Regional Science and Urban Economics 70 (2018) 155–173

Contents lists available at ScienceDirect

Regional Science and Urban Economics

journal homepage: www.elsevier.com/locate/regec

Measuring the spatial effect of multiple sites: An application to housing rent and public transportation in Tokyo, Japan

Taisuke Sadayuki [a,b,*,1]

[a] Research Institute for Environmental Economics and Management, Waseda University, Nishiwaseda 1-6-1, Shinjuku-ku, Tokyo 169-8050, Japan
[b] Regional Economics Applications Laboratory, University of Illinois at Urbana-Champaign, 607 S. Mathews Ave., Urbana IL 61801, USA

ARTICLE INFO

JEL codes:
R30
R40

Keywords:
Spatial analysis
Hedonic
Accessibility measure
Transportation
House price
Proximity ranking

ABSTRACT

Geographical relationships between a housing unit and its major surrounding sites, such as public transportation stops and crime scenes, are fundamental factors that determine housing value. This paper proposes a new parametric approach to estimate the aggregate spatial effect of multiple heterogeneous sites while providing fruitful interpretations of the effect of each of these sites. While the proposed method is developed based on a traditional accessibility measure, the way in which it addresses the role of the proximity order of sites in the spatial analysis is novel. The method is applied empirically using rental housing data in Tokyo, Japan to examine how the clustering of train and subway stations influences the rental prices in their vicinity. The results reveal a discounting impact of the order of each station's proximity, even after controlling for the effect of distance. In addition, the results reveal that using a traditional accessibility measure without considering the proximity order leads to serious estimation biases. The proposed methodology is applicable to various spatial topics, such as transportation, neighborhood externalities and polycentric urban structures.

1. Introduction

Geographical relationships between a housing unit and its major surrounding sites, such as public transportation stops, commercial facilities, schools, and crime scenes, as well as the characteristics of these sites are fundamental factors that determine the value of housing. In this paper, we propose an empirical model to estimate the aggregate spatial effect of multiple sites by accounting for the following three general assumptions: (A1) as the distance to a site decreases, the potential impact of the site increases; (A2) this impact may differ according to the characteristics of a site; and (A3) as the ranking of proximity to a site increases, the potential impact of the site increases. This new methodology enables us to gain fruitful and practical interpretations of the spatial effect of each of the surrounding heterogeneous sites. In particular, this is the first paper to address the third assumption (A3) that addresses multiple sites and shows in the application the diminishing effect of sites along their proximity order.

The main focus of the proposed methodology is to examine the spatial effect of multiple sites using point-to-point data accompanied by detailed site addresses. Previous studies in this field have predominantly used three types of proximity variables, namely, (i) the distance between a housing unit and its closest site (Ahlfeldt, 2011a,b; Ahlfeldt and Wendland, 2011; Mejia-Dorantes et al., 2012; Gibbons and Machin, 2005; Troy and Grove, 2008),[2] (ii) the number of sites within a certain distance from

* Research Institute for Environmental Economics and Management, Waseda University, Nishiwaseda 1-6-1, Shinjuku-ku, Tokyo 169-8050, Japan.
E-mail address: tai.sadayuki@gmail.com.

[1] The research of the paper was done mostly at the University of Illinois.

[2] Troy and Grove (2008), for example, compute the distance to the nearest park from each housing unit in Maryland and examine the impact of the crime rate at the park on neighboring housing values. Mejia-Dorantes et al. (2011) and Gibbons and Machin (2005) estimate the impact of the public transportation infrastructure by comparing the coefficients of the distance to the closest stations before and after the infrastructure is completed. Ahlfeldt (2011a,b) and Ahlfeldt and Wendland (2011) include minimum distances to various locations, such as a transportation station, main road, school, water space, green space, and industrial area, to estimate the land price.

https://doi.org/10.1016/j.regsciurbeco.2018.03.002
Received 19 May 2017; Received in revised form 5 March 2018; Accepted 6 March 2018
Available online 10 March 2018

a housing unit (Bak and Hewings, 2017; Gerardi et al., 2015; Harding et al., 2009; Immergluck and Smith, 2006; Leonard and Murdoch, 2009; Lin et al., 2009; Rogers and Winter, 2009; Schwartz et al., 2006; Schuetz et al., 2008; Srour et al., 2002),[3] and (iii) an indicator of whether any site is located within a certain distance from a housing unit (Hoen, 2010; Bowes and Ihlanfeldt, 2001; Cui and Walsh, 2015; Forrest et al., 1996; Kahn, 2007; Linden and Rockoff, 2008).[4] None of the three traditional proximity variables satisfies all three general assumptions listed above (Table 1). The use of each of these variables is justifiable under strict criteria, and failure to meet these criteria can lead to a biased estimate. For instance, using only the first type of proximity variable (i.e., the distance to the closest site) in the hedonic estimation assumes that the second and third closest sites have no influence on housing value. A possible way to address the effect of multiple sites is to regress the value of the housing unit on the unit's distances to the closest, second closest, and third closest sites, and so forth. However, adding multiple distances to the hedonic model would lead to a serious multicollinearity problem, preventing us from drawing reliable and meaningful interpretations of the spatial effect.[5] Another possible remedy is to coordinate the second type of proximity variable with the first type or to use a distance-decayed sum of sites within a certain area (Campbell et al., 2011; Kok et al., 2014).[6] A concern with using these practices is the choice of an adequate buffer, which researchers typically determine in an arbitrary manner. Several studies attempt to avoid problems associated with multiple sites and spatial heterogeneity by restricting housing samples to housing units located very close to sites instead of implementing variables to account for multiple sites (McMillen and McDonald, 2004; Pope, 2008).[7]

