What Does "Location, Location, Location" Mean in the Context of House Price Indexes

Robert J. Hill and Michael Scholz
Department of Economics, University of Graz

1 Introduction

The interconnectedness of the housing market and the broader economy has been clearly demonstrated by the recent global financial crisis. It is important therefore that governments, central banks and market participants are able to monitor developments in housing markets.

In recent years there has been a proliferation of house price indexes (HPIs) (see Hill, 2013). But the sensitivity of house price indexes to the method of construction is a potential source of confusion (see Silver, 2012). Several difficulties are well recognized in the literature.[1] For example, dwellings are sold only sporadically. Furthermore, every house is different both in terms of its physical characteristics and its location. It is important that house price indexes take account of this heterogeneity. Otherwise they will confound price changes with differences in quality. Hedonic methods – which express house prices as a function of a vector of characteristics – are ideally suited for constructing quality-adjusted house price indexes.

The hedonic method can be implemented in different ways (see Hill, 2013). In this note, we emphasis the real estate professionals' mantra: "location, location, location", in the context of hedonic house price indexes. The discussion focuses primarily on recent developments where semi-parametric hedonic equations are estimated and afterwards combined with the imputation method and a Fisher price index formula (see Hill and Scholz, 2014).

2 How Can We Incorporate Location in HPIs?

One of the key determinants of house prices is location. It means that houses which are identical in their structural characteristics vary in value due to their location.

[1] For a comprehensive overview of conceptual and practical issues related to the compilation of price indexes for residential properties see de Haan and Diewert (2013).

The structure of a house could be changed, the layout altered, but it cannot be moved. It is attached to its property as a part of its surrounding neighborhood. The explanatory power of the hedonic model can therefore be significantly improved by exploiting information on the location of each property. Desirable locations are, inter alia, top-rated school districts, close to recreation areas, shopping, public transportation, health care, work places or homes with views. On the other hand, it would be less desirable, for example, living close to industries, railroad tracks, freeways or in crime-ridden and economically depressed neighborhoods. Some of the aforementioned locational attributes, like distances or views, are easy to measure and to incorporate as dummy variables or distances to amenities in HPIs. But usually, the available data are not detailed enough to capture the location value precisely. For this reason more sophisticated approaches are needed. Nowadays, exact addresses of sold properties are available. Thus, exact coordinates, measured in longitudes and latitudes, can be used in non- and semi-parametric methods to proxy most of the missing locational information.

2.1 Dummy Variables

Probably the simplest way of incorporating geospatial information in the hedonic model is to include postcode identifiers, dummy variables for regions, school districts or other neighborhood characteristics, etc. for each house. However, such an approach misses the final details of locational effects, particularly when the identified geographical zones are quite large. Furthermore, Hill and Scholz (2014) show that an HPI based on postcode or region dummies can be biased when the average locational quality of the houses sold within postcodes changes over time.

2.2 Distances to Amenities

Given the availability of geospatial data, the distance of each house to prime spots such as the city center, recreation areas, the airport, nearest public transportation or hospital can be measured. These distances can be included as further characteristics in the hedonic model.

However, using such distances to amenities is problematic in hedonic models for a few reasons. First and most importantly, it makes only limited use of the available geospatial data, and hence leaves out a lot of potentially useful information. One example could be housing externalities which refer to the effects the characteristics of a house have on other residents and, potentially, businesses (see Rossi-Hansberg and Sarte, 2012) and vice versa. Second, for each distance one needs the correct specification of the functional form. The impact of distance to an amenity on price of a house may be quite complicated and not necessarily monotonic. For example, it may be desirable for a house to be neither too close nor too far away from the airport, city center, etc. Third, geographic direction may matter as well as distance. For example, a house's position relative to the flight path is at least as important as the actual distance from an airport. Fourth, a perhaps more informative alternative to distance to the city center is the commuting time (see Shimizu, 2014) as well as measures of urban transportation performance or of congestion.

2.3 Spatial-Autoregressive Models

Locational effects can be captured more effectively by a spatial autoregressive model. The main effect of explicitly modeling the spatial dependence in a hedonic equation is that it should help offset the locational omitted variables problem (see Hill, 2013). Thus, it accounts for housing externalities and other common factors shared by neighborhoods. A prominent example is the (first-order) autoregressive spatial model with (first-order) autoregressive errors, referred to as the SARAR(1,1) model (see Anselin 1988; Corrado and Fingleton 2012).

One problem with this approach is that when the model is estimated over a number of years of data the spatial weights matrix should be replaced by a spatiotemporal weights matrix. That is, the magnitude of the dependence between observations depends inversely on both their spatial and temporal separation.

Replacing a spatial weights matrix with a spatiotemporal weights matrix significantly increases the computational burden and complicates the derivation of price indexes (see Nappi-Choulet and Maury, 2009). One response to this problem is to use the adjacent-period (AP) version of the time-dummy method. In this case the temporal separation between observations never gets that large and hence it is more defensible to use a spatial weights matrix instead of the theoretically preferred spatiotemporal weights matrix (see Hill, Melser and Syed, 2009).

The main problem with spatial autoregressive models is that they impose a lot of prior structure on the spatial dependence.

2.4 Non- and Semi-Parametric Approaches

Non- and semi-parametric methods provide a different and potentially more flexible alternative for modeling spatial dependence. They can be used to construct a topographical surface describing how prices vary by location, ceteris paribus. For example, Hill and Scholz (2014) estimate with thin-plate regression splines the following *generalized additive model (GAM),* which consist in period t of a parametric part defined over the physical characteristics Z and a fully flexible nonparametric part (an unknown function g_t) defined on the geospatial data – the longitude z_{long} and latitude z_{lat} of each observed house in period t:

$$y = Z\beta_t + g_t(z_{lat}, z_{long}) + \varepsilon$$

where y is the vector of log prices.

