JOHANNES KEPLER UNIVERSITY LINZ

**DEPARTMENT OF ECONOMICS**
**JOHANNES KEPLER UNIVERSITY OF LINZ**

# How Small are Small Markets?

## Location Choice and Geographical Market Size for Child Care Services

by

Astrid PENNERSTORFER
Dieter PENNERSTORFER

Working Paper No. 1814
October 2018

**Johannes Kepler University of Linz**
**Department of Economics**
**Altenberger Strasse 69**
**A-4040 Linz - Auhof, Austria**
**www.econ.jku.at**

dieter.pennerstorfer@jku.at

# How Small are Small Markets?

## Location Choice and Geographical Market Size for Child Care Services

Astrid Pennerstorfer* and Dieter Pennerstorfer†

October 24, 2018

**Abstract**

In this article we propose an innovative way of delineating geographical markets based on easily accessible data. We apply this concept for the day care industry and investigate providers' location choices relative to local market characteristics to evaluate the widespread presumption that local markets for child care services are geographically very small. Using a panel of all day care centers for the metropolitan region of Vienna, Austria, for nearly a decade, as well as geographically extremely disaggregated data on the spatial distribution of children under the age of six at the 250m $\times$ 250m grid cell level, we find that the location of children and day care centers are strongly related, but this relationship diminishes as soon as the distance between a child's place of residence and the day care center's location increases. We conclude that geographical markets for day care services in metropolitan regions are indeed very small (about 500m or 550 yards).

Keywords: spatial market definition, location choice, market entry, child care, grid data

JEL Classification: R30, R53, L10, H44

*WU Vienna University of Economics and Business. E-mail: astrid.pennerstorfer@wu.ac.at. Welthandelsplatz 1, Building D4, 1020 Vienna, Austria. Declarations of interest: none.

†Corresponding author. Johannes Kepler University Linz. Altenberger Street 69, 4040 Linz, Austria. E-mail: dieter.pennerstorfer@jku.at. Declarations of interest: none

# 1 Introduction

Empirical contributions investigating market conduct of child care services usually suspect that geographical markets are very small (see, e.g., Cleveland and Krashinsky, 2009, Hotz and Xiao, 2011, or Xiao, 2010), because 'transportation costs of child care are extraordinarily high. The service is not typically brought to the child, but the other way round; parents have to transport the child to and from the service every day of the working week' (Cleveland and Krashinsky, 2009, p. 441). Home-child care journeys induce costs for parents in terms of time and other expenditures, and as parents usually have to make these trips back and forth twice a day, spatial proximity is crucial. Empirically, the spatial accessibility of child care has not only been found to be an important factor in the parental decision making process (e.g. Johansen et al., 1996; Kim and Fram, 2009; Teszenyi and Hevey, 2015), but also influences parental life satisfaction (Yamauchi, 2010) and maternal labor market participation rates (Compton and Pollak, 2014; Herbst and Barnow, 2008; Kawabata and Abe, 2018).[1]

While many existing studies concerned with day care choice agree that the location of day care centers is of key importance for parents (e.g. Johansen et al., 1996; Kim and Fram, 2009; Teszenyi and Hevey, 2015), and although child care markets have been described as small (e.g. Cleveland and Krashinsky, 2009; Hotz and Xiao, 2011; Xiao, 2010), there are surprisingly few attempts to quantify the geographical size of these local markets. Surveys on the parental choice of day care institutions are rather vague regarding the preferences for proximity, and include questions whether there are day care centers in the 'right location' (Yamauchi, 2010, p. 236), child care services in 'a reasonable distance' (Herbst and Barnow, 2008, p. 129) or 'close to home' vs. necessitating 'a bit of a drive' (Rose and Elicker, 2008, p. 1174), which impedes quantifying e.g. maximum commuting distances parents are willing

[1]Empirical evidence on the positive relationship between availability (or low costs) of child care and female labor market participation without focusing on the spatial dimension is widespread, see e.g. Baker et al. (2008), Bick (2016) or Haan and Wrohlich (2011).

to endure. Empirical evidence on how spatial considerations affect parents' perception of different providers being considered as close substitutes, and on how spatial proximity affects rates of parents' use of child care services thus remains scarce.

Consequently, articles investigating firm and market conduct in the day care industry are very heterogeneous regarding the applied spatial market delineation, ranging from census tracts (Queralt and Witte, 1998, 1999; Small and Stark, 2005) to zip or postal codes (Hotz and Xiao, 2011; Lam et al., 2013; Noailly and Visser, 2009; Xiao, 2010) and school zones (Owens and Rennhoff, 2014). The specific level of regional aggregation is usually not based on empirical considerations and the issue of geographic market definition often receives only limited attention in this literature. This is not specific to articles investigating child care, but applies more generally to empirical contributions investigating the causes or the consequences of market entry in various industries. Two approaches are commonly found to circumvent the issue of delineating geographical markets: First, entry models of (and in the spirit of) Bresnahan and Reiss (1990, 1991) often restrict their analyses to competition in small and geographically isolated markets, where spatial market definition is straightforward (see, e.g., Collard-Wexler, 2014). However, this approach impedes analyzing larger cities. Additionally, as argued by Aguirregabiria and Suzuki (2016), extrapolating results obtained from investigating rural markets to urban ones is generally not plausible. The second commonly applied strategy is picking a particular regional level of aggregation and showing that the results are robust to alternative geographical market definitions, as done, e.g., by Matsa (2011) or Waldfogel (2008).

One main reason why spatial market definition receives only limited attention is clearly related to data restrictions. To delineate markets based on demand substitutability and supply reactions, information on prices and quantities is necessary, data that are typically hard to obtain at the store level. In particular, the necessity to obtain this kind of data would thwart one very attractive feature of (structural) entry models, namely that these analyses are usually feasible with easily accessible data (basically the number of firms and the population size of local markets).

In this article we address the issue of geographical market definition by investigating location and entry decisions of day care centers relative to local market characteristics. We have access to the entire population of child care facilities covering nearly a decade for the metropolitan region of Vienna, Austria, including the day care centers' precise locations. Additionally, and even more importantly, we can draw on geographically extremely disaggregated data on the spatial distribution of children under the age of six (and other demand characteristics) at the 250m $\times$ 250m (about 270 yards) grid cell level. For each location (grid cell) we calculate the number of children in distance bands (of various sizes) around that location and examine the relationship between the number and the entry of day care centers in a specific location and (potential) demand in these distance bands. As argued by Waldfogel (2008), we can expect a strong relationship between the number of providers in a location and the number of consumers in a specific geographical area, if demand is drawn from that area. We thus reason that (easily observable) entry and exit decisions reveal information about the underlying market fundamentals (in our case, spatial demand substitutability). Based on these regression results we are able to infer the geographical market size.

We contribute to the literature on entry, firm and market conduct by providing a rather simple approach to define spatial markets. While our data is very detailed from a spatial perspective, it is easily accessible and privacy protection is not a relevant issue, in line with the modest data requirements common to (structural) entry models. We also contribute to the literature on child care services by quantifying the distance parents are willing to travel to day care facilities and by providing empirical evidence on the geographical market size in urban areas. This is highly policy relevant, as the positive effects ascribed to child care provision (on e.g. female labor market participation) will only be realized if the service is provided sufficiently closely to the parents' homes.

The day care market and the utilized grid data are particularly well-suited for this kind of analysis: First, it is well-established that proximity is very important in this industry (despite the lack of quantification). Second, as each child can occupy only one place in a single day care center, the number of children is a good proxy for (potential) local demand. Last, regional statistical grid units are particularly useful for this kind of analysis, as they are very small and standardized. The small size of the grid cells allows a high degree of flexibility in defining distance bands, and facilitates applications for a large variety of industries providing spatially differentiated goods or services. Our approach can thus serve as an easily replicable blueprint to empirically evaluate the geographical market size when analyzing firm and market conduct.

The remainder of the article is structured as follows: Section 2 gives an overview of Vienna's child care market. Section 3 presents the data used and outlines the empirical strategy. We then go on to describe the main results and alternative models serving as a robustness check in Section 4, before concluding in Section 5.

# 2 Industry Background: Vienna's Child Care Market

The area under investigation comprises the entire city of Vienna, Austria, a city with more than 1.8 million inhabitants covering 414.87 $km^2$ (Statistik Austria, 2018). Vienna is a federal state, and as such is responsible for regulating child care for children up to the age of six. Vienna's market for day care is a mixed market and consists almost exclusively of public and private non-profit institutions. There are different types of center-based day care arrangements, consisting of nursery schools for children between zero and three years of age ('Kinderkrippen'), play schools ('Kindergärten') for three to six year old children, and day care for heterogeneous age groups, comprising children between zero and six ('altersgemischte Gruppen').

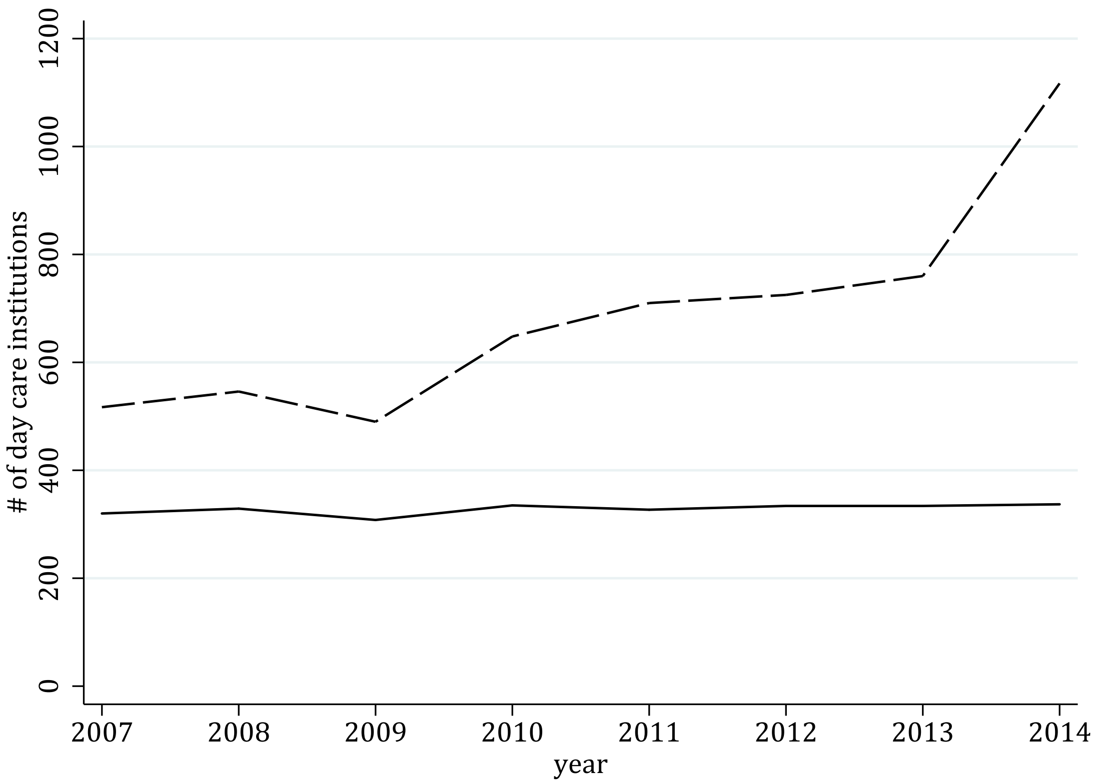
The sizes of day care centers are rather heterogeneous, and each institution hosts between one and 15 nursery groups (with 3.1 groups on average). It is very common that one center hosts groups of different types. Vienna's regulation and supply differ considerably from its surrounding federal state, Lower Austria, offering a more generous supply with longer opening hours and more available places, especially for children under three. As parents from outside Vienna are not subsidized and would have to fully pay for day care, it is plausible to assume that there is no (or only very little) influx of customers from outside Vienna.[2]

Vienna's child care market has recently experienced tremendous growth. The number of day care institutions increased by 74% between the years 2007 and 2014, growing from 837 day care centers to a total of 1,454 (see Figure 1). In the same time period the number of nursery groups increased from 2,675 to 3,921 (+47%). Figure 1 also indicates a change in the provider mix: Whereas 38% of all day care centers were run by public providers in 2007, this share declined to only 23% in 2014.

There are several reasons that can help explain the dynamics in this market. The Barcelona objectives, set by the European Union in 2002, envision the development of formal child care for young children in order to facilitate female labor participation (European Commission, 2013). These objectives encouraged federal and local governments to either provide day care centers themselves or to facilitate market entry of private institutions. From the demand side the strong increase can be explained only to a small extent by a general population growth in Vienna of about 8% (and of 12% of children below six) in this time period. More importantly, the share of children attending day care institutions in Vienna increased from only 23.1% of all children under three and 83.1% of all children between three and five in 2007 to 40.2% and 92.6% respectively in 2014 (Statistik Austria, 2015, p. 85).

[2]This assumption is also plausible the other way round. Additionally, for Viennese parents day care institutions outside Vienna are not very attractive due to generally inferior structural service quality (e.g. opening hours).

Figure 1: Growth of the Number of Day Care Institutions in Vienna




Notes: The solid (dashed) line denotes the number of public (private) day care institutions in Vienna.

