THREE ESSAYS IN APPLIED ECONOMICS

by

AMIT PATEL

B.A., Indiana University, 2010

B.S., Indiana University, 2010

M.A., University of Colorado, 2014

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Department of Economics

2017

This thesis entitled:
THREE ESSAYS IN APPLIED ECONOMICS
written by AMIT PATEL
has been approved for the Department of Economics

Prof. Jonathan Hughes, Chair

Prof. Scott Savage

Date ____________

The final copy of this thesis has been examined by the signatories, and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline.

PATEL, AMIT (Ph.D., Economics)

THREE ESSAYS IN APPLIED ECONOMICS

Thesis directed by Prof. Jonathan Hughes, Chair

This dissertation examines three topics in applied economics. In the first chapter, I examine the effects of market structure on ticket pricing in the professional sports industry. The professional sports industry is highly visible and generates notable discourse for local policymakers. The abundance of data in this industry provides a unique opportunity to study economic behavior. I use data on professional sports franchises in over 40 different markets to examine how competition affects ticket prices. I estimate a two-stage model to correct for potential endogeneity between prices and the number of firms in local markets. A first stage market structure model provides a correction term included in a second stage price regression. The results demonstrate a strong positive relationship between prices and market concentration. However, this effect is somewhat diminished when firms differentiate themselves, either by type, quality, or brand. Ignoring the endogeneity of market structure leads to biased estimates that understate the impact additional competitors have on price.

The second chapter, written with Bentley Clinton, focuses on how the joint venture with the International Olympic Committee (IOC) benefits the National Hockey League (NHL) through increases in popularity. Specifically, we study the impacts of participation in the Olympics on fan attendance at NHL games. While previous literature has examined the costs of participation, the benefits remain largely unstudied. We develop a censored difference-in-differences model that provides evidence that NHL participation in the Olympic Games lead to an increase in league-wide attendance of approximately 4.2%, equating to approximately 673 additional tickets sold per team per game. Furthermore, the boost in attendance is larger for teams with excess capacity. These results have implications on future decisions for the NHL to continue to work with the IOC, as well as, the most recent decision not to participate in the 2018 Winter Olympics in Pyeongchang,

South Korea.

In the final chapter of the dissertation, I investigate the price premium households are willing to pay for the ability to walk to nearby amenities. Locational characteristics (parks, schools, shopping, restaurants, etc...) are an important factor in the consumer's decision to purchase a particular home. A major emphasis has been placed on the walkability of the neighborhood and how close residents live to shopping and social activities. So much so, that almost all major real estate websites include a measure of the walkability of an address. On the whole, households are willing to pay more to live in more walkable neighborhoods. This price premium is largest at the top end of the Walk Score distribution. However, these effects diminish with the geographic size of the fixed effect suggesting there may exist nuisance effects, such as noise and congestion, in close proximity to these destinations. Understanding households willingness to pay for neighborhood walkability is paramount for local governments, city planners, developers, and policymakers in determining the optimal mixture of residential and commercial properties. They must also consider ways to minmize nuisance effects in order to realize the greatest potential benefits from neighborhood walkability.

Dedication

To my wife, Marcy Patel,

Without whom I do not know where I would be.

Acknowledgements

I would like to thank Jonathan Hughes and Scott Savage, as well as, the rest of my defense committee for all of their feedback, comments, and suggestions. They have given invaluable guidance to me in coursework and research. Also, I thank Patricia Holcomb for her efforts in aiding my progression through the program, making sure I met all of the necessary deadlines.

I also want to thank my family for all their love and support. Thank you to my son, Charlie, for motivating me to finish my dissertation as soon as possible. Finally, I would especially like to thank my wife, Marcy. What a crazy adventure this has been, with all the ups and downs. Thank you for always sticking by my side.

Contents

Chapter 1

Pricing and Market Structure in Professional Sports

1.1 Introduction

In January 2016, Stan Kroenke, owner of the Rams, announced that he would be moving his football franchise to Los Angeles. The announcement sent shockwaves throughout the sports world. The move would bring a professional football team back to the nation's second largest media market after an absence of over 20 years and as Los Angeles mayor, Eric Garcetti, claimed cement the city "as the epicenter of the sports world". Despite the significant enthusiasm from Angelinos and anger and disappointment from the people of St. Louis (the Rams now former residence), several questions remained, particularly from an economic standpoint. Will this move have a significant impact on the local economy or the quality of life of the areas' inhabitants? Empirical research suggests sports teams provide little to no external economic benefits, but may increase the quality of life or provide intangible benefits to local residents.[1] Another interesting question is how this move will affect the other franchises in the area? Particularly, will the increase in supply of sporting events lead to a decrease in ticket prices? Is the Los Angeles market large enough to support several franchises? Does it matter that the franchises play different sports or that their seasons may not overlap? This paper attempts to answer these questions by examining how competition among sports franchises, in both homogeneous and product differentiated markets, impacts ticket pricing behavior.

Competition in sports is generally viewed as that on the field, characterized by wins and losses.

[1] See Coates **et al.** (2008), Carlino & Coulson (2004), and Groothuis **et al.** (2004) for more details.

However, these franchises also compete with one another in the traditional economic sense. They compete for sponsorships, media access, but predominantly for consumers. Consumers looking to attend a sporting event have several options available to them. They could choose between different sports and/or different levels of competition.[2] In this study, consumers are assumed to choose between franchises at the highest level of quality and competition in their respective sports. This consists of teams from the five major sports leagues: Major League Baseball (MLB), Major League Soccer (MLS), the National Basketball Association (NBA), the National Football League (NFL), and the National Hockey League (NHL). Consumers derive utility from attending games and this utility is a function of price, among other factors.

In order to maximize profits, teams must optimally price tickets in the face of competition. Gate receipts make up a significant portion of the hundreds of millions of dollars a franchise takes in each year.[3] Sales to the actual sporting event also help to build the fan base. That is why most studies on the demand for sporting events specifically look at the effects on attendance (Krautmann & Berri, 2007; Fort, 2004a). While much of the research in this area attempts to control for competition, very few papers have explicitly attempted to quantify the effects of competition, and none have looked at the impact on ticket pricing from competition within local markets.

Much of the literature on profit maximizing decisions of sports franchises focuses on attendance and ticket pricing for an individual team. Assumed to function as a monopolist, we would expect these teams to price in the elastic region of demand. However, research generally finds that they actually price in the inelastic region of demand (Brook **et al.**, 2006; Coates & Humphreys, 2007; Noll, 1974; Ferguson **et al.**, 1991). Although, for many of these studies the null hypothesis of unit elastic pricing cannot be rejected.

Several alternative theories have been developed to explain these results. The "sportsman" hypothesis presumes owners may be maximizing utility rather than profits. Some have argued that teams set prices lower to generate goodwill from politicians for future policy considerations,

---

[2]Of course, they could also decide on some other form of entertainment.

[3]About 40% in NHL, 30% in MLB and NBA, and 20% in NFL.

specifically new stadium financing (Fort, 2004b). Others suggest that teams are maximizing long run profits and set prices lower today to generate more demand in the future (Ahn **et al.**, 2007). Most seem to believe that franchises also consider revenue from other sources, such as concessions and broadcast rights, and that under certain conditions a profit maximizing firm may set price in the inelastic region of demand (Krautmann & Berri, 2007; Coates & Humphreys, 2007). The majority of empirical work on individual team pricing decisions only considers competition secondarily and usually as a control for demand shifts. A more in depth analysis of competition could help explain some of these findings on pricing behavior.

The literature on competition between sports franchises is relatively sparse. Mills & Rosentraub (2014) look at border crossing data between Buffalo and Toronto and find an increase in vehicles traveling to Buffalo during Sabres home games. They also find variation in border crossings associated with differences in prices and team performance between the Buffalo Sabres and Toronto Maple Leafs, which could be because fans substitute between the two teams. A few studies have utilized the NHL lockout in 2004-05 to identify fan substitution across teams through changes in attendance between different levels of the sport (Winfree & Fort, 2008) and between the NBA and NHL (Rascher **et al.**, 2008). Winfree **et al.** (2004) estimate a negative relationship between attendance and proximity to the next closest team in Major League Baseball, providing evidence of competition for fans within a league. Furthermore, proximity may also impact ticket pricing. Using a Hotelling location framework, Henrickson (2012) finds that teams within a league compete spatially when pricing tickets and that teams whose closest competitors have higher prices will also price higher. Apart from Henrickson (2012), no other paper has considered the relationship between ticket prices and market structure, and none have analyzed pricing competition across leagues within markets.

The effect of market concentration on competitive outcomes has been studied extensively in the industrial organization literature. Early analysis examined the relationship between profits and market concentration and generally found that a higher concentration was correlated with higher profits. However, these studies were criticized due to measurement problems with calculating economic profits and also the inability to identify the direction of causation between profits and concentration (Demsetz, 1973). More recently, the research has transitioned to examine the relationship between prices and market structure. However, these price concentration regressions have their own econometric difficulties. Primarily, the fact that market structures are not randomly assigned. Firms make their entry decisions based on cost and demand factors, as well as, the entry behavior of potential competitors. A simple regression of price on market structure would yield biased estimates if there are unobserved cost and demand shocks that may influence both. The bias could be either positive or negative. If unobserved costs lead to higher prices and less entry the relationship between the two variables could be overstated. If, instead, unobserved positive demand shocks lead to higher prices and more entry, the regression would understate the relationship between prices and competition. In general, the literature finds that ignoring the endogeneity of the market structure leads to biased estimates.

One solution to the endogeneity issue would be to instrument for market structure. However, as Imbens & Wooldridge (2007) argue, in a non-linear model, such as this one, a control function approach provides more precise estimates of the treatment effect.[4] While this approach relies on stronger distributional assumptions, it is more efficient than the standard instrumental variables method given a correctly specified first stage model. Therefore, this paper uses a control function approach and implements a correction term to account for the possible correlation between market structure and the error term, similar to those used in other selection models (Heckman, 1979). This correction term is derived from a first stage estimation of the equilibrium number of firms in the market. Following the insights of Bresnahan & Reiss (1990) and Berry (1992), I develop a model of firm entry as an outcome of a strategic game between potential competitors. The results yield a correction term included in a second stage regression of price on market structure. A similar approach has been used to study the relationship between price and market structure in several industries including office supply stores (Manuszak & Moul, 2008). Singh & Zhu (2008) examine

[4]Another issue with the instrumental variables method is the difficulty in finding a suitable relevant, excluded variable.

the effects of market concentration on firm outcomes in the car rental industry, while allowing for firm heterogeneity. Mazzeo (2002) and Greenstein & Mazzeo (2006) extend the model to show how differentiation through firms' choice in product-type offerings can weaken competition, while Molnar & Savage (2017) find that competition also affects quality in the market for broadband internet.

In this paper, I analyze the effects of market concentration on firm pricing decisions using the professional sports industry. The industry plays a pivotal role in the economy generating over $28 billion in revenues in 2014, making it an important area of research. The industry presents a unique opportunity to examine the relationship between price and market structure due to the availability of data on each individual firm. Professional sports, in general, provide an abundance of information, with highly publicized decisions and identifiable consequences that can be used to test economic theories and offer insights on economic behavior.

To my knowledge, this paper is the first to analyze the effects of competition on ticket pricing for sporting events within markets. This paper adds to the literature on market concentration by examining the effects of competition in an industry that experiences significant public policy input, including subsidies, tax breaks, and antitrust privileges. The results in this paper not only have implications for pricing strategies, competitive conduct and relocation/expansion policies of leagues, but also in understanding firm behavior in similar type markets. Particularly, in other entertainment markets and markets with considerable government input and regulation such as the markets for loans and insurance. Furthermore, this research is significant in guiding antitrust and public policies, specifically those meant to attract firms. Apart from estimating the impact of market structure on price, this study provides further insights on how firms in general can limit price competition through differentiation in firm quality, as measured by team performance, and in firm brand strength, as measured by capacity rates. Additionally, there is less of a concern with measurements on price, quality, and definitions of market and product space, as there may be with previous studies due to data limitations.

1.2 Empirical Framework

In this section, I present a typical regression model relating prices to market structure. Following which, models of observed market outcomes are introduced with considerations for both homogeneous and differentiated firms. The first stage market structure models are used to construct correction terms included in the second stage price regression. These correction terms are meant to control for any potential endogeneity between prices and number of firms. In what follows, I borrow from the framework developed by Heckman (1979), Berry (1992), Mazzeo (2002), and Manuszak & Moul (2008).

1.2.1 Model of Prices and Competition

This paper examines the pricing behavior of sports franchises in the presence of competition. The typical price regression model is given by

$$p_{im} = X_{im}\beta + f(N_m, \delta) + \epsilon_{im} \tag{1.1}$$

where p_{im} is the log price for firm i in market m. The variable, X_{im}, includes market level characteristics representing cost and demand factors in each market and firm-level attributes that include measures of firm quality and brand or, in this case, fan strength. The function $f(N_m, \delta)$ characterizes the market structure, where N_m represents the number of firms in market m. Finally, ϵ_{im} reflects firm-market specific unobservables that influence price.

In this context, the dependent variable, p_{im}, is the average ticket price to attend a game for a particular sports franchise in each market. The X-variables in the price regression function primarily as controls and include demographics such as population size and income levels. Of particular importance is the function, $f(N_m, \delta)$, which captures the effect of competition in the market. Theory suggests that more competition results in lower prices. This hypothesis is tested by examining the incremental effect of each additional competitor on ticket prices.

One major concern with equation (1.1) is that the number of firms may not be exogenously assigned. Particularly, that there are unobserved demand and cost factors that influence both prices and the number of firms in the market. For instance, unobserved demand shocks could allow for both higher prices and a higher number of teams in the market. Similarly, if some markets have higher unobserved costs then we may see higher prices and fewer firms. The potential correlation between N_m and ϵ_{im} would lead to biased estimates. One solution would be the use of instrumental variables. An alternative, more efficient, solution to the endogeneity issue is to use a two step control function approach. In the first stage, a model of market structure estimates the determinants of market entry. These estimates are then used to derive a correction term that is included in the second stage price regression. In essence, the error term in (1.1) is split into a part that is correlated with the explanatory variables and a part that is not. The correlated term is then estimated explicitly in the price regression. While this approach relies on stronger distributional assumptions, it is more efficient than the standard instrumental variables approach given a correctly specified first stage market structure model.[5]

1.2.2 Empirical Framework with Homogeneous Firms

The empirical approach begins with the assumption that firms are homogeneous. I analyze market structure through a multi-agent game theoretic approach in which firms make discrete choices. Firms participate in a two-stage game, where in the first stage they decide whether or not to enter the market and then decide how to compete on price in the second stage. When making the entry decision, firms consider the profitability in the market, as well as, the actions of potential competitors. The number of firms in each market will depend on demand and cost factors in each location. Latent profits can be inferred by observing market structure, as firms will enter only if it is profitable to do so. Assume that firm i's latent profit function in market m with N_m firms is

[5] Results are similar using a two stage least squares approach, although less precise.

given by

$$\begin{aligned}\Pi(Z_m, N_m, \gamma) =& \pi(Z_m, N_m, \gamma) + u_m \\ =& Z_m\gamma + \alpha_1 \times 1(N_m > 1) + \alpha_2 \times 1(N_m > 2) \\ &+ \alpha_3 \times 1(N_m > 3) + \alpha_4 \times 1(N_m > 4) + u_m\end{aligned} \tag{1.2}$$

where Z_m are market factors that impact profitability such as population and income, the α's capture the impact on profits from each additional entrant in the market, and u_m are unobserved market characteristics. All firms are assumed to be homogeneous earning profits which are declining with the number of firms. The maximum number of firms in a market is capped at five for the purposes of estimation.[6] The market factors that influence firm entry, Z_m, include several variables that also affect prices. In theory, those variables could be identical. In which case, identification of the correction term would rely on the non-linearity of the market structure model. However, in this application these variables are allowed to differ. In particular, the first stage model includes a fixed cost component measured as the average home value in the market, which serves the exclusion restriction. Housing values provide a measure of the cost of land and construction in each market. This is an important criterion for making entry and exit decisions as sports teams need to build a stadium to house their franchise.[7] Any costs associated with stadium construction are assumed to have no effect on short term pricing decisions and is therefore excluded from the second stage price regression. This may not be the case if the team does not own the stadium and instead leases it from the city. However, leases would have to be negotiated each year if we are to believe that the facility costs influence short term pricing behavior. This is rarely, if ever, the case, as teams generally sign leases years, if not, decades in advance.

[6] There are only 3 markets with more than 5 teams: Chicago MSA with 6, San Francisco MSA with 8 and New York MSA with 10. The lack of variation makes it difficult to estimate additional effects of competition.

[7] Stadium costs have increased significantly, partly attributed to increasing land values. The cost of building a new stadium was one of the reasons the Seattle Sonics moved to Oklahoma City.

An equilibrium of this model states that all firms in the market expect nonnegative profits and any additional firm would find entry unprofitable. The probability of observing n number of firms in market m is given by

$$
\begin{aligned}
Pr[N_m = 0] &= Pr[\pi(N_m = 1) + u_m < 0] \\
&= Pr[u_m < -\pi(N_m = 1)] \\
&= F[-\pi(N_m = 1)] \\
Pr[N_m = n] &= Pr[\pi(N_m = n) + u_m > 0, \pi(N_m = n+1) + u_m < 0] \\
&= Pr[-\pi(N_m = n) < u_m < -\pi(N_m = n+1)] \\
&= F[-\pi(N_m = n+1)] - F[-\pi(N_m = n)] \\
Pr[N_m = n_{max}] &= Pr[\pi(N_m = n_{max}) + u_m > 0] \\
&= Pr[u_m > -\pi(N_m = n_{max})] \\
&= Pr[u_m < \pi(N_m = n_{max})] \\
&= F[\pi(N_m = n_{max})]
\end{aligned}
\tag{1.3}
$$

where $F(.)$ is the cumulative distribution function of u_m. The first equality states that a market will have no firms if it is unprofitable for a monopolist to locate there. The second equation states that exactly n firms will exist when each makes nonnegative profits, but that any additional entrant would find it unprofitable to operate in the market. Finally, the last equality suggests the maximum number of firms in the market is observed when each is profitable. Assuming the error term, u_m, to be distributed standard normal, the parameters of the latent payoff function can be estimated using maximum likelihood. The model used is an ordered probit routine with the dependent variable representing the number of firms in the market. By imposing the profitability condition, I can make inferences on the factors that affect a firm's decision to enter the market.

Following the estimation of the parameters of the latent payoff function, the correction term included in the price regression is derived by imposing distributional assumptions on the error terms. Specifically, the correlation between the errors in the entry and price equations is given by

$$\left\{\begin{array}{c}\epsilon_{im}\\ \\ u_m\end{array}\right\} \sim N\left\{\begin{array}{ccc}0, & \sigma_\epsilon^2 & \\ & & \\ 0, & \sigma_{\epsilon u} & 1\end{array}\right\} \tag{1.4}$$

where ϵ_{im} and u_m are the error terms for the price and payoff regressions, respectively, and $\sigma_{\epsilon u}$ is the covariance between the two. Given the distributional assumptions, we can represent the expectation of the error term in the price regression conditional on firm and market characteristics and the number of firms in the market as

$$E[\epsilon_{im}|X_{im}, Z_m, N_m] = \sigma_{\epsilon u} E[u_m|X_{im}, Z_m, N_m] \tag{1.5}$$

Plugging this into the price regression we have

$$p_{im} = X_{im}\beta + f(N_m, \delta) + \sigma_{\epsilon u} E[u_m|Z_m, N_m] + v_{im} \tag{1.6}$$

where $v_{im} = \epsilon_{im} - \sigma_{\epsilon u} E[u_m|Z_m, N_m]$ is the uncorrelated idiosyncratic error term assumed to impact price. The specific functional form of the price regression to be estimated is given by

$$\begin{aligned} p_{im} =& X_{im}\beta + \delta_1 \times 1(N > 1) + \delta_2 \times 1(N > 2) + \delta_3 \times 1(N > 3) \\ &+ \delta_4 \times 1(N > 4) + \sigma_{\epsilon u} E[u_m|Z_m, N_m] + v_{im} \end{aligned} \tag{1.7}$$

so that δ_n measures the incremental effect of the n^{th} competitor on price.

$E[u_m|Z_m, N_m]$ is derived from the first stage market structure model and $\sigma_{\epsilon u}$ becomes an additional parameter to be estimated in the price regression. The correction term is computed using the following equation[8]

$$E[u_m|Z_m, N_m] = \begin{cases} \frac{\phi[\pi(Z_m,n,\gamma)] - \phi[\pi(Z_m,n+1,\gamma)]}{\Phi[\pi(Z_m,n,\gamma)] - \Phi[\pi(Z_m,n+1,\gamma)]} & \text{if } 0 < n < n_{max} \\ \\ \frac{\phi[\pi(Z_m,n_{max},\gamma)]}{\Phi[\pi(Z_m,n_{max},\gamma)]} & \text{if } n = n_{max} \end{cases}$$

The correction term considers the possibility that $E[\epsilon_{im}|N_m] \neq 0$, which would lead to biased estimates in a simple naive price regression.

[8]See Appendix for full derivation of estimated equations and correction terms.

1.2.3 Empirical Framework with Heterogeneous Firms

In this specification, firms are differentiated based on their product offerings. In particular, firms are categorized by the type of product. In the sports industry, teams and leagues differentiate themselves by providing different types and levels of athletic competition. One can certainly make the argument that the NFL differentiates itself from the other major sports leagues by offering a more distinguished product. The NFL generates the most interest, the most revenue, the highest television ratings, and has the highest attendance as a percent of stadium capacity.[9] For these reasons, I make a distinction between firms offering a football product and all other firms. To the extent that consumers view these two types as imperfect substitutes, product differentiation will significantly affect entry and price decisions.

As was the case with homogeneous firms, here we assume that profits are declining with the number of firms. However, with heterogeneity, we expect a smaller decrease in profits and prices if firms are differentiated. The profit functions for firms offering football(F) and other(O) product types are given by

$$\begin{aligned}\pi_O &= Z_m\beta_O + g(\theta_O, N_m) + \epsilon_O \\ &= Z_m\beta_O + \theta_{OO1} \times 1(O > 1) + \theta_{OO2} \times 1(O > 2) + \theta_{OO3} \times 1(O > 3) \\ &\quad + \theta_{OF1} \times 1(F > 0) + \theta_{OF2} \times 1(F > 1) + \epsilon_O\end{aligned} \tag{1.8}$$

$$\begin{aligned}\pi_F &= Z_m\beta_F + g(\theta_F, N_m) + \epsilon_F \\ &= Z_m\beta_F + \theta_{FF1} \times 1(F > 1) + \theta_{FO1} \times 1(O > 0) + \theta_{FO2} \times 1(O > 1) + \epsilon_F\end{aligned}$$

where Z_m are market characteristics that affect profitability and the function $g(\theta, N_m)$ captures the effects of competition on profits for each product type. In theory, the market variables affecting profitability for each type could differ. For the purposes of this paper the same variables are included, although their effects are allowed to differ between types. The dummy variables represent

[9] In the previous season, capacity utilization in the NFL was 97%, while in all other leagues the average was 84%.

the incremental effects on profits from the presence of both similar and different type competitors.[10] For the purposes of estimation, the number of low-type firms in the market is capped at three, while for the high-type, only a maximum of two firms is ever observed in the data. The effect from the alternative product firms is limited to the first two. Lastly, the error terms represent market-type specific unobservables that may influence profits.

Each firm makes a decision of whether or not to enter the market and what product type to offer. Within a given market, firms of the same type earn the same profits, while profits across types may differ. In determining whether or not to enter the market, firms play a Stackelberg game.[11] In each round, the highest profit type firm makes an entry decision taking into consideration the optimal choice of subsequent potential entrants. The last firm of each type will enter as long as it is profitable to do so. An equilibrium of this model is represented by

$$\begin{aligned}
&\pi_O(O,F) > 0, \pi_O(O+1,F) < 0 \\
&\pi_O(O,F) > \pi_F(O-1,F+1) \\
&\pi_F(O,F) > 0, \pi_F(O,F+1) < 0 \\
&\pi_F(O,F) > \pi_O(O+1,F-1)
\end{aligned} \tag{1.9}$$

where (O,F) denotes the number of other and football type firms in the market. The first and the third set of inequalities represent the conditions under which no additional firms would find entry profitable for either type. The second and fourth inequalities, represent the no switching conditions, and state that no firm would want to switch product type offerings.

Assuming a distribution for each error term, the number of football and other type firms operating in a market can be estimated. Furthermore, by allowing for a nonzero correlation between the error terms in the price and profit equations, a correction term for each type can be derived to

[10] The models are as flexible as possible, while allowing for feasible estimation.

