If You Build It, They Will Come: A Demand Analysis of Household-Level Childcare Choices*

Flávio Cunha[1], Jung Youn Lee[2] and Ritika Sethi[1]

[1]Department of Economics, Rice University
[2]Jones Graduate School of Business, Rice University

October 20, 2024

Abstract

This study investigates the phenomenon of childcare deserts by integrating three components. First, we introduce a social planner model that captures the trade-off between maximizing market shares for high-quality programs and prioritizing economically vulnerable locations. Second, we harness a unique dataset that links families with childcare providers, enabling the establishment of proximity-based choice sets and estimating demand for childcare seats. Third, we explore interventions two local governments undertake to identify the social planner's preferences. Our framework offers insights into expansion locations for high-quality programs while considering potential cannibalization effects.

*Flávio Cunha thanks the Powell Foundation for supporting this research and expresses sincere appreciation to the Texas Workforce Commission for sharing its data with the Texas Policy Lab. This article results from approved research using data from the Texas Workforce Commission. The results, opinions, recommendations, or points of view expressed in this document represent the work and consensus of the authors and do not necessarily represent the official position or policies of the Texas Workforce Commission or their funding organizations.

1 Introduction

Childcare enterprises typically operate as small-scale entities offering two primary services. First, these establishments provide children with a secure and supervised environment while their parents or guardians engage in work or educational pursuits. Extensive research indicates that this service significantly encourages women's participation in the labor force (Blau and Tekin, 2007; Posadas and Vidal-Fernandez, 2013; Cascio, Haider, and Nielsen, 2015). Second, childcare programs deliver early education and foster social interaction among children. This environment is crucial in nurturing cognitive, socio-emotional, and motor skills, preparing children for future educational endeavors. Notably, high-quality childcare environments have the potential to significantly enhance the lifelong trajectories of the children they serve (Herbst and Tekin, 2012; Heckman, Pinto, and Savelyev, 2013; Duncan and Sojourner, 2013; Baker, Gruber, and Milligan, 2019; Gray-Lobe, Pathak, and Walters, 2023).

Given the profound implications for parents and children, scholars and policymakers are deeply concerned about the spatial inequalities observed in the distribution of childcare programs. Of particular concern is the prevalence of "childcare deserts" in socioeconomically vulnerable areas, characterized by a severe lack of accessible and affordable childcare options, especially for families with young children (Bassok, Fitzpatrick, and Loeb, 2011; Bassok and Galdo, 2016; The Urban Child Institute, 2016). These childcare deserts manifest in various forms, including limited availability, affordability, and accessibility to high-quality programs (The Urban Institute, 2019; Child Care Aware of America, 2022).

Researchers currently identify childcare deserts by combining two distinct combining available datasets. The first dataset includes information about all registered childcare programs, containing details such as program location and capacity, typically sourced from government agencies responsible for industry regulation. The second dataset, exemplified by the American Community Survey, provides estimates of the number of children in specific geographic areas, such as census tracts. Researchers employ particular criteria to define childcare deserts, considering factors like childcare seat-to-child ratios, the proximity of providers to households, and the quality of childcare services. Within this academic context, Davis, Lee, and Sojourner (2019) have developed a methodology for identifying and delineating childcare deserts, representing the state-of-the-art in this area of research.

Our contribution focuses on the critical need to distinguish between two essential explanations for the uneven distribution of childcare firms across geographical regions. Firstly, economic barriers can pose significant obstacles, preventing childcare enterprises from entering markets with an apparent excess demand for childcare slots. For example, these businesses may struggle to secure loans from credit markets, which is essential for expansion into these areas. Despite the high demand, the constraint lies in the insufficient supply. This scenario exemplifies a "childcare desert," where the latent demand is ample to sustain new entrants financially, underscoring a gap in service provision.

Conversely, certain geographical regions might exhibit low demand for high-quality childcare programs despite a significant number of families with young children. This situation could stem from financial limitations or a lack of awareness among households regarding the availability of such services. Unlike a "childcare desert," introducing new providers in these markets may inadvertently lead to competition over the existing limited demand. This competition can jeopardize the economic viability of both new entrants and established providers. Our study sheds light on this complex interplay between market demand, financial barriers, and the geographical distribution of childcare services.

We improve the precision of identifying childcare deserts by innovating over existing approaches in three critical ways. First, we formulate a social planner problem to optimize the expansion of childcare provision, either by establishing new high-quality programs or by upgrading the quality of existing ones. Our method redefines a childcare desert as an area with substantial potential for enrolling children in new programs while minimizing the adverse impact on the market share of existing high-quality programs. This refined definition offers a more comprehensive assessment of childcare accessibility and market viability.

To implement this model, we require two critical inputs: household demand for childcare services and the social planner's preferences, specifically their trade-off between maximizing market shares and expanding supply in socioeconomically vulnerable areas. Together, the model and its two inputs allow us to prioritize areas for spatial interventions effectively.

A broader contribution of this social planner approach is pinpointing under-served areas that maximize social welfare while maintaining market viability. Beyond its application in childcare, this approach is relevant for addressing equity gaps in various sectors, such as education deserts (Hillman, 2019), healthcare deserts (Gregg and Peiser, 2023), and food deserts (Allcott, Diamond, Dubé, Handbury, Rahkovsky, and Schnell, 2019). Our approach provides valuable insights into designing effective place-based policies within the broader context of regional economics (Neumark and Simpson, 2015; Holmes and Sieg, 2015).

The second innovation of our methodology involves leveraging a state-owned dataset that matches families to childcare businesses. This dataset plays a crucial role in delineating childcare markets by identifying recurring choice patterns, with a particular focus on factors such as the geographical proximity between households and childcare facilities. Our approach to defining childcare markets involves identifying overlapping geographical regions, specifically delineated by rings with a 10-mile radius centered around the centroid of a given census tract. Based on this dataset, our analysis underscores the effectiveness of our childcare market definition, as it successfully encompasses most cases within our comprehensive dataset.

We employ our dataset and the defined childcare market framework to estimate household demand for childcare services. Recent research highlights a common phenomenon wherein families receiving childcare subsidies often opt for options that may not represent the highest quality (Johnson, Ryan, and Brooks-Gunn, 2012). Additionally, high-quality childcare programs tend to be concentrated in more affluent neighborhoods (Hatfield, Lower, Cassidy, and Faldowski, 2015). This observation raises pertinent questions regarding the potential impact of expanding the supply of high-quality childcare services near disadvantaged families to encourage their enrollment in such programs. Addressing this question involves tackling two pivotal challenges inherent in estimating childcare demand.

First, our childcare choice model incorporates provider-specific unobserved heterogeneity, necessitating the introduction of provider fixed effects. These fixed effects are crucial for capturing inherent disparities in unobservable provider attributes that influence families' decision-making processes. However, estimating these fixed effects for a large number of providers poses significant computational challenges. We employ a two-step maximum likelihood estimation method that integrates a contraction mapping technique (Berry, Levinsohn, and Pakes, 1995; Goolsbee and Petrin, 2004).

Second, our model addresses the potential endogeneity of provider quality ratings. Our dataset's overlapping structure of childcare markets allows us to construct instrumental variables to mitigate this issue (Fan, 2013). We categorize competitors into two groups: direct competitors, found within the choice sets of at least one common household, and indirect competitors, present in the choice sets of households with no common selections with direct competitors. To construct our instrument for quality ratings, we leverage variations in provider characteristics and market demographics associated with indirect competitors. Our identifying assumption relies on the independence of the variation in quality ratings attributed to the presence of indirect competitors (and the markets they serve) from the variation in unobserved heterogeneity in preferences.

Our analysis uncovers several crucial insights into families' childcare choices. First, proximity emerges as an essential factor influencing parents' decisions when selecting childcare options. However, we also find that families are willing to travel longer distances when higher-quality childcare programs are available. This finding indicates that parents prioritize quality and are willing to make trade-offs regarding distance to access better care for their children. Furthermore, our findings underscore disparities in enrollment in high-quality childcare programs. Even after accounting for provider fixed effects and addressing endogeneity concerns, families in socially vulnerable neighborhoods are less likely to choose high-quality childcare options. This highlights the persistent barriers these families face in accessing and selecting optimal childcare services for their children.

The third innovation of our analysis involves direct interventions in the childcare market within two counties in Texas, which provide insights into the preferences of the social planner. Our model allows the social planner to assign varying degrees of importance to different geographic locations. For example, the social planner may prioritize expanding market share for high-quality childcare programs in areas characterized by notable socioeconomic vulnerability over other regions. To illustrate this, we examine two specific cases where local governments established childcare programs or purchased childcare seats in locations identified as "childcare deserts," each with varying socioeconomic vulnerability. By aligning our social planner's model with the decisions made by local governments, we can identify and quantify these preferences effectively.

Together, these three innovations enable us to construct a comprehensive framework that policymakers can use to strategically select optimal locations for expanding early care and education supply. Our analysis emphasizes the contribution of demand dynamics and the social planner's preferences. Our findings consistently highlight that understanding the demand for childcare seats is crucial for identifying "childcare deserts." For instance, our analysis quantifies the potential impact of supply-side expansion initiatives in addressing disparities in access to high-quality childcare between socially advantaged and disadvantaged households. While socially disadvantaged households may show slightly weaker preferences for high-quality childcare, our findings suggest this difference is relatively minor. Therefore, in areas with excess demand, supply-side policy interventions prove more effective in increasing high-quality childcare enrollment among socially vulnerable households than demand-side interventions. In summary, our results underscore the importance of avoiding locations with high cannibalization rates to promote equitable access to high-quality childcare. This approach ensures that expansion efforts effectively meet the needs of all families, especially those in socioeconomically vulnerable areas.

The policy implications of this study are particularly pertinent in the context of the benefits associated with universal early childhood education (ECE) programs. Extensive research consistently demonstrates the positive effects of high-quality ECE programs on children's cognitive, social, and emotional development, as well as their long-term educational outcomes (Gormley Jr, Gayer, Phillips, and Dawson, 2005; Baker, Gruber, and Milligan, 2019; Durkin, Lipsey, Farran, and Wiesen, 2022; Silliman and Mäkinen, 2022; Gray-Lobe, Pathak, and Walters, 2023). Programs like Head Start, Perry Preschool, and the Infant Health and Development Program have been extensively evaluated and shown both short-term and long-term advantages for participating children (Ludwig and Miller, 2007; Kline and Walters, 2016; Heckman, Pinto, and Savelyev, 2013; Duncan and Sojourner, 2013).

Living in a childcare desert, where access to high-quality childcare is limited or nonexistent, imposes high costs on families and children, particularly among vulnerable populations (Decker and Kelly, 2022). This lack of access can detrimentally affect children's early development, well-being, and school readiness, potentially perpetuating inequalities in educational and social outcomes. Addressing these disparities through targeted policy interventions to expand access to high-quality childcare is crucial for promoting equitable opportunities and reducing early childhood inequalities.

Our study contributes to the existing literature by exploring the implications of ECE expansion policies within the childcare market. The field of research on the childcare market has seen various studies employing diverse methodologies and data sources to investigate different aspects of childcare. Some recent studies have examined parents' preferences and analyzed the distributional effects of ECE expansion policies using provider-level data from specific regions, such as Minnesota and Pennsylvania (Borowsky, 2019; Bodéré, 2022). Others have utilized nationally representative samples to understand how parents make childcare choices (Berlinski, Ferreyra, Flabbi, and Martin, 2020; Borowsky, Brown, Davis, Gibbs, Herbst, Sojourner, Tekin, and Wiswall, 2022). Additionally, research has focused on the supply side of the market, using establishment-level data and licensure information to examine impacts on childcare provision (Hotz and Xiao, 2011; Bassok, Fitzpatrick, and Loeb, 2014; Brown, 2018). These studies shed light on important aspects of the childcare landscape and provide complementary insights to our paper.

The remainder of this paper is organized as follows. Section 2 provides an overview of the institutional background, and Section 2.2 describes the data employed in the study. Section 3 presents the model and estimation method, and Section 5 presents the model estimates. Section 5.3 presents the policy simulations, followed by the conclusion in Section 6.

2 Institutional Background and Data

2.1 Institutional Background

Our analysis focuses on how families make decisions regarding childcare, specifically examining the subsidized childcare initiative in Texas—the Childcare Services (CCS) program, managed by the Texas Workforce Commission (TWC) and its Local Workforce Development Boards. This program, funded by the Child Care Development Fund (CCDF), assists eligible families by easing the financial burden of childcare expenses, enabling parents to work or pursue education. Registered or licensed childcare providers can participate and receive reimbursement from TWC for offering care to eligible families. Within this program, eligible families can choose from participating providers.

Local Workforce Development Boards: Local Workforce Development Boards (LWDBs) are at the forefront of aiding low-income parents in pursuing education and training. As crucial intermediaries between TWC and local communities, LWDBs collaborate with the TWC to tailor some CCS program parameters to the unique needs of their regions. The Online Appendix A.1 provides further details on the specific geographic jurisdictions of the LWDBs across Texas. LWDBs are essential in disseminating critical information about the CCS program to eligible families. These boards facilitate access to childcare services by providing extensive and detailed information regarding various options. To ensure widespread accessibility of this information, it is disseminated through digital (LWDB's official website) and physical (on-site at childcare facilities) platforms. A key focus is placed on the quality of childcare services within the CCS program, with providers prominently displaying various quality indicators to inform and reassure parents. The nature of these specific quality indicators within the CCS program warrants a more in-depth examination, which will be addressed in the subsequent sections of this paper.

LWDBs receive applications for childcare subsidies from families. If the family gets a subsidy, the family chooses a subsidy-accepting provider, and the LWDBs authorize details such as start and end dates, days and hours of care, and applicable parent co-payments.

Providers: Childcare providers participating in the CCS program are categorized into four distinct models: licensed center-based programs, licensed home-based programs, registered home-based programs, and relative care. Licensed center-based programs are operated in non-residential, dedicated facilities. These environments are managed by professional staff and teachers who offer structured care and educational activities specifically tailored to cater to various age groups. Such programs are characterized by their emphasis on a structured environment and a curriculum designed to promote child development.

In contrast, licensed home-based programs are conducted in the permit holder's residence and are designed to accommodate up to twelve children. Providers of these programs must meet specific director qualifications, ensuring a standard of care and educational oversight. Registered home-based programs, slightly smaller in scale, offer care for up to six children, with the option to extend services to an additional six school-age children outside regular school hours. Unlike licensed home-based programs, these do not necessitate the provider to meet director qualifications, allowing for more flexible care arrangements.

Lastly, relative care represents a more personalized form of childcare, typically provided by family members within the child's extended family network.

The CCS childcare programs enter into agreements with TWC to receive reimbursement for their childcare services to eligible families. Some (but not all) childcare programs participate in the CCS. Providers accepting CCS referrals must serve children within their licensed age group and cannot discriminate based on parental income, public assistance receipt, or child protective service status. Additionally, they must adhere to their licensed capacity.

Providers in the CCS program can participate in the Texas Rising Star (TRS) certification program, emphasizing specific quality indicators.[1] TRS operates on a tiered system (2 to 4 Stars), evaluating staff qualifications, nutrition, parent involvement, and more. Licensed/registered childcare homes receive alternative TRS ratings – deferred status, provisionally certified, and fully certified ratings – that align with TRS 2-, 3-, and 4- Stars. Providers may also demonstrate quality through Texas School Ready, National Accreditation, or other indicators, automatically qualifying them for TRS 4-Star.