Besides a general trend of a continuous increase in female (and in particular: mothers') labor market participation, two policy interventions strongly encouraged the parents' use of institutional child care. First, a change in the funding mode of institutional day care: Since 2009 day care institutions have received lump-sum subsidies and a certain fee per tended child from the local government. The new funding mode was widely promoted as 'cost-free childcare for all', and although some day care institutions still charge parents a small monthly fee for e.g. offering additional activities or prolonged opening hours, this change reduced costs for parents significantly. Second, in 2010 a law was introduced stipulating that all children in their final year before school (five-to-six-year-olds) are obliged to attend a formal day care institution for at least 16 hours a week (Stadt Wien, 2015, p. 20).

In particular the substantial reduction of costs for parents has led to a sharp increase in demand, but growth in supply could not keep up at the same pace. An empirical investigation for Vienna has shown that the main challenges for founders of new day care centers lie in finding suitable premises and personnel as well as in observing all legal regulations (Schinkowitsch, 2014). Consequently, in the first years after 2009 growth of day care institutions has lagged behind and led to an increase in waiting lists, especially for children under the age of three.

# 3 Data and Empirical Strategy

## 3.1 Data

The aim of this article is to provide an estimate on the geographical scope of local markets in the day care industry. To do so, we utilize two different data sources for our empirical analysis, namely information on the day care facilities (supply side characteristics) and population data (local demand).

Information on day care centers is provided by the public administration of Vienna (Magistratsabteilung 23) and comprises data on all day care facilities from 2007 to 2014. The data include information on the institutions running the facility and the numbers and types of groups (nursery schools, play schools or day care for heterogeneous age groups) and were collected each year in October. Geographical information was only provided at the level of registration districts ('Zählbezirke') and was thus supplemented by the exact postal addresses of all day care facilities using additional, publicly available data sources.[3] The postal addresses were geocoded and could thus be linked with spatial data on demand indicators.

[3] In order to supplement the data with the exact postal addresses the day care institutions were linked to data reported in 'Vienna day care guides' published by the 'Wiener Familienbund' in 2005 and 2011, as well as to open government data (published by the public administration). Vienna is divided into 250 registration districts, each of these administrative entities hosting only a small number of day care centers (4.1 on average). As the types of institutions running the facilities as well as other characteristics of the day care centers (like opening hours, for example) are reported in both data sets, linking the data was time-consuming, but straightforward work. We are grateful to Julia Groiß for providing assistance in linking the two data sets.

To get accurate measures of local demand we utilize detailed information on the spatial distribution of the population. The Austrian Statistical Office ('Statistics Austria') places regional statistical grid units over the entire territory of Vienna. The grids are independent of administrative boundaries and the size of one grid cell is 250m $\times$ 250m. Each person is assigned to exactly one cell based on his / her postal address. Note that this provides very detailed information on the spatial distribution of the population, as one square-kilometer (square-mile) is represented by 16 (41) cells. The population is categorized by eleven different age cohorts, allowing us to identify the number of children younger than six. Aligning the detailed information on the age structure with the places of residence we can derive accurate measures of (potential) demand at a very small-scale regional level. This annual information is provided by Statistics Austria for the years 2007 to 2014.[4] In addition to the age distribution, information on socio-demographic variables like the country of birth (COB) and employment status by sex (for the year 2007) and the highest completed level of education (for the year 2011) are available at the same grid level. Data on the number of jobs at the location of work were collected in 2001 and 2011 and are interpolated for the years in between. These data are again provided by Statistics Austria. We supplement the sample with open government data on the (annually provided) location of subway stations.[5]

To link data on supply with demand characteristics we use spatial grid cells as observation units and refer to grid cells as 'locations ($l$)' henceforth. The number of day care centers and the number of nursery groups are thus aggregated at the location level and linked with demand characteristics.

The sample area of Vienna consists of 6,962 grid cells. To avoid investigating undevelopable areas (parks, river, ...) we follow Nishida (2015) and discard all empty

[4]Data is usually collected on January 1$^{st}$ of the respective year. Only the first year is an exception, where data were collected on October 31$^{st}$ 2006 instead of January 1$^{st}$ 2007. Data on the population is thus always collected before information on day care facilities is surveyed.

[5]Data on the location of subway stations is available at Bundesministerium für Digitalisierung und Wirtschaftsstandort (2018).

cells, which amounts to nearly half of them. While information on the number of residents is always provided, data on socio-demographic characteristics is only provided if the number of inhabitants exceeds 30 (until 2011) or three people (afterwards). As some variables cannot be calculated if socio-demographic characteristics are not reported we also discard these observations, leaving a balanced panel of $N = 3{,}066$ grid cells over $T = 8$ years from 2007 to 2014, comprising 98.9% of the population and 95.8% of all nursery groups. To ensure that the panel is balanced we keep the observations for all time periods if socio-demographic characteristics are observed at this location in the first year of the sample, i.e. in 2007.

Descriptive statistics on the variables are provided in Table 1 below. On average, we find 0.32 day care centers (0.97 nursery groups) in each location, with a total number of 550 residents and 33 inhabitants younger than six. Each grid cell hosts about 260 jobs, and only 2% of all grid cells are provided with a subway station. Information on the female employment rate, on the country of birth, and on the highest educational diploma are calculated as the share of the total female population, the total population, or the total population $\geq 15$ years old in location $l$, respectively.

We expect that the number of children (younger than six) in the neighborhood is strongly related to the number of nursery groups in the respective location. If parents prefer a day care institution close to work (rather than close to their place of residence), the number of jobs will serve as an additional indicator of (local) demand. We, however, expect that the location of jobs and day care are only weakly correlated, as most working parents seem to prefer child care close to their place of residence rather than to their workplace: Queralt and Witte (1998) mention a survey for Maryland, where 76% indicate a preference for child care close to home (in contrast to 3% preferring child care close to work), and Michelson (1985) finds that the mothers' trips from home to child care are shorter than from child care to work. Following existing research, female labor participation is expected to be positively correlated with local child care provision (e.g., Johansen et al., 1996; Kim and Fram, 2009; Rose and Elicker, 2008). The educational background (e.g., Kim and Fram, 2009) and the country of origin of the residents in the vicinity may also affect the preference for local day care facilities. A dummy for subway stations serves as an indicator for easy accessibility, and we expect a positive relation between subway stations and the number of (new) nursery groups.

Table 1: Descriptive Statistics

| Variable name | Variable description | N | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|---|---|
| # of day care institutions | | 24,528 | 0.32 | 0.68 | 0 | 9 |
| # of nursery groups | | 24,528 | 0.97 | 2.21 | 0 | 20 |
| # of residents | | 24,528 | 550.87 | 566.62 | 2 | 4,116 |
| # of children aged $\leq 5$ | | 24,528 | 32.83 | 38.36 | 0 | 262 |
| # of jobs | # of employed individuals working at location $l$ | 24,528 | 257.46 | 552.10 | 0 | 8,649 |
| # of subway stations | | 24,528 | 0.02 | 0.15 | 0 | 1 |
| Share employed women | Share of employed female residents in location $l$ (in % of residential female population) | 24,528 | 40.06 | 9.20 | 0 | 78.95 |
| Share COB[1] Austria | Share of residents in location $l$ born in Austria (in % of total residential population) | 24,528 | 77.86 | 11.96 | 6.54 | 100 |
| Share COB[1] other EU country | Share of residents in location $l$ born in EU country other than Austria (in % of total residential population) | 24,528 | 8.25 | 4.10 | 0 | 48.32 |
| Share COB[1] ROW[2] | Share of residents in location $l$ born in a non-EU country (in % of total residential population) | 24,528 | 13.89 | 10.38 | 0 | 86.73 |
| Share high school diploma | Share of residents in location $l$ with high school diploma as highest educational attainment (in % of residential population $\geq 15$ years old) | 24,528 | 19.94 | 6.93 | 0 | 73.57 |
| Share college degree | Share of residents in location $l$ with a college degree as highest educational attainment (in % of residential population $\geq 15$ years old) | 24,528 | 18.15 | 12.05 | 0 | 62.50 |

Notes: [1]) country of birth; [2]) rest of the world.

## 3.2 Empirical Strategy

In order to estimate the geographic size of local day care markets, we explain the level and the entry of nursery groups at a particular location $l$ by demand characteristics (in particular the number of children) at the respective location, but also by demand indicators in various distance bands around that location. The idea to relate the number of firms in a location to potential demand in concentric circles around that location is outlined in Seim (2006), who was the first to investigate the explicit location choice within (geographically larger) markets. With this approach the locations of stores are not isolated and both competition (number of stores) and demand (population) of locations nearby influence each other.[6]

Waldfogel (2008) uses a similar approach when investigating the restaurant market. Instead of using circles around each location he uses different levels of zip code areas. The number of restaurants in a five-digit zip code area is thus estimated as a function of the population in this (five-digit zip code) area, as well as the population in the remaining parts of the (larger) four-digit and three-digit zip code regions. Using a reduced form approach he shows that the coefficient on the number of inhabitants in the (smallest) five-digit zip is large, while the estimated paremeter for the remaining four-digit zip code (comprising consumers living outside the five-digit zip code under investigation) is much smaller, and the respective parameter for the (largest) three-digit zip code area is virtually zero. Waldfogel (2008) argues that we can expect a strong relationship between the number of nursery groups in a location and the number of consumers in a specific geographic area, if demand is drawn from that area. To our knowledge, he is thus the first to infer the size of the (geographical) market from regression results on market entry.

We follow Seim (2006) and calculate demand indicators in distance bands around particular locations and link this procedure with the argument put forward by Wald-

[6] Seim (2006) and other contributions by Datta and Sudhir (2013), Nishida (2015) and Zhu and Singh (2009) analyzing various retail markets all find that the number of consumers in the vicinity increases firms' sales or profitability, but the effect declines with distance.

fogel (2008): We expect the number of providers in a location to be strongly associated with the number of children nearby, but presume the relation to weaken as the distance between the children's place of residence and the providers' location increases. When the relationship approaches zero we have reached the boundaries of the geographic market. The spatial market size is thus revealed by the observed location choice of day care centers relative to the spatial distribution of (potential) demand.

Note that this estimation strategy is only feasible if providers are responsive to demand. In Waldfogel (2008), analyzing the restaurant market, and most (structural) entry models following Bresnahan and Reiss (1990, 1991) this is ensured by assuming firms to maximize profits. Vienna's market for day care is dominated by public and non-profit providers, with for-profit firms being virtually nonexistent, so that the assumption of profit-maximizing firm behavior seems to be unjustified.[7] Even if public and non-profit providers do not maximize profits, it is plausible to assume that meeting local need or neighborhood demand is part of the objective function of both types of providers.[8]

To estimate the size of the catchment area and to evaluate the influence of distance on the size of the effects of demand indicators on market structure, we calculate all variables reported in Table 1 for 'distance bands' of different length around each location $l$. Distance serves as a proxy for travel time, with both variables being closely related, as the entire sample region is densely populated. Distance band $d_1$ describes all other locations surrounding location $l$ within an airline distance of $\leq$ 500m.[9] Variables calculated at distance band $d_1$ therefore comprise information on all grid cells $m \in \{1, ..., M\}$, characterized by 250m $\leq d_{lm} \leq$ 500m, with $d_{lm}$ denot-

[7] Owens and Rennhoff (2014), investigating competition between for-profit and non-profit child care providers, assume non-profit firms to maximize profits as well.

[8] Note that the empirical approach of relating day care centers' locations to local demand also serves as a test for this assumption.

[9] To determine the distance between two locations we calculate the Euclidean distance between the centroids of the respective grid cells. We thus implicitly assume that all day care facilities and the entire population are located in the centroid of each grid cell.

ing the Euclidean distance between locations $l$ and $m$. Distance band $d_2$ captures all cells within a distance between $500\text{m} < d_{lm} \leq 1\text{km}$, and $d_3$ within a distance between $1\text{km} < d_{lm} \leq 2\text{km}$. The respective variables are aggregated or averaged among all grid cells in the respective distance band.[10]

By including all variables aggregated or averaged at the distance bands around location $l$ (in addition to the variables based on characteristics of location $l$) it is possible to evaluate (i) which variables are correlated with the location choice of day care centers, (ii) up to which distance these variables are associated with entry decisions, and (iii) how the strength of these relationships depends on distance to location $l$. We are thus able to evaluate the geographical market size and the catchment area in the child care market.

The spatial structure of the variables is illustrated in Figures 2 to 5 below. Figure 2 shows a map section of Vienna, including the locations of day care facilities (in 2014) as well as district borders. Figure 3 includes the spatial grid pattern and Figure 4 reports the distribution of the number of children aged $\leq 5$ (also for 2014). Finally, Figure 5 illustrates the calculation of the variables around a particular location, depicted by the darkest shade of gray. All nursery groups at this location (grid cell) are aggregated. All explanatory variables are calculated both for location $l$ and for various distance bands around this location, indicated by lighter shades of gray in Figure 5 (distance band $d_3$ is suppressed for convenience).