[11] Tamer (2003) shows that a simultaneous move game with two types leads to multiple equilibria.

control for the endogeneity in the price regression. Assume the errors are distributed as

$$\left\{\begin{matrix} \epsilon_{im} \\ \epsilon_O \\ \epsilon_F \end{matrix}\right\} \sim N \left\{\begin{matrix} 0 & \sigma_m^2 & & \\ 0 & \sigma_{mO} & 1 & \\ 0, & \sigma_{mF} & 0 & 1 \end{matrix}\right\} \tag{1.10}$$

where σ_{mO} represents the covariance between the error term in the price regression and the other type profit error and σ_{mF} represents the covariance between the error term in the price regression and the football type profit error. Additionally, the two profit error terms are assumed to be uncorrelated.[12] Given the distributional assumptions, we can represent the expectation of the error term in the price regression conditional on firm and market characteristics and the number of firms in the market as

$$E[\epsilon_{im}|X_{im}, Z_m, N_m] = \sigma_{mO}E[\epsilon_O|X_{im}, Z_m, N_m] + \sigma_{mF}E[\epsilon_F|X_{im}, Z_m, N_m] \tag{1.11}$$

Plugging this into the price regression we have

$$p_{im} = X_{im}\beta + f(N_m, \delta) + \sigma_{mO}E[\epsilon_O|X_{im}, Z_m, N_m] + \sigma_{mF}E[\epsilon_F|X_{im}, Z_m, N_m] + e_{im} \tag{1.12}$$

where the covariances are additional parameters to be estimated. More importantly, the new error term e_{im} is mean zero and uncorrelated with the regressors. The effects of competition on price can now be estimated without bias. The specific functional form of the price regression to be estimated is given by

$$\begin{aligned} p_{im} =& X_{im}\beta + \Delta_{S1} \times 1(Same > 0) + \Delta_{S2} \times 1(Same > 1) + \Delta_{D1} \times 1(Diff > 0) \\ &+ \Delta_{D2} \times 1(Diff > 1) + \sigma_{mO}E[\epsilon_O|X_{im}, Z_m, N_m] + \sigma_{mF}E[\epsilon_F|X_{im}, Z_m, N_m] + e_{im} \end{aligned} \tag{1.13}$$

so that Δ_S measures the additional effect of each similar product type competitor on price and Δ_D measures the additional effect of each different product type in the market.

1.2.4 Data

To estimate the second stage price regressions, I use a cross-section of data on the pricing decisions and characteristics of franchises from the five major sports leagues in the United States:

[12] The results in table 6 provide evidence that the error terms are indeed uncorrelated

Major League Baseball (MLB), National Basketball Association (NBA), National Football League (NFL), National Hockey League (NHL), and Major League Soccer (MLS). Alternatively, a panel data approach could potentially identify the relationship between market structure and prices. In which case, market fixed effects could be used to address the endogeneity issue by accounting for any unobserved time invariant factors that influence both prices and market structure. However, such an approach would require actual variation in market structure to identify parameters, which is not the case in this instance. Since 2010, only 6 such markets have experienced entry or exit of firms. Furthermore, as Manuszak & Moul (2008) point out, this approach still requires that these unobservable factors be relatively short run so as not to affect the entry and exit decisions of firms. With market fixed effects, it could still be possible that changes in the error term are correlated with changes in market structure. For these reasons, I contend that the cross sectional approach is better suited for this data and question and yields consistent estimates, provided the first stage market structure model is correctly specified.

Ticket pricing data for each team in the MLB, NBA, NFL, and NHL come from Team Marketing Report, which provides data on pricing, sponsorship, and marketing deals. For the MLS, ticket pricing data come from razorgator.com, a ticket broker. All ticket pricing data represent a weighted average of prices and seats available at the beginning of the 2014 season. Ticket sales make up a significant portion of total revenue for each team in each league, and therefore, serve as the strong measure of competitive conduct. Figure 1.1 plots ticket prices for teams in all leagues by market structure, while Figure 1.2 breaks out ticket prices by league. Apart from the MLS, it seems ticket prices increase with the number of competitors. This is counter-intuitive to the belief that more competition would lead to lower prices and suggests a concern for the endogeneity of market structure. This point is further illustrated in Table 1.1, which provides a summary of the observed market structure in 2014, along with the average ticket price for each type of market split out by league. This table also includes data on population and average household income. The number of teams in a market increases with population and income, suggesting that market structure may not be randomly assigned, but rather, depend on demand and cost factors, some of which could be unobserved. Table 1.2 presents the average ticket price by market structure and product type. Interestingly, the prices for the other-type tend to increase with like competition and decrease with the presence of football-type competitors, while prices for the football-type tend to increase with competition no matter the type.

In addition to ticket prices, I also collect team-specific data. This data is used to create measures of firm quality and firm brand strength. While firm quality may be difficult to quantify in many industries this is not the case in sports. A team's quality can be easily measured by its performance on the field. Team performance has been shown to significantly influence ticket sales. However, it can be difficult to compare performance across sports leagues, since some leagues use a points system while others use winning percentage. Furthermore, the leagues may differ on the difficulty of winning, likely due to differences in the dispersion of talent and rules governing the game.[13] Instead, a more useful measure of the relative performance or quality of a franchise is whether or not the team made the playoffs in its respective league. For these reasons, I include a dummy variable for playoff appearance in the previous season to examine the effect of firm quality on price in the face of competition. The second important firm specific variable is a measure of brand strength. For sports teams this is a gauge of fan strength or fan loyalty. I use an extension of measure from Henrickson (2012) and quantify this as the average attendance as a percentage of capacity for the previous 3 seasons. This capacity utilization measure captures fan's preferences for the product beyond that explained by team performance. Despite a poor quality of the product on the field, some teams continue to maintain higher levels of demand because of their brand, and this measure attempts to capture this effect. For example, the Chicago Cubs, maintained a high attendance rate despite nearly a century of futility. This was largely due to the strength of their brand and the loyalty they created with consumers. Data on team performance and attendance comes from sports-reference.com, while data on stadium capacity is collected from ballparks.com.

[13] There exists a substantial amount of literature on differences in competitive balance across leagues. For example see Fort & Maxcy (2003).

Figure 1.1: Ticket Prices by # of Teams in Market

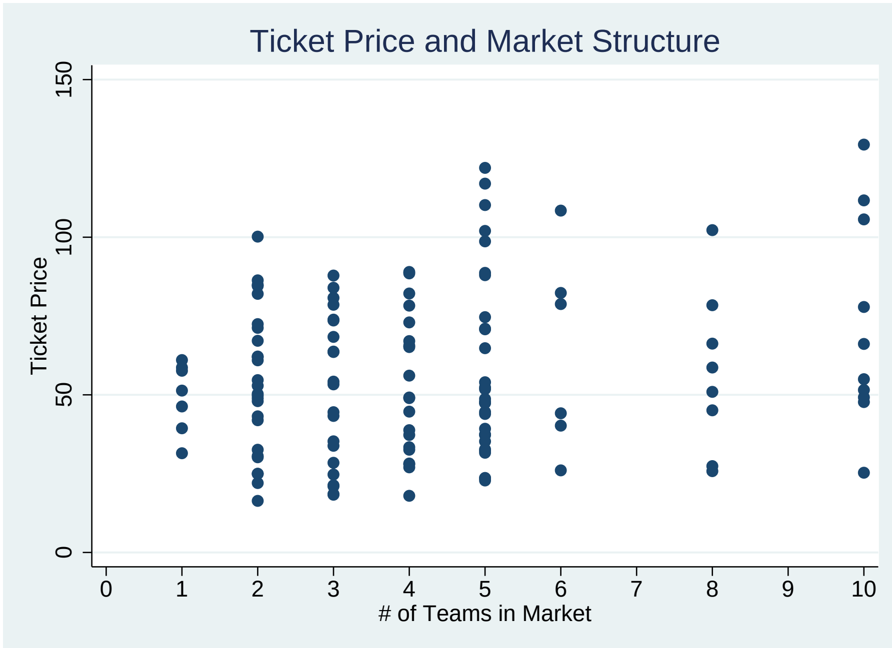


Figure 1.2: Ticket Prices in Each League by Market Structure

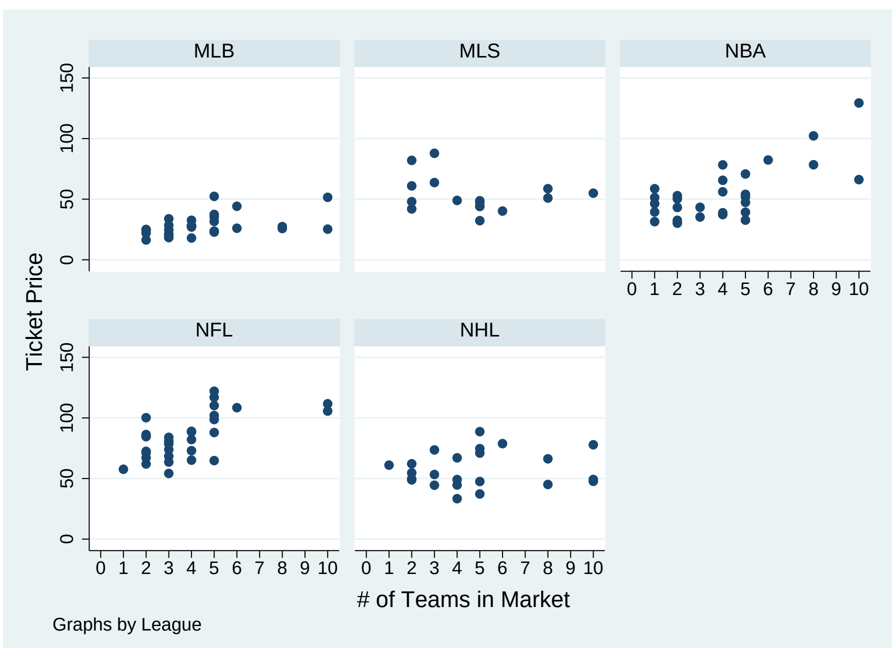


Table 1.1: Average Ticket Prices by Market Structure

Market Structure	N	MLB	MLS	NBA	NFL	NHL	Pop	Avg Inc
No firms	219	-	-	-	-	-	421,462	40,730
Monopoly	7	-	-	45.44	57.65	61.04	1,748,081	43,427
Duopoly	13	22.08	58.24	39.96	78.60	53.87	1,980,957	48,460
3 Firms	7	23.74	75.78	39.28	71.90	57.14	3,071,184	47,570
4 Firms	5	26.76	48.97	55.21	79.56	48.57	4,940,179	48,111
5+ Firms	9	32.38	46.90	68.59	102.84	62.19	8,225,491	57,884

Table 1.2: Average Ticket Price by Each Product Type Market

Product Type Market (O, F)	N	O-Type Price	F-Type Price
(1,0)	6	48.04	-
(0,1)	1	-	57.65
(1,1)	8	33.58	78.60
(2,0)	5	50.79	-
(2,1)	7	40.55	71.91
(3,1)	5	43.54	79.56
(3,2)	1	33.95	90.90
(4+,0)	1	56.85	-
(4+,1)	6	49.15	104.88
(4+,2)	1	62.76	108.69

A relevant market is defined as a metropolitan statistical area (MSA) determined by the Office of Management and Budget.[14] MSA's are economically integrated regions, generally, spanning several counties with a dense population. This definition of a market is typical in the sports literature as it encompasses the vast majority of the population likely to attend games and consume the product. Research has shown the MSA population to significantly affect attendance at sporting events (Brook **et al.**, 2006; Coates & Humphreys, 2007; Rascher **et al.**, 2008). The data consist of 128 teams scattered across 41 different markets[15], with additional markets having no teams included in the first-stage estimation.

The dataset also includes important market characteristics. The full list of these variables can be found in Table 1.3. The first set consists of typical measures aimed at capturing the cost and demand factors in a particular market. The first three, population, income, and housing value are measured in thousands. Data on income and population for 2014 comes from the Bureau of Economic Analysis, while data on housing values comes from the 2014 American Community Survey (ACS). The 2014 ACS also provides data on the percent of population that is Hispanic, percent of population that is White, and the region the MSA is located. Additional MSA level data includes number of large companies, defined as those that have more than 500 employees, from the 2013 County Business Patterns survey. This variable serves as another measure of the local economy. Larger businesses and corporations also tend to purchase many tickets to sporting events in both the general seating and luxury suites, influencing the demand and pricing of tickets.

[14] Canadian teams are excluded from the dataset.

[15] The data excludes the Green Bay Packers, who are located in a unique market in Green Bay, Wisconsin.

Table 1.3: Summary Statistics

Variables	Mean	Std. dev.
Population (1000s)	963.530	1,928.936
Income (1000s)	42.109	9.029
Housing Value (1000s)	232.917	123.203
Large Companies	811.785	743.343
% White	80.993	11.29
% Hispanic	15.367	17.722
Northeast	0.154	
Midwest	0.227	
South	0.385	
Playoffs	0.484	
% Att\Cap	86.290	16.535

1.3 Results

This section presents the results from the estimation of the models discussed in the previous section. First, I discuss the market structure and price regressions under the assumption of homogeneous firms. This is followed by the results of the model when the assumption is redefined to allow for firm heterogeneity.

1.3.1 Model Estimates with Homogeneous Firms

The market structure model is estimated using an ordered probit routine.[16] Table 1.4 presents the results from the maximum likelihood estimation of the first stage model. Most of the parameter estimates are reasonable and have the predicted sign. Based on the estimates, the number of firms is increasing in market size as reflected by population and the number of large companies. Markets with higher levels of income are also associated with higher profits. There is weak evidence that a larger population of Whites and a smaller population of Hispanics support more teams. The parameter estimate for housing values suggest land prices significantly impact a firm's entry

[16]The variables population, income, housing values, and large companies are scaled by: $Z_m^* = ln[{}^{Z_m}/_{\bar{Z}}]$, to facilitate estimation. A value equal to the mean is transformed to 0.

decision.[17] Regional dummies capture any additional differences in demand and cost factors for different parts of the country. When controlling for market size and demographics, the parameter estimates on these dummies suggest there is no significant difference in the likelihood of entry due to regional differences. The coefficients on competition suggest that the number of rivals in the market significantly impact profitability, with the largest decrease coming from the entry of the third firm. After which, the effect of additional entrants on profits declines.

Table 1.4: Market Structure Model for Homogeneous Firms

Variable	Estimate	Standard Error
Constant	4.320***	1.059
Population	1.503	1.087
Income	6.557***	2.030
Housing Value	-2.310**	1.110
Large Companies	4.464**	2.050
Hispanic	-0.401*	0.234
White	0.456	1.435
Northeast	-0.809	0.845
Midwest	-0.876	0.729
South	-1.209*	0.663
$\alpha_1(N > 1)$	-0.720***	0.260
$\alpha_2(N > 2)$	-2.025***	0.402
$\alpha_3(N > 3)$	-1.421***	0.409
$\alpha_4(N > 4)$	-1.043***	0.386
Log-Likelihood	-510.44	
Observations	260	

Note: Variables have been standardized to facilitate estimation

Table 1.5 displays the results from the estimation of the price regression given in equation (1.7). The first column provides the results from the uncorrected regression, while the second column is estimated with the inclusion of the correction term. The regressions include several market specific variables, as well as, team specific variables of quality and brand strength. In both regressions, population and income positively influence price, as does the percentage of the population that is white. Surprisingly, the number of large companies in the market, conditional on

[17]The equivalent F-statistic for this model is 11.08.

population and income levels, has no effect on price. More importantly, the coefficients on playoffs and attendance percentage provide evidence that team quality and brand strength significantly influence demand for the product. Teams that made the playoffs in the previous season increase their ticket prices by roughly 18% compared to non-playoff teams. Beyond simply performance, the degree of fans' preferences or loyalty to the team also leads to an increase in the price of admission. The results suggest that beyond the market characteristics, individual teams can differentiate themselves based on the quality and strength of their product and this can help mitigate the effects of competition.

The naive regression suggests that the first three competitors impact prices negatively, but that the presence of a fourth competitor may actually lead to an increase in price. In fact, the parameter estimates indicate that the first three competitors have a similar negative effect on price of about 17-18%. However, when controlling for the endogeneity in the market, the results indicate a much stronger competitive effect on price for the first two additional firms and a slightly smaller effect from the presence of a 3rd competitor. The first competitor leads to a decrease in price of 26%. The presence of a second competitor leads to an additional 19% decrease in price, while a third competitor in the market further decreases price by 15%.[18] The positive effect from the fourth competitor also becomes negative, although this is not statistically significant. The estimated coefficient on the control variable is positive and statistically significant at the 5% level. The sign on the control variable is consistent with the idea that there are unobserved demand factors that lead to both higher than expected ticket prices and a higher than expected number of teams in the market.

[18]Tests of the coefficients reveal that the parameter estimates for the first competitor are statistically different in both regressions. They also reveal that the estimate of the first competitor is statistically different from the estimate of the third competitor in the corrected regression.

Table 1.5: Price Regression for Homogenous Firms

	Uncorrected	Corrected
Constant	2.7596***	2.2603***
	(0.3884)	(0.6752)
Population	0.0208*	0.0063
	(0.0115)	(0.0567)
Income	0.0061*	0.0109***
	(0.0034)	(0.0004)
Large Companies	-0.0000	0.0001
	(0.0001)	(0.0003)
% White	0.0067*	0.0084***
	(0.0034)	(0.0005)
% Hispanic	0.0008	-0.0006
	(0.0031)	(0.0015)
Northeast	0.0822	0.0402
	(0.0863)	(0.1647)
Midwest	0.0188	-0.0225
	(0.0860)	(0.1286)
South	0.0111	-0.0202
	(0.0594)	(0.0456)
Playoffs	0.1712**	0.1827***
	(0.0703)	(0.0097)
% Att\Cap	0.0012***	0.0012***
	(0.0001)	(0.0002)
$\delta_1(N > 1)$	-0.1839*	-0.2585***
	(0.1054)	(0.0651)
$\delta_2(N > 2)$	-0.1753*	-0.1948***
	(0.0985)	(0.0017)
$\delta_3(N > 3)$	-0.1707**	-0.1469**
	(0.0690)	(0.0592)
$\delta_4(N > 4)$	0.0163	-0.0346
	(0.0715)	(0.1489)
$\sigma_{\epsilon u}$		0.0766**
		(0.0381)
Adj r^2	0.2464	0.2469
Observations	128	128

Note: Bootstrapped clustered standard errors at the MSA level in parenthesis

1.3.2 Model Estimates for Heterogeneous Firms

This section evaluates the effects of competition when firms are able to differentiate themselves through their choice in offering a specific type of product. The two different product types are football (F) and all other sports (O). Table 1.6 provides the results from a bivariate ordered probit regression. Note, the correlation coefficient, ρ, is not statistically significant. The hypothesis that $\rho = 0$ cannot be rejected, suggesting it is not necessary to run a bivariate ordered probit. Instead, separate ordered probit routines are utilized for each product type outcome and these are used to calculate the correction terms. These estimates can be found in Tables 1.7 and 1.8. The results are similar to those in Table 1.6.

The estimates reveal some similarities and some differences in profitability across both types. Income has a positive and significant effect on profits of both types and the coefficient on housing suggests less entry in markets where land is more expensive. The number of large companies is only significant for the other-types, while the percent of Hispanics in the market only negatively impacts the presence of football-type firms. The constant terms suggest that football teams are more likely to enter the market before any other sports team, all else equal. Among both types, the effects on profits from same-type competitors is much larger than from different type competitors. For other-type firms, the first three similar competitors have a significant negative impact on profits, while the first two different type firms have a small, though not statistically significant, effect. Similar to the homogeneous firm model, the largest decrease in profits comes from presence of the second same-type competitor. For the football-type firm the effect on payoffs due to the presence of another firm is more than double if that competitor is a football-type versus an other-type. However, the first other-type competitor does have a significant effect on the likelihood of entry by football-type firms, whereas the opposite effect is not as strong.

Table 1.9 presents the results from the price regressions with heterogeneous firms. The first column provides the uncorrected estimates, while the second displays the corrected estimates with the inclusion of correction terms derived from the first stage ordered probit regressions for each firm

Table 1.6: Market Structure Model for Product Type Markets

Variable	Estimate	Standard Error
Other Type		
Constant	3.761***	0.0.860
Population	0.900	1.088
Income	5.673***	2.029
Housing Value	-2.161**	1.099
Large Companies	4.314**	1.806
% Hispanic	-0.0635	0.230
Northeast	-0.657	0.838
Midwest	-0.714	0.774
South	-1.257*	0.654
Football Type		
Constant	3.986***	0.953
Population	1.026	1.065
Income	5.926***	1.587
Housing Value	-1.910**	0.919
Large Companies	2.502	1.665
% Hispanic	-0.780***	0.194
% White	0.604	1.370
Northeast	-0.607	0.774
Midwest	-0.975	0.724
South	-0.367	0.641
ρ	0.0912	0.203
Log-Likelihood	-740.06	
Observations	260	

Note: Variables have been standardized to facilitate estimation

Table 1.7: Market Structure Model for Other Product Type Markets

Variable	Estimate	Standard Error
Constant	3.691***	0.914
Population	0.795	1.052
Income	5.580***	2.132
Housing Value	-2.051*	1.112
Large Companies	4.277**	1.857
% Hispanic	-0.0205	0.236
% White	-0.281	1.597
Northeast	-0.676	0.844
Midwest	-0.707	0.779
South	-1.308**	0.658
$\theta_{OO1}(O > 1)$	-1.502***	0.324
$\theta_{OO2}(O > 2)$	-1.827***	0.384
$\theta_{OO3}(O > 3)$	-1.211***	0.381
$\theta_{OF1}(F > 0)$	-0.231	0.476
$\theta_{OF2}(F > 1)$	-0.668	0.575
Log-Likelihood	-470.54	
Observations	260	

Note: Variables have been standardized to facilitate estimation

Table 1.8: Market Structure Model for Football Product Type Markets

Variable	Estimate	Standard Error
Constant	4.732***	0.974
Population	1.211	1.172
Income	6.458***	2.025
Housing Value	-1.724*	1.031
Large Companies	2.084	1.879
% Hispanic	-0.885***	0.221
% White	1.257	1.628
Northeast	-0.548	0.917
Midwest	-0.896	0.806
South	-0.215	0.786
$\theta_{FF1}(F > 1)$	-3.617***	0.679
$\theta_{FO1}(O > 0)$	-1.514**	0.742
$\theta_{FO2}(O > 1)$	-0.826	0.536
Log-Likelihood	-230.25	
Observations	260	

Note: Variables have been standardized to facilitate estimation type. The coefficients on the market level characteristics are similar to those in the regression with homogeneous firms. While population does not have a significant effect on price, the number of establishments does. An additional 100 large sized companies leads to a 3% increase in price. Each thousand dollar increase in income levels leads to an increase in price of about 1.2%. Race also affects ticket prices, with the percent of whites in the market leading to an increase. Unlike the case with homogeneous firms, location has an impact on price for heterogeneous firms. The coefficients on the northeast and south regions suggest higher prices in those areas. In the northeast, this is most likely due to higher average costs relative to the west, while in the south this is most likely due to differences in demand ,relative to the west, not accounted for by the other variables. This may be the case if we believe western states have lower demand due to the availability of other forms of entertainment. The firm specific variables again illustrate that both firm quality and firm brand strength or fan loyalty have significant positive effects on price, all else equal. The estimates for these two variables are similar to those in the homogeneous price regression, and again, suggest that firms can limit competition by further differentiating themselves along these lines.

For the uncorrected price regression, the coefficients on competitors imply a much larger negative effect from the first differentiated firm than from the first same-type competitor and only the coefficient on the differentiated firm is significant. These effects are reversed when controlling for potential endogeneity. In fact, the magnitude on the same type competitors increases, while it decreases for the different types. The first similar competitor decreases price by about 31%, while the presence of the first different-type firm decreases price by roughly 26%. This is in line with the hypothesis that competition is lessened when firms are differentiated. Additional firms of either type do not seem to affect price much at all and are not significantly different from zero. Finally, the coefficients on the correction terms suggest a positive correlation between unobservables that affect price and market structure. Unobserved demand and cost factors support higher than normal prices in markets with higher than expected number of firms. Disregarding the endogeneity issue leads to biased estimates and erroneous conclusions about the nature of competition in these markets.