CCS childcare providers receive reimbursement from TWC for providing care to children who receive subsidy assistance. LWDBs determine the reimbursement rates based on a market rate survey conducted by the Texas Institute for Child and Family Wellbeing at the University of Texas

[1]Starting October 2023, all CCS providers must participate in TRS.

at Austin. This annual survey collects data on childcare prices and reports daily rates in percentiles to capture their distribution in each local childcare market. The reimbursement rates consider factors such as the provider types (e.g., childcare center, childcare home), child age groups, care hours (e.g., full-time or part-time care), and the provider's quality indicator (e.g., TRS certification level). The Online Appendix A.1 provides an example of how reimbursement rates are currently determined in Texas. Center programs receive higher reimbursement rates than licensed/registered childcare homes and relatives. Generally, higher-rated TRS providers receive higher reimbursement rates to reward higher-quality childcare services. In addition, full-time care generally receives higher reimbursement rates than part-time care. Reimbursement rates decrease as the age of the child increases.

Families: Eligible families receive financial assistance to cover childcare costs, facilitating access to affordable childcare services. The CCS program supports families across various income levels, with eligibility determined by income, family size, and work or education requirements.

The eligibility criteria, as per the Texas Administrative Code §809.41 (2007), include:

1. Age: The CCS program provides childcare support to families with children aged thirteen or less.

2. Work/School: Parents must be actively employed or participating in a job training/educational program to qualify, ensuring assistance aligns with families' employment and skill development needs.

3. Income: The family's income should not exceed 85% of the state median income or 150% of the federal poverty income level, targeting assistance to financially constrained families. Temporary increases beyond 85% of the state median income don't impact eligibility.

Eligible parents choose a childcare provider upon receiving benefits, adhering to federal and state laws allowing informed decisions. The LWDB then contacts the selected provider to confirm space availability and authorize childcare services.

Subsidized families may pay a copay, determined by income, family size, and the number of children needing care. Copay amounts are detailed in Table A.2 in Online Appendix A. Notably, the copay is independent of the childcare program the family decides to enroll their children in. Certain circumstances exempt families from copay, including participation in the Choices program, SNAP E&T program, homelessness, or receipt of protective services.

2.2 Data

In this section, we present the foundational data that supports our analysis to identify high-quality childcare deserts. At the heart of our research lies the comprehensive administrative dataset on childcare subsidies provided by the TWC. To enrich our understanding and add layers to our analysis, we further integrate this primary dataset with detailed census-level data regarding the Social Vulnerability Index (SVI) obtained from the Centers for Disease Control and Prevention (CDC). Additionally, we incorporate data on spatial interventions within the childcare market executed by local government agencies. This multifaceted approach allows us to gain a deeper, more nuanced understanding of the landscape of high-quality childcare availability.

2.2.1 Administrative Childcare Subsidy Data

Our analytical framework draws upon a rich family-provider matched dataset acquired through a collaborative data-sharing agreement with the TWC.[2] This dataset encompasses the years 2015 to 2019 and pertains to households, children, and childcare providers participating in the CCS program in Texas. This dataset includes comprehensive geographical addresses for households and childcare providers, which distinguishes it, establishing a distinctive and invaluable linkage between them. This high level of granularity serves as the linchpin of our analysis, enabling not only the precise matching of households with the childcare programs their children are presently enrolled in but also facilitating the exploration of alternative choices these households could have potentially made.

Providers Table 1 presents key statistics summarizing childcare home or center programs participating in Texas' subsidized childcare program between 2015 and 2019, totaling 11,445 programs. Predominantly, over seventy-five percent are childcare center-based programs, while the remainder consists of childcare home-based programs.

Regarding ownership, around half of the programs are affiliated with for-profit private organizations, with thirty-four percent being individually owned or operating as sole proprietorships. Non-profit private organizations represent seventeen percent, and government entities own approximately three percent of the childcare programs.

Regarding TRS quality ratings, the majority (eighty percent) lack a TRS rating. Only two

[2] This study marks the inaugural phase of a comprehensive methodology aimed at assessing the impact of the Texas Rising Star program on child outcomes.

and three percent hold TRS 2 and 3 Star ratings, respectively, while fifteen percent have a TRS 4 Star rating. TRS 2 Star includes deferred status home-based programs, TRS 3 Star includes provisionally certified home-based programs, and TRS 4 Star includes fully certified home-based programs and nationally accredited programs.

Table 1. Descriptive statistics: Provider characteristics

	Mean	SD	Min	Max
Program Type				
Center Program	0.76	0.42	0.00	1.00
Home Program	0.24	0.42	0.00	1.00
Ownership				
Governmental entity	0.03	0.17	0.00	1.00
Non-Profit Private Organization	0.17	0.37	0.00	1.00
For-Profit Private Organization	0.47	0.50	0.00	1.00
Individual	0.34	0.47	0.00	1.00
Quality Rating				
No TRS rating	0.80	0.40	0.00	1.00
TRS 2 Star	0.02	0.15	0.00	1.00
TRS 3 Star	0.03	0.17	0.00	1.00
TRS 4 Star	0.14	0.35	0.00	1.00
Licensed Capacity				
Licensed Capacity	77.74	66.23	9.00	322.00
Obs		11,445		

Note: The analysis includes 11,445 childcare programs that participated in Texas' subsidized childcare program between the years 2015 and 2019.

Source: Texas Workforce Commission

On average, a program has a licensed capacity of around 80 childcare seats, ranging from 9 to 322 seats. It's important to note that this licensed capacity pertains to the general market, encompassing all families and not exclusively those seeking subsidized care.

Households Table 2 presents key statistics summarizing households seeking childcare under Texas' CCS program between 2015 and 2019, totaling 113,368 households. These households have at least one child below the age of three. About 45% of households are Hispanic, and 33% are Black. Households are primarily single-parent families with an average of two dependents and a monthly family income of $1,156.

We look at the first time a family enrolls with a childcare provider as part of the CCS program. Approximately half of the families are exempt from copay, potentially indicating eligibility for additional financial assistance. On average, households pay $19 in copay, including those exempt.

The maximum copay in this sample is \$444. Families, on average, utilize childcare for about 12 days per month.

Table 2. Descriptive statistics: Household characteristics

	Mean	SD	Min	Max
Demographics				
White	0.16	0.37	0.00	1.00
Black	0.33	0.47	0.00	1.00
Hispanic	0.45	0.50	0.00	1.00
Single parent	0.88	0.33	0.00	1.00
Number of dependents	2.06	1.27	0.00	6.00
Income	1155.13	1163.13	0.00	4372.52
Childcare Program Usage				
Exempt from copay	0.48	0.50	0.00	1.00
Copay	18.95	42.90	0.00	444.00
Number of days	11.62	8.17	1.00	92.00
Enrolled with multiple programs	0.03	0.17	0.00	1.00
Center Program	0.97	0.17	0.00	1.00
Privately-owned Program	0.20	0.40	0.00	1.00
Individually-owned Program	0.65	0.48	0	1.00
TRS 2 Star Program	0.04	0.19	0.00	1.00
TRS 3 Star Program	0.07	0.25	0.00	1.00
TRS 4 Star Program	0.33	0.47	0.00	1.00
Distance	3.83	3.85	0.02	24.62
Obs.	113,368			

Note: The analysis includes 113,368 households with at least one child below three enrolled in Texas' subsidized childcare program between 2015 and 2019.

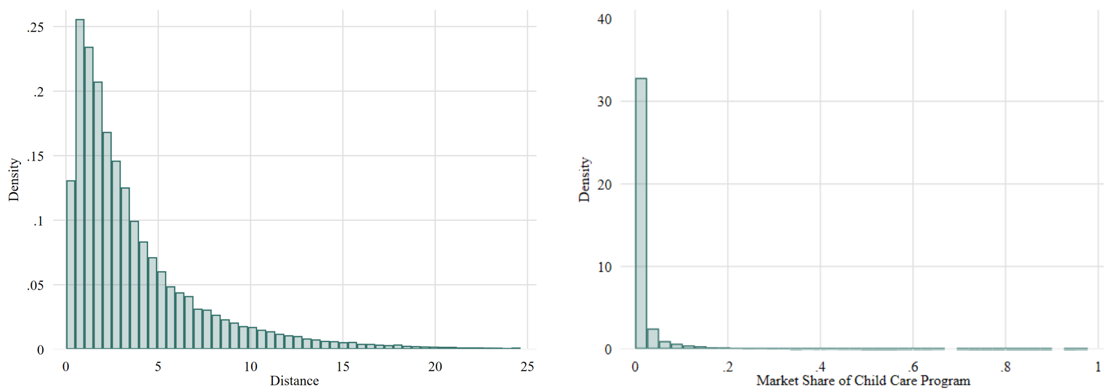
Source: Texas Workforce Commission

A small proportion of families (about three percent) in the sample are concurrently enrolled in multiple childcare programs. We define the "primary" childcare provider of such a household as the one where the household pays the highest copay and uses the most childcare days.

Table 2 further reveals that ninety-seven percent of households are enrolled in center programs. Fifteen percent opt for government or non-profit-owned programs, twenty percent for individually owned, and sixty-five percent for privately owned programs. About ten percent are enrolled in TRS 2 or 3 Star-rated programs, while thirty percent choose TRS 4 Star programs.

Understanding travel distances to childcare is crucial to understanding the alternatives available and the trade-offs households make. We measure the distance between the family's residence and the selected childcare provider by calculating the miles between the centroid of the census tract where the family resides and the geographic location of the chosen provider. The average distance households travel to their enrolled childcare program is 3.83 miles. Figure 1 (a) depicts that over ninety percent of families choose programs within a 10-mile radius, guiding our definition of childcare markets as overlapping regions within this distance. This approach allows us to explore factors influencing families' decisions within a reasonably defined geographic scope.

Figure 1. Distribution of Distance and Market Shares



Note: Figure (a) plots the distribution of the distance (in miles) between families and the childcare program that their children are enrolled in. Figure (b) plots the distribution of the market shares of the childcare programs.
Source: Texas Workforce Commission

Table 3 outlines families' available options within a 10-mile radius, with an average of 167 childcare programs. Predominantly, families encounter center-based programs owned by for-profit entities, often without TRS ratings. The nearest program is typically about 0.97 miles away, with center-based programs owned by for-profit entities and programs with no TRS ratings being closer than other categories.

Figure 1 (b) plots the distribution of the market shares of the childcare programs in our sample. Market share is defined as the number of families enrolled in the program divided by the total number of families within 10 miles of the program who could have enrolled. Below, we will discuss in more detail how we construct these market shares. As shown in the figure, most programs have small market shares. We include only those programs with a non-zero market share in our analysis, reducing the number of programs to 7,688. The necessity of excluding programs with a zero market share will become apparent when we discuss the model estimation.

Table 3. Descriptive statistics: Choice Set

	Mean	S.D.	Min	Max
Number of Programs				
All	167.26	121.07	1.00	494.00
Program Type: Home	34.77	30.18	0.00	128.00
Program Type: Center	132.48	97.07	0.00	395.00
Ownership: Government or non-profit	20.89	17.57	0.00	71.00
Ownership: Individual	60.73	56.55	0.00	267.00
Ownership: For-profit Organization	85.64	63.59	0.00	270.00
TRS Rating: None	112.84	86.76	0.00	348.00
TRS Rating: 2 Or 3 Star	22.80	21.40	0.00	98.00
TRS Rating: 4 Star	31.62	23.85	0.00	101.00
Distance				
All	0.97	1.28	0.02	10.00
Program Type: Home	1.76	1.76	0.03	9.99
Program Type: Center	1.07	1.33	0.02	10.00
Ownership: Government or non-profit	2.11	1.96	0.03	10.00
Ownership: Individual	1.52	1.69	0.02	9.97
Ownership: For-profit Organization	1.22	1.42	0.02	10.00
TRS Rating: None	1.13	1.38	0.02	10.00
TRS Rating: 2 Or 3 Star	2.04	1.83	0.04	10.00
TRS Rating: 4 Star	1.68	1.67	0.03	9.99
Obs.		113,368		

Note: The analysis includes 113,368 households with at least one child below three enrolled in Texas' subsidized childcare program between 2015 and 2019. We summarize the characteristics of the childcare programs in households' choice sets (i.e., programs located 10 miles around the household.)

Source: Texas Workforce Commission

2.2.2 Social Vulnerability Index

To understand the patterns of childcare program utilization across families of various socioeconomic conditions, we merge the above TWC data with information on the Social Vulnerability Index (SVI). Developed by the Center for Disease Control and Prevention (CDC), the SVI ranks census tracts on unemployment, racial and ethnic minority status, and disability. Higher SVI values indicate a greater level of social vulnerability in a tract.

Specifically, the SVI is calculated using a range of variables from the American Community Survey spanning 2014 to 2018. These variables capture various aspects of socioeconomic status, including the number of individuals living below the poverty line, the number of unemployed individuals, per capita income, and the number of individuals without a high school diploma. Household composition and disability variables are also considered, such as the number of individuals aged 17 or younger, the number of individuals with disabilities, and the number of individuals living in single-parent households. Additionally, minority status and language-related variables are included, such as the percentage of individuals classified as a minority and those who speak English "less than well." Housing types and transportation variables, such as multi-unit structures, mobile homes, crowding, lack of a vehicle, and group quarters, are also factored into the SVI calculation.

2.2.3 Local Government Interventions

Next, we describe two American Rescue Plan Act (ARPA) funded interventions to inform the trade-off between increasing enrollment in high-quality programs and prioritizing highly socially vulnerable locations.

The first intervention involves a Request for Applications (RFA) targeting political subdivisions such as cities and school districts to fund childcare facilities in County A, Texas. These facilities must meet specific criteria, including a minimum size of 12,200 square feet, eight classrooms, and an indoor play area ranging between 360 to 1,800 square feet. Priority is given to areas identified as childcare deserts, utilizing maps produced by Children at Risk[3]. This initiative requires the contribution of land or buildings, either through donation or via a no or low-cost, long-term lease of at least 30 years, as a condition for application consideration. Our analysis includes data on the winning applications, including the geographical locations of the new sites.

The second intervention we explore is an initiative by County B in Texas, which strategically addresses the need for accessible childcare by purchasing seats directly from childcare providers and allocating them to families that meet specific eligibility criteria. This approach allows center-based childcare programs to sell seats to the county, which are then assigned to eligible families through a meticulous application process.

On the family side, seat applications are submitted online, requiring proof of identification and eligibility. A county official reviews each application to determine a child's eligibility based on age, residential address, and household income. Children who meet the eligibility criteria are enrolled in one of the participating center-based programs, ensuring that families in need have access to quality childcare services.

The selection process for childcare centers is rigorous and based on a comprehensive prioritization matrix. This matrix evaluates each application on several critical factors: the quality of the center as indicated by its TRS rating, the SVI of its location, and whether the area qualifies as a childcare desert according to the desert map produced by the Texas Policy Lab at Rice Univer-

[3]See the Children at Risk childcare desert map here: https://childrenatrisk.org/childcaredeserts/.

sity[4]. Centers that score highly undergo an unannounced visit from county representatives, who assess various aspects, including the center's capacity, the quality and availability of equipment and materials, group sizes and staff-to-child ratios, the playground area, and overall center practices. During these visits, interviews with center directors are also conducted.

Following this assessment, the highest-rated centers are offered a contract to participate in the program. We obtained information on the locations of the 23 center-based programs that were awarded this contract.

Table 4 summarizes the Social Vulnerability Index (SVI) for all the census tracts across Texas in our sample and the two counties involved in the interventions. County B has a relatively higher SVI, indicating greater social vulnerability. Both interventions strategically select locations using childcare desert maps generated from current data, which notably do not incorporate the demand for subsidized, high-quality childcare programs. This critical observation is integrated into our analysis, allowing us to more accurately estimate the preferences of the social planner within our framework.