[10] To be precise, all count data (i.e. the number of day care institutions, nursery groups, residents, children, jobs or subway stations) are simply aggregated among all grid cells in the respective distance bands around location $l$, while the variables on the shares of socio-demographic characteristics are calculated for the respective distance bands in the same way as for location $l$.

Figure 2: Map Section of Vienna


Figure 3: Map Section of Vienna with Grid Pattern

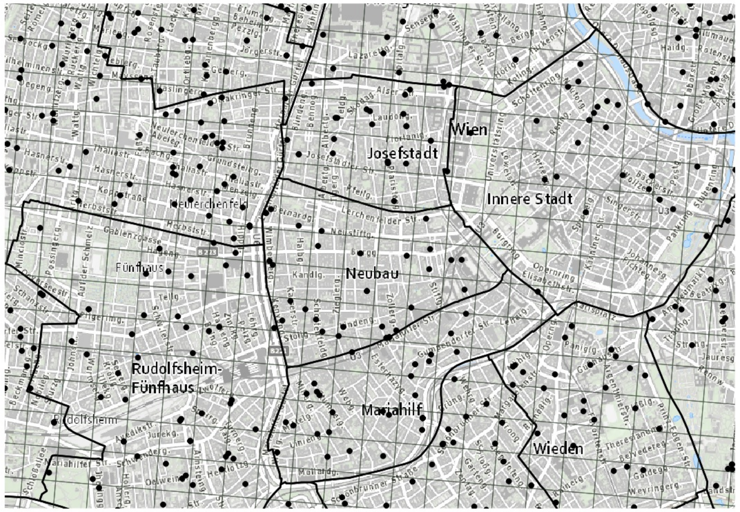



Figure 4: Spatial Distribution of Children Aged $\leq 5$

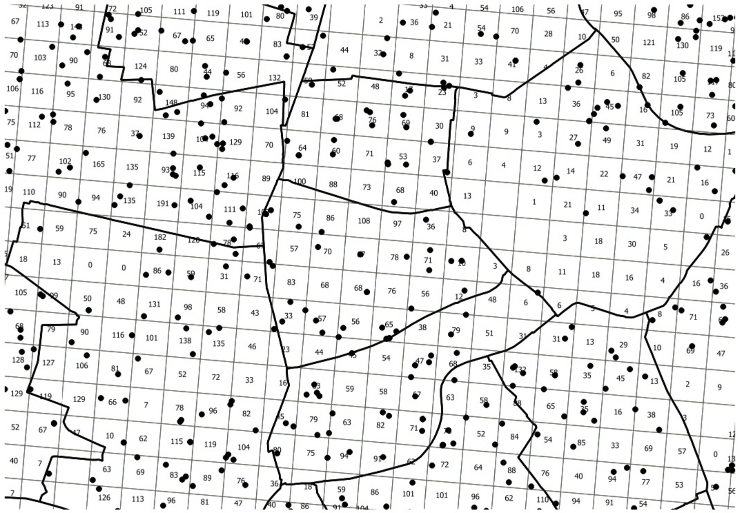
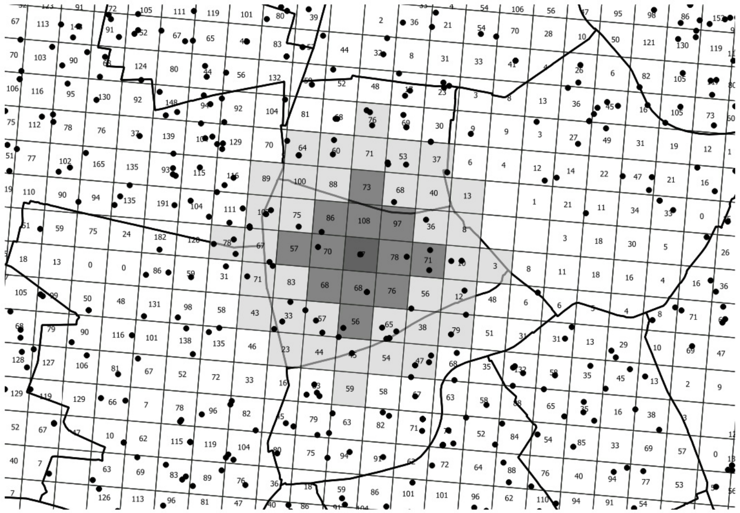


Figure 5: Spatial Aggregation of Variables



Notes: Figures 2 to 5 all illustrate the same map section of Vienna. Black dots denote the locations of day care centers in 2014 and bold lines indicate district borders. The 250m × 250m grid cells in Figures 3 to 5 are depicted by thin lines. The figures in the centers of all grid cells in Figures 4 and 5 indicate the number of children younger than six living in the respective grid cell in 2014. Locations without figures are unpopulated. In Figure 5, distance bands $d_1$ and $d_2$ around a particular location (dark gray) are colored in gray and light gray, respectively.

To relate the explanatory variables to both market structure and market dynamics we regress (i) the number of nursery groups (level) and (ii) the change in the number of nursery groups (net entry) at location $l$ on the explanatory variables calculated at the respective location and at various distance bands around this location. To estimate the number of nursery groups we are interested in the following relationship:

$$E(N_{lt}^{l}|\boldsymbol{X}_{lt}, c_r) = f(\boldsymbol{X}_{lt}\boldsymbol{\beta} + c_r) \tag{1}$$

with $N_{lt}^{l}$ as the number of nursery groups at location $l$ at time $t$. The subscript $l$ indicates that the variable is calculated at (or around) location $l$, while the superscript $l$ depicts that only characteristics of location $l$ are utilized to calculate this variable. The variable $c_r$ denotes region-specific fixed effects and the vector $\boldsymbol{X}_{lt}$ comprises all other explanatory variables calculated at location $l$ or at different distance bands around the respective location, with vector $\boldsymbol{\beta}$ including the respective parameters to be estimated. $\boldsymbol{X}_{lt}$ can be split into variables calculated at location $l$ ($\boldsymbol{X}_{lt}^{l}$) and at distance bands $d_1$ ($\boldsymbol{X}_{lt}^{d_1}$), $d_2$ ($\boldsymbol{X}_{lt}^{d_2}$) and $d_3$ ($\boldsymbol{X}_{lt}^{d_3}$). Thus, $\boldsymbol{X}_{lt}\boldsymbol{\beta} = \boldsymbol{X}_{lt}^{l}\boldsymbol{\beta}^{l} + \boldsymbol{X}_{lt}^{d_1}\boldsymbol{\beta}^{d_1} + \boldsymbol{X}_{lt}^{d_2}\boldsymbol{\beta}^{d_2} + \boldsymbol{X}_{lt}^{d_3}\boldsymbol{\beta}^{d_3}$, with $\boldsymbol{\beta}^{l}$, $\boldsymbol{\beta}^{d_1}$, $\boldsymbol{\beta}^{d_2}$, $\boldsymbol{\beta}^{d_2}$ as the parameter vectors to be estimated. $f(.)$ denotes some function, depending on the method applied to estimate the relationship of interest. As the number of nursery groups is count data, equation (1) is estimated by a Poisson regression model and function $f(.)$ is the therefore the exponential function.

Note that we refrain from including the (lagged) number of nursery groups in location $l$ and in the distance bands around this location in the main specification, because the location of nursery groups exhibits a strong path dependency. It is therefore difficult to isolate the negative competitive effect exerted by other (rival) nursery groups on market entry. However, we will pick up this issue again in the sensitivity analysis and address the competitive effect of child care institutions in the vicinity directly when investigating market dynamics, which we turn to now.

When estimating (net) market entry, the Poisson regression framework cannot be applied, because net entry can also be negative, while $\exp(\boldsymbol{X}_{lt}\boldsymbol{\beta} + c_r)$ can only be positive. We thus follow Cameron and Trivedi (2005), suggesting that '[o]rdered models ... are particularly useful when the count can also take negative values as may occur when modeling a net change, such as the net change in the number of firms in an industry' (p. 682), and use an ordered probit model to estimate the net change of the number of nursery groups. The relationship to be estimated can thus be stated as follows:

$$E(\Delta N^l_{l,t+1} | \boldsymbol{N}_{lt}, \boldsymbol{X}_{lt}, c_r) = g(\boldsymbol{N}_{lt}\boldsymbol{\alpha} + \boldsymbol{X}_{lt}\boldsymbol{\gamma} + c_r) \tag{2}$$

with $\Delta N^l_{l,t+1} = N^l_{l,t+1} - N^l_{lt}$ as the change of the number of nursery groups at location $l$ between year $t$ and $t+1$, and with $\boldsymbol{N}_{lt}\boldsymbol{\alpha} = N^l_{lt}\alpha^l + N^{d_1}_{lt}\alpha^{d_1} + N^{d_2}_{lt}\alpha^{d_2} + N^{d_3}_{lt}\alpha^{d_3}$ indicating the level of nursery groups at location $l$ ($N^l_{lt}$) and in the vicinity ($N^{d_1}_{lt}$, $N^{d_2}_{lt}$ and $N^{d_3}_{lt}$), as well as the corresponding parameters $\boldsymbol{\alpha} = (\alpha^l, \alpha^{d_1}, \alpha^{d_2}, \alpha^{d_3})'$ to be estimated. $\boldsymbol{\gamma}$ denotes the vector of parameters for the same explanatory variables, summarized in vector $\boldsymbol{X}_{lt}$, as in equation (1), and $c_r$ are again regional fixed effects, and $g(.)$ is some function.[11]

We mostly interpret the regression results in a descriptive rather than a causal way. While the results are expected to be mainly driven by local demand affecting the number and the entry of nursery groups, households could move to areas where the supply of child care is superior, inducing the variables on the number of children in the vicinity to be endogenous (reversed causality). We cannot account for this directly, because the data precludes tracking single households over time. When investigating the determinants of market dynamics reversed causality is less of an

[11] In an ordered probit framework the probability that $\Delta N^l_{l,t+1}$ takes the value $j \in J$, with $J$ as the entire set of net entry observed in any location, is characterized by $Pr(\Delta N^l_{l,t+1} = j) = Pr(\kappa_{j-1} < \boldsymbol{N}_{lt}\boldsymbol{\alpha} + \boldsymbol{X}_{lt}\boldsymbol{\gamma} + c_r + \upsilon_{lt} \leq \kappa_j)$. $\upsilon_{lt}$ is the error term and is assumed to be normally distributed. The probability that net entry $\Delta N^l_{l,t+1}$ takes the value $j$ equals the probability that the linear function of the explanatory variables plus the error term lies within the interval of the respective cut points $\kappa_{j-1}$ and $\kappa_j$. Ordered probit models provide estimates for these cut points in addition to the parameters. See e.g. Cameron and Trivedi (2005) for details.

issue, because it seems very unlikely that households move to neighborhoods because they expect day care institutions to enter there in the future. When analyzing the market structure (i.e. the number of nursery groups), however, causality may run in both directions.[12] We think that sorting induced by differences in the availability of child care is quantitatively relatively small, partly because empirical evidence suggests that, rather than local amenities, local employment opportunities seem to be more important for households with (potentially) economically active persons (Chen and Rosenthal, 2008; Scott, 2010). However, even if households do move to areas providing good child care, this does not impair our empirical strategy of defining the geographical market size: Spatially small markets may cause that either day care centers enter in the vicinity of children, or that households move to the proximity of child care institutions.

# 4 Results

## 4.1 Main Results: Market Structure

Table 2 reports the results of a regression model where the number of nursery groups in location $l$ is explained by characteristics of location $l$ as well as by exactly the same characteristics in the distance bands $d_1$, $d_2$ and $d_3$ (see equation (1)). Model [1] accounts for all possible differences between the 23 districts of Vienna and thus includes district-fixed effects, whereas Model [2] accounts for regional heterogeneity at an even smaller (registration district) level.

The parameter estimate on the number of children aged $\leq 5$ in location $l$ is positive, rather large and significantly different from zero. The coefficient of 0.00875 corresponds to a marginal effect (calculated at sample means) of 0.00594, which means that an increase in the number of children younger than 6 by one standard deviation (of 38.36, see Table 1) is associated with an increase in the number of

[12] See also the discussion on this issue in Waldfogel (2008).

nursery groups by 0.23 on average. The parameter estimate on the number of children in distance band $d_1$ (i.e. outside location $l$ but within a distance of up to 500m) is again positive and significantly different from zero at the 5% significance level, but the size of the coefficient drops to 0.00049. Therefore, the coefficient on the number of children in distance band $d_1$ is only 6% compared to the size of the parameter on children located directly at location $l$.[13] The size of the coefficient declines further to 0.00020 and 0.00002 in distance bands $d_2$ and $d_3$, respectively. Both parameter estimates are no longer significantly different from zero.

Turning to other parameter estimates reported in Table 2 reveals that most variables calculated at location $l$ have a statistically significant effect on the number of nursery groups at that location (at least at the 10% significance level). In line with our expectations a higher share of employed women living in the grid cell is associated with more nursery groups. While the parameter estimate on the number of jobs is positive and statistically significant, the size of the coefficient is rather small: The parameter estimate on the number of children is 55 times larger than the coefficient on the number of jobs, suggesting that day care institutions at the workplace location play a minor role compared to day care centers close to the place of residence. Easy accessibility, as measured by the number of subway stations, also has a positive influence on the number of nursery groups (at least at the 10% significance level). The share of residents with a high school diploma are negatively associated with day care provision, while the share of college graduates and the residents' countries of origin are not related with the number of nursery groups.

