Applications of Demand Estimation to the Child Care Market

A DISSERTATION

SUBMITTED TO THE FACULTY OF

THE UNIVERSITY OF MINNESOTA

BY

Jonathan Borowsky

IN PARTIAL FULFILLMENT

OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

Thomas J. Holmes, Advisor

July, 2020

Acknowledgments

I gratefully acknowledge the following:

The Office for Program Research and Evaluation in the Administration for Children and Families provided generous financial support through Child Care Research Scholars Grant #90YE0211.

The Minnesota Department of Human Services not only provided the data on which these projects are based, and also contributed meaningful to the development of the research program presented in this dissertation. I am particularly grateful to the DHS staff including Sheila Garceau, Peder Garnaas-Halvorson, and Lyn Rhodes, for both preparing the data extracts and helping me understand them, and to Elizabeth Davis, Won Fy Lee, and Aaron Sojourner for allowing me to build on their existing research datasets.

My advisor, Tom Holmes, has been patient, encouraging, and kind through the process of developing my research agenda and preparing my dissertation. I am thankful for his grace and intelligence. Joel Waldfogel, David Rahman, and Aaron Sojourner have been extremely generous with their time, have always been ready to talk to me about ideas, and have dispensed excellent advice, and I am thankful to them. Amil Petrin was an intuitive teacher of demand estimation, and I owe him a debt of thanks for introducing me to this powerful set of methods. My friends and colleagues in the University of Minnesota Economics PhD program, especially Aradyha Sood, Sergiy Golovin, Conor Clarke, Veronica Postal, Hasan Tosun, and Shahid Ahmed, have been a tremendous intellectual and personal resource. Elizabeth Davis has been my most important mentor and in all substantive ways has been a second advisor to me. I have benefited from her intelligence, kindness, and deep knowledge of the subject matter. I could not have done this without her, and I am incredibly grateful.

I am very thankful to Brian and Bernadette Mitchell, who have welcomed me into their home while I was writing this dissertation, and have been extremely supportive and encouraging. I am also very thankful to my mother, Toni Borowsky, who taught me to value persistence and intellectual seriousness.

My partner, Lauren Mitchell, has contributed vitally to this dissertation. She has been like a climbing partner, belaying the rope so that I do not fall, and helping me analyze the route and find the holds that will allow me to reach the summit.

Contents

List of Tables

List of Figures

Introduction

This dissertation presents three studies that use discrete choice demand estimation methods to analyze policy-relevant questions regarding the child care market. The three studies that comprise this dissertation are tied together by a common focus on the Minnesota child care market, and a common methodological approach based on combining discrete choice methods with spatial data on both consumers and producers.

Chapter 1, “Who Benefits From Child Care Quality Ratings” analyzes the effects on consumers of Minnesota’s Parent Aware Quality Rating and Improvement System, a voluntary ratings scheme for child care providers. This study has two research questions. First, do consumers respond to the ratings? Second, how are the benefits of ratings geographically distributed? The study uses panel data on prices, enrollments, and characteristics at Minnesota child care centers to estimate the demand system. I use the approach of Berry et al. (1995) to invert the demand system by simulating the assignment of consumers to providers, allowing the model to be estimated on provider-level data. My results suggests that the ratings have a modest but meaningful effect on child care provider choices. Welfare analysis suggests that the benefits of the ratings are highest in dense urban regions, including regions with high concentrations of low-income consumers.

Chapter 2, “How Far Will They Go?”, examines trade-offs between distance and provider characteristics in the child care choices of subsidized consumers from Minnesota’s Child Care Assistance Program. This study represents joint work with Elizabeth Davis. At the center of this study is estimation of a nested logit demand system that models the choice between center-based and family child care simultaneously with the choice of individual providers. We estimate a model using matched consumer-provider administrative data from the CCAP program, supplemented with a range of additional data on licensed child care providers. We present two applications of the estimated model. First, we define a construct to measure the relevant area that would be impacted by adding a hypothetical new center to the model, and analyze how this impact area varies in urban, suburban and rural locations. Second, we examine the extent to which race-based patterns in the use of center-based and family child care can be explained by differences in what type of providers are locally available.

Chapter 3, “Fewer Options or More Substitution”, uses a simplified version of the nested logit demand model as a lens to understand trends in the child care choices of Minnesota subsidized consumers over time. This study represents joint work with Elizabeth Davis. The proportion of CCAP consumers using family child care providers dramatically decreased between 2010 and 2020. To what extent can this be attributed to declining availability of family providers? We estimate the model on a sequence of cross-sections from simulated counterfactuals designed to estimate the importance of different factors in the declining use of family providers in this population. We find that two factors are particularly important: the decline in availability of family providers, and changes in the location and demographic composition of the CCAP consumer population that favor use of center-based providers. The effects of changes in the estimated choice model coefficients - which might represent preference either shifts or changes in average price or unobserved characteristics of the providers of different types - are also substantial. Increased availability of centers is much less important.

Chapter 1

Who Benefits From Quality Ratings?

1.1 Introduction

At least 42 states have implemented child care Quality Rating and Improvement Systems, (QRIS) which give providers the opportunity to be evaluated by regulators and assigned a quality rating. This paper uses geocoded panel data on child care providers and consumers to measure the impacts of one such program, Minnesota's ParentAware ratings. The ParentAware program is in large part an informational intervention, designed to improve the quality of child care by creating quality information and distributing it to consumers. I find that the ratings provide a yearly direct value to consumers of more than $3 million dollars, more than justifying the cost of creating and administering ParentAware.

If parents value quality child care but have difficulty discerning high from low quality providers, and if high quality is more expensive to produce, the child care market may suffer from a lemons problem which leads the market to undersupply high quality care. Given the body of evidence linking early childhood experiences to social and intellectual development, improving the quality of child care is a priority for policymakers concerned with inequality and opportunity. Crude or poorly designed measures can perversely distort provider incentives, however systems that rely on detailed assessment and measurement are expensive. For example, ParentAware involves qualified assessors conducting a site visit and administering a scale that measures the quality of the developmental environment, as well as many other measures. According to the Minnesota Office of the Legislative Auditor, this program cost more than $9 million to implement. It is therefore imperative to be able to measure the effects of these programs and assess whether the value that they provide to households is worth the cost. My estimates suggest that, in the case of ParentAware, the benefits are worth the costs.

It is also important to understand heterogeneity in the benefits that informational interventions provide. *Ex ante*, the effect of quality ratings on socioeconomic disparities in child care quality is ambiguous and depends of the relative magnitudes of different effects. The benefit from programs like ParentAware depends on the marginal impact of the provided information on choices, and the amount of benefit will be larger for consumers who have a higher chance of their choice being affected by the additional information. The benefit from the ratings will thus be higher for consumers whose choice set includes many providers whose quality levels are varied and lower for consumers with few choices or whose choices have less variation in quality. On the one hand, less privileged households may be more likely to live in denser urban areas where they can access a range of providers, and may have more difficulty independently discerning quality levels. On the other hand, higher quality providers may be concentrated in affluent neighborhoods and less privileged households cannot be induced to use high quality providers if none are accessible to them. The value of the ratings to households varies geographically and depends on the role of travel costs in household decision-making and the spatial distribution of providers of different types. In order to know how the value of the ratings to different consumers in different areas, we need a model of who shops where.

This paper will estimate the value of Minnesota's ParentAware ratings to households in different locations. I use panel data on provider enrollments, prices, and characteristics, collected by Child Care Aware of Minnesota, a non-profit that surveys all Minnesota licensed child care providers in order to provide data to child care resource and referral agencies, and assembled by Davis et al. (2019) to study geographic variation in child care access. Using this data I estimate a mixed logit model of child care provider choice of the Berry et al. (1995) type. While the classic application of this method is to implement a model with random coefficients on the taste for product characteristics, it also provides a powerful tool for modeling service markets where location is important to choice. A random coefficient can be thought of as a multiplicative interaction between the product characteristic and a latent taste variable. Travel costs are also an interaction, whose form is not multiplicative but is a function of the distance between the provider and the household. The parameters of a travel cost function can be estimated in the same way as the variance parameters of a random coefficients model. The result is a demand model that accounts for spatial substitution patterns where nearby providers compete more intensely with one another than distant ones.

This paper is not the first to use a structural model of demand to calculate the value of an informational intervention. My approach is similar to that used by Jin and Sorensen (2006) to determine the value of National Center for Quality Assurance health plan ratings. Nor is it the first paper to assess informational interventions in the child care setting. Xiao (2010) uses a structural model of child care demand to evaluate the National Association for the Education of Young Children child care accreditation system, finding modest welfare benefits from accreditation system.[1] Herbst (2016) examines the impacts of QRIS systems using national and a difference-in-differences strategy based on differential timing in QRIS adoption. Among other results, he finds that QRIS adoption causes an increase in child care employment, which can be taken as a proxy for child care utilization, and an increase in the educational credentials of people employed in child care, which can be taken as a proxy for quality levels.

In the context of the existing literature, this paper makes two contributions.

First, using geocoded provider data and the Berry et al. (1995) estimation framework, I solve the thorny market definition question that must be addressed in any rigorous study of the child care market. People tend to use child care that is close to their home. Estimates from the National Survey of Early Care and Education, conducted in 2016 by the federal Department of Health and Human Services, suggests the average distance between home and provider, among children using a center-based child care provider, is 4.6 miles for children 0-3 and 3.9 miles for children 4-5, a small distance compared to the size of most population centers. (National Survey of Early Care and Education Project Team (2016)). As a result, child care "markets" are overlapping rather than discrete. Most likely due to data limitations, previous studies like Xiao (2010) have defined child care markets using statistical areas such as ZIP codes. Since child care providers may be distributed

[1] The relationship between this paper and Xiao (2010) is discussed in greater detail in the literature review section below.

unevenly across close-together statistical regions, using statistical regions as a proxy for child care markets risks providing an inaccurate picture of child care access and variety in different locations. This is particularly problematic since the diameter of ZIP codes and other statistical areas varies in systematic ways, making it likely that any measurement error is correlated with neighborhood characteristics that may also be related to child care demand. In this paper, I estimate demand in a single statewide market with travel costs, replacing assumptions about market boundaries with assumptions about the structure of travel costs. The parameters of the travel cost function are estimated within the model. I propose that this approach provides a much more flexible and realistic picture of the child care market than would be possible with arbitrary market boundaries.

Second, I explore variation in the welfare benefits provided by the quality ratings system. As I have argued above, the benefits of quality ratings are inherently heterogenous, as the probability that ratings information will affect the choice of provider depends on the characteristics of the available choice set. I model household heterogeneity using demographic information from the 2011-15 American Consumer Survey (ACS). Each of the more than four thousand Minnesota census block groups that, in the ACS estimates, includes at least one child age 0-5 is a different consumer type with a location at the population centroid of the census block group and a weight equal to the number of children 0-5. Not only is this detailed cross-sectional demand variation helpful for estimating the model, it also means that I am able to conduct welfare analysis at the block group level and use the model to estimate regional variation in the benefits from the quality ratings.

My approach for assigning a money value to the ratings is like that of Jin and Sorensen (2006). That is to say, I compare choices with the ratings to the model's counterfactual choices for a world without the ratings information, both evaluated according to the utility function of a consumer informed by the ratings information. The question in the welfare counterfactual is "what would a consumer informed by the ratings need to be paid in order to have their decision made by a consumer with identical preferences, who did not have access to the ratings." I make an incremental addition to their method by noting that it is a special case of the framework that Train (2015) derives for welfare calculations when a consumer makes choices based on incomplete information. This provides me with closed-form expressions for the welfare quantities, rather than needing to simulate choice draws.

There are two critical problems of endogeneity that must be resolved in order to use a structural model to measure the value of quality ratings, especially in a market as rich with product variation as the child care market.

First, as Jin and Sorensen (2006), Xiao (2010), and Dranove and Jin (2010) all emphasize, we must expect that quality ratings are endogenous. After all, they are an attempt to measure product quality, and it's reasonable to expect them to be correlated with other quality information that consumers may observe, for example through advertising or provider reputation. Any econometric specification that ignores this will systematically overestimate the impact of quality ratings. In their influential study of restaurant hygiene score cards, Jin and Leslie (2003) use an intuitive panel data strategy to address this, comparing the revenue of highly rated restaurants before and after their rating is disclosed. I use a similar strategy, controlling in the demand model for the time-invariant component of provider quality that is observed by households, but unobserved by the researcher and associated with the rating ultimately assigned.

Second, I expect that price will also be endogenous, as it will be in any model of a product market where consumers have access to information about product quality or characteristics that is not present in the data available to the researcher. I address this using an adaptation of the instruments strategy used by Berry et al. (1995). For each provider, I construct price instruments based on a distance-weighted sum of the characteristics of nearby competing providers, capturing the expected inverse relationship between markup and local competition.

In section 2 of this paper, I survey the empirical literature on quality ratings and other informational interventions. In section 3, I introduce the data that will be used to estimate the model. In section 4, I present the economic model. This section has two components. First is the model of choice of child care provider that will be estimated, which is a standard logit discrete choice model of demand, with allowance for household heterogeneity based on varying demographics and based on unobserved variation in tastes. Second is an exposition of the method for calculating welfare quantities, following Train (2015). In section 5, I discuss the estimation methodology. I use the method of Berry et al. (1995) to estimate the demand model by the Generalized Method of Moments. The inversion of Berry (1994) is used to linearize the model and enable the use of instruments for price. A modified version of the Berry et al. (1995) instruments are used, where the instruments are weighted averages of the characteristics of nearby competitors. In section 6 I give a detailed presentation of the results from estimation, which show that the 4 star rating is positive, and other ratings are negative. Overall the ratings are valuable to consumers.

1.2 Literature Review

This paper contributes to the literature on the effects of quality ratings or "score cards", much of which is reviewed in Dranove and Jin (2010). The existing research suggests that disclosure of product attributes can have a big impact on consumer choice. Jin and Leslie (2003) consider Los Angeles County's introduction of a rule requiring restaurants to post a letter grade that reflects their performance on a health inspection. They find substantial evidence that this requirement led to an improvement in the hygiene performance of affected restaurants. Bollinger et al. (2011) study the impact of New York City rules that require restaurants to post calorie values of menu items on sales in Starbucks stores, finding that these these rules led consumers to choose lower-calorie foods and increased the sales of food in Starbucks establishments that were close to competing Dunkin Donuts stores. Disclosure rules are thus an attractive intervention for policy-makers.

Evidence on the effectiveness of *voluntary* disclosure regimes is more mixed. In their study, Jin and Leslie (2003) compare the effectiveness of the mandatory letter grade regime to a transitional period when some municipalities in Los Angeles County required the letter grades to be posted, but others did not, characterizing the latter regime as voluntary disclosure. They find that the effects of the voluntary disclosure regime are much less. Similarly, Mathios (2000) studies the effect of the Nutrition Labeling and Education Act, which mandates disclosure of nutritional information, on the sales of different types of salad dressing. This paper compares the salad dressing market before the introduction of NLEA, when some products displayed nutritional information labels to a period after the introduction of NLEA, when such labeling became mandatory, and finds substantial differences in consumer behavior. Hotz and Xiao (2013) provide a theoretical treatment that illustrates conditions where firms choose not to participate in quality disclosure for strategic reasons related to the effect of quality disclosure on markups through changed competition patterns.

A close comparison can be made between the present study and Jin and Sorensen (2006). That paper examines the National Center for Quality Assurance ratings, which are a voluntary ratings system for health plans. Consumers may be able, at least to some extent, to discern quality without the availability of the ratings, and in that paper, the authors use non-public ratings of health plans to separately identify the treatment effect of the ratings from the ratings' correlation with already-available information about quality. The present approaches the same problem by using data on the period before the ratings are available to control for the already-known information about quality that is likely to be correlated with the ratings.

A close comparison can also be made between the present study and Xiao (2010). That paper examines the privately administered system of child care center accreditation managed by the National Association for the Education of Young Children (NAEYC). Like the present study, that paper estimates the value to consumers of the ratings in a discrete choice framework. The present study differs from Xiao's work in three ways. First, I am able to use true panel data, consisting of multiple observations over time of the child care providers, including data on the enrollment of rated centers before they were rated. My strategy for controlling for endogeneity in the ratings, using these pre-rating observations, is more direct than Xiao's instrumental variables strategy. Second, I model consumer heterogeneity.

On the other hand, Xiao (2010) endogenizes the weight placed by consumers on the ratings. In her model agents have two different signals of quality, one coming from the ratings, and one that is "reputation" independent of the ratings. The longer a provider has existed the more informative the "reputation" signal is. The weight placed by the agent on each signal depends on the informativeness of that signal, and hence in Xiao's model the effect of ratings is moderated by the operational age of the provider, with ratings making the biggest difference for providers that are more recent entrants. I do not yet have data on how long the providers in my sample have existed, but expect to be able to include the relationship between operational age and quality ratings in future versions of this paper.

Jin (2005) considers the strategic incentives to disclose quality information, concluding that HMOs use participation in National Center for Quality Assurance ratings to distinguish themselves from competitors, and in the ratings in highly competitive markets are less likely to participate in the ratings.

An interesting thread in the literature deals with heterogeneity in consumers' use of disclosed product information. One way to put it is that not all customers are listening. This is of particular interest in the child care quality setting because of policymakers' interest in the choices made by low income families. Milyo and Waldfogel (1999) do not study quality disclosure, but rather price disclosure in the form of price advertising. Using a difference-in-differences strategy made possible by the *44 Liquormart* case, which legalized liquor store advertising in Rhode Island, they find limited effects of advertising on prices. The subset of stores who advertise after legalization cut some prices, but only on the products that they advertise or those advertised by competitors.

1.3 Data

Provider Panel

I use a unique panel data set based on an annual census of Minnesota licensed child care providers.

Child Care Aware of Minnesota maintains a database of child care providers as part of NACCRAware, a national child care data system designed for use by child care referral agencies. Providers are surveyed annually by Child Care Aware of Minnesota, on a rolling basis, in order to keep this database up to date. Information is self-reported by the providers, and includes enrollment numbers by age group, price by age group, quality rating information, as well as several other provider characteristics such as accreditations and nonprofit status. Davis et al. (2019) have prepared panel data on Minnesota child care providers based on periodic pulls from this database, merged with additional information from state government sources and licensing records.

The NACCRAware data contains information on price and enrollment in four age groups: infant, toddler, pre-school, and school-age. To simplify the analysis, I treated the infant, toddler, and pre-school categories as a single market. I calculated enrollment as the sum of infant, toddler, and pre-school enrollment, and calculated a price for each provider as a weighted average, using the enrollment from each age group as the weights for each provider. I do not make any use of the school-age group in this paper.

The panel data covers fiscal years 2012-2015 and includes information about child care centers, family day cares, and certain public child care programs, such as Head Start and pre-K programs in schools. This paper focuses on licensed child care centers.

Table 1.1: Center Descriptive Statistics, Fiscal Years 2014-2016

	N	Mean	Std. Dev.	Min	25%	Med.	75%	Max
Enrollment	3426.0	57.25	35.87	2.00	34.50	52.50	69.23	392.67
Weekly Price	3426.0	219.03	58.53	12.08	182.07	224.52	253.51	517.00
Licensed Capacity	3426.0	82.21	55.42	0.00	46.00	73.00	109.00	1140.00
USDA Food Prog	3426.0	0.34	0.47	0.00	0.00	0.00	1.00	1.00
Non-Profit	3426.0	0.37	0.48	0.00	0.00	0.00	1.00	1.00
Accreditation	3426.0	0.33	0.47	0.00	0.00	0.00	1.00	1.00
Rating Stars	3426.0	1.31	1.82	0.00	0.00	0.00	4.00	4.00

Descriptive Statistics

Table 1.1 shows descriptive statistics for the subsample used in the estimation, which consists of licensed child care centers in fiscal years 2014-2016. Additionally, cross-sectional descriptive statistics for each of the years used in estimation are presented in supplemental section 1.7, below. The data show that there is substantial variety in firm size and price. The standard deviation in weekly price is 27% of the mean weekly price of $219.03. The standard deviation in enrollment is 63% of the mean enrollment of 57.25.

Data on Child Care Ratings

The NACCRAware data contains data on participation in child care accreditation systems such as NAEYC, and data on ParentAware ratings from fiscal year 2014-2016.

Table 1.2 shows the fraction of centers participating in different information systems for each year used in the estimation. The data show a steady increase in the fraction of centers participating in both Accreditation and in ParentAware.

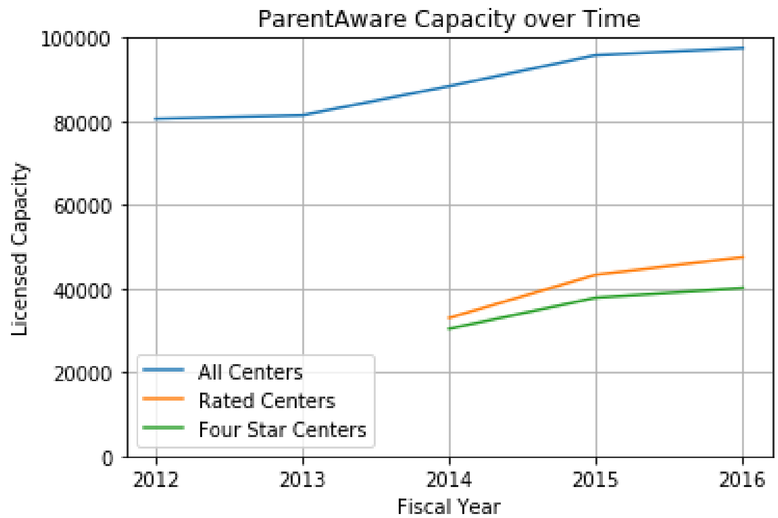
Table 1.2: Fraction of Centers with Different Ratings

Year	N	Accreditation	4 Star Rating	Any Star Rating
2014	1063	0.317	0.262	0.298
2015	1194	0.315	0.304	0.363
2016	1169	0.349	0.326	0.409

Figure 1.1 shows the total licensed capacity in centers of various categories. Almost all of the

rated capacity is in centers that have the highest 4-star rating.