Table 4. Descriptive statistics: Social Vulnerability Index

	N	Mean	SD	Min	Max
All census tracts	4,911	0.52	0.28	0.01	1.00
County A	350	0.47	0.29	0.01	1.00
County B	742	0.58	0.30	0.01	1.00

Data Source: Center for Disease Control and Prevention

3 Model

This section defines the social planner problem of selecting locations to improve access to high-quality childcare. We begin with an overview of the optimal location problem, emphasizing the goal of identifying sites with the highest expected market share while minimizing the cannibalization of existing high-quality programs. Next, we present the problem of a family choosing a childcare provider from their available options, discussing the demand model specification, identification, and estimation in detail. Finally, we revisit the social planner's problem with a more comprehensive analysis.

[4] See the Texas Policy Lab childcare desert maps here: https://tplapps.rice.edu/shiny/texas-county-child-care-deserts-03/.

3.1 The Problem of the Social Planner

A local government agency is contemplating establishing a new TRS 4 Star childcare center program. The primary objective is to select a location that maximizes the anticipated number of enrollments for the new program. The emphasis is on serving higher SVI neighborhoods while minimizing the potential impact on the market shares of existing TRS 4 Star childcare programs. We formulate the optimal location problem of the social planner as follows:

$$\mathbf{a}^* = \{\mathbf{a}_\mathbf{k}^*\}_{\mathbf{k}\in\mathcal{C}} = \arg\max_{\mathbf{a}} \sum_{k\in\mathcal{C}} a_k \sum_{c\in\mathcal{C}_k} w\left(y_c;\tau\right) \sum_{i\in\mathcal{I}_c} \left[p_{ick}\left(\mathbf{a};\theta\right) - \sum_{\ell\in\mathcal{T}_c} \Delta p_{ic\ell}\left(\mathbf{a};\theta\right) \right]$$
$$\text{subject to } a_k \in \{0,1\}; \sum_{k\in\mathcal{C}} a_k = 1. \tag{1}$$

In our framework, the agency's decision-making process involves determining an assignment vector $\mathbf{a} \equiv \{\mathbf{a}_\mathbf{k}\}_{\mathbf{k}\in\mathcal{C}}$ where $\mathcal{C} = \{1,\ldots,C\}$ represents the collection of census tracts within the agency's jurisdiction. Here, $a_k \in \{0,1\}$ signifies whether a new program will be inaugurated in the census tract $k \in \mathcal{C}$. This binary decision framework effectively maps out the geographical distribution of the new ECE program across the specified census tracts.

The function $w(y_c;\tau)$ denotes the weight assigned to the census tract c, based on its SVI (y_c), which ranges from 0 to 1, as defined in Section 2.2.2. This index quantifies the relative socioeconomic vulnerability of geographic locations, with higher values indicating greater socioeconomic vulnerability. The agency uses this index to assign priority weights to census tracts, integrating social equity considerations into the strategic placement of new ECE programs. These priority weights are crucial to ensure that the allocation of ECE resources aligns with broader social objectives, such as reducing inequalities in access to high-quality early childhood education. We assume that the local agency uses the following weighting function:

$$w\left(y_c;\tau\right) = \frac{\exp\left(\frac{y_c}{\tau}\right)}{\sum_{c'\in\mathcal{C}} \exp\left(\frac{y_{c'}}{\tau}\right)}, \tag{2}$$

where a lower value of $\tau > 0$ indicates the weights are more sensitive to socioeconomic vulnerability. We use two supply-side interventions to identify a range of values for τ, which we discuss in detail below.

The designation of $\mathcal{C}_c \equiv \{c : d_{ck} \leq r\}$ identifies the set of census tracts within a radius r of the census tract k, encompassing those who are within the practical reach of the proposed new ECE facility. This proximity criterion is critical to assess the accessibility of the new program to potential beneficiaries.

The function $p_{ick}(\mathbf{a};\theta)$ denotes the probability that the household i residing in census tract c, i.e., $i \in \mathcal{I}_c$ enrolls in the new TRS 4 Star center program that opens in the census tract k. given the assignment policy $\mathbf{a}$ and the demand parameters θ, which we derive below. It is important to recognize that the social planner regards the demand function as fixed within our framework. The model posits that the primary mechanism available to the social planner to influence household behavior is altering the geographic distribution of high-quality childcare services.

The term $\sum_{\ell \in \mathcal{T}_c} \Delta p_{ic\ell}(\mathbf{a};\theta)$ is the potential cannibalization of the demand for the incumbent TRS 4-Star programs, i.e., $\mathcal{T}_c = \{\ell : d_{c\ell} \leq r, \ell \text{ is TRS 4 Star}\}$. This term captures the interaction between the current landscape of a childcare market and expansion initiatives, as existing childcare programs are likely to affect the market share of the new programs and vice versa.

Cannibalization by the new program in the demand of the household i residing in census tract c for the incumbent TRS 4 Star program $\ell \in \mathcal{T}_c$ is the difference between the demand of the household i for the program ℓ if a new program does not open at all, that is, $a_k = 0 \; \forall \; k \in \mathcal{C}$ and the demand of the household i for the program ℓ given new assignment $\mathbf{a} = \{\mathbf{a_k}\}_{\mathbf{k}\in\mathcal{C}}$:

$$\Delta p_{ic\ell}(\mathbf{a};\theta) \equiv p_{ic\ell}(\{a_k = 0\}_{k\in\mathcal{C}};\theta) - p_{i\ell}(\mathbf{a};\theta) \tag{3}$$

Thus, the cannibalization of the enrollment of household i in all incumbent TRS 4-Star programs is $\sum_{\ell \in \mathcal{T}_c} \Delta p_{ic\ell}(\mathbf{a};\theta)$.

Our social planner approach contrasts with the dominant decentralized problem in the literature. For example, Borowsky, Brown, Davis, Gibbs, Herbst, Sojourner, Tekin, and Wiswall (2022) study the childcare market through the lens of a general equilibrium framework, in which firms operate in a perfectly competitive market. Berlinski, Ferreyra, Flabbi, and Martin (2020) considers a static equilibrium model in which firms make decisions under monopolistic competition. These two studies focus on quantifying the responses to interventions on the demand side, either by changes to the childcare subsidy program or the provision of direct cash transfers. Bodéré (2022) combines a dynamic model of providers' entry, exit, and quality investments with a static model of spatial competition and quantifies how equilibrium allocations change in response to demand and supply-side interventions. In particular, the intervention on the supply side considers start-up grants and higher reimbursement rates for high-quality providers.

In contrast, our study adopts a centralized approach, aligning with our goal of identifying regions deficient in high-quality subsidized childcare. We theorize that a locality can be defined as a high-quality subsidized childcare desert if adding a new premier childcare provider to the market barely affects the demand for existing services. In such under-served areas, introducing a new high-quality provider would likely benefit families on waiting lists for existing, high-quality, subsidy-accepting childcare programs, thereby pinpointing where the demand for quality childcare surpasses the available supply, particularly for those in need of subsidized options.

Next, we discuss our formulation and derivation of the demand function for childcare services.

3.2 The Problem of the Family

3.2.1 The Childcare Choice Set

We start by defining the childcare alternatives that a family can choose from. Family i resides in census tract c. Let d_{cj} denote the distance between the centroid of the census tract c and provider j. Guided by the distribution of distances between families and their selected providers in our sample (Figure 1), we define family i's choice set as the set of childcare programs located within 10 miles from its census tract of residence, i.e., $r = 10$:

$$\mathcal{J}_c \equiv \{\{j : d_{cj} \leq r\} \cup \{0\}\}, \tag{4}$$

where $j = 0$ denotes the outside option. We remark that the household takes the choice set as given.

3.2.2 The Utility Function

The utility of family i, which resides in census tract c, from enrolling its children in the childcare program $j \in \mathcal{J}_c \backslash \{0\}$ is:

$$u_{icj} = \beta_0 + \alpha_0 d_{cj} + \sum_{k=1}^{K} (\beta_k + \alpha_k d_{cj} + \gamma_k y_c) x_{jk} + \xi_j + \varepsilon_{ij}. \tag{5}$$

where $\boldsymbol{x}_j$ is a vector of observed provider characteristics, d_{cj} is the distance between census tract c and the provider j, and y_c is the SVI of the census tract. The term ξ_j captures the unobserved characteristics specific to the provider j, and ε_{ij} is the unobserved idiosyncratic taste of the family i for the provider j. The vector $\boldsymbol{x}_j \equiv \{x_{jk}\}_{k\in\{1,\dots,K\}}$ consists of the K characteristics of the childcare program that we observe in the data. These characteristics include the program type (whether the childcare program is center-based or home-based), ownership type (whether the program is owned by a government entity, a non-profit organization, a for-organization, or an individual), and the TRS rating of the program (i.e., whether the program has a TRS 4 Star rating, a TRS 3 Star rating, a TRS 2 Star rating, or no TRS rating at all).

Our goal is to estimate the parameter vector (α, β, γ). The term α_0 captures how the marginal utility of household i from enrolling with program j varies with the distance between them. The term β_k captures the marginal utility from the k^{th} characteristic. The term α_k captures how the marginal utility from the k^{th} characteristic of the program varies by distance. The term γ_k captures how the marginal utility from the k^{th} characteristic varies by the SVI of the census tract that the family resides in.

3.2.3 The Demand Function

Normalizing the utility from the outside option to zero, family i chooses provider j if and only if $u_{icj} > u_{icj'}\ \forall j' \neq j$, which gives us the probability that family i chooses provider j (conditional on $\boldsymbol{x}_j, d_{cj}, y_c$) as the following:

$$p_{icj} = \int_{\{e_{ij}: u_{icj} > u_{icj'} \forall j' \neq j\}} dF(e_{ij}), \tag{6}$$

where $e_{ij} = \varepsilon_{ij} + \xi_j$ represents the composite unobserved taste. In the case that ξ_j correlates with the observed characteristics, e_{ij} will not be independent of $(\boldsymbol{x}_j, d_{cj}, y_c)$, thereby resulting in biased estimates for the parameters in Equation (5). For instance, center programs may offer better customer service that parents value, but we do not observe these features in the data. Another unobserved characteristic of a program is its capacity, and it is possible that high-quality programs have binding capacity constraints. In both cases, ignoring the correlation between unobserved and observed provider characteristics can lead to overestimating the value parents place in center programs.

Our empirical strategy to address this potential correlation is to include provider dummies δ_j that subsume the impact of both observable and unobservable provider-specific characteristics (Berry, Levinsohn, and Pakes, 2004; Goolsbee and Petrin, 2004):

$$\delta_j = \beta_0 + \sum_{k=1}^{K} \beta_k x_{jk} + \xi_j. \tag{7}$$

The utility of family i from choosing provider $j \in \mathcal{J}_i \setminus \{0\}$ can now be written as:

$$u_{icj} = \delta_j + \alpha_0 d_{cj} + \sum_{k=1}^{K} (\alpha_k d_{cj} + \gamma_k y_c) x_{jk} + \varepsilon_{ij}. \tag{8}$$

Once we condition on δ_j, the part of e_{ij} that correlates with the observed characteristics, the residual component ε_{ij} does not correlate with the observed characteristics. An advantage of this approach is the imposition of minimal restrictions on the mean or variance of the unobserved provider characteristics or the covariance of the unobserved characteristics across providers. However, estimating the fixed effects of each provider, which amounts to around 10,000 fixed effects, is computationally demanding. Although, in principle, these fixed effects can be estimated using maximum likelihood estimation, just like the other demand parameters, locating the maximum in such a high-dimensional parameter space presents challenges in practice. To make this problem more tractable, we employ Berry, Levinsohn, and Pakes (1995)'s algorithm, which locates the fixed effects conditional on other model parameters to concentrate them out during estimation instead of estimating them together in a single step.

Next, we will discuss the estimation of p_{icj}, the probability that family i residing in census tract c enrolls in the childcare program j. In essence, this step is the estimation of the demand for a childcare program in the family's choice set.

3.3 Computational Approximation of the Social Planner's Problem

4 Identification and Estimation

4.1 Demand for Childcare Services

Our estimation consists of two steps. First, we maximize the likelihood function using household-level data and including the separate provider-specific dummies. This step identifies all parameters except β_0 and $\{\beta_k\}_{k=1}^{K}$. Second, we estimate the remaining parameters by regressing the estimated fixed effects on the provider characteristics in Equation (7). We explain each step in detail below.

For each family i, we observe whether the family chooses the childcare provider j, which we denote by $b_{ij} \in \{0,1\}$. Let $\mathcal{C}_j \equiv \{;\ d_{cj} \leq r\}$ denote the market to which the provider j caters, i.e., the set of census tracts within r miles of the provider. Let $\mathcal{I}_c$ denote the set of families residing in census tract c. We define the observed and predictive market shares of provider j as:

$$\text{Observed market share: } \hat{s}_j = \frac{1}{\sum_{c \in \mathcal{C}_j} \mathcal{I}_c} \sum_{c \in \mathcal{C}_j} \sum_{i \in \mathcal{I}_c} b_{icj},$$
$$\text{Predicted market share: } s_j = \frac{1}{\sum_{c \in \mathcal{C}_j} \mathcal{I}_c} \sum_{c \in \mathcal{C}_j} \sum_{i \in \mathcal{I}_c} p_{icj}, , \tag{9}$$

where p_{icj} denotes the probability that family i residing in census tract c chooses provider $j \in \mathcal{J}_c \backslash \{0\}$, under the assumption that the ϵ's are drawn from the Type I extreme value distribution:

$$p_{cj} = \frac{\exp\left(\delta_j + \alpha_0 d_{cj} + \sum_{k=1}^{K} (\alpha_k d_{cj} + \gamma_k y_c) x_{jk}\right)}{1 + \sum_{j' \in \mathcal{J}} \exp\left(\delta_{j'} + \alpha_0 d_{cj'} + \sum_{k=1}^{K} (\alpha_k d_{cj'} + \gamma_k y_c) x_{j'k}\right)}. \tag{10}$$

Note that this probability expression only depends on census tract and provider characteristics, i.e., families residing in the same census tract c are equally likely to enroll in a child care program. This is a reasonable assumption because families in the subsidized child care program will not vary much in terms of socioeconomic status and policymakers will often work with information known only at the neighborhood level. This assumption also leads to computational ease because now we are working with a vector of conditional choice probabilities of the size of the number of census tracts rather than the number of families. In the Online Appendix D, we consider a wider array of individual characteristics, such as a race and income, in the model specification.

