Empirical Issues on the Econometrics of Hedonic House Price Indexes

Esmeralda A. Ramalho and Joaquim J.S.Ramalho (Universidade de Évora, Portugal)

The construction of housing price indexes raises many conceptual and practical problems, mainly because houses are heterogenous assets whose prices can only be observed when they are sold. Hedonic pricing methodologies are a prominent approach in this area, which allow not only to compile house price data controlling for quality differences but also the assessment of the value of a particular property in absence of market transactions.

Hedonic pricing methodologies builds upon the idea that different characteristics of a good or service impact differently on their evaluation by consumers. Thus, the first stage of this methodology consists of specifying an hedonic price function, i.e. a function relating transaction prices to the relevant characteristics of the good or service. Some of the model specification issues raised by hedonic regression are common to any econometric analysis. In fact, the specification of the hedonic function requires both the choice of the house attributes to use as explanatory variables and of the model functional form. As it is well known, omitting a relevant covariate or using an incorrect functional form leads in general to inconsistent estimation of the parameters of interest and, hence, of the implicit prices of the house attributes. Other common model specification issue is the presence of heteroskedastic errors, which typically invalidates the usual inference procedures used in regression analysis but does not distort the parameter estimates. An additional issue is the definition of the house strata and/or the time period for which the hedonic function is valid. Finally, hedonic house price empirical studies are likely to be affected by several data issues. For example, in some cases it will be necessary to use weighting adjustments to account for the likely endogeneity of the sampling. In other cases, practitioners will have to deal with outliers, measurement errors, missing data and multicollinearity. All these potential sources of misspecification or misuse of hedonic regression models are carefully reviewed in detail in the paper, in terms of consequences, detection and possible solutions to circumvent and/or disentangle the data limitations. In fact, one of the major aims of this paper is instigating a change in the current practice in the applied hedonic house price literature of choosing a priori a particular hedonic model and overlooking the specification analysis at all.

Using regression techniques, the parameters in the hedonic price function, which may be interpreted as the implicit marginal prices for each characteristic (or some function of it), are estimated. Based on the estimated marginal prices, housing prices can be straightforwardly adjusted in order to remove the idiosyncratic influence of those sources of heterogeneity. Typically, the hedonic regression model is estimated by ordinary least squares (OLS) but more complex methods may be employed. The choice of the appropriate estimation method depends on the functional form adopted and on various model specification (e.g. heteroskedasticity) or data (e.g. outliers, measurement errors, missing data, multicollinearity) issues. This paper reviews the most commonly applied methods to estimate hedonic functions like OLS, weighted least squares and robust estimation methods to data issues, and analyze the main consequences of using those methods in presence of several sources of misspecification.

Given the large variety of misspecification issues that may affect hedonic analyses, it is somewhat surprising that most of the hedonic pricing studies implemented so far have not applied formal specification tests to assess the validity of the assumptions underlying the hedonic regression. However, any hedonic analysis should comprehend the application of a battery of statistical tests to the estimated models, including general tests for model misspecification like the RESET, which are sensitive to a wide variety of departures from the postulated model, and specific tests, which aim to detect particular forms of misspecification

problems of which researchers may be especially suspicious like the instability of the parameters in space and time and endogenous stratification. An integrated approach for both type of tests is proposed, where all tests are implemented as checks for the relevance of a set of artificial regressors, in a robust version that accounts for the presence of heteroskedasticity. The paper also reviews other procedures that may also be used to evaluate the postulated model and discuss some model selection criteria and tests that may be applied to select the best formulation for the hedonic function in cases where the competing formulations are non-nested.

The major aim of this paper is to provide a simple but comprehensive survey of the econometric procedures that are most relevant for the construction of housing price indexes, irrespective of the type of price index and hedonic method employed. This survey is oriented towards the applied economist estimating hedonic models, rather than specialized theorists.