which equals one for a census tract in the years following the closure of either Walgreens or Rite Aid after the merger approval and zero otherwise. The outcome of interest is the total number of pharmacies. Column (1) includes a full sample with never treated as an untreated group. Column (2) includes a full sample with not-yet-treated as an untreated group. Column (3) includes census tract with the presence of either Walgreens or Rite Aid prior year to merger approval with never treated as an untreated group. Column (4) census tract with the presence of either Walgreens or Rite Aid prior year to merger approval with not-yet-treated as an untreated group. Standard errors are clustered at the census tract level. Significance levels are denoted by + p<0.10, * p<0.05, ** p<0.01, *** p<0.001.

3.7 Figures

Figure 10: Walgreens and Rite Aid Merger Events Timeline

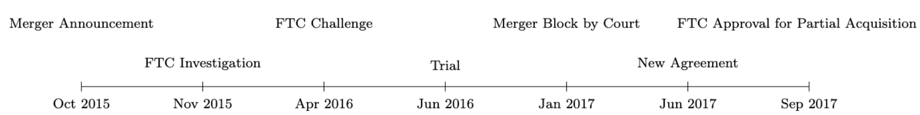
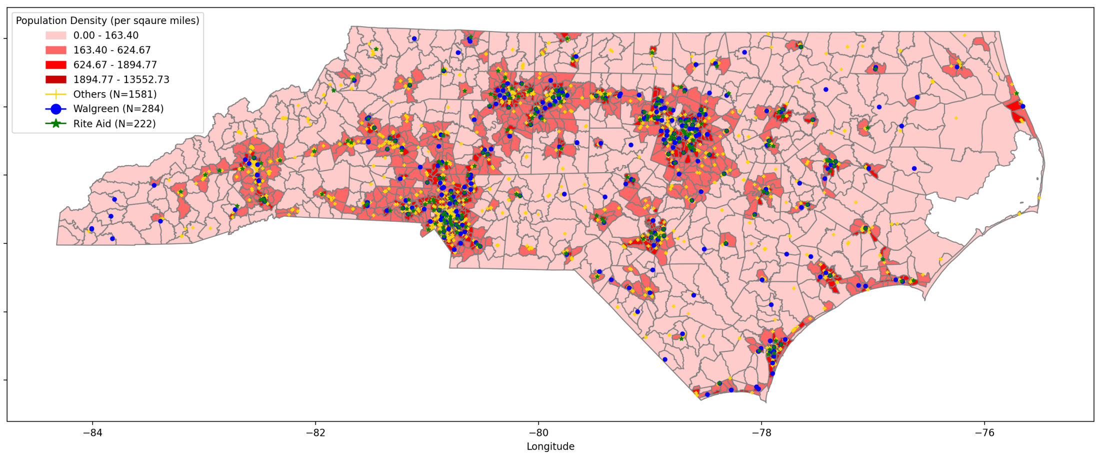
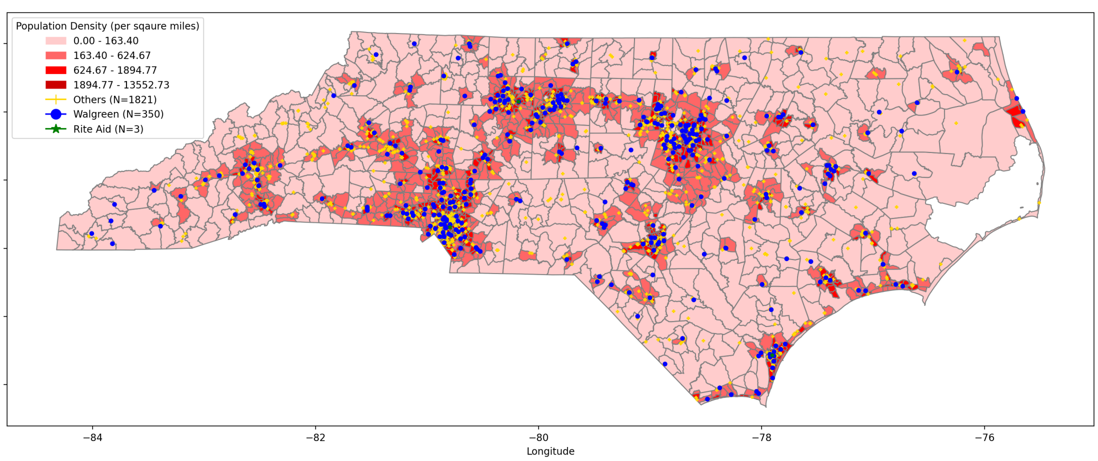


Figure 11: Geographic Distribution of Stores Location Before/After the Horizontal Merger



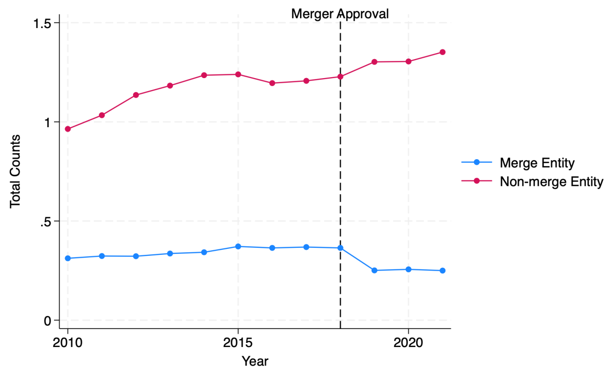
(a) Year: 2017 (Before the merger approval)



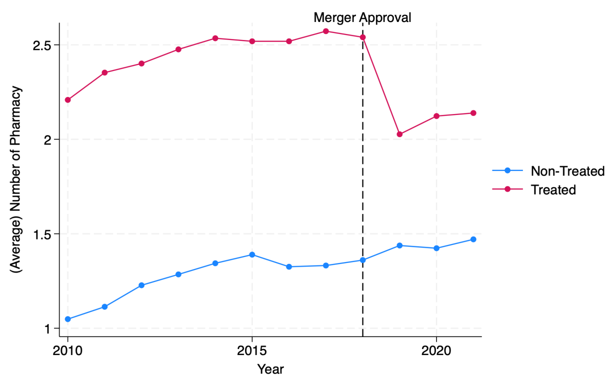
(b) Year: 2021 (After the merger approval)

Notes: The figure shows the geographic distribution of Walgreens, Rite Aid, and other pharmacies in North Carolina before (2018) and after the horizontal merger between the two companies. The map is color-coded by population density with four quartiles, with darker colors indicating higher population density.

Figure 12: Number of Stores Before/After the Horizontal Merger



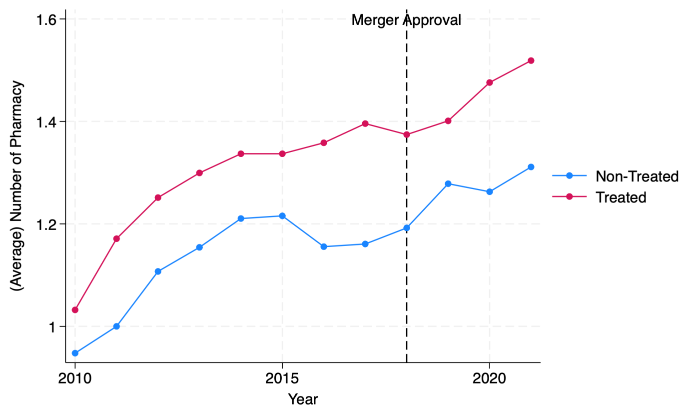
(a) Average Number of Stores by Firm: Merged Firms and Non-merged Firms



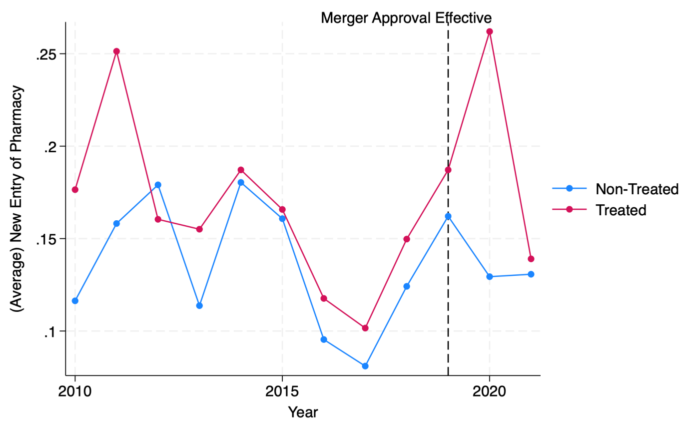
(b) (Average) Number of Stores by Treated/Untreated Towns

Notes: Figure 12a represents the average number of stores in 952 census tracts in North Carolina for Merged Entity (Walgreens and Rite AId) and Non-merge Entity, which is all other competitors. Figure 12b represents the average number of stores in the 952 census tract, conditional on the treated census tract or untreated census tract. The treated census tract is defined as if either Walgreens or Rite Aid was active in 2017. The untreated census tract is defined if neither Walgreens or Rite Aid were active in 2017.

Figure 13: Number of Stores Before/After the Horizontal Merger



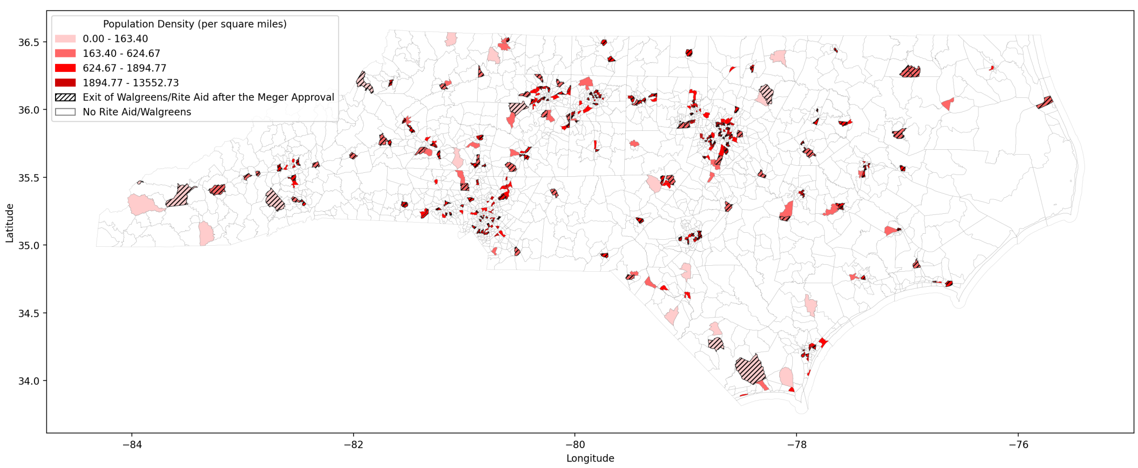
(a) Average Number of Non-Merged Stores by Treatment