Table 1.9: Price Regression for Heterogenous Firms

	Uncorrected	Corrected
Constant	3.3807***	2.2409***
	(0.8432)	(0.4645)
Population	0.0067	-0.0049
	(0.0114)	(0.0077)
Income	0.0012	0.0118***
	(0.0081)	(0.0020)
Large Companies	0.0001	0.0003**
	(0.0001)	(0.0001)
% White	0.0063	0.0101**
	(0.0049)	(0.0043)
% Hispanic	0.0034	-0.0001
	(0.0039)	(0.0004)
Northeast	0.2165*	0.1091***
	(0.1106)	(0.0038)
Midwest	0.0950	0.0124
	(0.0826)	(0.0872)
South	0.0938	0.0474***
	(0.0793)	(0.0092)
Playoffs	0.1692**	0.1894***
	(0.0656)	(0.0102)
% Att\Cap	0.0009**	0.0008***
	(0.0003)	(0.0001)
1st Same (Δ_{S1})	-0.1863	-0.3133***
	(0.1298)	(0.0216)
2nd Same (Δ_{S2})	0.0204	-0.0221
	(0.1251)	(0.1205)
1st Diff (Δ_{D1})	-0.2863***	-0.2645**
	(0.0960)	(0.1226)
2nd Diff (Δ_{D2})	0.1741	0.0417
	(0.1068)	(0.0811)
σ_{mF}		0.0233
		(0.0273)
σ_{mO}		0.1370**
		(0.0665)
Adj r^2	0.2893	0.2983
Observations	128	128

Note: Bootstrapped clustered standard errors at the MSA level in parenthesis

1.3.2.1 Subsample Analysis

Additionally, we may be interested in the effects of competition on price separately for each type. That is to say, does differentiated competition affect the pricing behavior of football teams differently than all other sports teams? The type-specific price regressions reveal some of these differences. Table 10 presents the results for ticket prices of only the other-type firms. Compared to the grouped regression in Table 9, the first same and first different type firms have a larger negative effect on price. This suggests other-type firms may compete more heavily with one another, lowering price more to avoid losing customers to similar type teams. Table 11 presents the price regression results for football franchises. None of the variables are statistically significant, which is likely due to the small number of observations. The data set only consists of 31 NFL teams. Nevertheless, the parameters on the competitive variables suggest smaller effects on price due the presence of other firms. There is a stronger effect from similar type firms providing some evidence of product differentiation as a means to reduce competition.

The estimates also reveal that these teams face stronger competition from the presence of an NFL team than vice versa. This is as expected since the NFL is considered the dominant league in the United States. Furthermore, the results indicate that other-type firms can mitigate these competitive effects slightly more than football-types by providing a higher quality product or increasing brand strength as revealed by the coefficient estimates on the playoffs dummy and attendance.

The separate regressions also reveal how the demographic variables affect price differently for either type firm. The magnitudes on most of these variables are larger for other-type firms implying they influence pricing decisions for all other sports teams more than for football teams. Specifically, income level has twice the effect on prices for teams in the NBA, NHL, MLB, and MLS compared to the NFL. The number of large companies in a metropolitan area influences the ticket prices for football teams more than all other types. This is as expected, as football teams rely more heavily on revenue from selling tickets to companies.

Table 1.10: Price Regression for Other Type Firms

	Uncorrected	Corrected
Constant	2.3132***	1.5137**
	(0.4971)	(0.6345)
Population	0.1284***	0.0575
	(0.0470)	(0.0692)
Income	0.0121**	0.0202***
	(0.0053)	(0.0062)
Large Companies	-0.0002*	0.0001
	(0.0001)	(0.0002)
% White	0.0127***	0.0131**
	(0.0046)	(0.0060)
% Hispanic	-0.0039	-0.0042
	(0.0039)	(0.0040)
Northeast	0.0950	0.0937
	(0.1410)	(0.1290)
Midwest	-0.0264	-0.0219
	(0.1183)	(0.1044)
South	0.0637	0.0544
	(0.1013)	(0.1162)
Playoffs	0.1928**	0.2115***
	(0.0807)	(0.0364)
% Att\Capt	0.0010**	0.0009**
	(0.0004)	(0.0004)
1st Same (Δ_{S1})	-0.2761**	-0.3695***
	(0.1352)	(0.0643)
2nd Same (Δ_{S2})	0.1123	0.0716
	(0.1412)	(0.1579)
1st Diff (Δ_{D1})	-0.3611***	-0.3306**
	(0.0999)	(0.1300)
2nd Diff (Δ_{D2})	0.1553	0.0704
	(0.1301)	(0.0751)
σ_{mF}		0.0543
		(0.0396)
σ_{mO}		0.1263**
		(0.0642)
Adj r^2	0.3858	0.3876
Observations	97	97

Note: Bootstrapped clustered standard errors at the MSA level in parenthesis

Table 1.11: Price Regression for Football Type Firms

	Uncorrected	Corrected
Constant	2.9598***	3.1253
	(0.5554)	(3.4129)
Population	-0.0097	0.0053
	(0.0175)	(0.2293)
Income	0.0118**	0.0111
	(0.0056)	(0.0205)
Large Companies	0.0002*	0.0002
	(0.0001)	(0.0006)
% White	-0.0010	0.0006
	(0.0049)	(0.0237)
% Hispanic	-0.0003	-0.0018
	(0.0033)	(0.0176)
Northeast	0.0890	-0.0116
	(0.1266)	(0.7712)
Midwest	0.0010	-0.0687
	(0.1235)	(0.7047)
South	-0.0552	-0.0943
	(0.1112)	(0.6585)
Playoffs	0.0395	0.0777
	(0.0947)	(0.2237)
% Att\Cap	0.0006	0.0007
	(0.0006)	(0.0016)
1st Same (Δ_{S1})	-0.2310	-0.2659
	(0.1405)	(1.7647)
1st Diff (Δ_{D1})	0.1408	-0.1600
	(0.0871)	(0.8606)
2nd Diff (Δ_{D2})	-0.1217	-0.1645
	(0.0913)	(0.2510)
σ_{mF}		-0.0995
		(0.5156)
σ_{mO}		0.0770
		(0.2031)
Adj r^2	0.5173	0.5038
Observations	31	31

Note: Bootstrapped clustered standard errors at the MSA level in parenthesis

1.4 Discussion

Overall, these results indicate a simple regression of price on market structure would yield biased estimates, because observed markets may not be exogenously assigned. The uncorrected estimates suggest that the first three competitors have a similar effect on price, biasing downward the effect of the first competitor and overstating the effect of the third competitor when firms are treated as homogeneous. For heterogeneous firms, the simple regression implies the first differentiated firm induces tougher competition than a similar type firm, understating the impact of same-type competition. Conclusions drawn from the naive regressions would be misleading.

The results from the corrected regression provide a better understanding of competition within markets in the professional sports industry. The first two competitors have a significant impact on price and differentiation by product type may somewhat mitigate this effect. For sports teams, the results demonstrate the benefits of differentiation and the impacts of entry and exit of franchises. There is some evidence of teams increasing ticket prices when other franchises leave the market.[19] For local governments, the results demonstrate the benefits of having multiple teams in terms of providing more variety, but also in lower ticket prices. The results can also be useful in providing predictions on market structure and outcomes due to changes in exogenous variables, whether they are long term or short term. For example, a short term positive shock to income levels would lead to higher prices, however, a longer term increase would lead to more entry and potentially lower prices.

The market structure model for homogeneous firms demonstrates the degree of competition between sports franchises. The incremental effect on payoffs is largest from the second competitor suggesting markets are significantly less likely to support three teams versus two teams. Additional competitors have a decreasing effect on profits. A franchise would need approximately 600,000 more residents to offset the impact on profits from the first competitor or an increase in median

[19] For example, both the Seattle Seahawks and the Seattle Mariners increased ticket prices following the Seattle Sonics move to Oklahoma City and the Atlanta Hawks and the Atlanta Braves increased prices the season after the Atlanta Thrashers departed for Winnipeg.

income level of about 5,000 dollars, all else equal.[20] For the market structure for different product types, the results indicate that income significantly influences the presence of both types. However, comparing the coefficients suggests an increase in income is relatively more likely to induce entry of a football franchise.

The methodological approach utilized here provides an extension of previous methods by examining firm differentiation in not just product-type offerings, but also on the basis of quality and brand strength. Both of these effects indicate that franchises can mitigate price competition by improving team performance and building a loyal fan base. A playoff caliber team can significantly diminish the effect the first competitor has on ticket prices and can almost completely offset the additional effect from a second competitor.

1.5 Conclusion

This paper investigates the relationship between prices and market structure in the professional sports industry. I provide strong evidence that sports franchises do in fact compete on price in local markets. The results presented here have significant implications on pricing strategies for teams, relocation and expansion decisions for teams and leagues, and can inform government policies directed toward professional sports franchises, as well as, firms in similar type industries. Additionally, the empirical approach provides insights on the determinants of market structure in the sports industry. There is some evidence that heterogeneous firms consider the presence and product type of competitor firms when making entry decisions.

Similar to previous literature on the effects of competition on price, I present further evidence that ignoring the endogeneity of market structure would yield biased estimates because observed markets may not be exogenous. Any conclusions drawn from a model that ignores the endogeneity issues can be severely misguided. Instead, I adopt a control function approach to correct for the endogeneity between prices and the number of firms in the market. Additionally, my approach allows firms to differentiate themselves through product-type offerings, product quality, and product

[20] Assuming population and income levels were are their sample means.

brand strength. All three mechanisms are pivotal for firms trying to diminish the effects of competition on price and profits.

Chapter 2

Double Shifting: NHL Attendance Impacts of Player Participation in the Winter Olympics

2.1 Introduction

In 1995, representatives of the NHL, the NHLPA, and the IOC announced that NHL players would participate in the Winter Olympic Games for the first time. One of the stated goals of this joint venture was to increase exposure to the game of hockey, and to generate interest in professional leagues. NHL commissioner Gary Bettman saw the impact the Dream Team had on the popularity of the NBA and was looking for a similar effect.[1] The level of competition at the Winter Olympic Games in Nagano, Japan in 1998 was well received and NHL players had participated in four consecutive Olympic Games since. While reactions to NHL players joining their respective Olympic teams have been generally favorable, participation comes at the risk of both season fatigue that may lead to a decline in popularity and player fatigue, which may lead to injury or harm performance. As the Winter Olympic Games take place in the middle of the NHL regular season, costs incurred by participation have the potential to be significant for both players and owners alike. Prior studies have scrutinized the costs of NHL involvement in the Olympic Games with respect to both competitive balance (Longley, 2012; Cairney **et al.**, 2015) and the potential for player injury (Engebretsen **et al.**, 2016). These potential costs may be a major contributing factor in the NHL's decision not to participate in the next Winter Olympics in South Korea. To date, there has been no rigorous study of the primary benefit referenced in the NHL's and NHLPA's

[1] In 1992, commissioner Bettman worked as a league executive in the NBA.

initial justification for allowing participation: increased exposure to the sport. Our study addresses this shortfall by analyzing one dimension of changes in exposure. We assess the impact of NHL player participation in the Olympic Games on NHL game attendance.

We choose to examine the impact on attendance for two reasons. First, gate receipts are the single largest contributor to total revenue, accounting for roughly 40% of a team's income. In 2014, the average NHL team took in nearly $50 million in ticket sales.[2] Second, the direct link between attendance and ticket sales allows us to easily quantify any impacts on attendance in dollar terms. While it is possible that NHL participation in the Olympic Games may also have an impact on television viewership, the prohibitive cost of national and regional ratings data makes this dimension difficult to analyze.

Cursory examinations of attendance data indicate NHL participation in the Olympics may, on average, have a positive effect on attendance. Table 2.1 reports average game attendance and average attendance as a percent of arena capacity for Olympic and non-Olympic years before and after NHL participation. While these results are suggestive of a possible boost in attendance with NHL player eligibility, a portion of this increase may be attributable to general increases in attendance over time. This trend is evident in Figure 2.1. In addition, factors such as league-wide competitiveness, upgraded arenas, and fluctuations in general regional preferences may drive the observed differences in attendance displayed in Table 2.1. We appeal to econometric estimation methods to isolate the impact of Olympic participation on NHL attendance by controlling for confounding cross-sectional and temporal factors.

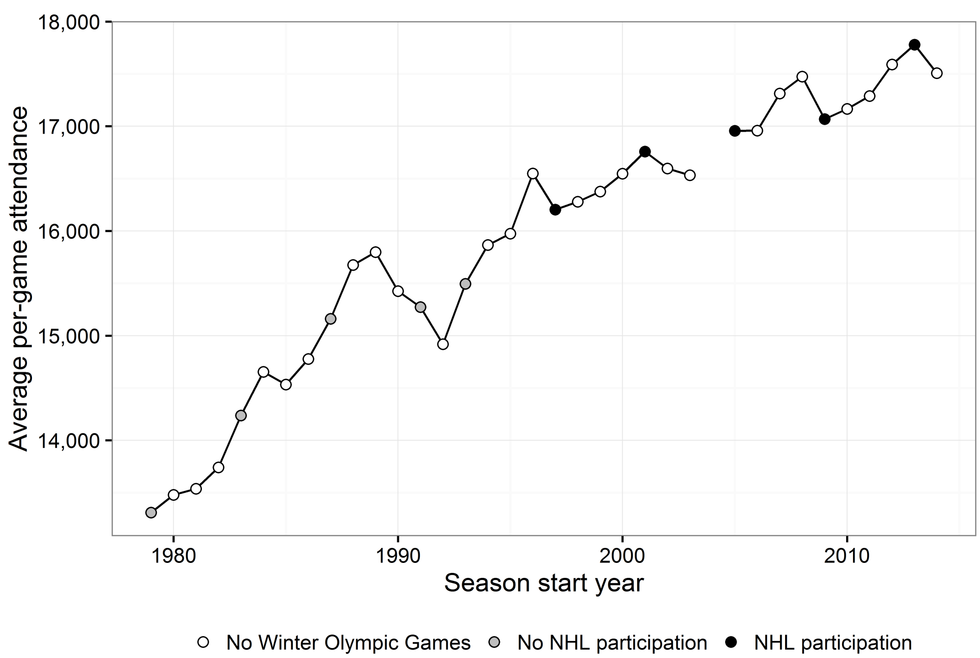
We employ a censored difference-in-differences estimation strategy and reveal a causal link between NHL Olympic participation and average game attendance. Our strategy accounts for censoring of data due to stadium capacity constraints. Our model also takes into account changes in the popularity of the NHL before and after the policy change. After controlling for a range of factors, we find evidence that between 1987 and 2004 NHL participation in the Olympic Games

[2] To our knowledge, the NHL's revenue sharing program allows most teams to retain all of the revenue from their home games and redistributes a small percentage away from the top 10 revenue-generating teams.

Table 2.1: Summary of Per-game Attendance

	1979-2014	1987-2004
Panel A. Average per-game attendance		
Non-Winter Olympic Games seasons	16,103	16,088
Winter Olympic Games seasons		
Without NHL participation	14,736	15,322
With NHL participation	16,991	16,502
Panel B. Average percent of capacity		
Non-Winter Olympic Games seasons	90.0	90.5
Winter Olympic Games seasons		
Without NHL participation	87.2	89.8
With NHL participation	92.0	90.0

Figure 2.1: Average Per-game Attendance by Season



leads to an increase in league attendance of approximately 4.2% in Olympic years. As NHL teams during this sample period averaged 16,020 fans per game, this suggests an extra 673 tickets sold each game. Assuming a similar boost to attendance today, an average ticket price of $65, and 41 home games per season, this amounts to over $1,790,000 in additional attendance revenue per team in Olympic year seasons. Impacts of this magnitude suggest NHL participation in the Olympic Games is a valuable revenue generator for NHL teams.

Ignoring the censoring issue due to stadium capacity biases our estimates toward zero. Additionally, excluding observations where game attendance was restricted due to capacity, we find that NHL participation led to an even larger boost to attendance of 5.9% on average. This suggests that teams that failed to fill their stadiums enjoyed a greater benefit in terms of increased attendance from this partnership with the Olympics. Expanding the dataset to include all years following the NHL-WHA merger in 1978 results in a slightly smaller but still significant increase of about 3.5%. Analyzing the impact for only the years following the announcement of the partnership indicates an Olympic boost on attendance of about 1.8% relative to non-Olympic seasons. However, looking at the two most recent Olympic seasons, the impact is almost double. This could be because of the relatively better performances of the US and Canadian teams. Finally, our results indicate that the boost in attendance is likely confined to Olympic years with NHL participation, as we do not find significant spillover effects into non-Olympic years.

By cooperating rather than competing, the NHL and the Olympics realize benefits by providing a better hockey product. Furthermore, the NHL enjoys an increase in attendance throughout the year due to its participation in the Olympics.[3] These results provide some empirical evidence of the spillover effects that come from joint ventures. By working together to provide a product, firms can generate increased notoriety that boost sales of its other products.

Our study builds on literature that assesses determinants of sporting event demand (i.e., attendance). The majority of existing studies focus on determinants of attendance in standard

[3]The Olympics likely receive a greater spillover benefit, in the form of increased viewership of other Olympic events, from this partnership with the NHL

season play and do not explicitly account for player participation in international events. Villar **et al.** (2009) and Schofield (1983) provide extensive surveys. Literature focused on Olympic participation is limited. Most relevant are works by Longley (2012) and Cairney **et al.** (2015) that examine player impacts of participation and assess the extent to which team performance is impacted by the number of roster players that participate in the Olympic Games. Our analysis also lends empirical support for a developing line of theoretical literature on player participation in international events (Gürtler **et al.**, 2015). Beyond this performance-based assessment, no current research examines the league-wide effect of participation on attendance. Understanding these demand-side consequences is essential from the perspective of the NHL, players, and team owners who must weigh these benefits with possible costs when deciding to participate in the Winter Olympics in future seasons.

2.2 Demand for Professional Sporting Events

Our work contributes to existing literature on the topic of sporting event demand. As most studies in this area discuss, demand for professional sporting events is based on a standard goods demand model where consumers make decisions about consumption subject to their economic status and informed by the observable (to the consumer) attributes of the good.[4] A survey of prior research in this area can be found in Villar **et al.** (2009) and Schofield (1983). In the context of professional sports, demand is often measured in the form of game (or match) attendance. Villar **et al.** (2009) classify the determinants of attendance into four broad categories: i) expected quality, ii) economic aspects, iii) uncertainty of outcome, and iv) opportunity cost and other factors. We use these four general classifications to guide our analysis and tailor our model specifically to a study of attendance at NHL games. Villar **et al.** (2009) directly cite several studies of league attendance in North American hockey markets. We turn to these and other NHL-focused analyses to inform our empirical specification.

Expected quality measures considered by prior studies include home and visitor team quality,

[4] For a detailed discussion of the professional sports club's profit maximization see Ferguson **et al.** (1991).

club rankings, and value of superstars (Coates & Humphreys, 2012; Jones & Ferguson, 1988). Each of these attributes is a quality-based signal to prospective hockey "consumers" and informs their decision to consume the good. In addition to increased exposure to the game through international competition, this is another possible mechanism through which NHL participation in the Olympic Games can influence game attendance. That is, competition at the national level provides a clearer signal of a domestic team's talent level. We test the extent to which this mechanism drives attendance in an extension of our model.

With respect to economic aspects influencing league attendance, prior studies of NHL attendance indicate mixed results with respect to importance of income and pricing components. Jones & Ferguson (1988) include consumer income and find that consumers treat attendance at NHL games as an inferior good. An assessment of game pricing is made difficult by the relative paucity of ticket pricing data and is further complicated by the fact that both attendance and price may be endogenous.[5] To the extent that price discrimination occurs within an arena and attendance is driven differentially by changes in the distribution of ticket prices, studies to date have not been able to isolate, and therefore control for, this effect.

Lastly, opportunity costs and other effects such as time, weather, and new arenas could also influence attendance.Leadley & Zygmont (2006) expand the attendance model of Jones & Ferguson (1988) and find positive impacts of new arenas on event attendance. A number of empirical studies account for the impact of increased league popularity over time by introducing a time trend to the analysis (Leadley & Zygmont, 2006; Carlton **et al.**, 2004). Olympic participation may play a role in driving this trend through increased exposure and therefore increased overall popularity, however the lack of a counterfactual league precludes us from estimating the magnitude of this long-run effect.

As the main focus of our study is Olympic participation's impact on NHL game attendance, we contribute to a sparse literature on the consequences of complementary international sporting

[5] Existing price data consists of either ticket prices set in advance of the season based on expected demand or estimated from total revenues.

event participation. Longley (2012) points out that there are few parallels in professional sports to the NHL's participation in the Winter Olympic Games. The NHL is unique in its opportunity and willingness to participate in an international competitive event during its standard playing season. International competition for most other major sports is either nonexistent (NFL), or takes place during the offseason (MLB participation in the World Baseball Classic, NBA participation in the summer Olympic Games, FIFA World Cup).[6] Longley's main focus deviates from our paper in that it examines the extent to which Olympic participation impacts competitive balance in the NHL. The author's findings focus primarily on team-level point production, a metric that can be interpreted as impacting competitive balance and game quality that indirectly affects interest in attending live games. In contrast, our study looks at the direct impact of the Olympics on attendance.

Recent work by Gürtler **et al.** (2015) examines the interactions of national teams and local teams in the context of professional soccer. The authors explore optimal bargaining between the two parties. An important foundation of this theoretical work is the fact that local clubs already deem themselves international brands and are disincentivized to participate. In the NHL, teams are not considered international brands and the league itself is therefore incentivized to grow its brand both domestically and globally.

While none of the aforementioned studies directly address the impact of Olympic participation on the demand for North American professional hockey, all form the foundation from which our model is specified. We turn now to a discussion of our empirical approach and a description of the available data. A discussion of implications and potential future work concludes.

2.3 Empirical Approach

Our main specification analyzes the extent to which NHL participation in the Winter Olympics impacts attendance over the course of a season. We utilize a censored difference-in-differences framework to identify the effect from NHL player participation in the Winter Olympics. We

[6] It should be noted that there are some exceptions in the area of international soccer competition. For example, the 2022 FIFA World Cup scheduled for Qatar will take place during several major professional league seasons.

empirically estimate:

$$\begin{aligned} ln(Att_{it}) =&\beta_1 Oly_t + \beta_2 NHLPart_t + \beta_3 Oly_t * NHLPart_t \\ &+ \gamma X_{it} + \delta OtherEvent_t + \alpha_i + Trend + \epsilon_{it}. \end{aligned} \tag{2.1}$$

The dependent variable is the log of average game attendance for team i in season t. The variable Oly is an indicator dummy for Olympic seasons and $NHLPart$ represents the years following the NHL's decision to participate in the Winter Olympics, which began with the 1997-1998 season. The interaction term, $Oly * NHLPart$, indicates the seasons in which NHL players participated in Olympic hockey. Thus, β_1 provides an estimate of the average Olympic effect on attendance, β_2 represents the effect on attendance following the policy change, and β_3 is our coefficient of interest, the effect of NHL involvement in the Olympics. One issue in identifying the effect of NHL participation in the Olympics on attendance is the fact that attendance levels for some teams are constrained due to stadium capacity. Failure to account for these capacity constraints could lead us to understate the effect of participation. To address this concern, we estimate the attendance regression using a censored regression framework with arena capacity as the right censoring value. The tobit procedure assumes the error terms are normally distributed with constant variance and uses maximum likelihood to estimate the parameters of the model. We also include a discussion of the more conservative, uncensored estimation results.

The remaining variables in our model function as controls. As discussed previously, team quality can significantly impact attendance. For this reason, we introduce several team performance variables captured in the X_{it} term.[7] One such attribute, previous season performance, may have a carryover effect on attendance next season. To address this concern, we include a team's points earned as a percent of total possible points, as well as indicators for whether the team made the playoffs, reached the finals, and won the Stanley Cup in the previous season.[8] Current season performance indicators include the team's total points percentage and whether the team made the playoffs that year. These performance variables are similar to those found in previous literature

[7]For a complete list of controls, see Table 2.2.

[8]The use of previous season data excludes the first year of expansion or relocated teams from our dataset.

on attendance.[9] Leadley & Zygmont (2006) find that new NHL arenas themselves can positively influence attendance. The data show significant increase in attendance for the first two years of a new arena, so we include dummy variables for each of those years. We include the log of arena capacity to account for any stadium size effects on attendance.[10] We also account for other international competitions that involved NHL players during this time. This includes both the Canada Cup and World Cup competitions, which took place in 1987, 1991, and 1996. The inclusion of team fixed effects controls for time-invariant differences across franchises that may affect attendance. Finally, to account for the fact that NHL participation occurs later in our data set and that the NHL has seen a general increase in popularity over our sample, we incorporate a time trend in our empirical specification.

Standard errors for all model specifications that rely on team-level data are clustered by season. This level of clustering addresses concerns due to the aggregate nature of the treatment we study. Specifically, concerns about studying policies applied to all groups within a sample. We discuss these concerns and related, auxiliary results in section 2.4.2.

2.3.1 Data

Figure 2.1 provides a view of average per-game attendance by season from 1979-2014. There is a clear upward trend in attendance throughout this period, likely due to increases in popularity and arena capacities, which provides support for the inclusion of an arena capacity and trend variable in our regressions. Gray dots in the figure represent Olympic seasons without NHL participation and black dots represent Olympic seasons with NHL participation. Table 2.1 summarizes the average attendance data for the entire period and from a subset of data around the time the decision to participate was made. A simple, descriptive difference-in-differences analysis of these data indicates a boost in both attendance and capacity utilization from Olympic participation.

[9]For example, seeFerguson **et al.** (1991); Brook **et al.** (2006); Villar **et al.** (2009).