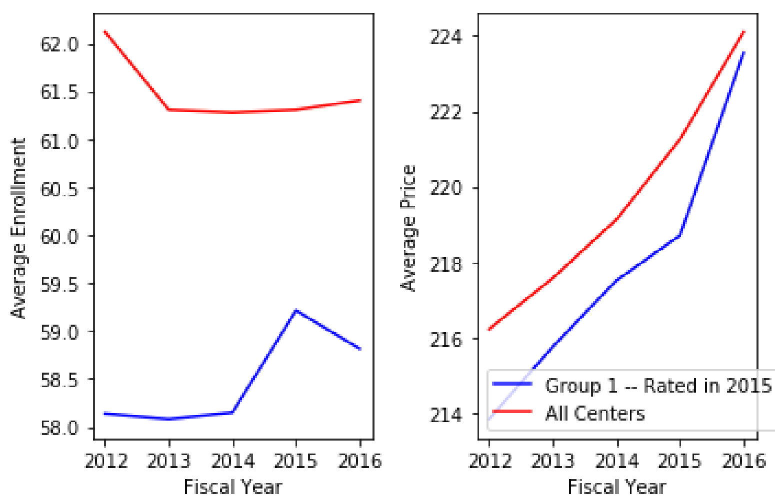
Figure 1.1: Capacity in Rated Centers



Provider Locations and Supplementary Data

Provider location is geocoded based on provider address from the Child Care Aware data and licensing records. Provider data is supplemented by block group level household demographic data

Figure 1.2: Experience of Centers First Rated 4 Stars in FY 2015



from the American Community Survey 2011-15 estimates, accessed through the NHGIS geocoded census data system. Household "locations" are actually the block group centroids. Each location is weighted by the ACS estimates of the number of children 0-5 in that block group. "Percent low income" is fraction of population in the block group under 200% of the poverty level. "Percent college" is fraction of adult population in the block group with bachelors degree or greater education. Household-provider distances are straight-line distance between the block group centroid and the provider address, calculated using Vincenty's formula.

1.4 Model

Demand

I represent choice of child care provider using a standard discrete choice model of product choice. Specifically, consider household i's decision over what child care arrangement to use. Suppose that i can choose any provider within R miles, where R is a distance radius around household i's location that is chosen to be large compared to child care travel distances, such as 50 miles. Then, the choice set for household i in period t is

$$J_{it} = \{j \in J_t | d_{ij} < R\} \cup \{0\}$$

where d_{ij} is distance between household i and provider j, and choice 0 is an outside option. The outside option represents any choice not explicitly represented in the choice set. Here, that would include parental care, care by friends and family, care in a licensed family day care, or care in a publicly provided center such as a Head Start program.

Following Berry, Levinsohn, and Pakes (1995) (hereinafter, "BLP",) I allow for a choice-specific utility that depends on both provider characteristics and characteristics (observed or postulated) of households.

$$u_{ijt} = X_{jt}\beta - \alpha p_{jt} + \xi_{jt} + \gamma d_{ij} + \sum_h \sigma_h x_{jht} \nu_{ih} + \epsilon_{ijt}$$

Here, ξ_{jt} is a "structural" error term representing unobserved information about provider j in period t that is relevant to all households, and ϵ_{ijt} is error term that is "idiosyncratic" to the household.

Choice-specific household utility depends on household "type" in two ways. First, random coefficients may be implemented through ν_{ih}, a random draw from a unit normal distribution, so that consumer i's "taste" for characteristic h is distributed $N(\beta_h, \sigma_h)$.[2] Second, through the effect of the household-provider distance term d_{ij}.

The purpose of allowing choice-specific utility to be different for different "types" of household is twofold. First, it allows for more patterns of substitutability between providers that are more complex than what could be represented in a non-mixed specification. Second, it allows for the model to incorporate important information about households that affects those substitution patterns.

Here it is worth saying a little bit about the role of household-provider distance in the model. Because provider market shares are observed only at the aggregate level, we do not directly observe the distance between households and the providers they choose. Households are assigned to providers endogenously through the demand model. However, the inclusion of household-provider distance allows the model to treat providers that are near to one another as closer substitutes than providers that are far away from one another, in a way that is shaped by specific information on where households live. In this way, the model's treatment of household-provider distance is analogous to the way BLP treat household income.

In the discrete choice model, each household chooses the provider in their choice set that provides the highest choice specific utility. Assuming that the idiosyncratic error term ϵ_{ijt} has the extreme value distribution, and is i.i.d., and normalizing the utility value of the outside option to be centered around zero, the probability that household i will choose provider j is given by the logit choice probability formula.

$$P_{ijt} = \frac{e^{X_{jt}\beta - \alpha p_{jt} + \xi_{jt} + \gamma d_{ij} + \sum_h \sigma_h x_{jht} \nu_{ih}}}{1 + \sum_{k \in J_i} e^{X_{kt}\beta - \alpha p_{kt} + \xi_{kt} + \gamma d_{ik} + \sum_h \sigma_h x_{kht} \nu_{ih}}}$$

Demand, stated as market shares for each of the providers, has the form

$$s_{jt} = \int_i \frac{e^{X_{jt}\beta - \alpha p_{jt} + \xi_{jt} + \gamma d_{ij} + \sum_h \sigma_h x_{jht} \nu_{ih}}}{1 + \sum_{k \in J_i} e^{X_{kt}\beta - \alpha p_{kt} + \xi_{kt} + \gamma d_{ik} + \sum_h \sigma_h x_{kht} \nu_{ih}}} w_i di$$

where w_i is a weight function capturing the proportion of households of each type, based on the proportion of households at each location and the assumed distribution of the random taste parameter.

[2] This version of the paper, however, does not report the results of any random coefficients specifications.

Calculation of Welfare Quantities

To measure the value of the ParentAware ratings to consumers, we wish to determine compensating variation. Compensating variation for quality ratings is the answer to the question "how much would consumers with the ratings information be willing to pay in order to avoid giving the information up?"

Jin and Sorensen (2006) point out an important subtlety about the calculation of welfare quantities for ratings systems and other informational interventions. When compensating variation is calculated for a price change, or the introduction of a new product, the change in market conditions affects the consumer's choice set but not their utility function. In contrast quality information has a direct effect on willingness to pay. To correctly value the counterfactual where consumers give up the ratings information, it is necessary to value the choices that would be made without the ratings information according to the utility function of the consumer who has the ratings information.

Let A and B be the information regimes without and with the ratings, respectively, and let $a(\varepsilon)$ and $b(\varepsilon)$ be the optimal choice functions corresponding to each regime. The compensating variation value of the ratings is:[3]

$$\frac{1}{\alpha}[V^B(b) - V^B(a)]$$

To illustrate why it is important to evaluate the choices according to the with-ratings utility function, it may be helpful to consider the hypothetical of the low-rated provider's loyal client. Suppose that there is a provider whose quality rating indicates a quality level lower than what consumers would have expected without the rating. Further suppose that there are some households who choose that provider despite the low rating. (Perhaps the cost is also low, or the household receives a good idiosyncratic utility draw for that provider.) We now ask how well that consumer would be if the quality information was not available. For this consumer, the quality information cannot have been marginal. If the chosen provider is the best choice even with the negative quality information, it will also be the best choice without the negative quality information. This consumer's choice is unaffected by the quality ratings, and their welfare should thus also be unaffected. However, if the

[3] More rigorously, compensating variation value of market conditions B over market conditions A is defined as the money transfer that would have to be paired with B in order to make the consumer indifferent between A and B-minus-transfer. Like Jin and Sorensen (2006), I am considering a model where the marginal utility of money is uniform across consumers, in which case compensating variation simplifies to the difference in consumer surplus divided by the marginal utility of money.

welfare value of ratings to this were naively calculated as $V^B(b(\varepsilon)) - V^A(a(\varepsilon))$, then this consumer would be considered to have been *harmed* by the ratings. Using the with-ratings utility function to calculate the welfare value of the without-ratings choice function gives the desirable property that a consumer cannot be harmed by additional information.

Train (2015) considers the more general question of how to calculate welfare quantities when an agent bases their choice on inaccurate or incomplete estimates of value, and derives expressions for these welfare quantities. Jin and Sorensen (2006) calculate $V^B(a)$ by simulation, taking random draws from ϵ to simulate choices and then evaluating those choices according to V^B. However the quantity they are calculating can also be as a special case of the framework explained by Train, who shows that $V^B(a)$ can be written as the sum of $V^A(a)$ and an adjustment term. In Appendix 1.9, I use Train's framework to derived closed-form expressions for the compensating variation of quality ratings.

1.5 Estimation

Instruments

As is usual in a model of demand for differentiated products, we expect that price is likely to be endogenous. There are many inputs that we do not directly observe that would be expected to determine quality or other dimensions of desirability of child care providers to consumers, and it is reasonable to suppose that the "high-quality" providers should also be higher-priced.

Let x_j^x be a provider characteristics. A price instrument z_j^c is constructed by taking the distance-weighted sum of x_k^c across other providers in the market.

$$z_j^c = \sum_{j \in J_t, k \neq j} \frac{x_k^c}{d_{jk}}$$

Five instruments are constructed in this manner. Instruments are constructed from provider characteristics licensed capacity, nonprofit status, accreditation status, and ParentAware rating status. An instrument is also constructed using a constant for x_k^c, which provides a measure of the density of other nearby centers.

One test of instruments is to examine the results of the "first stage" regression of the endogenous variables on the instruments. If the instruments do not have explanatory power, they cannot be very satisfactory. Appendix 1.8 shows the results of the first stage regression .The specification uses year fixed effects to control for time trends in child care prices. Taken as a whole, the instruments provide substantial additional explanatory power. An F-test of the restriction dropping the five instruments from the model is rejected at the 1% significance level.

Berry Inversion and Estimation Strategy

I follow the strategy devised by Berry (1994) for how to use instruments to estimate a discrete choice model. The strategy involves inverting the market shares function in order to get a linear problem.

In order to explain how this works, it is helpful to re-write the choice-specific utility function as the sum of three terms

$$u_{ijt} = \delta_j + \mu_{ij} + \epsilon_{ij}$$

The first term, δ_j, captures the purely "vertical" dimension of differentiation between providers. That is, δ_j is everything about provider j that is valued the same by all households, including the unobserved quality term ξ_j, and the value placed on j's observed characteristics by a consumer with average tastes. The second term, μ_{ij} captures the "horizontal" dimension of differentiation between providers; those parts of the household's valuation that depend on the household's type. This might include the effect on choice of the distance between household and provider. Even if all households place the same value on distance, the values of d_{ij} will be different for households at different locations. μ_{ij} will also include the effects of taste variation as expressed through random coefficients. Finally, ϵ_{ij} is the random idiosyncratic error term.

Given that μ_{ij} is a function of the data and some unknown parameters θ, and given a vector of values, $\delta = \{\delta_j\}$, fitted market shares can be computed, conditional on θ and $\{\delta_j\}$, using the formula

$$s_j(\delta;\theta) = \sum_i w_i \frac{e^{\delta_j + \mu_{ij}}}{1 + \sum_k e^{\delta_k + \mu_{ik}}}$$

The strategy for estimating the model has three parts. First, given a candidate value for θ, determine the vector of values $\delta(\theta)$ that matches the fitted market shares $s_j(\delta;\theta)$ to the observed market shares in the data. Second, considering the expression,

$$\delta_j = X_j\beta - \alpha p_j + \xi_j$$

estimate the vector of structural errors $\xi(\theta)$ as the residual of an instrumental variables regression with $\delta(\theta)$ as the left hand side. Third, using this vector of structural errors as an input into a GMM objective function, and re-computing δ and ξ for each candidate value of θ find the value of θ that minimizes that objective function. In the current version of the analysis, the only moment condition is the one from the demand equations, $E(\xi_j|Z_j) = 0$. The objective function I use to estimate θ is thus relatively simple.

$$\hat{\theta} = \arg\min_{\theta} \xi(\theta)^T \xi(\theta)$$

The BLP contraction

The foregoing discussion assumes that there is a method of determining the vector δ that matches the fitted market shares $s(\delta;\theta)$ to the observed market shares s^0. In order to do so I use the contraction described by BLP. This method uses a process of iterative adjustment. Define the operator $T\cdot$ by

$$T \cdot \delta_j = \delta_j + \ln s_j^0 - \ln s_j(\delta;\theta)$$

BLP show that this operator is a contraction and thus that it has a unique fixed point, which can be found by iteratively applying it to an initial “guess”. Since the fixed point occurs when $s_j(\delta;\theta) = \ln s_j^0$, this provides a computational method for calculating $\delta(\theta)$. An initial value of δ is set, and then the contraction iteratively applied until the differences between δ and $T \cdot \delta$ are small compared to a specified tolerance.

Identifying a Treatment Effect

In specifying the model, the ParentAware star ratings are included in the provider characteristics matrix X_{jt}. We should, however, be concerned that the demand unobservable ξ_{jt} will be correlated with the ParentAware ratings variables. ξ_{jt} captures whatever information about providers that is unobserved by the researcher but that households know and incorporate into their choice decisions. If, as seems reasonable to expect, the providers that have high values of ξ_{jt} in the data generating process are more likely to be assigned a four star rating, then the coefficient on the four star rating parameter may be biased upwards.

In order to address this, I follow a difference-in-differences strategy. Implicit in this strategy is the assumption that the component of ξ_{jt} that is correlated with the eventual rating is stable over time. Each provider is assigned to a "group" based on the highest ParentAware rating they receive. Thus, a provider whose highest star rating is four stars is assigned to the four stars group. The providers that are never rated in the data are their own group. A dummy variable is assigned to each group and included in X_{jt}. These dummy variables account for the average differences between the providers that receive, for example, four stars, when they are rated, and those that receive three stars, or that never choose to become rated.

As part of this difference-in-differences strategy, I also include year fixed effects by incorporating year dummies into the characteristics matrix X_{jt}. Year fixed effects are equivalent to allowing the value of the outside option to be different in different years. This is necessary because the overall demand for child care, and for center-based care, is not constant over time, it is increasing. Since the likelihood of being rated is correlated with time – the ratings are only present in the later years – if we did not account for this overall demand trend it might bias the ratings coefficients upwards. An alternative to using year fixed effects would be including a linear or quadratic time trend.

1.6 Results

Baseline Demand Models

In this subsection, I compare the estimates from the mixed logit demand model to logit models that do not account for the role of geography and travel costs in child care demand.

Table 1.3 shows these results. In all three specifications, the outcome variable can be understood as a score that captures the utility value to an average consumer of choosing product j. For models A and B the outcome variable is defined as $\ln s_j - \ln s_0$, the linearization of the logit model that Berry (1994) suggests in order to allow instrumental variables to be employed. Coefficients for models A and B are by ordinary (OLS) and instrumental variables (2SLS) least squares respectively. Model C is a mixed logit model where the household type varies with location, with a single household-product utility interaction term, a linear travel cost. Model C is estimated using nonlinear two stage least squares using the Berry et al. (1995) estimation algorithm.[4] The "outcome" variable in a mixed logit model is a mean utility score that is most intuitively understood by noting that if the utility specification contains no household-product interaction terms then the model specializes to the $\ln s_j - \ln s_0$ model of A and B. Thus if the model C estimation algorithm is run with the coefficient on travel distance dropped (or equivalently, fixed at zero), the results are the same as for model B. As a consequence of this, the coefficients in all three columns are on the same scale and can be compared directly.

All of these specifications include year fixed effects to control for population growth and statewide trends in the demand for child care.

Comparing the price coefficients in Table 1.3, we can see that price has a coefficient that is positive and statistically significant in model A, which was estimated with ordinary least squares. Taken at face value, the coefficient implies that consumers prefer to spend more rather than less on child care providers. This was to be expected given our presumption that price is correlated with unobserved quality. Model B uses instrumental variables. Here, the price coefficient is small and its difference from zero is not statistically significant. In that model there is no accounting for travel

[4] That is to say, I use BLP's nonlinear GMM estimation algorithm, but choose the weighting matrix that specializes the GMM estimator into 2SLS.

Table 1.3: Baseline Models

	OLS (A)	2SLS (B)	Nonlinear 2SLS /BLP (C)
Distance (Miles)	.	.	-0.093^{***}
			(0.002)
Price (Weekly)	0.001^{***}	-0.00004	-0.035^{***}
	(0.0002)	(0.0003)	(0.001)
Star Rating – 1	-0.182^{*}	-0.186^{*}	-0.425^{*}
	(0.096)	(0.096)	(0.249)
Star Rating – 2	-0.00001	-0.005	-0.440^{**}
	(0.075)	(0.075)	(0.194)
Star Rating – 3	0.107	0.094	-0.377
	(0.151)	(0.151)	(0.391)
Star Rating – 4	-0.007	-0.002	0.423^{***}
	(0.053)	(0.053)	(0.138)
Group – 1	0.079	0.067	-0.284
	(0.082)	(0.082)	(0.212)
Group – 2	0.059	0.049	-0.117
	(0.065)	(0.065)	(0.168)
Group – 3	0.058	0.052	-0.327
	(0.119)	(0.119)	(0.307)
Group – 4	0.037	0.033	0.131
	(0.052)	(0.052)	(0.134)
Capacity	0.006^{***}	0.006^{***}	0.011^{***}
	(0.0002)	(0.0002)	(0.0005)
USDA Food Program	-0.099^{***}	-0.103^{***}	0.035
	(0.021)	(0.021)	(0.055)
Nonprofit	0.014	0.002	-0.364^{***}
	(0.021)	(0.021)	(0.055)
Accreditation	0.070^{**}	0.099^{***}	0.671^{***}
	(0.030)	(0.033)	(0.085)
Constant	-9.345^{***}	-9.231^{***}	0.497^{***}
	(0.045)	(0.071)	(0.184)
Year F.E.	Yes	Yes	Yes
Observations	3,426	3,426	3,426

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

costs, this model treats all providers as equally substitutable with one another regardless of distance. Model C includes travel costs, so the estimates are based on a demand structure where the degree of substitutability depends on geography. In the coefficient estimates for this model, distance and price are both statistically significant and have the expected sign. The relative size of the coefficients, $\theta_{Distance}/\beta_{Price} = 2.64$, multiplying by 52 to obtain a yearly quantity, suggests that a typical consumer is willing to pay about \$137 extra per year to avoid an additional mile of distance between home and the child care provider. This substantial distance penalty is consistent with our expectation that consumers prefer child care that is within a few miles of the home.

The key coefficients of interest are the ones associated with the star ratings. The omitted category is "unrated", and so the ratings coefficients can be understood as differences from the unrated category. In Model C, the coefficient on the highest rating, 4 Stars, is 0.423. Scaling this to the coefficient on price, $\beta_{4Star}/\beta_{Price} = \$12.09$, and multiplying this by 52 implies that a typical consumer is willing to pay about \$628 extra per year in order to use a 4 Star rated provider compared to an unrated provider. On the other hand, the coefficients on the lower rating, 1-3 stars are all negative, and all have values around -0.4. The results imply that providers that receive ratings lower than 4 Stars are perceived as worse than unrated providers. Scaling these coefficients in the same way as above suggests that, for example, a typical consumer would require about \$630 of compensation per year in order to accept a 1 Star rated provider rather than an unrated. All of the coefficients on the star ratings are statistically significant except the one on 3 Stars. The non-significant coefficient on the 3 Star rating may be attributed to the fact that there are comparatively few providers with this rating.

Models With Demographics

Table 1.4 shows the results of specifications where household demographics enter into household demand but do not interact with provider characteristics. One way to think about the interpretation of this is that the household demographics affect the likelihood of choosing the outside option, but do not affect the relative attractiveness of the various inside goods.

The parameter estimates resulting from these specifications are a little perplexing. Column D

Table 1.4: Estimates with Consumer Demographics

	(D)	(E)	(F)
Distance (Miles)	-0.497***	-0.093***	-0.093***
	(0.032)	(0.008)	(0.023)
% Low Income	-5.242***	.	0.000
	(0.509)		(1.59)
% College	.	0.001	0.000
		(0.310)	(0.569)
Price (Weekly)	−0.034***	−0.035***	−0.035***
	(0.001)	(0.001)	(0.001)
Star Rating – 1	−0.477*	−0.425*	−0.425*
	(0.253)	(0.249)	(0.249)
Star Rating – 2	−0.509***	−0.440**	−0.440**
	(0.197)	(0.194)	(0.194)
Star Rating – 3	−0.425	−0.377	−0.377
	(0.397)	(0.391)	(0.391)
Star Rating – 1	0.452***	0.423***	0.423***
	(0.140)	(0.138)	(0.138)
Group – 1	−0.407*	−0.284	−0.284
	(0.216)	(0.212)	(0.212)
Group – 2	−0.076	−0.117	−0.117
	(0.171)	(0.168)	(0.168)
Group – 3	−0.490	−0.327	−0.327
	(0.312)	(0.307)	(0.307)
Group – 4	0.203	0.131	0.131
	(0.136)	(0.134)	(0.134)
Capacity	0.010***	0.011***	0.011***
	(0.0005)	(0.0005)	(0.0005)
USDA Food Program	−0.031	0.035	0.035
	(0.056)	(0.055)	(0.055)
Nonprofit	−0.326***	−0.364***	−0.364***
	(0.056)	(0.055)	(0.055)
Accreditation	0.525***	0.671***	0.671***
	(0.086)	(0.085)	(0.085)
Constant	4.036***	0.499***	0.498***
	(0.187)	(0.184)	(0.184)
Year F.E.	Yes	Yes	Yes
Observations	3,426	3,426	3,426

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

shows a specification that includes the low income variable, defined as the percentage of households in the block group whose income is less than 200% of the poverty level. The coefficient on this demographic variable is large, negative, and statistically significant, which would suggest that demand for center-based child care is in block groups with many low-income families. The estimate on the distance parameter jumps up to -0.497, compared to the value of -0.093 that was estimated for Column C of Table 1.3. However, while the model with the income demographic has these distinctive results, the models that include the college variable, defined as the percentage of people over 25 in the block group whose education level is at least a bachelors degree, are essentially the same as Column C of Table 1.3. Column E is the estimates for the model that only includes the college variable. Column F is the model that includes both demographic variables, nesting both Column D and Column E. In both of these cases the estimated coefficients on the demographics are close to zero, and hence the other parameter estimates are almost identical to Column C of Table 1.3.