Let $\theta_1 \equiv \left(\alpha_0, \{\alpha_k\}_{k=1}^{K}, \gamma_0, \{\gamma_k\}_{k=1}^{K}\right)$ and $\theta_2 \equiv \left(\beta_0, \{\beta_k\}_{k=1}^{K}\right)$, where the latter set of parameters is subsumed into the fixed effects. The estimation consists of two steps. The first step nests an inner loop contraction mapping in δ into the parameter search for θ_1, and the second step recovers θ_2 from the estimated $\delta(\theta_1)$. In the first step, for any candidate values of θ_1 and an initial guess for the vector of fixed effects $\delta^{(1)}$, we construct the predicted market shares $\forall j \in \{1, \ldots, J\}$:

$$s_j\left(\theta_1, \delta^{(1)}\right) = \frac{1}{\sum_{c \in \mathcal{C}_j} |\mathcal{I}_c|} |\mathcal{I}_c| \sum_{c \in \mathcal{C}_j} p_{cj}\left(\theta_1, \delta^{(1)}\right). \tag{11}$$

We then solve for:

$$\delta_j^{(h+1)} = \delta_j^{(h)} + \log(\hat{s}_j) - \log\left[s_j\left(\theta_1, \delta^{(h)}\right)\right], \tag{12}$$

for $h = \{1, 2, \ldots\}$ and stop only when $\left\|\delta_j^{h+1} - \delta_j^{h}\right\|$ is small enough. We then construct the log-likelihood function values at $\left(\theta_1, \hat{\delta}(\theta_1)\right)$:

$$\mathcal{L}_{icj}\left(\theta_1, \hat{\delta}\left(\theta_1\right)\right) = b_{icj} \log\left[p_{cj}\left(\theta_1, \hat{\delta}\left(\theta_1\right)\right)\right] + \left(1 - b_{icj}\right) \log\left[1 - p_{cj}\left(\theta_1, \hat{\delta}\left(\theta_1\right)\right)\right], \tag{13}$$

and the parameters θ_1 are estimated by maximizing the sum of the log-likelihoods:

$$\hat{\theta}_1 = \arg\max_{\theta_1} \sum_{j \in \mathcal{J}_i} \sum_{c \in \mathcal{C}_j} \sum_{i \in \mathcal{I}_c} \mathcal{L}_{icj}\left(\theta_1, \hat{\delta}\left(\theta_1\right)\right). \tag{14}$$

The second step of the estimation involves regressing the estimated fixed effects $\hat{\delta}(\hat{\theta}_1)$ on provider characteristics $\{x_{jk}\}_{k=1}^{K}$ to recover θ_2. The empirical challenge at this step is that one of the variables in $\boldsymbol{x}_j$, the provider's TRS ratings, may correlate with some unobserved characteristics of the provider (see discussion above). To this end, we recover θ_2 using a 2SLS regression, which we explain in detail next.

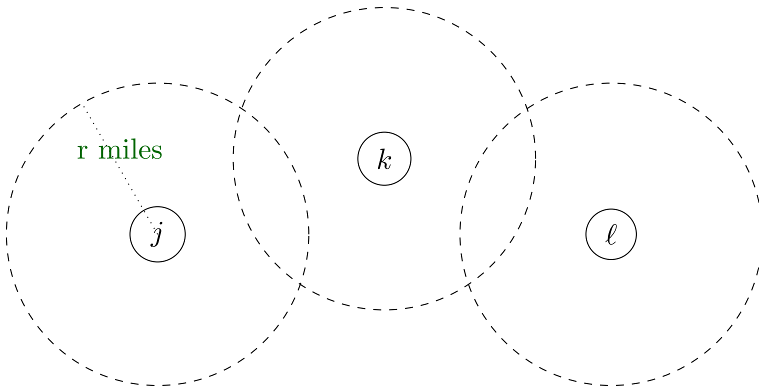
We leverage the overlapping nature of childcare markets to develop a set of instruments to address the endogeneity in TRS ratings. We construct instruments that assess the degree of differentiation between the provider and its direct competitors and indirect (or excluded) competitors and their respective markets (Gandhi and Houde, 2019). We argue that the exogenous characteristics and market conditions of the indirect competitors can only impact the quality of the provider indirectly through the quality of the direct competitors (Fan, 2013).

Figure 2 illustrates how we construct these instruments. First, we define the market of the provider j as the set of census tract within an r-mile radius of the provider's location, denoted as $\mathcal{C}_j \equiv \{i : d_{cj} \leq r\}$. The direct competitors of provider j are defined as the set of providers that overlap some with its market, denoted $\mathcal{D}_j \equiv \{k : \mathcal{C}_j \cap \mathcal{C}_k \neq \emptyset\}$. On the other hand, the indirect competitors of the provider j are the providers that do not overlap with its market but do overlap with the market of its direct competitors, denoted $\mathcal{E}_j \equiv \{\ell : \ell \in \mathcal{D}_k \setminus \mathcal{D}_j \text{ and } k \in \mathcal{D}_j\}$.

Next, we construct several instruments based on the total count, variations in providers' exogenous characteristics, and variations in market demographics. These instruments capture the differences in exogenous characteristics and market conditions between the provider and its direct and indirect competitors. We also construct the squares of these differences and the interactions among these differences. We incorporate the Gaussian transformation of the distance as weights, i.e., competitors and indirect competitors located closer to the provider are assigned higher weights in the construction of these instruments. The details of these instruments are summarized in the Online Appendix B.

It is important to note that since the endogenous variable in our setting is categorical, representing whether the provider is TRS 2-Star, 3-Star, or 4-Star, we follow the two-step instrumental variable (IV) method proposed by Wooldridge (2010). We discuss this step in more detail in the Online Appendix C.

Figure 2. Overlapping Markets and Construction of Instruments



4.2 Social Planner's Preferences

We now discuss how we recover the values of τ using supply-side interventions. Consider a local government agency with $|\mathcal{C}|$ census tracts under its jurisdiction. It is considering opening K new child care programs.[5] As we now demonstrate, these interventions allow us to measure the amount of weight policymakers assign to SVI. However, to do so, we need to change the social planner's problem to account for the fact that these interventions used existing childcare desert maps, which ignore cannibalization and the demand function for childcare services.

Ignoring cannibalization effects means that the term $\sum_{\ell \in \mathcal{T}_c} \Delta p_{ic\ell}(\mathbf{a}; \theta)$ in Equation (1) is identically zero. The fact that childcare desert maps ignore demand implies that household i in census tract c is equally likely to enroll in any provider $k \in \mathcal{J}_c$ such that $p_{ick} = \frac{1}{|\mathcal{J}_c|}$. Hence, the demand in census tract c for child care program that opens in census tract k would then be:

$$\sum_{i \in \mathcal{I}_c} \left[p_{ick}(\mathbf{a}; \theta) - \sum_{\ell \in \mathcal{T}_c} \Delta p_{ic\ell}(\mathbf{a}; \theta) \right] = \sum_{i \in \mathcal{I}_c} \frac{1}{|\mathcal{J}_c|} = \frac{|\mathcal{I}_c|}{|\mathcal{J}_c|} \tag{15}$$

Policymakers typically have relied on only measuring the demand for child care within the proposed neighborhood. Therefore, given that local government agencies had access to childcare desert maps that are currently available, we argue that the locations of the new facilities were chosen by solving

[5] The exercise also applies to choosing locations for contracted slots. To fix on the critical ideas, our presentation focuses on the intervention that expands childcare infrastructure in a specific county, but this choice is without loss of generality.

the following (demand) model-free social planner problem:

$$\begin{aligned}
\mathbf{a}^* = \{\mathbf{a}^*_{\mathbf{k}}\}_{\mathbf{k}\in\mathcal{C}} &= \underset{\mathbf{a}}{\operatorname{argmax}} \sum_{\mathbf{k}\in\mathcal{C}} \mathbf{a}_{\mathbf{k}} \mathbf{w}\left(\mathbf{y}_{\mathbf{k}};\tau\right) \frac{\mid \mathcal{I}_{\mathbf{k}} \mid}{\mid \mathcal{J}_{\mathbf{k}} \mid} \\
\text{subject to } a_k \in \{0,1\};\ &\sum_{k\in\mathcal{C}} a_k = K.
\end{aligned} \tag{16}$$

Hence, in this formulation, the policymaker aims to identify the top K census tracts with the highest SVI-adjusted excess demand. Our goal is to estimate τ. To achieve this, we match the locations predicted by Equation 16 with the actual locations selected by the local body to open new high-quality childcare programs. We calculate the probability that a census tract receives one of the K new childcare programs based on identifying the top K locations with the highest SVI-adjusted excess demand. To facilitate this process, we construct continuous and differentiable approximations for the predicted assignment (Jang, Gu, and Poole, 2016).

Here, we explain how we derive the probability that census tract k is chosen to open one of the K new childcare programs per the predicted distribution based on Equation (16). Given τ, consider the problem of opening only one program. We define a shorthand representation of the SVI-adjusted excess demand in census tract k as: $e_k(\tau) = w_k\left(y_k;\tau\right)\frac{|\mathcal{I}_k|}{|\mathcal{J}_k|}$. The probability that a new program opens in census tract k is then formulated as:

$$q_k = \frac{\exp\left(\frac{e_k(\tau)}{\kappa}\right)}{\sum_{k'\in\mathcal{C}} \exp\left(\frac{e'_k(\tau)}{\kappa}\right)} \tag{17}$$

which corresponds to a higher probability for a census tract with a higher SVI-adjusted excess demand for childcare. Here, a lower value for $\kappa > 0$ brings the vector q closer to a binary vector, i.e., the social planner problem becomes less noisy and is mostly based on the SVI-adjusted excess demand.

Now, consider the problem of opening two programs. Initialize $\chi_k^{(1)} = e_k(\tau)$ and $\psi_k^{(1)} = 0$ for all $k \in \mathcal{C}$. The probability that a new program opens in census tract k in the first round is given by:

$$q_k^{(1)} = \frac{\exp\left(\frac{\chi_k^{(1)}}{\kappa}\right)}{\sum_{k'\in\mathcal{C}} \exp\left(\frac{\chi_{k'}^{(1)}}{\kappa}\right)}$$

This expression is equivalent to the probability of opening only one program, as derived in Equation (17). Now, to derive the probability of opening the second program, update $\psi_k^{(2)} = \psi_k^{(1)} + q_k^{(1)}$ and

$\chi_k^{(2)} = \chi_k^{(1)} + \log(1-\psi_k^{(2)}) - \max_{k'}\{\chi_{k'}^{(1)} + \log(1-\psi_{k'}^{(2)})\}$ for all $k \in \mathcal{C}$. The probability that a new program opens in census tract k in the second round is given by:

$$q_k^{(2)} = \frac{\exp\left(\frac{\chi_k^{(2)}}{\kappa}\right)}{\sum_{k'\in\mathcal{C}} \exp\left(\frac{\chi_k^{(2)}}{\kappa}\right)} \tag{18}$$

We can interpret $1-\psi_k$ as the weight assigned to census tract k in deriving these probabilities. When we are searching for the top-1 location, each census tract gets assigned the same weight, $1-\psi_k^{(1)} = 1$. Suppose that the probability of a new program opening in the first round is the highest for census tract ℓ_1, i.e., $q_{\ell_1}^{(1)} \sim 1$, then $\psi_{\ell_1}^{(2)} = 1$ and $\chi_{\ell_1}^{(2)} = \chi_{\ell_1}^{(1)} - \max_{c'} \chi_{c'}^{(1)}$. Note that $\chi_{\ell_1}^{(2)} = 0$ because $\max_{k'} \chi_{k'}^{(1)} = \chi_{\ell_1}^{(1)}$, implying $q_{\ell_1}^{(2)}$ will be the lowest. Therefore, census tract ℓ_1 is will be chosen in the second round with a very small probability given that it was chosen in the first round, which is ensured by updating the weight assigned to census tract ℓ_1 in the second round to zero, i.e., $\psi_{\ell_1}^{(2)} = 0$.

In this way, we can solve the problem of opening K programs by recursively solving for $h \in \{1, \ldots, K\}$:

$$q_c^{(h)} = \frac{\exp\left(\frac{\chi_k^{(h)}}{\kappa}\right)}{\sum_{k'\in\mathcal{C}} \exp\left(\frac{\chi_k^{(h)}}{\kappa}\right)} \text{ subject to} \tag{19}$$

$$\chi_k^{(h)} = \begin{cases} e_k(\tau) & \text{if } h = 1 \\ \chi_k^{(h-1)} + \log(1-\psi_k^{(h)}) - \max_{k'}\{\chi_{k'}^{(h-1)} + \log(1-\psi_{k'}^{(h)})\} & \text{if } h > 1 \end{cases} \tag{19a}$$

$$\text{and } \psi_k^{(h)} = \begin{cases} 0 & \text{if } h = 1 \\ \psi_k^{(h-1)} + q_k^{(h-1)} & \text{if } h > 1 \end{cases} \tag{19b}$$

Finally, let $q_k = \sum_h q_k^{(h)}$ denote the probability that one of the K programs opens in census tract c. Note that this is equivalent to $\psi_k^{(K)}$. We estimate τ by solving for maximum likelihood estimation:

$$\tau^*(\kappa) \equiv \arg\max_{\tau} \left\{ \log \left[\prod_k q_k(\tau)^{\tilde{q}_k} \left(1 - q_k(\tau)\right)^{(1-\tilde{q}_k)} \right] \right\} \tag{20}$$

where $\tilde{q}_k \in \{0, 1\}$ denotes the true assignment i.e., whether a program opened in census tract k.

5 Results and Applications

5.1 Demand for Childcare Services Parameter Estimates

This section presents the estimation results of the demand model and the social planner's preferences outlined in Section 3. Table 5 reports the results of three specifications. Column (1) reports the results of the specification that does not account for unobserved quality or the endogeneity in TRS ratings. The specification in column (2) accounts for unobserved quality, but not for the endogeneity in TRS ratings. The specification in column (3) accounts for unobserved quality as well as for the endogeneity in TRS ratings. All specifications include LWDB fixed effects.

We find several notable patterns in household preferences regarding the choice of childcare. Primarily, families exhibit a preference for geographical proximity in their childcare choices. Additionally, families tend to favor center-based childcare programs over home-based ones, opt for programs owned by individuals or private organizations as opposed to programs owned by the government or a non-profit organization, and show a preference for programs with a TRS rating. Although families value proximity, we find that they are willing to extend their travel distances for childcare programs with a TRS 4-star rating.

These patterns persist across all three specifications. However, upon accounting for unobserved quality, we observe that the marginal utilities derived from center-based, individual, or privately-owned childcare providers remain positive but decrease. Notably, when adjusting for unobserved quality differences among providers, higher SVI households exhibit lower marginal utility from enrolling in TRS 4-rated programs. This suggests that households facing higher SVI may prioritize aspects of quality not fully captured by TRS ratings. It is plausible that these households encounter informational barriers that hinder their understanding of childcare quality as assessed by TRS ratings.

For instance, Forry, Isner, Daneri, and Tout (2014) survey a subset of families participating in the Minnesota Child Care Choices study, finding that parents with lower education and income invest less time in selecting childcare providers. They also note that these parents, who allocate less time to their search, tend to prioritize factors of convenience of access. Other studies support this finding, indicating that low-income families often consider fewer childcare options (Anderson, Ramsburg, and Scott, 2005) and spend less time on the selection process (Layzer, Goodson, and Brown-Lyons, 2007). Moreover, several studies consistently show that cost considerations are negatively correlated with family income, maternal employment status, and educational attain-

Table 5. Estimation results

	(1)	(2)	(3)
Constant	0.18	0.19	-0.96
	(0.15)	(0.28)	(0.34)
Distance to Childcare Program	-0.55	-0.58	-0.58
	(0.00)	(0.00)	(0.00)
Program Type: Center	1.97	1.69	1.46
	(0.02)	(0.04)	(0.10)
Program Owner: Individual	0.26	0.14	0.88
	(0.01)	(0.06)	(0.13)
Program Owner: Pvt Org	0.46	0.34	0.78
	(0.01)	(0.06)	(0.09)
Program TRS Rating: 2 or 3 Star	0.51	0.56	0.28
	(0.03)	(0.05)	(0.58)
Program TRS Rating: 4 Star	0.73	0.78	4.36
	(0.02)	(0.04)	(0.60)
TRS 2 or 3 Star $\times$ Distance	-0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)
TRS 4 Star $\times$ Distance	0.02	0.02	0.02
	(0.00)	(0.00)	(0.00)
TRS 2 or 3 Star $\times$ SVI	0.10	-0.06	-0.06
	(0.03)	(0.04)	(0.04)
TRS 4 Star $\times$ SVI	0.03	-0.22	-0.22
	(0.03)	(0.04)	(0.04)
Program-specific Unobserved Quality	No	Yes	Yes
Instrument for Observed Quality	No	No	Yes

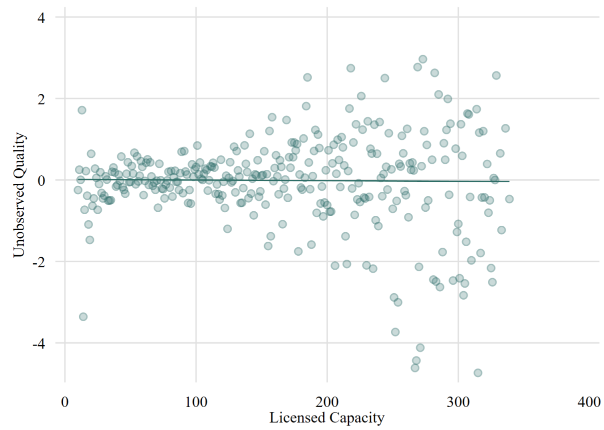
Note: Table presents the two-step nested fixed point MLE estimates of the preference parameters from Equation 8. Column (1) does not account for the unobserved heterogeneity across childcare programs. Column (2) accounts for this heterogeneity but does not account for the endogeneity in providers' TRS (quality) ratings. Column (3) accounts for both. Standard errors are reported in parentheses. The first-stage regression results for the instruments are provided in the Online Appendix C. All specifications include LWDB fixed effects.

ment (Leslie, Ettenson, and Cumsille, 2000; Peyton, Jacobs, O'Brien, and Roy, 2001; Early and Burchinal, 2001; Kensinger Rose and Elicker, 2008; Kim and Fram, 2009).