[10]We note that Villar **et al.** (2009) suggest this is a measure of potential demand rather than a control for actual demand, however in the absence of a better control, we employ this capacity measure in our specification.

This effect is illustrated in Figure 2.2. The top panel of the figure presents the log of average league-wide attendance. The bottom panel includes detrended residuals. Olympic years after NHL participation show a clear increase over years without NHL participation and without Olympic games. While this graphical evidence is of value in assessing participation impacts on attendance, for the reasons outlined previously, we rely on regression analysis with cross-sectional and temporal controls to estimate the causal effect.

Figure 2.2: League-Wide Average Attendance by Season

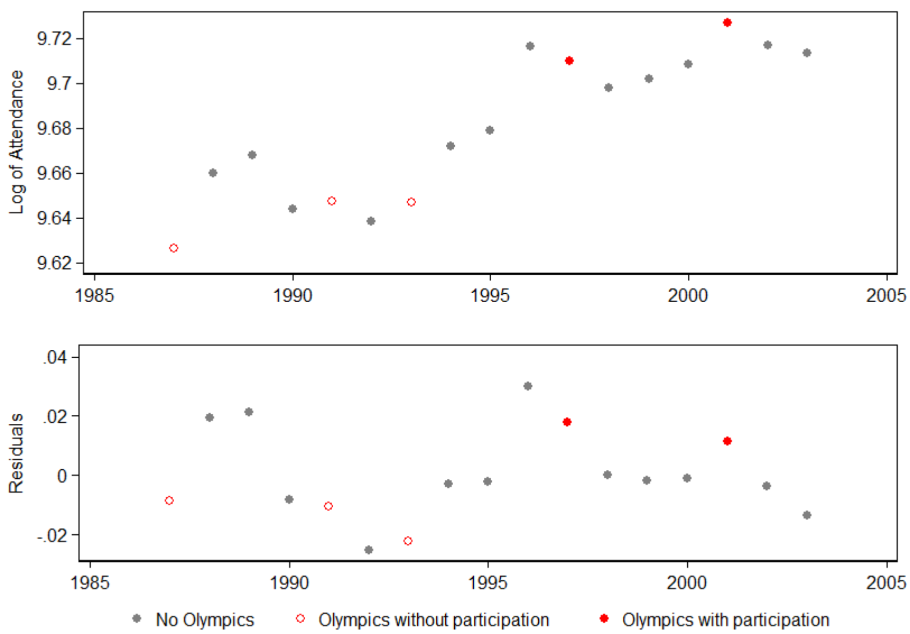


Table 2.2 provides summary statistics for the data used in our season-level models. Our preferred specification includes data from the 1987-88 through 2003-04 seasons. Following the 2003-04 season, the subsequent NHL lockout resulted in the cancellation of a full season. This time frame includes five Winter Olympics (1988, 1992, 1994, 1998, and 2002), the last two of which involved NHL players.[11] Average game attendance by season was 16,020 and ranged from 7,863 to 21,002 tickets sold, while average arena capacity was 17,839 seats. Four percent of team-seasons

[11] As a sensitivity, we estimate the above model on the full time series of available data. Results are presented in Table 5 and discussed in more detail below.

were spent in a new arena. Attendance data come from two sources: Rod Fort's collection of sports data and hockeydb.com. All team performance data, Olympic participation, and other international competitions data, were obtained from hockey-reference.com. Data on NHL arenas and seating capacities are from hockey.ballparks.com.

Table 2.2: Summary Statistics

Variable	Mean	Std. Dev.	Min.	Max.
Average per-game attendance	16,020	2,424	7,863	21,002
Olympics	0.28	0.45	0	1
NHL participation	0.13	0.34	0	1
Points percentage	0.512	0.098	0.143	0.799
Playoff team	0.64	0.48	0	1
Prior season points percentage	0.507	0.1	0.143	0.799
Prior season playoff team	0.64	0.481	0	1
Prior season champions	0.04	0.197	0	1
Prior season runners up	0.04	0.197	0	1
Arena capacity	17,839	1,899	10,585	26,000
First-year arena	0.04	0.20	0	1
Second-year arena	0.05	0.22	0	1
Other international event	0.16	0.37	0	1

2.4 Effect of Olympic Participation on NHL Attendance

2.4.1 Main Results

The results from the estimation of equation (2.1) are presented in Table 2.3. Column 1 of Table 2.3 provides the results with the inclusion of an overall time trend. Column 2 presents results when we include team fixed effects to account for any time-invariant team or city effects that influence attendance. The coefficient on the Olympic variable suggests that increased competition from the Winter Olympics had a negative impact on NHL attendance when league players were not involved. We hypothesize that hockey fans substitute away from the NHL and consume more international hockey or other winter sports competitions. This trend is reversed, however, when NHL players participate in the international competition. In fact, NHL attendance increased by over 4% in those seasons compared to previous Olympic seasons. This could be the result of the expansion of hockey to a larger audience and the increased promotion of the NHL and its players from the Olympics. The results indicate that the league achieved its stated goal of generating interest in the NHL through the Winter Olympics. The coefficients on the control variables indicate that the quality of the team and the arena they play in are also significant in explaining variation in attendance.

In column 3, the overall time trend is replaced with team specific time trends. This allows for popularity effects over time to vary by team. There could be significant local economic or demographic changes throughout this period that may affect attendance. For example, increases in income levels or population over time for a city with an NHL franchise could have significant impacts on game attendance. This may, in turn, affect the estimated relationship between the interaction term and the dependent variable. This could also be the case if other sports franchises move in and out of the area.[12] As displayed in column 3, the results with this specification are fairly consistent. In fact, these results are stable across several specifications, including those with team specific trends and team fixed effects found in column 4.

2.4.2 Robustness

To test whether capacity crowds impact our estimates, we estimate a model that does not account for censoring. Table 2.4 presents these results. As expected, ignoring the censoring issue produces marginally smaller parameter estimates for most of the variables. The coefficients on the interaction term suggest a similar effect to those from the censored regression. As an alternative to censored estimation, we present results for regressions excluding observations where average game attendance over the course of the season is larger than the stated capacity of the arena. These results are displayed in Table 2.5. Average game attendance is greater than arena seating capacity for 34 team-seasons or about 8% of the observations. Twenty-four of those observations come from just four teams suggesting these teams either sold significant standing room only tickets or there is

[12] Only six cities saw movement in and/or out of other major franchises: St. Louis, Los Angeles, Vancouver, Phoenix, Tampa Bay, and Minneapolis.

Table 2.3: Censored Regression Results (1987-2004 seasons)

	(1)	(2)	(3)	(4)
Olympics	-0.026^{***}	-0.029^{***}	-0.030^{***}	-0.030^{***}
	(0.007)	(0.006)	(0.008)	(0.007)
NHL Participation	-0.020^{**}	-0.001	-0.015	0.011
	(0.008)	(0.008)	(0.014)	(0.012)
NHL Participation $\times$ Olympics	0.050^{***}	0.042^{***}	0.049^{***}	0.037^{***}
	(0.011)	(0.007)	(0.009)	(0.009)
Points percentage	0.487^{***}	0.336^{***}	0.388^{***}	0.381^{***}
	(0.095)	(0.077)	(0.086)	(0.058)
Playoff team	-0.038^{**}	-0.015	-0.030^{**}	-0.025^{***}
	(0.018)	(0.013)	(0.015)	(0.009)
Prior season points percentage	0.293^{***}	0.182^{***}	0.152^{*}	0.316^{***}
	(0.089)	(0.065)	(0.080)	(0.063)
Prior season playoff team	0.026	0.041^{***}	0.041^{**}	0.019
	(0.021)	(0.014)	(0.018)	(0.013)
Prior season champions	0.008	0.005	0.020	0.008
	(0.026)	(0.015)	(0.013)	(0.013)
Prior season runners up	0.011	0.038	0.058^{**}	0.043^{*}
	(0.026)	(0.029)	(0.025)	(0.022)
Log(arena capacity)	0.681^{***}	0.518^{***}	0.538^{***}	0.596^{***}
	(0.051)	(0.063)	(0.070)	(0.095)
First-year arena	0.072^{**}	0.078^{***}	0.071^{**}	0.081^{***}
	(0.030)	(0.026)	(0.030)	(0.030)
Second-year arena	0.069^{**}	0.052^{**}	0.048^{*}	0.031
	(0.033)	(0.024)	(0.025)	(0.024)
Other international event	0.006	0.010^{*}	0.009	0.010
	(0.007)	(0.006)	(0.007)	(0.006)
Time Trend	Yes	Yes	No	No
Team FE	No	Yes	No	Yes
Team Time Trend	No	No	Yes	Yes
Log Likelihood	207	369	334	454
Observations	422	422	422	422

Standard errors clustered at the season level in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

some measurement error in stadium capacities.[13] As displayed in Table 2.5, the coefficients on the interaction term are larger by roughly 1%. This result suggests, unsurprisingly, that teams with excess capacity may see larger attendance increases than capacity constrained teams as a result of NHL participation in the Winter Olympics.

Expanding the number of seasons to include the full dataset from 1979-2014 yields a slightly smaller impact on attendance of about 3.5%. These estimates can be found in Table 2.6. We chose the 1979 season as the cutoff, since this was the first year following the merger of the NHL and WHA. This move effectively made the NHL the only major professional hockey league in North America. The 2004 lockout and subsequent renegotiation of the collective bargaining between the NHL and NHLPA led to changes in game rules and changes to how points were allocated.[14] Although this may affect the estimates on the points percentage variables, it should not affect our estimates of the coefficients on the Olympics, NHL participation, and the interaction term, since they are uncorrelated with how team performance is measured. In this specification, we include variables to control for the effect of the lockout season and other shortened seasons due to work stoppages.[15]

We also consider the effects over select time periods. These results are shown in Table 2.7. The first four columns examine the effect for years 1995-2014, while the fifth column looks at the more recent seasons from 2008-2014. The latter includes two Olympic years, 2010 and 2014. We hypothesize that the attendance effect may change over time if the partnership has lost some of its novelty or is expected by consumers. It could also be because the league has employed new channels (e.g., social media, online streaming) to reach consumers and promote league events and no longer realizes the benefits of showcasing its players in the Winter Olympics. Since every Olympic season during this time includes NHL player participation, we exclude the Olympic and NHL participation

[13]Teams with attendance consistently over capacity include the Calgary Flames, Chicago Blackhawks, Philadelphia Flyers, and Toronto Maple Leafs.

[14]Particularly, teams were awarded an additional point for winning a shootout rather than the game being declared a tie as it had been prior to the lockout.

[15]These include the 1992, 1994, and 2012 seasons.

Table 2.4: Uncensored Regression Results (1987-2004 seasons)

	(1)	(2)	(3)	(4)
Olympics	-0.023^{***}	-0.026^{***}	-0.027^{***}	-0.026^{***}
	(0.007)	(0.006)	(0.007)	(0.006)
NHL Participation	-0.027^{***}	-0.008	-0.023^{*}	0.002
	(0.008)	(0.008)	(0.012)	(0.012)
NHL Participation × Olympics	0.049^{***}	0.042^{***}	0.049^{***}	0.036^{***}
	(0.011)	(0.008)	(0.009)	(0.009)
Points percentage	0.422^{***}	0.293^{***}	0.349^{***}	0.335^{***}
	(0.085)	(0.073)	(0.081)	(0.056)
Playoff team	-0.032^{*}	-0.013	-0.027^{*}	-0.019^{*}
	(0.016)	(0.011)	(0.015)	(0.009)
Prior season points percentage	0.247^{***}	0.163^{**}	0.140^{*}	0.290^{***}
	(0.074)	(0.058)	(0.070)	(0.062)
Prior season playoff team	0.027	0.039^{**}	0.037^{*}	0.021
	(0.021)	(0.014)	(0.018)	(0.012)
Prior season champions	0.005	0.005	0.018	0.004
	(0.022)	(0.014)	(0.012)	(0.013)
Prior season runners up	0.004	0.025	0.040^{*}	0.026
	(0.019)	(0.026)	(0.023)	(0.021)
Log(arena capacity)	0.710^{***}	0.554^{***}	0.552^{***}	0.635^{***}
	(0.051)	(0.062)	(0.071)	(0.097)
First-year arena	0.050^{*}	0.060^{**}	0.054^{**}	0.065^{**}
	(0.024)	(0.021)	(0.024)	(0.024)
Second-year arena	0.049	0.039	0.039	0.021
	(0.029)	(0.022)	(0.025)	(0.026)
Other international event	0.006	0.010^{**}	0.009	0.011^{**}
	(0.005)	(0.005)	(0.006)	(0.005)
Time Trend	Yes	Yes	No	No
Team FE	No	Yes	No	Yes
Team Time Trend	No	No	Yes	Yes
Adjusted R^2	0.494	0.746	0.693	0.811
Observations	422	422	422	422

Standard errors clustered at the season level in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

Table 2.5: Regression results (1987-2004 seasons, excluding capacity constrained teams)

	(1)	(2)	(3)	(4)
Olympics	-0.024^{*}	-0.029^{***}	-0.031^{**}	-0.028^{***}
	(0.012)	(0.009)	(0.012)	(0.009)
NHL Participation	-0.035^{**}	-0.008	-0.019	0.002
	(0.014)	(0.010)	(0.015)	(0.015)
NHL Participation × Olympics	0.059^{***}	0.052^{***}	0.059^{***}	0.044^{***}
	(0.018)	(0.011)	(0.013)	(0.012)
Points percentage	0.430^{***}	0.308^{***}	0.372^{***}	0.374^{***}
	(0.096)	(0.076)	(0.094)	(0.074)
Playoff team	-0.035^{*}	-0.016	-0.032^{*}	-0.025^{**}
	(0.018)	(0.013)	(0.017)	(0.011)
Prior season points percentage	0.225^{**}	0.147^{**}	0.121	0.310^{***}
	(0.079)	(0.067)	(0.076)	(0.071)
Prior season playoff team	0.029	0.042^{**}	0.040^{**}	0.021
	(0.023)	(0.015)	(0.019)	(0.014)
Prior season champions	0.015	0.011	0.025^{*}	0.014
	(0.025)	(0.014)	(0.013)	(0.015)
Prior season runners up	0.010	0.038	0.057^{*}	0.039
	(0.022)	(0.034)	(0.030)	(0.028)
Log(arena capacity)	0.744^{***}	0.651^{***}	0.604^{***}	0.713^{***}
	(0.053)	(0.046)	(0.067)	(0.061)
First-year arena	0.043	0.052^{*}	0.048	0.065^{*}
	(0.036)	(0.028)	(0.031)	(0.031)
Second-year arena	0.034	0.031	0.032	0.020
	(0.031)	(0.030)	(0.028)	(0.030)
Other international event	0.012	0.016^{*}	0.016	0.012
	(0.011)	(0.009)	(0.012)	(0.009)
Constant	2.024^{***}	3.043^{***}	3.496^{***}	2.370^{***}
	(0.505)	(0.410)	(0.636)	(0.590)
Time Trend	Yes	Yes	No	No
Team FE	No	Yes	No	Yes
Team Time Trend	No	No	Yes	Yes
Adjusted R^2	0.497	0.736	0.682	0.800
Observations	370	370	370	370

Standard errors clustered by season in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

Table 2.6: Censored Regression Results (1979-2014 seasons)

	(1)	(2)	(3)	(4)
Olympics	−0.022*	−0.019**	−0.016*	−0.013
	(0.011)	(0.009)	(0.008)	(0.010)
NHL Participation	−0.052***	−0.023*	−0.054***	−0.051***
	(0.013)	(0.013)	(0.012)	(0.016)
NHL Participation × Olympics	0.042**	0.035**	0.035**	0.030*
	(0.018)	(0.015)	(0.014)	(0.016)
Points percentage	0.356***	0.344***	0.355***	0.400***
	(0.096)	(0.075)	(0.076)	(0.060)
Playoff team	0.011	0.000	0.000	−0.011
	(0.014)	(0.011)	(0.013)	(0.009)
Prior season points percentage	0.275***	0.264***	0.280***	0.325***
	(0.080)	(0.067)	(0.074)	(0.061)
Prior season playoff team	0.039***	0.028**	0.031**	0.021*
	(0.014)	(0.012)	(0.013)	(0.011)
Prior season champions	0.045**	0.063***	0.050***	0.037**
	(0.021)	(0.019)	(0.014)	(0.017)
Prior season runners up	0.050**	0.043**	0.045**	0.042**
	(0.020)	(0.019)	(0.018)	(0.018)
Log(arena capacity)	0.693***	0.521***	0.456***	0.466***
	(0.057)	(0.066)	(0.061)	(0.073)
First-year arena	0.069***	0.085***	0.069**	0.065**
	(0.027)	(0.031)	(0.034)	(0.029)
Second-year arena	0.058**	0.064**	0.050*	0.031
	(0.029)	(0.029)	(0.029)	(0.022)
Other international event	0.007	0.008	0.010	0.008
	(0.010)	(0.009)	(0.007)	(0.009)
Shortened season	0.017***	0.012**	0.008	0.013
	(0.006)	(0.006)	(0.011)	(0.013)
Season after 2004 lockout	−0.010	−0.004	−0.008	−0.016
	(0.013)	(0.011)	(0.010)	(0.012)
Time Trend	Yes	Yes	No	No
Team FE	No	Yes	No	Yes
Team Time Trend	No	No	Yes	Yes
Log Likelihood	274	503	482	631
Observations	883	883	883	883

Standard errors clustered by season in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

variables and only include their interaction. The parameter estimates for the 1995 to 2014 seasons indicate a similar effect of Olympic participation on attendance relative to non-Olympic seasons. However, more recently, it seems like the NHL does enjoy a significant boost in attendance levels of about 3% when NHL players participate in the Olympics. A similar effect in 2018 would lead to an increase in average attendance of 525 per NHL game. As our results indicate, the NHL's participation in the Olympic games provide a moderate boost to demand for the domestic league's games.

The trends observed in Figures 2.1 and 2.2 raise potential questions about the role capacity expansion played over the period of study. While league-wide capacity did not expand dramatically over this period, we nonetheless apply our censored regression model replacing the dependent variable with attendance as a percent of capacity. Results are presented in Table 2.8 and are qualitatively similar to those produced by the model of log attendance.

As a final sensitivity, we investigate possible deflation of standard errors that may arise from the fact that our primary model applies a league-wide policy to individual, team-level observations. As documented in Moulton (1986); Bertrand **et al.** (2004), this setting may lead to over-rejection of the null hypothesis - a troubling result when interpreting the significance of the interaction term of our model. As a first-order approach to addressing the concern, we cluster standard errors for all models at the season level. We also apply two alternative estimation approaches to confirm the magnitude and significance of our primary result.

First, we aggregate data to the season level and thus regress average league-wide attendance by season on our treatment and pre/post period variables. This aggregation effectively collapses our sample to 17 observations. Results of a parallel, censored regression on the team-level data (i.e., constraining γ and δ to zero in equation 2.1) and the season-level averages are included in Table 2.9. Results of the two estimation strategies are similar in magnitude and significance, lending support to our primary result.

As a second approach, we construct a series of placebo policy interventions. For this test, we randomly assign an NHL participation period and Olympic years to the data and apply our

Table 2.7: Estimation Results for Select Time Periods

	1995-2014				2008-2014
	(1)	(2)	(3)	(4)	(5)
NHL Participation × Olympics	0.017	0.014	0.018	0.015	0.030***
	(0.012)	(0.009)	(0.011)	(0.012)	(0.002)
Points percentage	0.225*	0.283***	0.324***	0.336***	0.003
	(0.129)	(0.098)	(0.098)	(0.085)	(0.004)
Playoff team	0.020	−0.002	−0.008	−0.005	−0.004
	(0.020)	(0.016)	(0.018)	(0.013)	(0.003)
Prior season points percentage	0.292***	0.302***	0.347***	0.316***	0.180***
	(0.105)	(0.097)	(0.092)	(0.093)	(0.004)
Prior season playoff team	0.039**	0.016	0.016	0.022	−0.000
	(0.016)	(0.014)	(0.014)	(0.015)	(0.003)
Prior season champions	0.036	0.093***	0.065***	0.069**	
	(0.035)	(0.023)	(0.025)	(0.028)	
Prior season runners up	0.038	0.051**	0.040**	0.067**	0.040***
	(0.029)	(0.021)	(0.020)	(0.028)	(0.007)
Log(arena capacity)	0.852***	0.227**	0.268***	0.241***	−0.044***
	(0.125)	(0.103)	(0.101)	(0.083)	(0.000)
First-year arena	0.030	0.057	0.058*	0.075***	
	(0.026)	(0.037)	(0.030)	(0.027)	
Second-year arena	0.035	0.053	0.061*	0.046**	
	(0.035)	(0.035)	(0.035)	(0.021)	
Other international event	0.040***	0.023***	0.038***	0.033**	
	(0.010)	(0.007)	(0.008)	(0.015)	
Shortened season	0.030***	0.024***	0.038***	0.041***	0.028***
	(0.007)	(0.007)	(0.008)	(0.009)	(0.001)
Season after 2004 lockout	−0.011	−0.002	−0.008	−0.012	
	(0.012)	(0.009)	(0.012)	(0.012)	
Time Trend	Yes	Yes	No	No	Yes
Team FE	No	Yes	No	Yes	Yes
Team Time Trend	No	No	Yes	Yes	No
Log Likelihood	157	319	313	419	134
Observations	545	545	545	545	179

Standard errors clustered by season in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

Table 2.8: Olympic Participation Impact on Capacity Rates (1987-2004 seasons)

	(1)	(2)	(3)	(4)
Olympics	-0.022^{***}	-0.024^{***}	-0.026^{***}	-0.025^{***}
	(0.005)	(0.005)	(0.007)	(0.006)
NHL Participation	-0.014^{*}	0.001	-0.006	0.009
	(0.008)	(0.007)	(0.008)	(0.010)
NHL Participation × Olympics	0.038^{***}	0.033^{***}	0.038^{***}	0.029^{***}
	(0.009)	(0.006)	(0.008)	(0.008)
Points percentage	0.391^{***}	0.265^{***}	0.310^{***}	0.296^{***}
	(0.074)	(0.060)	(0.068)	(0.047)
Playoff team	-0.028^{*}	-0.010	-0.023^{*}	-0.017^{**}
	(0.015)	(0.010)	(0.012)	(0.008)
Prior season points percentage	0.249^{***}	0.159^{***}	0.140^{**}	0.261^{***}
	(0.077)	(0.054)	(0.066)	(0.051)
Prior season playoff team	0.020	0.032^{***}	0.031^{**}	0.016
	(0.017)	(0.011)	(0.014)	(0.010)
Prior season champions	0.008	0.005	0.017	0.008
	(0.022)	(0.013)	(0.011)	(0.010)
Prior season runners up	0.009	0.027	0.046^{**}	0.035^{*}
	(0.023)	(0.023)	(0.020)	(0.018)
Log(arena capacity)	-0.270^{***}	-0.415^{***}	-0.391^{***}	-0.343^{***}
	(0.041)	(0.061)	(0.058)	(0.094)
First-year arena	0.061^{**}	0.064^{***}	0.060^{**}	0.062^{***}
	(0.025)	(0.021)	(0.025)	(0.023)
Second-year arena	0.058^{**}	0.041^{**}	0.038^{*}	0.021
	(0.029)	(0.019)	(0.021)	(0.019)
Other international event	0.004	0.008	0.007	0.008
	(0.005)	(0.005)	(0.007)	(0.005)
Time Trend	Yes	Yes	No	No
Team FE	No	Yes	No	Yes
Team Time Trend	No	No	Yes	Yes
Log Likelihood	276	446	415	534
Observations	422	422	422	422

Dependent variable is per game attendance as percent of total team capacity. Standard errors clustered by season in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

regression specification. Figure 2.3 is a histogram of point estimates for the model's interaction term. The vertical line in the figure is the estimate obtained from the censored regression result in column 1 of Table 2.9. The distribution of these placebo estimates serves as further confirmation of the result of our censored methodology.

