1st European Symposium on Quantitative Methods in Transportation Systems, 4-8th September 2012, Lausanne, Switzerland

Impact of Data Structure on Difference-In-Differences Methods – A Case Study Using Dubai Metro Effects on Land/Property Values

Sara I. Mohammad[a], Daniel J. Graham[†a], Patricia C. Melo[a]
[a]Railway and Transport Strategy Centre, Centre for Transport Studies, Dept of Civil and Environmental Engineering, Imperial College London, London SW7 2AZ

Extended Abstract

A large number of studies have examined the effect of rail systems on land and property values. The conventional way of estimating this relationship is by regressing the reported property value on various factors that are expected to have an impact on the result. The empirical literature has generally adopted hedonic models to estimate the relationship controlling for the effect of the variables that are expected to have an impact on land/property values – that is, property characteristics, the exact property location, access to a rail station, train service level frequency, neighborhood amenities, environmental factors including unobserved local-area condition factors.

Although results of these types of models are indicative, they do not reflect causal relationships. For example, stations may be at high-valued areas like a commercial hub zone, and hedonic models typically do not account for the impact of these local effects on the reported values. In addition, unobserved heterogeneity across land/properties is likely to be present and if correlated with the distance to the train station and the property price, results will be biased. As an improvement to the hedonic model, a number of studies have recently started using an innovation-based model. The effect of rail on land/property values is estimated by examining land/property values before and after the innovation (or treatment) occurred for the properties that experienced the innovation (treated) compared to those that did not (controlled). This method is called the difference-in-differences (DID) estimation.

One of the issues for estimating the effect of rail on land/property values over a period of time is the lack of availability of panel data. To our best knowledge, no study conducting such an analysis was been successful in obtaining enough transactions for the same land and properties over a period of time. This poses an issue with the unobserved factors that need to be accounted

[†]Author for correspondence: RTSC, Centre for Transport Studies, Imperial College London, London SW7 2AZ, UK, Tel: +44 (0)20 7594 6088, Fax: +44 (0)20 7594 6107, Email: d.j.graham@imperial.ac.uk.

for to reduce biasness in results obtained from regression models. One way of solving this issue is by constructing a pseudo panel data structure (Deaton, 1985; Verbeek and Nijman, 1992; Collado, 1997; Cameron and Trivedi, 2005). A few studies have grouped the data into cohorts and used the mean values of these cohorts as individuals in the new sample, forming some sort of panel structure. A value for each cohort is then available for several periods of time even if the individual observations in that cohort are not the same over time. One important assumption is that the observations within a given cohort share the same unobserved characteristics. The literature has demonstrated that in order to reduce bias by grouping data at least 100 observations per cohort are required (Verbeek and Nijman, 1992). This was done in some studies that lacked panel data structure and constructed a pseudo panel data (see e.g. Dargay and Vythoulkas, 1999; Bourguignon et al., 2004; Russel and Fraas, 2005). As an example for a study estimating the effect of rail on land/property values using the DID model, Gibbons and Machin (2005) grouped data into postcode units but for a smaller number of observations. On average 2.5 observations constituted one cohort and each cohort contained 10 to 15 houses. On the other hand, Bowes and Ihlanfeldt (2001) used every single observation in the DID model and included fixed effect variables to account for unobserved heterogeneity.

We use the newly operated Dubai Metro as a case study to estimate the effect of the rail system on land and property values. Similar to other cities, panel data is not available for land/property transactions. Although we attempt to construct a pseudo panel data structure, the number of resulting cohorts (communities in Dubai) composing the pseudo panel was very small and hence could not be considered. As a result, we apply the DID method on the original dataset of individual properties by combining the property level data with community level data that can account for unobserved heterogeneity. This gives a motivation to study the impact of alternative data structures on the estimates obtained from DID models. We compare the results obtained from the estimation of two DID models; one based on a limited number of groups in a pseudo panel data, the other based on a large dataset of individual cross-sectional property data. In order to account for unobserved heterogeneity, suitable variables are included in the estimation of both models. Certain assumptions are needed when using cross-sectional data in a DID approach, and these are carefully discussed and justified. The consistency of the estimates obtained in each approach is discussed and conclusions are drawn respectively.

REFERENCES

BOURGUIGNON, F., GOH, C. C. & KIM, D. I. 2004. Estimating individual vulnerability to poverty with pseudo-panel data. *World Bank Policy Research Working Paper 3375.* The World Bank and Seoul National University.

BOWES, D. R. & IHLANFELDT, K. R. 2001. Identifying the impacts of rail transit stations on residential property values. *Journal of Urban Economics,* 50**,** 1-25.

CAMERON, A. C. & TRIVEDI, P. K. 2005. *Microeconometrics: Methods and applications*, Cambridge.

COLLADO, M. D. 1997. Estimating dynamic models from time series of independent cross-section. *Journal of Econometrics,* 82**,** 37-62.

DARGAY, J. M. & VYTHOULKAS, P. C. 1999. Estimation of a dynamic car ownership model - A pseudo-panel approach. *Journal of transport economics and policy,* 33**,** 287-301.

DEATON, A. 1985. Panel data from time series of cross-sections. *Journal of Econometrics,* 30**,** 109-126.

GIBBONS, S. & MACHIN, S. 2005. Valuing rail access using transport innovations. *Journal of Urban Economics,* 57**,** 148-169.

RUSSEL, J. E. & FRAAS, J. W. 2005. An application of panel regression to pseudo panel data. *Multiple Linear Regression Viewpoints,* 31.

VERBEEK, M. & NIJMAN, T. 1992. Can cohort data be treated as genuine panel data. *Empirical Economics,* 17**,** 9-23.