Most explanatory variables in the surrounding distance band $d_1$ are also statistically significant. The parameter estimates on the count variables (i.e. the number of jobs and subway stations) in distance band $d_1$ are smaller than in location $l$, although the distance decay is less pronounced compared to the number of children.

[13]Note that in Poisson regressions coefficients of different variables in the same regression can be compared that way, because the ratio between parameter estimates of two variables is the same as the ratio of the marginal effects of the respective variables (see Cameron and Trivedi, 2005, p. 669).

Only two parameter estimates for variables calculated for distance bands $d_2$ (up to 1 km) and $d_3$ (up to 2 km) are significantly different from zero at the 10%-level, and none at the 5% significance level. This finding is strengthened by a $\chi^2$-test on the joint significance of all variables within particular distance bands: While the null-hypothesis that the coefficients of all variables in a particular distance band are zero is clearly rejected for location $l$ and distance band $d_1$, it is not rejected for distance bands $d_2$ and $d_3$ at any conventional significance levels. We thus conclude that a distance of 500m is a reasonable threshold radius for the relevant catchment area, because demand indicators outside this area are not (significantly) related to location choices of the day care centers.

Comparing the two models presented in Table 2 reveals only marginal differences, and the results remain quite stable once heterogeneity between registration districts is controlled for.

Table 2: Regression Results regarding the Number of Nursery Groups

| | Model [1] | | | Model [2] | | |
|---|---|---|---|---|---|---|
| | Coeff. | Std. Err. | Sign. | Coeff. | Std. Err. | Sign. |
| Location $l$ | | | | | | |
| # of children aged ≤ 5 | 0.00875 | (0.00098) | *** | 0.00777 | (0.00097) | *** |
| # of jobs | 0.00016 | (0.00006) | *** | 0.00015 | (0.00006) | ** |
| # of subway stations | 0.31361 | (0.16669) | * | 0.29236 | (0.18371) | |
| Share employed women | 0.01644 | (0.00536) | *** | 0.01824 | (0.00568) | *** |
| Share COB Austria | 0.00378 | (0.00574) | | 0.00598 | (0.00623) | |
| Share COB other EU country | −0.01325 | (0.01332) | | −0.00787 | (0.01463) | |
| Share high school diploma | −0.02597 | (0.00985) | *** | −0.03243 | (0.01048) | *** |
| Share college degree | −0.00005 | (0.00822) | | 0.00524 | (0.00870) | |
| Distance band $d_1$ | | | | | | |
| # of children aged ≤ 5 | 0.00049 | (0.00024) | ** | 0.00072 | (0.00028) | ** |
| # of jobs | 0.00003 | (0.00002) | * | 0.00003 | (0.00002) | |
| # of subway stations | 0.12074 | (0.06526) | * | 0.10691 | (0.07577) | |
| Share employed women | 0.04210 | (0.01293) | *** | 0.04438 | (0.01516) | *** |
| Share COB Austria | 0.02044 | (0.00982) | ** | 0.02404 | (0.01125) | ** |
| Share COB other EU country | 0.11924 | (0.02977) | *** | 0.15353 | (0.03980) | *** |
| Share high school diploma | −0.06504 | (0.02794) | ** | −0.08240 | (0.03364) | ** |
| Share college degree | −0.00281 | (0.01939) | | 0.00192 | (0.02291) | |
| Distance band $d_2$ | | | | | | |
| # of children aged ≤ 5 | 0.00020 | (0.00013) | | 0.00023 | (0.00015) | |
| # of jobs | −0.00000 | (0.00001) | | −0.00001 | (0.00001) | |
| # of subway stations | 0.06177 | (0.04238) | | 0.06618 | (0.05146) | |
| Share employed women | −0.02166 | (0.01744) | | −0.00195 | (0.02475) | |
| Share COB Austria | 0.02712 | (0.01631) | * | 0.04063 | (0.02065) | ** |
| Share COB other EU country | 0.09350 | (0.06994) | | 0.11731 | (0.09505) | |
| Share high school diploma | 0.02271 | (0.03854) | | −0.00720 | (0.05732) | |
| Share college degree | −0.00529 | (0.02651) | | 0.00625 | (0.03585) | |
| Distance band $d_3$ | | | | | | |
| # of children aged ≤ 5 | 0.00002 | (0.00005) | | 0.00001 | (0.00007) | |
| # of jobs | −0.00000 | (0.00000) | | −0.00000 | (0.00000) | |
| # of subway stations | −0.01090 | (0.02369) | | −0.01375 | (0.02830) | |
| Share employed women | −0.04699 | (0.02543) | * | −0.00918 | (0.04827) | |
| Share COB Austria | −0.01229 | (0.01626) | | 0.01858 | (0.02749) | |
| Share COB other EU country | 0.06123 | (0.08541) | | 0.17962 | (0.16573) | |
| Share high school diploma | −0.03838 | (0.06301) | | −0.08244 | (0.11802) | |
| Share college degree | −0.00190 | (0.03397) | | 0.01032 | (0.05978) | |
| Constant | −4.38595 | (2.23486) | ** | −11.44703 | (4.16057) | *** |
| # of observations | 24,528 | | | 24,528 | | |
| Time effects | Yes – 7 | | | Yes – 7 | | |
| District effects | Yes – 22 | | | No | | |
| Registration district effects | No | | | Yes – 242 | | |
| Log-likelihood | –39,906 | | | –36,089 | | |
| $\chi^2$ test statistic for $l$ | 107.55 ($df = 8$) [$p = 0.000$] | | | 91.76 ($df = 8$) [$p = 0.000$] | | |
| $\chi^2$ test statistic for $d_1$ | 36.17 ($df = 8$) [$p = 0.000$] | | | 32.63 ($df = 8$) [$p = 0.000$] | | |
| $\chi^2$ test statistic for $d_2$ | 9.21 ($df = 8$) [$p = 0.325$] | | | 7.01 ($df = 8$) [$p = 0.536$] | | |
| $\chi^2$ test statistic for $d_3$ | 11.39 ($df = 8$) [$p = 0.181$] | | | 3.97 ($df = 8$) [$p = 0.860$] | | |

Notes: Standard errors are reported in parentheses and are based on standard errors that are clustered at the grid cell level. *** significant at 1 %, ** significant at 5 %, * significant at 10 % level. $l$ denotes values of the respective variables in the grid cell, $d_1$ values aggregated or averaged for all cells within a distance [250m, 500m], $d_2$ values aggregated or averaged for all cells within a distance (500m, 1,000m], and $d_3$ values aggregated or averaged for all cells within a distance (1,000m, 2,000m]. The reported $\chi^2$-statistics test the hypotheses whether the coefficients of all variables for a particular distance band are jointly significant (H0: all coefficients are zero).

## 4.2 Main Results: Market Dynamics

Table 3 presents regression results regarding the change in the number of nursery groups in location $l$. In contrast to the models on market structure reported in Table 2, the regressions on net entry include the number of nursery groups at location $l$ as well as in its vicinity to account for competitive pressure. As expected, the number of nursery groups has a negative effect on net entry in location $l$: If the number of nursery groups is high, net market entry will be lower (or even negative), ceteris paribus. The effect diminishes in absolute terms as the distance from location $l$ increases, but the parameter estimates are significantly different from zero at least at the 10% significance level for both distance bands $d_1$ and $d_2$. The coefficients of the other variables are qualitatively similar to those of Table 2: The number of children, jobs and subway stations in location $l$ are not only positively associated with the level, but also with the change in the number of nursery groups at this location. Again, the size of the estimated coefficients diminishes with distance, whereby particularly the number of children is significantly related to net entry, even in distance band $d_2$ (up to 1km). The share of employed women in location $l$ and distance band $d_1$ is also associated with higher market entry.

$\chi^2$ test statistics on the joint significance of all variables at a particular distance band again suggest that the variables calculated at location $l$ or at distance band $d_1$ are jointly significant, while the null-hypothesis stating that the coefficients of all variables in distance band $d_3$ are zero cannot be rejected. The results are less clear-cut for distance band $d_2$, where the null-hypothesis is rejected in one specification (Model [4]), but not in the other one (Model [3]).

Table 3: Regression Results regarding the Change of the Number of Nursery Groups

| | Model [3] | | | Model [4] | | |
|---|---|---|---|---|---|---|
| | Coeff. | Std. Err. | Sign. | Coeff. | Std. Err. | Sign. |
| Location $l$ | | | | | | |
| # of nursery groups | −0.05648 | (0.00667) | *** | −0.06885 | (0.00673) | *** |
| # of children aged $\leq 5$ | 0.00200 | (0.00041) | *** | 0.00204 | (0.00038) | *** |
| # of jobs | 0.00005 | (0.00002) | ** | 0.00005 | (0.00002) | ** |
| # of subway stations | 0.18956 | (0.07500) | ** | 0.22667 | (0.07721) | *** |
| Share employed women | 0.00206 | (0.00086) | ** | 0.00270 | (0.00090) | *** |
| Share COB Austria | −0.00138 | (0.00116) | | −0.00087 | (0.00120) | |
| Share COB other EU country | 0.00042 | (0.00247) | | 0.00122 | (0.00252) | |
| Share high school diploma | −0.00131 | (0.00146) | | −0.00240 | (0.00150) | |
| Share college degree | 0.00093 | (0.00138) | | 0.00105 | (0.00136) | |
| Distance band $d_1$ | | | | | | |
| # of nursery groups | −0.00299 | (0.00173) | * | −0.00803 | (0.00211) | *** |
| # of children aged $\leq 5$ | 0.00019 | (0.00008) | ** | 0.00029 | (0.00009) | *** |
| # of jobs | 0.00001 | (0.00001) | * | 0.00001 | (0.00001) | * |
| # of subway stations | 0.06761 | (0.02106) | *** | 0.07534 | (0.02754) | *** |
| Share employed women | 0.00446 | (0.00187) | ** | 0.00707 | (0.00198) | *** |
| Share COB Austria | 0.00069 | (0.00111) | | 0.00129 | (0.00123) | |
| Share COB other EU country | 0.00197 | (0.00472) | | 0.00544 | (0.00521) | |
| Share high school diploma | −0.00504 | (0.00391) | | −0.00828 | (0.00438) | * |
| Share college degree | 0.00071 | (0.00289) | | 0.00111 | (0.00320) | |
| Distance band $d_2$ | | | | | | |
| # of nursery groups | −0.00220 | (0.00116) | * | −0.00450 | (0.00147) | *** |
| # of children aged $\leq 5$ | 0.00012 | (0.00004) | *** | 0.00018 | (0.00006) | *** |
| # of jobs | −0.00000 | (0.00000) | | 0.00000 | (0.00000) | |
| # of subway stations | 0.02424 | (0.01354) | * | 0.02716 | (0.01906) | |
| Share employed women | 0.00225 | (0.00259) | | 0.00405 | (0.00304) | |
| Share COB Austria | 0.00050 | (0.00152) | | 0.00497 | (0.00175) | *** |
| Share COB other EU country | −0.00546 | (0.00724) | | −0.00542 | (0.00815) | |
| Share high school diploma | −0.00137 | (0.00572) | | −0.00315 | (0.00657) | |
| Share college degree | 0.00549 | (0.00405) | | 0.00541 | (0.00448) | |
| Distance band $d_3$ | | | | | | |
| # of nursery groups | 0.00057 | (0.00060) | | 0.00069 | (0.00072) | |
| # of children aged $\leq 5$ | −0.00000 | (0.00002) | | 0.00000 | (0.00002) | |
| # of jobs | −0.00000 | (0.00000) | | −0.00000 | (0.00000) | |
| # of subway stations | −0.00750 | (0.00705) | | −0.01171 | (0.00985) | |
| Share employed women | −0.00418 | (0.00356) | | −0.00030 | (0.00401) | |
| Share COB Austria | 0.00147 | (0.00280) | | 0.00292 | (0.00329) | |
| Share COB other EU country | −0.00335 | (0.01078) | | −0.00280 | (0.01410) | |
| Share high school diploma | −0.00179 | (0.00884) | | −0.00580 | (0.01150) | |
| Share college degree | −0.00111 | (0.00527) | | −0.00153 | (0.00645) | |
| # of observations | 21,462 | | | 21,462 | | |
| Time effects | Yes – 6 | | | Yes – 6 | | |
| District effects | Yes – 22 | | | No | | |
| Registration district effects | No | | | Yes – 242 | | |
| Log-likelihood | –10,830 | | | –10,748 | | |
| Pseudo-$R^2$ | 0.0209 | | | 0.0283 | | |
| $\chi^2$ test statistic for $l$ | 101.52 ($df = 9$) [$p = 0.000$] | | | 137.77 ($df = 9$) [$p = 0.000$] | | |
| $\chi^2$ test statistic for $d_1$ | 25.77 ($df = 9$) [$p = 0.002$] | | | 38.84 ($df = 9$) [$p = 0.000$] | | |
| $\chi^2$ test statistic for $d_2$ | 13.62 ($df = 9$) [$p = 0.136$] | | | 22.14 ($df = 9$) [$p = 0.008$] | | |
| $\chi^2$ test statistic for $d_3$ | 10.94 ($df = 9$) [$p = 0.280$] | | | 7.29 ($df = 9$) [$p = 0.607$] | | |

Notes: Standard errors are reported in parentheses and are based on standard errors that are clustered at the grid cell level. *** significant at 1 %, ** significant at 5 %, * significant at 10 % level. $l$ denotes values of the respective variables in the grid cell, $d_1$ values aggregated or averaged for all cells within a distance [250m, 500m], $d_2$ values aggregated or averaged for all cells within a distance (500m, 1,000m], and $d_3$ values aggregated or averaged for all cells within a distance (1,000m, 2,000m]. The reported $\chi^2$-statistics test the hypotheses whether the coefficients of all variables for a particular distance band are jointly significant (H0: all coefficients are zero).