Our proposed proximity measure is based on another type of measure, namely, an "accessibility measure," which is the sum of gravity-based functions that are decreasing in distance and increasing in the destination's attractiveness. Among the numerous studies related to the accessibility measure, which was developed in fields of study such as land use and transportation (to name a few examples, Hansen, 1959; Ottensmann and Lindsey, 2008; Iacono et al., 2010; Salze et al., 2011; Song, 1996), the number of studies that apply it to the hedonic approach has increased in the past two decades (Appendix A provides a detailed discussion). Most of the accessibility measures in these hedonic analysis studies are based on zone-to-zone measures rather than point-to-point measures, i.e., the distances used in these measures are computed between zones (such as zip code areas, transportation analysis zones, and voting precincts) rather than between detailed addresses of housing units and sites. This approach was chosen because the major purpose of these studies is to assess a city's accessibility to employment opportunities in other cities by counting the number of employment or job opportunities in each area, thereby addressing the significance of a polycentric urban structure in determining housing value.

In comparison with these studies, our focus is more local, as we examine the spatial effect of multiple sites, such as public transportation stops, parks, supermarkets, foreclosures, and crime scenes, that are likely to affect only those who live in the neighborhood. The accessibility measure can be useful in such examinations for two reasons. First, the measure provides more flexibility in the functional form than the three types of proximity variables listed earlier. Secondly, the number of parameters in the measure is not affected by the number of sites; thus, the measure can provide useful implications about the spatial effect of each site without facing a serious multicollinearity problem.

That said, the accessibility measure still fails to address the third assumption (A3). It is assumed that under this traditional accessibility measure, the spatial impact of a site does not depend on the proximity order. To gain a better understanding of the role of proximity order, consider a case in which the closest public transportation station is located one mile away from a housing unit. Suppose that a new station on a different line will be constructed just half a mile away from the housing unit such that this new station will become the closest one and the previously closest station will be the second closest. Will the impact of the previously closest station on the housing unit's value remain the same after the new station is constructed? If the answer is yes, the proximity order to each station may not be important. If the impact of the former station decreases because of the presence of the new station, then the order of proximity to each site should be considered when estimating the housing price. To the best of our knowledge, none of the previous studies using the accessibility measure has considered the proximity ranking perspective in practice.

In this paper, we develop a new proximity measure that satisfies all three assumptions (Table 1) by taking several steps, as demonstrated in the following section. In brief, we first redefine the traditional accessibility measure to fit within the context of point-to-point spatial analysis. Additionally, we demonstrate an alternative measure that has potential advantages over the traditional accessibility measures. Then, we introduce the proposed proximity measure by adding a new parameter to these measures to address the proximity ranking perspective. The proposed proximity measure is more general in its functional specification than the traditional accessibility measure and can be applied to various fields of study that deal with the spatial effects of clustered sites.

In section 3, we illustrate an application of the relationship between the housing rental value and the clustering of train and subway stations in Tokyo, Japan. In general, including a greater number of neighboring stations in an empirical model should yield a better estimation result if the model is correctly specified. However, in the application, we observe that the estimation results deteriorate using the traditional accessibility measure when a larger number of stations is included in the model. This result is due to the misspecification of the spatial effect by failing to account for the proximity order. The proposed measure solves this problem, and the estimation result improves as the number of stations considered in the model increases, suggesting that the proximity order of nearby stations plays a significant role in determining the value of housing.

To our knowledge, all existing hedonic analyses of the real estate market in Japanese urban areas have taken into account only the distance to the nearest station (see, for example, Diewert and Shimizu, 2016; Gao and Asami, 2001; Nakagawa et al., 2007; Shimizu and Nishimura, 2007; Shimizu et al., 2010; Yamagata et al., 2016). We show in the following application that at least the first three closest stations should be considered in order to obtain a better estimate of the housing value in Tokyo, whereas including more than the five closest stations in the model does

[3] Many studies on the impact of foreclosures on neighborhoods use this type of variable (Bak and Hewings, 2017; Gerardi et al., 2015; Harding et al., 2009; Immergluck and Smith, 2006; Leonard and Murdoch, 2009; Lin et al., 2009; Rogers and Winter, 2009; Schwartz et al., 2006; Schuetz et al., 2008). Instead of simply counting the sites, Srour et al. (2002) estimate the impact of social recreation areas, shopping centers, and workplaces by counting the number of retail jobs and total jobs and by measuring the area of park spaces.

[4] Linden and Rockoff (2008) use dummy variables indicating whether any sex offender lives within 0.1 miles or within 0.1–0.3 miles from a transactional housing unit to estimate its impact on the property value. Other papers using this type of variable include studies on the impacts of rail transit stations on housing values (Bowes and Ihlanfeldt, 2001; Forrest et al., 1996; Kahn, 2007) and on the effect of foreclosures on crime (Cui and Walsh, 2015).

[5] In Appendix B, the results of our application show that variance inflation factors (VIFs) of distance variables exceed the value of ten when we include distances to the first three closest sites.

[6] Campbell et al. (2011) study the impact of foreclosures and use two types of proximity variables as controls. One variable is a count of foreclosures within 0.25 miles from each transactional housing unit. The other variable is a distance-weighted sum of foreclosures within 0.01 mile. Kok et al. (2014) examine the determinants of land prices in which a distance-weighted sum of job opportunities at each CBD is used as an explanatory variable to control for the accessibility to CBDs.

[7] Pope (2008) excludes housing units that have more than one sex offender living within 0.15 miles. McMillen and McDonald (2004) estimate the impact of the Midway rapid transit line infrastructure in Cook County, Illinois, by excluding housing units that are located more than 1.5 miles from the Midway line or closer to other lines.