(b) (Average) Number of New Entries of Stores by Treated/Untreated Towns

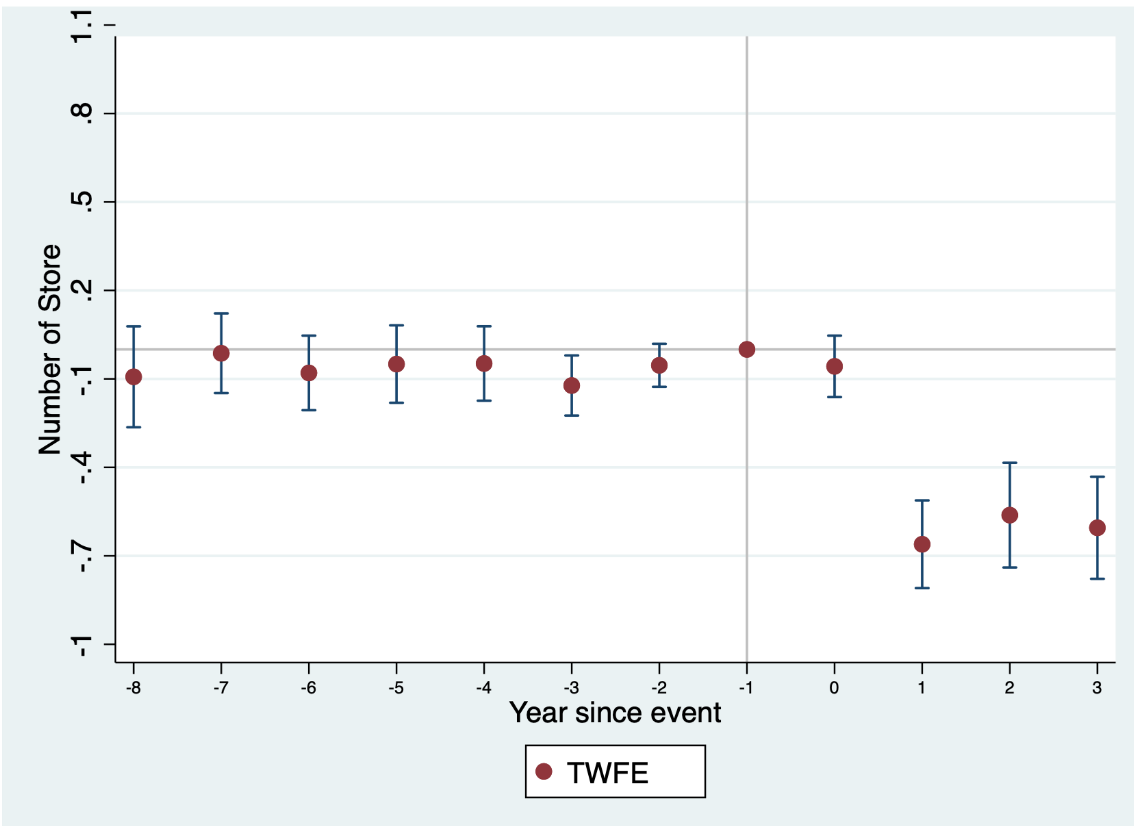
Notes: Figure 13a represents the average number of non-merged stores in 952 census tracts in North Carolina for Merged Entity (Walgreens and Rite Aid) and Non-merge Entity, which is all other competitors. Figure 13b represents the average number of new entries by non-merged firms in the 952 census tract, conditional on the treated census tract or untreated census tract. The treated census tract is defined as if either Walgreens or Rite Aid were active in 2017.

Figure 14: Geographic Distribution of Exit of Stores Before/After the Horizontal Merger



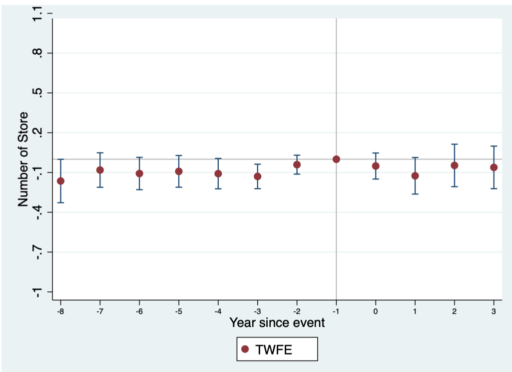
Notes: The figure shows the geographic distribution of closing stores of Walgreens and Rite Aid in North Carolina after the horizontal merger between the two companies. The map is color-coded by population density with four quartiles, with darker colors indicating higher population density.

Figure 15: Event Study: The effects of the horizontal merger on the total number of pharmacies

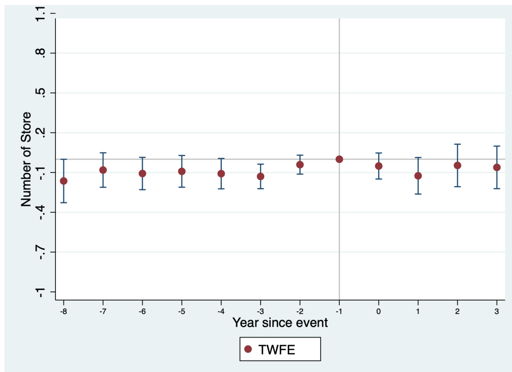


Notes: Coefficient plots from event-study difference-in-differences analyses that regress the number of independent pharmacies in a census tract on year fixed effects, census tract fixed effects, control variables, and market× year fixed effects. The sample consists of census tracts between 2010 and 2021. The omitted baseline period is $t = -1$, which is the last pre-treatment period. Standard errors are clustered at the census-tract level and error bars represent 95 confidence intervals.

Figure 16: Event Study: The effects of the horizontal merger on the new entries by non-merged firms



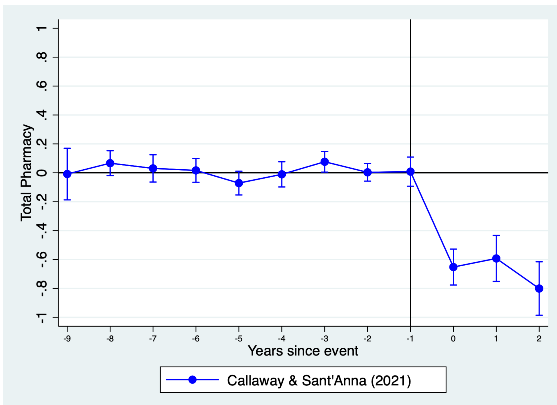
(a) (Average) Number of Non-Merged Stores by Treatment/Untreated Groups

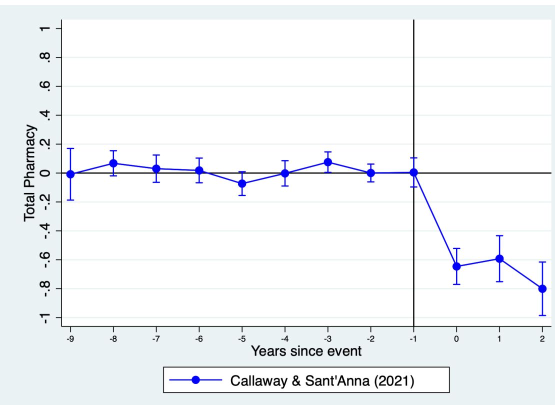


(b) (Average) Number of New Entires of Stores by Treated/Untreated Towns

Notes: Coefficient plots from event-study difference-in-differences analyses that regress the number of new entrants by non-merged firms in a census tract on year fixed effects, census tract fixed effects, control variables, and market× year fixed effects. The sample consists of census tracts between 2010 and 2021. The omitted baseline period is $t = -1$, which is the last pre-treatment period. Standard errors are clustered at the census-tract level and error bars represent 95 confidence intervals.

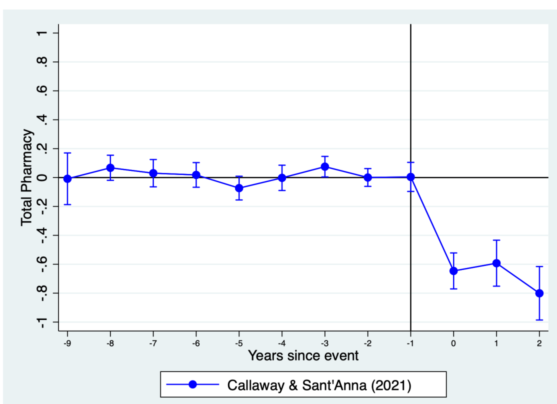
Figure 17: Horizontal Mergers and Pharmacy: Event Study

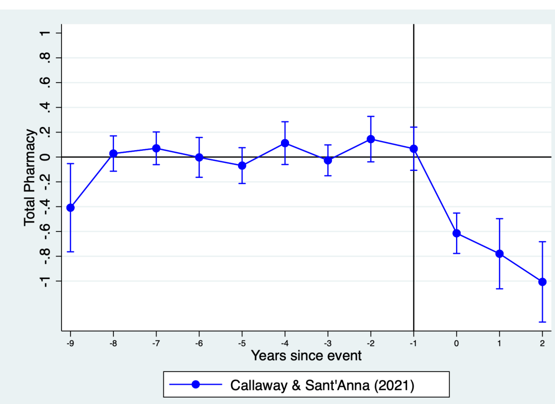




(a) Full sample
Untreated: Never treated

(b) Full sample
Untreated: Not-yet-treated





(c) Sub-sample
Untreated: Never treated

(d) Sub-sample
Untreated: Not-yet-treated

Notes: Coefficient plots from event-study difference-in-differences analyses that regress the number of new entrants by non-merged firms in a census tract on year fixed effects, census tract fixed effects, control variables, and market× year fixed effects. The full sample consists of census tracts between 2010 and 2021. The sub-sample includes census tracks where a Rite Aid or Walgreens store was present before the horizontal merger approval. The omitted baseline period is $t = -1$, which is the last pre-treatment period. Standard errors are clustered at the census-tract level and error bars represent 95 confidence intervals.

Appendix A Review of Discrete Static Games

A.1 Discrete Games Details

Under strategic interaction settings[1], 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.

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

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

$$\sigma_{-i}^*(a_{-i} = 1|s) =$$

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

[1] 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.

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 mapped out the sequence of events in the game, which is illustrated in Figure 18. 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 $\sigma_{-i}(a_{-i}|s)$. Having gathered all this information, players then finalize their decisions, denoted a_i, and the game progresses to the succeeding period, $t+1$.

