Review of: "The Residential Property Price Impact of Luas Investments"

Clara Simón de Blas[1]

1 Universidad Rey Juan Carlos

Potential competing interests: No potential competing interests to declare.

The above article propose a model to measure the price impact of developments in light-rail network in Dublin. The paper could benefit of some considerations.

Overall the paper is too long and the author fails in providing a concrete model proposal. Some details of the model are given in the introduction, some others in section 2, 3 and so long. It is quite difficult to follow the proposal. I would suggest a small summary in section 3. Even more, more robust statistical analysis is required to support the conclusions.

Besides, in scientific papers is unusual to use sentences such as "I have done XXX". I would recommend the author to read some scientific papers related to the item under studio.

Introduction and related literature

The article would benefit of a map of the city of Dublin, highlighting the inner and outer area and depicting lines A1 and B1.

Please, explain the rationale under "My model uses the semi-log form where the dependent variable is the log of sales price and the independent variables enter the regression model linearly." Has your dependent variable a problem of exponential variability? Others?

Please explain the rationale under "The database includes an individual calendar day for each sale, but for modelling simplicity I convert the sale dates into quarterly frequency from 1 to 40, covering the 40 quarters (ten calendar years) of the data." Data harmonization frequently summarizes original data into weekly / monthly /annual totals, averages, etc. How this data was aggregated? Why 40?

The paper would benefit of a representation of this data "In the raw data, there were a small number of extreme values. For example, the highest sales price recorded in the database is €139 million. I delete the three observations in the dataset with sales prices greater than €50 million." In a boxplot graph or similar. Where this deleted data anomalous?

"First, I compute the linear distance between each property and each Luas station; then, I find the Luas station that is closest to each property and its distance in kilometres." Do you refer to Euclidean distance? Please clarify

"Finally, I take each property sale within 3 kilometres of the closest Luas and using the Google Distance Matrix API calculate the walking time of each property to the closest Luas." How was the walking travel time estimated? Average speed may be reported, Have the author consider road conditions, safety, etc. that may dissuade citizens to chose the

shortest path given by google?

I suggest the author to explore ARIMAX models, instead of linear regression combined with the nonparametric pricing function in place of the postal zone dummy variables

“difference-in-differences test” I assume the author refers to T-Student average comparison test? This test requires the verification of the normal assumption undelying the data distribution. Has the author test this assumption? I would suggest in this case the use of experimental factors design.