I do not have a compelling explanation for why the specification in column D should be so different from the nesting specification in column F. This issue merits further investigation and consideration to ensure that the estimates presented are reliable.

Welfare Calculations

By Year

Table 1.5: Welfare Quantities

FY	Actual	Counterfactual	Adjustment	Benefit
2014	$100,975,800.42	$93,104,316.91	-$5,768,092.55	$2,103,390.96
2015	$115,533,611.15	$106,222,502.29	-$6,270,825.54	$3,040,283.31
2016	$115,804,854.12	$107,021,278.21	-$5,313,216.28	$3,470,359.63

Table 1.5 shows the calculation of the total welfare benefit from the ParentAware ratings. The first column, "Actual" shows a money scaling of the total expected utility calculated from the model as follows. First, for each block group type i, I calculate $\frac{52}{\beta_{Price}} \log \sum_j e^{V_{ij}}$, the expected utility from the model to a consumer of that type evaluated over possible values of the idiosyncratic utility draw ε_i, scaled to a money value by dividing by the price coefficient, and translated into a yearly value by multiplying by 52. These per-person values are summed, weighted by the number of children

0-5 in that block group, to give the values shown in the table. To calculate the second column, "Counterfactual", I calculate counterfactual choice utilities by starting with the estimated model and setting the coefficients on the ratings to zero. I then calculate the same log sum calculation, scaling, and summing as for the previous quantity, yielding an estimate of the expected utility from the model to a consumer choosing without the ratings information. However, I wish to calculate the value of the without-ratings choices according to the with-ratings utility values. The third column, "Adjustment", is calculated at the consumer type level as $\sum_j P_{ij} D_{ij}$, where D_{ij} is the difference between choice utility with the ratings coefficients set to zero, and the choice utility with the estimated coefficients; and P_{ij} is the probability that i will choose j in the counterfactual with the ratings coefficients set to zero; scaled and summed in the same way. "Counterfactual" - "Adjustment" gives the total expected value to consumers of their choices without the ratings, evaluated using the with-ratings utilities. "Actual" - "Counterfactual" + "Benefit" gives the compensating variation of the ratings. That is, the amount of money that would compensate consumers who had the ratings for having their choices made by consumers with identical preferences except for the ratings.

The results show a rapid growth in the total value of the ParentAware ratings to consumers, up to a yearly total value of almost \$3.5 million in fiscal year 2016. This reflects the increase in the number of rated providers, and consequently in the number of consumers whose choice is influenced by the ratings.

By Location

The value of the ratings is inherently heterogenous. Consumers whose set of available choices includes a variety of different classifications of providers are more likely to have their choice affected by the ratings. Consumers with few choices, or whose choices had uniform ratings, would be unlikely to have the ratings affect their choice and would gain little from the ratings information.

Table 1.6 shows the per-child expected benefit per consumer, aggregated to the county level. The table includes sixteen Minnesota counties. The counties shown in the table are all the counties whose share of the children age 0-5 is at least 1%, according to the ACS. The column labeled "Benefit" is calculated by evaluating the yearly expected benefit at the block group level, and then averaging over the block groups in each county, weighted by the number of children age 0-5 in each block

Table 1.6: Per-person Benefit by County (FY 2016)

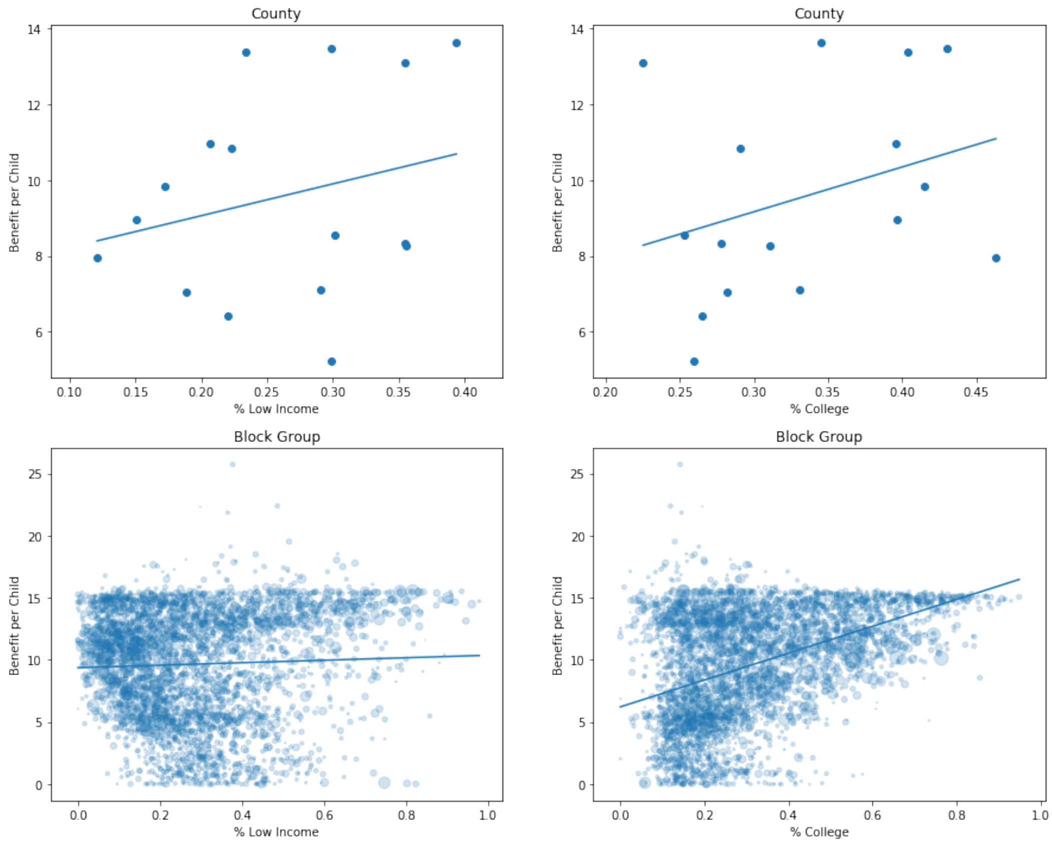
County	MSA	Benefit ($)	Child Share (%)	Benefit Share (%)
Ramsey County	Twin Cities	13.64	10.67	14.64
Hennepin County	Twin Cities	13.48	22.65	30.74
Olmsted County	Rochester	13.39	3.06	4.12
Crow Wing County	(No MSA)	13.10	1.07	1.41
Dakota County	Twin Cities	10.96	7.76	8.56
Anoka County	Twin Cities	10.84	6.17	6.73
Washington County	Twin Cities	9.84	4.39	4.35
Scott County	Twin Cities	8.95	2.95	2.65
Stearns County	St. Cloud	8.55	2.75	2.36
St. Louis County	Duluth	8.34	2.99	2.51
Blue Earth County	Mankato	8.27	1.06	0.88
Carver County	Twin Cities	7.97	1.83	1.47
Clay County	Fargo-Moorhead	7.11	1.20	0.86
Wright County	Twin Cities	7.05	2.84	2.01
Sherburne County	Twin Cities	6.41	1.77	1.14
Rice County	(No MSA)	5.24	1.06	0.56

group. The column labeled "Benefit Share" is the total expected value of ParentAware to consumers in the county divided by the statewide total, and the column labeled "Child Share" is the number of children age 0-5 in the county divided by the statewide total.

Generally, the benefits of the ratings are greatest in the urban counties. The chart is topped by the counties containing Minnesota's three most populous cities: Minneapolis (Hennepin County), St. Paul (Ramsey County), and Rochester (Olmsted County). This is in line with our expectations that the benefits will be greater in dense cities where consumers have a wider variety of child care choices. The pattern is not absolute, however. Crow Wing, (a central Minnesota county whose principle city is Brainerd) has a larger than usual number of four star centers and consequently a high per-child expected benefit from ParentAware, exceeding that of suburban counties in the Minneapolis-St. Paul metropolitan area.

Figure 1.3 describes the relationship between the estimated benefits from ParentAware and some demographic characteristics. It should be noted that these estimates are based on the model that includes the distance cost but not any demographics in the choice utility specification, and therefore the variation in local benefits depicted in figure 1.3 arises wholly from differences in local child

Figure 1.3: Local Benefits of ParentAware, by Demographics



care choice sets and provider characteristics, rather than the direct effects of demographics. The top two subplots show county level aggregates, and the bottom two subplots show block group level estimates. In the left two subplots, the horizontal axis variable is the percent low income, defined as the proportion of households whose income is less than or equal to 200% of the federal poverty level. In the right two subplots, the horizontal axis variable is the percent college, defined as the proportion of individuals aged 25 and older whose highest education level is at least a bachelors degree. For each of these variables the county level quantity has been computed by taking a weighted average of the block group level variables, with the weight depending on the number of children in the block group (rather than the number of households or adults).

The left group of subplots of figure 1.3, describing the relationship between income and the benefits of ParentAware, are particularly interesting. At the county level, there is a loose positive relationship between the income variable and the estimated benefits from ParentAware. I interpret this as reflecting the effect of density. Many block-groups with a higher percentage of low-income people are urban block groups that are denser and where there is a greater variety of child care options. Consequently, there is a higher likelihood that the ratings will affect the choice of child care. At the block group level, however, the relationship between the income variable and the benefit variable is weaker. I interpret this as reflecting a balance of the effects of density against the fact that low-income neighbourhoods contain fewer highly rated centers.

The right group of subplots of figure 1.3 describe the relationship between education and the benefits of ParentAware. They show a positive relationship at both the county and block group level. This is in line with expectations. More college-educated people live in urban areas where the variety of child care options is greater and ratings are more likely to have an effect on choice. Furthermore, areas containing more college-educated people are also more likely to contain highly rated centers.

1.7 Additional Descriptive Statistics

Table 1.7: Provider Descriptive Statistics, Fiscal Year 2014

	N	Mean	St. Dev.	Min	25%	Med.	75%	Max
Enrollment	1063.0	56.67	34.55	2.00	34.00	50.50	70.40	302.17
Weekly Price	1063.0	216.76	59.13	12.68	178.29	221.37	253.97	517.00
Licensed Capacity	1063.0	83.14	47.23	0.00	48.00	74.00	110.00	372.00
USDA Food Prog	1063.0	0.35	0.48	0.00	0.00	0.00	1.00	1.00
Non-Profit	1063.0	0.37	0.48	0.00	0.00	0.00	1.00	1.00
Accreditation	1063.0	0.32	0.47	0.00	0.00	0.00	1.00	1.00
Rating Stars	1063.0	1.11	1.75	0.00	0.00	0.00	4.00	4.00

Table 1.8: Provider Descriptive Statistics, Fiscal Year 2015

	N	Mean	St. Dev.	Min	25%	Med.	75%	Max
Enrollment	1194.0	56.86	36.65	4.00	34.09	52.58	68.5	392.67
Weekly Price	1194.0	217.95	58.28	12.08	181.93	223.37	253.2	441.99
Licensed Capacity	1194.0	80.21	59.21	0.00	43.00	70.00	108.0	1140.00
USDA Food Prog	1194.0	0.33	0.47	0.00	0.00	0.00	1.0	1.00
Non-Profit	1194.0	0.37	0.48	0.00	0.00	0.00	1.0	1.00
Accreditation	1194.0	0.31	0.46	0.00	0.00	0.00	1.0	1.00
Rating Stars	1194.0	1.33	1.83	0.00	0.00	0.00	4.0	4.00

Table 1.9: Provider Descriptive Statistics, Fiscal Year 2016

	N	Mean	St. Dev.	Min	25%	Med.	75%	Max
Enrollment	1169.0	58.17	36.24	2.38	35.33	54.00	69.23	371.65
Weekly Price	1169.0	222.21	58.13	12.53	186.01	228.76	253.48	475.68
Licensed Capacity	1169.0	83.39	58.23	0.00	46.00	73.00	110.00	1140.00
USDA Food Prog	1169.0	0.33	0.47	0.00	0.00	0.00	1.00	1.00
Non-Profit	1169.0	0.37	0.48	0.00	0.00	0.00	1.00	1.00
Accreditation	1169.0	0.35	0.48	0.00	0.00	0.00	1.00	1.00
Rating Stars	1169.0	1.46	1.85	0.00	0.00	0.00	4.00	4.00

1.8 First Stage Regression of Price on Instruments

Table 1.10: Coefficients From First Stage Regression

	Price
Capacity	0.191^{***}
	(0.013)
USDA Food Prog.	-3.006^{*}
	(1.570)
Nonprofit	-15.386^{***}
	(1.494)
Accreditation	29.730^{***}
	(2.123)
Rating – 1 Star	-7.494
	(7.085)
Rating – 2 Star	-5.379
	(5.513)
Rating – 3 Star	-12.214
	(11.129)
Rating – 4 Star	12.482^{***}
	(3.915)
Group – 1 Star	-15.552^{***}
	(6.034)
Group – 2 Star	-14.526^{***}
	(4.767)
Group – 3 Star	-12.501
	(8.752)
Group – 4 Star	-4.313
	(3.826)
Z from Constant	-0.094^{***}
	(0.032)
Z from Capacity	0.010^{***}
	(0.001)
Z from Nonprofit	0.008
	(0.040)
Z from Accreditation	0.379^{***}
	(0.146)
Z from Rated Status	-0.641^{***}
	(0.130)
Constant	151.704^{***}
	(2.190)
Year F.E.	Yes
Observations	3,426
R^2	0.536
Note:	$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

1.9 Closed Form Calculation of Welfare Quantities

Let A and B be the two different information regimes. A is without the ratings, and B is with the ratings. $a(\varepsilon)$ and $b(\varepsilon)$ are decision rules, that we will shortly associate with A and B. ζ_j^A and ζ_j^B are the mean utility values. Following Jin and Sorensen (2006), the value of the ratings is

$$\frac{1}{\alpha}\left(V^B(b) - V^B(a)\right).$$

This quantity can be calculated by simulating many draws of ε. However, another way is to derive closed form expressions. In order to do this it is helpful to write $\zeta_j^B = \zeta_j^A + D_j$. Then we can break down $V^B(a)$,

$$V^B(a) = E_\varepsilon\left[\zeta_j^B + \varepsilon_j | \zeta_j^A + \varepsilon_j = \max_k \zeta_k^A + \varepsilon_k\right].$$

Then, splitting up ζ^B,

$$V^B(a) = E_\varepsilon\left[\zeta_j^A + \varepsilon_j | \zeta_j^A + \varepsilon_j = \max_k \zeta_k^A + \varepsilon_k\right] + E_\varepsilon\left[D_j | \zeta_j^A + \varepsilon_j = \max_k \zeta_k^A + \varepsilon_k\right],$$

Substituting well-known expressions, (and writing out $D_j = \zeta_j^B - \zeta_j^A$)

$$V^B(a) = \log\left(\sum_k \exp \zeta_k^A\right) + \sum_j \left(\frac{\exp \zeta_j^A}{\sum_k \exp \zeta_k^A}\right)(\zeta_j^B - \zeta_j^A)$$

Then the value of the ratings is

$$\frac{1}{\alpha}\left[\log\left(\sum_k \exp \zeta_k^B\right) - \log\left(\sum_k \exp \zeta_k^A\right) - \sum_j \left(\frac{\exp \zeta_j^A}{\sum_k \exp \zeta_k^A}\right)(\zeta_j^B - \zeta_j^A)\right)$$

Chapter 2

How Far Will They Go?

2.1 Introduction

Most children under age 5 attend child care or early learning programs on a regular basis (Laughlin, 2013), and unlike K-12 public schools, parents typically must provide transportation to get their children there and back home again. For a child attending five days per week, that can mean ten trips between home (or work) and the child care provider's location per week.[1] Not surprisingly, then, parents identify convenience and location as key factors in determining which child care provider they use (Tang et al., 2012). While studies show that parents on average use child care settings close to home, no studies to date have been able to directly measure the trade-off between distance and other provider characteristics in choosing child care. Given the importance of early childhood experiences for child development and later outcomes (Chaudry et al. (2017); Duncan and Sojourner (2013)), understanding the role of distance in parents' decisions may have implications for policy. If distance greatly affects parent decisions, then local supply matters a great deal. While there is a large literature examining factors influencing parents' use of child care, these studies typically have limited data on the numbers and types of providers locally available near families. Previous studies tend to observe each family's chosen provider but not the set of providers from which the family chose this provider, implicitly assuming that all families face the same set of options.

[1]This chapter is based on coauthored work with Elizabeth Davis

How far are households willing to travel in order to access child care providers? Evidence from the 2012 National Survey of Early Care and Education confirms that child care choice is very local and most consumers using market-based care use a provider within five miles from the home (National Survey of Early Care and Education Project Team, 2016). The fact that chosen arrangements tend to be close to the home may reflect either a propensity of households to choose a provider that is close, or the fact that most households do not need to travel in order to access a desirable provider. From a policy perspective, what matters most is the *potential* distance that families would travel, if needed, rather than the actual distance that they do travel. Yet there has been relatively little direct study of the role of distance in child care provider choice.

This paper addresses that gap in understanding using matched consumer/provider data from Minnesota's Child Care Assistance Program ("CCAP"). We study all children ages 0-4 who received subsidies from CCAP in September 2015. CCAP is Minnesota's main assistance program implementing the federal Child Care Development Block Grant (CCDBG) Act, and receives funding from both the federal government and the state of Minnesota. Eligibility for CCAP depends on income and parental activity requirements that can be satisfied by employment, education, training, or employment search.

We estimate a nested logit model of provider demand, where the probability that a particular consumer chooses a particular provider depends on the consumer's characteristics, the provider's individual characteristics as well as provider type, and interactions between consumer and provider variables. Importantly, we include consumer/provider distance as a determinant of the demand function, modeling the fact that consumers may trade distance against other desirable or undesirable provider characteristics. Estimates from the model summarize the relationship between the characteristics of the provider, including the distance between the provider and a particular household, and the probability that household will use that provider. This provides information about how households trade distance against desirable and undesirable provider characteristics, and about how variation in what is locally available shapes the characteristics of the providers that are chosen.

Our research also informs the question of what market definitions or geographic ranges are most appropriate in order to evaluate local differences in access. Market definition in the child care market is of importance for numerous research and practical applications. A social scientist wishing to understand patterns in the use of center-based care may want to control for local differences in what is available. A subsidy program administrator may want to estimate differences in local access in order to carry out a legislative mandate to promote equality of access. A policy analyst seeking to estimate the effect of a local program may want to assign locations as treatment or control group based on proximity to the program site. Suppose, for example, an analyst is studying the effects of public pre-K programs on private providers, and wishes to assign which providers are "treated" by the establishment of a particular pre-K. The geographic range over which that program will have an effect depends on the distribution of consumers and competing providers. A program in a rural area that is less dense can be expected to have an effect over a wider area. This model provides a lens for understanding the relevant geographic range for policy analysis and how it changes depending on local characteristics. We develop a measure of the relevant geographic range for a policy intervention based on the estimated choice model and discuss how this measure varies in different areas. Our results suggest that in most areas, a six mile radius would be required to capture 50% of the impact of introducing a new center provider. This impact radius is smaller in some urban areas and in rural cities and towns. In rural areas, the most relevant local region may be the catchment area around each city or town.

Finally, the model allows us to examine how variation in what is locally available shapes the type of provider chosen. For example, Black families tend to choose childcare centers more frequently than family child care providers. However, the reasons for this choice pattern are unclear. One possible explanation is geographic variation in the availability of center-based care. Our estimates allow us to asses how much of the variation in use of center-based care depends on variation in the types of care that are locally available, and how much variation remains after controlling for what is locally available. We show that controlling for locally available supply through the nested logit model substantially reduced estimated differences between race groups in the propensity to use family provider care, compared to alternatives that do not control for the composition of local supply or that attempt to control for the composition of local supply using zip-code or county-based numbers of providers. By simulating a counterfactual, we determine the relative importance of location and demographic differences, compared to remaining behavioral differences after local supply is controlled for. We find that location, income, and work status together are much more important than the unexplained behavioral differences, explaining about four-fifths of the observed gaps in use of center-based care between White and Black households and between White and Hispanic households.

2.2 Literature Review

The existing literature on child care examines many factors that play a role in the decisions families make with regard to non-parental care for their children, including whether to use non-parental care, what type and how much (e.g., Blau and Hagy (1998); Tekin (2005); Tekin (2007); Connelly (1992)). The factors shown to be important in child care selection include time and resource constraints as well as child and family needs and parental preferences (Coley et al. (2014); Davis and Connelly (2005); Henly and Lambert (2005); Li-Grining and Coley (2006)). Observed patterns of child care use differ across income groups, with many studies finding that low-income families are less likely to use center-based care than high-income families (Burstein and Layzer (2007); Chaudry (2004); Chaudry et al. (2011); Lippman et al. (2008); Tang et al. (2012). The type of care used has also been shown to be related to a number of other family and child characteristics, including child age, mother's education, race and ethnicity, and family structure as well as parent's stated preferences (Brewster and Padavic (2002); Coley et al. (2014), Coneus et al. (2008); Early and Burchinal (2001); Han (2004); Kim and Fram (2009); Liang et al. (2000); Loeb et al. (2004); Lowe and Weisner (2004); Morrissey (2017); Shlay et al. (2005)).