Table 2.9: Moulton Problem Sensitivities (1987-2004 seasons)

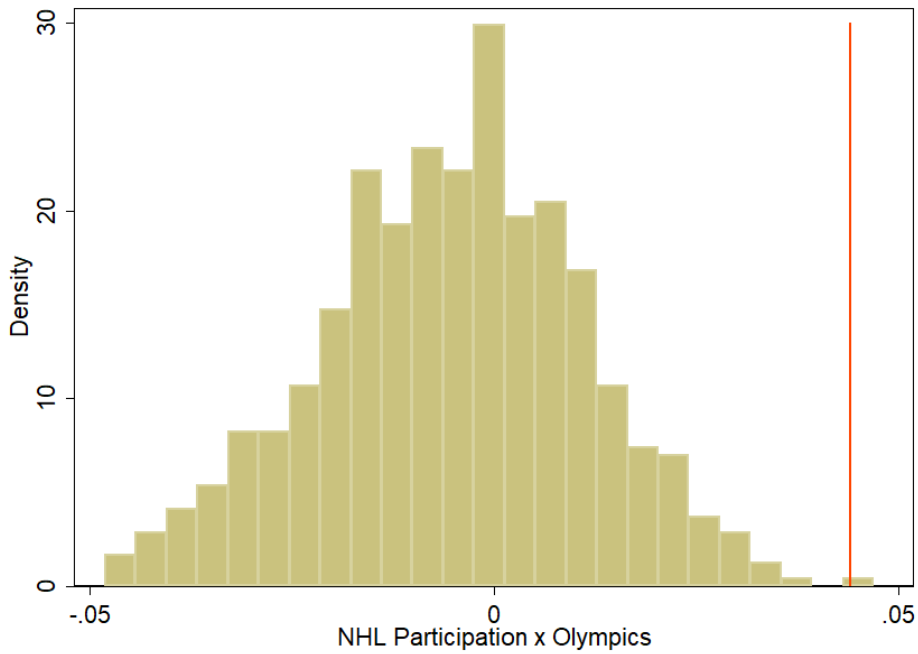
	Censored (1)	Annual (2)
Olympics	-0.025^{**}	-0.020^{**}
	(0.010)	(0.008)
NHL Participation	0.002	0.000
	(0.019)	(0.018)
NHL Participation × Olympics	0.044^{***}	0.037^{***}
	(0.011)	(0.009)
Constant	9.460^{***}	9.476^{***}
	(0.066)	(0.064)
Time Trend	Yes	Yes
Adjusted R^2		0.811
Observations	422	17

Column 1 standard errors clustered by season. Column 2 standard errors robust to heteroskedasticity.
* p<0.1, ** p<0.05, *** p<0.01.

2.4.3 Extensions

The effect of NHL participation and Olympic performance may have residual impact after Olympic years. To address this concern, we allow treatment effects to vary by the amount of time following an Olympic season. Table 2.10 presents these results. The magnitude of the effect in the Olympic year ($t = 0$) is similar to results of our main specification, on the order of 3 to 4 percent. Tests of joint significance of the interactions for one year and two years after an Olympic year in two of the models indicate that the effect of NHL participation may differ over time. Our estimates suggest that outside of Olympic years, NHL participation did not have a significant effect on attendance (i.e., a potential residual effect). The effect of additional advertising and fan interest surrounding the NHL and its players seems to be contained to Olympic years. That is not to say

Figure 2.3: Interaction Term Point Estimates from Placebo Simulation



that Olympic participation has not increased the overall popularity of the NHL. The longer term effect could simply be captured in the positive time trend present in Figure 2.1.

The results from the main specification suggest on average a 4 percent boost in attendance for NHL games in seasons where NHL players participate in the Winter Olympics. Considering hockey's popularity in Canada, it is possible that Olympic participation has a different impact on attendance based on whether the franchise is located in the United States or Canada. The estimates of a differential effect can be found in Table 2.11. Columns 1 and 2 show the results from estimation of equation 1 separately for U.S. and Canadian teams. For these separate regressions, our point estimates suggest the effect of NHL participation on attendance may be larger for franchises in the United States, however, we cannot conclude this effect is significantly different from that of Canadian teams. Column 3 and 4 present a full interaction model allowing for the effect of the Olympics and NHL participation to vary by country, with column 4 including individual team trends. The estimates show there may be a slight boost in attendance of professional hockey games

Table 2.10: Residual Effect of Olympic Participation (1987-2004 seasons)

	(1)	(2)	(3)	(4)
Olympics	-0.022^{***}	-0.025^{***}	-0.026^{***}	-0.027^{***}
	(0.008)	(0.006)	(0.010)	(0.007)
Olympics$_{t-1}$ (one year prior)	0.006	0.005	0.005	0.001
	(0.009)	(0.008)	(0.008)	(0.008)
Olympics$_{t-2}$ (two years prior)	0.013	0.015^{*}	0.015^{*}	0.013
	(0.009)	(0.008)	(0.008)	(0.011)
NHL Participation	-0.017^{*}	0.014	-0.000	0.019
	(0.009)	(0.010)	(0.011)	(0.014)
After NHL participation × Olympics	0.047^{***}	0.030^{***}	0.035^{***}	0.031^{***}
	(0.009)	(0.008)	(0.010)	(0.010)
After NHL participation × Olympics$_{t-1}$	0.002	-0.012	-0.013	-0.003
	(0.009)	(0.009)	(0.011)	(0.010)
After NHL participation × Olympics$_{t-2}$	-0.016^{*}	-0.029^{***}	-0.029^{***}	-0.022^{*}
	(0.010)	(0.009)	(0.011)	(0.013)
Points percentage	0.489^{***}	0.336^{***}	0.388^{***}	0.383^{***}
	(0.095)	(0.076)	(0.084)	(0.058)
Playoff team	-0.038^{**}	-0.015	-0.030^{**}	-0.025^{***}
	(0.018)	(0.012)	(0.015)	(0.009)
Prior season points percentage	0.293^{***}	0.180^{***}	0.149^{*}	0.314^{***}
	(0.089)	(0.065)	(0.081)	(0.065)
Prior season playoff team	0.026	0.041^{***}	0.042^{**}	0.019
	(0.021)	(0.014)	(0.018)	(0.013)
Prior season champions	0.008	0.005	0.020	0.008
	(0.026)	(0.015)	(0.013)	(0.012)
Prior season runners up	0.011	0.038	0.058^{**}	0.043^{*}
	(0.026)	(0.030)	(0.025)	(0.023)
Log(arena capacity)	0.680^{***}	0.518^{***}	0.538^{***}	0.591^{***}
	(0.052)	(0.064)	(0.070)	(0.095)
First-year arena	0.073^{**}	0.079^{***}	0.072^{**}	0.082^{***}
	(0.030)	(0.026)	(0.030)	(0.030)
Second-year arena	0.069^{**}	0.050^{**}	0.046^{*}	0.030
	(0.034)	(0.024)	(0.025)	(0.024)
Other international event	0.010	0.015^{**}	0.014	0.013^{*}
	(0.008)	(0.007)	(0.009)	(0.008)
Time Trend	Yes	Yes	No	No
Team FE	No	Yes	No	Yes
Team Time Trend	No	No	Yes	Yes
Log Likelihood	207	370	335	455
F-stat (t-1 vs. t-2 interaction)	1.79	4.62	4.06	1.49
Observations	422	422	422	422

Standard errors clustered by season in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

for teams located in the U.S., though the coefficient is not statistically different from zero.

Additionally, we may believe the number of Olympians on an NHL team affects the size of the attendance boost (e.g., an "all-star" effect). Table 2.12 provides the results of several models examining the impact of the number of Olympic NHL players per team on attendance. The first four columns include estimates of the impact during Olympic years. Column 1 presents the results from a simple regression of attendance on the number of Olympic players on a team's roster. Results indicate a statistically significant, positive impact on attendance. However, this measure is likely highly correlated with team quality and performance measures. Columns 2 through 4 include current and past team performance variables under specifications including temporal controls and team-level fixed effects. The coefficient is positive and stable across these specifications but is not significant. Finally, we expand the data set to include all years from 1995 to 2015, similar to the analysis in the extension above. Results indicate that each additional Olympian increases average season attendance by about 0.3% in Olympic seasons compared to non-Olympic seasons. During this time frame teams averaged roughly 4 to 5 Olympians, which equates to about a 1.5% boost in overall attendance in seasons when players were allowed to participate in the Winter Olympics. This result parallels the findings in Table 2.7.

2.5 Conclusions

The NHL has decided not to participate in the next Winter Olympics in Pyeongchang, South Korea in 2018. As mentioned above, concerns from league officials and owners regarding the consequences of an elongated break in the middle of the season with respect to overall costs and player health and performance seem to have weighed heavily on this decision. Prior studies have examined some of these costs to participation, however, there are no such studies on the benefits of NHL involvement in the Olympics regarding increased exposure and popularity. We examine the impacts of Olympic participation on attendance in the NHL. We find evidence that NHL participation in the Olympic Games is associated with an increase in league attendance of approximately 4.2%, equating to approximately 673 additional tickets sold per game. The league's

Table 2.11: Censored Regression Results by Country (1987-2004 seasons)

	USA only	CAN only	Interaction	
	(1)	(2)	(3)	(4)
Olympics	-0.028^{***}	-0.047^{***}	-0.033^{*}	-0.035^{*}
	(0.007)	(0.017)	(0.019)	(0.019)
NHL Participation	-0.015	0.058^{*}	0.037^{*}	0.043
	(0.009)	(0.031)	(0.019)	(0.029)
NHL Participation × Olympics	0.048^{***}	0.031^{*}	0.023	0.021
	(0.008)	(0.018)	(0.022)	(0.024)
USA city			0.111^{**}	-0.063^{***}
			(0.047)	(0.024)
USA × Olympics			0.007	0.005
			(0.024)	(0.023)
USA × NHL Participation			-0.048^{**}	-0.073^{*}
			(0.020)	(0.038)
USA × NHL Part. × Olympics			0.022	0.037
			(0.026)	(0.027)
Points percentage	0.436^{***}	0.041	0.330^{***}	0.385^{***}
	(0.100)	(0.075)	(0.076)	(0.085)
Playoff team	-0.033^{**}	0.040^{**}	-0.015	-0.028^{*}
	(0.016)	(0.017)	(0.013)	(0.015)
Prior season points percentage	0.200^{**}	0.182	0.179^{***}	0.146^{*}
	(0.098)	(0.114)	(0.063)	(0.078)
Prior season playoff team	0.040^{**}	0.055^{**}	0.042^{***}	0.043^{**}
	(0.020)	(0.021)	(0.014)	(0.018)
Prior season champions	-0.002	0.028	0.011	0.013
	(0.016)	(0.029)	(0.013)	(0.011)
Prior season runners up	0.061^{*}	-0.114^{***}	0.038	0.058^{**}
	(0.031)	(0.024)	(0.029)	(0.025)
Log(arena capacity)	0.531^{***}	0.461^{**}	0.488^{***}	0.552^{***}
	(0.073)	(0.204)	(0.063)	(0.071)
First-year arena	0.081^{***}	0.046	0.082^{***}	0.073^{**}
	(0.030)	(0.030)	(0.027)	(0.032)
Second-year arena	0.056^{**}	0.030	0.054^{**}	0.050^{*}
	(0.024)	(0.029)	(0.024)	(0.027)
Other international event	0.013^{*}	0.018	0.011^{*}	0.009
	(0.007)	(0.018)	(0.006)	(0.008)
Time Trend	Yes	Yes	Yes	No
Team FE	Yes	Yes	Yes	No
Team Time Trend	No	No	No	Yes
Log Likelihood	270	114	371	337
Observations	309	113	422	422

Standard errors clustered by season in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

Table 2.12: Effect of Olympic Player Quantity on Attendance

	Olympic years				1995-2014
	(1)	(2)	(3)	(4)	(5)
No. of Olympic players	0.028***	0.007	0.008	0.009	0.003*
	(0.008)	(0.008)	(0.007)	(0.006)	(0.001)
Playoff team		0.005	0.031	−0.037	−0.002
		(0.033)	(0.038)	(0.033)	(0.016)
Points percentage		0.105	−0.050	0.270	0.271***
		(0.248)	(0.289)	(0.220)	(0.095)
Prior season points percentage		0.473**	0.314	0.086	0.296***
		(0.206)	(0.232)	(0.151)	(0.099)
Prior season playoff team		0.042	0.069*	0.048	0.016
		(0.027)	(0.038)	(0.031)	(0.014)
Prior season champions		0.110*	0.125*	0.140***	0.093***
		(0.061)	(0.066)	(0.043)	(0.023)
Prior season runners up		0.049	0.060	0.018	0.052**
		(0.086)	(0.088)	(0.061)	(0.022)
Log(arena capacity)					0.223**
					(0.103)
First-year arena					0.058
					(0.037)
Second-year arena					0.055
					(0.035)
Other international event					0.023***
					(0.008)
Year FE	No	No	Yes	Yes	No
Team FE	No	No	No	Yes	Yes
Team Time Trend	No	No	No	No	Yes
Log Likelihood	16	29	33	104	318
Observations	145	145	145	145	545

Standard errors clustered by season in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

decision to not participate in the 2018 games may mean team owners are forgoing millions of dollars in additional gate revenue.

Chapter 3

The Value of Neighborhood Walkability

3.1 Introduction

In an attempt to revitalize urban areas, cities across the country have introduced more mixed use developments. This New Urbanism movement has led to the creation of new communities composed of a mixture of residential and commercial properties. Proponents of this movement argue that households want to live in pedestrian-friendly areas. More recently, a major emphasis has been placed on the walkability of the neighborhood and how close residents live to shopping and social activities. In fact, a recent survey by the National Association of Realtors found that a majority of Americans wanted homes and businesses to "be built closer together, so that stores and shops are within walking distance and don't require the use of an automobile" (Ulm, 2007). If homeowners truly value the walkability of a location we should see this capitalized in housing prices. More walkable areas should result in higher prices compared to similar homes in less walkable areas. In this paper, I estimate the walkability premium using a hedonic housing model that also controls for both observed and unobserved locational characteristics.

The concept of walkable neighborhoods is not new, urban planners dating back to the 1960s have argued that communities should be designed to be pedestrian-friendly. A major component of these walking oriented developments is the accessibility of desired destinations. Studies have shown that locations with more amenities lead to more walking behavior. Coupled with greater access to public transit, this also leads to less overall driving (Handy **et al.**, 2005). Research has also shown several other benefits to pedestrian-oriented neighborhoods including improvements to the environment (Frank & Engelke, 2005) and better overall health. Residents who can walk to nearby amenities experience lower body weight and obesity rates (Doyle **et al.**, 2006) and lower body mass index (Brown **et al.**, 2009). Du Toit **et al.** (2007) find that more walkable neighborhoods also lead to more social interactions and a better sense of community. These areas are also associated with lower crime rates (Foster **et al.**, 2011). These vast benefits suggest land-use policies should be designed to improve walkability in neighborhood developments. With all these benefits, aside from simply living close to local amenities, we would think consumers would be willing to pay a premium to live in more walkable neighborhoods. Estimating this value has implications for the amount of public investment allocated to creating more of these types of developments. Some of the public costs could also be offset by increases in property tax revenue generated from these locations.

A common approach to valuing locational characteristics is through the use of a hedonic model. The hedonic model applied to housing attempts to assign a value to the individual characteristics of a home. These characteristics can be either structural or locational. The hedonic approach has been used extensively to study the impacts of transit-oriented development on housing values. As Bowes & Ihlanfeldt (2001) point out, the value of public transit could be either positive or negative, depending on whether the benefits of the commuting alternative outweigh the added noise and congestion. The literature, in general, finds an overall positive effect of increased transit options on property values, particularly for rail lines (Hess & Almeida, 2007; Billings, 2011). However, there is some evidence to suggest that the property value impact is significantly reduced within the immediate area of a transit stop due to increases in noise, congestion, crime, and other nuisance effects (Diaz, 1999). These can also be areas of concern for pedestrian-oriented developments.

The research on whether walkability increases house prices is mixed. In one of the early studies on the New Urbanism movement, Tu & Eppli (2001) find that consumers pay anywhere from a 4-15% premium to live in these new communities compared to more conventional developments. However, they are unable to truly disentangle the effect of walkability since they only identify whether or not a property is located in one of these neighborhoods.[1] They do not have a specific measure of

[1]The premium most likely incorporates walkability, as well as, other location characteristics associated with

walkability and only capture the overall premium associated with these new developments.

To be able to measure walkability, it is important to first define the concept. One way to define walkability could be based on the quality or suitability of walking available, taking into consideration factors that influence the pleasantness of the walk. For example, the street congestion, the ability to cross intersections, the presence of trees, safety, and other pedestrian services. However, as the previous literature mentions, accessibility factors play a much larger role in influencing walking behavior. These factors include distance to desired destinations, land-use, street connectivity, and density measures. All of these factors were influential in the creation of Walk Score. Walk Score measures the walkability of an address by analyzing the distance to a number of local amenities. The score has been shown to be an appropriate indicator of walkability and walking behavior and has been used in previous studies on walkability.

Using the Walk Score metric, Pivo & Fisher (2011) find that walkability leads to higher office and retail property values. This suggests more walkable neighborhoods lead to more business. Of course, the value of a walkable neighborhood could be different for commercial versus residential properties. Cortright (2009) analyzes the relationship between Walk Score and home values in 15 different markets in the United States. His research indicates a significant positive correlation between Walk Score and house prices. However, his analysis is done at the Metropolitan Statistical Area level and the specification only controls for differences in income levels, available jobs, and distance to the central business district. There could be several additional locational factors associated with both higher Walk Scores and higher prices that are unaccounted for, leading to biased estimates. For example, both the amount of nearby open space (Irwin, 2002) and the quality of the local school (Black, 1999) have been shown to lead to higher house prices. Using a similar approach, Rauterkus & Miller (2011) find that land values generally increase with Walk Score. However, again, they fail to account for any confounding locational factors associated with Walk Scores that could potentially influence these land values.

In order to address this issue, Boyle **et al.** (2014) use a fixed effects approach to control

walkability.

for any unobserved heterogeneity in neighborhoods. They use fixed effects at the subdivision, township range section[2], and zip code levels and conclude that walkability's impact on housing values is statistically insignificant. Their data consist of 3,423 observations of homes in Miami, Florida. One reason they find statistical insignificance could be due to the lack of power from a low number of observations and lack of variation in Walk Score. This is particularly true at the subdivision and township range sections, where the mean number of observations per area is 2 and 12.8, respectively.[3] This could also explain why some other variables in the regressions, such as number of bathrooms, number of bedrooms, and distance to central business district are also insignificant.[4] The zip code level fixed effect model provides a lot more observations per neighborhood and more variation in Walk Scores within each area. Interestingly, the estimated coefficient on Walk Score moves from negative to positive when increasing the area of the fixed effect, although it continues to remain statistically insignificant. This could simply be because of the model specification and the need for more observations and\or more explanatory variables. Furthermore, their estimates could be the result of the use of the appraised value of the home as the dependent variable, rather than the actual sales price. If households truly value walkability we are more likely to see this in how much they actually pay for the house, the true market value, rather than what is determined by an appraiser. After all, the appraised value can often lag behind the market value for some years.

Instead, this paper uses the actual sales price of homes in Denver, Colorado to estimate the walkability premium, while also controlling for unobserved heterogeneity across locations through the use of fixed effects. The model includes several additional housing and locational variables. The results indicate that, on the whole, households are willing to pay more to live in more walkable areas. This price premium is largest at the top end of the Walk Score distribution. However, these effects diminish with the geographic size of the fixed effect. This could simply be due to the reduction in

[2]This is defined as a one square mile plot of land.

[3]The min and max observations per area suggest the median number of observations is even smaller.

[4]The literature generally finds that these variables significantly influence house prices (Sirmans **et al.**, 2006)

variation in Walk Scores at smaller geographic areas or, similar to the transit access literature, it could suggest there exists a larger negative impact from nuisance effects in close proximity to these destinations. While consumers value walkability, they may choose to avoid the main traffic areas, instead opting to live a block or two away from many of the local amenities. This way they are still able to walk to many amenities, while avoiding the increased traffic, noise pollution, and other negative externalities associated with these areas. Local governments and property developers must consider these nuisances when building these walkable developments.

3.2 Conceptual Framework

The model follows the standard hedonic theory developed by Lancaster (1966) and Rosen (1974). A house can be broken down into a bundle of attributes from which consumers derive utility. Consumers purchase the home that maximizes utility subject to their budget constraint. Assuming the supply of housing is fixed at any given point in time, we can estimate the implicit price of each housing characteristic using regression techniques. The hedonic pricing function for homes is given by:

$$\mathbf{Price} = \mathbf{f}(\mathbf{Structural\ Characteristics},\ \mathbf{Locational\ Characteristics}) \tag{3.1}$$

This equation states that the price of a house is a function of its structural characteristics (square footage, bedrooms, bathrooms, age, etc.) and local or neighborhood characteristics (school quality, open space, crime, walkability, etc.). The regression of house price on these characteristics provides an estimate of the implicit price or marginal willingness to pay for each individual attribute.

The specific functional form of the model estimated in this paper is given by:

$$P_i = X_i\gamma + Z_i\delta + \beta WalkScore_i + \epsilon_i \tag{3.2}$$

where P_i is the log sales price of house i. The correct functional form of the hedonic pricing function is debatable. The general consensus is to use a semi-log form with the log of house price as the dependent variable. This allows us to interpret the coefficients as the percent change in house price due to a unit change in the attribute. As Malpezzi **et al.** (1980) point out the semi-log form allows the dollar price of each characteristic to vary with price and helps to minimize the issue of heteroskedasticity. The vector, X_i, consists of several structural housing characteristics. In a meta-analysis on hedonic housing models, Sirmans **et al.** (2006) found that the most common housing characteristics that influence price are square footage, lot size, age, bedrooms, bathrooms, garage, a swimming pool, a fireplace, and air conditioning. Their research also shows that after controlling for time, the effects of housing characteristics on price remain relatively stable. I use this as a basis to determine which variables to include in my model, making slight adjustments to account for the specific housing environment in Denver, Colorado.[5] Table 3.1 provides a description of all variables included in the regression. I also include year and quarter of year fixed effects to account for any time and seasonal differences in average house prices as the sample spans a few years.

It is reasonable to assume that more walkable neighborhoods are also associated with other locational characteristics and excluding them could bias the estimated effect of walkability on house prices. Therefore, I incorporate several other measures of neighborhood characteristics as controls in the regression. These are captured in the vector, Z_i, and include the amount of nearby open space, the quality of the elementary school, the population density, the median income, and the distance to the central business district (CBD). All of these variables have been shown to influence housing values in previous research and are potentially correlated with Walk Score. Finally, location fixed effects are included to account for any unobserved differences in localities that could be associated with differences in Walk Scores. These location fixed effects vary from zip code to neighborhood to elementary school to census tract.

The variable of interest is the walkability of the neighborhood as determined by the Walk Score (www.walkscore.com). Walk Score is a measure that computes the walkability of the area and analyzes routes to nearby locations and overall pedestrian friendliness. In essence, this score provides a measure of how little a homeowner would be required to drive to complete daily errands.

[5] I do not include swimming pools, but add in whether the home is attached or detached.

Table 3.1: Description of Variables

Variable	**Definition**
Sale Price	Sale price (in 2015 dollars)
Bedrooms	Total number of bedrooms
Baths	Total number of bathrooms
Sq Ft	Square footage
Lot Size	Lot size (in square feet)
Detached	1 if detached home
Garage	1 if home has garage
Basement	1 if home has a basement
Fireplace	1 if home has a fireplace
Central AC	1 if home has central air conditioning
Age	Age of home
Age^2	Squared age of home
Walk Score	Score assessing walkability
Transit Score	Score assessing availability of public transit
Distance to CBD	Distance to central business district (in miles)
Open Space	Percent of open space within 1/2 mile of home
School Quality	3 year total performance score for elementary school
Pop Density	Number of people per square km ($100/km^2$)
Median Income	Median income of Census block group (in $1000s$)

They use a patented algorithm to generate a Walk Score for each address on a scale of 0-100. Points are awarded based on the distance a home is to amenities in several categories. The categories include restaurants and bars, schools, parks, grocery stores, coffee shops, shopping, and entertainment. Each category is weighted equally with maximum points given to amenities within a quarter mile. A decay function is used to determine points to amenities further away. Locations that require more than a 30 minute walk (1.5 miles) are awarded zero points. Additionally, the calculation takes into consideration land-use, street connectivity, block length, and intersection density.

The scores are grouped into categories. A score of 90-100 is considered a Walker's Paradise, 70-89 is Very Walkable, 50-69 is Somewhat Walkable, and 0-49 is Car Dependent. Figures 3.1-3.3 provide a look at the data available for particular locations on walkscore.com. The addresses are all from the Berkeley neighborhood in Denver and demonstrate the differences in Walk Score categories. The maps provide a look at the various amenities located near an address. Additional information includes commute times, Bike Score, and Transit Score. The Transit Score measures the availability of public transit for a particular location on a similar scale of 0 to 100. The algorithm takes into consideration the distance to nearby transit routes, the 'usefulness' of nearby routes, the frequency of service, as well as, the type of service. Similar to how Walk Score can provide a measure of how necessary a car is to complete daily tasks, the Transit Score provides a measure of necessary a car may be for commuting.

Of course, Walk Score is not without its criticisms. The measure weighs each category equally suggesting households value each type of amenity the same. Also, the tool does not take into account the quality of each amenity nor the density of each type of amenity. The extent to which households value categories differently or whether there exists significant variation in the quality of amenities across ranges of Walk Scores is not known. Nevertheless, Walk Score seems to provide a reasonable measure of the walkability of a neighborhood.