### 4.3 Sensitivity Analysis

In order to confirm that the results are not driven by spatial dependence or regional heterogeneity not accounted for in the main specifications, by means of particular sub-samples or omitted variables, this section provides a number of sensitivity analyses. We group the robustness exercises by potential limitations of our analysis that may bias the results and/or affect the interpretation of the findings. The signs and statistical significance of most explanatory variables of interest are hardly affected by the model variations provided in the sensitivity analyses, strengthening the confidence in the main conclusion regarding the geographical market size of about 500m. The primary findings of the robustness exercises are thus only briefly discussed in the main part of this article, while the regression results are relegated to Appendix B.

*Spatial Dependence:* We are aware of the locations (the grid cells) being spatially dependent. The issue of spatial dependence in local markets is in particular stressed by Yan et al. (2014), who point out that discarding the issues of spatial dependence and spatial heterogeneity 'may affect the validity of the conclusions' (p. 250). The regression results reported in Tables 2 and 3 account for spatial dependence and spatial heterogeneity by considering demand (and partly supply) characteristics of other locations in the vicinity, as well as by including regional fixed effects (at different regional levels), when estimating the number and the entry of nursery groups. Nevertheless, this may not account for all spatial dependence in the data. We thus perform Moran's $I$ tests, developed by Moran (1950),[14] as well as heteroscedasticity robust LM tests, proposed by Born and Breitung (2011), based on the residuals of the regressions reported in Tables 2 and 3. The respective test statistics show that there is no further spatial autocorrelation left in the residuals, suggesting that spatial dependence and spatial heterogeneity are adequately accounted for in the empirical models. A detailed description of the performed tests is provided in Appendix A.

[14]See, e.g. Anselin (1988) for an introduction to spatial dependence and spatial econometrics.

*Growth in the Number vs. the Capacity of Providers:* In the main part of the article the number of nursery groups serves as the dependent variable. Market entry at some location, however, could either occur by new day care centers entering this location, or by existing day care institutions expanding their capacity (i.e. opening up additional nursery groups within an existing day care center). Thus, the number of day care centers is used rather than the number of nursery groups in this robustness exercise, to evaluate whether the results are sensitive to this distinction. The respective results are reported in Table B.1 (market structure) and Table B.2 (market dynamics) in Appendix B. The findings are qualitatively very similar compared to the main results: In all four specifications the number of children at location $l$ and in distance band $d_1$ are positively and significantly related to the number or the change in the number of day care centers (at least at the 5% significance level), while the coefficients on the number of children further away are much smaller. Again, the results indicate a strong distance decay in the size of the coefficients similar to the main results reported in Sections 4.1 and 4.2 above.

*Relaxing Distance Bands:* Throughout the analysis demand indicators are calculated in distance bands around each location to allow for heterogeneous effects of the number of children (in different distance bands) on the number of nursery groups. However, this approach restricts the coefficients to be the same for all children living in different locations within each distance band, irrespective of whether the children's place of residence is close to the inner or near the outer border of the respective distance band. Thus, for example, demand indicators in locations at a distance of 550m are restricted to have the same influence as these variables in grid cells at a distance of 950m, because both locations are grouped in the same distance band $d_2$. This restriction is relaxed in this sensitivity analysis and the number of children is included at all possible distances separately. As distances are calculated as Euclidean distances between the centroids of grid cells, we get 30 distinct distances for all locations within 2km. As explanatory variables the regression thus includes the number of children at the respective location, at locations within a distance of exactly 250m, within a distance of exactly $\sqrt{250^2 + 250^2}$m $\approx$ 354m, and so on. We keep the distance bands for calculating all other explanatory variables, so that the number of parameters to be estimated does not grow too large, and include registration district and time fixed effects. The model thus resembles Model [2] in Table 2, but includes information on the number of children in the vicinity in a more detailed way.

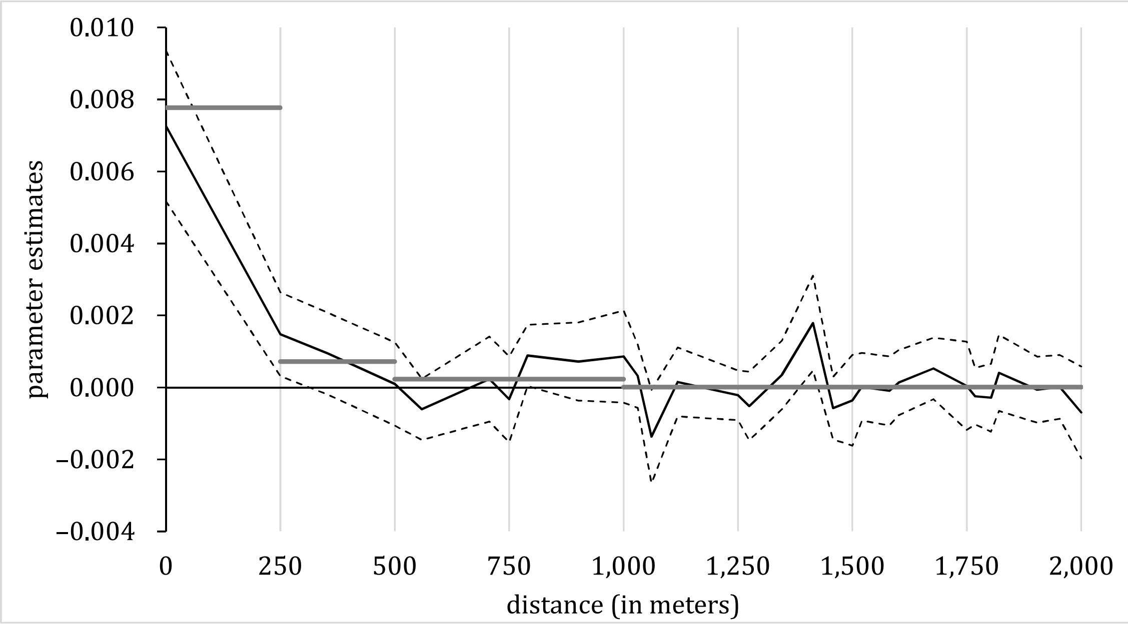
Estimation results regarding the number of children depending on the exact distance between the children's residences and day care centers' locations are illustrated in Figure 6 below (black solid line), along with the 95% confidence interval (dashed lines) and the point estimates of the main specification [2] (thick gray line).[15] The parameter estimate on the number of children in the same location $l$ is 0.0073, and thus very similar to the main specifications in Table 2. As distance increases, the size of the estimated coefficients declines quickly. The parameter estimate for neighboring locations (with a distance of 250m) is significantly positive, but the coefficients are not significantly different from zero for 354m and 500m. For distances between 500m and 2km the estimated parameters fluctuate around zero without an obvious trend, and for only three out of 26 variables the corresponding parameter estimates are statistically different from zero at the 10% significance level.

*Heterogeneous Market Size:* Throughout the analysis we implicitly assumed the geographical market size to be identical throughout the entire city of Vienna. While regional heterogeneity was controlled for by including district or registration district effects, the distance decay was restricted to be the same for all locations analyzed. As the outskirts of the Vienna are rather suburban areas, the sample is split in an 'urban' and a 'suburban' region, based on the population density around each location.[16] Unsurprisingly, the results reported in Table B.3 in Appendix B suggest

[15] Both parameter estimates and significance levels of all other variables are very similar compared to the results reported in Model [2] in Table 2. The results are thus not reported, but available from the authors upon request.

[16] The sample was split based on the median population within a distance of 2km around all

Figure 6: Parameter Estimates on the Distance-Specific Number of Children




Notes: The figure graphically illustrates the parameter estimates on the number of children in specific distances. The solid line is based on the point estimates and the dotted lines indicate the 95% interval. The thick gray line indicates the point estimates on the number of children in location $l$ and distance bands $d_1$, $d_2$ and $d_3$ of the main specification (reported in the second column of Table 2).

that distance is more important in more densely populated areas. In 'urban' areas the number of children in location $l$ is significantly related to the number of nursery groups, while this is not the case for the number of children in distance band $d_1$. In 'suburban' areas, the parameters on the number of children both in location $l$ and distance band $d_1$ are significantly positive. Additionally, the point estimate on the number of children in location $l$ is 26 times as large as the point estimate on this variable in distance band $d_1$ in 'urban' areas, while the ratio between these two point estimates is only eight for 'suburban' regions of the city. $\chi^2$ tests again suggest that the variables calculated at location $l$ are jointly significant at the 1% level, while the parameter estimates on the variables calculated at the more distant bands $d_2$ and $d_3$ are not jointly significant at any reasonable significance levels. The variables in the narrowest distance band $d_1$ are jointly significant for 'suburban' regions, but

locations in order to get two sub-samples of similar size.

not for 'urban' areas.

*Omitted Variables* When analyzing the market structure omitted variables may bias the results, due to neglecting (i) variables measuring competitive pressure or (ii) other, unobserved variables. (i) When investigating the market structure, the (lagged) number of nursery groups both at location $l$ as well as in the various distance bands around this location is omitted (see Table 2), because the location of day care institutions exhibits a strong degree of path dependency. It is therefore difficult to isolate the (negative) competition effect. In this sensitivity analysis one regression includes the lagged number of nursery groups (see Model [11] in Table B.4 in Appendix B). The results indicate a positive relation between the number of nurseries at location $l$ in the previous year and the number of groups in the current year, due to path dependency, as expected. The coefficient on the number of nursery groups in year $t-1$ in distance band $d_1$ is significantly negative, while the parameter estimates on this variable calculated at the more distant bands $d_2$ and $d_3$ are not significantly different from zero at the 5% level. The estimated parameters on the number of children are again significantly positive for location $l$ and distance band $d_1$, but insignificant for areas further away. $\chi^2$ tests again suggest that the geographical market size is about 500 meters.

(ii) The location choice of day care centers may be influenced by variables that are unobservable at this regionally highly disaggregated grid level. While the data sample used in the analysis allows for capturing spatial variation in demand quite accurately, this is not the case for cost differentials. There is no systematic difference in wages in this industry across locations, because wages are based on collective agreements at a federal level, whereas housing and rental prices differ substantially and may thus influence entry decisions at particular locations. Further, it is possible that rental prices and the population density of children are correlated (in one way or the other).[17] In the main specifications regional heterogeneity is accounted for by

[17]Note that rental prices at location $l$ should only affect entry at that location directly, but not at other locations in the vicinity, and should therefore influence parameter estimates (if at all) only

including fixed effects at the district or at the registration district level. In this final sensitivity analysis we include location-fixed effects instead to control for unobserved heterogeneity at an even smaller spatial scale. The results of this specification, presented in Model [12] of Table B.4 in Appendix B, suggest that the number of children as well as the number of jobs, and furthermore the existence of a subway station at location $l$ are positively correlated with the number of nursery groups, as in the main specifications reported in Table 2 above.

# 5 Discussion

In this article we investigate the geographical market size for child care services in an urban context. It is often assumed that local markets in this industry are very narrow due to high transportation costs. Previous studies on child care choices based on survey data present evidence that the location of day care centers is of key importance for parents, but this literature provides little attempts to quantify the commuting distance parents are willing to endure to travel to day care centers. Empirical contributions analyzing firm and market conduct in this industry use different regional entities to delineate local markets, but usually devote little attention to identifying the catchment area of child care facilities.

We contribute to this literature by defining the size of geographical markets based on the location choice of day care centers. To do so, we utilize a panel of all day care facilities in Vienna covering nearly a decade as well as geographically extremely disaggregated data on demand at the 250m $\times$ 250m grid cell level. The number and the entry of nursery groups in a grid cell are estimated as a function of potential demand (the number of children under the age of six) in the respective location and in distance bands around that location. We use the notion that researchers can expect to find a strong relationship between the number of firms in a location

on variables calculated at a particular location, but not in the respective distance bands around this location.

and the number of consumers in a specific geographical area, if demand is drawn from that area. The results suggest that potential local demand is strongly and significantly correlated with market structure and market entry, but the size of the coefficients diminishes quickly with increasing distance. The results also indicate a strong distance decay concerning the competitive effect resulting from rival providers in the vicinity. From these results we can infer the day care centers' catchment areas, suggesting that local markets are indeed geographically very small (about 500m), consistent with the (suspected) high transportation costs.