Numerous studies also show that there is variation in the supply and the characteristics of that supply across communities (Gordon and Chase-Lansdale (2001), Davis et al. (2019)). Yet few studies of child care selection decisions take into account the characteristics of the providers nearby the family. In particular, the role of distance in child-care decision-making is under-studied. Most non-parental child care takes place outside the child's home, and the location of particular providers (and available modes of transport) likely play an important role in the family's decision about which care provider to use. Indeed, studies estimate that families do not travel far for child care. In a national study, on average, center-based arrangements were within five miles of home for those using out-of-home care for children age 5 and younger (NSECE Team 2016). In a study of families in Minnesota using child care outside the home, the average travel time (one way) for the youngest child was under ten minutes (Chase and Shelton, 2001). While descriptive information suggests parents use child care providers close to home, no studies directly measure the trade-off between distance and other provider characteristics in choosing child care. Further, the literature on child care selection decisions has largely been focused on the decision between types or modes of care, typically categorized as center-based care, family child care, and care by friends and relatives. These care types differ in ways that are important to parents, including the cost, home or school-like setting, number and age-mix of children, and child-adult ratio. But this approach is limited in helping us understand why parents choose a specific provider, that is, how the parents make trade offs among the characteristics they care about, including convenience (location and hours), quality, and price in the context of what is locally available to them.

This study builds on the existing literature in three main ways. First, most studies of child care choice are missing measures of distance to nearby providers, which is likely to be an important driver of the choices parents make. The distance measure we use is based on the locations of home and provider, rather than self-reports by parents of driving time or distance. Second, few studies have measures of provider quality, and others use measures or proxies for quality that are self-reported by parents. The provider data in Minnesota include ratings from the Parent Aware Quality Rating Information System (QRIS). Finally, we analyze parents' use of a specific provider, whereas all previous studies have analyzed only the type of care chosen. Thus we build on the prior literature on child care choice by accounting more fully for the amount and characteristics (including distance and quality) of the supply near the family and how that influences their decision to use a specific child care provider.

2.3 Data

Data Sources

We draw on four different data sources for details about CCAP-subsidized consumers, their provider choices, and the providers themselves.

Our data source for information on subsidized consumers and their providers is administrative data from the CCAP program, made available under a data sharing agreement with Minnesota DHS.

This administrative data consists of matched child, household, and provider level data pertaining to a period that includes September 2015. The data files contain a variety of child and household characteristics such as the child's race, the household's address, the household's monthly income reported to DHS for eligibility purposes, and flags for whether the reported income includes earnings from work. The data summarizes subsidy payments associated with every CCAP child, including the amount of subsidy for each provider associated with that child for subsidized care in each month. This allows us to determine which providers are used by each child in each month.

We assembled data on Minnesota child care providers from three sources. First, DHS provided us with licensing administrative data for all licensed child care centers and licensed family providers active during our study period. This data set included the most complete information on which providers were active in September 2015, their addresses and types, and the providers' total licensed capacity.

Second, Minnesota DHS provided administrative data on Parent Aware ratings for all participating providers. Parent Aware is Minnesota's QRIS. Participating providers are assessed by the regulator and assigned a rating of One, Two, Three, or Four Stars. These ratings are prominently displayed on the Parent Aware provider search website maintained by the state.

Third, we used a "snapshot" of NACCRRAware in September 2015. NACCRRAware is an administrative database maintained by Child Care Aware of America, primarily for child care resource and referral agencies in providing referral services. This data set provides information on provider characteristics that is not available from other sources, such as whether a provider is willing to accept child care assistance, whether the provider has a quality accreditation from a private accreditation agency such as the National Association for the Education of Young Children (NAEYC), some information about the provider's schedule of services, and for family providers some demographic information that includes race. The NACCRAware data includes information on the prices charged by individual providers that is collected as part of Minnesota's market rate survey process. Variables in this data set are self-reported by the providers.

Data on Providers

We define our provider sample as all Minnesota licensed child care providers active in September 2015 who were accepting CCAP consumers. Minnesota licensed providers are either child care centers or family child care providers. Child care centers are generally larger, are located on business premises, and may have multiple staff members and classrooms. Family child care providers are generally smaller, with a maximum total licensed capacity of 14 children, are located in the provider's home, and are staffed by either one or two providers.

Since not all licensed providers accept CCAP, we used two complementary criteria to evaluate whether a provider should be included in the sample. First, we used a variable from the NACCRRAware data that described whether providers were willing to accept child care assistance. A provider was included if their value for this variable indicated either that they were ”Willing to Accept Child Care Assistance” or that they were ”Currently Serving Child Care Assistance Children”. 8979 providers, including 1281 centers and 7428 providers, met this criterion. Second, we included any providers who, according to the CCAP data, were actually used by at least one CCAP consumer, regardless of how they were listed in the NACCRRAware data. This added 269 additional providers to the sample, including 83 centers and 187 family providers. Of the 1364 centers included in the sample, 798 or 55% actually served at least one CCAP consumer according to the CCAP data. Of the 7615 family providers included in the sample, 1696 or 22% actually served at least one CCAP consumer.

Table 2.1 shows some descriptive statistics on the provider characteristic variables used in the analysis.

The first group of variables are measures of provider quality. All of these measures are binary dummy variables. Rating 1-4 refers to Parent Aware Star Ratings. It can be seen that a large proportion of centers in the sample are rated, and that among rated centers, Four Stars is the most common rating. Ratings are much less common among family child care providers in the sample, and family providers are more likely to have a One Star or Two Star rating than a Three Star or Four Star rating. Accreditation refers to any of a number of private quality accreditation systems for child care providers. A provider was coded as accredited if the accreditation variable in the NACCRRAware

Table 2.1: Descriptive Statistics on Providers

	Center		Family		All	
	mean	std	mean	std	mean	std
Rating 1	0.021	0.144	0.033	0.178	0.031	0.174
Rating 2	0.049	0.216	0.030	0.170	0.033	0.178
Rating 3	0.016	0.126	0.012	0.108	0.012	0.111
Rating 4	0.301	0.459	0.019	0.135	0.061	0.240
Accreditation	0.363	0.481	0.005	0.072	0.060	0.237
Building Quality	0.059	0.235	0.054	0.225	0.054	0.227
Chain	0.317	0.466	0.000	0.000	0.048	0.214
Nonprofit	0.444	0.497	0.002	0.043	0.069	0.254
Hours / day	10.308	3.038	10.809	1.981	10.733	2.182
Days / week	4.964	0.578	5.025	0.356	5.016	0.398
Margin = min{Price - subsidy,0}	25.813	23.916	13.260	17.550	15.167	19.192
Margin imputed	0.531	0.499	0.266	0.442	0.306	0.461
Capacity	75.722	49.914	11.864	1.519	21.565	30.093
log(Capacity)	4.105	0.701	2.465	0.133	2.714	0.660
FCC Asian	0.000	0.000	0.010	0.100	0.009	0.092
FCC Black	0.000	0.000	0.023	0.149	0.019	0.138
FCC Hispanic	0.000	0.000	0.009	0.096	0.008	0.089
FCC Multi/Native/Oth./Unrep.	0.000	0.000	0.080	0.271	0.068	0.251
FCC White	0.000	0.000	0.878	0.328	0.744	0.436

data had any value, or if the provider was listed in the Parent Aware data as participating in the Accelerated Pathway rating subprogram, which requires accreditation. 36.3% of centers in the sample are accredited, compared to only 0.5% of family providers. Building Quality is a dummy variable for participation in a Parent Aware program that gives providers grants to obtain ratings or increase their ratings.

The second group of variables describe additional provider characteristics. The chain variable identifies multi-establishment firms. It is coded based on a license holder entity identifier present in the licensing administrative data. We classify a provider as belonging to a chain if it belongs to a license holder identifier number that is associated with more than one provider. 31.7% of centers in the data are coded as chains. The variables "Hours/day" and "Days/week" are coded based on scheduling information in the NACCRRAware data. Hours/day is the difference between the closing time and opening time reported for the provider, averaged over the days that the provider was open. This schedule information was missing for about 3.2% of the providers in the sample. These missing values were filled using the sample mean for each variable.

"Margin" is the price variable in the data, representing the difference between the price charged by the provider and the maximum subsidy reimbursement available from CCAP. Child care prices generally depend on age group (e.g. infant, toddler, and preschool) and timing mode (hourly, daily, weekly). In order to have a single price for each provider, we used the weekly toddler price since the toddler age group was most numerous in our consumer sample. When a provider did not report a weekly price mode, but did report a daily price mode we converted daily to weekly by multiplying the value by five. When only an hourly price mode was reported we converted to weekly by multiplying the value by 40. Since our consumer sample consists of subsidized households, we reduced the price by the amount of the subsidy. CCAP reimbursement rates are set for child care centers and family providers at the county level. In addition, higher reimbursement rates are paid to providers that have certain Parent Aware ratings, accreditation, or (for family providers) child-related educational accreditation. For each provider, we determined the applicable weekly reimbursement cap for the toddler age group based on the provider's county, type, and characteristics. If the weekly toddler price was higher than this reimbursement cap, we set the margin equal to the difference. If the weekly toddler price was lower than the reimbursement cap we set the margin to zero.

Unfortunately, the NACCRRAware price data is very incomplete. 53.1% of the centers in our sample and 26.6% of the family providers did not have a toddler price value. In these cases we set the margin variable equal to the average margin for providers of the same type (center or family provider) in the same county, for providers which had price data was available. In the small number of cases when there were no providers of the same type in the same county we set the margin equal to the sample average for providers of that type. Observations where the margin was filled in this way are marked with the "margin imputed" dummy variable.

The last group of variables describes the race of family providers in the data. Family provider race was coded based on a more detailed race description in the Nware data. Notably 87.8% of family providers are White, and only 2.3% are Black.

We determine provider locations by using ArcGIS software to geocode the provider addresses that were reported in the licensing administrative data.

Consumers

We define our consumer[2] sample as all children age 0-4 who received CCAP subsidies in September 2015 who used a Minnesota licensed provider. We focused on consumers age 0-4 in order to exclude school age children, who may tend to use different types of providers and have different choice patterns than younger children.

For compatibility with the available provider data, we restricted our attention to consumers using providers present in the licensing data. Aside from Minnesota licensed centers and family providers, other classes of provider that may receive CCAP subsidies are providers licensed by other states, providers licensed by Native American Tribes, license exempt child care centers (which include some school based programs, Head Starts, and programs that primarily serve school age children), and legal nonlicensed providers (a category that Minnesota regulations define as providers who serve only a single family and relatives providing paid care, and who meet certain other regulatory standards). We dropped 568 or 3.9% of the available consumer observations because they used providers other than those in the licensing data. Of these the largest group was legal nonlicensed providers, who were used by 336 dropped consumers.

Table 2.2: Descriptive Statistics on Consumers

	Count	Mean	Std
Family Income (thousands)	13880	2.144	1.227
Work earnings dummy	13880	0.873	0.333
Asian/Pacific	13880	0.020	0.140
Black	13880	0.449	0.497
Hispanic	13880	0.054	0.226
Multiracial	13880	0.065	0.247
Native American	13880	0.012	0.109
Unreported	13880	0.067	0.250
White	13880	0.333	0.471
Rural	13880	0.273	0.446

Table 2.2 shows some descriptive statistics for consumer demographics used in the analysis.

[2] Throughout this paper we use the term "consumer" rather than "household" to refer to the decision-making agent. In the child care market, the beneficiary of the service will generally be the child, and the decision-maker the parents or the household as a whole. Some of our demographic regressors pertain to the child, and some pertain to the household. To simplify the terminology we call both the child and household the "consumer".

Family income is monthly and measured in thousands of dollars. Income is based on values reported to CCAP for the purpose of determining eligibility. For a small number of observations income was missing or had unreasonable values in the original data. The missing values reflect situations where CCAP did not have valid eligibility records for that household in that month. 139 missing income values, approximately 1% of observations, were filled by linear interpolation observations of the same household in prior and subsequent months that did have income data. Observations where it was not possible to interpolate an income value were dropped from the data. Two outlying observations were dropped because of extreme values. One of these recorded a negative income, and the other an income that was too high to be compatible with CCAP eligibility for the family size reported.

The work earnings dummy is an indicator for whether any of the reported income was based on work earnings, which we include as a measure of whether a parent in the household is employed.

We used derived race categories present in the CCAP administrative data that combine reported race and Hispanic ethnicity into a single variable. Thus, “White” means “non-Hispanic White”, “Black” means “non-Hispanic Black” and “Hispanic” means “Hispanic of any race”. The largest racial groups in the data are Black (44.9%) and White (33.3%).

We assigned locations to the households by geocoding the address fields available in the restricted access administrative data using ArcMap and an address locator provided by USpatial at the University of Minnesota. To preserve household anonymity, we deidentified the exact locations by reassigning each household to the nearest location on a 1/4 mile grid.

We assigned each de-identified location a rurality status. We assigned rurality based on US Census Bureau classifications of census blocks into Urbanized Areas, Urban Clusters, and residual Rural Areas. These classifications provide a much more fine-grained classification of urbanization than alternative definitions that are based on grouping counties. We classified a location as rural if according to the Census definitions the location is in a census block that is not part of an Urbanized Area. In other words, we identified both locations that are Rural or those that are part of an Urban Cluster as rural. Urban Clusters include areas of population density whose total population is between 2,500 and 50,000, and can reasonably be thought of as small towns.

Table 2.3 shows the number of CCAP children by race and rurality. Rurality is correlated with

Table 2.3: Consumers by Race and Rurality

	Urban	Rural
Asian/Pacific	268	11
Black	5993	235
Hispanic	524	227
Multiracial	666	241
Native	74	92
Unreported	686	246
White	1877	2740

race in this population. An overwhelming majority of Black CCAP children are urban (96%). In contrast a majority (59%) of White CCAP children are rural.

For each child in the data we determined a single "primary" provider. 10.9% of the children in the sample used more than one provider in September 2015. There are two reasons why this might be the case. First, the family might simultaneously use multiple providers for that child, for example using one provider for standard hours care and a different provider for after-hours care. Second, the child might have switched providers during September 2015. Whenever a child was associated with more than one provider, we selected the provider that received the largest subsidy payment as the primary provider for the month.

Distances

For each household location and each provider location, we computed the geodesic distance in miles using the geopy python package. Geopy calculates distance from latitude and longitude using the algorithm of Karney (2013) based on an ellipsoidal model of the earth that is more accurate than great-circle distance.

Figure 2.1 shows a frequency histogram of the actual distances between consumers and the providers they use, by type of provider. For 24.9% of consumers the provider is within 1 mile of the home. For 76.5% of consumers the provider is within 5 miles of the home. Longer travel distances are more common for consumers using centers than consumers using family providers. Figure 2.2 shows a similar frequency histogram by whether the household is urban or rural. Interestingly, very

Figure 2.1: Distribution of Consumer/Provider Distances by Type of Provider

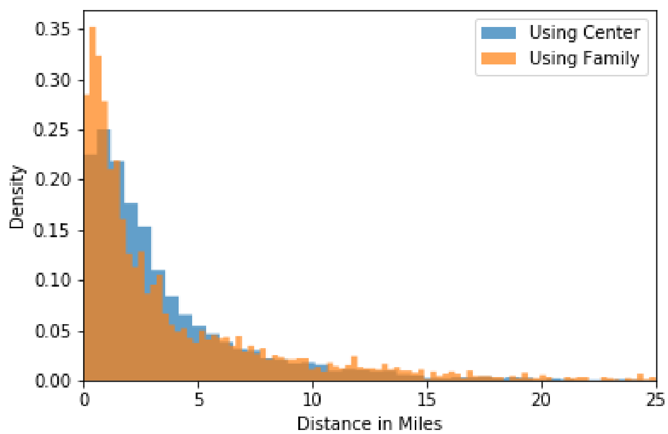
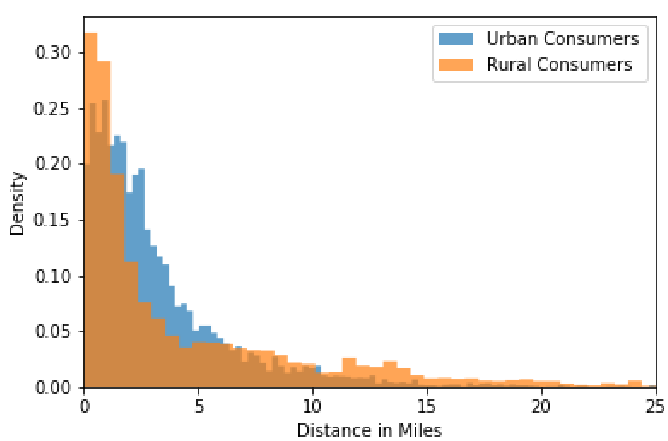
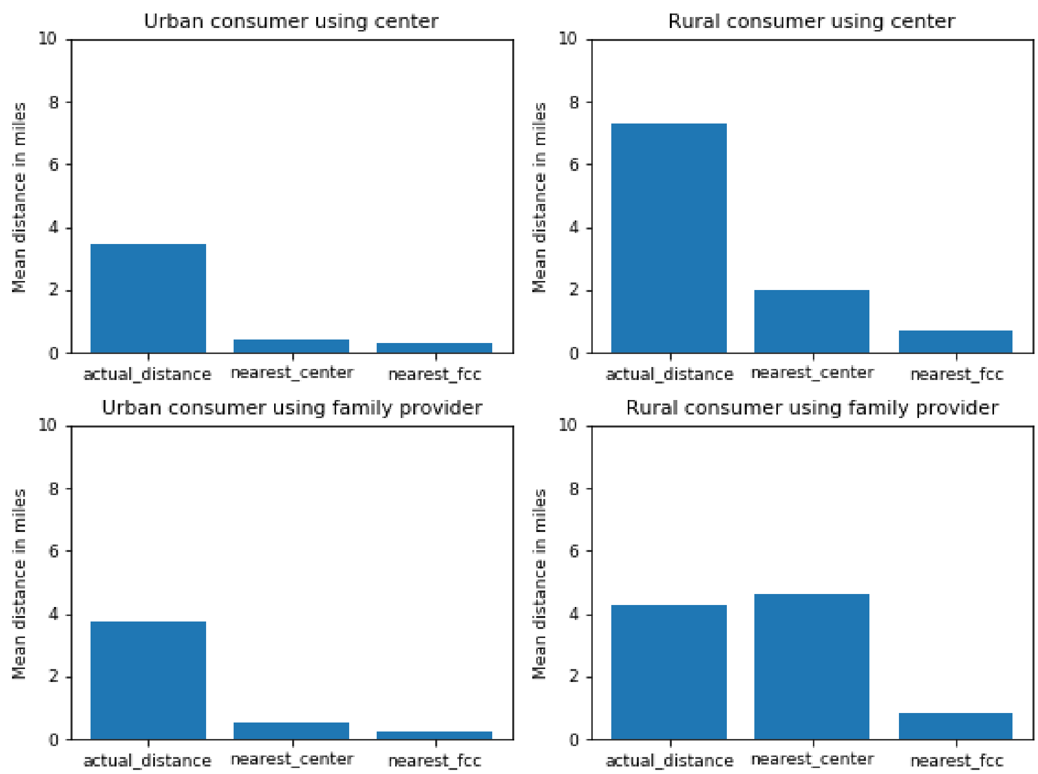


Figure 2.2: Distribution of Consumer/Provider Distances by Urban/Rural Consumer



short and very long distances are both more common among rural consumers than urban consumers. We conjecture that consumers living in small rural cities and towns are more limited to the care options within that small city or town, whereas consumers in larger cities have more scope to access a variety of options in nearby areas.

Figure 2.3: Actual and Minimum Distances by Provider Type and Urban/Rural



Even though most travel distances are short, consumers usually do not use the closest provider. Figure 2.3 compares the mean distance to the nearest provider of each type compared to the mean actual distance among consumers using a provider of that type. Among all urban CCAP consumers, the mean distance to the nearest center and family provider are 0.43 and 0.30 miles respectively. Among all rural CCAP consumers, the mean distance to the nearest center is 3.6 miles but the mean distance to the nearest family provider is only 0.8 miles.

2.4 Model

Following a standard random utility approach to discrete choice, we assume that each consumer $i \in I$ can choose any provider $j \in J$, where the set J comprises all the providers in the data, plus an outside option indexed by zero. The utility when consumer i chooses provider j has an observable and an unobservable component, $U_{ij} = V_{ij} + \varepsilon_{ij}$. We make assumptions on the distribution of the ε_{ij} so that the probability that j gives the highest utility to consumer i is given by the nested logit link function. The values are normalized by the outside option so that $V_{i0} = 0$. V_{ij} is a function of observables specified by the regression equation.

In nested logit, providers are assumed to be partitioned into nests. Here the nests are centers, family providers, and the outside option. Let M be the set of nests and J_m the set of providers in nest m.

Link Function

The link function can be understood in two levels. Consider the probability that consumer i chooses provider j, belonging to nest n:

$$P_{ij} = \frac{e^{V_{ij}/\lambda_n}}{\sum_{k \in J_n} e^{V_{ik}/\lambda_n}} \quad \frac{\left(\sum_{k \in J_n} e^{V_{ik}/\lambda_n}\right)^{\lambda_n}}{\sum_{m \in M} \left(\sum_{k \in J_m} e^{V_{ik}/\lambda_m}\right)^{\lambda_m}}$$

The lower model is the conditional probability; the probability of j given that the consumer chooses some provider in nest n. This conditional probability depends only on the characteristics of the providers in nest n, and is given by the standard multinomial logit choice probabilities.

The upper model is the marginal probability; the probability that nest n is chosen. This probability depends on the characteristics of all providers in the choice set. However the total nest value $\sum_{k \in J_k} e^{V_{ik}/\lambda_n}$ is a sufficient statistic for the role of providers in each nest in determining the marginal probability.

The nest parameters λ_n, whose values are typically between 0 and 1, govern the extent to which consumers would substitute away from nest n if n's nest value is reduced. In the special case where $\lambda_n = 1$ for all $n \in M$, the denominator of the lower model cancels with the numerator of the upper model, leaving the ordinary multinomial logit choice probabilities (without nests). In the opposite special case where $\lambda_n \to 0$ for all $n \in M$, the marginal probabilities are fixed and consumers will not substitute between nests. λ_n is closely related to the extent to which the choice error term ε_{ij} is correlated within the nest. λ_n = 1 corresponds to no correlation of the values of ε_{ij} and ε_{ik} for j, k in the same nest n. $\lambda_n = 0$ corresponds to perfect correlation of the values of ε_{ij} and ε_{ik} for all $j, k \in J_n$.