Accounting for the endogeneity in TRS ratings, we note that the marginal utility from center-based programs is still positive, but lower, while that from individual or private-organization ownership is higher. The marginal utility from enrolling in TRS 2- or 3-star programs is positive but statistically not significant, while that from TRS 4 Star programs is much higher now.

For a more accessible interpretation of the estimation findings, we turn our attention to the marginal effects, revealing four key insights – households value proximity, they exhibit a distinct preference for TRS 4-star ratings, they are willing to trade proximity for TRS 4-star rating, and higher SVI households have a weaker preference for TRS 4-star rating. First, consider a household with SVI 1, which is presented with two alternatives – both privately owned, center-based programs with no TRS rating. One is situated at 4 miles, while the other is at 5 miles. Given the model

Figure 3. Correlation Between Program-specific Unobserved Quality and Licensed Capacity



Note: Figure showcases the relationship between these residuals and the licensed capacity of the childcare program.

estimates, the household is 9 percentage points more likely to choose the 4-mile program over the 5-mile one.

Second, consider a household with an SVI of 1 facing three alternatives: all privately-owned, center-based programs located within a 5-mile radius. These options vary in quality ratings – no TRS rating, TRS 2 or 3 stars, and TRS 4 stars. The household exhibits a 90 percentage point higher likelihood of enrolling in the TRS 4-star program compared to the program without a TRS rating. Similarly, they are 89 percentage points more likely to enroll in the TRS 4-star program compared to the program rated TRS 2 or 3 stars.

Third, consider the same household facing a choice between a privately-owned, center-based program with no TRS rating, situated at 4 miles, and a privately-owned and center-based program with a TRS 4-star rating, located at 5 miles. The household demonstrates an equal likelihood of choosing either program. These results indicate that the household is willing to drive up to an extra mile for a TRS 4-star program.

Finally, in a comparison of two households, one with SVI 1 and the other with SVI 0, both are faced with two alternatives—privately-owned, center-based programs located at 5 miles. The first is unrated, while the second has a TRS 4-star rating. The SVI 1 household is 92 percentage points more likely to choose the TRS 4-star program. Conversely, the SVI 0 household is 93 percentage points more likely to choose it over the no-rating program.

Figure 3 explores how the unobserved quality of a childcare program is related to its licensed capacity. They do not appear to be correlated, suggesting that the unobserved quality is not an

Table 6. Social Planner's Preference Estimates

	Estimate	S.E.	Distance Measure
County A			
$\kappa = 1.0$	0.03	0.01	3.39
$\kappa = 0.1$	0.11	0.02	2.03
$\kappa = 0.01$	0.45	0.02	2.56
County B			
$\kappa = 1.0$	0.33	1.41	2.93
$\kappa = 0.1$	0.81	1.87	2.43
$\kappa = 0.01$	2.23	0.14	2.45

Note: This table presents the results of the maximum likelihood estimation in Equation (20) to recover τ while keeping κ fixed. The distance measure represents the square root of the sum of squared distances between the predicted and true location assignments.

artifact of capacity constraints. In table D, in the vector $\mathbf{x_i}$ we include indicators for whether the family's race is Black, whether their race is Hispanic, and their income in addition to their SVI. We also include fixed effects for the year in which the family enrolls in subsidized childcare.

5.2 Social Planner's Preference Parameter Estimates

Now, we discuss the results of the maximum likelihood estimation defined in Equation (20) to derive τ. Note the presence of another parameter, κ, in the derivation of the probability that a census tract is selected to open new programs, as defined in Equation (19). The probability distribution approaches a binary assignment vector as the value of κ decreases. We fix $\kappa = 0.01$, resulting in $\tau = 0.45$ for County A and $\tau = 2.23$ for County B, indicating that SVI is more salient in County B's location choice.

For the sake of exposition, we consider multiple values of κ, fixing each one to solve for τ, and subsequently compute the sum of squared distances between the predicted locations k^* and the true location assignments $\tilde{k}$, denoted as $\sqrt{\sum_{k^*,\tilde{k}} d^2_{k^*,\tilde{k}}}$. This measure assesses, on average, how closely we are able to match the predicted locations to the true locations.

We conduct this analysis for three values of $\kappa \in \{0.01, 0.1, 1.0\}$. For County A, we find that the minimum distance measure is achieved at $\kappa = 0.1$, resulting in an average distance of 2.03 miles and estimating $\tau = 0.11$. At $\kappa = 0.01$, the distance measure is 2.56 miles, corresponding to $\tau = 0.45$. This indicates that, on average, predicted locations are within 2.56 miles from the true locations. For County B, the distance measure was 2.43 miles at $\kappa = 0.1$, yielding $\tau = 0.81$, and 2.45 miles at $\kappa = 0.01$.

5.3 Decomposition

In this section, we perform a decomposition analysis to examine the impact of various components within the social planner problem. Our thought experiment is to choose the locations of $K = 5$ new high-quality programs and then register how our choices influence the market shares of high-quality programs, including both new entrants and existing providers. This initiative envisions the establishment of center-based 4-star programs, which will be managed by private entities while housed in government-built facilities. This model reflects the intervention in County A.

The analysis begins with the selection of five locations based on the comprehensive framework of the social planner problem; we do so for our two estimates of the social planner's preferences. This selection process is then replicated twice more, each time omitting a specific model component: first, the cannibalization factor is excluded, followed by the removal of the demand function. These steps facilitate the measurement of the impact attributable to the preferences of the social planner, the cannibalization effects, and the demand for childcare services. The purpose of employing this analytical approach is to elucidate the contributions of each factor to a more refined method to identify areas with excess demand for quality childcare.

In addition, a key aspect of this endeavor is the determination of the unobserved quality of the new program, denoted by the parameter ξ. Our assumption posits that ξ aligns with the $90th$ percentile value within its distribution for the TRS 4 Star program, a presumption that provides a balanced benchmark for quality expectations. This distribution is derived from the second stage of the two-stage least squares (2SLS) regression analysis, discussed in Section 3.

Full Social Planner's Problem: The primary objective of selecting locations for the new programs revolves around maximizing prospective enrollment figures while simultaneously striving to minimize the negative impact on market shares of existing high-quality childcare programs. Thus, the local agency solves the following social planner problem described by Equation (1). For convenience, we reproduce this equation here, subject to some modifications:

$$\mathbf{a}^* = \{\mathbf{a}_\mathbf{k}^*\}_{\mathbf{k}\in\mathcal{C}} = \arg\max_{\mathbf{a}} \sum_{k\in\mathcal{C}} a_k \sum_{c\in\mathcal{C}_k} w\left(y_c;\tau\right) \mid \mathcal{I}_c \mid \left[\int_\xi p_{ck}\left(\mathbf{a},\xi;\theta\right) dF\left(\xi\right) - \int_\xi \sum_{\ell\in\mathcal{T}_c} \Delta p_{c\ell}\left(\mathbf{a},\xi;\theta\right) dF\left(\xi\right)\right]$$
$$\text{subject to } a_k \in \{0,1\}; \sum_{k\in\mathcal{C}} a_k = K. \tag{21}$$

The term $p_{ck}\left(\mathbf{a}, \xi; \theta\right)$ denotes the predicted probability of a family residing in census tract c enrolling in the new program that opens in the census tract k given the location assignment $\mathbf{a}$, the distribution of the unobserved quality of the existing programs (ξ), and model estimates θ. The term $\sum_{\ell \in \mathcal{T}_c} \Delta p_{c\ell}\left(\mathbf{a}, \xi; \theta\right)$ denotes the change in the predicted probability of a family in census tract c enrolling in an existing TRS 4 Star program. The social planner aims to maximize the predicted market share of the new program while minimizing the loss in the market share of the existing TRS 4 Star programs, adjusted for the SVI weights, $w(y_c; \tau)$. Note that the social planner is unaware of the ξ distribution and assumes that $\xi_\ell \in N(\mu_\ell, \sigma_\ell^2)$ where μ_k and σ_ℓ are the mean and standard deviation of the distribution obtained from the 2SLS step in Section 3 for TRS quality ratings $\ell \in \{0, 2, 3, 4\}$. Additionally, we assume that the value of ξ_k for the new program that opens in the census tract k will be at the $90th$ percentile of the observed distribution of ξ's for TRS 4 Star programs.

We solve the social planner problem for $K = 5$ programs. We solve for this problem sequentially, i.e., we start with solving for the optimal location for selecting a location for one program. Then, we solve for the optimal location of the second program, and so on.

Next, we apply our framework to identify areas for improvement in the quality ratings of existing ECE programs. In this alternative scenario, the local government agency is tasked with identifying an existing childcare program eligible for technical support to upgrade to a TRS 4-star rating. The primary goal of this initiative is to enhance the quality of the program, thus maximizing prospective enrollments without significantly affecting the market share of existing high-quality childcare programs. This approach aims to foster a childcare environment that improves the overall quality of the service without undermining the sustainability of established providers. The local government's objective function is:

$$\mathbf{a}^* = \{\mathbf{a_k^*}\}_{\mathbf{k} \in \mathcal{J}} = \arg\max_{\mathbf{a}} \sum_{j \in \mathcal{J}} a_j \sum_{c \in \mathcal{C}_j} w\left(y_c; \tau\right) \mid \mathcal{I}_c \mid \left[\int_\xi p_{ck}\left(\mathbf{a}, \xi; \theta\right) dF\left(\xi\right) - \int_\xi \sum_{\ell \in \mathcal{T}_c} \Delta p_{c\ell}\left(\mathbf{a}, \xi; \theta\right) dF\left(\xi\right) \right]$$
$$\text{subject to } a_k \in \{0,1\}; \sum_{k \in \mathcal{J}} a_k = K. \tag{22}$$

i.e., the local government agency solves for an assignment vector $\mathbf{a} \equiv \{\mathbf{a_k}\}_{\mathbf{k} \in \mathcal{J}}$ where $\mathcal{J}$ denotes the set of childcare programs located in the local government agency's jurisdiction and $a_k \in \{0, 1\}$ denotes whether program k's rating is upgraded to TRS 4-star.

Turning Cannibalization Off: Next, we study how the market shares of high-quality programs change when we choose K locations without considering the potential for cannibalization of the demand for the services of incumbent firms in the market. To do this, we maximize the following objective function:

$$\begin{aligned} \mathbf{a}^* = \{\mathbf{a}_\mathbf{k}^*\}_{\mathbf{k}\in\mathcal{C}} = \arg\max_{\mathbf{a}} \sum_{k\in\mathcal{C}} a_k \sum_{c\in\mathcal{I}_{\|}} w_c\left(y_c;\tau\right) \mid \mathcal{I}_c \mid \int_\xi p_{ck}\left(\mathbf{a},\xi;\theta\right) dF(\xi) \\ \text{subject to } a_c \in \{0,1\}; \sum_{k\in\mathcal{C}} a_k = K. \end{aligned} \tag{23}$$

$$\begin{aligned} \mathbf{a}^* = \{\mathbf{a}_\mathbf{k}^*\}_{\mathbf{k}\in\mathcal{J}} = \arg\max_{\mathbf{a}} \sum_{j\in\mathcal{J}} a_j \sum_{c\in\mathcal{C}_j} w\left(y_c;\tau\right) \mid \mathcal{I}_c \mid \int_\xi p_{ck}\left(\mathbf{a},\xi;\theta\right) dF\left(\xi\right) \\ \text{subject to } a_k \in \{0,1\}; \sum_{k\in\mathcal{J}} a_k = K. \end{aligned} \tag{24}$$

Turning Demand Function Off: Next, we assume that the demand function is uniform across providers in the market. As we derived in Section 4.2, this leads to the (demand) model-free version of the problem of the social planner of building new high-quality programs:

$$\begin{aligned} \mathbf{a}^* = \{\mathbf{a}_\mathbf{k}^*\}_{\mathbf{k}\in\mathcal{C}} = \arg\max_{\mathbf{a}} \sum_{\mathbf{c}\in\mathcal{C}} \mathbf{a}_\mathbf{k}\mathbf{w}\left(\mathbf{y}_\mathbf{k};\tau\right) \frac{\mid \mathcal{I}_\mathbf{k} \mid}{\mid \mathcal{J}_\mathbf{k} \mid} \\ \text{subject to } a_k \in \{0,1\};\ \sum_{k\in\mathcal{C}} a_k = K. \end{aligned} \tag{25}$$

Similarly, we explore a model-free version of the problem of the social planner of upgrading the quality of existing programs:

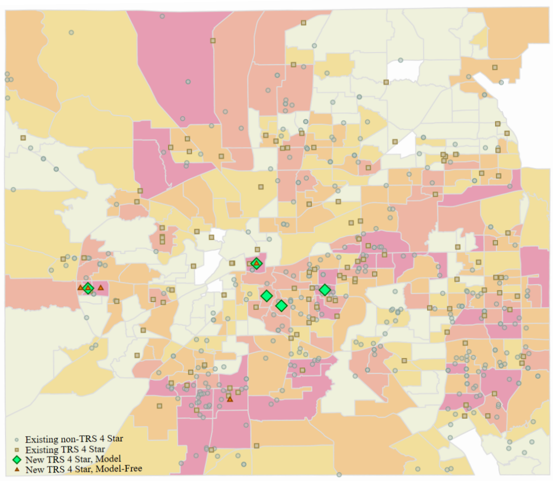
$$\begin{aligned} \mathbf{a}^* = \{\mathbf{a}_\mathbf{k}^*\}_{\mathbf{k}\in\mathcal{J}} = \arg\max_{\mathbf{a}} \sum_{k\in\mathcal{J}} a_k \sum_{c\in\mathcal{C}_k} w\left(y_c;\tau\right) \frac{\mid \mathcal{I}_c \mid}{L_k} \\ \text{subject to } a_j \in \{0,1\}; \sum_{j\in\mathcal{J}} a_j = K. \end{aligned} \tag{26}$$

where L_k denotes the number of licensed seats at program k. We construct a measure of excess demand for this program: number of families located around the program divided by the number of licensed seats at the program. A program with a higher excess demand is more likely to be picked. Families are weighted by the SVI of their census tract of residence.