Several studies have been conducted to test the validity and reliability of Walk Score as an appropriate measure of walkability. Carr **et al.** (2010) calculate distances to over 4,000 walkable

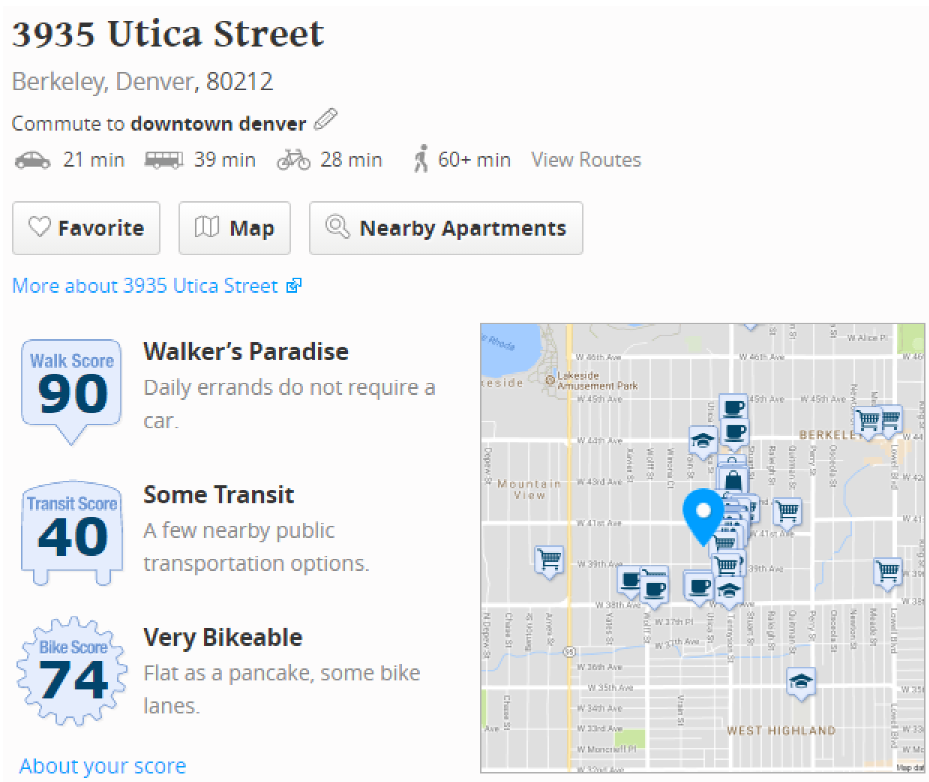


Figure 3.1: Walk Score Example - Walker's Paradise

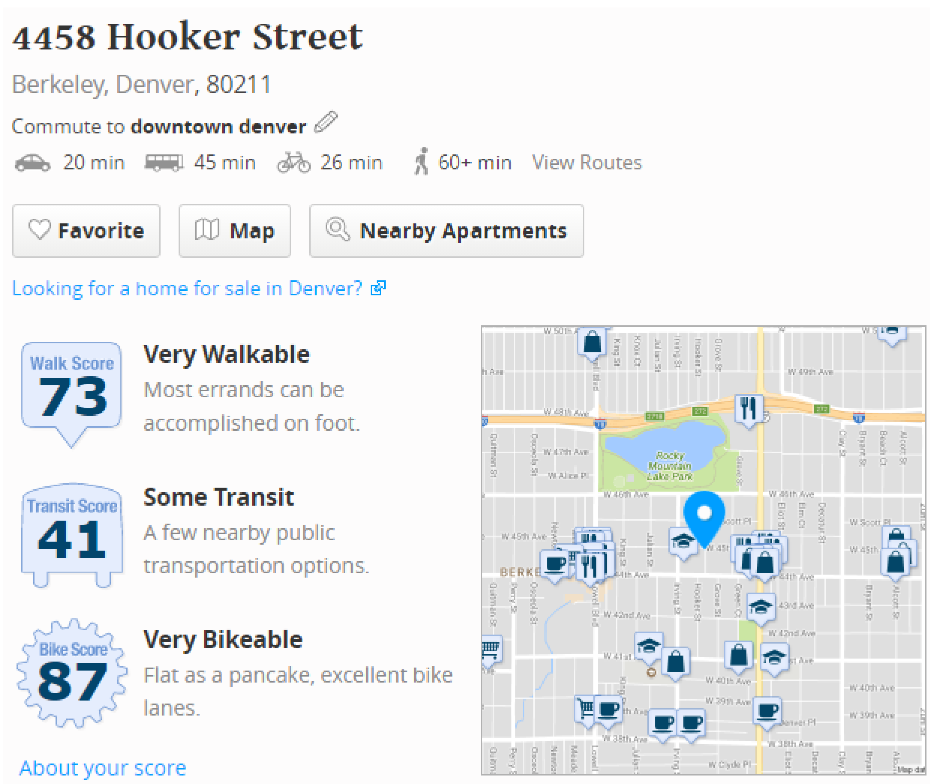


Figure 3.2: Walk Score Example - Very Walkable

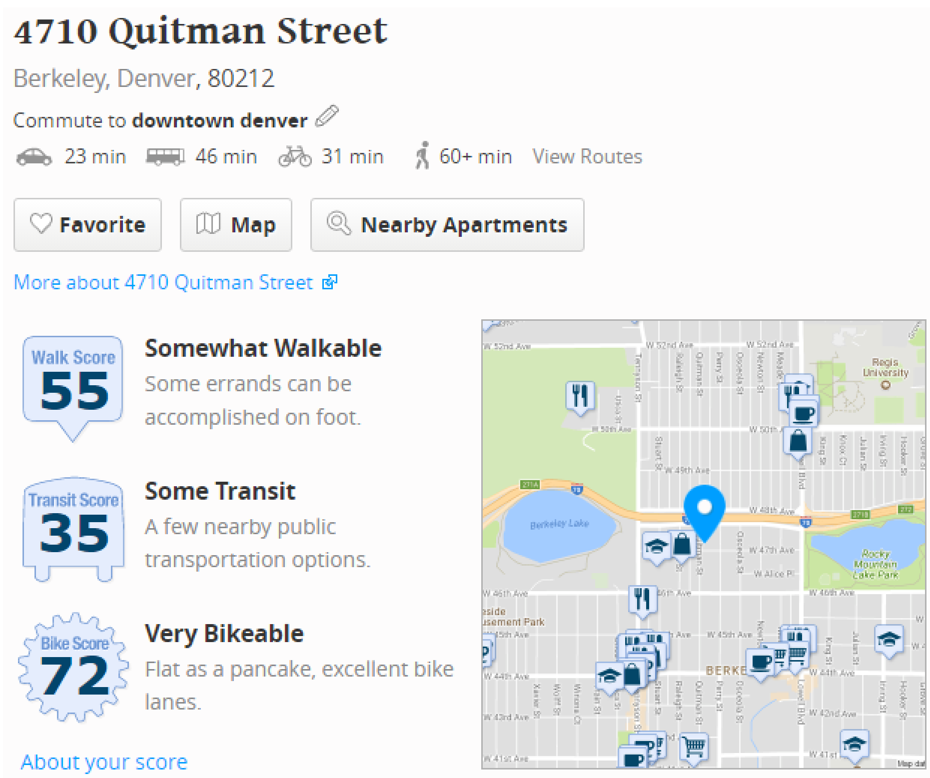


Figure 3.3: Walk Score Example - Somewhat Walkable

amenities in 13 different categories for homes in Rhode Island. They find a significant correlation between Walk Score and the number of amenities in each category within a 1 mile buffer from each location. In a similar study, Duncan **et al.** (2011) calculate the density of walkable amenities for several categories by counting the total number of destinations within several buffer distances to an address and compare that to the Walk Score. They do this for four different metropolitan areas across the United States at several buffer lengths up to roughly 1 mile. They too find significant correlations between their neighborhood walkability measures and Walk Score and conclude that "Walk Score is a valid measure of estimating neighborhood walkability" for several locations at various spatial distances. Subsequent studies have also confirmed the use of Walk Score as a valid measure of neighborhood walkability (Duncan **et al.**, 2013; Manaugh & El-Geneidy, 2011).

3.2.1 Data

The sample consists of over 28,000 single family home transactions between 2012 and 2015 for the city and county of Denver. The housing data come from the multiple listing service (MLS) and include several housing characteristics, as well as, the Walk Score at the time of sale. Table 3.2 provides the overall summary statistics. The average home in the sample sold for roughly $365,000 and consisted of 3 bedrooms, 2 bathrooms, and 1,550 square feet of living space. The majority of homes were detached and included a garage, a basement, and a fireplace. The average Walk Score for a residence was about 58 and ranged from as low as 1, meaning a car was required to complete any sort of daily task, to as high as 98, suggesting the location was a 'walkers paradise.' Figure 3.4 provides a plot of the full distribution of Walk Scores in the sample, while Figure 3.5 shows a heat map of Walk Scores throughout Denver.

Table 3.2: Summary Statistics

Variable	**Mean**	**Std. Dev.**	**Min.**	**Max.**	**Walk Score Correlation**
Sale Price	365,355	268,700	25,246	4,839,042	0.3257
Bedrooms	3.25	0.96	1	9	-0.1019
Baths	2.37	1.11	1	8	-0.1001
Sq Ft	1,550	758	600	8,894	-0.0496
Lot Size	7,321	25,436	1,000	982,651	-0.0062
Detached	0.84		0	1	-0.0702
Garage	0.72		0	1	-0.0917
Basement	0.73		0	1	0.2049
Fireplace	0.55		0	1	0.0713
Central AC	0.46		0	1	-0.0942
Age	61	35	0	140	0.5218
Age^2	4,952	4,416	0	19,600	-
Walk Score	57.87	19.42	1	98	-
Transit Score	44.84	10.27	0	96	0.5785
Distance to CBD	4.40	2.92	0	13.68	-0.6252
Open Space	7.64	6.62	0	52.65	-0.1798
School Quality	64.86	18.32	30.60	99	0.0564
Pop Density	31.09	18.74	0.26	1,001.25	0.0698
Median Income	67.63	33.72	8.27	210.15	-0.0770

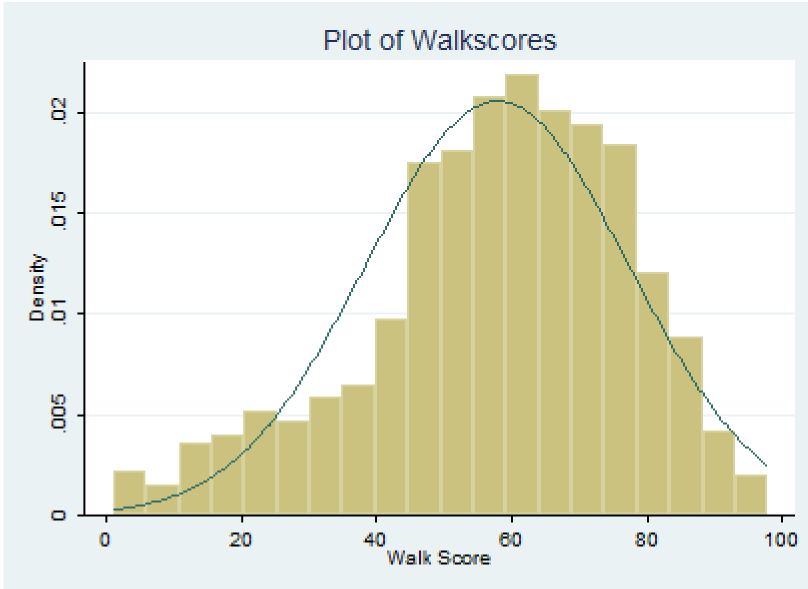


Figure 3.4: Walk Scores

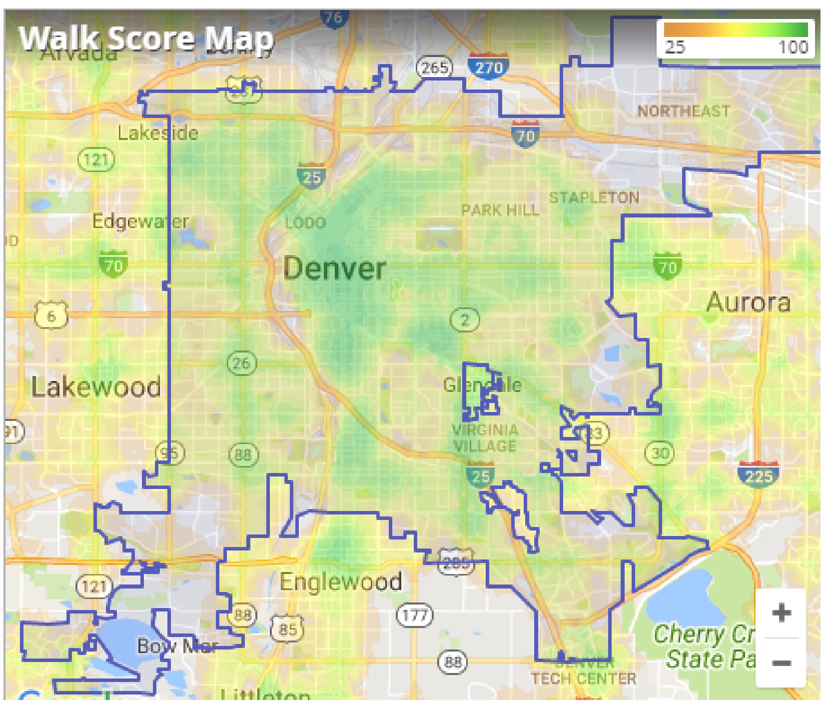


Figure 3.5: Map of Denver Walk Scores

Apart from Walk Score, several other neighborhood variables were also collected. Geographic information systems (GIS) software was used to calculate the distance of each address to the central business district. The CBD is the commercial core and center of commerce and economic activity for Denver, Colorado. The area defining the CBD comes from the community planning and development department for Denver. GIS software was also used to calculate the percent of open space within one mile of each home. Data on open space come from Colorado Ownership, Management and Protection (COMAP) version 8 developed by the United States Geological Survey and acquired from Colorado State University.

The variable used to measure academic quality is a comprehensive indicator developed by the Colorado Department of Education. The State of Colorado has developed a school performance framework which seeks to “hold districts and schools accountable for performance on the same, single set of indicators and measures.” Attainment is measured in four major categories: academic achievement, academic growth, academic gaps, and postsecondary and workforce readiness. The variable used in this study is a 3 year total percentage constructed by aggregating each of the categories and dividing by the total possible points for each elementary school within Denver. Each address was assigned an elementary school using the shapefile from the School Attendance Boundary Survey. The survey is conducted by the National Center for Education Statistics who define a catchment area as the “geographic area from which students are eligible to attend a local school”. Altogether, the sample data consist of 78 elementary schools with an average 3 year total achievement of 65%.

Demographic and socioeconomic data come from the US Census and Colorado State Demography Office. Each house is geo-coded using the Census Geocoder, which assigns a census block and also provides data on housing units, population, and land area. This data are used to calculate housing and population density at the block level. Each home is also matched to demographic data from the American Community Survey at the census block group level. This includes information on the median income level of the block group.

Table 3.2 also presents the correlations between Walk Score and all other variables. Higher Walk Scores are correlated with higher home prices. They are also correlated with higher Transit Scores, which is to be expected as areas with many accessible amenities are likely to have more transit options available. Also, as expected, there is a significant negative correlation between walkability and distance to CBD. We would expect the proximity of amenities to decrease the further we get from the city center. Interestingly, there is relatively strong positive correlation between Walk Score and the age of the home. This makes sense if we believe residential growth spurs commercial growth and increases in public infrastructure.

The regression analysis is done at several geographic fixed effects levels. Table 3.3 provides a summary of these levels including the number of areas in the sample, the average land size, and the average standard deviation in Walk Score at each level. The areas range from as large as zip code to as small as census tract. Since amenities within a quarter mile are given the maximum number of points, it is important to ensure that the fixed effect areas are large enough to maintain necessary variation in Walk Scores. For example, the average census tract has an area of 1.21 square kilometers. If the typical census tract in the data where a square plot of land each side would be roughly .7 miles long, ensuring an area large enough to have reasonable variation in Walk Scores.[6] While the average variation in Walk Score for a census tract is a little less than half of the total sample, I contend it is still reasonably large enough to identify any price effect. The average range of scores for a typical census tract is 45 points and the typical census tract contains over 200 observations.

Table 3.3: Summary of Geographic Fixed Effects

Geographic Level	**Count**	**Mean # of Obs**	**Avg Land Size (Sq Km)**	**Walk Score St. Dev.**
Zip Code	47	616	3.63	12.10
Neighborhood	76	381	2.24	10.07
Elem School	78	371	2.19	10.10
Census Tract	141	205	1.21	8.92

[6]The average area of a census block or block group is much smaller resulting in insufficient variation in Walk Scores

3.3 Results

Early studies using Walk Score to estimate the effect of walkability on housing values have found a positive relationship between 1%-10% for each 10 point increase in Walk Score(Cortright, 2009; Gilderbloom **et al.**, 2015). Using a similar ordinary least squares methodology, I find that in Denver each 10 point increase in Walk Score leads to an increase in the price of a home of 6%. This result can be found in column 1 of Table 3.4. However, as Boyle **et al.** (2014) point out, this estimate can be significantly biased because it does not take into account the heterogeneity of neighborhoods. For this reason, I estimate the effect of Walk Score on sales price using fixed effects at varying geographic levels. These include zip code, neighborhood, elementary school, and census tract. Even though I include several other neighborhood characteristics as controls, location fixed effects account for any unobserved differences in neighborhoods that could potentially bias the results.

Columns 2-5 of Table 3.4 provide results of the regressions with fixed effects. The coefficient on Walk Score decreases significantly once fixed effects are added, but remains positive and significant. The results indicate that households are willing to pay around a 1-2% premium or about $3,000-$7,000 for each 10 point increase in Walk Score, or to essentially move up a tier in Walk Score's categorization of walkability. Interestingly, the coefficient on Walk Score becomes smaller with the geographic size of the fixed effect. This could simply be due to the reduction in variation in Walk Scores within smaller geographic areas, particularly the census tract, or it could suggest that walkability is less important to households when making location decisions within a small defined area. It could also be the case that while households do value the opportunity to walk to nearby amenities, they also dislike noise, traffic, and other negative factors associated with these nearby amenities. These two opposing effects indicate that households may want to be close to amenities, but not too close. They may prefer to live a few blocks from the main traffic areas of a location. The extra distance required to walk to desired destinations may not preclude walking behavior. The coefficient on population density reinforces this idea suggesting that households would prefer to live in less populated areas, possibly to avoid pedestrian congestion and pollution, all else equal.

To further test this hypothesis, I include Transit Score in the regressions presented in Table 3.5. Transit Score measures the availability of public transit for a particular location. Higher scores indicate proximity to transit stops and more frequent routes. Higher score locations are more likely to be on main roads and traffic areas. If households truly dislike noise, congestion, and pollution, we should expect higher Transit Scores to lead to a lower price. This is in fact what is observed in the regressions in Table 3.5. A 10 point increase in Transit Score leads to a 1-2% decrease in the value of the home. This provides additional evidence for the idea that while households value walkable access to amenities, there also exist negative factors associated with more walkable areas that need to be taken into consideration. Consumers want to live close, but not too close to amenities and services. Households may value the accessibility but would rather not live in front of a bus stop or have to deal with the noise associated with living too close to a train stop. The largest benefit may come from creating walkable developments that limit noise and congestion problems.

The coefficients on the structural characteristics in Tables 3.4 and 3.5 all have the predicted sign and for the most part remain fairly stable at the various fixed effect levels. This is to be expected as the features of the house should be uncorrelated with any unobserved differences in the neighborhood. The coefficients on the usual housing attributes are similar in magnitude and direction to those found in the literature (Sirmans **et al.**, 2006). Apart from Walk Score, distance to the central business district, percent of open space, the quality of the elementary school, population density and the income of the area all significantly affect the value of the home. Each mile away from the CBD leads to a decrease in price by about 6-7%. The coefficient on elementary school quality remains positive and significant, although there is some fluctuation in the point estimate at different fixed effect levels. The coefficient estimates on population density suggests homeowners are willing to pay more to live in less dense areas and the estimates for income levels suggest more affluent neighborhoods have higher house prices.

Table 3.4: House Price Regressions on Walk Score

	(1)	(2)	(3)	(4)	(5)
Bedrooms	0.0243***	0.0344***	0.0377***	0.0346***	0.0376***
	(0.0026)	(0.0023)	(0.0021)	(0.0022)	(0.0021)
Baths	0.0949***	0.0864***	0.0824***	0.0844***	0.0818***
	(0.0032)	(0.0028)	(0.0026)	(0.0027)	(0.0025)
Sq Ft	0.0002***	0.0002***	0.0002***	0.0002***	0.0002***
	(5.1e-06)	(4.6e-06)	(4.3e-06)	(4.3e-06)	(4.3e-06)
Lot Size	6.2e-10***	8.8e-10***	9.0e-10***	7.3e-10***	7.2e-10***
	(7.5e-11)	(7.9e-11)	(7.7e-11)	(8.7e-11)	(8.2e-11)
Detached	0.1295***	0.1592***	0.1741***	0.1658***	0.1781***
	(0.0108)	(0.0085)	(0.0085)	(0.0086)	(0.0081)
Garage	0.0943***	0.0674***	0.0625***	0.0654***	0.0630***
	(0.0044)	(0.0037)	(0.0034)	(0.0035)	(0.0034)
Basement	0.1145***	0.1022***	0.0984***	0.0945***	0.0950***
	(0.0058)	(0.0049)	(0.0044)	(0.0045)	(0.0042)
Fireplace	0.1057***	0.0845***	0.0634***	0.0675***	0.0607***
	(0.0046)	(0.0039)	(0.0035)	(0.0036)	(0.0034)
Central AC	0.1179***	0.0895***	0.0788***	0.0842***	0.0774***
	(0.0038)	(0.0032)	(0.0029)	(0.0030)	(0.0028)
Age	-0.0020***	-0.0018***	-0.0010***	-0.0008***	-0.0009***
	(0.0004)	(0.0003)	(0.0003)	(0.0003)	(0.0003)
Age^2	-2.0e-03***	-1.8e-03***	-1.0e-03***	-8.4e-04***	-9.3e-04***
	(4.0e-04)	(2.9e-04)	(2.7e-04)	(2.8e-04)	(2.6e-04)
Walk Score	0.0058***	0.0022***	0.0009***	0.0014***	0.0004**
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Distance to CBD	-0.0528***	-0.0589***	-0.0567***	-0.0784***	-0.0671***
	(0.0014)	(0.0038)	(0.0046)	(0.0054)	(0.0068)
Open Space	0.0025***	0.0023***	0.0005	0.0020***	0.0014***
	(0.0004)	(0.0004)	(0.0004)	(0.0005)	(0.0005)
School Quality	0.0042***	0.0025***	0.0003	(.)	0.0011***
	(0.0002)	(0.0002)	(0.0002)	(.)	(0.0003)
Pop Density	-0.0020***	-0.0013***	-0.0011***	-0.0010***	-0.0011***
	(0.0006)	(0.0002)	(0.0001)	(0.0003)	(0.0001)
Median Income	0.0034***	0.0024***	0.0012***	0.0019***	0.0010***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Constant	10.8777***	11.3134***	11.6149***	11.6454***	11.6470***
	(0.0403)	(0.0320)	(0.0333)	(0.0347)	(0.0412)
Observations	28955	28955	28955	28955	28955
Adj r^2	0.8132	0.8630	0.8827	0.8759	0.8865
FE Level	-	Zip Code	Neighborhood	Elem School	Census Tract

Note: All regressions include year and quarter fixed effects.

Bootstrapped clustered standard errors at the census block level in parenthesis.

* p<0.1, ** p<0.05, *** p<0.01.

