Journal of Property Research ›
Volume 36, 2019 - Issue 3

Articles

Sales comparison approach, multiple regression analysis and the implicit prices of housing

Gaetano Lisi

Pages 272-290 | Received 11 Mar 2019, Accepted 29 Jul 2019, Published online: 16 Aug 2019

Cite this article https://doi.org/10.1080/09599916.2019.1651755

ABSTRACT

By using two of the main evaluation methods, namely the sales comparison approach and the multiple regression analysis, this paper points out the key role of implicit prices of housing characteristics in real estate appraisal. Firstly, this paper confirms the close link existing between the sales comparison approach and the multiple regression analysis. Furthermore, unlike examples in the related literature, this paper highlights the difference between marginal and implicit prices of housing characteristics. Although the related literature expresses awareness of the methodology for calculating marginal and implicit prices, there appears to be less awareness that implicit prices can: 1) lead to an estimate of the house price that is different from the regression model-predicted price; 2) replace the crucial and non-trivial phase of adjustment and reconciliation that characterises the main appraisal method, namely, the popular Sales Comparison

Approach. Finally, an empirical analysis provides evidence of the key role of implicit prices in estimating house value.

KEYWORDS:

Implicit prices | marginal prices | house value | sales comparison approach | multiple regression analysis

Notes

1. The small change should be related to the unit of measurement of each housing characteristic. For example, in the case of the number of bathrooms, the monetary price change expresses the change in the house price when the number of bathrooms changes by one unit (unit change).

2. Recently, Pace and Zhu (2017) focus on the comparison between explicit and implicit prices. They define both in the same (and general) way, namely, the values of housing characteristics. The difference is that the implicit or latent prices are obtained by using mortgage data, while the explicit prices are based on the traditional approach that uses sale prices of housing. Those authors use the term of implicit price because house price does not explicitly appear in the model when using mortgage data (unlike in the standard hedonic model). Concisely, Pace and Zhu (2017) compare two different approaches and do not give two (different) definitions of implicit and explicit price.

3. Because of their precision and time- and cost-saving advantages, real estate mass appraisal methods that employ multiple regression-based models, known as automated valuation models (AVMs), are becoming increasingly prominent in industry practice and have received attention from the academic community (International Association of Assessing Officers (IAAO), 2018). AVMs are used in a host of industries – both public and private – including loan origination, fraud detection, and portfolio valuation (Lou & Robson, 2007), and are promoted and advanced by such organizations as the International Association of Assessing Officers (IAAO).

4. The key insight of this paper is that transaction data varies over time and locations, while two-thirds of homeowners have mortgages that they pay each month. Also, the decision of whether to pay, or not, the mortgage has little variation over the macroeconomic cycle or across locations and, therefore, may have less selection bias

than transaction prices. Precisely, Pace and Zhu (2017) find that their approach, based on mortgage data, works at its best when default is more prevalent (such as in a bust) and works at its worst during a boom. In contrast, the traditional approach based on transaction data performs at its best when transaction volumes are high (such as in a boom) and performs at its worst in a bust. The two approaches, therefore, are not mutually exclusive. Moreover, they are highly correlated.

5. The method proposed by Lai et al. (2008), the so-called 'replication method', aims at determining the optimal weights of comparable property attributes that best duplicate the subject property. Lai et al. (2008) show that when the number of comparable properties is large compared to the number of attributes, the replication method weakly outperforms both the traditional general least squares regression and grid method.

6. This reasoning can be extended to the qualitative variables (such as the quality of the landscape), when they are transformed into discrete variables or into binary variables.

7. Haupt et al. (2010) show that the null hypothesis of the correct specification of the parametric model, proposed by Anglin and Gençay (1996), against the alternative of parametric misspecification, cannot be rejected at any reasonable level of significance. Also, they show that the parametric model proposed by Anglin and Gençay (1996) is a better predictor than the nonparametric specification proposed by Henderson et al. (2007).

8. X and Z, therefore, define the degree or intensity of the housing characteristics such as the number of square metres of the lot size and the number of bathrooms, respectively).

9. Of course, in the case of undesirable characteristics (such as pollution or location of the property in a degraded area), the relative implicit price enters with negative sign.

10. In order to distinguish it from the 'the regression-model predicted price' which is primarily a mathematical-statistical value.

11. The purely mathematical meaning of the constant term (the intercept) in Equation (1) can be proved in a straightforward way by performing the MRA with the dataset used in the paper. The intercept switches from negative to positive values, as well as switches from high to low values, simply by modifying the number of variables (housing features) included in the regression model.

12. Note that only in the case of a linear model, the regression coefficients are monetary price changes or marginal price.

13. Indeed, regression analysis is typically proceeded by univariate analyses to identify the degree of linear relationship between the dependent and independent variables, which helps the analyst to decide what kind of transformation is appropriate in order to improve the goodness of fit of the model.

14. On the important issue of the weights to be assigned to the selected comparable properties in the case of the weighted average, see for example, Colwell et al. (1983), Gau, Lai, and Wang (1992), Epley (1997).

15. A binary variable or dummy variable is a variable that is used in empirical models to represent a non-quantitative characteristic, such as gender, race or precisely location (see, e.g. Hill et al., 2011).

16. Bidanset and Lombard (2014) compare these two popular spatial regression models and shows that GWR achieves more uniform results (precisely, a lower coefficient of dispersion) overall than SLM.

17. Instead, the linear model does not overcome any fundamental statistical test, i.e. correct specification of the model, absence of heteroscedasticity and normal data.

18. In reality, a simple Excel spreadsheet can be used for obtaining the regression coefficients in Table 1. A more sophisticated statistical software (like the one used in this analysis, i.e. STATA 11) is only necessary to carry out more detailed tests on the correctness of the chosen regression model.

19. The regression model-predicted prices, namely both Piˆloglin and Pˆilog, are obtained in a straightforward way using the STATA command ‘levpredict’ that generates predictions of the levels of the dependent variable in logarithmic models, without the bias that arises when the values of the dependent variable in log are (simply) exponentiated. The regression analysis is indeed a statistical rather than a mathematical method. Thus, the error term must also be considered when generating the prices with respect to levels (dollar, euro, etc.).

20. As regards the model choice, this result is not trivial. Note that the logarithmic model-predicted prices are more efficient than the log-linear model-predicted prices (see again Table 3). Furthermore, the logarithmic model has an R^2-adjusted that is

(slightly) higher than the log-linear model (see the Appendix). It follows that the MRA suggests a different result, namely, that the logarithmic model should be preferred.

21. For the sake of simplicity, we do not use the ‘correction’ proposed by Halvorsen and Palmquist (1980) for the coefficients of the dummy variables. However, for a small value of the coefficient of the dummy variable, the ‘correction’ is irrelevant.

22. Note that vj corresponds to Xi in the case of continuous variables, to Zi in the case of discrete variables and to Di in the case of binary or dummy variables.

Additional information

Notes on contributors

Gaetano Lisi

Gaetano Lisi is Associate Professor of Economic policy at the e-Campus University (Novedrate, Italy) and Functionary with expertise in statistical-economic analysis at the Revenue Agency (Rome, Italy).