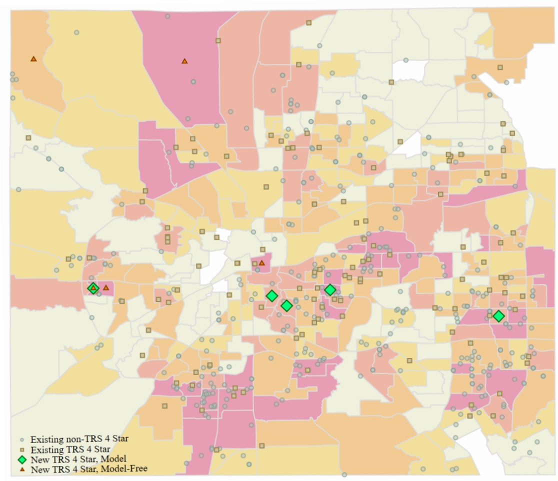
Decomposition Results: The sites for new programs, as pinpointed by the above methods are illustrated in Figure 4. The four versions of the social planner's problem (full model vs model-free approaches for building new high-quality programs and upgrading the quality of existing programs) choose different locations for the new facilities.

Figure 4. ECE Expansion: Social Planner Problem vs Model-Free Approaches of Building New High Quality Programs and Upgrading the Quality of Existing Programs

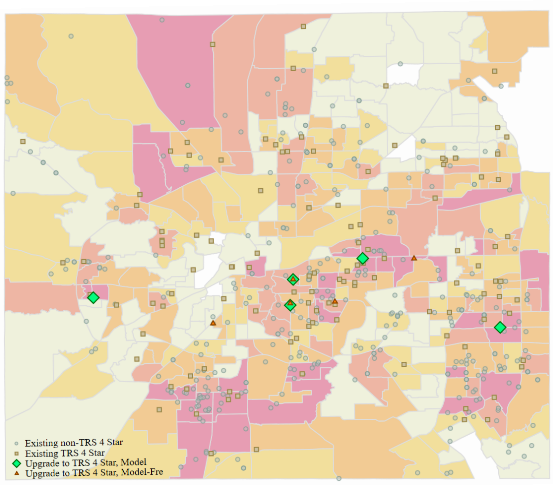
(a) Build New, $\tau = 0.45$



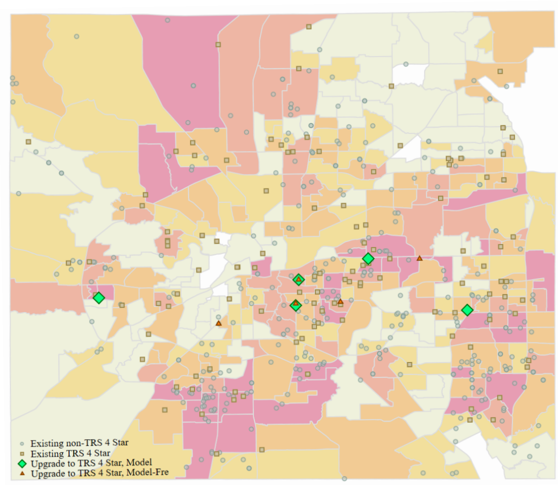
(b) Build New, $\tau = 2.23$



(c) Upgrade Existing, $\tau = 0.45$



(d) Upgrade Existing, $\tau = 2.23$



Note: Figure (a) illustrates the sites recommended by solving for the social planner problem of building new high-quality programs in Equation (21) vs the model-free approach in Equation (25) for $\tau = 0.45$. Figure (b) shows these sites for $= 2.23$. Figures (c) and (d) compare the sites recommended by social planner problem of upgrading the quality of existing programs in Equation (22) vs the model-free approach in Equation (26). A darker shade of the polygon denotes a higher number of families seeking subsidized child care.

Figure 5 (a) examines the market share impact of introducing five new TRS 4-star programs sites recommended by solving for the social planner problem of building new high-quality programs in Equation (21) vs the model-free approach in Equation (25) for $\tau = 0.45$.

Initially, high-quality programs collectively command a market share of around 46.2% before any new introductions. By solving for the social planner problem of selecting locations for new programs in Equation (21), the overall market share for high-quality programs rises to about 49.6%. This reflects an increase of 7.4%. When opening five new programs based on locations with the highest excess demand (the model-free approach in Equation (25)), the market share of TRS 4-star programs increases by about 6.2%. At, $\tau = 2.23$, the corresponding increases are 7.4% and 5.5%, as shown in Figure 5 (b). Overall, at a larger τ, i.e., when the priority weights are less sensitive to SVI, the improvement in predicted market share of high-quality programs under the demand model approach over the model-free approach is more prominent.

Figure **??** (a) and (b) examine the market share impact of five sites recommended by solving for the social planner problem of upgrading the quality of existing programs in Equation (22) vs the model-free approach in Equation (26). The increase in the market share of TRS 4 Star programs is predicted to be 9.4% in the demand model approach vs 4.4% under the model-free approach. Similar numbers under both values of τ.

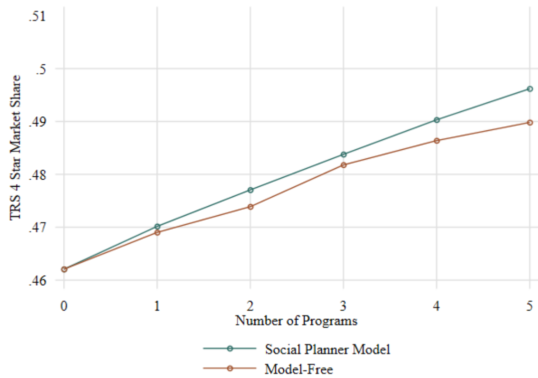
Table 7 compares cumulative market shares of the new programs, average distances traveled to the new programs, and average SVI compositions at the new programs for the demand model and model-free approaches of opening new high-quality programs.

The demand model approach selects sites that are expected to have slightly higher market shares compared to the sites selected under the model-free approach. The former are expected to have families travel a longer distance to avail child care. The former are expected to have higher SVI families enroll with them. The improvement in the market share of new programs under the demand model approach over the model-free approach is expected to be higher for higher τ. The expected enrollment of high-SVI families is lower at higher τ but still higher than that under the model-free approach.

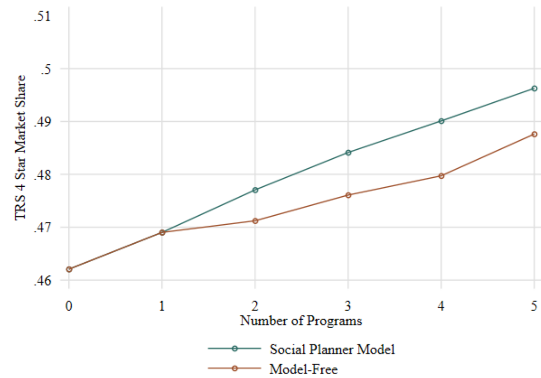
As we see in Table 8, the expected market share of programs whose quality is upgraded to TRS 4 Star is much higher under the demand model approach compared to the model-free approach, 6.7% vs 3.2%. The distance that families are expected to travel to enroll at the new programs is also lower under the model approach and so is the expected SVI composition.

Figure 5. Market Share of High-Quality Childcare Programs

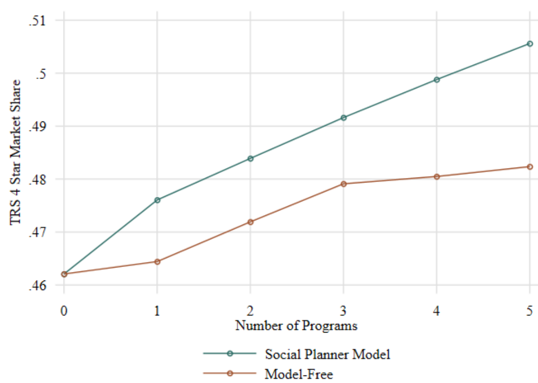
(a) Build New, $\tau = 0.45$



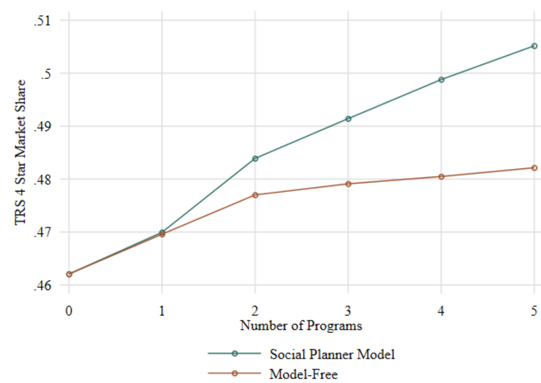
(b) Build New, $\tau = 2.23$



(c) Upgrade Existing, $\tau = 0.45$



(d) Upgrade Existing, $\tau = 2.23$



Note: These figures illustrate the predicted market share of TRS 4-star childcare programs based on the model estimates in Section 5 when expanding the supply of high-quality childcare, grouped by considering three different approaches: solving the social planner problem to open new programs in Equation (21), solving the social planner problem to open new programs in Equation (22), and (c) counting the SVI-weighted excess ratios in Equation (25), while setting $\tau = \{0.45, 2.23\}$ in Equation (2)

Table 7. Market Share, Distance and SVI composition after High-Quality ECE Expansion

	(1)	(2)	(3)	(4)	(5)	(6)
	$\tau = 0.45$			$\tau = 2.23$		
K	Market Share	Distance	SVI	Market Share	Distance	SVI
Panel A: Social Planner Model						
1	0.013	2.809	0.847	0.014	1.639	0.684
2	0.027	2.198	0.765	0.027	2.198	0.765
3	0.039	2.445	0.785	0.040	2.398	0.779
4	0.050	2.544	0.790	0.051	2.444	0.760
5	0.060	2.560	0.797	0.061	2.555	0.772
Panel B: Model-Free						
1	0.014	1.639	0.684	0.014	1.639	0.684
2	0.024	1.870	0.673	0.017	1.627	0.617
3	0.038	2.092	0.737	0.027	1.841	0.632
4	0.047	2.206	0.718	0.032	1.809	0.558
5	0.053	2.222	0.706	0.046	2.009	0.645

Note: This table compares cumulative market shares, average distances traveled, and average SVI compositions for different policies under two values of $\tau \in \{0.45, 2.23\}$. K denotes the number of new programs.

Table 8. Market Share, Distance and SVI composition after High-Quality ECE Expansion

	(1)	(2)	(3)	(4)	(5)	(6)
		$\tau = 0.45$			$\tau = 2.23$	
K	Market Share	Distance	SVI	Market Share	Distance	SVI
Panel A: Social Planner Model						
1	0.015	2.964	0.840	0.014	1.866	0.677
2	0.029	2.437	0.763	0.029	2.437	0.763
3	0.042	2.610	0.786	0.043	2.577	0.747
4	0.056	2.666	0.768	0.056	2.666	0.768
5	0.067	2.649	0.753	0.066	2.711	0.750
Panel B: Model-Free						
1	0.004	2.918	0.809	0.012	3.309	0.642
2	0.015	3.212	0.681	0.024	3.126	0.730
3	0.027	3.114	0.739	0.027	3.114	0.739
4	0.029	3.130	0.737	0.029	3.130	0.737
5	0.032	3.124	0.743	0.032	3.076	0.733

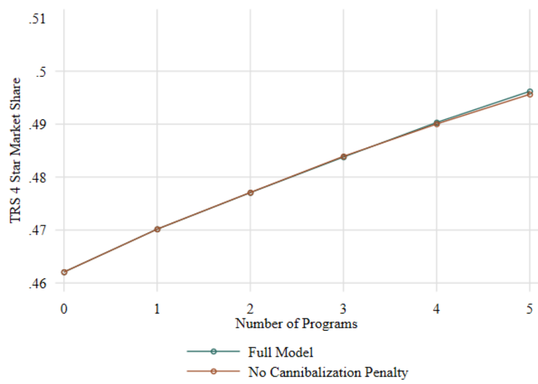
Note: This table compares cumulative market shares, average distances traveled, and average SVI compositions for different policies under two values of $\tau \in \{0.45, 2.23\}$. K denotes the number of upgraded programs.

Figure 6 (a) presents the expected market share of TRS 4 Star programs under the full model in Equation (21) vs when the cannibalization penalty is turned off in Equation (23). There is not much difference at $\tau = 0.45$ but as we in Figure 6 (b), third program onward, the market share of TRS 4 Star programs is expected to be slightly higher when the full model is considered i.e., ignoring the cannibalization penalty would ignore to an overall lower enrollment at TRS 4 Star programs.

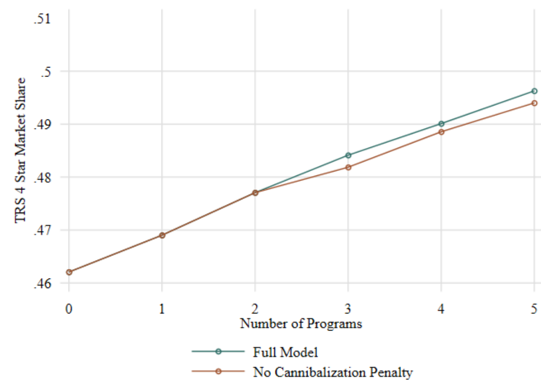
Figure 6 (c) and (d) present the expected market share of TRS 4 Star programs under the full model in Equation (22) that solves for the selection of existing programs for quality upgrading vs when the cannibalization penalty is turned off. The expected market shares are marketdly higher when full model is considered.

Figure 6. Cannibalization

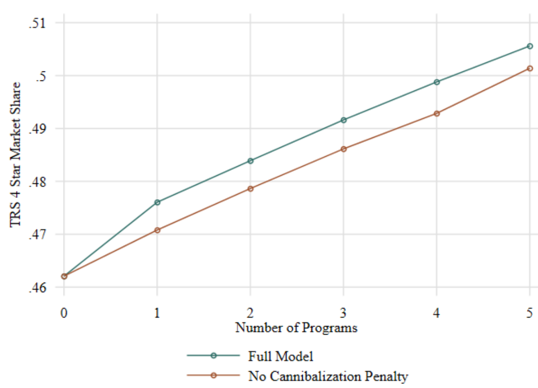
(a) Build New, $\tau = 0.45$



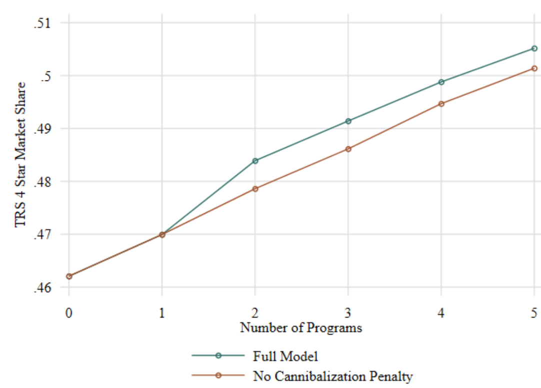
(b) Build New, $\tau = 2.23$



(c) Upgrade Existing, $\tau = 0.45$



(d) Upgrade Existing, $\tau = 2.23$



Note: These figures illustrate the predicted market share of TRS 4-star childcare programs based on the model estimates in Section 5 when expanding the supply of high-quality childcare, grouped by considering four different approaches: solving the social planner problem to open new programs in Equation (21) vs without the cannibalization penalty in Equation (23), solving the social planner problem to upgrade existing programs in Equation (22) vs without the cannibalization penalty, while setting $\tau = \{0.45, 2.23\}$ in Equation (2)

Supply-side vs Demand-side Policy Interventions: Using model estimates, we can dissect the observed variations in TRS 4-star enrollment between low-SVI and high-SVI groups, attributing them to underlying supply-side factors (availability of TRS 4-star programs) and demand-side factors (preferences for TRS 4-star rating). In this analysis, we nullify the interaction term between TRS rating and SVI, replacing the estimated negative values with zero. At the baseline, 47% of high-SVI families are expected to enroll in a TRS 4-star program. By setting the TRS-SVI preference parameters to zero, we observe an increase in TRS 4-star enrollment among high-SVI families to 50%. To contextualize this outcome against supply-side interventions, the opening of a four TRS

4-star program would suffice to achieve a 51% TRS 4-star enrollment among high-SVI families.

