Transportation Research Part B 64 (2014) 24–40

Contents lists available at ScienceDirect

Transportation Research Part B

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

A spatial Difference-in-Differences estimator to evaluate the effect of change in public mass transit systems on house prices

Jean Dubé [a,*], Diègo Legros [b], Marius Thériault [c], François Des Rosiers [d]

[a] Université du Québec à Rimouski (UQAR), 300, Allée des Ursulines, Rimouski G5L 3A1, Canada
[b] Université de Bourgogne, 2, Boulevard Gabriel, 21066 Dijon Cedex, France
[c] Université Laval, 2325, rue des Bibliothéques, Pavillon Félix-Antoine-Savard, Québec G1V 0A6, Canada
[d] Université Laval, 2325, rue de la Terrasse, Pavillon Palasis-Prince, Québec G1V 0A6, Canada

ARTICLE INFO

Article history:
Received 22 July 2013
Received in revised form 26 February 2014
Accepted 27 February 2014

Keywords:
Mass transit
Policy evaluation
Hedonic pricing model
Difference-in-Differences estimator
Spatial econometrics

ABSTRACT

Evaluating the impact of public mass transit systems on real-estate values is an important application of the hedonic price model (HPM). Recently, a mathematical transformation of this approach has been proposed to account for the potential omission of latent spatial variables that may overestimate the impact of accessibility to mass transit systems on values. The development of a Difference-in-Differences (DID) estimator, based on the repeat-sales approach, is a move in the right direction. However, such an estimator neglects the possibility that specification of the price equation may follow a spatial autoregressive process with respect to the dependent variable. The objective of this paper is to propose a spatial Difference-in-Differences (SDID) estimator accounting for possible spatial spillover effects. Particular emphasis is placed on the development of a suitable weights matrix accounting for spatial links between observations. Finally, an empirical application of the SDID estimator based on the development of a new commuter rail transit system for the suburban agglomeration of Montréal (Canada) is presented and compared to the usual DID estimator.

1. Introduction

There is an important interest in the literature for establishing the willingness-to-pay (WTP) for being located close to a public mass transit (PMT) system. The transit-oriented development (TOD) concept is a good example of such a concern (Ma and Lo, 2013) and some endeavor has been devoted to explicitly include it in the modeling process (Li et al., 2012). The WTP can be assessed through either a stated preference (SP) or revealed preference (RP) approach (Hensher, 2010).[1] In empirical applications, the RP method is less common than the SP approach since data is usually difficult to collect (Peer et al., 2013). The hedonic price method applied on real estate transactions is a good example of how the RP approach may be implemented for evaluating the impact of proximity to PMT on house prices.

* Corresponding author. Tel.: +1 418 318 6998.
E-mail addresses: jean_dube@uqar.ca (J. Dubé), diego.legros@u-bourgogne.fr (D. Legros), marius.theriault@esad.ulaval.ca (M. Thériault), francois.desrosiers@fsa.ulaval.ca (F. Des Rosiers).

[1] There is a special issue of Transportation Research Part B published in 2010 (Volume 44, Issue 6) that addresses this topic.

http://dx.doi.org/10.1016/j.trb.2014.02.007
0191-2615/

Hedonic pricing model (HPM) applications have evaluated the effect of accessibility to PMT systems on single family house values (Des Rosiers et al., 2010; McMillen and McDonald, 2004), multifamily house values (Celik and Yankaya, 2006), apartment rent (Pan and Zhang, 2008; Cao and Hough, 2007), office rent (Ryan, 2005; Weinberger, 2001), industrial plant (Cervero and Duncan, 2002), and vacant land prices (Knaap et al., 2001).

The magnitude of the estimated effect is not homogenous and may depend on the type of public transportation as well as on the structure of the city (Mohammad et al., 2013; Debrezion et al., 2007, 2011). However, the interpretation of the PMT mode on real estate prices may be biased if the spatial amenities influencing the price determination process are not accounted for and are revealed to be largely correlated with the proximity of PMT. This explains why emphasis has recently been placed on the development of a Difference-in-Differences (DID) estimator to isolate the effect of PMT on house prices. Such an approach is based on a comparison of the difference in the prices of houses sold before and after the implementation of a PMT system and appears to be one of the best ways to evaluate the effect in the medium to long term (Wardrip, 2011; Gibbons and Machin, 2008).[2]

The Difference-in-Differences (DID) approach adequately controls for the possible omission of significant variables correlated with the PMT service descriptors. The DID estimator is also a convenient way to deal with the omission of a latent constant spatial structure uncorrelated with the independent variables and generating spatial autocorrelation among residuals (Dubé and Legros, 2013b, 2011). What is less clear, however, is what happens if spatial autocorrelation is generated by an autoregressive process over the dependent variable.