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:

$$\begin{aligned} \sigma_i(a_i = 1|s) = \Psi_i(\Pi_i(a_i, s)) &:= \frac{\exp(\pi_i(a_i = 1, s))}{1 + \exp(\pi_i(a_i = 1, s))} \\ \sigma_{-i}(a_{-i} = 1|s) = \Psi_{-i}(\Pi_{-i}(a_{-i}, s)) &:= \frac{\exp(\pi_{-i}(a_{-i} = 1, s))}{1 + \exp(\pi_i(a_{-i} = 1, s))} \end{aligned} \tag{A.1.1}$$

where equilibrium functions Ψ map the choice-specific value function into choice probabilities. By the rational expectation assumption, 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).

A.2 Discussion on Multiple Equilibria and Estimation Approach

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 1.2.4, delineating the relationship between equilibrium choices, σ_i, and equilibrium beliefs for each player $i = 1, \dots, n$. 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[2].

The second approach, pioneered by Hotz and Miller (1993), Bajari et al. (2010b), and dynamic games (Berry and Reiss (2007)) employs a two-step method. This method is computationally light and uses weaker assumptions about multiple equilibria compared to the NXFP algorithm. [3] In the first stage, I non-parametrically es-

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

[3]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.

timate conditional expectation $\gamma_{-i} = E[a_{-i} = 1|s]$ from observed choices and market characteristics d.[4] In the second stage, the econometrician estimates a single-agent random utility model. This model incorporates both market characteristics, s, and the equilibrium beliefs, $\gamma{-}i$, 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 A.2.1 (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

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

a set of structural parameters without the need for repeated model solving, leading to a substantial reduction in computation time.

A.3 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 2 (Identification)**.** *Deterministic payoff components $\pi(a_i, a_{-i}, s)$ are identified if different deterministic payoff components $\sigma_i(a_i = 1|s) \neq \tilde{\sigma}_i(a_i = 1|s)$ yield alternative equilibrium probabilities $\pi(a_i, a_{-i}, s) \neq \tilde{\pi}(a_i, a_{-i}, s)$.*

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:

$$
\begin{aligned}
&\underbrace{\Pi_1(a_1 = 1|s)}_{\text{unknown}} = \sigma_2(a_2 = 1|s)\,\underbrace{\pi_1(a_1 = 1, a_2 = 1|s)}_{\text{unknown}} \\
&+ (1 - \sigma_2(a_2 = 1|s))\,\underbrace{\pi_1(a_1 = 1, a_2 = 0|s)}_{\text{unknown}} \\
&\underbrace{\Pi_1(a_1 = 0|s)}_{\text{known}} = 0 \\
&\underbrace{\Pi_2(a_2 = 1|s)}_{\text{unknown}} = \sigma_1(a_1 = 1|s)\,\underbrace{\pi_2(a_1 = 1, a_2 = 1|s)}_{\text{unknown}} \\
&+ (1 - \sigma_1(a_1 = 1|s))\,\underbrace{\pi_1(a_1 = 0, a_2 = 1|s)}_{\text{unknown}} \\
&\underbrace{\Pi_2(a_2 = 0|s)}_{\text{known}} = 0
\end{aligned}
\tag{A.3.1}
$$

After fixing d, the left-hand side of equation (A.3.1) comprises two unknown components: the deterministic utilities $\Pi_1(a_1 = 1, s)$ and $\Pi_2(a_2 = 1, s)$. In accordance with the assumption 1.2.2, the expected payoff of remaining inactive, $\Pi_1(a_1 = 0, s)$ and $\Pi_2(a_2 = 0, s)$, is known to the econometrician, which normalized to zero. Conversely, on the right-hand side of the equation, there exist four unknowns: $\pi_1(a_1 = 1, a_2 = 1, s)$, $\pi_1(a_1 = 1, a_2 = 0, s)$, $\pi_2(a_1 = 1, a_2 = 1, s)$, and $\pi_2(a_1 = 0, a_2 = 1, s)$. This results in an under-identified scenario.[5]

The utilization of exclusion restrictions is a common strategy for disentangling the system of equations in (A.3.1) to satisfy identification condition.[6] The exclusion restriction requires that the relevant state variable d can be split into two components: one that is universal across all players within the same market, referred to

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

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

as s_x, and player-specific shocks denoted as s_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 A.3.1 (Exclusion Restriction)**.**

$$\pi_i(a_i, a_{-i}, s) = \pi(a_i, a_{-i}, s_x, s_i).$$

Given the imposition of exclusion restrictions, let $d = (s_x, s_1, s_2)$, where s_x represents common state variables for players 1 and 2 in the same markets, s_1 is player 1's specific state variable, and s_2 is player 2's specific state variable. When s_x 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 $\Pi_i(a_i = 1|s_1 = H, s_2 = H) := \Pi_i(a_i = 1|H, H)$ and

$\sigma_i(a_i = 1, a_{-i} = 1|s_1 = H, s_2 = H) := \sigma_i(a_i = 1, a_{-i} = 1|H, H)$ for brevity.

$$
\begin{aligned}
\Pi_1(a_1 = 1|H, H) &= \sigma_2(a_2 = 1|H, H)\underbrace{\pi_1(a_1 = 1, a_2 = 1|s_1 = H, s_2 = H)}_{=\pi_1(a_1=1,a_2=1|s_1=H)} \\
&\quad + (1 - \sigma_2(a_2 = 1|H, H))\underbrace{\pi_1(a_1 = 1, a_2 = 0|s_1 = H, s_2 = H)}_{=\pi_1(a_1=1,a_2=0|s_1=H)} \\
\Pi_1(a_1 = 1|H, L) &= \sigma_2(a_2 = 1|H, L)\underbrace{\pi_1(a_1 = 1, a_2 = 1|H, L)}_{=\pi_1(a_1=1,a_2=1|s_1=H)} \\
&\quad + (1 - \sigma_2(a_2 = 1|H, L))\underbrace{\pi_1(a_1 = 1, a_2 = 0|H, L)}_{=\pi_1(a_1=1,a_2=0|s_1=1)} \\
\Pi_1(a_1 = 1|L, H) &= \sigma_2(a_2 = 1|L, H)\underbrace{\pi_1(a_1 = 1, a_2 = 1|L, H)}_{=\pi_1(a_1=1,a_2=1|s_1=L)} \\
&\quad + (1 - \sigma_2(a_2 = 1|L, H))\underbrace{\pi_1(a_1 = 1, a_2 = 0|L, H)}_{=\pi_1(a_1=1,a_2=0|s_1=L)} \\
\Pi_1(a_1 = 1|L, L) &= \sigma_2(a_2 = 1|L, L)\pi_1(a_1 = 1, a_2 = 1|L, L) \\
&\quad + (1 - \sigma_2(a_2 = 1|L, L))\underbrace{\pi_1(a_1 = 1, a_2 = 0|L, L)}_{=\pi_1(a_1=1,a_2=0|s_1=L)} \qquad \text{(A.3.2)}
\end{aligned}
$$

The exclusion restriction implies that $\pi_1(a_1 = 1, a_2 = 1|s_1 = H, s_2 = H) = \pi_1(a_1 =$

$1, a_2 = 1|s_1 = H, s_2 = L)$, which leads to the following equation:

$$\begin{aligned}
\Pi_1(a_1 = 1|H, H) &= \sigma_2(a_2 = 1|H, H)\pi_1(a_1 = 1, a_2 = 1|s_1 = H) \\
&+ (1 - \sigma_2(a_2 = 1|H, H))\pi_1(a_1 = 1, a_2 = 0|s_1 = H) \\
\Pi_1(a_1 = 1|H, L) &= \sigma_2(a_2 = 1|H, L)\pi_1(a_1 = 1, a_2 = 1|s_1 = H) \\
&+ (1 - \sigma_2(a_2 = 1|H, L))\pi_1(a_1 = 1, a_2 = 0|s_1 = H) \\
\Pi_1(a_1 = 1|L, H) &= \sigma_2(a_2 = 1|L, H)\pi_1(a_1 = 1, a_2 = 1|s_1 = L) \\
&+ (1 - \sigma_2(a_2 = 1|H, L))\pi_1(a_1 = 1, a_2 = 0|s_1 = L) \\
\Pi_1(a_1 = 1|L, L) &= \sigma_2(a_2 = 1|L, L)\pi_1(a_1 = 1, a_2 = 1|s_1 = L) \\
&+ (1 - \sigma_2(a_2 = 1|L, L))\pi_1(a_1 = 1, a_2 = 0|s_1 = L) \qquad \text{(A.3.3)}
\end{aligned}$$

The left-hand side of the system of equations in A.3.3 involves $dim(s_1) \times dim(s_2) = 2 \times 2 = 4$ unknowns. In contrast, the right-hand side of the equation encompasses four unknowns $(\pi_1(a_1 = 1, a_2 = 1|s_1 = H), \pi_1(a_1 = 1, a_2 = 0|s_1 = H), \pi_1(a_1 = 1, a_2 = 1|s_1 = L), \pi_1(a_1 = 1, a_2 = 0|s_1 = L))$. This implies that the equation A.3.3 is identified. A similar argument can be applied to demonstrate the identification of payoffs for player 2 based on equation A.3.4.

$$\begin{aligned}
\Pi_2(a_1 = 1|H, H) &= \sigma_1(a_1 = 1|H, H)\pi_2(a_1 = 1, a_2 = 1|s_2 = H) \\
&+ (1 - \sigma_1(a_1 = 1|H, H))\pi_2(a_1 = 1, a_2 = 0|s_2 = H) \\
\Pi_2(a_1 = 1|H, L) &= \sigma_1(a_1 = 1|H, L)\pi_2(a_1 = 1, a_2 = 1|s_2 = H) \\
&+ (1 - \sigma_1(a_1 = 1|H, L))\pi_2(a_1 = 1, a_2 = 0|s_2 = H) \\
\Pi_2(a_1 = 1|L, H) &= \sigma_1(a_1 = 1|L, H)\pi_2(a_1 = 1, a_2 = 1|s_2 = L) \\
&+ (1 - \sigma_1(a_1 = 1|H, L))\pi_2(a_1 = 1, a_2 = 0|s_2 = L) \\
\Pi_2(a_1 = 1|L, L) &= \sigma_1(a_1 = 1|L, L)\pi_2(a_1 = 1, a_2 = 1|s_2 = L) \\
&+ (1 - \sigma_1(a_2 = 1|L, L))\pi_2(a_1 = 1, a_2 = 0|s_2 = L) \qquad \text{(A.3.4)}
\end{aligned}$$