Regression Equation

The regression equation is a model of the component of the probability of consumer i choosing provider j that depends on observed characteristics of i and j. The regression equation can be interpreted as the mean utility of i choosing j, conditional on the observed characteristics.

$$\begin{aligned} V_{ik} &= X_{ij}\theta_n \\ &= \text{demographics}_i \alpha_n + \text{characteristics}_j \beta + \text{interactions}_{ij}\gamma \end{aligned}$$

Here, we use the term demographics to mean consumer-level attributes such as the age of the child or the household income. The demographics terms in the regression equation are multiplied by a nest-specific coefficient. These terms will cancel out in the lower model but not in the upper model. The coefficients can be interpreted as a relationship between household demographics and the propensity to choose a given type of provider.

We use the term characteristics to mean provider-level attributes such as the provider's QRIS rating.

We use the term interactions to mean any functions of consumer and provider attributes. Thus "interactions" include distance, which is a function of the consumer and provider locations.

2.5 Estimation

We estimate the model using maximum likelihood. The likelihood function is

$$\mathcal{L} = \prod_{i \in I} \prod_{j \in J} P_{ij}(\theta)^{Y_{ij}}$$

This gives the log-likelihood function

$$\log \mathcal{L} = \sum_{i \in I} \sum_{j \in J} Y_{ij} \log P_{ij}(\theta)$$

2.6 Demand Model Results

This section presents estimates from the nested logit choice model and examines how well the model represents patterns of distance in the data. We present coefficient estimates from two versions of the model: a baseline that assumes all consumers respond similarly to distance, and a variation that allows urban and rural consumers to respond to distance differently. The results show that all consumers are very sensitive to distance, but urban households are much more sensitive than rural ones. We show that the model does a good job at representing the relationship between proximity to providers of different types and the choice of center or family child care. However, there is substantial regional variation in average travel distances that the model does not explain. Finally, we present results from estimating the model separately on urban and rural consumers. These results confirm that rural consumers respond differently to distance.

Baseline Estimates

Table 2.4 presents the results from two specifications of the nested logit demand model. Specification (1) includes consumer demographics that modify the probability of choosing family providers, as well as a range of provider characteristics, and the distance between consumer and provider. Specification (2) includes all of the same regressors as specification (1), plus a dummy variable for rural households that interacts with distance.

The estimates of primary interest are the coefficients on distance. These estimates show that distance is an extremely important factor in provider choice. Since the probability of choosing any given provider is small, the coefficients have an approximate interpretation as semi-elasticities. Thus the magnitude of the coefficient on distance in specification (1) is that an extra mile of distance between consumer and provider will be associated with a 45.9% reduction in the probability that provider is chosen. The highly significant interaction between the rural consumer dummy and distance shows that rural households are much less sensitive to distance than urban households. Summing the dis-

Table 2.4: Results from Full Sample Estimation of Model

	(1) Baseline	(2) Interactions
Family	-2.5351**	-1.7598**
	(0.1335)	(0.2728)
Household income (thousand $) x Family	0.0533	0.0535
	(0.0403)	(0.039)
Work earnings dummy x Family	0.0964	0.079
	(0.1654)	(0.0986)
Asian/Pacific x Family	3.2541**	3.263**
	(0.2713)	(0.2431)
Hispanic x Family	0.4028	0.2384*
	(0.3037)	(0.1229)
Multiracial x Family	0.3728**	0.2316*
	(0.1457)	(0.1246)
Native American x Family	0.9105**	0.6528**
	(0.2331)	(0.2034)
Unreported x Family	0.5282**	0.3911
	(0.0866)	(0.3021)
White x Family	0.8833**	0.6467**
	(0.1391)	(0.1115)
Rural x Family		0.3782**
		(0.1485)
Rating 1	0.6408**	0.6777**
	(0.2696)	(0.1076)
Rating 2	0.6575**	0.6495**
	(0.126)	(0.1096)
Rating 3	0.3467**	0.2746
	(0.0372)	(0.2468)
Rating 4	0.8578**	0.8112**
	(0.0352)	(0.0907)
Accreditation	-0.2528**	-0.2834**
	(0.0749)	(0.086)
Building Quality	-0.4062**	-0.3532**
	(0.0478)	(0.0794)
Chain	-0.066	-0.0679
	(0.0625)	(0.0528)
Nonprofit	-0.9588**	-0.8914**
	(0.2506)	(0.0708)
Hours / day	0.1438**	0.1258**
	(0.0132)	(0.0118)
Days / week	0.3382**	0.3043**
	(0.0297)	(0.047)
Margin = min{Price - subsidy,0}	-0.0021*	-0.0022**
	(0.0012)	(0.001)
Margin imputed	-0.0658	-0.0695
	(0.0637)	(0.0501)
log(capacity)	1.0756**	1.0347**
	(0.0792)	(0.062)
Distance (miles)	-0.4588**	-0.5593**
	(0.0372)	(0.0343)
Distance x Rural		0.2707**
		(0.0316)
λ-Center	1.2918**	1.2556**
	(0.0261)	(0.0605)
λ-Family	1.3479**	1.1549**
	(0.0315)	(0.0871)

Robust (sandwich) standard errors in parentheses tance coefficient and the interaction, the approximate effect of an extra mile of distance for a rural consumer is a 28.9% reduction in the choice probability. Urban consumers, who are the reference category, are almost twice as responsive to distance, at 55.9%. The large difference between urban and rural consumers may reflect several different causal mechanisms. Access to transportation is likely to be an important component of this difference. The costs of car ownership are lower in rural areas, so low-income households in rural areas may be more likely to have a car. The use of a straight-line measure of distance may introduce a distortion in the measure of distance cost where urban miles are more costly to travel than rural miles because of traffic, speed limits, or road layouts. Finally, there may be a role for differences in preferences: perhaps people who don't mind driving a lot are more likely to locate or remain living in rural Minnesota. Further research is necessary to understand the large difference in distance-responsiveness between urban and rural consumers and how it may affect the measurement of child care access. Unless otherwise stated, subsequent parts of this paper will rely on the version of the model described in specification (2) of Table 2.4.

The coefficients on provider demographics can be interpreted as the relationship between those demographics and the propensity to choose family providers over centers. Neither the coefficient on household income nor the coefficient on work status is statistically significant. All of the race coefficients are positive, and most are statistically significant. Black consumers, the largest race group in the sample, are the reference category for these coefficients, so the estimates mean that consumers from other race groups are more likely to choose family child care providers than Black consumers. These race-based choice patterns will be discussed in more detail in a subsequent section of the paper.

The coefficients estimated on measures of provider quality generally have the expected signs, with some exceptions. All of the coefficients on Parent Aware quality ratings are positive, and the coefficients on ratings of 1, 2, or 4 Stars are highly statistically significant, indicating that consumers are more likely to choose rated providers than unrated providers, which are the reference category. The coefficient on the 3 Star rating is not statistically significant and has a smaller magnitude than any of the other ratings. Since 3 Stars is the least common rating, this coefficient is not as precisely estimated, and the 95% confidence interval includes all of the other ratings point estimates as well as zero. The relative magnitude of the coefficients on distance and the 4 Star rating suggest that a typical urban consumer is willing to travel about 1.5 additional miles in order to access a 4 Star provider. This suggests that quality, as measured by the ratings, plays an important role in the provider choices of subsidized consumers. It is surprising however, that the dummy variable for private quality accreditation is negative and highly statistically significant, as is the dummy variable for nonprofit providers. One possible explanation is that the providers characterized by these variables serve specialized populations other than the CCAP-subsidized population. Since 4 Star providers receive a higher subsidy reimbursement, and accredited providers have the option of an accelerated pathway to 4 Star rating, a provider that is accredited but not rated may tend to be one that, for reasons not observed in our data, does not expect to serve many CCAP consumers. The coefficient on Building Quality, a grants program and assistance program that is part of Parent Aware and is designed to help consumers improve their rating, is negative and highly significant. Consumers may view Building Quality participation as a negative quality signal, or the relationship could be confounded by less successful providers being more likely to participate in Building Quality.

The estimated coefficients on measures of provider availability are generally consistent with expectations. The coefficients on the provider schedule variables, "hours/day" and "days/week" are also positive and both practically and statistically significant. The log of total licensed capacity is included to proxy the likelihood of a place being available, and the confidence interval for this coefficient contains one, which is the value that would be expected if the vacancy rate is uniform across providers. The price variable in the regression is margin, defined as the number of dollars by which the provider's weekly preschool price exceeds the applicable maximum weekly reimbursement rate. The coefficient on margin is statistically significant and has the expected negative sign but has an extremely small magnitude. If taken at face value this coefficient would imply that a typical urban subsidized household would be willing to pay an additional $254 of weekly price to avoid an additional mile of distance. However, it is to be expected that price sensitivity will be underestimated. Price is endogenous, because providers that are higher quality in ways not observed in the data will also tend to charge higher prices. This will tend to bias the estimate in the positive direction. Furthermore the price measure used is subject to substantial measurement error and has a large proportion of imputed values, which will tend to bias the estimate towards zero. Credible estimation of choice parameters related to price would require either much more comprehensive data on provider characteristics or a causal identification strategy beyond the scope of this project.

Comparison of model and data

In this subsection, we examine the extent to which the model is able to replicate patterns of distance and provider choice in the data.

Figure 2.4: Center Proximity and Utilization

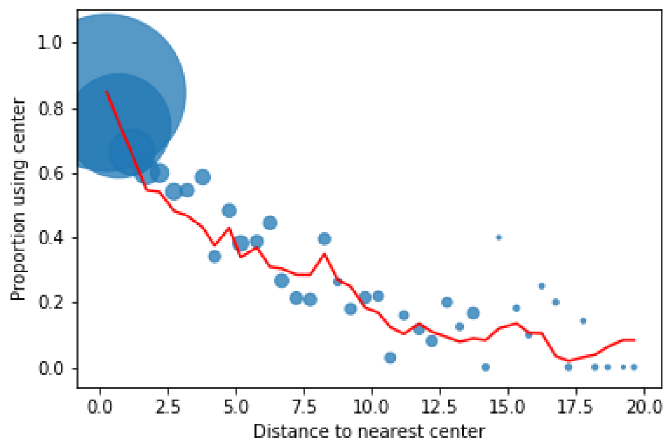
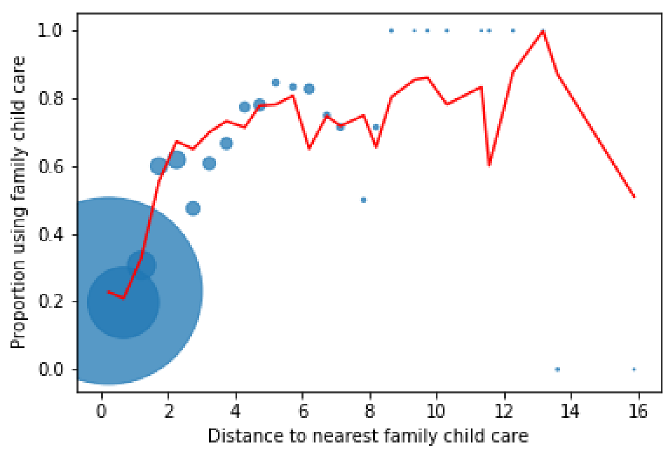


Figure 2.5: Family Child Care Proximity and Utilization



Figures 2.4 and 2.5 show the relationship between proximity to the nearest center or family provider and the probability of choosing a provider of that type. The blue circles are a binned representation of the actual consumer/provider distances in the data. The size of the circle represents the number of consumers in that bin, and the height of the circle represents the proportion of consumers choosing the focal type of provider. The red line represents predicted average rates of center and family child care utilization obtained by simulating the choice model. The predictions are not monotone in the horizontal axes because they reflect the actual locations, demographics, and mix of locally available supply for the consumers in each bin. Figure 2.4 shows that the model successfully replicates a key descriptive feature of the data, the strong relationship between use of centers and proximity to centers. The predicted rates of center use match the actual rates both for the majority of subsidized consumers who live less than two miles from the nearest center, and for the smaller group who are farther away. Figure 2.5 shows that the model also replicates the more counterintuitive relationship between family provider proximity and family provider utilization. Consumers with the lowest distances to family providers also tend to be be close to centers, and thus more likely to use centers. Consumers far away from the nearest family providers are in areas with less density and are more likely to use a family provider.

Table 2.5: Actual and Model Center Utilization

	actual % center	model % center
urban	89.4%	89.4%
rural	38.4%	38.4%

Table 2.6: Actual and Model Mean Distance

	actual mean distance	model expected distance
urban	3.51	3.52
rural	5.442	5.43

Tables 2.5 and 2.6 present actual and simulated values of the proportion using centers and the mean distance between consumer and utilized provider for urban and rural households. As should be expected given the inclusion of a rural consumer dummy variable, the model closely matches these moments for both urban and rural groups.

Table 2.7: Actual and Model Center Utilization

	actual proportion center	model proportion center	difference
Hennepin County	0.966	0.960	-0.004
Ramsey County	0.852	0.872	-0.020
Dakota County	0.970	0.905	0.0650
Anoka County	0.940	0.877	0.063
Olmsted County	0.773	0.772	0.001
St. Louis County	0.553	0.577	-0.024
Stearns County	0.759	0.743	0.016
Washington County	0.935	0.853	0.082
Crow Wing County	0.398	0.477	-0.079
Scott County	0.929	0.838	0.091

Table 2.7 shows the actual and predicted proportion using centers by county for the ten counties with the highest numbers of subsidized consumers. The model does a reasonably good job of matching county patterns in the data. The counties in the table with the largest positive differences, Washington (8.2% pts) and Scott (9.1% pts) are both outer suburban counties of the Twin Cities metropolitan area, whereas the county in the table with the largest negative differences, Crow Wing (-7.9% pts), is a micropolitan county in east central Minnesota centered on the city of Brainerd. One possible explanation is the role of employment location differences that are not part of the model: consumers in outer suburbs may be committing to jobs closer to the center of the city, giving them better access to centers than the model predicts.

Table 2.8: Actual and Model Mean Distance

	actual mean distance	model expected distance	difference
Hennepin County	3.14	3.57	-0.43
Ramsey County	3.24	3.84	-0.60
Dakota County	4.32	4.38	-0.06
Anoka County	4.19	4.46	-0.28
Olmsted County	3.88	2.90	0.98
St. Louis County	5.06	3.41	1.65
Stearns County	3.43	2.90	0.52
Washington County	4.67	4.75	-0.08
Crow Wing County	6.74	3.86	2.88
Scott County	5.24	4.74	0.50

Table 2.8 shows the actual and predicted mean distance between consumer and provider for the

same ten counties. These statistics show that there are some significant regional patterns that are not captured by the model. The model overestimates the distance for consumers in Hennepin (-0.43 mi) and Ramsey (-0.60 mi) counties, which are the urban cores of the Twin Cities metropolitan area, Minneapolis and St. Paul. Conversely the model substantially underestimates the distance for consumers in Crow Wing County (+2.88 mi), St. Louis County (+1.65 mi, Duluth), and Olmsted County (+0.98 mi, Rochester).

Estimation on Urban and Rural Subsets

To further explore the differences in distance sensitivity between urban and rural consumers, we estimated the model separately on each group of consumers. Table 2.9 shows the results from these versions of the model. It should be noted that no restriction is made on choice sets in these versions. A rural consumer remains free to access a provider in an urban location and vice versa.

Estimating on subsets of the data is equivalent to allowing *all* coefficients to interact with the subset variable, meaning that these versions are substantially less restricted than specification (2) from Table 2.4. The results show that this relaxation of constraints has little effect on estimates of the key coefficients on responsiveness to distance, which are extremely similar to those from the previous version, specification (2) in Table 2.4. The coefficient on distance for urban consumers is -0.590, compared to -.559 in the previous version. The coefficient on distance for rural consumers is -0.269, compared to -0.289 as implied by the coefficients in the previous version. The differences are well within the confidence intervals.

There are, however some suggestive differences in other estimated coefficients. For example, in the rural subsample, all of the magnitudes of the coefficients on Parent Aware are smaller than the magnitudes of corresponding coefficients estimated on the urban subsample. While the rural subsample rating coefficients are not very precisely estimated, the magnitude differences for ratings 1, 2, and 3 are nonetheless large enough to be statistically significant.

Table 2.9: Results Urban and Rural Subsamples

Variables	Urban	Rural
Family	-3.1678**	0.0882
	(0.2605)	(0.7074)
Household income (thousand $) x Family	0.0585	0.0537
	(0.052)	(0.0759)
Work earnings dummy x Family	0.1321	-0.0173
	(0.1777)	(0.6137)
Asian/Pacific x Family	3.3603**	0.0394
	(0.2442)	(1.3375)
Hispanic x Family	0.3166*	-0.0167
	(0.1915)	(0.1211)
Multiracial x Family	0.2336	0.0159
	(0.2214)	(0.3352)
Native American x Family	0.3465**	0.4562**
	(0.0957)	(0.1899)
Unreported x Family	0.4521**	0.0715
	(0.1026)	(0.3094)
White x Family	0.7951**	0.3029
	(0.071)	(0.1898)
Rating 1	0.8705**	0.0646
	(0.0668)	(0.2054)
Rating 2	0.8091**	0.0938
	(0.0949)	(0.236)
Rating 3	0.0191	0.2644
	(0.0519)	(0.1929)
Rating 4	0.9836**	0.2355
	(0.0356)	(0.1829)
Accreditation	-0.3594**	-0.0197
	(0.0315)	(0.2003)
Building Quality	-0.7624**	0.1742
	(0.0483)	(0.2034)
Chain	-0.0327	-0.0794
	(0.0308)	(0.11)
Nonprofit	-0.9583**	-0.6203**
	(0.0513)	(0.1151)
Hours / day	0.1295**	0.1243**
	(0.0042)	(0.0196)
Days / week	0.3242**	0.2549**
	(0.0201)	(0.1213)
Margin = min{Price - subsidy,0}	-0.0019**	-0.0093**
	(0.0005)	(0.0032)
Margin imputed	0.0213	-0.263**
	(0.0216)	(0.1086)
log(capacity)	0.9824**	1.3207**
	(0.0311)	(0.1211)
Distance (miles)	-0.5901**	-0.2693**
	(0.0208)	(0.0354)
λ-Center	1.2961**	1.2471**
	(0.0313)	(0.126)
λ-Family	1.4059**	1.02**
	(0.0378)	(0.1344)

Sandwich standard errors in parentheses

2.7 Market Definitions for Policy Analysis

This section presents the first application of the estimated demand model, using the predictions of the model to present a characterisation of the locally relevant region for child care policy analysis. We define a novel construct for defining a locally relevant region, which we call the p-impact radius. In order to characterize how the locally relevant region varies in different places, we calculate a specific version of this measure, the 50%-impact radius, at the population centroid of each census tract in Minnesota. The impact radii we calculate are generally higher in rural and suburban regions and lower in cities and towns. Even in urban areas, our impact areas are quite large compared to the market areas discussed in previous related studies that focused on extremely dense cities.

Definition of Impact Radius

Our construct for defining a locally-relevant region is based on the idea of adding a single hypothetical provider j_0 at a given location. What other providers would be affected? Specifically, if ΔP_{ik} is the change in the probability that consumer i uses provider k resulting from the introduction of j_0, we define

$$R_p(j_0) = \min R$$

s.t.

$$\sum_{k \in N_R j_0} \sum_{i \in I} \Delta P_{ik} \geq p \cdot \sum_{k \in J} \sum_{i \in I} \Delta P_{ik}$$

Thus $R_{50\%}(j_0)$ is the smallest radius where 50% of the impacts on other providers will be experienced by providers within the radius.

To implement the definition we make the additional hypothetical provider a center and give it characteristics that represent the average characteristics of centers in the sample.

This definition of a locally-relevant region is focused on potential substitution between providers, which is relevant for considering policy questions of crowd-out or for studying prices.

Variation in Impact Radius

To study how the locally relevant region is different in urban and rural areas, we computed the 50% impact radius at the population centroid of every census tract in Minnesota.[3]

Table 2.10: Impact Radii at Tract Centroids

	count	mean	std	min	25%	50%	75%	max
r	1338.0	6.24	3.89	0.31	3.86	5.4	7.79	35.6

Figure 2.6: Impact Radii at Tract Centroids

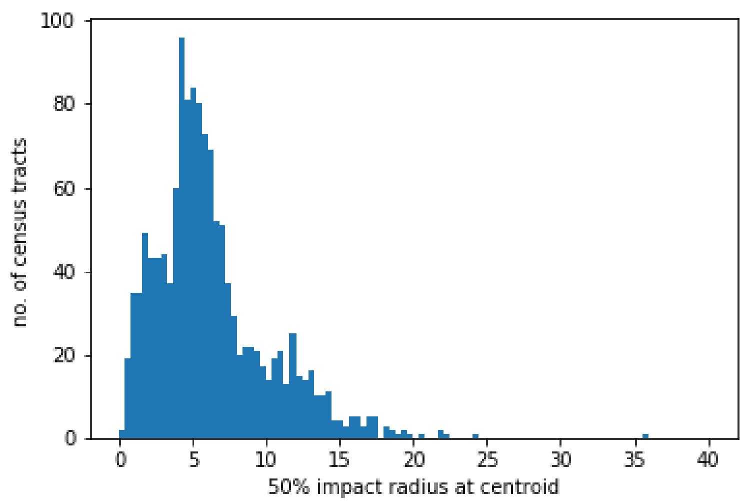
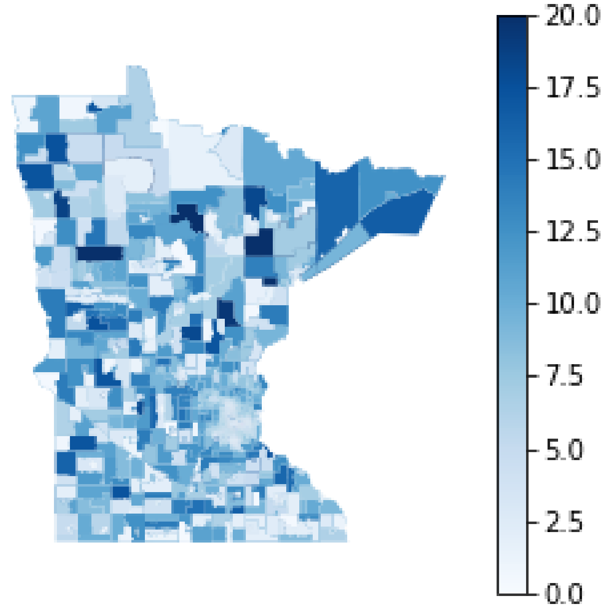


Table 2.10 shows some descriptive statistics for the radii calculated at the tract centroids, and Figure 2.6 is a histogram of the same values. These data show that most census tracts have impact radius values near the median value of 5.4 miles. However there are also many tracts with valuse less than five miles and also many with values greater than ten miles.