Note that the small size of geographical markets may stem from parents either not willing or not required to travel far to day care centers (due to a dense network of providers, for example). Thus, the small size of local markets must not be confused with a monopolist's catchment area. However, day care centers are different from each other in a number of (non-spatial) dimensions (e.g. opening hours or types of nursery groups). The main finding of our study, i.e. that day care centers draw their demand from a very narrow area, suggests that parents are only willing to accept a short extra distance to get an otherwise (i.e. in non-spatial dimensions) more preferable provider.

From a policy perspective, these results lead to the conclusion that in order to further encourage parents to opt for formal child care arrangements as envisioned the EU's 'Lisbon Strategy', it is important to offer a sufficiently dense network of providers. The Lisbon Strategy formulates clear objectives related to child care provision to 'remove disincentives to female labor force participation' (European Council, Barcelona, 2002, p. 12) and ranks high on the political agenda. In addition to an increase in female labor participation, 'childcare has gained renewed attention as a key part of the social infrastructure and increasingly as an educational resource in its own right' (Gallagher, 2017, p. 1). The conclusions of our empirical analysis suggest that the positive effects ascribed to child care provision will only be fully realized if the service is provided within a sufficiently short distance, close to the parents' homes. From an urban planning perspective it is important to note that finding suitable premises has been found to be one of the main obstacles when establishing new day care centers. The very small catchment areas for child care markets thus pose an additional challenge to be kept in mind.

It is, however, important to stress that the determined local market size of around 500m certainly depends on available modes of transportation in an area. In Vienna, the small market size points towards the fact that a common mode of transportation is walking, while common or preferred modes of travel may, of course, differ in other urban settings. In more rural areas, transportation costs are smaller due to different transport modes and less congested roads, and geographical markets are thus expected to be larger.

The main contribution of this article, related to the more general empirical literature investigating firm entry and market conduct, is to provide a simple and rather easily replicable approach to delineate geographical markets. Our data on potential demand is innovative, as regional statistical grid units provide a rather new way of organizing and utilizing spatial data that has received only limited attention in economic applications so far. However, this kind of data has become available in a number of countries in recent years. Besides Austria similar data are available e.g. for Finland (see Eerola and Lyytikäinen, 2015), Sweden and Norway (see Tammilehto-Luode et al., 2000) and will become even more widespread in the near future. While grid data are very detailed from a spatial perspective, they are easily accessible, whereby privacy protection is not a relevant issue, in line with the modest data requirements common to entry models.

While drawing on easily accessible data is one advantage of our study, linking consumer and provider data at an individual level would be a fruitful extension of our study. This information is recorded by the Viennese government, but not accessible to the authors due to privacy issues. The full population of all day care centers used in this analysis covers the (exhaustive) choice set, and 'linked consumer-provider' data would reveal information about actual parental choices regarding child care. This would allow researchers to estimate consumers' substitution patterns directly, as done, e.g., by Gowrisankaran et al. (2011, 2018) for the hospital market. With this data one could gain insights into whether preferences of having child care facilities nearby vary systematically depending on parents' socio-economic statuses or (mothers' or fathers') working hours, and how highly parents value e.g. prolonged opening hours, extraordinary activities or groups with a smaller number of children relative to spatial proximity.

**Acknowledgements:**
We thank Stefanie Peer, Biliana Yontcheva, conference participants of the 2018 Winter Seminar of the Gesellschaft für Regionalforschung and the 2018 ESPAnet Austria Research Conference as well as participants at workshops and seminars at the Vienna University of Economics and Business, at the University of Innsbruck and at the Austrian Institute of Economic Research (WIFO Vienna) for helpful comments on earlier versions of the article. We also thank the Magistratsabteilung 23 (City of Vienna) for providing data on child care institutions and Julia Groiß helping to link the data sets. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

# References

Aguirregabiria, V. and Suzuki, J. (2016). *Empirical games of market entry and spatial competition in retail industries*, pages 201–232. Edward Elgar Publishing, Cheltenham, UK; Northampton, MA; USA.

Anselin, L. (1988). *Spatial econometrics. Methods and models.* Kluwer Academic Publishers, Dordrecht, NL. 1. Edition.

Baker, M., Gruber, J., and Milligan, K. (2008). Universal child care, maternal labor supply, and family well-being. *Journal of Political Economy*, 116(4):709 – 745.

Bick, A. (2016). The quantitative role of child care for female labor force participation and fertility. *Journal of the European Economic Association*, 14(3):639 – 668.

Born, B. and Breitung, J. (2011). Simple regression-based tests for spatial dependence. *The Econometrics Journal*, 14(2):330–342.

Bresnahan, T. F. and Reiss, P. C. (1990). Entry in monopoly markets. *The Review of Economic Studies*, 57(4):531–553.

Bresnahan, T. F. and Reiss, P. C. (1991). Entry and competition in concentrated markets. *Journal of Political Economy*, 99(5):977–1009.

Bundesministerium für Digitalisierung und Wirtschaftsstandort (2018). U-Bahnnetz Bestand Wien. (Underground system in Vienna). Technical report. https://www.data.gv.at/katalog/dataset/stadt-wien_ubahnnetzbestandwien; Accessed: July 17, 2018.

Cameron, A. and Trivedi, P. (2005). *Microeconometrics: Methods and Applications*. Cambridge University Press.

Chen, Y. and Rosenthal, S. S. (2008). Local amenities and life-cycle migration: Do people move for jobs or fun? *Journal of Urban Economics*, 64(3):519 – 537.

Cleveland, G. and Krashinsky, M. (2009). The nonprofit advantage: Producing quality in thick and thin child care markets. *Journal of Policy Analysis and Management*, 28(3):440–462.

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

Compton, J. and Pollak, R. A. (2014). Family proximity, childcare, and women's labor force attachment. *Journal of Urban Economics*, 79(Supplement C):72–90.

Datta, S. and Sudhir, K. (2013). Does reducing spatial differentiation increase product differentiation? Effects of zoning on retail entry and format variety. *Quantitative Marketing and Economics*, 11(1):83–116.

Eerola, E. and Lyytikäinen, T. (2015). On the role of public price information in housing markets. *Regional Science and Urban Economics*, 53:74 – 84.

European Commission (2013). Barcelona targets. The development of childcare facilities for young children in Europe with a view to sustainable and inclusive growth. Report.

European Council, Barcelona (2002). Presidency conclusions. Report. http://www.consilium.europa.eu/uedocs/cms_data/docs/pressdata/en/ec/71025.pdf; Accessed: July 17, 2018.

Gallagher, A. (2017). The business of care: Marketization and the new geographies of childcare. *Progress in Human Geography*, forthcoming.

Gowrisankaran, G., Lucarelli, C., Schmidt-Dengler, P., and Town, R. (2011). The impact of the medicare rural hospital flexibility program on patient choice. *International Journal of Industrial Organization*, 29(3):342–344.

Gowrisankaran, G., Lucarelli, C., Schmidt-Dengler, P., and Town, R. (2018). Can amputation save the hospital? The impact of the medicare rural flexibility program on demand and welfare. *Journal of Health Economics*, 58:110–122.

Haan, P. and Wrohlich, K. (2011). Can child care policy encourage employment and fertility? Evidence from a structural model. *Labour Economics*, 18(4):498–512.

Herbst, C. M. and Barnow, B. S. (2008). Close to home: A simultaneous equations model of the relationship between child care accessibility and female labor force participation. *Journal of Family and Economic Issues*, 29(1):128–151.

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

Johansen, A., Leibowitz, A., and Waite, L. (1996). The importance of child-care characteristics to choice of care. *Journal of Marriage and the Family*, 58(3):759–772.

Kawabata, M. and Abe, Y. (2018). Intra-metropolitan spatial patterns of female labor force participation and commute times in Tokyo. *Regional Science and Urban Economics*, 68:291 – 303.

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

Lam, M., Klein, S., Freisthler, B., and Weiss, R. E. (2013). Child center closures: Does nonprofit status provide a comparative advantage? *Children and Youth Services*, 35(3):525–534.

Matsa, D. A. (2011). Competition and product quality in the supermarket industry. *The Quarterly Journal of Economics*, 126(3):1539–1591.

Michelson, W. (1985). *From Sun to Sun: Daily Obligations and Community Structure in the Lives of Employed Women and Their Families*. Rowman & Allanheld.

Moran, P. A. P. (1950). Notes on continuous stochastic phenomena. *Biometrika*, 37(1/2):17–23.

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

Noailly, J. and Visser, S. (2009). The impact of market forces on child care provision: Insights from the 2005 child care act in the Netherlands. *Journal of Social Policy*, 38(03):477–498.

Owens, M. F. and Rennhoff, A. D. (2014). Provision and price of child care services: For-profits and nonprofits. *Journal of Urban Economics*, 84(Supplement C):40–51.

Queralt, M. and Witte, A. (1999). Estimating the unmet need for services: A middling approach. *Social Service Review*, 73(4):524–559.

Queralt, M. and Witte, A. D. (1998). Influences on neighborhood supply of child care in Massachusetts. *Social Service Review*, 72(1):17–46.

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

Schinkowitsch, D. (2014). *Herausforderungen in der Gründung von elternverwalteten Kindergruppen in Wien. (Challenges for founding day care institutions in Vienna)*. FH Campus Wien. Master thesis.

Scott, A. J. (2010). Jobs or amenities? Destination choices of migrant engineers in the USA. *Papers in Regional Science*, 89(1):43–63.

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

Small, M. and Stark, L. (2005). Are poor neighborhoods resource deprived? A case study of childcare centers in New York. *Social Science Quarterly*, 86(5):1013–1036.

Stadt Wien (2015). Elementare Bildung und Betreuung in Wien. (Early childhood education and care in Vienna). Report, Wirtschaft, Arbeit und Statistik (Magistratsabteilung 23). https://www.wien.gv.at/statistik/pdf/statistik-journal-wien-1-2015.pdf; Accessed: July 17, 2018.

Statistik Austria (2015). *Kindertagesheimstatistik. (Statistic on child care facilities)*. Verl. Österreich, Wien.

Statistik Austria (2018). Bevölkerungsstand Wien 2017. (Population data for Vienna 2017). http://statcube.at/statistik.at/ext/statcube/jsf/tableView/tableView .xhtml; Accessed: April 13, 2018.

Tammilehto-Luode, M., Backer, L., and Rogstad, L. (2000). Grid data and area delimitation by definition towards a better European territorial statistical system. *Statistical Journal*, 17(2):109 – 117.

Teszenyi, E. and Hevey, D. (2015). Age group, location or pedagogue: Factors affecting parental choice of kindergartens in Hungary. *Early Child Development and Care*, 185(11-12, SI):1961–1977.

Waldfogel, J. (2008). The median voter and the median consumer: Local private goods and population composition. *Journal of Urban Economics*, 63(2):567–582.

Wiener Familienbund (2005). Wiener Kindergartenführer. (Vienna's day care guide). Report.

Wiener Familienbund (2011). Wiener Kindergartenführer. (Vienna's day care guide). Report.

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

Yamauchi, C. (2010). The availability of child care centers, perceived search costs and parental life satisfaction. *Review of Economics of the Household*, 8(2):231–253.

Yan, J., Guo, C., and Paarlberg, L. E. (2014). Are nonprofit antipoverty organizations located where they are needed? A spatial analysis of the Greater Hartford region. *The American Statistician*, 68(4):243–252.

Zhu, T. and Singh, V. (2009). Spatial competition with endogenous location choices: An application to discount retailing. *Quantitative Marketing and Economics*, 7:1–35.

# Appendix A. Spatial Dependence

This section provides a detailed description of the diagnostic tests on the residuals of the Poisson regressions on the number of nursery groups (see equation (1)) and of the ordered probit regressions on the change in the number of nursery groups (see equation (2)). The diagnostic tests investigate if any spatial structure is not accounted for in the models and thus left in the residuals. For the Poisson regressions we take the Pearson (standardized) residuals defined as $\frac{N^l_{lt}-\hat{N}^l_{lt}}{\sqrt{\hat{N}^l_{lt}}}$ (see, e.g. Cameron and Trivedi, 2005), with $\hat{N}^l_{lt} = e^{\boldsymbol{X}_{lt}\hat{\beta}+\hat{c}_r}$ as the expected value of $N^l_{lt}$. For the ordered probit models the residuals are calculated as $\Delta N^l_{l,t+1} - \Delta\hat{N}^l_{l,t+1}$, with $\Delta\hat{N}^l_{l,t+1} = \sum_{j\in J} jPr(\Delta\hat{N}^l_{l,t+1} = j)$ as the expected value of $\Delta N^l_{lt}$, $Pr(\Delta\hat{N}^l_{lt} = j)$ as the probability that $\Delta N^l_{lt}$ takes the value $j$, and $J$ as the set of all possible outcomes. We denote the estimated residuals by $\hat{u}_{lt}$ for both estimation techniques for notational convenience.

In all model specifications the residuals are clustered at the location (grid cell) level to account for the correlation of the residuals within locations over time. As clustering is unfeasible in the diagnostic tests discussed below, we regress the estimated residuals of the Poisson or the ordered probit models, $\hat{u}_{lt}$, on location-fixed effects $c_l$. We thus take the estimated residuals $\hat{\epsilon}_{lt}$ from the OLS regression $\hat{u}_{lt} = c_l + \epsilon_{lt}$.