More generally, the choice-specific value functions as expressed in equation (1.2.3) could lead to the following expression:

$$\underbrace{\Pi_i(a_i = 1, s)}_{\text{unknown}} = \sum_{a_{-i} \in A_{-i}} \underbrace{\sigma_{-i}(a_{-i}|s)}_{\text{known}} \underbrace{\pi_i(a_i = 1, a_{-i}, s)}_{\text{unknown}} \text{ for all } i = 1, ..., N. \tag{A.3.5}$$

This equation implies that, once s_x 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 $(\pi_i(a_i = 1, a_{-i}, s))$.[7] Consequently, without the introduction of exclusion restrictions, the system of equations in (A.3.5) cannot be identified. Next, armed with the assumption of exclusion restrictions as given in Assumption A.3.1, I can reformulate the choice-specific value function as follows:

$$\underbrace{\Pi_i(a_i = 1, s_i, s_{-i})}_{\text{unknown}} = \sum_{a_{-i} \in A_{-i}} \underbrace{\sigma_{-i}(a_{-i}|s_i, s_{-i})}_{\text{known}} \underbrace{\pi_i(a_i = 1, a_{-i}, s_i)}_{\text{unknown}} \text{ for all } i = 1, ..., N. \tag{A.3.6}$$

Evidently, the number of unknowns (free parameters) on the left-hand side has reduced from $\pi_i(a_i = 1, a_{-i}|s_i, s_{-i})$ to $\pi_i(a_i = 1, a_{-i}, s_i)$. This induces more variations on the right-hand side than the number of unknowns on the left-hand side flow utilities $\Pi_i(a_i = 1, s_i, s_{-i})$, which is the over-identified case.[8] It follows the necessary condition that the supports of beliefs $\sigma_{-i}(a_{-i} = 1|s_i, s_{-i})$ need to be sufficiently rich with 2^{n-1} points. It is evident that the incorporation of high-dimensional common market characteristics s_x does not impact the identification conditions, as long as the rank conditions for s_{-i} given s_i are satisfied. The crucial determinants of identification are the number of players, choices, and variations $s_{-i}|s_i$ involved in the system, rather than the dimensionality of the common market characteristics s_x.

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

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

Remark 1 (Identification)**.**

1. *Suppose that Assumption 1.2.2 and Assumption A.3.1 are satisfied. As long as there are* 2^{n-1} *points in the support of conditional distribution* $s_{-i}|s_i$ *related to* $\sigma_{-i}(a_{-i}|s_i, s_{-i})$*, then the necessary condition holds.*
2. *Allowing high dimensional market characteristics* s_x *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 s_{-i}, will directly impact player $-i$'s entry probability. Conversely, player i's entry probability is indirectly influenced by s_{-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 (A.3.6).

Appendix B Tables

Table 13: Summary Statistics: High Elderly Population 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: "High elderly population town" 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.

Table 14: Summary Statistics: Non-High Elderly Population 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	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-high elderly population 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 the township. [g] "Prop. Insurance" refers to the ratio of the population within each age group enrolled in health insurance.

Table 15: The New Entries in Chain Pharmacies with Different Distances and the Number of Independent Pharmacies.

	(1)	(2)	(3)	(4)
	Independent Stores	Independent Stores	Independent Stores	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. Standard errors are clustered at the town level. Significance levels are denoted by + p < 0.10, * p < 0.05, ** p <0.01, and *** p < 0.001.

Table 16: 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. Column (1) includes observable demographic variables and county-fixed effects. Column (2) includes observable demographic variables, county-fixed effects, and year-fixed effects. Standard errors are clustered at the town level. Significance levels are denoted by + p<0.10, * p<0.05, ** p<0.01.*** p<0.001.

Table 17: 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 18: Top 10 Importance Features from Xg boosting

(a) High Elderly Population 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-High Elderly Population 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 the high elderly population township/non-high elderly population township. #. denote the number and % denotes the share of demographic groups out of the total population in the towns.

Table 19: 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 20: 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<0.01.*** p<0.001.

Table 21: (Selected) Descriptive Statistics: Non-High Elderly Population 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-high elderly population 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 the township. [g] "Prop. Insurance" refers to the ratio of the population within each age group enrolled in health insurance.

Table 22: (Selected) Descriptive Statistics: High Elderly Population 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: "High elderly population township" is defined as townships with an age over 65 population ratio higher 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.

Table 23: Confusion Matrix

(a) High Elderly Population 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-High Elderly Population 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 24: Full Results for Bajari et al. (2010b)

	(1)	(2)
	Elderly Town	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. Unemployment	-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

Table 25: 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_e	The effect of store-specific shifter: Employees Size
β_x	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
s_{imt}	Player i specific shifter: Employees Size
s_{-imt}	Player $-i$ specific shifter: Employees Size
s_{xmt}	Common market characteristics
s_{xmt}^{pre}	Pre-selected common market characteristics
s_{xmt}^{pool}	Pool of richer common market characteristics
$s_{xmt}^{interaction}$	Interaction terms of s_{xmt}^{pool}

Table 26: First Stage Reduced Form CCP: Existing Method Bajari et al. (2010b)

	(1) High Elderly Population Towns	(2) Non-High Elderly Population Towns
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 Entry and Exit in township m and year t. Column (1) includes towns with a high elderly population. Column (2) includes towns with a non-high elderly population. Towns are defined as high elderly population towns when the percentage of residents aged over 65 is higher than 20 percent in the year 2000. Towns are defined as non-high elderly population towns if the percentage of residents aged over 65 is lower than 20 percent in the year 2000. Standard errors are clustered at the county level. Significance levels are denoted by + p<0.10, * p<0.05, ** p<0.01, *** p<0.001.

Table 27: 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: Samples include towns with a non-high elderly population in the years 2000-2019. In column (1), I use existing estimators based on the approach described in Bajari et al. (2010b). I use pre-selected market covariates, as described in Appendix 22. In column (2), I use my developed orthogonal estimators, which employ a data-driven approach to variable selection, I utilize a pool of market characteristics described in Appendix 13. I further use sample splitting and moment conditions based on equation (2.3.6) to remove biases from ML in the first stage of nuisance parameters estimation. Standard errors are clustered at the county level. Significance levels are denoted by + p<0.10, * p<0.05, ** p<0.01, *** p<0.001.

Table 28: 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

Table 29: Expected Number of Stores under Counterfactual Scenario 2 (Year: 2019)

	(Average) Independent Pharmacy Counts			
	Predicted	CF S2	$\triangle$	$\triangle$%
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

Table 30: Comparison of Rite Aid Stores

Year	Financial Reports (10-k)	Raw Data Sample
2012	228	231
2013	226	231
2014	224	231
2015	224	218
2016	225	218
2017 (Year before the merger approval)	224	222

Notes: I report the number of Rite Aid stores as disclosed by Rite Aid in their 10-K annual financial reports, and compare these figures with my samples. The results show that the numbers are quite close.

Table 31: Heterogeneity by Income Group: Horizontal Merger and Number of Pharmacy Stores

Outcome:	Total Pharmacies			
	(1)	(2)	(3)	(4)
Closure of Merged Pharmacy	-0.861***	-0.861***	-0.672***	-0.456***
	(0.149)	(0.149)	(0.162)	(0.112)
Census Tract FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	2,856	2,856	2,856	2,856
Outcome mean	1.409	1.719	1.549	1.405
Income Group	Quartile 0-0.25	Quartile 0.25-0.5	Quartile 0.5-0.75	Quartile 0.75-1

Notes: Estimates are from staggered TWFE Callaway and Sant'Anna (2021) regressions of the number of stores in census tract c in year t on an indicator called "closure of merged pharmacy," which equals one for a census tract in the years following the closure of either Walgreens or Rite Aid after the merger approval and zero otherwise. The outcome of interest is the total number of pharmacies. Column (1) includes the lowest income group. Column (2) includes the second quantile income group. Column (3) includes the third quantile income group. Column (4) includes the fourth quantile income group. Standard errors are clustered at the census tract level. Significance levels are denoted by + p<0.10, * p<0.05, ** p<0.01, *** p<0.001.

Table 32: Heterogeneity by Income Group: Horizontal Merger and Number of Non-merged Pharmacy Stores

Outcome:	Total Pharmacies			
	(1)	(2)	(3)	(4)
Closure of Merged Pharmacy	0.028	-0.113	0.053	0.239***
	(0.149)	(0.149)	(0.162)	(0.097)
Census Tract FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	2,856	2,856	2,856	2,856
Outcome mean	1.133	1.438	1.200	1.021
Income Group	Quartile 0-0.25	Quartile 0.25-0.5	Quartile 0.5-0.75	Quartile 0.75-1

Notes: Estimates are from staggered TWFE Callaway and Sant'Anna (2021) regressions of the number of non-merged stores in census tract c in year t on an indicator called "closure of merged pharmacy," which equals one for a census tract in the years following the closure of either Walgreens or Rite Aid after the merger approval and zero otherwise. The outcome of interest is the total number of pharmacies. Column (1) includes the lowest income group. Column (2) includes the second quantile income group. Column (3) includes the third quantile income group. Column (4) includes the fourth quantile income group. Standard errors are clustered at the census tract level. Significance levels are denoted by + p<0.10, * p<0.05, ** p<0.01, *** p<0.001.

Table 33: Heterogeneity by Income Group: Horizontal Merger and New Entries of Non-merged Pharmacy Stores

Outcome:	Total Pharmacies			
	(1)	(2)	(3)	(4)
Closure of Merged Pharmacy	-0.021	-0.113	0.098	0.085
	(0.080)	(0.097)	(0.063)	(0.069)
Census Tract FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	2,856	2,856	2,856	2,856
Outcome mean	0.121	0.163	0.145	0.143
Income Group	Quartile 0-0.25	Quartile 0.25-0.5	Quartile 0.5-0.75	Quartile 0.75-1

Notes: Estimates are from staggered TWFE Callaway and Sant'Anna (2021) regressions of the number of entries of non-merged stores in census tract c in year t on an indicator called "closure of merged pharmacy," which equals one for a census tract in the years following the closure of either Walgreens or Rite Aid after the merger approval and zero otherwise. The outcome of interest is the total number of pharmacies. Column (1) includes the lowest income group. Column (2) includes the second quantile income group. Column (3) includes the third quantile income group. Column (4) includes the fourth quantile income group. Standard errors are clustered at the census tract level. Significance levels are denoted by + p<0.10, * p<0.05, ** p<0.01, *** p<0.001.