Table 3.5: House Price Regressions on Walk & Transit Scores

	(1)	(2)	(3)	(4)	(5)
Bedrooms	0.0244***	0.0344***	0.0378***	0.0348***	0.0377***
	(0.0026)	(0.0023)	(0.0021)	(0.0021)	(0.0020)
Baths	0.0947***	0.0862***	0.0823***	0.0839***	0.0815***
	(0.0032)	(0.0028)	(0.0026)	(0.0027)	(0.0025)
Sq Ft	0.0002***	0.0002***	0.0002***	0.0002***	0.0002***
	(5.1e-06)	(4.6e-06)	(4.3e-06)	(4.3e-06)	(4.2e-06)
Lot Size	6.2e-10***	8.8e-10***	8.8e-10***	7.1e-10***	7.0e-10***
	(7.5e-11)	(8.0e-11)	(7.7e-11)	(8.7e-11)	(8.1e-11)
Detached	0.1282***	0.1575***	0.1732***	0.1640***	0.1769***
	(0.0108)	(0.0085)	(0.0085)	(0.0085)	(0.0081)
Garage	0.0941***	0.0673***	0.0624***	0.0651***	0.0629***
	(0.0044)	(0.0037)	(0.0034)	(0.0035)	(0.0033)
Basement	0.1148***	0.1027***	0.0989***	0.0954***	0.0958***
	(0.0058)	(0.0048)	(0.0044)	(0.0045)	(0.0042)
Fireplace	0.1060***	0.0847***	0.0632***	0.0670***	0.0603***
	(0.0046)	(0.0039)	(0.0035)	(0.0036)	(0.0034)
Central AC	0.1179***	0.0893***	0.0786***	0.0839***	0.0772***
	(0.0038)	(0.0032)	(0.0029)	(0.0030)	(0.0028)
Age	-0.0021***	-0.0019***	-0.0010***	-0.0009***	-0.0009***
	(0.0004)	(0.0003)	(0.0003)	(0.0003)	(0.0003)
Age^2	2.0e-05***	1.3e-05***	2.8e-06	1.3e-06	1.1e-06
	(3.3e-06)	(2.2e-06)	(2.1e-06)	(2.3e-06)	(2.0e-06)
Walk Score	0.0059***	0.0023***	0.0011***	0.0016***	0.0005***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Transit Score	-0.0008**	-0.0010***	-0.0016***	-0.0023***	-0.0019***
	(0.0004)	(0.0003)	(0.0003)	(0.0003)	(0.0003)
Distance to CBD	-0.0534***	-0.0610***	-0.0598***	-0.0862***	-0.0719***
	(0.0014)	(0.0040)	(0.0047)	(0.0056)	(0.0068)
Open Space	0.0025***	0.0022***	0.0001	0.0016***	0.0009**
	(0.0004)	(0.0004)	(0.0004)	(0.0005)	(0.0005)
School Quality	0.0042***	0.0025***	0.0003	(.)	0.0011***
	(0.0002)	(0.0002)	(0.0002)	(.)	(0.0003)
Pop Density	-0.0020***	-0.0013***	-0.0011***	-0.0010***	-0.0011***
	(0.0006)	(0.0002)	(0.0001)	(0.0003)	(0.0001)
Median Income	0.0034***	0.0023***	0.0012***	0.0018***	0.0010***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Constant	10.9060***	11.3659***	11.6940***	11.7834***	11.7525***
	(0.0443)	(0.0384)	(0.0376)	(0.0406)	(0.0451)
Observations	28955	28955	28955	28955	28955
Adj r^2	0.8133	0.8631	0.8829	0.8764	0.8868
FE Level	-	Zip Code	Neighborhood	Elem School	Census Tract

Note: All regressions include year and quarter fixed effects.
Bootstrapped clustered standard errors at the census block level in parenthesis.
* p<0.1, ** p<0.05, *** p<0.01.

3.3.1 Extensions

The effect of walkability on house prices may not be the same throughout the distribution of Walk Scores. More flexible forms of the Walk Score variable can help us better understand its effect. Table 3.6 presents the results from the regression of house prices on Walk Score splines. Walk Score is cut into 5 linear splines of equal length with knots at 20, 40, 60, and 80. The splines reveal some heterogeneous effects across the range of Walk Scores. Scores increasing from 0-20 actually have a negative effect on Walk Score. This may be because the added congestion outweighs any benefits of walkability. Walk Scores from 20-60 have a smaller positive marginal impact on house prices, while the largest positive effects are seen for the highest values of Walk Scores. In fact, the marginal effect for the highest scores are roughly two times greater than the average marginal effect estimated in Table 3.5. Each point increase in Walk Score from 80-100 leads to a .2% increase in the sale price of the home, suggesting that the largest impact comes from moving from a 'very walkable' area to a 'walker's paradise'.

Table 3.7 provides the results of regressions with Walk Score bins. Scores are grouped into 5 equal bins from 0-100. The 0-20 bin is excluded for estimation purposes. Similar to the spline regressions, an increase in Walk Score at the lower end may actually have a negative effect on price. The higher range of the walkability scores significantly impact house prices. Both the coefficients on the Walk Score 60-80 bin and 80-100 bin are statistically significant at all geographic levels except for the census tract. While the point estimates are as expected, the standard errors are relatively much larger, which is likely due to a lack of variation when scores are grouped together. The coefficient for the highest Walk Score bin across all specifications is slightly lower than what would be expected based on the constant marginal effect estimated in Table 3.5. For example, with zip code fixed effects moving from a score of 10 to a score of 90 leads to an increase in price of 11.2%. A similar increase using the unit effect estimated in Table 3.5 would lead to an increase in price of 18%.

Another interesting question is whether the price premium on walkability varies across the

Table 3.6: House Price Regressions on Walk Score Splines

	(1)	(2)	(3)	(4)
Bedrooms	0.0349***	0.0340***	0.0349***	0.0378***
	(0.0023)	(0.0021)	(0.0021)	(0.0020)
Baths	0.0866***	0.0840***	0.0840***	0.0815***
	(0.0028)	(0.0026)	(0.0027)	(0.0025)
Sq Ft	0.0002***	0.0002***	0.0002***	0.0002***
	(4.5e-06)	(4.3e-06)	(4.3e-06)	(4.2e-06)
Lot Size	9.1e-10***	8.5e-10***	6.9e-10***	6.9e-10***
	(8.3e-11)	(8.6e-11)	(8.9e-11)	(8.1e-11)
Detached	0.1628***	0.1699***	0.1655***	0.1769***
	(0.0084)	(0.0096)	(0.0085)	(0.0081)
Garage	0.0681***	0.0570***	0.0651***	0.0630***
	(0.0037)	(0.0035)	(0.0035)	(0.0033)
Basement	0.1020***	0.0984***	0.0949***	0.0960***
	(0.0048)	(0.0045)	(0.0045)	(0.0042)
Fireplace	0.0843***	0.0643***	0.0673***	0.0608***
	(0.0038)	(0.0035)	(0.0036)	(0.0034)
Central AC	0.0881***	0.0792***	0.0836***	0.0771***
	(0.0031)	(0.0029)	(0.0030)	(0.0028)
Age	-0.0014***	-0.0010***	-0.0008***	-0.0009***
	(0.0003)	(0.0003)	(0.0003)	(0.0003)
Age^2	7.7e-06***	2.4e-06	1.3e-07	6.5e-07
	(2.1e-06)	(2.1e-06)	(2.2e-06)	(2.0e-06)
Walk Score1	-0.0033***	-0.0036***	-0.0033***	-0.0033***
	(0.0009)	(0.0008)	(0.0010)	(0.0010)
Walk Score2	0.0010*	0.0017***	0.0014**	0.0012**
	(0.0006)	(0.0005)	(0.0006)	(0.0006)
Walk Score3	0.0008**	0.0006*	0.0009**	0.0001
	(0.0004)	(0.0003)	(0.0004)	(0.0004)
Walk Score4	0.0046***	0.0019***	0.0024***	0.0009**
	(0.0004)	(0.0004)	(0.0004)	(0.0004)
Walk Score5	0.0055***	0.0020**	0.0022**	0.0020**
	(0.0010)	(0.0010)	(0.0010)	(0.0010)
Transit Score	-0.0011***	-0.0016***	-0.0023***	-0.0019***
	(0.0003)	(0.0003)	(0.0003)	(0.0003)
Distance to CBD	-0.0635***	-0.0625***	-0.0874***	-0.0746***
	(0.0040)	(0.0048)	(0.0056)	(0.0070)
Open Space	0.0023***	0.0003	0.0016***	0.0010**
	(0.0004)	(0.0004)	(0.0005)	(0.0005)
School Quality	0.0025***	0.0005**	(.)	0.0011***
	(0.0002)	(0.0002)	(.)	(0.0003)
Pop Density	-0.0014***	-0.0010***	-0.0010***	-0.0011***
	(0.0002)	(0.0001)	(0.0003)	(0.0001)
Median Income	0.0023***	0.0011***	0.0018***	0.0010***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Constant	11.4991***	11.8031***	11.8804***	11.8263***
	(0.0386)	(0.0387)	(0.0435)	(0.0494)
Observations	28955	27649	28955	28955
Adj r^2	0.8643	0.8865	0.8767	0.8868
FE Level	Zip Code	Neighborhood	Elem School	Census Tract

Note: All regressions include year and quarter fixed effects.
Bootstrapped clustered standard errors at the census block level in parenthesis.
* p<0.1, ** p<0.05, *** p<0.01.

Table 3.7: House Price Regressions on Walk Score Bins

	(1)	(2)	(3)	(4)
Bedrooms	0.0344***	0.0377***	0.0347***	0.0377***
	(0.0023)	(0.0021)	(0.0021)	(0.0020)
Baths	0.0869***	0.0826***	0.0842***	0.0816***
	(0.0028)	(0.0026)	(0.0027)	(0.0025)
Sq Ft	0.0002***	0.0002***	0.0002***	0.0002***
	(4.6e-06)	(4.3e-06)	(4.3e-06)	(4.2e-06)
Lot Size	9.2e-10***	8.6e-10***	6.7e-10***	6.7e-10***
	(8.2e-11)	(7.7e-11)	(8.7e-11)	(8.1e-11)
Detached	0.1606***	0.1735***	0.1644***	0.1768***
	(0.0084)	(0.0085)	(0.0085)	(0.0081)
Garage	0.0675***	0.0624***	0.0651***	0.0629***
	(0.0037)	(0.0034)	(0.0035)	(0.0033)
Basement	0.1021***	0.0990***	0.0950***	0.0958***
	(0.0048)	(0.0044)	(0.0045)	(0.0042)
Fireplace	0.0839***	0.0635***	0.0671***	0.0606***
	(0.0038)	(0.0035)	(0.0036)	(0.0034)
Central AC	0.0889***	0.0786***	0.0839***	0.0772***
	(0.0031)	(0.0029)	(0.0030)	(0.0028)
Age	-0.0016***	-0.0010***	-0.0009***	-0.0009***
	(0.0003)	(0.0003)	(0.0003)	(0.0003)
Age^2	1.0e-05***	2.4e-06	9.6e-07	9.5e-07
	(2.1e-06)	(2.1e-06)	(2.2e-06)	(2.0e-06)
Walk Score20	-0.0034	-0.0151*	-0.0138	-0.0175**
	(0.0087)	(0.0084)	(0.0084)	(0.0088)
Walk Score40	-0.0018	0.0097	0.0091	0.0006
	(0.0103)	(0.0094)	(0.0102)	(0.0106)
Walk Score60	0.0515***	0.0354***	0.0460***	0.0119
	(0.0111)	(0.0101)	(0.0110)	(0.0111)
Walk Score80	0.1118***	0.0482***	0.0664***	0.0197
	(0.0128)	(0.0119)	(0.0131)	(0.0127)
Transit Score	-0.0009***	-0.0015***	-0.0022***	-0.0019***
	(0.0003)	(0.0003)	(0.0003)	(0.0003)
Distance to CBD	-0.0657***	-0.0636***	-0.0895***	-0.0739***
	(0.0040)	(0.0047)	(0.0056)	(0.0069)
Open Space	0.0020***	0.0000	0.0014***	0.0009**
	(0.0004)	(0.0004)	(0.0005)	(0.0005)
School Quality	0.0025***	0.0003	(.)	0.0012***
	(0.0002)	(0.0002)	(.)	(0.0003)
Pop Density	-0.0014***	-0.0011***	-0.0010***	-0.0011***
	(0.0002)	(0.0001)	(0.0003)	(0.0001)
Median Income	0.0023***	0.0011***	0.0018***	0.0010***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Constant	11.4776***	11.7462***	11.8562***	11.7833***
	(0.0381)	(0.0373)	(0.0404)	(0.0458)
Observations	28955	28955	28955	28955
Adj r^2	0.8636	0.8830	0.8764	0.8868
FE Level	Zip Code	Neighborhood	Elem School	Census Tract

Note: All regressions include year and quarter fixed effects.
Bootstrapped clustered standard errors at the census block level in parenthesis.
* $p<0.1$, ** $p<0.05$, *** $p<0.01$.

range of house prices. This could be because of differences in the preferences between buyers of higher priced homes and buyers of lower priced homes. Indeed, Zietz **et al.** (2008) use a quantile regression approach and find that certain housing characteristics, such as square footage and bathrooms, are priced differently for costlier houses compared to cheaper houses. A similar result is found for homes in Hong Kong (Kim **et al.**, 2015). Heintzelman (2010) finds some effect of a land use preservation policy on property values for some price ranges in Massachusetts. I apply a similar quantile regression approach to the hedonic model to estimate the heterogeneity in households' willingness to pay to live in more walkable neighborhoods.

Quantile regression can provide a richer characterization of the data. Unlike ordinary least squares (OLS) regression, which estimates the conditional mean of the distribution, quantile regression can be used to examine the heterogeneity in the effects of covariates across the entire distribution of the endogenous variable. This method is also more robust to outliers and assumptions about the distribution of the error terms (Koenker & Bassett Jr, 1978). Given the appropriate data generating process, quantile regression provides unbiased estimates of coefficients. This may not be the case if we simply segmented the observations based on the quantile of the dependent variable and ran OLS separately on each subset. As Heckman (1977) argues this sample selection issue would lead to biased estimates.

Table 3.8 presents the results from the quantile hedonic price regressions with neighborhood fixed effects. Regressions are run for the 10th, 25th, 50th, 75th, and 90th percentiles of house prices. The results indicate some variation in the effects of structural characteristics on house prices. The marginal effect from an additional bedroom is relatively higher for lower priced homes. Detached homes, presence of a garage, and central air conditioning have a larger relative effect on price at the lower range compared to the higher range. The effect of walkability on house price is about 1.3% for a 10 point increase for the lower end of the distribution. This effect is nearly cut in half for the regressions of the 75th and 90th percentiles, although the estimated change in dollars is larger.[7] The marginal effect of a change in Walk Score on price is $176 for the 10th quantile, $270

[7]Tests reveal that the coefficients on Walk Score at the 10th and 25th percentile are statistically different than

for the median quantile, and $442 for the 90th quantile. The price of a home at the 10th percentile is $155,000, while at the 90th percentile it is $650,000. It is important to consider the types of buyers of these differently priced homes and their constraints of income. The median income of the block group for homes in the bottom quartile of price is $48,000, while for the top quartile it is $94,000. Buyers of cheaper homes tend to be relatively younger and may value walkability more.[8] They are willing to pay a larger percentage to live in a more urban setting with access to bars, restaurants, and other social activities. They may not care about the quality of the elementary school as much as older households, as the regressions suggest. Surprisingly, the magnitude of the effect of a change in Transit Score decreases with price. This suggests that households purchasing cheaper homes may not value transit availability as much, or may have a greater distaste for noise and congestion, or both. This is counterintuitive to the idea that purchasers of lower priced homes are less likely to have a vehicle and therefore more apt to utilize public transit.

When building these mixed use developments, developers must consider how to maximize the value of the properties. One way is to increase the amount of walkable access to amenities, while limiting nuisance effects. This could be achieved by increasing the amount of nearby open space. The results in Table 3.9 attempt to answer this question by including an interaction term between Walk Score and open space. The coefficient on the interaction term indicates a positive price effect from higher walkability and more open space nearby. Although the coefficient on open space is now negative at the smaller fixed effect levels. Taking both coefficients together, the marginal effect of open space on property values is positive for Walk Scores above 36, which is the case for the vast majority of the sample. The results suggests that increasing the amount of nearby open space may increase the value of walkability by mitigating some of the nuisance effects, particularly congestion.

---

the coefficients at the 75th and 90th percentile at the 5% significance level.

[8] The median age is 32 for the bottom 25% of homes and 40 for the top 25% of homes. Buyers of cheaper homes are also less educated, with only 14%, on average, having a bachelor's degree compared to 50% at the top quartile of home prices.

Table 3.8: Hedonic House Price Regressions by Quantile

	q10	q25	q50	q75	q90
Bedrooms	0.0466***	0.0427***	0.0381***	0.0341***	0.0293***
	(0.0023)	(0.0023)	(0.0012)	(0.0015)	(0.0019)
Baths	0.0808***	0.0843***	0.0843***	0.0780***	0.0710***
	(0.0027)	(0.0016)	(0.0032)	(0.0029)	(0.0030)
Sq Ft	0.0002***	0.0002***	0.0002***	0.0002***	0.0003***
	(3.3e-06)	(3.0e-06)	(3.6e-06)	(2.7e-06)	(3.9e-06)
Lot Size	1.8e-09	1.3e-09	8.8e-10	5.2e-10	2.4e-10
	(1.6e-07)	(1.7e-07)	(1.2e-07)	(6.4e-08)	(6.3e-08)
Detached	0.1956***	0.1786***	0.1661***	0.1588***	0.1433***
	(0.0099)	(0.0095)	(0.0065)	(0.0042)	(0.0080)
Garage	0.0966***	0.0750***	0.0599***	0.0483***	0.0387***
	(0.0085)	(0.0058)	(0.0037)	(0.0032)	(0.0032)
Basement	0.1080***	0.0924***	0.0905***	0.0915***	0.0947***
	(0.0037)	(0.0047)	(0.0024)	(0.0039)	(0.0073)
Fireplace	0.0710***	0.0627***	0.0507***	0.0478***	0.0470***
	(0.0048)	(0.0040)	(0.0029)	(0.0026)	(0.0029)
Central AC	0.1145***	0.0935***	0.0728***	0.0565***	0.0455***
	(0.0042)	(0.0035)	(0.0042)	(0.0031)	(0.0039)
Age	-0.0016***	-0.0016***	-0.0012***	-0.0008***	-0.0004
	(0.0002)	(0.0003)	(0.0002)	(0.0002)	(0.0003)
Age^2	6.1e-06***	7.6e-06***	5.0e-06***	2.2e-06	-1.4e-06
	(1.8e-06)	(2.7e-06)	(1.4e-06)	(1.7e-06)	(2.5e-06)
Walk Score	0.0013***	0.0013***	0.0010***	0.0007***	0.0007***
	(0.0002)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Transit Score	-0.0019***	-0.0017***	-0.0014***	-0.0012***	-0.0014***
	(0.0003)	(0.0003)	(0.0002)	(0.0002)	(0.0002)
Distance to CBD	-0.0545***	-0.0566***	-0.0524***	-0.0515***	-0.0548***
	(0.0037)	(0.0026)	(0.0042)	(0.0051)	(0.0060)
Open Space	-0.0006**	-0.0005	0.0001	0.0004**	0.0004*
	(0.0003)	(0.0004)	(0.0004)	(0.0001)	(0.0003)
School Quality	-0.0002	-0.0001	0.0003	0.0005***	0.0007***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Pop Density	-0.0012***	-0.0010***	-0.0009***	-0.0010***	-0.0012***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0002)
Median Income	0.0013***	0.0012***	0.0011***	0.0010***	0.0009***
	(0.0001)	(0.0001)	(0.0001)	(0.0000)	(0.0001)
Constant	10.9576***	11.1749***	11.3241***	11.4464***	11.5456***
	(0.0509)	(0.0390)	(0.0254)	(0.0295)	(0.0294)
Observations	28955	28955	28955	28955	28955
Pseudo R^2	0.6191	0.6451	0.6734	0.6998	0.7219

Note: All regressions include year and quarter fixed effects.

Bootstrapped standard errors in parenthesis.

* $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table 3.9: House Price Regressions on Walk Score Interacted with Open Space

	(1)	(2)	(3)	(4)	(5)
Bedrooms	0.0243***	0.0345***	0.0377***	0.0346***	0.0376***
	(0.0026)	(0.0023)	(0.0021)	(0.0021)	(0.0020)
Baths	0.0945***	0.0862***	0.0821***	0.0838***	0.0815***
	(0.0032)	(0.0028)	(0.0026)	(0.0027)	(0.0025)
Sq Ft	0.0002***	0.0002***	0.0002***	0.0002***	0.0002***
	(5.1e-06)	(4.6e-06)	(4.3e-06)	(4.3e-06)	(4.2e-06)
Lot Size	6.1e-10***	8.9e-10***	8.6e-10***	7.1e-10***	7.0e-10***
	(7.5e-11)	(8.1e-11)	(7.8e-11)	(8.7e-11)	(8.1e-11)
Detached	0.1286***	0.1575***	0.1736***	0.1643***	0.1767***
	(0.0108)	(0.0085)	(0.0085)	(0.0086)	(0.0081)
Garage	0.0935***	0.0674***	0.0620***	0.0649***	0.0627***
	(0.0044)	(0.0037)	(0.0034)	(0.0035)	(0.0033)
Basement	0.1163***	0.1025***	0.0999***	0.0958***	0.0961***
	(0.0057)	(0.0048)	(0.0043)	(0.0044)	(0.0042)
Fireplace	0.1049***	0.0848***	0.0623***	0.0668***	0.0600***
	(0.0046)	(0.0039)	(0.0035)	(0.0036)	(0.0034)
Central AC	0.1178***	0.0892***	0.0786***	0.0840***	0.0771***
	(0.0038)	(0.0031)	(0.0029)	(0.0030)	(0.0028)
Age	-0.0021***	-0.0019***	-0.0010***	-0.0010***	-0.0009***
	(0.0004)	(0.0003)	(0.0003)	(0.0003)	(0.0003)
Age^2	2.0e-05***	1.3e-05***	2.3e-06	1.4e-06	8.4e-07
	(3.3e-06)	(2.2e-06)	(2.1e-06)	(2.3e-06)	(2.0e-06)
Walk Score	0.0055***	0.0024***	0.0005***	0.0012***	1.1e-05
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Walk x Open Space	5.4e-05***	-1.1e-05	6.1e-05***	4.2e-05***	6.2e-05***
	(1.5e-05)	(1.4e-05)	(1.5e-05)	(1.6e-05)	(1.6e-05)
Open Space	0.0001	0.0026***	-0.0022***	-0.0005	-0.0017**
	(0.0007)	(0.0007)	(0.0007)	(0.0008)	(0.0008)
Transit Score	-0.0009**	-0.0010***	-0.0017***	-0.0024***	-0.0020***
	(0.0004)	(0.0003)	(0.0003)	(0.0003)	(0.0003)
Distance to CBD	-0.0538***	-0.0608***	-0.0635***	-0.0872***	-0.0724***
	(0.0014)	(0.0040)	(0.0047)	(0.0056)	(0.0067)
School Quality	0.0043***	0.0025***	0.0004	(.)	0.0012***
	(0.0002)	(0.0002)	(0.0002)	(.)	(0.0003)
Pop Density	-0.0020***	-0.0013***	-0.0011***	-0.0010***	-0.0011***
	(0.0006)	(0.0002)	(0.0001)	(0.0003)	(0.0001)
Median Income	0.0033***	0.0023***	0.0011***	0.0018***	0.0010***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Constant	10.9326***	11.3608***	11.7366***	11.8120***	11.7806***
	(0.0464)	(0.0391)	(0.0387)	(0.0429)	(0.0455)
Observations	28955	28955	28955	28955	28955
Adj r^2	0.8135	0.8631	0.8831	0.8764	0.8868
FE Level	-	Zip Code	Neighborhood	Elem School	Census Tract

Note: All regressions include year and quarter fixed effects.

Bootstrapped clustered standard errors at the census block level in parenthesis.

* p<0.1, ** p<0.05, *** p<0.01.

3.4 Conclusion

The New Urbanism movement has led to the growth in mixed use developments within cities. Associated with these developments is the concept of creating walkable neighborhoods, reducing the need for households to rely on their vehicles to accomplish daily living activities. Instead, allowing households to have walkable access to nearby amenities. Using Walk Score as a measure of the walkability of a location, I find that households are willing to pay a premium to live in more walkable areas. A 10 point increase in Walk Score leads to a 1-2% increase in housing prices. With the average price in the sample of $365,000, this equates to anywhere from a $7,000 to $15,000 increase in price when moving from a 'somewhat walkable' area to a 'very walkable' area. The price effect is largest for the top range of Walk Scores, suggesting households are willing to pay even more at locations with a large number of destinations within walking distance.

The walkability premium diminishes somewhat with the geographic size of the fixed effect. One reason for this could be that the impact is simply absorbed into the location fixed effect. It could also suggest that walkability is less valuable to households when making location decisions within a smaller defined area. It may be the case that while households do value the opportunity to walk to nearby amenities, they also dislike noise, congestion, and other negative factors associated with heavily trafficked areas. These two opposing effects indicate that households want to be close to amenities, but not too close. They may prefer to live a few blocks from the main commercial area of the neighborhood. The extra distance required to walk to desired destinations could impose relatively small costs and may not preclude walking behavior. This point is further illustrated with the inclusion of transit availability in the regressions. The coefficients on Walk Score increase slightly, while the point estimates for Transit Score are negative, again, possibly suggesting a distaste for noise and congestion.

The results imply that there is a willingness to pay for more walkable developments in cities like Denver. However, urban planners need to consider ways to mitigate the nuisance effects associated with more access to local amenities. Previous literature may have overstated the value of neighborhood walkability by not accounting for confounding locational characteristics. A geographic fixed effects approach helps to control for local heterogeneity, although, it could come at the cost of reducing necessary variation in Walk Scores.