Many authors argue that the spatial autoregressive model (SAR) is preferable to the spatial error model (SEM) since it allows us to decompose the marginal effect into a direct and an indirect (spatial spillover) effect (Debarsy et al., 2012; LeSage and Pace, 2009). Taking advantage of this, LeSage and Pace (2009) showed that the SEM model can be transformed into a spatial Durbin version (Le Gallo, 2002; Elhorst, 2010) which is, to some extent, a natural extension of the SAR model (see also Elhorst, 2012).

This paper aims at developing a spatial Difference-in-Differences (SDID) estimator that allows for spatial autoregressive specification of the dependent variable. Based on the SAR specification of the hedonic price equation, a new estimator is proposed. The DID and SDID estimators are compared as part of a case study related to the Montréal (Canada) suburban commuter train. It shows how the SDID estimator may be preferable by allowing the marginal effect to be decomposed into two different components, while its advantage largely relies on the amplitude of the spatial autoregressive parameter estimated.

The paper is divided into five sections. The first proposes a brief review of the usual hedonic pricing model (HPM) and its extension to deal with the spatial considerations. The second section presents the DID estimator as well as its natural extension to the spatial case (SDID). Particular attention is paid to presenting the advantages of such approaches as well as the assumptions underlying these estimators. The third section propose a general discussion on the construction of the weights matrix for both models (spatial HPM and SDID) to make sure that modelers correctly measure the spatial spillover effect when data consist of real-estate transactions collected over time. This is necessary in order to avoid spurious spatial relation with respect to other temporal aspects such as perfect anticipation (Dubé and Legros, 2013a, 2011). The fourth section presents the statistics related to the development of a new Commuter Rail Train (CRT) on the northern suburban part of Montréal (Canada) and to the transaction data available to evaluate the effect of the implementation of a new PMT service. The DID and SDID estimators are applied to this case study and the results of the study are discussed and compared. The final section concludes the paper.

2. The hedonic pricing model

Hedonic theory (Rosen, 1974) states that the price of a complex good can be expressed as a function of its various extrinsic and intrinsic attributes, and that coefficients related to one specific characteristic represent its implicit (hedonic) price.

Statistically, the hedonic pricing model expresses the selling price of a complex good i sold at time t, $\mathbf{y}_{it}$, as a function of the various attributes of the good, summarized in a matrix of explanatory variables, $\mathbf{X}_{it}$, which includes a series of continuous descriptors (such as age, living area, and so on) as well as a series of dichotomous variables (presence/absence of housing attributes), a set of dummy time-variables, $\mathbf{D}_{it}$, representing time fixed effects controlling for differences in the composition of the sample in each time period (Wooldridge, 2000) as well as for temporal heterogeneity (Eq. (1)) vector of constant term, $\imath$.

$$\mathbf{y}_{it} = \imath\alpha + \mathbf{D}_{it}\delta + \mathbf{X}_{it}\beta + \epsilon_{it} \quad \forall\ i = 1, \ldots, N_T \tag{1}$$

A particularity of real-estate data used to estimate the model is that they are not strictly spatial and are clearly different from panel data (see Parent and LeSage, 2010). Transactions database consist of a set of cross-sectional data pooled over time: observations are collected over time, but individual sales are not systematically repeated in each time period. Recurrences of house sales are usually treated as random events. In such situations, the total number of observations is noted $N_T = \sum_{t=1}^{T} N_t$, where N_t is the total number of observations in one time period, t.

Thus, the vector $\mathbf{y}_{it}$ is of dimension $(N_T \times 1)$, the vector $\imath$ is of dimension $(N_T \times 1)$, the matrix $\mathbf{D}_{it}$ is of dimension $(N_T \times (T-1))$, the matrix $\mathbf{X}_{it}$ is of dimension $(N_T \times K)$, where K is the total number of independent variables, and the vector

[2] Mokhtarian and Cao (2008) propose is similar conclusion.