Table 34: Robustness Check: Adding Adjacent Markets: Horizontal Merger and Number of Pharmacy Stores

Outcome:	Total Pharmacies			
	(1)	(2)	(3)	(4)
Closure of Merged Pharmacy	-0.289***	-0.287***	-0.385***	-0.383***
	(0.049)	(0.049)	(0.142)	(0.114)
Census Tract FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	11,413	11,413	2,244	2,244
Outcome mean	1.520	1.520	1.682	1.682
Sample	Full sample	Full sample	Sub-sample	Sub-sample
Untreated Group	Never-treated	Not-yet-treated	Never-treated	Not-yet-treated

Notes: Estimates are from staggered TWFE Callaway and Sant'Anna (2021) regressions of the number of stores in census tract c in year t on an indicator called "closure of merged pharmacy," which equals one for a census tract in the years following the closure of either Walgreens or Rite Aid after the merger approval and zero otherwise. I extend the treatment of markets to adjacent markets where Walgreens or Rite Aid were present in 2017. The outcome of interest is the total number of pharmacies. Column (1) includes a full sample with never treated as an untreated group. Column (2) includes a full sample with not-yet-treated as an untreated group. Column (3) includes census tract with the presence of either Walgreens or Rite Aid prior year to merger approval with never treated as an untreated group. Column (4) census tract with the presence of either Walgreens or Rite Aid prior year to merger approval with not-yet-treated as an untreated group. Standard errors are clustered at the census tract level. Significance levels are denoted by + p<0.10, * p<0.05, ** p<0.01, *** p<0.001.

Table 35: Robustness Check of Adding Adjacent Markets: Horizontal Merger and Non-merged Pharmacy Stores

Outcome:	Total Number of Non-merging Pharmacies			
	(1)	(2)	(3)	(4)
Closure of Merged Pharmacy	0.045	0.047	-0.087	-0.073
	(0.041)	(0.041)	(0.144)	(0.115)
Census Tract FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	11,408	11,413	2,244	2,244
Outcome mean	1.198	1.198	1.210	1.21
Sample	Full sample	Full sample	Sub-sample	Sub-sample
Untreated Group	Never-treated	Not-yet-treated	Never-treated	Not-yet-treated

Notes: Estimates are from staggered TWFE Callaway and Sant'Anna (2021) regressions of the number of stores in census tract c in year t on an indicator called "closure of merged pharmacy," which equals one for a census tract in the years following the closure of either Walgreens or Rite Aid after the merger approval and zero otherwise. I extend the treatment of markets to adjacent markets where Walgreens or Rite Aid were present in 2017. The outcome of interest is the total number of non-merging pharmacies. Column (1) includes a full sample with never treated as an untreated group. Column (2) includes a full sample with not-yet-treated as an untreated group. Column (3) includes census tract with the presence of either Walgreens or Rite Aid prior year to merger approval with never treated as an untreated group. Column (4) census tract with the presence of either Walgreens or Rite Aid prior year to merger approval with not-yet-treated as an untreated group. Standard errors are clustered at the census tract level. Significance levels are denoted by + p<0.10, * p<0.05, ** p<0.01, *** p<0.001.

Table 36: Robustness Check: Adding Adjacent Markets: Horizontal Merger and Entries of Non-merged Pharmacy Stores

Outcome:	Total Number of Entries of Non-merging Pharmacies			
	(1)	(2)	(3)	(4)
Closure of Merged Pharmacy	-0.010	-0.010	-0.0010	0.011
	(0.029)	(0.029)	(0.061)	(0.044)
Census Tract FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	11,408	11,413	2,244	2,244
Outcome mean	0.143	0.143	0.152	0.152
Sample	Full sample	Full sample	Sub-sample	Sub-sample
Untreated Group	Never-treated	Not-yet-treated	Never-treated	Not-yet-treated

Notes: Estimates are from staggered TWFE Callaway and Sant'Anna (2021) regressions of the number of stores in census tract c in year t on an indicator called "closure of merged pharmacy," which equals one for a census tract in the years following the closure of either Walgreens or Rite Aid after the merger approval and zero otherwise. I extend the treatment of markets to adjacent markets where Walgreens or Rite Aid were present in 2017. The outcome of interest is the total entries of non-merging pharmacies. Column (1) includes a full sample with never treated as an untreated group. Column (2) includes a full sample with not-yet-treated as an untreated group. Column (3) includes census tract with the presence of either Walgreens or Rite Aid prior year to merger approval with never treated as an untreated group. Column (4) census tract with the presence of either Walgreens or Rite Aid prior year to merger approval with not-yet-treated as an untreated group. Standard errors are clustered at the census tract level. Significance levels are denoted by + p<0.10, * p<0.05, ** p<0.01, *** p<0.001.

Table 37: Robustness Check: Alternative Demographic Controls: Horizontal Merger and Number of Pharmacy Stores

Outcome:	Total Pharmacies			
	(1)	(2)	(3)	(4)
Closure of Merged Pharmacy	-0.685***	-0.682***	-0.833***	-0.795***
	(0.071)	(0.071)	(0.143)	(0.125)
Census Tract FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	11,408	11,413	2,244	2,244
Outcome mean	1.520	1.520	2.367	2.367
Sample	Full sample	Full sample	Sub-sample	Sub-sample
Untreated Group	Never-treated	Not-yet-treated	Never-treated	Not-yet-treated

Notes: Estimates are from difference-in-differences regressions of the number of stores in census tract c in year t on an indicator called "horizontal merger," which equals one for a census tract in the years following the closure of either Walgreens or Rite Aid after the merger approval and zero otherwise. I use population density instead of total population as a robustness check. The outcome of interest is the total number of pharmacies. Column (1) includes a full sample with never treated as an untreated group. Column (2) includes a full sample with not-yet-treated as an untreated group. Column (3) includes census tract with the presence of either Walgreens or Rite Aid prior year to merger approval with never treated as an untreated group. Column (4) census tract with the presence of either Walgreens or Rite Aid prior year to merger approval with not-yet-treated as an untreated group. Standard errors are clustered at the census tract level. Significance levels are denoted by + p<0.10, * p<0.05, ** p<0.01, *** p<0.001.

Table 38: Robustness Check: Alternative Demographic Controls: Horizontal Merger and Non-merged Pharmacy Stores

Outcome:	Total Number of Non-merging Pharmacies			
	(1)	(2)	(3)	(4)
Closure of Merged Pharmacy	0.070	0.072	0.041	0.049
	(0.057)	(0.057)	(0.062)	(0.063)
Census Tract FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	11,408	11,413	2,244	2,244
Outcome mean	1.198	1.198	1.329	1.329
Sample	Full sample	Full sample	Sub-sample	Sub-sample
Untreated Group	Never-treated	Not-yet-treated	Never-treated	Not-yet-treated

Notes: Estimates are from staggered TWFE Callaway and Sant'Anna (2021) regressions of the number of stores in census tract c in year t on an indicator called "closure of merged pharmacy," which equals one for a census tract in the years following the closure of either Walgreens or Rite Aid after the merger approval and zero otherwise. I use population density instead of total population as a robustness check. The outcome of interest is the total number of non-merging pharmacies. Column (1) includes a full sample with never treated as an untreated group. Column (2) includes a full sample with not-yet-treated as an untreated group. Column (3) includes census tract with the presence of either Walgreens or Rite Aid prior year to merger approval with never treated as an untreated group. Column (4) census tract with the presence of either Walgreens or Rite Aid prior year to merger approval with not-yet-treated as an untreated group. Standard errors are clustered at the census tract level. Significance levels are denoted by + p<0.10, * p<0.05, ** p<0.01, *** p<0.001.

Table 39: Robustness Check: Alternative Demographic Controls: Horizontal Merger and Entries of Non-merged Pharmacy Stores

Outcome:	Total Number of Entries by Non-merging Pharmacies			
	(1)	(2)	(3)	(4)
Closure of Merged Pharmacy	0.022	0.024	0.034	0.036
	(0.039)	(0.039)	(0.043)	(0.043)
Census Tract FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	11,408	11,413	2,244	2,244
Outcome mean	0.143	0.143	0.152	0.152
Sample	Full sample	Full sample	Sub-sample	Sub-sample
Untreated Group	Never-treated	Not-yet-treated	Never-treated	Not-yet-treated

Notes: Estimates are from staggered TWFE Callaway and Sant'Anna (2021) regressions of the number of stores in census tract c in year t on an indicator called "closure of merged pharmacy," which equals one for a census tract in the years following the closure of either Walgreens or Rite Aid after the merger approval and zero otherwise. I use population density instead of total population as a robustness check. The outcome of interest is the total entries of non-merging pharmacies. Column (1) includes a full sample with never treated as an untreated group. Column (2) includes a full sample with not-yet-treated as an untreated group. Column (3) includes census tract with the presence of either Walgreens or Rite Aid prior year to merger approval with never treated as an untreated group. Column (4) census tract with the presence of either Walgreens or Rite Aid prior year to merger approval with not-yet-treated as an untreated group. Standard errors are clustered at the census tract level. Significance levels are denoted by + $p<0.10$, * $p<0.05$, ** $p<0.01$, *** $p<0.001$.

Table 40: de Chaisemartin and Xavier D'Haultfoeuille's Staggered TWFE: Horizontal Merger and Pharmacy Stores

Outcome:	Pharmacies	Non-merging Pharmacies	Non-merging Pharmacy Entries
	(1)	(2)	(3)
Horizontal Merger	-0.657***	0.0556	0.003
	(0.0504)	(0.0426)	(0.0459)
Census Tract FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Controls	Yes	Yes	Yes
Observations	11,413	11,413	11,413
Outcome mean	1.520	1.198	0.143

Notes: Estimates are from staggered TWFE Callaway and Sant'Anna (2021) regressions of the number of stores in census tract c in year t on an indicator called "closure of merged pharmacy," which equals one for a census tract in the years following the closure of either Walgreens or Rite Aid after the merger approval and zero otherwise. Column (1) includes the total number of pharmacies. Column (2) includes the total number of pharmacies from non-merging firms. Column (3) includes the total number of new entries by non-merging pharmacies. Standard errors are clustered at the census tract level. Significance levels are denoted by + p<0.10, * p<0.05, ** p<0.01, *** p<0.001.

Table 41: 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 the 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 provides 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.