Figure 2.7 illustrates the geography of this variation through a chloropleth map, where each census tract is colored by the impact radius at the center. The map shows substantial variation. The area around the Twin Cities and around some of the other larger urban areas such as Rochester and

[3]The 2010 centers of population published by the Census Bureau.

Figure 2.7: 50% Impact Radius by Census Tract



St. Cloud such as show relatively low impact radii. The rest of Minnesota shows a patchwork of high and low values.

This patchwork reflects a number of factors, an important substantive descriptive feature of the calculated impact radii. The impact radii calculated for small cities in rural areas tend to be just as low as those calculated in cities. Figure 2.8 shows details of two locations: the Twin Cities metropolitan area, and Thief River Falls, a small city in Northwest Minnesota. The panels are at similar scales: each panel shows a rectangle whose dimensions are 1 degree of longitude by 1 degree of latitude. The Twin Cities Area detail map shows that impact radii tend to be lower in more dense central urban areas, and higher in suburban and outlying areas. The Thief River Falls detail map includes two small cities: Thief River Falls, in the center of the area, and Crookston, to the southeast. These small cities have impact radius values comparable to those in downtown Minneapolis. However there is a steep gradient from these cores to some much larger impact radii in some of the nearby tracts. Figure 2.9

Figure 2.8: 50% Impact Radius, Selected Locations

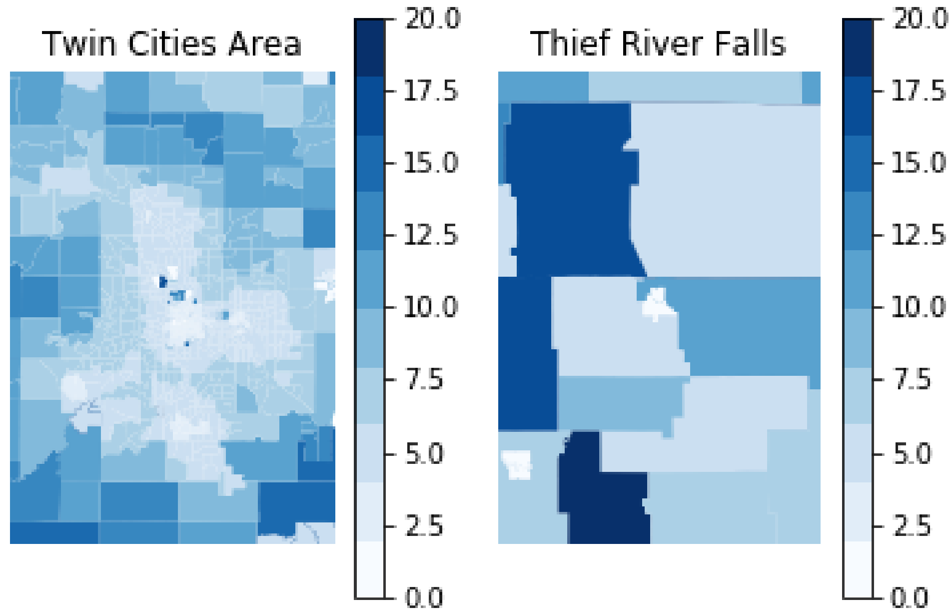
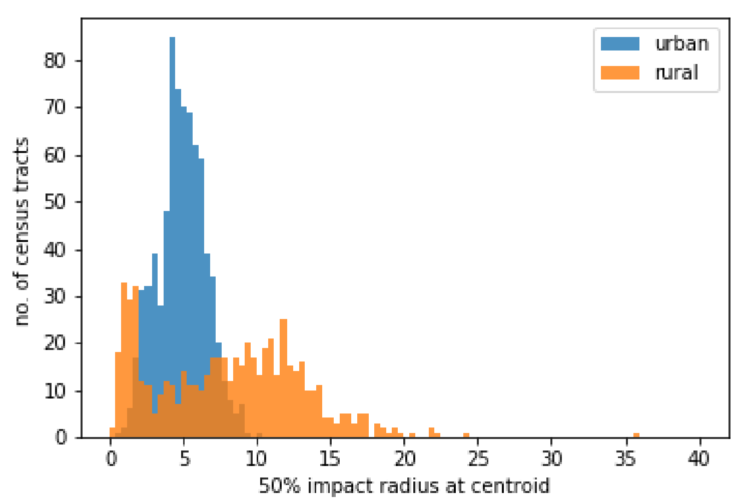
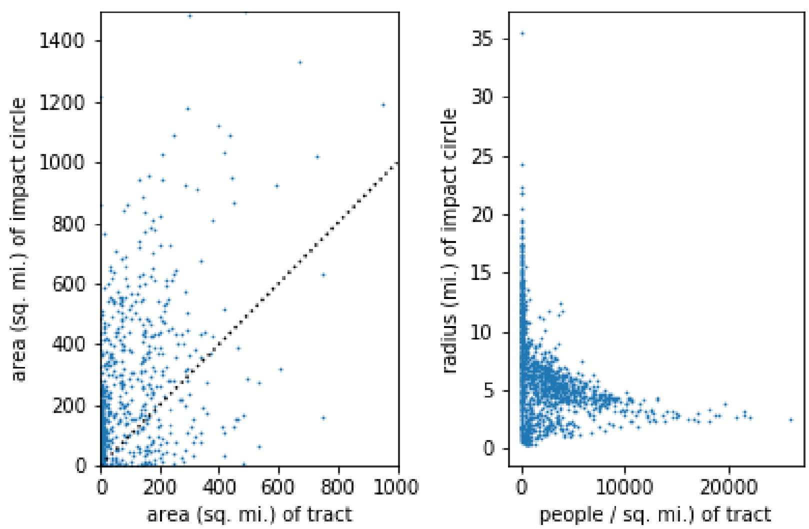


Figure 2.9: Impact Radii in Urban and Rural Tracts



shows the distribution of the impact radius measure for urban and rural census tracts respectively.[4] Rural tracts had both the highest and lowest values of the impact radius measure. Intuitively, this pattern can be explained by the fact that provider competition is contained within the catchment areas of small cities in rural areas. A new provider in Thief River Falls would compete chiefly with other providers within the small area of Thief River Falls, since there are relatively fewer options around. Thus these small cities have smaller provider impact radii than suburban areas that are more dense but less isolated.

Figure 2.10: Impact Area or Radius and Tract Geography



The geography of the tracts is not a particularly good proxy for the locally relevant areas, as measured using our impact radius definition. Figure 2.10 shows scatter plots relating the impact radius to geographic descriptors of the tracts. The first panel shows the area in square miles of the impact circle for each tract, calculated as $\pi R_{50\%}^2$, compared to the land area of the tract itself. Most of the points shown are above the dotted 45 degree line, indicating that the impact area is larger than the area of the tract.

The second panel of Figure 2.10 shows the impact radius value plotted against the tract population

[4] Sometimes tracts are partially inside and partially outside of Urbanized Areas. For this analysis, we classified a tract as urban if its centroid was in an Urbanized Area and rural otherwise.

density.[5] The most dense tracts tend to have impact radius values between 3 and 5 miles. For the 215 tracts with population density greater than 5000 people per square mile the 25th and 75th percentile are 3.19 miles and 4.63 miles.

The results of our analysis suggest that relatively large regions are appropriate for analyzing place-based policies or interventions. In rural areas and small cities, it may be most appropriate to consider the catchment area around the nearest city or town as the locally relevant region. In larger cities, the locally relevant region suggested by our analysis is a circle of several miles radius. Analysis that uses census tracts or zip codes as proxies for child care markets that are treated by a place-based intervention are likely to substantially underestimate the effects of those interventions, since most of these effects will occur outside those small geographic units.

2.8 Race Differences and Local Availability

This section presents a second application of the estimated demand model. We examine the extent to which race-based differences in the type of care used by subsidized households can be explained by differences in what is locally available. First, we compare the nested logit choice model to single-level logit models that use the same household demographics. We show that the additional information on local supply introduced by the nested logit model is relevant for predicting whether consumers use family child cares or centers, and that including this information reduces the estimated differences between White and Black consumers. Second, we present estimates of the choice model on Black, White, and Hispanic subgroups in the data, and introduce family provider race as an explanatory variable. Finally, we use the model to simulate a counterfactual that measures how much of the difference between Black, White, and Hispanic households can be explained by differences in what is locally available where consumers these groups live.

Comparison of Nested to Single-Level Logit Models

Table 2.11 presents coefficients estimated for several comparison specifications that use different methods to control for locally available supply. All four models use the same data and the same set

[5] Population density was calculated as the 2010 population divided by the land areas in square miles.

Table 2.11: Results from Comparison Models

	(1) no local controls	(2) zip counts	(3) county counts	(4) nested logit
Family (Constant)	-3.0344**	-2.6144**	-0.4793**	-2.9797**
	(0.0878)	(0.0983)	(0.1227)	(0.182)
Household income (thousand $) x Family	0.0406*	0.0384*	0.0469**	0.0409
	(0.0221)	(0.0218)	(0.0228)	(0.0503)
Work earnings dummy x Family	0.157*	0.1328	0.0944	0.0884
	(0.087)	(0.0876)	(0.0885)	(0.1765)
Asian/Pacific x Family	3.3931**	3.2279**	3.1227**	3.3838**
	(0.1394)	(0.1454)	(0.1436)	(0.1811)
Hispanic x Family	0.8111**	0.6186**	0.3875**	0.3814
	(0.1079)	(0.1103)	(0.1128)	(0.2963)
Multiracial x Family	0.7152**	0.5469**	0.4332**	0.2795**
	(0.1051)	(0.1052)	(0.1088)	(0.0745)
Native American x Family	1.5685**	1.3765**	1.1585**	0.978**
	(0.162)	(0.1763)	(0.1745)	(0.4796)
Unreported x Family	0.6682**	0.5696**	0.4712**	0.4114*
	(0.1029)	(0.1055)	(0.1115)	(0.2404)
White x Family	1.2671**	1.0823**	0.8294**	0.7742**
	(0.0701)	(0.0693)	(0.0727)	(0.1133)
Rural x Family x Nest Family	2.18**	1.8747**	0.7455**	1.0621**
	(0.0569)	(0.063)	(0.0754)	(0.2092)
Zip code centers		-0.0904**		
		(0.0078)		
Zip code family child cares		0.014**		
		(0.0012)		
County centers			0.0055**	
			(0.0006)	
County family child cares			-0.0071**	
			(0.0004)	
Distance (miles)				-0.4904**
				(0.1635)
Distance x Rural				0.2535**
				(0.1038)
λ-Center				1.1079**
				(0.3461)
λ-Family				0.92**
				(0.2504)

Robust (sandwich) standard errors in parentheses of consumer demographics as the nested logit models reported in Table 2.4. The first three columns are logit models. For consistency with the nested logit results, which used center as the omitted nest category, the outcome variable here is choice of a family provider (as opposed to child care center). The first column uses no controls for what is locally available other than the dummy variable for “rural” location. The second and third columns use the counts of child care centers and family child cares within a local region, defined as zip code and county respectively. The last column is a simplified version of the baseline nested logit provider choice model that uses location and provider type but none of the provider characteristics variables. Here, we are interested in whether controlling for differences in what is locally available reduces estimated race differences in the use of family child care. Thus the key coefficients are those on the race dummy variables. The reference category is Black, which is the largest group in the sample. The coefficients represent differences in the typical propensity of consumers in the named group to choose family providers, compared to the behavior of Black households.

Even without controlling for the characteristics of individual providers, the nested logit model (specification 4) estimates substantially smaller race coefficients compared to the logit specification that does not control for differences in local availability (specification 1). All of the race coefficients except the coefficient on the Asian/Pacific Islander are substantially reduced. This difference in estimates reflects the fact that the nested logit model effectively conditions choice of family-based care on a distance-weighted index of what is available close to the consumer. The estimates from specification (4) are very similar to those from specification (2) of Table 2.4, indicating that the additional provider characteristics used in the baseline model are less important for explaining race differences in family child care use than the information about the number and location of providers incorporated distance covariate and the structure of the model.

Specifications (2) and (3) are single-level logit models that use counts of local providers to control for differences in what is locally available. Specification (2), which uses zip-code level counts, yields estimates that are closer to the values estimated in specification (1). We conjecture that zip codes are too small compared to the locally relevant region for child care choice, and zip-level controls are too noisy a measure of what is locally available. Specification (3), which uses county-level counts, gives unexpected signs on the variables measuring local availability, suggesting that more centers are associated with a higher probability of using family providers, and vice-versa. These results suggest that neither zip code nor county level measures are sufficient to adequately control for local differences in availability.

Race Subsamples and Family Provider Race

In this subsection we further explore race-based choice patterns by estimating the model on Black, White, and Hispanic subgroups in the data, and supplementing the provider characteristics with additional variables describing the race of the family providers.

Table 2.12 shows the results from these estimating the model on sample subgroups corresponding to Black, White, and Hispanic consumers respectively.

The key coefficient of interest is that labeled Family, which represents the propensity to select a family provider compared to a center with otherwise similar characteristics. It is important for interpretation that for each subgroup, the reference category for family provider race is set up to match the race of the subgroup. Thus, for the Black subgroup, the coefficient represents the propensity to choose a Black family provider, for the White subgroup it represents the propensity to choose a White family provider, and for the Hispanic subgroup it represents the propensity to choose a Hispanic family provider. These results show that, conditional on local availability, the propensity for Black consumers to choose a Black family provider is similar to the propensity of White consumers to choose a White family provider.

Counterfactuals

How important are the modeled differences in local availability, compared to the remaining race differences that remain after controlling for them? We quantify the role of these differences by simulating a series of counterfactuals that apply the estimated behavioral parameters from the race groups to data that represents an alternative distribution of consumers and providers.

Table 3.4 summarizes the results from these counterfactuals. Each value in the table represents the proportion of consumers using centers under one of the counterfactual scenarios. The column labeled "Actual" gives the actual proportion of consumers in each of the three race groups who

Table 2.12: Estimates from Race Subsamples with FCC Race

	Black	White	Hispanic
Family	-0.4886**	-0.7502**	-0.1777
	(0.1754)	(0.2877)	(1.2302)
Household income (thousand $) x Family	0.1199**	0.0181	-0.063
	(0.0387)	(0.0513)	(0.1515)
Work earnings dummy x Family	-0.3373**	0.216*	0.558
	(0.1549)	(0.1278)	(0.4934)
Rural x Family	1.578**	0.2243*	0.4702
	(0.2641)	(0.1222)	(0.5376)
Rating 1	0.8201**	0.248	0.1054
	(0.091)	(0.1928)	(0.7745)
Rating 2	0.6609**	0.4555**	-0.1353
	(0.0476)	(0.1551)	(0.375)
Rating 3	-0.5378**	0.4402**	1.073
	(0.038)	(0.1848)	(0.725)
Rating 4	0.881**	0.5568*	0.8619**
	(0.0538)	(0.3109)	(0.1948)
Accreditation	-0.6167**	0.1685	0.2942
	(0.036)	(0.7126)	(0.6381)
Building Quality	-1.1347**	-0.1335	-0.0121
	(0.0509)	(0.1949)	(0.4686)
Chain	-0.5673**	0.3914**	0.7144**
	(0.0477)	(0.172)	(0.2779)
Nonprofit	-1.1595**	-0.548**	-0.4526
	(0.0474)	(0.1354)	(0.326)
Hours / day	0.1101**	0.1596**	0.0979**
	(0.0049)	(0.0185)	(0.0406)
Days / week	0.3532**	-0.0255	0.2122
	(0.0277)	(0.1184)	(0.2162)
Margin = min{Price - subsidy,0}	-0.0017**	-0.0065**	-0.0058
	(0.0003)	(0.0019)	(0.0062)
Margin imputed	0.2428**	-0.5289**	-0.8148
	(0.0215)	(0.0533)	(0.6945)
log(capacity)	1.0782**	1.1391**	0.5527
	(0.0488)	(0.0601)	(0.3626)
FCC race Asian/Pacific	-3.8624**	-1.2696**	-0.8543
	(0.3614)	(0.4152)	(1.1492)
FCC race Black		-0.6449**	-1.4879**
		(0.1256)	(0.6945)
FCC race Hispanic	-194.8641**	-0.9219	
	(8.557)	(0.6376)	
FCC race Multiracial/other/unknown/native	-1.4135**	0.8429**	-0.9975**
	(0.1344)	(0.1729)	(0.3657)
FCC race White	-3.4581**		-1.6446**
	(0.2874)		(0.6906)
Distance (miles)	-0.5957**	-0.4796**	-0.541**
	(0.0241)	(0.0731)	(0.158)
Distance x Rural	0.2993**	0.1833**	0.3268**
	(0.0222)	(0.0669)	(0.1036)
Lambda Center	1.2861**	1.2011**	1.2571**
	(0.0354)	(0.0713)	(0.4155)
Lambda Family	1.2597**	1.1648**	1.1108
	(0.0338)	(0.1155)	(0.7588)

Sandwich standard errors in parentheses

Table 2.13: Counterfactuals on Race and Local Availability

	Actual	C1	C2	C3
Black	93.3%	82.4%	82.7%	65.6%
White	53.8%	73.8%	74.6%	75.1%
Hispanic	75.2%	78.3%	77.1%	75.9%

use centers rather than family providers. The difference between Black and White consumers is an astonishing 39.5 percentage points. The difference between Hispanic and White consumers is 21.4 percentage points. The column labeled C1 is the baseline counterfactual, obtained by applying the coefficients estimated from each race group to the distribution of locations, household income, and work status of the entire CCAP consumers population. C1 asks: how much of the difference in use of center-based care would remain if there were not differences in the location, income and work status variables? Under the counterfactual, the difference between Black and White consumers is reduced to 8.6 percentage points (22% of its prior value). The difference between Hispanic and White consumers is reduced to 4.5 percentage points (21% of its prior value). We thus conclude that the differences in location, income and work status between CCAP consumers of different races are much more important for explaining differences in the use of center-based care than the differences in choice patterns that remain after controlling for local availability.

The final two columns of Table 2.13 assess the quantitative importance of the race distribution of family providers. Each of these columns represents the results of simulating a counterfactual where, in addition to equalizing in location, income and work status, we also randomly reassigned family provider race according to a reference distribution. In C2 we reassigned family provider races using draws from the race proportions of the Minnesota population as reported in the American Community Survey's 1-Year Estimates for 2015. In C3 we reassigned family provider races using the proportions from the CCAP consumer population. We independently drew samples and simulated the counterfactual 10 times, reporting the average. The counterfactual where the race distribution of family providers is reassigned to the Minnesota population distribution shows only minor differences from the baseline counterfactual. The gap between Black and White consumers reduces slightly to 8.1% and the gap between Hispanic and White consumers reduces to 2.5%. Under the more extreme counterfactual C3, where the race distribution of family providers is set to the race distribution of CCAP consumers, meaning that 44.9% of the family providers are Black, 33.3% are White, the pattern in predicted use of center-based care flips around so that Black consumers are 9.5 percentage points less likely to use centers than White consumers.

2.9 Conclusion

In this paper, we addressed three main research questions. First, how important is distance in determining child care choice? Our coefficient estimates suggest urban households respond more to distance than rural households. For an urban consumer, an extra mile of distance is typically associated with about a 55.9% reduction in the probability that a provider is chosen. For a rural consumer, an extra mile of distance is typically associated with a 28.9% reduction in the probability that a provider is chosen. Second, what is the relevant local region for assessing child care policy interventions? We use our estimated model to calculate a measure of this locally relevant region. We find that in most parts of Minnesota, a six mile radius would be required to capture 50% of the impact of introducing a new center provider. The impact radius is smaller in some urban areas and in rural cities and towns. Third, to what extent does variation in local supply explain different rates of use of family child care among consumers of different races? We show that our model produces smaller estimates of race differences in the propensity to use family care than comparison models that do not control for local availability, or that control for differences in local availability using counts of providers by county or zip code. We estimate counterfactuals to quantify the importance of remaining differences in choice patterns, compared to the importance of observed location and demographics. We find that location, work status, and income explain four-fifths of the gap between White consumers and Black or Hispanic consumers in the use of family providers. These findings contribute to a better understanding of the importance of distance and local availability in determining parents' choice of family or center child care providers.

Chapter 3

Fewer Options or More Substitution?

3.1 Introduction

Family child care has been declining in the United States. Family providers are formal, regulated providers who offer child care services within the provider's home.[1] Family providers disproportionately serve consumers in rural areas and consumers who require a non-standard care schedule. The decline in family providers has attracted increasing concern from child care advocates, researchers, and government administrators. In particular there is a concern that declining availability of family child care may compromise the ability of the low-income families targeted by child care subsidy programs to access care. This paper aims to describe the impact of decreased family child care availability on consumers subsidized by Minnesota's Child Care Assistance Program.