Bibliography

Ahn, Seung C, Lee, Young H, et al. 2007. Life-cycle demand for major league baseball. International Journal of Sport Finance, **2**(2), 79–93.

Al-Mosaind, Musaad A, Dueker, Kenneth J, & Strathman, James G. 1993. Light-rail transit stations and property values: a hedonic price approach.

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

Bertrand, Marianne, Duflo, Esther, & Mullainathan, Sendhil. 2004. How much should we trust differences-in-differences estimates? The Quarterly journal of economics, **119**(1), 249–275.

Billings, Stephen B. 2011. Estimating the value of a new transit option. Regional Science and Urban Economics, **41**(6), 525–536.

Black, Sandra E. 1999. Do better schools matter? Parental valuation of elementary education. The Quarterly Journal of Economics, **114**(2), 577–599.

Bolitzer, Benjamin, & Netusil, Noelwah R. 2000. The impact of open spaces on property values in Portland, Oregon. Journal of environmental management, **59**(3), 185–193.

Bowes, David R, & Ihlanfeldt, Keith R. 2001. Identifying the impacts of rail transit stations on residential property values. Journal of Urban Economics, **50**(1), 1–25.

Boyle, Austin, Barrilleaux, Charles, & Scheller, Daniel. 2014. Does Walkability Influence Housing Prices? Social Science Quarterly, **95**(3), 852–867.

Bresnahan, Timothy F, & Reiss, Peter C. 1990. Entry in monopoly market. The Review of Economic Studies, **57**(4), 531–553.

Brook, Stacey L, et al. 2006. Evaluating inelastic ticket pricing models. International Journal of Sport Finance, **1**(3), 140–150.

Brown, Barbara B, Yamada, Ikuho, Smith, Ken R, Zick, Cathleen D, Kowaleski-Jones, Lori, & Fan, Jessie X. 2009. Mixed land use and walkability: Variations in land use measures and relationships with BMI, overweight, and obesity. Health & place, **15**(4), 1130–1141.

Burger, John D, & Walters, Stephen JK. 2003. Market size, pay, and performance A general model and application to major league baseball. Journal of Sports Economics, **4**(2), 108–125.

Cairney, J, Joshi, D, Li, YC, & Kwan, MYW. 2015. The Impact of the Olympics on Regular Season Team Performance in the National Hockey League. J Athl Enhancement 4: 6. of, **7**, 2.

Carlino, Gerald, & Coulson, N Edward. 2004. Compensating Differentials and the Social Benefits of the NFL. Journal of Urban Economics, **56**(1), 25–50.

Carlton, Dennis W, Frankel, Alan S, & Landes, Elisabeth M. 2004. The control of externalities in sports leagues: an analysis of restrictions in the National Hockey League. Journal of Political Economy, **112**(S1), S268–S288.

Carr, Lucas J, Dunsiger, Shira I, & Marcus, Bess H. 2010. Validation of Walk Score for estimating access to walkable amenities. British Journal of Sports Medicine, bjsports69609.

Coates, Dennis, & Humphreys, Brad R. 2007. Ticket prices, concessions and attendance at professional sporting events. International Journal of Sport Finance, **2**(3), 161.

Coates, Dennis, & Humphreys, Brad R. 2012. Game attendance and outcome uncertainty in the National Hockey League. Journal of Sports Economics, 1527002512450260.

Coates, Dennis, Humphreys, Brad R, et al. 2008. Do economists reach a conclusion on subsidies for sports franchises, stadiums, and mega-events? Econ Journal Watch, **5**(3), 294–315.

Cocco, Angelo, & Jones, JCH. 1997. On going south: the economics of survival and relocation of small market NHL franchises in Canada. Applied Economics, **29**(11), 1537–1552.

Cortright, Joe. 2009. Walking the walk: How walkability raises home values in US cities.

Demsetz, Harold. 1973. Industry structure, market rivalry, and public policy. The Journal of Law & Economics, **16**(1), 1–9.

Des Rosiers, François, Lagana, Antonio, Thériault, Marius, & Beaudoin, Marcel. 1996. Shopping centres and house values: an empirical investigation. Journal of Property Valuation and Investment, **14**(4), 41–62.

Diaz, RB. 1999. IMPACTS OF RAIL TRANSIT ON PROPERTY VALUES. In: 1999 Commuter Rail/Rapid Transit Conference, Proceedings.

Doyle, Scott, Kelly-Schwartz, Alexia, Schlossberg, Marc, & Stockard, Jean. 2006. Active community environments and health: the relationship of walkable and safe communities to individual health. Journal of the American Planning Association, **72**(1), 19–31.

Du Toit, Lorinne, Cerin, Ester, Leslie, Evie, & Owen, Neville. 2007. Does walking in the neighbourhood enhance local sociability? Urban Studies, **44**(9), 1677–1695.

Duncan, Dustin T, Aldstadt, Jared, Whalen, John, Melly, Steven J, & Gortmaker, Steven L. 2011. Validation of Walk Score® for estimating neighborhood walkability: an analysis of four US metropolitan areas. International journal of environmental research and public health, **8**(11), 4160–4179.

Duncan, Dustin T, Aldstadt, Jared, Whalen, John, & Melly, Steven J. 2013. Validation of Walk Scores and Transit Scores for estimating neighborhood walkability and transit availability: a small-area analysis. GeoJournal, **78**(2), 407–416.

Engebretsen, Lars, Steffen, Kathrin, & Soligard, Torbjørn. 2016. Injury Risk in the Olympic Games. Pages 9–18 of: Arthroscopy and Sport Injuries. Springer.

Ferguson, Donald G, Stewart, Kenneth G, Jones, JCH, & Le Dressay, Andre. 1991. The pricing of sports events: do teams maximize profit? The Journal of Industrial Economics, 297–310.

Fort, Rodney. 2004a. Inelastic sports pricing. Managerial and decision economics, **25**(2), 87–94.

Fort, Rodney. 2004b. Subsidies as incentive mechanisms in sports. Managerial and Decision Economics, **25**(2), 95–102.

Fort, Rodney, & Maxcy, Joel. 2003. Competitive balance in sports leagues: An introduction. Journal of Sports Economics, **4**(2), 154–160.

Fort, Rodney, & Quirk, James. 1995. Cross-subsidization, incentives, and outcomes in professional team sports leagues. Journal of Economic Literature, **33**(3), 1265–1299.

Foster, Sarah, Giles-Corti, Billie, & Knuiman, Matthew. 2011. Creating safe walkable streetscapes: Does house design and upkeep discourage incivilities in suburban neighbourhoods? Journal of Environmental Psychology, **31**(1), 79–88.

Frank, Lawrence D, & Engelke, Peter. 2005. Multiple impacts of the built environment on public health: walkable places and the exposure to air pollution. International Regional Science Review, **28**(2), 193–216.

Geddert, Ronald L, Semple, R Keith, et al. 1985. Locating a Major Hockey Franchise: Regional Considerations. Regional Sciences Perspectives, **15**(1), 13–29.

Gilderbloom, John I, Riggs, William W, & Meares, Wesley L. 2015. Does walkability matter? An examination of walkabilitys impact on housing values, foreclosures and crime. Cities, **42**, 13–24.

Greenstein, Shane, & Mazzeo, Michael. 2006. The role of differentiation strategy in local telecommunication entry and market evolution: 1999–2002. The Journal of Industrial Economics, **54**(3), 323–350.

Groothuis, Peter A, Johnson, Bruce K, & Whitehead, John C. 2004. Public funding of professional sports stadiums: Public choice or civic pride? Eastern Economic Journal, **30**(4), 515–526.

Gürtler, Oliver, Lang, Markus, & Pawlowski, Tim. 2015. On the Release of Players to National teams. Journal of Sports Economics, **16**(7), 695–713.

Handy, Susan, Cao, Xinyu, & Mokhtarian, Patricia. 2005. Correlation or causality between the built environment and travel behavior? Evidence from Northern California. Transportation Research Part D: Transport and Environment, **10**(6), 427–444.

Heckman, James J. 1977. Sample selection bias as a specification error (with an application to the estimation of labor supply functions).

Heckman, James J. 1979. Sample Selection Bias as a Specification Error. Econometrica: Journal of the Econometric Society, 153–161.

Heintzelman, Martin D. 2010. Measuring the property-value effects of local land use and preservation referenda. Land Economics, **86**(1), 22–47.

Henrickson, Kevin E. 2012. Spatial competition and strategic firm relocation. Economic Inquiry, **50**(2), 364–379.

Hess, Daniel Baldwin, & Almeida, Tangerine Maria. 2007. Impact of proximity to light rail rapid transit on station-area property values in Buffalo, New York. Urban studies, **44**(5-6), 1041–1068.

Imbens, Guido, & Wooldridge, Jeffrey. 2007. whats new in econometrics lecture 5. NBER Summer Institute.

Irwin, Elena G. 2002. The effects of open space on residential property values. Land economics, **78**(4), 465–480.

Jones, J CH. 1984. Winners, losers and hosers: Demand and survival in the National Hockey League. Atlantic Economic Journal, **12**(3), 54–63.

Jones, J Colin H, & Ferguson, Donald G. 1988. Location and survival in the National Hockey League. The Journal of Industrial Economics, 443–457.

Kahane, Leo, & Shmanske, Stephen. 1997. Team roster turnover and attendance in major league baseball. Applied Economics, **29**(4), 425–431.

Kahane, Leo, Longley, Neil, & Simmons, Robert. 2013. The effects of coworker heterogeneity on firm-level output: assessing the impacts of cultural and language diversity in the National Hockey League. Review of Economics and Statistics, **95**(1), 302–314.

Kim, Hyung-Gun, Hung, Kwong-Chin, & Park, Sung Y. 2015. Determinants of Housing Prices in Hong Kong: A Box-Cox Quantile Regression Approach. The Journal of Real Estate Finance and Economics, **50**(2), 270–287.

Koenker, Roger, & Bassett Jr, Gilbert. 1978. Regression quantiles. Econometrica: journal of the Econometric Society, 33–50.

Krautmann, Anthony C, & Berri, David J. 2007. Can we find it at the concessions? Understanding price elasticity in professional sports. Journal of Sports Economics, **8**(2), 183–191.

Lancaster, Kelvin J. 1966. A new approach to consumer theory. The journal of political economy, 132–157.

Leadley, John C, & Zygmont, Zenon X. 2006. When is the honeymoon over? National Hockey League attendance, 1970-2003. Canadian Public Policy/Analyse de Politiques, 213–232.

Longley, Neil. 2012. The impact of international competitions on competitive balance in domestic leagues: The case of the National Hockey League's participation in the winter Olympics. International Journal of Sport Finance, **7**(3), 249.

Lutzenhiser, Margot, & Netusil, Noelwah R. 2001. The effect of open spaces on a home's sale price. Contemporary Economic Policy, **19**(3), 291–298.

Malpezzi, Stephen, Ozanne, Larry, & Thibodeau, Thomas. 1980. Characteristic prices of housing in 59 SMSAs. The Urban Institute.

Manaugh, Kevin, & El-Geneidy, Ahmed. 2011. Validating walkability indices: How do different households respond to the walkability of their neighborhood? Transportation Research Part D: Transport and Environment, **16**(4), 309–315.

Manuszak, Mark D, & Moul, Charles C. 2008. Prices and Endogenous Market Structure in Office Supply Superstores*. The Journal of Industrial Economics, **56**(1), 94–112.

Mazzeo, Michael J. 2002. Competitive outcomes in product-differentiated oligopoly. Review of Economics and Statistics, **84**(4), 716–728.

Metz, Neil. 2015. Effect of Distance to Schooling on Home Prices. The Review of Regional Studies, **45**(2), 151.

Mills, Brian M, & Rosentraub, Mark S. 2014. The National Hockey League and Cross-Border Fandom Fan Substitution and International Boundaries. Journal of Sports Economics, 1527002514535174.

Molnar, Gabor, & Savage, Scott J. 2017. Market Structure and Broadband Internet Quality. The Journal of Industrial Economics, **65**(1), 73–104.

Moulton, Brent R. 1986. Random group effects and the precision of regression estimates. Journal of econometrics, **32**(3), 385–397.

Noll, Roger G. 1974. Government and the sports business: Papers prepared for a conference of experts, with an introduction and summary. Brookings institution.

Paul, Rodney J. 2003. Variations in NHL attendance: the impact of violence, scoring, and regional rivalries. American Journal of Economics and Sociology, **62**(2), 345–364.

Pivo, Gary, & Fisher, Jeffrey D. 2011. The walkability premium in commercial real estate investments. Real Estate Economics, **39**(2), 185–219.

Rascher, Daniel A, Brown, Matthew T, Nagel, Mark S, & McEvoy, Chad D. 2008. Where did National Hockey League fans go during the 2004-2005 lockout? An analysis of economic competition between leagues. International Journal of Sport Management and Marketing, **5**(1-2), 183–195.

Rauterkus, Stephanie Yates, & Miller, Norman. 2011. Residential land values and walkability. Journal of Sustainable Real Estate, **3**(1), 23–43.

Rosen, Sherwin. 1974. Hedonic prices and implicit markets: product differentiation in pure competition. Journal of political economy, **82**(1), 34–55.

Schofield, John A. 1983. Performance and attendance at professional team sports. Journal of Sport Behavior, **6**(4), 196.

Seredynski, Glen, Jones, J Colin H, & Ferguson, Donald G. 1994. On team relocation, league expansion, and public policy: or, where do we put this hockey franchise and why would you care. Seton Hall J. Sport L., **4**, 663.

Singh, Vishal, & Zhu, Ting. 2008. Pricing and market concentration in oligopoly markets. Marketing Science, **27**(6), 1020–1035.

Sirmans, G Stacy, MacDonald, Lynn, Macpherson, David A, & Zietz, Emily Norman. 2006. The value of housing characteristics: a meta analysis. The Journal of Real Estate Finance and Economics, **33**(3), 215–240.

Sirmans, Stacy, Macpherson, David, & Zietz, Emily. 2005. The composition of hedonic pricing models. Journal of real estate literature, **13**(1), 1–44.

Stewart, Kenneth G, Ferguson, Donald G, & Jones, JCH. 1992. On violence in professional team sport as the endogenous result of profit maximization. Atlantic Economic Journal, **20**(4), 55–64.

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

Tu, Charles C, & Eppli, Mark J. 2001. An empirical examination of traditional neighborhood development. Real Estate Economics, **29**(3), 485–501.

Ulm, Gene. 2007. The key findings from a national survey of 1,000 adults. 1, **1**(1), 1.

Villar, Jaume García, Guerrero, Plácido Rodríguez, et al. 2009. Sports attendance: a survey of the literature 1973-2007. Rivista di Diritto e di Economia dello Sport, **5**(2), 112–151.

Vrooman, John. 2009. Theory of the perfect game: Competitive balance in monopoly sports leagues. Review of Industrial Organization, **34**(1), 5–44.

Wada, Roy, & Zahirovic-Herbert, Velma. 2013. Distribution of Demand for School Quality: Evidence from Quantile Regression. Journal of Housing Research, **22**(1), 17–31.

Winfree, Jason A, & Fort, Rodney. 2008. Fan substitution and the 2004-05 NHL lockout. Journal of Sports Economics.

Winfree, Jason A, McCluskey, Jill J, Mittelhammer, Ron C, & Fort, Rodney. 2004. Location and attendance in major league baseball. Applied Economics, **36**(19), 2117–2124.

Zietz, Joachim, Zietz, Emily Norman, & Sirmans, G Stacy. 2008. Determinants of house prices: a quantile regression approach. The Journal of Real Estate Finance and Economics, **37**(4), 317–333.

Appendix A

Chapter 1: Ticket Pricing and Market Structure

A.1 Estimation of Entry Model for Homogeneous Firms

The first stage entry model under the assumption that all firms are homogeneous is given by,

$$\Pi(Z_m, N_m, \gamma) = Z_m\gamma + \sum_{n=1}^{N-1} \alpha_n \times 1(N_m > n) + u_m \tag{A.1}$$

The variance of the latent payoff function, u_m, is normalized to one which allows the parameters of the profit function to be estimated using maximum likelihood for an ordered probit. Specifically, the equations to be estimated are,

$$
\begin{aligned}
Pr[N_m = 0] &= Pr[\pi(N_m = 1) + u_m < 0] \\
&= Pr[Z_m\gamma + u_m < 0] \\
&= Pr[u_m < -(Z_m\gamma)] \\
&= F[-(Z_m\gamma)] \\
&= \Phi[-(Z_m\gamma)] \\
&= 1 - \Phi[(Z_m\gamma)] \\
Pr[N_m = 1] &= Pr[\pi(N_m = 1) + u_m > 0, \pi(N_m = 2) + u_m < 0] \\
&= Pr[Z_m\gamma + u_m > 0, Z_m\gamma + \alpha_1 + u_m < 0] \\
&= Pr[u_m > -(Z_m\gamma), u_m < -(Z_m\gamma + \alpha_1)] \\
&= F[-(Z_m\gamma + \alpha_1)] - F[-(Z_m\gamma)] \\
&= \Phi[-(Z_m\gamma + \alpha_1)] - \Phi[-(Z_m\gamma)] \\
&= \Phi[(Z_m\gamma)] - \Phi[(Z_m\gamma + \alpha_1)] \\
Pr[N_m \geq 5] &= Pr[\pi(N_m \geq 5) + u_m > 0] \\
&= Pr[Z_m\gamma + \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 + u_m > 0] \\
&= Pr[u_m > -(Z_m\gamma + \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4)] \\
&= Pr[u_m < (Z_m\gamma + \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4)] \\
&= F[Z_m\gamma + \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4] \\
&= \Phi[Z_m\gamma + \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4]
\end{aligned}
\tag{A.2}
$$

A.2 Derivation of the Correction Term for Homogeneous Firms

Given the distributional assumptions of the error terms,

$$
\begin{pmatrix} \epsilon_{im} \\ u_m \end{pmatrix} \sim N \begin{pmatrix} 0, & \sigma_\epsilon^2 & \\ 0, & \sigma_{\epsilon u} & 1 \end{pmatrix}
\tag{A.3}
$$

it can be shown that conditional on u_m the distribution of ϵ_{im} is given by

$$\epsilon_{im}|u_m \sim N(\sigma_{\epsilon u}u, \sigma_\epsilon^2(1-\rho^2)) \tag{A.4}$$

$$\begin{aligned}
E[\epsilon|u] &= \mu_\epsilon + \rho\frac{\sigma_\epsilon}{\sigma_u}(u-\mu_u) \quad where\ \rho = \frac{cov(\epsilon u)}{\sigma_\epsilon\sigma_u} \\
&= \mu_\epsilon + \frac{\sigma_{\epsilon u}}{\sigma_\epsilon\sigma_u}\frac{\sigma_\epsilon}{\sigma_u}(u-\mu_u) \\
&= 0 + \frac{\sigma_{\epsilon u}}{\sigma_\epsilon}\sigma_\epsilon(u-0) \\
&= \sigma_{\epsilon u}u
\end{aligned}$$

$$\begin{aligned}
V[\epsilon|u] &= \int_{-\infty}^{\infty}(\epsilon - E[\epsilon|u])^2 h(\epsilon|u)d\epsilon \\
&= \int_{-\infty}^{\infty}(\epsilon - \sigma_{\epsilon u}u)^2 h(\epsilon|u)d\epsilon \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(\epsilon - \sigma_{\epsilon u}u)^2 h(\epsilon|u)f(u)d\epsilon du \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(\epsilon - \sigma_{\epsilon u}u)^2 g(\epsilon,u)d\epsilon du \\
&= E[\epsilon - \sigma_{\epsilon u}u]^2 \\
&= E[\epsilon^2 - 2\sigma_{\epsilon u}u\epsilon + \sigma_{\epsilon u}^2u^2] \\
&= \sigma_\epsilon^2 - 2\sigma_{\epsilon u}^2 + \sigma_{\epsilon u}^2\sigma_u^2 \\
&= \sigma_\epsilon^2 - \sigma_{\epsilon u}^2 \\
&= \sigma_\epsilon^2 - (\rho\sigma_\epsilon\sigma_u)^2 \\
&= \sigma_\epsilon^2 - \rho^2\sigma_\epsilon^2 \\
&= \sigma_\epsilon^2(1-\rho^2)
\end{aligned} \tag{A.5}$$

Using the law of iterated expectations, the expectation of the error term in the price regression conditional on the covariates is

$$
\begin{aligned}
E[\epsilon|X,Z,N] &= E[E(\epsilon|X,Z,N,U)|X,Z,N] \\
&= E[E(\epsilon|u)|X,Z,N] \\
&= E[\sigma_{\epsilon u} u|Z,N] \\
&= \sigma_{\epsilon u} E[u|Z,N]
\end{aligned}
\tag{A.6}
$$

The expression for the correction term has a closed form solution. Given the distributional assumptions on the error term, we know that

$$
\phi(u) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{u^2}{2}) \tag{A.7}
$$

and

$$
\begin{aligned}
Pr[-\pi(n) < u < -\pi(n+1)] &= \Phi(\pi(n)) - \Phi(\pi(n+1)) \\
Pr[u| - \pi(n) < u < -\pi(n+1)] &= \frac{\phi(u)}{\Phi(\pi(n)) - \Phi(\pi(n+1))}
\end{aligned}
\tag{A.8}
$$

so that

$$
\begin{aligned}
E[u| - \pi(n) < u < -\pi(n+1)] &= \int_{-\pi(n)}^{-\pi(n+1)} u \cdot Pr[u| - \pi(n) < u < -\pi(n+1)] du \\
&= \int_{-\pi(n)}^{-\pi(n+1)} u \cdot \frac{\phi(u)}{\Phi(\pi(n)) - \Phi(\pi(n+1))} du \\
&= \int_{-\pi(n)}^{-\pi(n+1)} \frac{u \cdot \frac{1}{\sqrt{2\pi}} \exp(-\frac{u^2}{2})}{\Phi(\pi(n)) - \Phi(\pi(n+1))} du \\
&= \int_{-\pi(n)}^{-\pi(n+1)} \frac{\frac{d}{du}(-\frac{1}{\sqrt{2\pi}} \exp(-\frac{u^2}{2}))}{\Phi(\pi(n)) - \Phi(\pi(n+1))} du \\
&= \frac{-[\frac{1}{\sqrt{2\pi}} \exp(-\frac{u^2}{2})]_{-\pi(n)}^{-\pi(n+1)}}{\Phi(\pi(n)) - \Phi(\pi(n+1))} \\
&= \frac{-[\phi(u)]_{-\pi(n)}^{-\pi(n+1)}}{\Phi(\pi(n)) - \Phi(\pi(n+1))} \\
&= \frac{\phi(\pi(n)) - \phi(\pi(n+1))}{\Phi(\pi(n)) - \Phi(\pi(n+1))}
\end{aligned}
\tag{A.9}
$$

and

$$
\begin{aligned}
E[u|u > -\pi(5)] &= \int_{-\pi(5)}^{\infty} u \cdot Pr[u|u > -\pi(5)]du \\
&= \int_{-\pi(5)}^{\infty} u \cdot \frac{\phi(u)}{\Phi(\pi(5))}du \\
&= \frac{-[\phi(u)]_{-\pi(5)}^{\infty}}{\Phi(\pi(5))} \\
&= \frac{\phi(\pi(5))}{\Phi(\pi(5))}
\end{aligned}
\tag{A.10}
$$

A.3 Derivation of the Correction Term for Heterogenous Firms

Using the law of iterated expectations, the expectation of the error term in the price regression conditional on the covariates is given by

$$
E[\epsilon_{im}|X_{im}, Z_m, N_m] = \sigma_{mO} E[\epsilon_O|X_{im}, Z_m, N_m] + \sigma_{mF} E[\epsilon_F|X_{im}, Z_m, N_m] \tag{A.11}
$$

having estimated the values of the profit-function parameters and given the distributional assumptions above, the probability of realizing a specific market outcome can be predicted. This in turn is utilized to calculate the correction term which can be written as

$$
\begin{aligned}
E[\epsilon_{im}|X_{im}, Z_m, N_m] =& \sigma_{mO} \frac{\iint\limits_{n=(O,F)} \epsilon_O f(\epsilon_O, \epsilon_F) d\epsilon_O d\epsilon_F}{\iint\limits_{n=(O,F)} f(\epsilon_O, \epsilon_F) d\epsilon_O d\epsilon_F} \\
&+ \sigma_{mO} \frac{\iint\limits_{n=(O,F)} \epsilon_F f(\epsilon_O, \epsilon_F) d\epsilon_O d\epsilon_F}{\iint\limits_{n=(O,F)} f(\epsilon_O, \epsilon_F) d\epsilon_O d\epsilon_F}
\end{aligned}
\tag{A.12}
$$

where $\iint\limits_{n=(O,F)} f(\epsilon_O, \epsilon_F) d\epsilon_O d\epsilon_F$ represents the probability of the specific market outcome n=(O,F).