Appendix C Figures

Figure 18: Timing of the Game

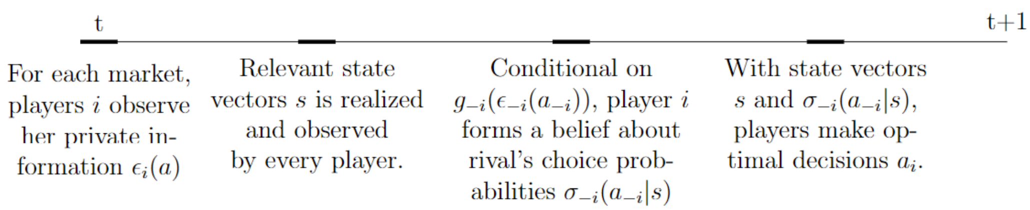
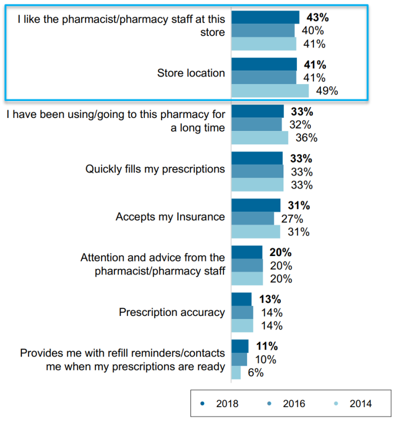
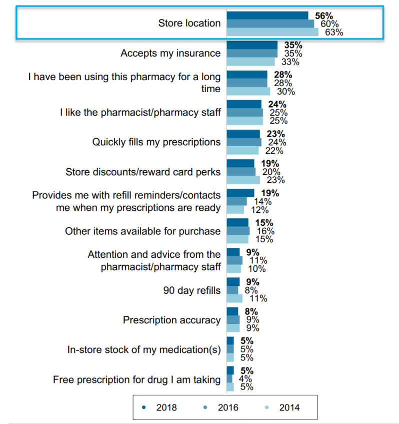


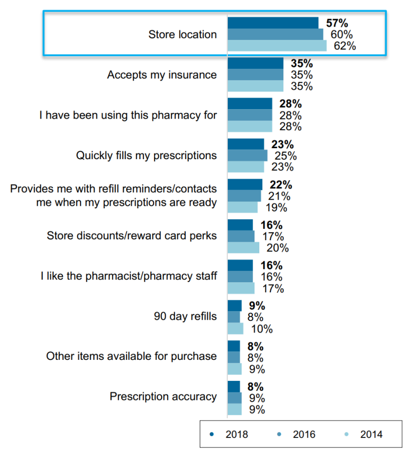
Figure 19: Industry Survey: Main Reason for Choosing Primary Pharmacy



(a) Independent Pharmacy



(b) Supermarket Pharmacy



(c) Chain Pharmacy

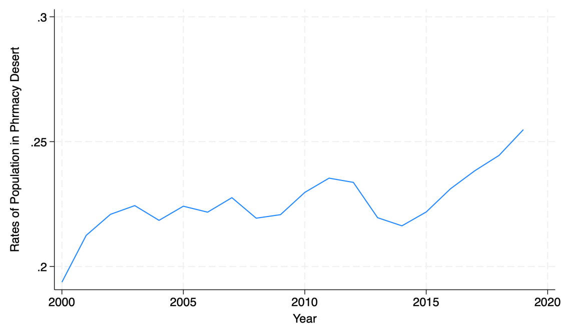
Sources: Pharmacy Satisfaction Data Summary Report, 2018 Boehringer Ingelheim Pharmaceuticals, Inc

Figure 20: Industry Survey: Main Reason for Switching Primary Pharmacy

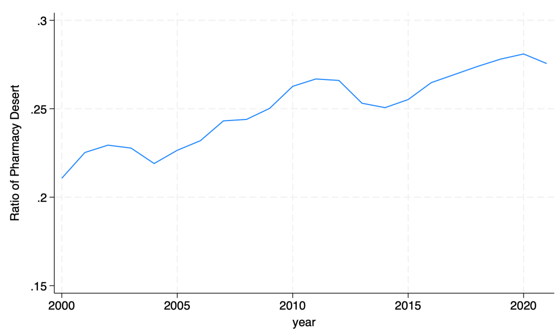
	Total 2018 (n=2,244)	Total 2016 (n=2,035)
Close to my house/work	33%	34%
Health insurance requires me to use this pharmacy	21%	19%
I moved	18%	20%
90 day refills	18%	19%
Like the pharmacy staff	18%	13%
Refill reminders/contacts me when Rx ready	17%	14%
Has other items available for purchase	12%	12%
Offers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

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

Figure 21: Trends in Pharmacy Deserts: Alternative Definition



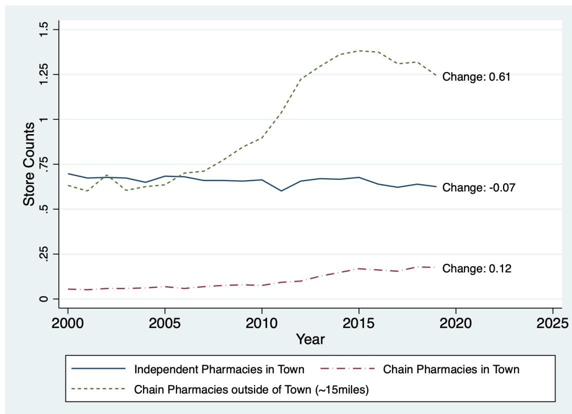
(a) Population Weight-Average



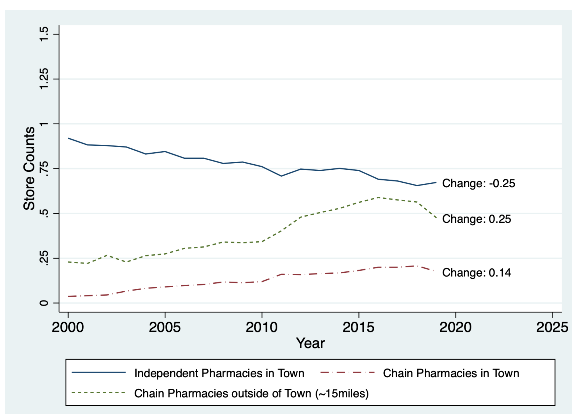
(b) Within 5 miles

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.

Figure 22: (Average) Number of Independent/Chain Pharmacies between 2000-2019 by Age Group

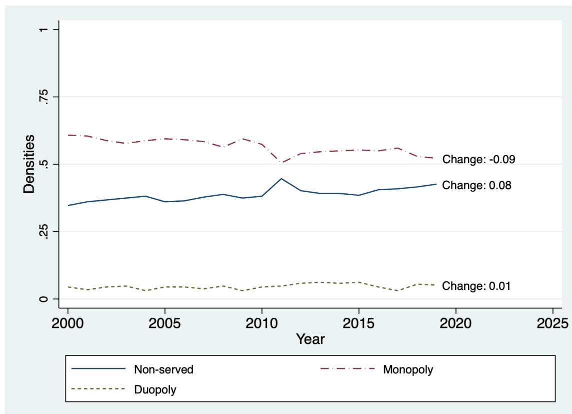


(a) Non-High Elderly Population Town

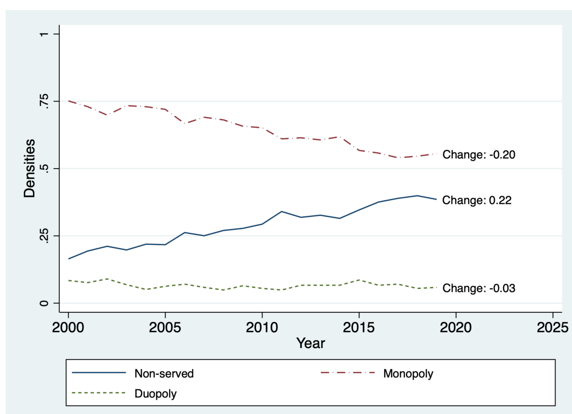


(b) High Elderly Population Town

Figure 23: Distribution of Market Structure of Independently-Owned Pharmacy by Age Group

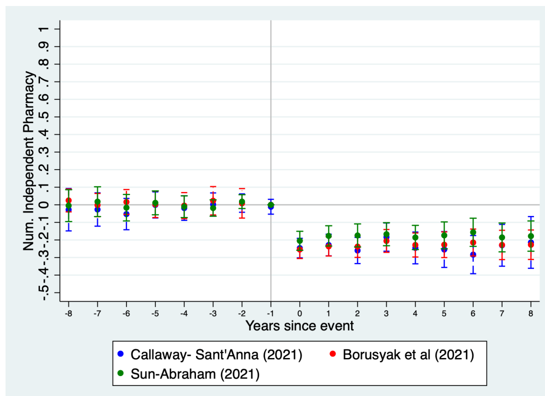


(a) Non-High Elderly Population Township



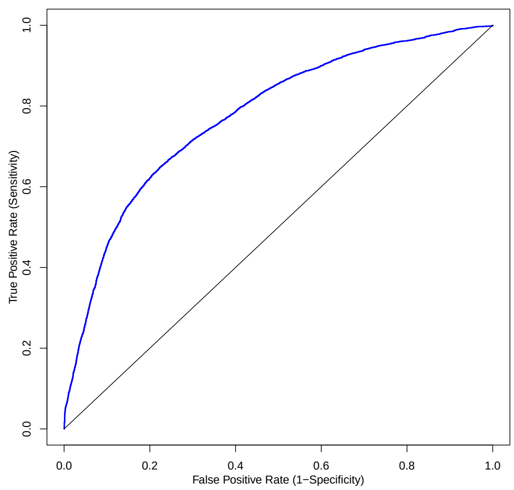
(b) High Elderly Population Township

Figure 24: More Event Study: The effects of chain pharmacy entry on local independent pharmacy (Robustness Check)

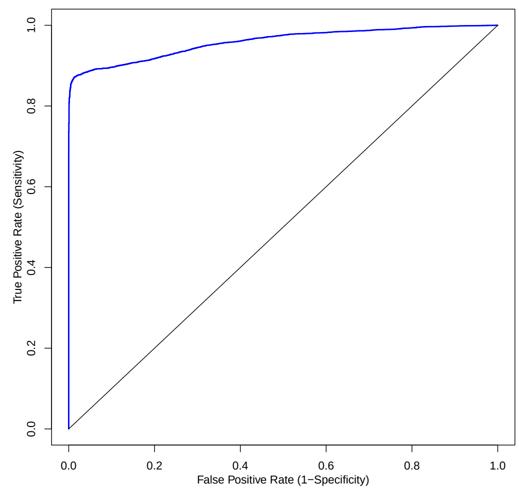


Note: This figure presents coefficient plots from event-study difference-in-differences analyses, which regress the number of independent pharmacies in a township on year-fixed effects, town fixed effects, control variables, and market× year-fixed effects. For robustness check purposes, I experiment with the staggered treatment frameworks proposed by Callaway and Sant'Anna (2021), Borusyak et al. (2021), and Sun and Abraham (2021). The sample consists of 802 townships between 2000 and 2019. The baseline period, omitted in this analysis, is $t = -1$, representing the last pre-treatment period. Standard errors are clustered at the town level, and error bars represent 95% confidence intervals.