Imputed prices for each house in each period can then be obtained by inserting its particular mix of characteristics (including the longitude and latitude) into the estimated hedonic model. Finally, the use of the Fisher price index formula gives the HPI.

Other estimation methods for similar problems are also available. For example, Kolbe et al. (2012) estimate their semi-parametric hedonic model with a two-step procedure. After splitting house prices into building and land components, the location value surface is estimated with adaptive weight smoothing. Other examples are Clapp et al. (2002) or McMillen and Redfearn (2010). However, most of the applied non- and semi-parametric approaches in estimating the hedonic model do not compute price indexes. Kagie and van Wezel (2007) estimate their hedonic model using a decision tree model called boosting, but without the use of geospatial data.

The main advantage of non-and semi-parametric methods is that they do not need to assume a functional form in the hedonic model (or in parts of it), and hence avoiding problems of misspecification of the functional form. Comparisons with pure parametric models almost invariably find that non- and semi-parametric models outperform the simple linear models in terms of goodness-of-fit measures or even when, for example, repeat-sales are used as a benchmark (see Hill and Scholz, 2014). Furthermore, these methods include geospatial data in an intuitive and natural way, while other often used approaches as the average characteristics method cannot include these kind of information at all.[2]

3 Practical Considerations

Another demonstration of the flexibility of non- and semi-parametric methods is the ability of dealing with missing characteristics. Hill and Scholz (2014) show the importance of including all available observations (even those missing some of the structural characteristics), otherwise a sample selection bias may occur. If, for example, the age, or number of floors of a house are unknown, but the dwelling is located in a more or less homogeneous neighborhood, then location is highly correlated with both of them. Thus, the estimation of a topographical surface will also capture such effects, even if some characteristics are missing.

The difference of the constructed HPI based on a fully parametric model using, for example, post code dummies as locational characteristics, and a HPI derived from a semi-parametric hedonic-GAM based on longitudes and latitudes, depends partly on how finely defined are the identifiable locational zones in a city. Hill and

[2] It measures intuitively the change in the price of the average house over time. But averaging longitudes and latitudes does not make sense. For example, in the case of a city like Sydney built round a natural harbor the average location may be under water!

Scholz (2014) show that the use of broader geographical areas can induce a bias in the HPI. Using data from Sydney, Australia, from 2001-2011 they find a downward bias during the housing market boom in the first part of the observed period. This can be attributed to systematic changes over time within each zone in the locational quality of houses sold.

4 Concluding Remarks

Location is one of the driving factors in house prices. Hence location should likewise play a prominent role in the construction of house price indexes. The increasing availability of geospatial data has the potential therefore to significantly improve the accuracy of HPIs. However, a consensus has not yet emerged regarding the best way of using geospatial data. In our opinion, a hedonic model estimated using flexible non- and semi-parametric techniques used in combination with the hedonic imputation method are an attractive option in this regard.

References

Anselin, L. (1988), Spatial Econometrics: Methods and Models. Dordrecht: Kluwar Academic Publishers.

Clapp J.M., H.J. Kim, and A.E. Gelfand (2002), Predicting spatial patterns of house prices using LPR and Bayesian smoothing Real Estate Economics 30(4): 505–532.

Corrado, L. and B. Fingleton (2012), Where Is the Economics in Spatial Econometrics? Journal of Regional Science 52(2), 210–239.

De Haan, J. and W.E. Diewert (eds.) (2013), Handbook on Residential Property Price Indices (RPPIs), Luxembourg: Eurostat.

Hill, R. J. (2013), Hedonic Price Indexes for Housing: A Survey, Evaluation and Taxonomy, Journal of Economic Surveys 27(5), 879–914.

Hill, R. J., D. Melser and I. Syed (2009), Measuring a Boom and Bust: The Sydney Housing Market 2001-2006, Journal of Housing Economics 18(3), 193–205.

Hill, R. J. and M. Scholz (2014), Incorporating Geospatial Data in House Price Indexes: A Hedonic Imputation Approach with Splines, Mimeo.

Kagie, M. and M. van Wetzel (2007), Hedonic price models and indices based on boosting applied to the Dutch housing market, Intelligent Systems in Accounting, Finance and Management 15, 85–106.

Kolbe J., R. Schulz, M. Wersing, and A. Werwatz (2012), Location, location, location: Extracting location value from house prices, SFB 649 Discussion Paper 2012-040.

McMillen D. and C. L. Redfearn (2010), Estimation and Hypothesis Testing for Nonparametric Hedonic House Price Functions, Journal of Regional Science 50(3), 712–733.

Nappi-Choulet, I. and T. Maury (2009), A Spatiotemporal Autoregressive Price Index for the Paris Office Property Market, Real Estate Economics 37(2), 305–340.

Rossi-Hansberg, E. and P.D. Sarte (2012), Economics of Housing Externalities, International Encyclopedia of Housing and Home, Vol. 2, 47–50.

Silver, M. (2012), Why House Price Indexes Differ: Measurement and Analysis, IMF Working Paper, WP/12/125.

Shimizu, C. (2014), Estimation of Hedonic Single-Family House Price Function Considering Neighborhood Effect Variables, Sustainability 6, 2946–2960.