6 Conclusion

High-quality childcare provision is essential for both women's labor force participation and children's development. Given the importance of childcare, scholars and policymakers are concerned about spatial inequalities in the distribution of childcare programs, particularly the prevalence of "childcare deserts" in socioeconomically vulnerable areas. These deserts lack accessible and affordable childcare options, limiting opportunities for families. Researchers use datasets detailing childcare programs and population estimates to identify these deserts, focusing on factors such as childcare seat-to-child ratios.

This paper innovates over existing approaches to identifying childcare deserts in three significant ways. It formulates a social planner problem to optimize the expansion of childcare provision, either by establishing new programs or upgrading existing ones. This model redefines a childcare desert as an area with substantial potential for new enrollments while minimizing adverse impacts on existing programs. The model requires inputs on household demand and the social planner's preferences for maximizing market shares and expanding supply in vulnerable areas. These preferences are estimated by relying on direct interventions in the childcare market within two Texas counties.

The methodology leverages a state-owned dataset that matches families to childcare businesses, identifying overlapping geographical regions to define childcare markets. This approach utilizes rich administrative data from the subsidized childcare program in Texas, matching families to childcare businesses. The estimation of households' demand for childcare considers trade-offs between provider characteristics such as quality ratings and proximity. To account for unobserved quality, the model incorporates program-specific fixed effects, and a contraction mapping is employed to recover these fixed effects efficiently. Instrumental variables are constructed to address the potential endogeneity of quality ratings, relying on the unique definition of childcare markets as geographically overlapping regions. The analysis confirms that proximity is a key factor in parental choice, but parents are willing to travel further for higher-quality programs, indicating a strong preference for quality over convenience.

Lastly, we compare the proposed methods with benchmark methods currently in use for a county in Texas. The comparison focuses on capturing market shares for the new programs, the average distance families will have to travel to the new programs, and the average SVI of the families expected to be enrolled in the new programs. Our findings indicate that the proposed methods are superior. They achieve similar goals as the current methods with lower deployment and are better at targeting more socioeconomically vulnerable communities. We also explore the importance of adjusting the priority weights assigned to SVI and the significance of penalizing the cannibalization of the market shares of incumbent high-quality programs.

This study highlights the salience of quality ratings in parents' childcare choices. Future research involves studying the outcomes for children, such as their educational progress from enrolling in high-quality childcare, understanding how parents' perceptions of quality differ from official quality ratings, and exploring what motivates childcare programs to participate in the subsidized childcare program. These areas of inquiry will provide further insights into optimizing childcare provision to meet the needs of all families effectively.

References

Allcott, H., R. Diamond, J.-P. Dubé, J. Handbury, I. Rahkovsky, and M. Schnell (2019). Food deserts and the causes of nutritional inequality. *The Quarterly Journal of Economics 134*(4), 1793–1844.

Anderson, S. G., D. M. Ramsburg, and J. Scott (2005). Illinois study of license-exempt child care. *Department of Health and Human Services*.

Baker, M., J. Gruber, and K. Milligan (2019). The long-run impacts of a universal child care program. *American Economic Journal: Economic Policy 11*(3), 1–26.

Bassok, D., M. Fitzpatrick, and S. Loeb (2011). Disparities in child care availability across communities: Differential reflection of targeted interventions and local demand. *Center for Education Policy Analysis*, 1–13.

Bassok, D., M. Fitzpatrick, and S. Loeb (2014). Does state preschool crowd-out private provision? the impact of universal preschool on the childcare sector in oklahoma and georgia. *Journal of Urban Economics 83*, 18–33.

Bassok, D. and E. Galdo (2016). Inequality in preschool quality? community-level disparities in access to high-quality learning environments. *Early Education and Development 27*(1), 128–144.

Berlinski, S., M. M. Ferreyra, L. Flabbi, and J. D. Martin (2020). Child care markets, parental labor supply, and child development. *Parental Labor Supply, and Child Development*.

Berry, S., J. Levinsohn, and A. Pakes (1995). Automobile prices in market equilibrium. *Econometrica: Journal of the Econometric Society*, 841–890.

Berry, S., J. Levinsohn, and A. Pakes (2004). Differentiated products demand systems from a combination of micro and macro data: The new car market. *Journal of Political Economy 112*(1), 68–105.

Blau, D. and E. Tekin (2007). The determinants and consequences of child care subsidies for single mothers in the usa. *Journal of Population Economics 20*, 719–741.

Bodéré, P. (2022). Dynamic spatial competition in early education: an equilibrium analysis of the preschool market in pennsylvania.

Borowsky, J. (2019). Who benefits from child care ratings? evidence from minnesota's parentaware program. *Department of Economics, University of Minnesota, Minneapolis and Saint Paul, MN*.

Borowsky, J., J. H. Brown, E. E. Davis, C. Gibbs, C. M. Herbst, A. Sojourner, E. Tekin, and M. J. Wiswall (2022). An equilibrium model of the impact of increased public investment in early childhood education. Technical report, National Bureau of Economic Research.

Brown, J. H. (2018). Does public pre-k have unintended consequences on the child care market for infants and toddlers? *Princeton University Industrial Relations Section Working Paper 626*.

Cascio, E. U., S. J. Haider, and H. S. Nielsen (2015). The effectiveness of policies that promote labor force participation of women with children: A collection of national studies.

Child Care Aware of America (2022). The cost of child care 2022.

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

Decker, P. and K. Kelly (2022). Should childcare subsidies be universal or targeted? *Journal of Policy Analysis and Management 41*(3), 921–943.

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

Durkin, K., M. W. Lipsey, D. C. Farran, and S. E. Wiesen (2022). Effects of a statewide pre-kindergarten program on children's achievement and behavior through sixth grade. *Developmental Psychology*.

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

Fan, Y. (2013). Ownership consolidation and product characteristics: A study of the us daily newspaper market. *American Economic Review 103*(5), 1598–1628.

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

Gandhi, A. and J.-F. Houde (2019). Measuring substitution patterns in differentiated-products industries. *NBER Working paper* (w26375).

Goolsbee, A. and A. Petrin (2004). The consumer gains from direct broadcast satellites and the competition with cable tv. *Econometrica 72*(2), 351–381.

Gormley Jr, W. T., T. Gayer, D. Phillips, and B. Dawson (2005). The effects of universal pre-k on cognitive development. *Developmental Psychology 41*(6), 872.

Gray-Lobe, G., P. A. Pathak, and C. R. Walters (2023). The long-term effects of universal preschool in boston. *The Quarterly Journal of Economics 138*(1), 363–411.

Gregg, A. and J. Peiser (2023, October 22). Drugstore closures are leaving millions without easy access to a pharmacy. *The Washington Post*.

Hatfield, B. E., J. K. Lower, D. J. Cassidy, and R. A. Faldowski (2015). Inequities in access to quality early care and education: Associations with funding and community context. *Early Childhood Research Quarterly 30*, 316–326.

Heckman, J., R. Pinto, and P. Savelyev (2013). Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. *American Economic Review 103*(6), 2052–86.

Herbst, C. M. and E. Tekin (2012). The geographic accessibility of child care subsidies and evidence on the impact of subsidy receipt on childhood obesity. *Journal of Urban Economics 71*(1), 37–52.

Hillman, N. (2019). Place matters: A closer look at education deserts. *Washington: Third Way*.

Holmes, T. J. and H. Sieg (2015). Structural estimation in urban economics. In *Handbook of regional and urban economics*, Volume 5, pp. 69–114. Elsevier.

Hotz, V. J. and M. Xiao (2011). The impact of regulations on the supply and quality of care in child care markets. *American Economic Review 101*(5), 1775–1805.

Jang, E., S. Gu, and B. Poole (2016). Categorical reparameterization with gumbel-softmax. *arXiv preprint arXiv:1611.01144*.

Johnson, A. D., R. M. Ryan, and J. Brooks-Gunn (2012). Child-care subsidies: Do they impact the quality of care children experience? *Child development 83*(4), 1444–1461.

Kensinger Rose, K. and J. Elicker (2008). Parental decision making about child care. *Journal of Family Issues 29*(9), 1161–1184.

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

Kline, P. and C. R. Walters (2016). Evaluating public programs with close substitutes: The case of head start. *The Quarterly Journal of Economics 131*(4), 1795–1848.

Layzer, J., B. Goodson, and M. Brown-Lyons (2007). National study of care for low-income families: Care in the home: A description of family child care and the experiences of the families and children that use it—final report.

Leslie, L. A., R. Ettenson, and P. Cumsille (2000). Selecting a child care center: What really matters to parents? In *Child and Youth Care Forum*, Volume 29, pp. 299–322. Springer.

Ludwig, J. and D. L. Miller (2007). Does head start improve children's life chances? evidence from a regression discontinuity design. *The Quarterly Journal of Economics 122*(1), 159–208.

Neumark, D. and H. Simpson (2015). Place-based policies. In *Handbook of regional and urban economics*, Volume 5, pp. 1197–1287. Elsevier.

Peyton, V., A. Jacobs, M. O'Brien, and C. Roy (2001). Reasons for choosing child care: Associations with family factors, quality, and satisfaction. *Early childhood research quarterly 16*(2), 191–208.

Posadas, J. and M. Vidal-Fernandez (2013). Grandparents' childcare and female labor force participation. *IZA Journal of Labor Policy 2*, 1–20.

Silliman, M. and J. Mäkinen (2022). Childcare, social skills, and the labor market.

Texas Workforce Commission (2023). Workforce development boards websites.

The Urban Child Institute (2016). All child care is not created equal.

The Urban Institute (2019). Executive summary: Assessing child care subsidies through an equity lens.

Wooldridge, J. (2010). Wooldridge. econometric analysis of cross section and panel data.

Online Appendix

Quantifying the Trade-Offs between Proximity and Quality in Childcare Provision: A Demand Analysis of Household-Level Choices

A. Childcare Services Program in Texas

B. Construction of Instruments

C. First-Stage Regressions

D. Alternative Specifications

A Childcare Services Program in Texas

The following tables and figures provide detailed insights into the geographical distribution, reimbursement rates, and parent copayment structure of the CCS program in Texas.

Figure A.1 presents the geographical distribution of the CCS program in Texas. Panel (a) provides an overview of the locations of the 28 local boards that are responsible for the day-to-day operations of the CCS program (Texas Workforce Commission, 2023). These boards play a vital role in coordinating and implementing the program at the local level. Panel (b) displays the geographical locations of the providers who actively participated in the CCS program between the years 2015 and 2019.

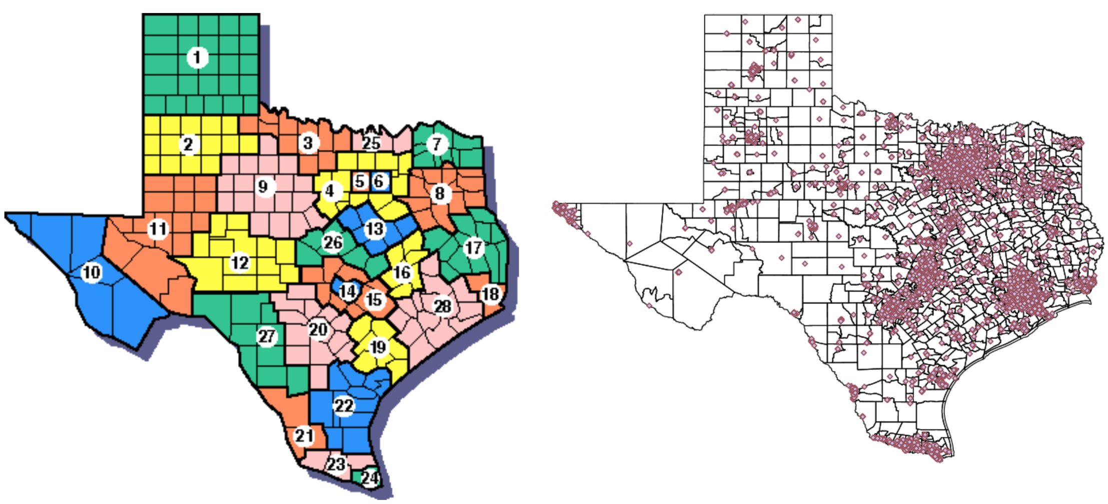
Table A.1 offers information regarding the reimbursement rates set by one specific board for a given administrative year. This table serves as an illustrative example, presenting the reimbursement rates categorized by various factors such as provider program type (licensed center, licensed home, registered home), program rating (ranging from no rating to TRS 2-star, TRS 3-star, TRS 4-star, and Texas School Ready), care schedule (full-time or part-time), and child age group (infant, toddler, pre-school, school-age). By providing these detailed reimbursement rates, the table allows for a comprehensive understanding of the financial aspects associated with different types of childcare programs and their specific characteristics.

Table A.2 complements the information presented in Table A.1 by showcasing an example of the sliding fee scale for parent copayments set by a particular board for a given administrative year. This table outlines the fee structure for parents, including the charges for the first child and any additional children. Moreover, it presents the income thresholds associated with each income bracket, taking into account the family size. These details offer insights into the affordability and cost-sharing arrangements for families participating in the CCS program, helping to understand the financial implications for parents based on their income levels and family size.

It shows the fee charged to the parent for the first child and every additional child. It shows the income threshold for each income bracket for each family size.

Figure A.1. Geographical Jurisdictions of LWDBs and Geographical Locations of Childcare Programs in Texas

(a) Geographic Jurisdictions of LWDBs (b) Geographic Locations of Childcare Programs



Source: Texas Workforce Commission

Table A.1. An example of reimbursement rates for childcare providers

Type	Rating	Infant		Toddler		Preschool		School	
		Full Time	Part Time	Full Time	Part Time	Full Time	Part Time	Full Time	Part Time
Licensed Center	Regular	45.20	40.20	40.60	36.00	36.80	28.40	34.60	26.00
	TRS2	47.46	42.21	42.63	37.80	38.64	29.82	36.33	27.30
	TRS3	48.42	43.56	43.45	38.70	39.38	30.96	37.08	28.44
	TRS4	53.80	48.40	48.00	43.00	43.40	34.40	41.20	31.60
	TSR					38.64	29.82		
Licensed Home	Regular	40.00	36.00	36.60	32.60	34.00	28.00	30.40	26.00
	TRS2	42.00	37.80	38.43	34.23	35.70	29.40	31.92	27.30
	TRS3	43.20	39.24	39.42	35.46	36.38	30.42	32.94	28.44
	TRS4	48.00	43.60	43.80	39.40	40.40	33.80	36.60	31.60
	TSR					35.70	29.40		
Registered Home	Regular	38.20	33.40	35.00	30.00	31.20	24.80	27.20	22.20
	TRS2	40.11	35.07	36.75	31.50	32.76	26.04	28.56	23.31
	TRS3	41.58	41.00	42.00	36.40	37.40	30.40	33.00	27.60
	TRS4	46.20	41.00	42.00	36.40	37.40	30.40	33.00	27.60
	TRS2	–	–	–	–	32.76	26.04	–	–

Note: Table presents an example of reimbursement rates from an LWDB for a given administrative year. *Source*: Texas Workforce Commission.