Figure 25: ROC in First Stage Estimation of CCP: High Elderly Population Town



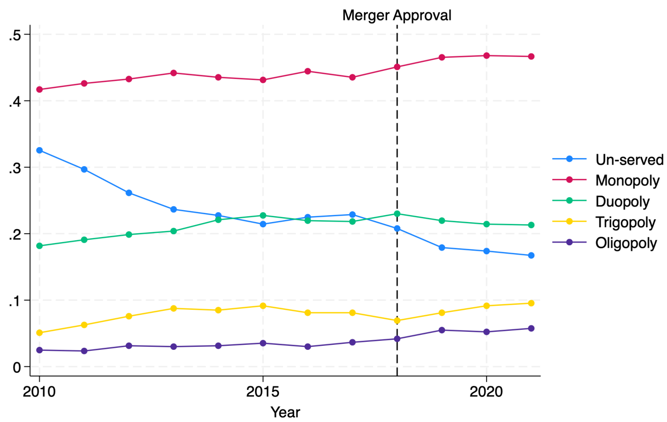
(a) Logit
AUC: 0.7766



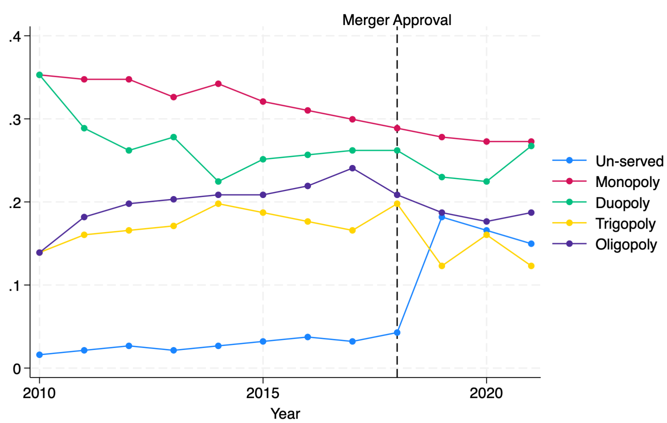
(b) XG Boosting
AUC: 0.9592

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

Figure 26: Change in Market Structures with Before/After the Merger Approval



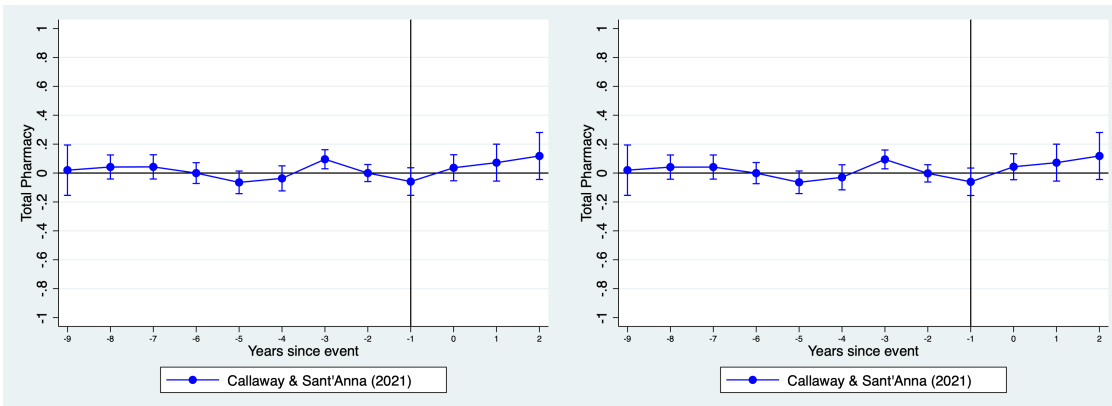
(a) Towns without Walgreens or Rite Aid in 2017



(b) Towns with Walgreens or Rite Aid in 2017

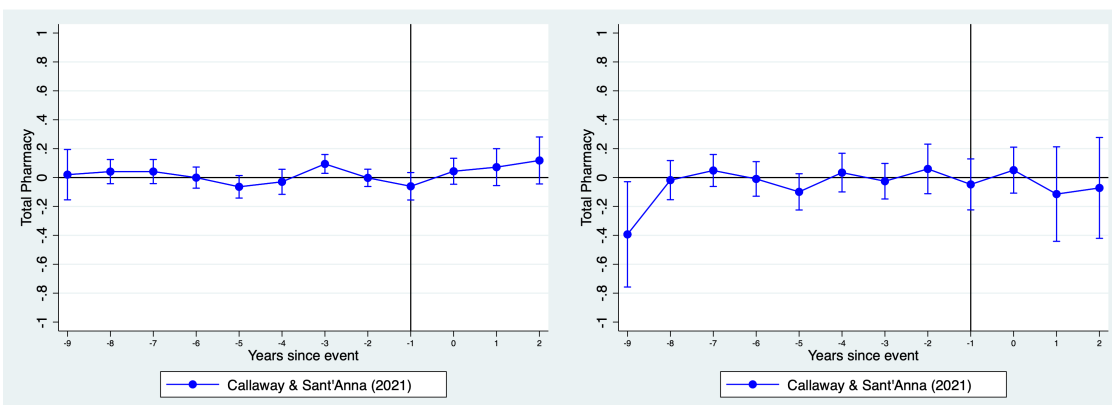
Note: This figure illustrates the market structure changes following the approval of a horizontal merger. Figure 26a represents towns without Walgreens or Rite Aid in 2017 (prior year to merger approval) and Figure 26b represents towns with Walgreens or Rite Aid in 2017. The market structures are categorized by the level of competition, ranging from unserved markets (no providers) to oligopoly (≥ 4). The vertical dashed line labeled "Merger Approval" denotes the year 2018 when the merger was approved, providing a clear before-and-after comparison.

Figure 27: Dynamic Effects of Horizontal Merger on Non-merging Pharmacies



(a) Full sample
Untreated group: Never treated

(b) Full sample
Untreated group: Not-yet-treated

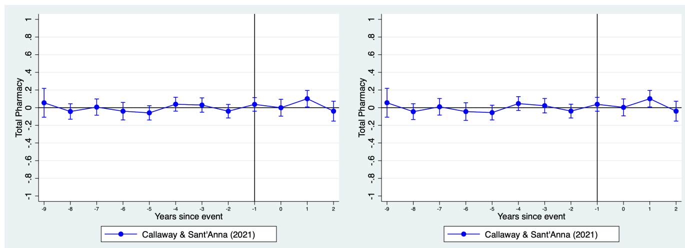


(c) Sub-sample
Untreated group: Never treated

(d) Sub-sample
Untreated group: Not-yet-treated

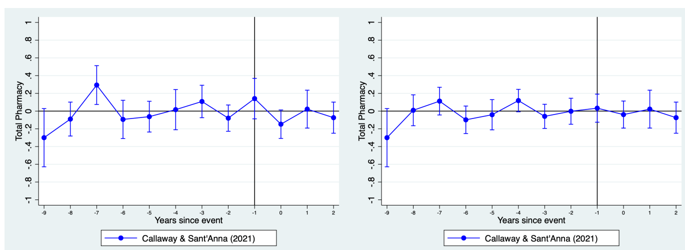
Notes: Coefficient plots from event-study difference-in-differences analyses that regress the number of new entrants by non-merged firms in a census tract on year fixed effects, census tract fixed effects, control variables, and market× year fixed effects. The full sample consists of census tracts between 2010 and 2021. The sub-sample includes census tracks where a Rite Aid or Walgreens store was present before the horizontal merger approval. The omitted baseline period is $t = -1$, which is the last pre-treatment period. Standard errors are clustered at the census-tract level and error bars represent 95 confidence intervals.

Figure 28: Dynamic Effects of Horizontal Merger on Entries of Non-merging Pharmacies



(a) Full sample Untreated group: Never treated

(b) Full sample Untreated group: Not-yet-treated

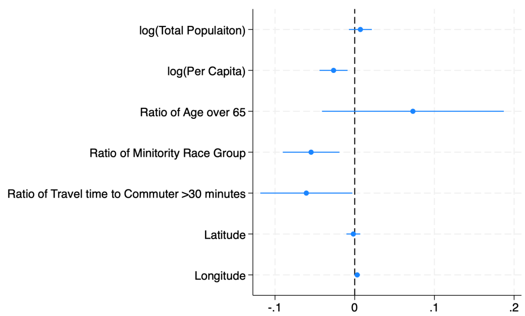


(c) Sample from Walgreens and Rite Aid Stores in 2017 Untreated group: Never treated

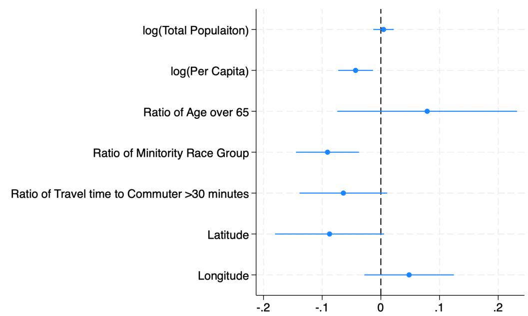
(d) Sample from Walgreens and Rite Aid Stores in 2017 Untreated group: Not-yet-treated

Notes: Coefficient plots from event-study difference-in-differences analyses that regress the number of new entrants by non-merged firms in a census tract on year fixed effects, census tract fixed effects, control variables, and market× year fixed effects. The full sample consists of census tracts between 2010 and 2021. The sub-sample includes census tracks where a Rite Aid or Walgreens store was present before the horizontal merger approval. The baseline period is $t = -1$, which is the last pre-treatment period. Standard errors are clustered at the census-tract level and error bars represent 95 confidence intervals.

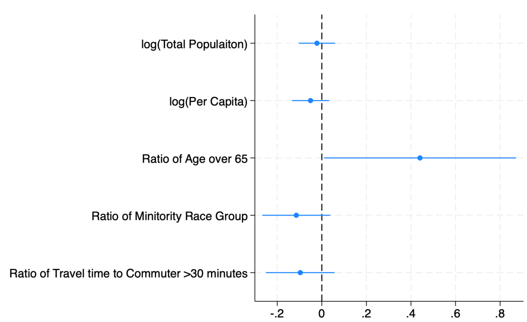
Figure 29: Balancing Test of the Exit of Merged Firms' Treatment on Pre-Market Characteristics



(a) Census Tract Cluster-Standard Error



(b) Adding County FE



(c) Census Tract FE

Notes: Regression of the treatment indicator (exit of Walgreens/Rite Aid) on observable pre-market characteristics in the year prior to the exit of merged firms. The coefficients of these regressions with their 95% confidence interval are plotted in each sug figure which uses census tract level cluster standard error (a), add county fixed effects (b), and use census tract fixed effects (c) with census tract clustered standard errors.

