

## Combining Property Transaction Data with a Representative Sample to Obtain a Biased Correct Property Price Index

Alim Furkan Kalay  
The University of Queensland  
[a.kalay@uq.edu.au](mailto:a.kalay@uq.edu.au)

Alicia N. Rambaldi  
The University of Queensland

A number of hedonic-based methods are available to compute property price indices (see Hill and Rambaldi, 2021 for a recent review). It is well understood that property transactions data suffer from sample composition bias, and only some available methods can partially control the sale composition bias by modeling assumptions. The issue arises as it is not possible to choose a random or representative sample of the population of dwellings to sell in a given month/quarter/year. Some types of housing sell in larger proportions depending on the market conditions at the time.

In this paper we use recently developed methods to obtain bias corrected estimates of the hedonic model. The method is double robust. We then compute price indices using standard hedonic imputation methods. The estimation method relies on integrating the sample of market transactions with dwelling characteristics information from a representative auxiliary sample. The paper relies on recent developments from a growing literature in data integration methods (e.g., Chen et al., 2020; Kalay, 2021). We adapt a mass-imputation approach to integrating transactions data and the auxiliary sample. Our approach allows controlling for the composition bias, and it can be implemented with both parametric and nonparametric methods. The model estimation method requires weaker assumptions than conventional methods due to the use of an auxiliary sample.

Our approach requires two data sets: (1) market sales transactions data (containing prices and property characteristics) and (2) an auxiliary sample that contains the same property characteristics but more closely represents the stock of properties in the area/region. We combine individual residential property transactions information for the period 2011 to 2016 sourced from administrative land titles, with the characteristics of each property as reported by dwelling inhabitants in the 2016 Australia Census. This is our auxiliary sample. The linking and de-identification is undertaken by the Australian Bureau of Statistics and covers the state of Victoria in Australia. The data are developed under an Australian Research Council Linkage Project involving the Universities of Queensland and Melbourne, and the Victorian Department of Treasury and Finance. While measures of dwelling stocks are not typically readily available, the use of auxiliary samples, linked data and new statistical methods, provides the possibility to correct one of the major sources of bias in the computation of property price indices.

## REFERENCES

Chen, Y., Li, P., & Wu, C. (2020). Doubly robust inference with nonprobability survey samples. *Journal of the American Statistical Association*, 115(532), 2011-2021.

Hill, R. J., & Rambaldi, A. N. (2021). Hedonic Models and House Price Index Numbers (No. WP152021). School of Economics, University of Queensland, Australia.

Kalay, A. F. (2021). Double Robust Mass-Imputation with Matching Estimators. arXiv preprint arXiv:2110.09275.