Two test statistics are provided based on the residuals of the main regressions: First, we perform a Moran's $I$ test, developed by Moran (1950) and described e.g. in Anselin (1988). The test statistic is defined as $I = \frac{\hat{\boldsymbol{\epsilon}}' W \hat{\boldsymbol{\epsilon}}}{\hat{\boldsymbol{\epsilon}}'\hat{\boldsymbol{\epsilon}}}$. $\hat{\boldsymbol{\epsilon}}$ is the estimated vector of residuals described above and $\boldsymbol{W}$ is a spatial weights matrix of dimension $NT$, capturing the spatial relations between locations in the vicinity. Its characteristic element $w_{at,bs}$ describes the relation between locations $a$ and $b$ at time periods $t$ and $s$. The element $w_{at,bs}$ is based on $w^*_{at,bs} = 1$, if the distance between $a$ and $b$ is equal or below a defined threshold distance, $a \neq b$ and $t = s$ (i.e. we only investigate contemporaneous spatial correlation), and otherwise zero. The matrix $\boldsymbol{W}$ is row-normalized and therefore $w_{at,bs} = \frac{w^*_{at,bs}}{\sum_b \sum_s w^*_{at,bs}}$. Second, we perform an LM test on spatial autocorrelation of the residuals that is robust to heteroscedasticity, as outlined in Born and Breitung (2011). The LM-test is $\chi^2$-distributed with one degree of freedom. In both test statistics the null-hypothesis is the absence of spatial autocorrelation. The threshold distance (and thus the spatial weights matrix $\boldsymbol{W}$), has to be specified exogenously. As any choice is somewhat arbitrary we use a short threshold distance of 500m and a long one with 2km in alternative specifications of the test statistics.

The results of the diagnostic tests for all four model specifications reported in the main part of this article are provided in Table A.1 below. Using the short threshold distance of 500m the Moran's $I$ statistics, which can take values between $-1$ and $+1$, are very close to zero and not significantly different from zero at the 10% significance level. The p-values of the LM tests are slightly higher compared to those of the Moran's $I$ tests, suggesting a moderate degree of heteroscedasticity in the residuals. Thus, the LM tests do also not rejected the null hypothesis at the 10% level in any model specification. The null hypothesis of no spatial autocorrelation is rejected in two out of eight tests at the 10% level if the weights matrix $\boldsymbol{W}$ is based on the longer threshold distance of 2km. We nevertheless conclude that spatial dependence and spatial heterogeneity are adequately accounted for in our models and that no (or hardly any) spatial structure is left in the residuals.

Table A.1: Diagnostic Statistics on Spatial Autocorrelation

| | Threshold distance | Table 2 Model [1] | Table 2 Model [2] | Table 3 Model [3] | Table 3 Model [4] |
|---|---|---|---|---|---|
| Moran's $I$ | 500 m | $-0.0012$ | 0.0028 | 0.0003 | $-0.0015$ |
| Std. dev. | | 0.0032 | 0.0032 | 0.0034 | 0.0034 |
| $z$-score | | 0.3555 | 0.8898 | 0.1016 | 0.4354 |
| p-value | | 0.7222 | 0.3736 | 0.9191 | 0.6633 |
| LM-test | 500 m | 0.0682 | 0.5436 | 0.0025 | 0.0665 |
| p-value | | 0.7939 | 0.4609 | 0.9599 | 0.7964 |
| Moran's $I$ | 2 km | $-0.0018$ | $-0.0006$ | 0.0004 | 0.0004 |
| Std. dev. | | 0.0009 | 0.0009 | 0.0010 | 0.0010 |
| $z$-score | | 1.8722 | 0.6622 | 0.4046 | 0.4337 |
| p-value | | 0.0612 | 0.5079 | 0.6857 | 0.6645 |
| LM-test | 2 km | 4.0400 | 0.4753 | 0.0699 | 0.0825 |
| p-value | | 0.0444 | 0.4906 | 0.7914 | 0.7739 |

Notes: $z$-scores are $t$-distributed with $N-1 = 24,527$ (Model [1] and [2]) or $N-1 = 21,462$ (Model [3] and [4]) degrees of freedom, respectively. LM-test statistics are $\chi^2$ distributed with 1 degree of freedom.

# Appendix B. Results of Sensitivity Analyses

Table B.1: Regression Results regarding the Number of Nurseries

| | Model [5] | | | Model [6] | | |
|---|---|---|---|---|---|---|
| | Coeff. | Std. Err. | Sign. | Coeff. | Std. Err. | Sign. |
| Location $l$ | | | | | | |
| # of children aged ≤ 5 | 0.00852 | (0.00086) | *** | 0.00797 | (0.00087) | *** |
| # of jobs | 0.00016 | (0.00005) | *** | 0.00015 | (0.00005) | *** |
| # of subway stations | 0.30281 | (0.14323) | ** | 0.23255 | (0.16461) | |
| Share employed women | 0.01280 | (0.00475) | *** | 0.01530 | (0.00505) | *** |
| Share COB Austria | −0.00106 | (0.00501) | | 0.00020 | (0.00536) | |
| Share COB other EU country | −0.01556 | (0.01170) | | −0.01312 | (0.01291) | |
| Share high school diploma | −0.01655 | (0.00851) | * | −0.02086 | (0.00910) | ** |
| Share college degree | −0.00120 | (0.00687) | | 0.00248 | (0.00746) | |
| Distance band $d_1$ | | | | | | |
| # of children aged ≤ 5 | 0.00061 | (0.00021) | *** | 0.00086 | (0.00025) | *** |
| # of jobs | 0.00003 | (0.00001) | ** | 0.00003 | (0.00002) | |
| # of subway stations | 0.12523 | (0.05701) | ** | 0.08954 | (0.06895) | |
| Share employed women | 0.02826 | (0.01126) | ** | 0.03916 | (0.01363) | *** |
| Share COB Austria | 0.01807 | (0.00906) | ** | 0.02041 | (0.01016) | ** |
| Share COB other EU country | 0.08405 | (0.02876) | *** | 0.09964 | (0.03606) | *** |
| Share high school diploma | −0.05701 | (0.02344) | ** | −0.07001 | (0.02808) | ** |
| Share college degree | 0.00890 | (0.01586) | | 0.01134 | (0.01851) | |
| Distance band $d_2$ | | | | | | |
| # of children aged ≤ 5 | 0.00015 | (0.00010) | | 0.00023 | (0.00013) | * |
| # of jobs | 0.00000 | (0.00001) | | 0.00000 | (0.00001) | |
| # of subway stations | 0.03458 | (0.03769) | | 0.03505 | (0.04607) | |
| Share employed women | −0.01046 | (0.01580) | | 0.02054 | (0.02155) | |
| Share COB Austria | 0.02563 | (0.01449) | * | 0.04612 | (0.01847) | ** |
| Share COB other EU country | 0.09441 | (0.06114) | | 0.12564 | (0.08221) | |
| Share high school diploma | 0.02780 | (0.03724) | | −0.01645 | (0.04996) | |
| Share college degree | −0.02595 | (0.02493) | | −0.00643 | (0.03202) | |
| Distance band $d_3$ | | | | | | |
| # of children aged ≤ 5 | 0.00005 | (0.00004) | | 0.00003 | (0.00006) | |
| # of jobs | −0.00000 | (0.00000) | | −0.00001 | (0.00000) | |
| # of subway stations | 0.00394 | (0.02096) | | 0.01037 | (0.02541) | |
| Share employed women | −0.04072 | (0.02341) | * | 0.03498 | (0.04518) | |
| Share COB Austria | −0.00777 | (0.01516) | | 0.01437 | (0.02135) | |
| Share COB other EU country | 0.02829 | (0.08278) | | 0.09325 | (0.13140) | |
| Share high school diploma | −0.02742 | (0.05699) | | −0.10952 | (0.09808) | |
| Share college degree | 0.01085 | (0.02990) | | 0.04811 | (0.05109) | |
| Constant | −5.89698 | (2.04058) | *** | −13.76036 | (3.25571) | *** |
| # of observations | 24,528 | | | 24,528 | | |
| Time effects | Yes – 7 | | | Yes – 7 | | |
| District effects | Yes – 22 | | | No | | |
| Registration district effects | No | | | Yes – 242 | | |
| Log-likelihood | –15,458 | | | –14,474 | | |
| $\chi^2$ test statistic for l | 127.36 ($df = 8$) [$p = 0.0000$] | | | 111.15 ($df = 8$) [$p = 0.0000$] | | |
| $\chi^2$ test statistic for $d_1$ | 36.51 ($df = 8$) [$p = 0.0000$] | | | 31.78 ($df = 8$) [$p = 0.0001$] | | |
| $\chi^2$ test statistic for $d_2$ | 8.77 ($df = 8$) [$p = 0.3625$] | | | 10.63 ($df = 8$) [$p = 0.2236$] | | |
| $\chi^2$ test statistic for $d_3$ | 9.27 ($df = 8$) [$p = 0.3197$] | | | 3.45 ($df = 8$) [$p = 0.9030$] | | |

Notes: Standard errors are reported in parentheses and are based on standard errors that are clustered at the grid cell level. *** significant at 1 %, ** significant at 5 %, * significant at 10 % level. $l$ denotes values of the respective variables in the grid cell, $d_1$ values aggregated or averaged for all cells within a distance [250m, 500m], $d_2$ values aggregated or averaged for all cells within a distance (500m, 1,000m], and $d_3$ values aggregated or averaged for all cells within a distance (1,000m, 2,000m]. The reported $\chi^2$-statistics test the hypotheses whether the coefficients of all variables for a particular distance band are jointly significant (H0: all coefficients are zero).

Table B.2: Regression Results regarding the Change of the Number of Nursery Groups

| | Model [7] | | | Model [8] | | |
|---|---|---|---|---|---|---|
| | Coeff. | Std. Err. | Sign. | Coeff. | Std. Err. | Sign. |
| Location $l$ | | | | | | |
| # of nurseries | −0.24636 | (0.03761) | *** | −0.30864 | (0.03721) | *** |
| # of children aged $\leq 5$ | 0.00308 | (0.00054) | *** | 0.00330 | (0.00050) | *** |
| # of jobs | 0.00007 | (0.00003) | ** | 0.00007 | (0.00003) | ** |
| # of subway stations | 0.31367 | (0.10688) | *** | 0.29696 | (0.11984) | ** |
| Share employed women | 0.00278 | (0.00137) | ** | 0.00451 | (0.00144) | *** |
| Share COB Austria | −0.00368 | (0.00183) | ** | −0.00378 | (0.00188) | ** |
| Share COB other EU country | −0.00238 | (0.00377) | | −0.00198 | (0.00399) | |
| Share high school diploma | 0.00180 | (0.00203) | | 0.00009 | (0.00222) | |
| Share college degree | 0.00193 | (0.00194) | | 0.00228 | (0.00206) | |
| Distance band $d_1$ | | | | | | |
| # of nurseries | −0.00108 | (0.00778) | | −0.02563 | (0.00884) | *** |
| # of children aged $\leq 5$ | 0.00027 | (0.00012) | ** | 0.00044 | (0.00014) | *** |
| # of jobs | 0.00001 | (0.00001) | | 0.00001 | (0.00001) | |
| # of subway stations | 0.10490 | (0.03214) | *** | 0.08204 | (0.04261) | * |
| Share employed women | 0.00123 | (0.00275) | | 0.00679 | (0.00315) | ** |
| Share COB Austria | −0.00129 | (0.00185) | | −0.00042 | (0.00202) | |
| Share COB other EU country | −0.01546 | (0.00786) | ** | −0.01286 | (0.00889) | |
| Share high school diploma | −0.00141 | (0.00677) | | −0.00884 | (0.00776) | |
| Share college degree | 0.00328 | (0.00488) | | 0.00633 | (0.00571) | |
| Distance band $d_2$ | | | | | | |
| # of nurseries | 0.00568 | (0.00537) | | −0.00136 | (0.00628) | |
| # of children aged $\leq 5$ | 0.00006 | (0.00006) | | 0.00018 | (0.00008) | ** |
| # of jobs | −0.00000 | (0.00000) | | 0.00000 | (0.00001) | |
| # of subway stations | 0.00874 | (0.02008) | | −0.00830 | (0.02918) | |
| Share employed women | 0.00787 | (0.00423) | * | 0.01517 | (0.00515) | *** |
| Share COB Austria | −0.00144 | (0.00245) | | 0.00692 | (0.00280) | ** |
| Share COB other EU country | 0.00295 | (0.01214) | | −0.00374 | (0.01361) | |
| Share high school diploma | −0.00313 | (0.00934) | | −0.01147 | (0.01052) | |
| Share college degree | 0.00483 | (0.00642) | | 0.00744 | (0.00698) | |
| Distance band $d_3$ | | | | | | |
| # of nurseries | 0.00535 | (0.00232) | ** | 0.00853 | (0.00277) | *** |
| # of children aged $\leq 5$ | −0.00001 | (0.00003) | | −0.00000 | (0.00003) | |
| # of jobs | −0.00000 | (0.00000) | ** | −0.00001 | (0.00000) | ** |
| # of subway stations | 0.00429 | (0.01030) | | −0.01179 | (0.01435) | |
| Share employed women | 0.00270 | (0.00587) | | 0.01053 | (0.00753) | |
| Share COB Austria | 0.00622 | (0.00401) | | 0.00875 | (0.00506) | * |
| Share COB other EU country | −0.04597 | (0.01839) | ** | −0.04971 | (0.02206) | ** |
| Share high school diploma | 0.00346 | (0.01401) | | −0.01286 | (0.01846) | |
| Share college degree | 0.00570 | (0.00888) | | 0.01071 | (0.01136) | |
| # of observations | 21,462 | | | 21,462 | | |
| Time effects | Yes – 7 | | | Yes – 7 | | |
| District effects | Yes – 22 | | | No | | |
| Registration district effects | No | | | Yes – 242 | | |
| Log-likelihood | –4,185 | | | –4,093 | | |
| Pseudo-$R^2$ | 0.0963 | | | 0.1160 | | |
| $\chi^2$ test statistic for $l$ | 118.03 ($df = 9$) [$p = 0.0000$] | | | 150.43 ($df = 9$) [$p = 0.0000$] | | |
| $\chi^2$ test statistic for $d_1$ | 24.56 ($df = 9$) [$p = 0.0035$] | | | 27.84 ($df = 9$) [$p = 0.0010$] | | |
| $\chi^2$ test statistic for $d_2$ | 10.58 ($df = 9$) [$p = 0.3059$] | | | 25.20 ($df = 9$) [$p = 0.0160$] | | |
| $\chi^2$ test statistic for $d_3$ | 19.772 ($df = 9$) [$p = 0.0194$] | | | 28.63 ($df = 9$) [$p = 0.0028$] | | |