In Minnesota, there has been a profound composition shift in the way that subsidized care is delivered. In July 2010, approximately 12,000 children received CCAP payments to support attendance in licensed family child cares. By June 2018, this number had declined to less than 5000. It is natural to associate this decline with the contemporaneous decline in the availability of family child care providers. However this association is complicated by the presence of several other plausibly important trends. Over the same period, the availability of child care centers has been increasing. If centers and family child cares are close substitutes then this could in principle explain why fewer

[1] This chapter is based on coauthored work with Elizabeth Davis

subsidized consumers are using family providers. At the same time, the demographics of households participating in CCAP have shifted. Black consumers are generally found to be less likely to use family child care, and the proportion of subsidized households who are Black has been rising. Similarly, the average income of subsidized households has been increasing, and it is possible that higher-income households are less inclined to use family child care, which is generally a lower-priced option. Finally consumers' beliefs and priorities for early childhood care may be shifting. Consumers more attuned to school readiness may place a higher value on center-based care because it is perceived as a more school-like option.

Our strategy for attending to these competing explanations is to examine child care choices conditional on consumer demographics and on the relative local availability of family and center-based care. Using administrative data from the CCAP program, we estimate a nested logit demand model where each consumer's choice of family or center provider depends on the number of family and center providers, the distance to each individual provider, and consumer demographics that modify the relative attractiveness of family and center providers. Assuming that any systematic shifts in demand are statewide, this approach allows those demand shifts to be disentangled from changes due to the composition of supply. Stated simply, to the extent that the declining use of family providers tracks declining availability of family providers, we treat the decline as plausibly driven by supply. To the extent that the propensity to use family providers declines uniformly, even in areas where the composition of supply doesn't change, we treat the decline as driven by demand.

Our results suggest that declining availability of family child care providers is indeed an important factor in explaining trends in the provider types used by CCAP consumers. We simulate the demand model under a sequence of counterfactuals designed to quantify how much family provider availability, center availability, and consumer demographics and location contribute to the overall trend. We find that the two most import factors for explaining this trend are family provider availability and changes in the location and demographics of CCAP consumers. Changes in conditional choice behavior, which might represent preference changes or changes in prices or other characteristics of providers, are a much less important factor. Surprisingly, increased availability of centers has only a negligible impact on the statewide trend.

3.2 Literature Review

This study is relevant to three literatures: First, literature on child care choice and the reasons why consumers choose center or family care. Second, the recent and growing policy discussion on the causes and consequences of decreased availability of family child care providers. Third, the small but important literature on how variation in local supply shapes child care consumption choices.

Center versus Family Child Care

A long-standing literature compares center and family child care and investigates the reasons why families may choose one or the other. This literature generally associates centers with educational or developmental quality. Kim and Fram (2009) use a latent class analysis to characterize households seeking child care, identifying a "learning and quality"-oriented group who are more likely to use centers, and a "practicality"-oriented group who are more likely to use family providers. Davis and Connelly (2005) find that employed mothers rarely use family child care, suggesting that family child care serves primarily to facilitate mothers' participation in the labor force. Analyzing a large observational study of providers of different types Bassok et al. (2016), document that centers provide more frequent reading and math activities, fewer hours of TV, and are more likely to be staffed by caregivers with an early childhood degree. However, Bassok et al. (2013) show that educational attainment and earnings increased substantially among family providers over the 1990-2010 period, decreasing the quality gap between centers and family providers. Rather than examining the differences between family child care and centers, this paper investigates substitutability between these types of care, asking whether decline in the use of family care in the subsidized population in Minnesota can be explained by increased availability of centers.

Decline of Family Child Care Availability

The nationwide decline in the number of family child care providers has attracted the attention of commentators, policy-makers, and researchers. Using data on the child care workforce from the Current Population Survey, Bassok et al. (2013) document substantial declines in family child care workers and increases in child care center workers over the 1990-2010. A report published by the

National Center for Early Childhood Quality Assurance documents the nationwide decline in the number of family child cares, arguing that decreased availability of family providers may have negative impacts on rural households, minority groups, and families with nonstandard work schedules. [NCECQA, 2019] A report published by the Minnesota Center for Rural Policy and Development characterizes the declining availability of family child care in rural Minnesota as a “crisis” that has been “quietly brewing”, pointing out that the number of family child cares in Minnesota declined 27% from 2006-2015 and pointing out that growth in the number of child care centers over the same period was concentrated in the Twin Cities metropolitan area.

Effect of Local Supply on Child Care Consumption

Finally we add to the small but important literature on how child care consumption patterns are shaped by local availability. Gordon and Chase-Lansdale (2001) analyzed census data sources on the availability of centers and family providers in different parts of the United States, showing that communities with more center workers for each child have higher rates of center use, and communities with more family providers for each child have higher rates of family provider use, results that have been confirmed in a more recent study by Coley et al. (2014). Davis and Connelly (2005) incorporate county-level measures of local availability into a model of child care choice but do not find any effect of local availability on type of care chosen.

3.3 Data

Data Sources

We draw on two sources of administrative data. For information on subsidized consumers and their provider choices, we use administrative data from the CCAP program. This data, provided under a data sharing agreement with the Minnesota Department of Human Services (DHS) consists of deidentified monthly records of subsidy payments at the child and child-provider level linked to child and family characteristics including race and household income. For information on providers, we use administrative data from DHS’s licensing of child care providers. This data provides us with monthly records of which providers are active, their type, and their licensed capacity. We focus on five focal months extracted from the data, spread over a nine-year period: September 2010, September 2012, September 2014, September 2016, and September 2018.

Consumers

For each consumer, we designated the provider that received the highest subsidy payment in the focal month as that consumer's primary provider. We limit our consumer sample to CCAP recipients who used a Minnesota licensed child care provider as their primary provider. This excludes consumers who used providers licensed by a state other than Minnesota, tribally licensed providers, license exempt providers (such as certain programs provided by schools and community centers) and legal non-licensed providers (providers taking care of only related children, or of children from a single family).

Table 3.1: Descriptive Statistics on Consumers

	count	mean	std	min	25%	50%	75%	max
Family Income (thousand $ mo.)	71951	1.994	1.262	0.0	1.083	1.88	2.727	13.07
Work	71951	0.843	0.364	0.0	1.000	1.00	1.000	1.00
Asian/Pacific	71951	0.023	0.149	0.0	0.000	0.00	0.000	1.00
Black	71951	0.403	0.491	0.0	0.000	0.00	1.000	1.00
Hispanic	71951	0.066	0.249	0.0	0.000	0.00	0.000	1.00
Multi/Unreported	71951	0.122	0.327	0.0	0.000	0.00	0.000	1.00
Native American	71951	0.014	0.119	0.0	0.000	0.00	0.000	1.00
White	71951	0.372	0.483	0.0	0.000	0.00	1.000	1.00
Using Family Provider	71951	0.280	0.449	0.0	0.000	0.00	1.000	1.00

Table 3.1 shows some descriptive statistics on consumer characteristics used in the analysis. Evaluated over all of the observations, the proportion of consumers using a family provider as their primary provider is 28.0%. Figure 3.1 shows how this proportion has declined over the sample period. In September 2010, the proportion of consumers using a family provider as their primary provider is 40.5%. In September 2018, that proportion is only 16.7%

Family income and the work indicator are based on records used to determine CCAP eligibility. In a small proportion of observations, where income was missing in the data, we imputed a monthly income value by linear interpolation between non-missing values in prior and subsequent months. Consumers for whom income could not be interpolated because it was missing in all months were

Figure 3.1: Changes in Use of Family Providers

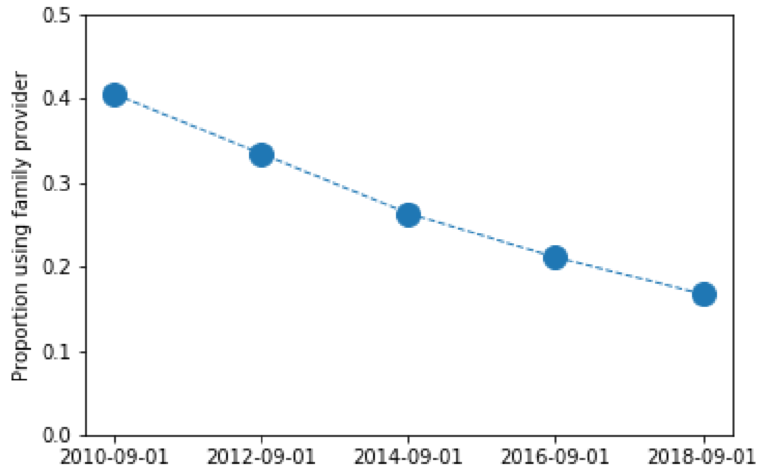
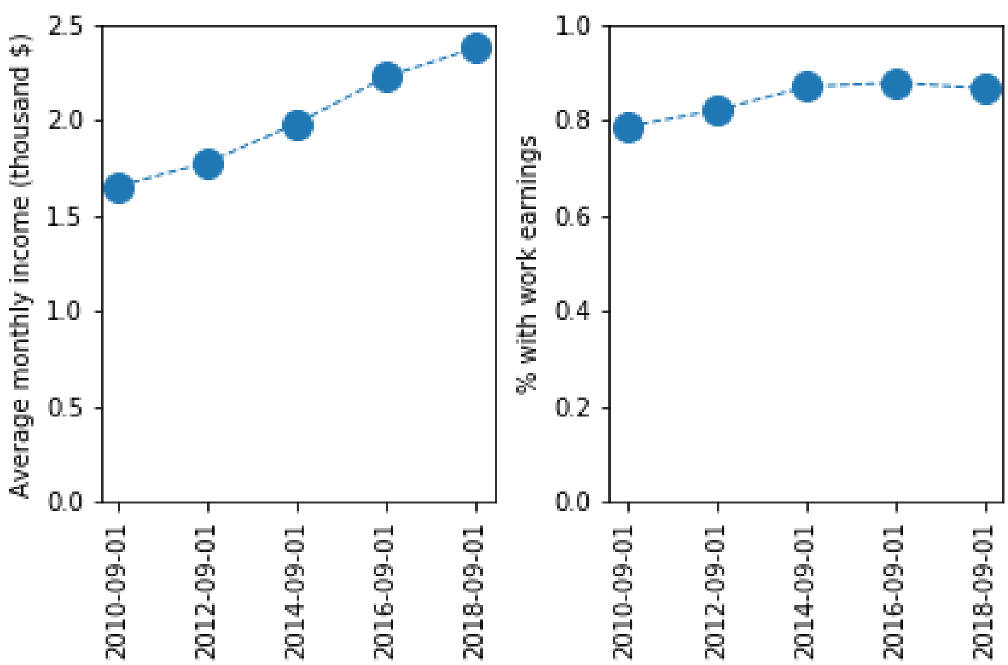


Figure 3.2: Income and Work Status of CCAP Consumers



dropped from the sample. Figure 3.2 shows how the average household income of CCAP consumers, and the proportion of CCAP consumers with a parent working, have changed over the sample period. The work indicator is based on an indicator in the data for whether any of the household's income was earnings from work. There has been a marked increase in the household income of CCAP consumers. In September 2010 the average monthly household income was \$1,665. In September 2018 the average was \$2,379. This implies an annual growth rate of about 5.3%. This growth is slightly faster than nominal median household income in Minnesota, which grew by about 4.7% annually between January 2010 and January 2018. There has also been an increase in the proportion of CCAP consumers with income from work, from 78.7% in September 2010 to 86.7% in September 2018.

Figure 3.3: Race of CCAP consumers

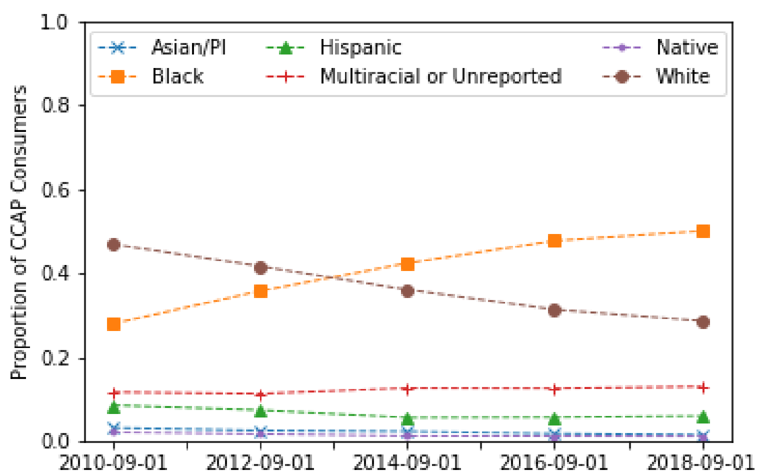
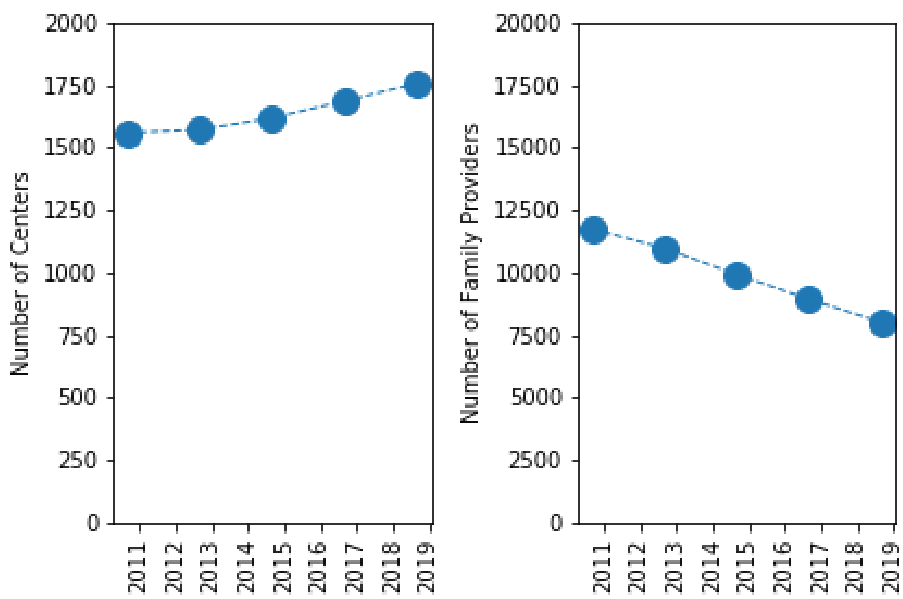


Figure 3.3 shows the proportion of CCAP consumers in different race groups over time. We used derived race categories that combine information on race and Hispanic ethnicity, so that White means non-Hispanic White, and Hispanic means Hispanic of any race. Throughout the sample period, the largest race groups among consumers have been White and Black. However the proportions of White and Black consumers have changed substantially over time. In September 2010, 46.8% of CCAP-subsidized consumers were White and 27.9% were Black. In September 2018, the proportion of

White consumers had decreased to 28.6% and the proportion of Black consumers had increased to 50.0%.

Providers

Figure 3.4: Number of Centers and Family Providers



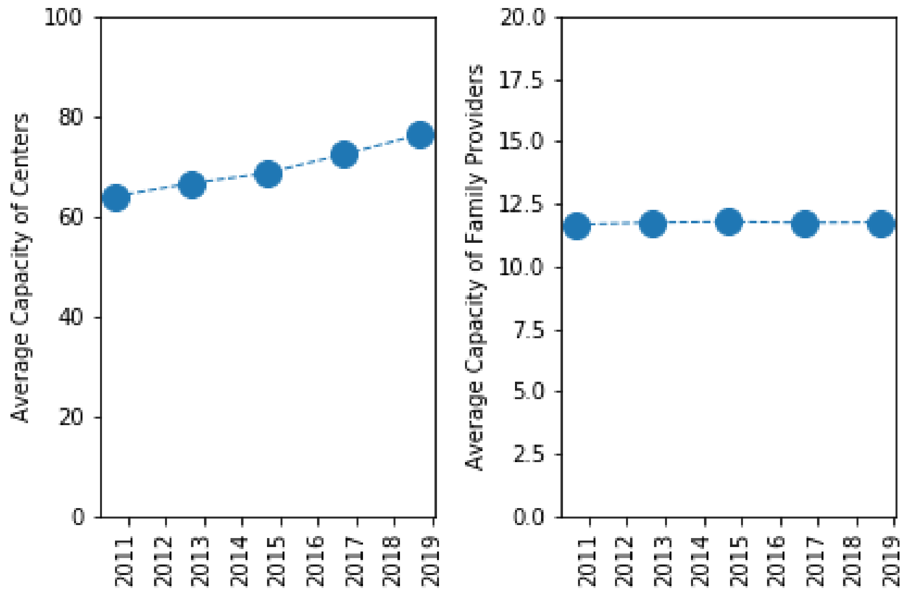
Our provider sample included all Minnesota licensed family providers and child care centers that existed during the focal months.

Figure 3.4 shows the number of centers and family providers over the sample period. The number of centers increased from 1,559 in September 2010 to 1,756 in September 2018. This corresponds to an average increase of 1.6% per year. Over the same period the number of family providers fell from 11,755 to 7,993, an average decrease of 4.0% per year.

Table 3.2: Descriptive Statistics on Provider Capacity

	count	mean	std	min	25%	50%	75%	max
Centers	8190.0	69.75	50.53	4.0	30.0	59.0	97.0	579.0
Family	49556.0	11.74	1.55	1.0	10.0	12.0	12.0	14.0
All	57746.0	19.97	27.82	1.0	10.0	12.0	14.0	579.0

Figure 3.5: Average Provider Capacity



In order to account for the fact that the size of providers varies, and may be different in different regions or different over time, we used total licensed capacity as a measure of provider size. Table 3.2 shows some descriptive statistics on the capacity of providers in the sample, and Figure 3.5 shows changes over time in the average capacity of centers and family providers. The average capacity of family providers is relatively stable at just under 12. The average capacity of centers is increasing over time, from 64.0 in September 2010 to 76.3 in September 2018, an average increase of 2.4% per year.

Locations and Distances

We obtained consumer and provider locations by geocoding addresses in the administrative data using ArcGIS software. We deidentified the consumer locations by associating each consumer with the nearest point on a 1/4 mile grid.

For each focal month, we calculated distances between every consumer location and every provider location. We used straight-line distances calculated using the GeoPy Python package. This package uses an ellipsoidal geodesic model of the earth that is more accurate than great-circle distances.

3.4 Methods

Overview

Our object is to determine how much of the change in use of family providers can be explained by changes over time in what is locally available. Our strategy is in two steps. First, we estimate a cross-sectional choice model that models the provider chosen as a function of the providers available. Second, we use the estimated choice model to compute counterfactuals.

Choice Model

Specifics

We use a nested logit model where the probability that a particular consumer chooses a particular provider is a function of the consumer's demographics and the characteristics of all available providers.

The choice probability function for nested logit is

$$P_{ij}(\theta) = e^{V_{ij}(\theta)/\lambda_n} \frac{\left(\sum_{k \in J_n} e^{V_{ik}(\theta)/\lambda_n}\right)^{\lambda_n - 1}}{\sum_{m \in M} \left(\sum_{k \in J_m} e^{V_{ik}(\theta)/\lambda_m}\right)^{\lambda_m}}$$

The consumer, indexed by i, may choose any provider, indexed by j, from the choice set J. The choice set is partitioned into nests $\{J_m\}$, that represent categories of providers that we assume may be relevant for the determination of consumers' substitution patterns. Here, those nests are family providers and center providers.

The nested logit model can be interpreted as a random utility model where the utility of consumer i choosing provider j nest J_n has the form

$$U_{ij} = V_{ij} + \varepsilon_{ijn}.$$

The consumer's choice of provider j is interpreted to mean that after the realization of ε_{ijn}, j is the provider that gave highest utility. Assuming that ε_{ijn} has a generalized extreme value distribution where the error terms may be correlated for providers that are in the same next, but are independent for providers in different nests, the choice probability function above gives the probability that $U_{ij} \geq U_{ik} \forall k \in J$.

Under the random utility interpretation, the λ_n terms can be interpreted as capturing the relative importance of the provider factor and the nest factor in ε_{ijn}. When $\lambda_n = 1$ then there is no correlation between the error terms for providers in nest n. If $\lambda_n = 1\,\forall n$ then independence of irrelevant alternatives holds and the choice probability formula becomes multinomial logit. As $\lambda_n \to 0$ the demand for nest n becomes perfectly inelastic, as the number and characteristics of providers available in nest n becomes irrelevant to whether or not a consumer chooses a provider in that nest.

We let $\theta = (\gamma, \delta, \phi)$ and assume a linear regression equation for $V_{ij}(\theta)$:

$$V_{ij}(\theta) = C_j\gamma_n + D_i\delta_n + F_{ij}\phi.$$

C_j are provider characteristics such as capacity, D_i are consumer demographics, and F_{ij} are interactional characteristics such as the distance between the consumer and the provider.

We estimate the choice model at four evenly spaced cross-sections: 2010, 2012, 2014, 2016, and 2018. For each of these years, we use the choice data from September. We estimate the model separately for each data cross-section.

We define the choice set to include all Minnesota-licensed center and family providers that exist during the focal month, plus other center and family providers who are used by at least one subsidized consumer. This specification is driven by data availability, and is likely to be both over-inclusive and under-inclusive. Some Minnesota licensed providers may not be available to subsidized consumers. Not all providers accept child care assistance. Furthermore, at the time that a particular subsidized consumer is choosing a provider, not all providers will have vacancies. On the other hand, there may be providers that are not Minnesota-licensed, such as license-exempt providers and providers that are licensed by tribes or neighboring states that would be willing to accept CCAP-subsidized consumers but do not happen to be chosen and thus do not appear in the data.

Estimation

At each time period we estimate the model on a cross-section of the data using maximum likelihood.

Purpose of the Choice Model

We estimate a model of child care choice so that we can make predictions about the proportion of subsidized consumers who would use family providers under various counterfactuals. The model specifies the relationship between the set of locally available providers and the choice of provider type.