Table A.2. An example of the sliding fee scale

First Child	30	80	120	150	185	220	255	300
Additional Child	20	35	50	60	70	85	110	125
Income Threshold	50%	75%	100%	125%	150%	175%	200%	85%
Threshold Criterion	FPG†	FPG	FPG	FPG	FPG	FPG	FPG	SMI‡
Family size	Income amounts							
2	718	1,078	1,437	1,796	2,155	2,514	2,873	3,954
3	905	1,358	1,810	2,263	2,715	3,168	3,620	4,884
4	1,092	1,638	2,183	2,729	3,275	3,821	4,367	5,814
5	1,278	1,918	2,557	3,196	3,835	4,474	5,113	6,744
6	1,465	2,198	2,930	3,663	4,395	5,128	5,860	7,675
7	1,652	2,478	3,303	4,129	4,955	5,781	6,607	7,849
8	1,838	2,758	3,677	4,596	5,515	6,434	7,353	8,024
9	2,025	3,038	4,050	5,063	6,075	7,088	8,100	8,198
10	2,212	3,318	4,423	5,529	6,635	7,741	-	8,372
11	2,398	3,598	4,797	5,996	7,195	8,394	-	8,547
12	2,585	3,878	5,170	6,463	7,755	-	-	8,721
13	2,772	4,158	5,543	6,929	8,315	-	-	8,896
14	2,958	4,438	5,917	7,396	8,875	-	-	9,070
15	3,145	4,718	6,290	7,863	-	-	-	9,244

Note: Copay is determined by family income and number of children in care. Income thresholds are established by the Boards. Table presents an example of the sliding fee scale.
†FPG: Federal Poverty Guideline, ‡SMI: State Median Income. *Source*: Texas Workforce Commission

B Construction of Instruments

To address the issue of endogeneity in providers' quality, we construct a set of instruments using direct and indirect competitors' characteristics and market information. We denote the set of direct competitors, indirect competitors, and the market for provider j as $\mathcal{C}_j$, $\mathcal{E}_j$, and $\mathcal{I}_j$ respectively. Additionally, $d_{j\ell}$ represents the distance between provider j and provider ℓ, cap_j denotes the licensed capacity of provider j, SVI_i represents the SVI of family i's residential census tract, and x_j represents a characteristic of provider j such as program type or owner type.

Below, we provide a detailed explanation of how we construct the instruments using direct and indirect competitors' characteristics and market information. We follow the same steps to construct analogous instruments using direct and indirect competitors' characteristics and market information.

Total distance-adjusted count (IVs 1 and 13): For both the direct and indirect competitors, this instrument gives a measure of the overall level of competition faced by the provider. We take the Gaussian transformation of the distance between the provider and its competitors and sum them up, placing higher emphasis on competitors that are located closer to the provider:

$$\sum_{\ell \in \mathcal{C}_j} \exp(-d_{j\ell}).$$

Variations in distance-adjusted Capacity-to-Demand ratios (IVs 2 and 14): This instrument measures competition in terms of excess supply. It provides a quantitative measure of how much additional capacity a provider has compared to its competitors. Providers with higher excess supply are likely to face a more competitive environment as they compete for a smaller customer base for their childcare spots.

First, we calculate the capacity-to-demand ratio for each provider. This ratio is obtained by dividing the licensed capacity of the provider by the total number of families in its market. In constructing this ratio, we place higher emphasis on families that are located closer to the provider.

Next, we calculate the difference between the provider's ratio and each competitor's ratio, placing higher emphasis on closely located competitors, and sum these differences across all competitors:

$$\sum_{\ell \in \mathcal{C}_j} \left(\frac{cap_j}{\sum_{i \in \mathcal{I}_j} n_i \exp(-d_{ji})} - \frac{cap_\ell}{\sum_{i \in \mathcal{I}_\ell} n_i \exp(-d_{\ell i})} \right) \exp(-d_{j\ell}).$$

Squared Variations in distance-adjusted Capacity-to-Demand ratios (IVs 3 and 15): The instrument is constructed to capture the squared differences of the capacity-to-demand ratios, adjusted for distance, among a provider and its competitors.

To begin, we repeat the steps followed in the construction of the previous instrument to calculate the capacity-to-demand ratio for each provider. Next, we compute the squared difference between the focal provider's capacity-to-demand ratio and the competitor's capacity-to-demand ratio, adjusted by the Gaussian distances between them:

$$\sum_{\ell \in \mathcal{C}_j} \left(\frac{cap_j}{\sum_{i \in \mathcal{I}_j} n_i \exp(-d_{ji})} - \frac{cap_\ell}{\sum_{i \in \mathcal{I}_\ell} n_i \exp(-d_{\ell i})} \right)^2 \exp(-d_{j\ell})$$

Variations in distance-adjusted Capacity-to-SVI ratios (IVs 4 and 16): This instrument measures competition in terms of excess supply while taking into account the socioeconomic vulnerability of the market. It gives a measure of the potentially unequal distribution of resources and services across different communities.

To begin, we calculate the capacity-to-SVI ratio for the focal provider. This ratio is determined by dividing the provider's licensed capacity by the sum of the number of families in its market, adjusted by the SVI of each family as well as the proximity to the provider. Next, we compute the difference between the focal provider's capacity-to-SVI ratio and the competitor's capacity-to-SVI ratio. Similar to the previous instruments, we adjust these differences by the Gaussian distances:

$$\sum_{\ell\in\mathcal{C}_j}\left(\frac{cap_j}{\sum_{i\in\mathcal{I}_j} n_i \exp(-\frac{1}{2}(1-SVI_i)-\frac{1}{2}d_{ji})}-\frac{cap_\ell}{\sum_{i\in\mathcal{I}_\ell} n_i \exp(-\frac{1}{2}(1-SVI_i)-\frac{1}{2}d_{\ell i})}\right)\exp(-d_{j\ell}).$$

Squared variations in distance-adjusted Capacity-to-SVI ratios (IVs 5 and 17): This instrument measures the squared differences in the capacity-to-SVI ratios, adjusted for distance, between a provider and its competitors.

Repeating the steps in the construction of the previous instrument, we calculate the capacity-to-SVI ratio for each provider. Next, we compute the squared differences between the focal provider's capacity-to-SVI ratio and the competitor's capacity-to-SVI ratio. Similar to the previous instruments, we adjust these ratios by the Gaussian distances:

$$\sum_{\ell\in\mathcal{C}_j}\left(\frac{cap_j}{\sum_{i\in\mathcal{I}_j} n_i \exp(-\frac{1}{2}(1-SVI_i)-\frac{1}{2}d_{ji})}-\frac{cap_\ell}{\sum_{i\in\mathcal{I}_\ell} n_i \exp(-\frac{1}{2}(1-SVI_i)-\frac{1}{2}d_{\ell i})}\right)^2\exp(-d_{j\ell}).$$

Characteristics (IVs 6, 7, 18 and 19): These instruments measure the competition specific to the program type and ownership type of the provider. They count the number competitors that share the same program and ownership characteristics as the focal provider, adjusting for the distance between them.

For instance, let the program type of the provider be denoted as x_j, i.e., whether the provider is a center or a home program. The expression $\mathbb{1}\{x_j = x_\ell\}$ equals 1 if the program type of the focal provider (x_j) is the same as the program type of a competitor (x_ℓ), and 0 otherwise. By summing these over all competitors, we count the number of competitors that share the same program type as the provider. We also adjust this count by the Gaussian distance:

$$\sum_{\ell\in\mathcal{C}_j}\mathbb{1}\{x_j = x_\ell\}\exp(-d_{j\ell}).$$

We repeat these steps to construct the ownership type instrument.

Interaction between Capacity-to-Demand ratio and characteristics (IVs 8, 9, 20 and 21): These instruments involve interactions between the capacity-to-demand ratios and the program/ownership characteristics of providers. These instruments allow us to capture the combined effect of these variables on competition.

$$\sum_{\ell\in\mathcal{C}_j}\left(\frac{cap_j}{\sum_{i\in\mathcal{I}_j} n_i \exp(-d_{ji})}-\frac{cap_\ell}{\sum_{i\in\mathcal{I}_\ell} n_i \exp(-d_{\ell i})}\right)\mathbb{1}\{x_j = x_\ell\}\exp(-d_{j\ell})$$

Interaction among characteristics (IVs 10 and 22): This instrument captures the interaction between the program and ownership types. This instrument allows us to examine how the joint effect of these characteristics influences competition. $\sum_{k'\neq k}\sum_{\ell\in\mathcal{C}_j}\mathbb{1}\{x_{jk} = x_{\ell k}\}\mathbb{1}\{x_{jk'} = x_{\ell k'}\}\exp(-d_{j\ell})\ \forall k\in\{1,\ldots,K\}$

Interaction between Capacity-to-SVI ratio and characteristics (IVs 11, 12, 23 and 24): These instruments capture the interaction between the capacity-to-SVI ratio and the characteristics of providers, including program types and ownership types:

$$\sum_{\ell\in\mathcal{C}_j}\left(\frac{cap_j}{\sum_{i\in\mathcal{I}_j} n_i \exp(-\frac{1}{2}(1-SVI_i)-\frac{1}{2}d_{ji})}-\frac{cap_\ell}{\sum_{i\in\mathcal{I}_\ell} n_i \exp(-\frac{1}{2}(1-SVI_i)-\frac{1}{2}d_{\ell i})}\right)\mathbb{1}\{x_{jk} = x_{\ell k}\}\exp(-d_{j\ell}).$$

C First-Stage Regressions

To address the endogeneity of the categorical variable representing the provider's TRS rating, we employ the two-step instrumental variable method proposed by Wooldridge (2010).

In our framework, we denote the endogenous characteristic of provider j as x_j^1, the exogenous characteristics as x_j^2, and the instruments as z_j. We first estimate the logit model $P(x_j^1|x_j^2, z_j) \equiv G(x_j^2, z_j; \zeta)$ using maximum likelihood estimation. This model captures the relationship between the provider's characteristics and the probability of having a specific TRS rating, conditional on the exogenous characteristics and instruments. The estimated logit model provides us with the predicted probabilities of being TRS 2-Star, 3-Star, or 4-Star.

In the second step, we perform a two-stage least squares (2SLS) regression of the estimated fixed effects $\hat{\delta}(\hat{\theta})$ using the instruments, the predicted probabilities from the logit model, and the exogenous characteristics x_j^2. Table C.1 reports the first stage of the 2SLS regression, providing details on the instrumental variable estimation. The reported F-statistics indicate the strength of the instruments and alleviate concerns about weak instruments, as they are large. The analysis in Column 1 focuses on whether a provider is TRS two or three star, and we include the provider's program and ownership characteristics and predicted probabilities of each TRS rating as explanatory variables. Similarly, Column 3 of the table presents the logit analyses for TRS 4-Star rating.

Table C.1. Results: First stage regression

	(1) TRS 2 or 3 Star	(2) TRS 4 Star
Predicted Probability: TRS 2 or 3 Star	1.08	-0.07
	(0.09)	(0.09)
Predicted Probability: TRS 4 Star	-0.13	0.97
	(0.08)	(0.10)
Program Type: Center	-0.00	0.01
	(0.02)	(0.02)
Owner Type: Individual	-0.03	-0.00
	(0.02)	(0.03)
Owner Type: Private Organization	-0.02	-0.00
	(0.02)	(0.02)
Mean	0.149	0.186
F	28.032	20.204
Obs	7,688	7,688

Note: Table presents the results of the first stage of the second step of the model estimation (2SLS) as described in Section 4.1. The predicted probabilities of TRS 2 or 3 Star and of TRS 4 Star are based on estimating the logit model predicting the probability of having a specific TRS rating given the exogenous characteristics (program type, program ownership) of the 7,688 childcare programs and the instruments.

D Alternative Specifications

In the x_i vector in the baseline specification in Equation (5), we include only the SVI of the census tract that household i resides in. Now, we explore how the results may change if we expand x_i to include the household's race/ethnicity and income. Additionally, we include fixed effects for the first year that the family enrolls in child care. The results remain qualitatively similar, but the utility from enrolling in the TRS 4 star is stronger, and so is the disutility from enrolling in TRS 2 or 3 Star both in magnitude and statistical significance. The willingness to trade off proximity for TRS 4 Star is also stronger in magnitude but only slightly. In the baseline specification, higher SVI households appeared to be less likely to enroll in TRS 4 Star compared to programs with no TRS rating. After including household characteristics, we see that higher SVI households now appear to be more likely to enroll in TRS 2 or Star programs compared to programs with no TRS rating. Both results point toward the same conclusion that higher SVI households are more likely to enroll in lower-quality programs.

In the x_i vector in the baseline specification in Equation (5), we include only the SVI of the census tract that household i resides in. Now, we explore how the results may change if we expand x_i to include the household's race/ethnicity and income. Additionally, we include fixed effects for the first year that the family enrolls in child care. The results remain qualitatively similar, but the utility from enrolling in the TRS 4-star program is stronger, and so is the disutility from enrolling in TRS 2 or 3-star programs, both in magnitude and statistical significance. The willingness to trade off proximity for TRS 4-star programs is also stronger in magnitude but only slightly.

In the baseline specification, households with higher SVI appeared to be less likely to enroll in TRS 4-star programs compared to programs with no TRS rating. After including household characteristics, we see that higher SVI households now appear to be more likely to enroll in TRS 2 or 3-star programs compared to programs with no TRS rating. Both results point toward the same conclusion: higher SVI households are more likely to enroll in lower-quality programs.

Table D.1. Estimation results

	(1)	(2)	(3)
Constant	0.28	-8.14	-9.66
	(0.15)	(0.20)	(0.26)
Distance	-0.55	-0.58	-0.58
	(0.00)	(0.00)	(0.00)
Center	1.97	1.76	1.79
	(0.02)	(0.05)	(0.10)
Owner: Individual	0.26	0.13	1.23
	(0.01)	(0.07)	(0.12)
Owner: Pvt Org	0.46	0.35	1.00
	(0.01)	(0.06)	(0.08)
TRS 2 or 3 Star	0.59	0.49	-2.38
	(0.03)	(0.05)	(0.58)
TRS 4 Star	0.78	0.60	5.80
	(0.03)	(0.05)	(0.50)
TRS 2 or 3 Star $\times$ Distance	-0.00	0.01	0.01
	(0.00)	(0.00)	(0.00)
TRS 4 Star $\times$ Distance	0.02	0.03	0.03
	(0.00)	(0.00)	(0.00)
TRS 2 or 3 Star $\times$ SVI	0.12	0.07	0.07
	(0.04)	(0.03)	(0.03)
TRS 4 Star $\times$ SVI	0.03	0.03	0.03
	(0.03)	(0.02)	(0.02)
TRS 2 or 3 Star $\times$ Black	-0.14	-0.03	-0.03
	(0.02)	(0.02)	(0.02)
TRS 4 Star $\times$ Black	-0.10	-0.06	-0.06
	(0.02)	(0.02)	(0.02)
TRS 2 or 3 Star $\times$ Hispanic	-0.03	-0.00	-0.00
	(0.02)	(0.02)	(0.02)
TRS 4 Star $\times$ Hispanic	0.04	0.06	0.06
	(0.02)	(0.02)	(0.02)
TRS 2 or 3 Star $\times$ Income	-0.01	-0.01	-0.01
	(0.00)	(0.00)	(0.00)
TRS 4 Star $\times$ Income	-0.01	-0.01	-0.01
	(0.00)	(0.00)	(0.00)
Provider f.e.	No	Yes	Yes
Instrument for quality	No	No	Yes

Note: Table presents the two-step nested fixed point MLE estimates of the preference parameters from Equation 8. Column (1) does not account for the unobserved heterogeneity across childcare programs. Column (2) accounts for this heterogeneity but does not account for the endogeneity in providers' TRS (quality) ratings. Column (3) accounts for both. Standard errors are reported in parentheses. The first-stage regression results for the instruments are provided in the Online Appendix C.C. All specifications include LWDB and year fixed effects.