Notes: Standard errors are reported in parentheses and are based on standard errors that are clustered at the grid cell level. *** significant at 1 %, ** significant at 5 %, * significant at 10 % level. $l$ denotes values of the respective variables in the grid cell, $d_1$ values aggregated or averaged for all cells within a distance [250m, 500m], $d_2$ values aggregated or averaged for all cells within a distance (500m, 1,000m], and $d_3$ values aggregated or averaged for all cells within a distance (1,000m, 2,000m]. The reported $\chi^2$-statistics test the hypotheses whether the coefficients of all variables for a particular distance band are jointly significant (H0: all coefficients are zero).

Table B.3: Regression Results regarding the Number of Nursery Groups in Urban vs. Suburban Areas

| | Model [9] | | | Model [10] | | |
|---|---|---|---|---|---|---|
| | Coeff. | Std. Err. | Sign. | Coeff. | Std. Err. | Sign. |
| Location $l$ | | | | | | |
| # of children aged ≤ 5 | 0.00749 | (0.00111) | *** | 0.01323 | (0.00154) | *** |
| # of jobs | 0.00012 | (0.00006) | ** | 0.00062 | (0.00025) | ** |
| # of subway stations | 0.25042 | (0.17613) | | −0.25256 | (0.60285) | |
| Share employed women | 0.01268 | (0.00717) | * | 0.01846 | (0.00828) | ** |
| Share COB Austria | 0.01140 | (0.00662) | * | −0.01765 | (0.00922) | * |
| Share COB other EU country | −0.01432 | (0.01603) | | −0.02174 | (0.02369) | |
| Share high school diploma | −0.02503 | (0.01339) | * | −0.01825 | (0.01423) | |
| Share college degree | 0.00441 | (0.01025) | | −0.01149 | (0.01369) | |
| Distance band $d_1$ | | | | | | |
| # of children aged ≤ 5 | 0.00029 | (0.00026) | | 0.00173 | (0.00049) | *** |
| # of jobs | 0.00003 | (0.00002) | * | 0.00015 | (0.00007) | ** |
| # of subway stations | 0.05980 | (0.07732) | | 0.05219 | (0.16145) | |
| Share employed women | 0.01186 | (0.01972) | | 0.07535 | (0.01938) | *** |
| Share COB Austria | −0.02181 | (0.01429) | | 0.07841 | (0.02239) | *** |
| Share COB other EU country | −0.00712 | (0.05510) | | 0.20506 | (0.05247) | *** |
| Share high school diploma | −0.00348 | (0.03517) | | −0.15539 | (0.05334) | *** |
| Share college degree | 0.00288 | (0.02333) | | 0.06066 | (0.03372) | * |
| Distance band $d_2$ | | | | | | |
| # of children aged ≤ 5 | 0.00014 | (0.00014) | | 0.00032 | (0.00037) | |
| # of jobs | −0.00000 | (0.00001) | | 0.00000 | (0.00004) | |
| # of subway stations | 0.03848 | (0.05413) | | 0.01498 | (0.10427) | |
| Share employed women | 0.02568 | (0.02910) | | −0.04454 | (0.02585) | * |
| Share COB Austria | 0.00935 | (0.02466) | | 0.02289 | (0.03015) | |
| Share COB other EU country | 0.01690 | (0.11443) | | 0.04834 | (0.11556) | |
| Share high school diploma | 0.03989 | (0.06012) | | 0.04202 | (0.06821) | |
| Share college degree | −0.01001 | (0.03711) | | 0.00760 | (0.04623) | |
| Distance band $d_3$ | | | | | | |
| # of children aged ≤ 5 | −0.00007 | (0.00006) | | 0.00031 | (0.00018) | * |
| # of jobs | 0.00000 | (0.00000) | | 0.00001 | (0.00002) | |
| # of subway stations | −0.02315 | (0.02943) | | 0.05597 | (0.06565) | |
| Share employed women | −0.06783 | (0.06311) | | 0.01764 | (0.04284) | |
| Share COB Austria | −0.03406 | (0.03467) | | 0.02935 | (0.02381) | |
| Share COB other EU country | −0.00930 | (0.20440) | | −0.11902 | (0.12623) | |
| Share high school diploma | −0.06043 | (0.13433) | | −0.04882 | (0.10757) | |
| Share college degree | 0.04016 | (0.06462) | | 0.03477 | (0.05842) | |
| Constant | 2.68287 | (4.46007) | | −11.67499 | (4.02381) | *** |
| # of observations | 12,274 | | | 12,254 | | |
| Time effects | Yes – 7 | | | Yes – 7 | | |
| District effects | Yes – 22 | | | Yes – 22 | | |
| Log-likelihood | –25,738 | | | –12,784 | | |
| Sample | Urban areas | | | Suburban areas | | |
| $\chi^2$ test statistic for $l$ | 60.95 ($df = 8$) [$p = 0.0000$] | | | 97.56 ($df = 8$) [$p = 0.0000$] | | |
| $\chi^2$ test statistic for $d_1$ | 12.65 ($df = 8$) [$p = 0.1245$] | | | 44.99 ($df = 8$) [$p = 0.0000$] | | |
| $\chi^2$ test statistic for $d_2$ | 9.69 ($df = 8$) [$p = 0.8200$] | | | 12.74 ($df = 8$) [$p = 0.5830$] | | |
| $\chi^2$ test statistic for $d_3$ | 10.91 ($df = 8$) [$p = 0.2873$] | | | 11.29 ($df = 8$) [$p = 0.1213$] | | |

Notes: Standard errors are reported in parentheses and are based on standard errors that are clustered at the grid cell level. *** significant at 1 %, ** significant at 5 %, * significant at 10 % level. $l$ denotes values of the respective variables in the grid cell, $d_1$ values aggregated or averaged for all cells within a distance [250m, 500m], $d_2$ values aggregated or averaged for all cells within a distance (500m, 1,000m], and $d_3$ values aggregated or averaged for all cells within a distance (1,000m, 2,000m]. The reported $\chi^2$-statistics test the hypotheses whether the coefficients of all variables for a particular distance band are jointly significant (H0: all coefficients are zero). The sample is split based on the median populations of 58,000 residents within 2 km distance around location $l$.

Table B.4: Results regarding the Number of Nursery Groups Considering Omitted Variables

| | Model [11] | | | Model [12] | | |
|---|---|---|---|---|---|---|
| | Coeff. | Std. Err. | Sign. | Coeff. | Std. Err. | Sign. |
| Location $l$ | | | | | | |
| # of nurseries (at $t-1$) | 0.32622 | (0.00844) | *** | | | |
| # of children aged $\leq 5$ | 0.00349 | (0.00063) | *** | 0.00303 | (0.00083) | *** |
| # of jobs | 0.00011 | (0.00004) | *** | 0.00025 | (0.00015) | * |
| # of subway stations | 0.20999 | (0.11308) | * | 0.90249 | (0.27033) | *** |
| Share employed women | 0.01147 | (0.00353) | *** | | | |
| Share COB Austria | −0.00954 | (0.00325) | *** | | | |
| Share COB other EU country | −0.01129 | (0.00924) | | | | |
| Share high school diploma | −0.00900 | (0.00637) | | | | |
| Share college degree | −0.00008 | (0.00477) | | | | |
| Distance band $d^1$ | | | | | | |
| # of nurseries (at $t-1$) | −0.00760 | (0.00343) | ** | | | |
| # of children aged $\leq 5$ | 0.00082 | (0.00019) | *** | 0.00035 | (0.00021) | |
| # of jobs | 0.00001 | (0.00001) | | 0.00001 | (0.00003) | |
| # of subway stations | 0.02155 | (0.04702) | | −0.13234 | (0.09092) | |
| Share employed women | 0.01940 | (0.00972) | ** | | | |
| Share COB Austria | 0.03246 | (0.00756) | *** | | | |
| Share COB other EU country | 0.07532 | (0.02568) | *** | | | |
| Share high school diploma | −0.07336 | (0.02067) | *** | | | |
| Share college degree | 0.02558 | (0.01349) | * | | | |
| Distance band $d^2$ | | | | | | |
| # of nurseries (at $t-1$) | −0.00014 | (0.00212) | | | | |
| # of children aged $\leq 5$ | 0.00004 | (0.00010) | | 0.00021 | (0.00012) | * |
| # of jobs | 0.00001 | (0.00001) | | −0.00001 | (0.00001) | |
| # of subway stations | −0.06261 | (0.02844) | ** | 0.08055 | (0.04785) | * |
| Share employed women | 0.00974 | (0.01574) | | | | |
| Share COB Austria | 0.01186 | (0.01180) | | | | |
| Share COB other EU country | 0.07566 | (0.05070) | | | | |
| Share high school diploma | −0.03900 | (0.03464) | | | | |
| Share college degree | −0.00894 | (0.02244) | | | | |
| Distance band $d^3$ | | | | | | |
| # of nurseries (at $t-1$) | −0.00203 | (0.00106) | * | | | |
| # of children aged $\leq 5$ | 0.00002 | (0.00005) | | −0.00004 | (0.00006) | |
| # of jobs | 0.00000 | (0.00000) | | 0.00000 | (0.00001) | |
| # of subway stations | 0.02868 | (0.02072) | | −0.01243 | (0.01866) | |
| Share employed women | 0.02177 | (0.03197) | | | | |
| Share COB Austria | 0.01722 | (0.01591) | | | | |
| Share COB other EU country | −0.04548 | (0.09262) | | | | |
| Share high school diploma | −0.04540 | (0.06971) | | | | |
| Share college degree | 0.00185 | (0.03884) | | | | |
| Constant | −6.30863 | (2.41907) | *** | | | |
| # of observations | 24,528 | | | 7,408 | | |
| Time effects | Yes – 7 | | | Yes – 7 | | |
| Registration district effects | Yes – 242 | | | No | | |
| Location effects | No | | | Yes – 926 | | |
| Log-likelihood | –21,725 | | | –9,123 | | |
| $\chi^2$ test statistic for $l$ | 1,654.47 ($df = 9$) [$p = 0.0000$] | | | 28.71 ($df = 3$) [$p = 0.0000$] | | |
| $\chi^2$ test statistic for $d^1$ | 39.15 ($df = 9$) [$p = 0.0000$] | | | 4.76 ($df = 3$) [$p = 0.1903$] | | |
| $\chi^2$ test statistic for $d^2$ | 10.16 ($df = 9$) [$p = 0.3376$] | | | 7.36 ($df = 3$) [$p = 0.0614$] | | |
| $\chi^2$ test statistic for $d^3$ | 8.78 ($df = 9$) [$p = 0.4574$] | | | 0.88 ($df = 3$) [$p = 0.8303$] | | |

Notes: Standard errors are reported in parentheses and are based on standard errors that are clustered at the grid cell level. *** significant at 1 %, ** significant at 5 %, * significant at 10 % level. $l$ denotes values of the respective variables in the grid cell, $d_1$ values aggregated or averaged for all cells within a distance [250m, 500m], $d_2$ values aggregated or averaged for all cells within a distance (500m, 1,000m], and $d_3$ values aggregated or averaged for all cells within a distance (1,000m, 2,000m]. The reported $\chi^2$-statistics test the hypotheses whether the coefficients of all variables for a particular distance band are jointly significant (H0: all coefficients are zero). Number of observations is reduced in regression including location $l$ fixed effects, because all locations with only zero outcomes (i.e. without nursery groups) over the entire period have to be dropped. Time invariant variables are not identified anymore in the fixed effects model and are thus neglected.