Even though the main outcome of interest is the choice of family providers as a group, we model the choice of individual provider. The reason for modeling individual provider choice as well as modality choice is to incorporate variation in what is locally available into the model at a structural level. Defining $S_{in} = \sum_{j \in J_n} e^{V_{ij}}$, the marginal probability of choosing a provider in the family nest n_f is given by

$$M_{if}(\theta) = \frac{S_{in_f}^{\lambda_{n_f}}}{S_{in_c}^{\lambda_{n_c}} + S_{in_f}^{\lambda_{n_f}}}.$$

The sum terms S_{in_f}, S_{in_c} can be interpreted as an index of the local availability to consumer i of family providers and center providers respectively. These terms depend on the number and desirable characteristics of the available providers, with each provider weighted by its proximity to the consumer. An alternative would be to dispense with the provider level of the choice model and instead include a reduced-form index of the local availability, such as the count of centers and family providers in the consumer's county. This would require us to make assumptions about which providers are locally relevant to which consumer and in what degree of importance. Under our approach the degree of relevance each provider has to each consumer is a function of relational characteristics F_{ij} such as distance, and the parameters that govern this relationship are estimated from the data.

Limitations

Our model abstracts away from a realistic child care choice process in several significant ways.

First, we assume that each every consumer considers every provider. This is unrealistic for three reasons. First, consumers are only actually able to select providers that have a vacancy at the time the consumer is searching for child care. Second, child care search is costly. Low income working parents may have little time and limited ability to vet potential providers. Indeed, research on child care search suggests that many subsidized consumers only consider one or two potential options before choosing a providers (Forry et al., 2014).

Second, we ignore the dynamic aspects of child care choice. The costs of child care search, combined with the benefits of a stable consumer-provider relationship, make provider choice a long-term decision. Consumers choosing a provider can thus be expected to consider not only whether a provider meets their current needs, but also whether the provider will be able to meet their needs in the future.

Third, we do not model the relationship between provider choice and the parents' employment.

Counterfactuals

The counterfactuals we compute are designed to estimate the relative contribution of changes in local availability to providers of different types, consumer demographics, and "preferences".

$$Q_n(I, J, \theta) = \sum_{i \in I} \sum_{j \in J_n} P_{ij}(\theta)$$

and let I^t and J^t be the consumer and provider data corresponding to time t, and $\hat{\theta}^t$ be the model parameters estimated from time t.

In order to estimate the contribution of changes in local availability, we compute $Q_n(I^t, J^{2010}, \hat{\theta}^t)$ for each focal time period. That is, we simulate the model having replaced the choice set with that of 2010, but leaving everything else the same. To further decompose the contribution of changes in family provider and center availability, we estimate additional counterfactuals where only the family providers or only the center providers are replaced in the choice set. For example to estimate the contribution of center availability we compute $Q_n(I^t, J^t_{n_c} \cup J^{2010}_{n_c} \cup \{0\}, \hat{\theta}^t)$, fixing the family providers at 2010 levels.

In a similar way, we estimate the contribution of demographic changes by computing

$$Q_n(I^{2020}, J^t, \hat{\theta}^t)$$

for each focal time period. That is, we estimate the choice model using the estimated parameters and choice set from each period, but replacing the consumer population with that of 2010.

We estimate the contribution of "preferences" by using the choice set and consumer population belonging to each period but substituting the estimated parameters from 2010, $Q_n(I^t, J^t, \hat{\theta}^{2010})$.

3.5 Results

Results from Choice Model Estimation

Table 3.3: Parameter Estimates

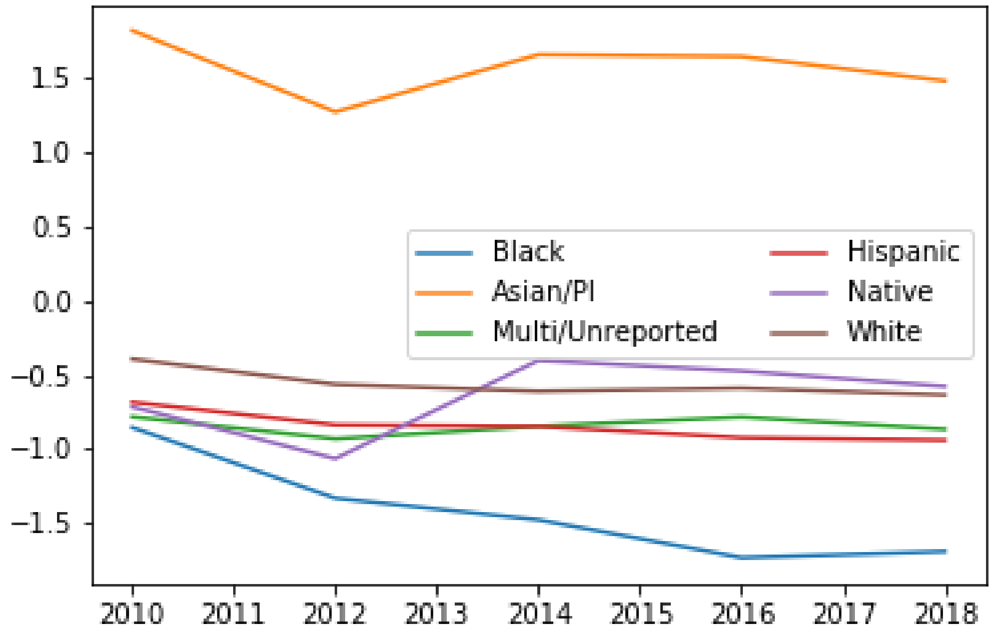
Variables	2010	2012	2014	2016	2018
Family	-0.8536**	-1.3303**	-1.4764**	-1.7301	-1.6899**
	(0.2661)	(0.0124)	(0.4069)	(1.4953)	(0.0477)
Household Income (1000 $ monthly) x Family	0.0053	0.0415	0.0268	0.0225	-0.004
	(0.0406)	(0.0307)	(0.0597)	(0.6011)	(0.0337)
Work x Family	0.3861**	0.3603**	0.3639**	0.2135	0.2008
	(0.1374)	(0.0679)	(0.1585)	(1.3642)	(0.1419)
Asian/PI x Family	2.6724**	2.6021**	3.1339**	3.3752**	3.1712**
	(0.1787)	(0.0247)	(0.707)	(0.4107)	(0.0916)
Hispanic x Family	0.0722	0.4001**	0.6309**	0.9471*	0.8245**
	(0.1253)	(0.0159)	(0.3012)	(0.4954)	(0.1297)
Multiracial/Unreported x Family	0.1706	0.4942**	0.6299**	0.8077**	0.7506**
	(0.1107)	(0.0583)	(0.1599)	(0.3803)	(0.0965)
Native American x Family	0.1393	0.2668**	1.0747	1.2568	1.1134**
	(0.3363)	(0.0546)	(0.7467)	(1.0823)	(0.0396)
White x Family	0.4612**	0.7672**	0.867**	1.1416*	1.0552**
	(0.1089)	(0.0556)	(0.2367)	(0.6132)	(0.0823)
log(capacity)	1.5764**	1.6386**	1.5841**	1.5381	1.3864**
	(0.1004)	(0.0136)	(0.1328)	(1.012)	(0.0357)
Distance (miles)	-0.4574**	-0.5064**	-0.5133**	-0.4458	-0.4268**
	(0.0319)	(0.0172)	(0.032)	(0.283)	(0.0125)
λ-Center	1.1674**	1.2593**	1.341**	1.2823	1.2092**
	(0.0704)	(0.03)	(0.0768)	(0.8417)	(0.0337)
λ-Family	1.2248**	1.3233**	1.3411**	1.3244	1.2327**
	(0.0942)	(0.0124)	(0.1199)	(1.2363)	(0.0326)

Robust (sandwich) standard errors in parentheses

Table 3.3 shows estimated demographic coefficients from each cross-sectional estimate of the choice model. These coefficients summarize the relationship between child and household demographics parameters and the propensity to choose care of each type.

The first group of estimated coefficients relates the consumers' demographics to the propensity to choose family providers rather than centers. There is a positive relationship between the work earnings dummy and use of family providers, which is consistent with prior research relating the use of family child care to employment.

Figure 3.6: Changes over time in Race x Family Coefficients



The reference group for the race variables is Black consumers, who are the largest race group in most of the sample years. The constant term labeled Family represents the propensity of Black consumers to choose family provider care. The coefficients on other groups represent the differences between consumers in these groups and Black consumers. Changes over time in the propensity of consumers of different races to use family providers are illustrated in Figure 3.6. In all years, Black consumers have the lowest propensity to use family providers and Asian or Pacific Islander consumers have the highest propensity to use family providers. The behavior of most of the race groups is quite stable over time with two market exceptions. First, the Native American group has a large increase in propensity to use family care between 2012 and 2014. Second, the Black group has large decreases in propensity to use family care between 2010 and 2016.

The coefficient on log capacity is significantly greater than one in four out of five sample years. Since the probability of choosing any particular provider is quite small, the point estimate for 2014, 1.58, can be approximately interpreted as an elasticity, such that a 1% increase in licensed capacity is associated with a 1.58% increase in that provider's market share. This coefficient having an estimated value greater than one suggests that capacity is correlated with desirable unobserved characteristics of the provider such as schedule flexibility.

Counterfactuals

Figure 3.7: Results from Counterfactuals

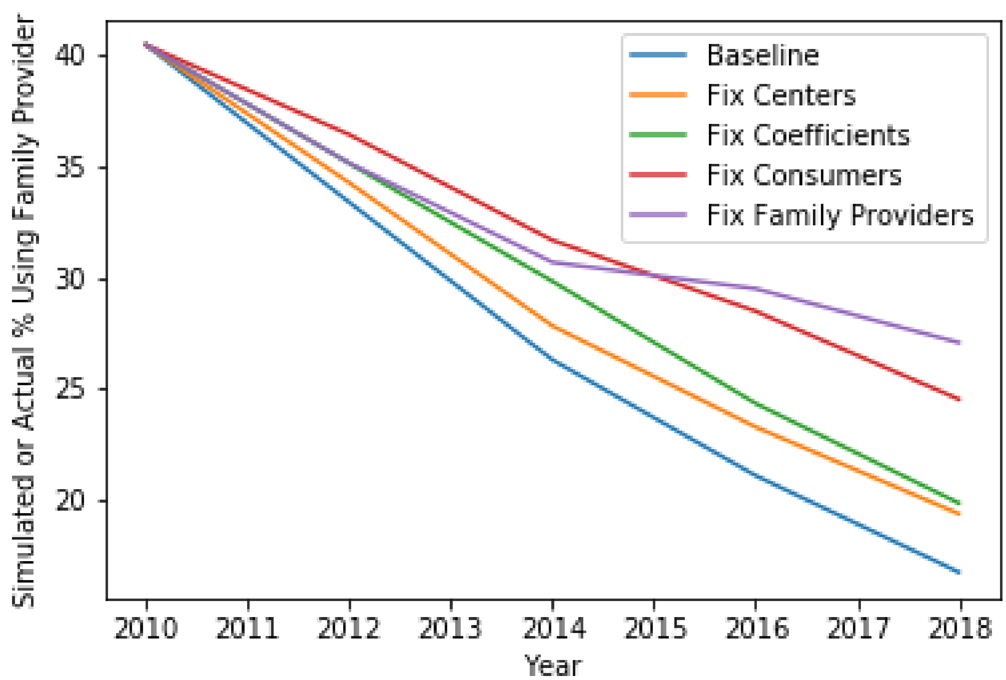


Table 3.4: Results from Counterfactuals

	Baseline	Fix Centers	Fix Coefficients	Fix Consumers	Fix Family Providers
2010	40.47%	40.47%	40.47%	40.47%	40.47%
2012	33.39%	34.27%	35.14%	36.44%	35.16%
2014	26.30%	27.83%	29.83%	31.68%	30.69%
2016	21.09%	23.28%	24.34%	28.49%	29.50%
2018	16.74%	19.37%	19.85%	24.53%	27.08%

Figure 3.7 shows the results from simulating counterfactuals and Table 3.4 presents the same information in table form. Each counterfactual is defined by fixing one aspect of the data at its 2010 value, while allowing other aspects to vary. The difference between the counterfactual path and the baseline path is a measure of how important the factor that has been fixed is to the baseline decline in use of family child care.

The counterfactual labeled “fix family providers” represents simulations of a market where the family providers sector is frozen at its 2010 state. This counterfactual answers the question, “what proportion of CCAP consumers would use family providers if the family providers of 2010 were available?”. Under this scenario, the use of family providers declines to 27.08% of the CCAP consumers population rather than 16.74% of the CCAP consumers in the baseline. Thus approximately 44% of the 2010-2018 decline in the use of family providers can be attributed to decreasing availability of family providers.

The counterfactual labeled “fix consumers” represents simulations of a market where the consumer locations and demographics (race, income and work status) from 2010 are used but all other aspects of the market are allowed to vary. The results show that demographic and location shifts in the CCAP population are almost as important as declining availability of family providers, and more important over the 2010-2014 period. Over the entire 2010-2018 period approximately 33% of the decline in the use of family providers can be attributed to shifts in the location and demographics of consumers.

The remaining counterfactuals show that increased availability of centers and statewide preference shifts are much less important. Changes in center availability contribute to only about 11% of the decline in the use of family providers. The counterfactual labeled “Fix Coefficients” uses the estimated coefficients from 2010 to estimate choices using the market features of the other focal years. This can be interpreted as capturing statewide preference shifts together well as shifts in the average characteristics of unobserved provider attributes, including changes in provider prices. However this factor can only account for about 13% of the decline. The results of our analysis thus suggest that changing type of care preferences and changing prices are much less important than declining numbers of family providers in explaining the declining use of family care in this population.

Conclusion

The central question of this paper is the extent to which the decline in use of family child care by subsidized consumers in Minnesota can be attributed to decline in the availability of family child care providers. We estimate a nested logit demand model on five focal months of choice data in two year increments from September 2010 to September 2018, and used these estimates to simulate counterfactuals to evaluate how important family provider availability is compared to a number of competing explanations. Our results show that the decline in family child care use in this population cannot be well explained by increased availability of center or by statewide shifts in preferences. Declining availability of family providers and demographic and location shifts in the population of subsidized consumers are both important contributing factors. We estimate that declining availability of family providers accounts for about 44% of the decline in the rate of family provider use, and changing locations and demographics accounts for about 33%.

Bibliography

D. Bassok, M. Fitzpatrick, S. Loeb, and A. S. Paglayan. The early childhood care and education workforce in the united states: Understanding changes from 1990 through 2010. *Education Finance and Policy*, 8(4):581–601, 2013.

D. Bassok, M. Fitzpatrick, E. Greenberg, and S. Loeb. Within-and between-sector quality differences in early childhood education and care. *Child Development*, 87(5):1627–1645, 2016.

S. Berry, J. Levinsohn, and A. Pakes. Automobile prices in market equilibrium. *Econometrica*, 63 (4):841–890, 1995.

S. T. Berry. Estimating discrete-choice models of product differentiation. *The RAND Journal of Economics*, pages 242–262, 1994.

D. M. Blau and A. P. Hagy. The demand for quality in child care. *Journal of Political Economy*, 106(1):104–146, 1998.

B. Bollinger, P. Leslie, and A. Sorensen. Calorie posting in chain restaurants. *American Economic Journal: Economic Policy*, 3(1):91–128, 2011.

K. L. Brewster and I. Padavic. No more kin care? change in black mothers' reliance on relatives for child care, 1977-94. *Gender & Society*, 16(4):546–563, 2002.

N. Burstein and J. I. Layzer. National study of child care for low-income families: Patterns of child care use among low-income families. *MA: Abt Associates*, 2007.

R. Chase and E. Shelton. *Child Care Use in Minnesota: Report of the 1999 Statewide Household Child Care Survey*. Minnesota Department of Children, Families & Learning, 2001.

A. Chaudry. *Putting Children First: How Low-wage Working Mothers Manage Child Care*. Russell Sage Foundation, 2004.

A. Chaudry, J. M. Pedroza, H. Sandstrom, A. Danzinger, M. Grosz, M. Scott, and S. Ting. Child care choices of low-income working families. *Urban Institute*, 2011.

A. Chaudry, T. Morrissey, C. Weiland, and H. Yoshikawa. *Cradle to Kindergarten: A New Plan to Combat Inequality*. Russell Sage Foundation, 2017.

R. L. Coley, E. Votruba-Drzal, M. A. Collins, and P. Miller. Selection into early education and care settings: Differences by developmental period. *Early Childhood Research Quarterly*, 29(3): 319–332, 2014.

K. Coneus, K. Goeggel, and G. Muehler. Maternal employment and child care decision. *Oxford Economic Papers*, 61(suppl_1):i172–i188, 2008.

R. Connelly. The effect of child care costs on married women's labor force participation. *The Review of Economics and Statistics*, pages 83–90, 1992.

E. E. Davis and R. Connelly. The influence of local price and availability on parents' choice of child care. *Population Research and Policy Review*, 24(4):301–334, 2005.

E. E. Davis, W. F. Lee, and A. Sojourner. Family-centered measures of access to early care and education. *Early Childhood Research Quarterly*, 47:472–486, 2019.

D. Dranove and G. Z. Jin. Quality disclosure and certification: Theory and practice. *Journal of Economic Literature*, 48(4):935–63, 2010.

G. J. Duncan and A. J. Sojourner. Can intensive early childhood intervention programs eliminate income-based cognitive and achievement gaps? *Journal of Human Resources*, 48(4):945–968, 2013.

D. M. Early and M. R. Burchinal. Early childhood care: Relations with family characteristics and preferred care characteristics. *Early Childhood Research Quarterly*, 16(4):475–497, 2001.

N. Forry, T. K. Isner, M. P. Daneri, and K. Tout. Child care decision making: Understanding priorities and processes used by low-income families in minnesota. *Early Education and Development*, 25(7):995–1015, 2014.

R. A. Gordon and P. L. Chase-Lansdale. Availability of child care in the united states: A description and analysis of data sources. *Demography*, 38(2):299–316, 2001.

W.-J. Han. Nonstandard work schedules and child care decisions: Evidence from the nichd study of early child care. *Early Childhood Research Quarterly*, 19(2):231–256, 2004.

J. R. Henly and S. Lambert. Nonstandard work and child-care needs of low-income parents. In S. M. Bianchi, L. M. Casper, and B. King, editors, *Work, Family, Health, and Well-being*. Lawrence Erlbaum Associates Publishers, 2005.

C. M. Herbst. The impact of quality rating and improvement systems on families' child care choices and the supply of child care labor. IZA Discussion Papers 10383, Institute for the Study of Labor (IZA), 2016.

V. J. Hotz and M. Xiao. Strategic information disclosure: The case of multiattribute products with heterogeneous consumers. *Economic Inquiry*, 51(1):865–881, 2013.

G. Z. Jin. Competition and disclosure incentives: An empirical study of hmos. *RAND Journal of Economics*, 36(1):93–112, 2005.

G. Z. Jin and P. Leslie. The effect of information on product quality: Evidence from restaurant hygiene grade cards. *The Quarterly Journal of Economics*, 118(2):409–451, 2003.

G. Z. Jin and A. T. Sorensen. Information and consumer choice: The value of publicized health plan ratings. *Journal of Health Economics*, 25:248–275, 2006.

J. Kim and M. S. Fram. Profiles of choice: Parents' patterns of priority in child care decision-making. *Early Childhood Research Quarterly*, 24(1):77–91, 2009.

L. Laughlin. Who's minding the kids? child care arrangements: Spring 2011. Technical report, US Census Bureau, 2013.

C. P. Li-Grining and R. L. Coley. Child care experiences in low-income communities: Developmental quality and maternal views. *Early Childhood Research Quarterly*, 21(2):125–141, 2006.

X. Liang, B. Fuller, and J. D. Singer. Ethnic differences in child care selection: The influence of family structure, parental practices, and home language. *Early Childhood Research Quarterly*, 15 (3):357–384, 2000.

L. Lippman, S. Vandivere, J. Keith, and A. Atienza. Child care use by low-income families: Variations across states. research brief. publication# 2008-23. *Child Trends*, 2008.

S. Loeb, B. Fuller, S. L. Kagan, and B. Carrol. Child care in poor communities: Early learning effects of type, quality, and stability. *Child Development*, 75(1):47–65, 2004.

E. D. Lowe and T. S. Weisner. 'you have to push it—who's gonna raise your kids?': Situating child care and child care subsidy use in the daily routines of lower income families. *Children and Youth Services Review*, 26(2):143–171, 2004.

A. D. Mathios. The impact of mandatory disclosure laws on product choices: An analysis of the salad dressing market". *The Journal of Law and Economics*, 43(2):651–678, 2000.

J. Milyo and J. Waldfogel. The effect of price advertising on prices: Evidence in the wake of 44 liquormart. *American Economic Review*, 89(5):1081–1096, 1999.

T. W. Morrissey. Child care and parent labor force participation: A review of the research literature. *Review of Economics of the Household*, 15(1):1–24, 2017.

National Survey of Early Care and Education Project Team. Fact sheet: How far are early care and education arrangements from children's homes? Technical Report OPRE Report No. 2016-10, Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services, 2016.

N. C. on Early Childhood Quality Assurance. Addressing the decreasing number of family child care providers in the united states. Technical report, National Center on Early Childhood Quality Assurance, 2019.

A. B. Shlay, H. Tran, M. Weinraub, and M. Harmon. Teasing apart the child care conundrum: A factorial survey analysis of perceptions of child care quality, fair market price and willingness to pay by low-income, african american parents. *Early Childhood Research Quarterly*, 20(4): 393–416, 2005.

S. Tang, R. L. Coley, and E. Votruba-Drzal. Low-income families' selection of child care for their young children. *Children and Youth Services Review*, 34(10):2002–2011, 2012.

E. Tekin. Child care subsidy receipt, employment, and child care choices of single mothers. *Economics Letters*, 89(1):1–6, 2005.

E. Tekin. Single mothers working at night: Standard work and child care subsidies. *Economic Inquiry*, 45(2):233–250, 2007.

K. Train. Welfare calculations in discrete choice models when anticipated and experienced attributes differ: A guide with examples. *Journal of choice modelling*, 16:15–22, 2015.

M. Xiao. Is quality accreditation effective? evidence from the child care market. *International Journal of Industrial Organization*, 28(6):708–721, 2010.