Appendix D Data Description

D.0.1 Construction of Data

To construct my final dataset, I leverage 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 a 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 a number of chains within 5-30 miles.

D.0.2 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 14 and Table 13.

D.0.3 Health related variables

To account for potential time-varying 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 the "Medicaid Expansion" dummy variable.[1] 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).

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

D.1 Additional Discussions

D.1.1 Strategic Decision of Pharmacies

Entry and Exit as a strategic decision by pharmacy: Qualitative evidence suggests that entry and exit (i.e., opening and closing of outlets) are the most important strategic decisions for pharmacies, which I capture in my models. Pharmacies compete in granular geographical markets as consumers consider the location when deciding where to shop. Industry reports [2] highlight that the most important factor for consumers is the location proximity, followed by their health insurance acceptance, and then the quality of service received. In Appendix 19, I show that across various types of pharmacies (independent, chain, and, mass merchants), location consistently emerges as the most important factor determining consumers' pharmacy preferences.[3] These qualitative anecdotes suggest that entry/exit is the main consideration along with the demographic characteristics of the market, which aligns with my empirical application.

I also address concerns regarding the extent to which pharmacy sales contribute to prescription drug revenues. Industry reports indicate that 94% of sales from independent pharmacies are derived from prescription medications, while 70-75% of sales from major chain pharmacies originate from prescription drugs. [4]

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

[3]Admittedly, while health insurance and pricing do play roles, location remains the predominant factor of consideration. Based on the anecdotal evidence, I abstract away from decisions on other dimensions - prices, product variety, health insurance in-network/out-of-cost, and qualities. Appendix 20 supports the qualitative evidence, underscoring location as the most important factor when consumers switch pharmacies.

[4]Source: Industry Report and 10-K issued by Rite Aid, Walgreens, and CVS.

D.1.2 Pharmacy Profits and Machine Learning Approach

I provide the discussions for the motivation of the machine learning approach in pharmacy entry/exit games.

1. Limited Understanding of Market Characteristics: There is little knowledge of which market characteristics are relevant to the opening of pharmacies for econometricians. Instead of relying on an ad-hoc selection process for market covariates, I adopt a data-driven approach, incorporating a rich set of covariates from socioeconomic and health-related characteristics. This allows the model to identify which characteristics are pertinent to the underlying payoffs of pharmacies. Given the high multicollinearity of market covariates, machine learning methods with regularization are well-suited for my application.

2. Unknown functional form: The true functional form of payoffs of independent pharmacies is also unknown to empirical researchers. I relax the commonly used assumption that the profit function of a pharmacy is a linear function of observable covariates. In the first stage, I employ fully non-parametric machine learning methods to enhance prediction power over beliefs about rivals' conditional choice probabilities. In the second stage, I utilize flexible functional forms, like interaction terms, to better capture the underlying payoffs.

D.2 Estimation Details for Neyman Orthogonal Estimators

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

$$\hat{\gamma}_{-imtk} = E[a_{-imt}|c_{mt}, s_{-imt}, s_{imt}, s^{pool}_{xmt}, y_t, \text{county}_f] \tag{D.2.1}$$

where I use various modern machine learners such as linear-based Machine learning (Lasso, Ridge, and Elastic net), Random Forests Classifiers, and Xg Boosting Classifiers. For β, I use Logistic Lasso following Belloni et al. (2016). To learn μ, I use Lasso based on the equations D.2.3. Note that the estimators of nuisance parameters are required to have convergence rates that are faster than $N^{-1/4}$.

Estimation of (β_{ek}, β_{xk})**:** Given $\hat{\gamma}_{-imtk}$ in hand, I estimate the nuisance parameters β_{ek} and β_{xk}. I accommodate flexible interactions between these richer independent variables $s^{interaction}_{xmtk}$. Due to the nature of high-dimensional settings, I construct the following Logit Lasso specification and minimize the function by searching a set of parameters, then keep β_{ek}, β_{xk}.

$$\begin{aligned}&(\hat{\theta}_{\gamma k}, \hat{\theta}_{ck}, \hat{\beta}_{ek}, \hat{\beta}_{xk}, \hat{\alpha}_{tk}, \hat{\alpha}_{\text{countyk}}) \in \\ &\underset{\theta_\gamma,\theta_c,\beta_e,\beta,\alpha_t,\alpha_{\text{county}}}{\arg\min} \left[E_n[\Lambda_i(\theta_\gamma, \theta_c, \beta_e, \beta_x, \alpha_t, \alpha_{\text{county})}] + \frac{\lambda_1}{n} ||(\theta_\gamma, \theta_c, \beta_e, \beta_x)||_1 \right]\end{aligned} \tag{D.2.2}$$

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

Estimation of $\mu_{\gamma k}, \mu_{ck}$**:** The parameters $(\mu_{\gamma k}, \mu_{ck})$ are additional nuisance parameters, not present in the original method. They are introduced in the process of constructing the moment function to ensure the orthogonality property. Define $f_i \equiv \sqrt{\Lambda(x)(1-\Lambda(x))}$, where $\Lambda(x)$ denotes the choice probabilities of being active, as induced by the original moment function. The bias correction term is derived from the linear projection of $z_{imt} = (c_{mt}, s_{-imt})$ on $x_{imt} = (s_{imt}, s_{xmt})$, applying Lasso regression with ℓ_1 penalties determined through cross-validation:

$$f_i z_{imt} = f_i x'_{imt}\mu + u_{imt}, \quad E[f_i x_{imt} u_{imt}] = 0 \tag{D.2.3}$$

[5]As I want to keep the year fixed effects and county fixed effects, I do not allow penalty terms for these two fixed effects

Second Stage Cluster Standard Error of Structural Parameters 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.

Bibliography

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 and Peter Reiss, "Empirical models of entry and market structure," *Handbook of industrial organization*, 2007, *3*, 1845–1886.

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.

Buchmueller, Thomas C, Mireille Jacobson, and Cheryl Wold, "How far to the hospital?: The effect of hospital closures on access to care," *Journal of health economics*, 2006, *25* (4), 740–761.

Cabral, Luıs MB, "Horizontal mergers with free-entry: why cost efficiencies may be a weak defense and asset sales a poor remedy," *International Journal of Industrial Organization*, 2003, *21* (5), 607–623.

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

Caoui, El Hadi, Brett Hollenbeck, and Matthew Osborne, "The impact of dollar store expansion on local market structure and food access," *Available at SSRN 4163102*, 2022.

Chen, Chia-Wen, "Estimating the foreclosure effect of exclusive dealing: Evidence from the entry of specialty beer producers," *International Journal of Industrial Organization*, 2014, *37*, 47–64.

Chenarides, Lauren, Clare Cho, Rodolfo M Nayga Jr, and Michael R Thomsen, "Dollar stores and food deserts," *Applied Geography*, 2021, *134*, 102497.

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.

Chiang, Harold D, Kengo Kato, Yukun Ma, and Yuya Sasaki, "Multiway cluster robust double/debiased machine learning," *Journal of Business & Economic Statistics*, 2022, *40* (3), 1046–1056.

Collard-Wexler, Allan, "Mergers and sunk costs: An application to the ready-mix concrete industry," *American Economic Journal: Microeconomics*, 2014, *6* (4), 407–447.

Davidson, Carl and Arijit Mukherjee, "Horizontal mergers with free entry," *International Journal of Industrial Organization*, 2007, *25* (1), 157–172.

de Chaisemartin, Clément 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 Paul LE Grieco, "Wal-Mart and the geography of grocery retailing," *Journal of Urban Economics*, 2013, *75*, 1–14.

_ **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.

Gaynor, Martin, "Health care industry consolidation," *Statement before the Committee on Ways and Means Health Subcommittee US House of Representatives*, 2011.

_ **, Kate Ho, and Robert J Town**, "The industrial organization of health-care markets," *Journal of Economic Literature*, 2015, *53* (2), 235–284.

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.

Haas-Wilson, Deborah and Christopher Garmon, "Hospital mergers and competitive effects: Two retrospective analyses," *International Journal of the Economics of Business*, 2011, *18* (1), 17–32.

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.

_ **and Nathan Yang**, "Unobserved heterogeneity in dynamic games: Cannibalization and preemptive entry of hamburger chains in Canada," *Quantitative Economics*, 2016, *7* (2), 483–521.

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

Kim, Hyungjin, "Rural Pharmacy Access and Competition: Static Games with Machine Learning," February 2023. Available at SSRN 4377695.

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

Lopez, Rigoberto, Keenan Marchesi, and Sandro Steinbach, "Dollar Store Expansion, Food Retail Competition, and Rural Employment," in "Annual Meeting, January," Vol. 6 2023, p. 8.

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.

Mansley, Ryan, Nathan H Miller, Gloria Sheu, and Matthew C Weinberg, "A Price Leadership Model for Merger Analysis," *International Journal of Industrial Organization*, 2023, p. 102975.

Marino, Anthony M and Ján Zábojník, "Merger, ease of entry and entry deterrence in a dynamic model," *The Journal of Industrial Economics*, 2006, *54* (3), 397–423.

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.

Miller, Nathan H and Matthew C Weinberg, "Understanding the price effects of the MillerCoors joint venture," *Econometrica*, 2017, *85* (6), 1763–1791.

_ **, Marc Remer, Conor Ryan, and Gloria Sheu**, "Upward pricing pressure as a predictor of merger price effects," *International Journal of Industrial Organization*, 2017, *52*, 216–247.

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.

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.

Tenn, Steven, "The price effects of hospital mergers: a case study of the Sutter–Summit transaction," *International Journal of the Economics of Business*, 2011, *18* (1), 65–82.

Thompson, Aileen, "The effect of hospital mergers on inpatient prices: a case study of the New Hanover–Cape Fear transaction," *International Journal of the Economics of Business*, 2011, *18* (1), 91–101.

Weinberg, Matthew, "The price effects of horizontal mergers," *Journal of competition Law and Economics*, 2008, *4* (2), 433–447.

Werden, Gregory J and Luke M Froeb, "The entry-inducing effects of horizontal mergers: An exploratory analysis," *The Journal of Industrial Economics*, 1998, *46* (4), 525–543.

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.

Young, Eric H, Samantha Pan, Alex G Yap, Kelly R Reveles, and Kajal Bhakta, "Polypharmacy prevalence in older adults seen in United States physician offices from 2009 to 2016," *PLoS One*, 2021, *16* (8), e0255642.