# Mass appraisal without statistical estimation: a simplified comparable sales approach based on a spatiotemporal matrix[9]

Sonia Yousfi, PhD Cand.[§]; Jean Dubé, Prof. [¤]; Diego Legros, Prof. [§]; Sotirios Thanos, LECT[¥]
[§]Université de Bourgogne, [¤]Université Laval, [¥] University of Manchester

## Abstract

For mass appraisal in real estate, the hedonic pricing method (HPM) tends to be most commonly used by academics, while the comparable sales approach (CSA) is mostly preferred by professionals. The CSA takes into account information on individual characteristics identifying similar complex goods, spatial proximity reflecting similar spatial amenities and temporal constraints by selecting past sales only.
This paper shows how CSA is a constrained version of a spatial autoregressive model that can be implemented by simple matrix calculations. Using US transaction data, we compare CSA to "a-spatial" HPM results and conduct an out-of-sample exercise to gauge the prediction performance of the two approaches.
The findings suggest that CSA is a very useful tool for mass appraisal, especially when the number of independent variables available is limited.

## Main objective and key idea

- CSA price prediction is based on identifying real estate goods with similar characteristics.
- Focus on the spatial and temporal similarity between comparables that have identical housing attributes and not on finding the most similar comparables in terms of structural characteristics (given the large literature on the topic).[5,6]
- Main aim: provide **a simple and intuitive method for implementing CSA** based on a spatiotemporal weight matrix specification[3] for the selection of comparables or, in this case, spatiotemporal nearest neighbors.
- CSA is shown to be a constrained version of a simple spatial autoregressive (SAR) model extended to contain temporal connections.

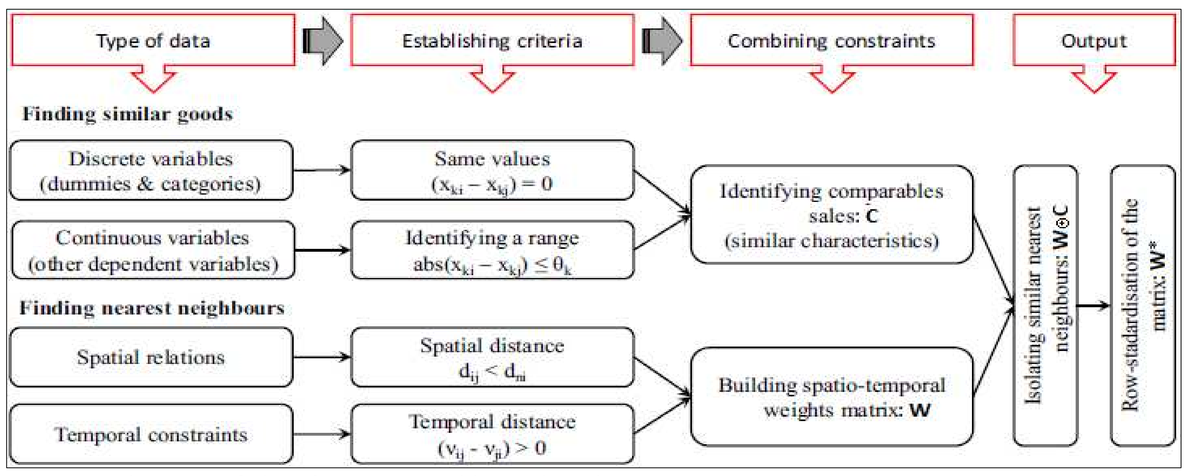


**Figure 1.** The steps of identifying comparables and predicting prices

## Methods and Materials

- Empirical illustration: US data set from Lucas County (Ohio) spanning a timeframe from 1993 to 1998.
- Estimation and comparison of prediction performance conducted by employing an out-of-sample one-step-ahead forecasting approach.
- The empirical application compares the predictive power of the spatiotemporal nearest neighbors CSA to an "a-spatial" HPM.
  - For HPM predictions, we have:

$$\mathrm{Ln}(y_{i,t}) = \alpha + d_{i,t}\delta_t + \sum_k x_{k,i,t}\beta_k + \varepsilon_{i,t}$$

The procedure is repeated for all time periods ($t + 1$, …, $t + s$, …, $t + S$) until the last transaction. This iterative approach results in the coefficients changing over time, which points to a framework similar to CSA, ensuring consistency of comparisons. A simple extension is also proposed and is based on the introduction of a nonlinear effect, known as the gamma transformation.[1,2]

  - For CSA predictions, **Figure 1** illustrates the criteria for identifying similar houses. We have:

$$y_{i,t} = \mathbf{W}^* y_{i,t-s}$$

- Comparison based on the actual sale price (in log), $\ln(y_{i,t})$, and the predicted sale price (in log), $\ln(\hat{y}_{i,t})$ in each time period to predict house prices for the subsequent time period. Seven different statistics are calculated and can be found in **Tab. 1**.

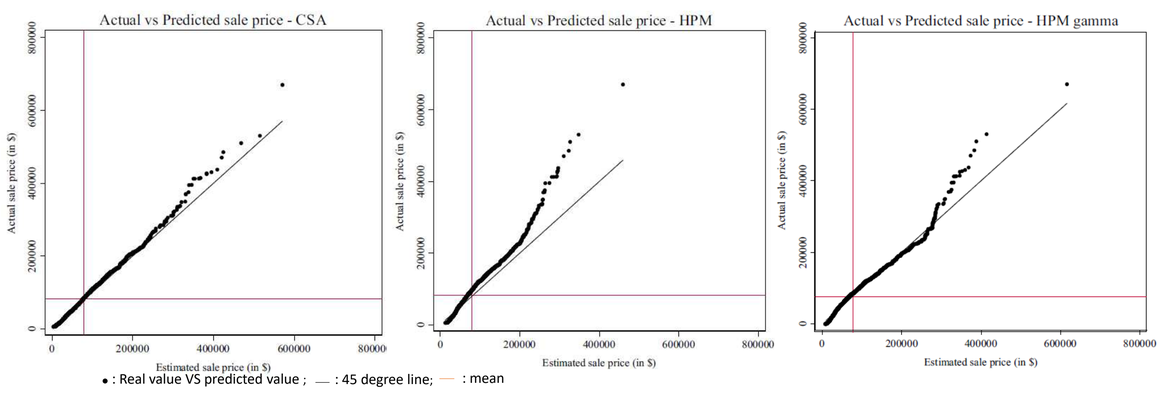


**Figure 2.** CSA and HPM performance comparisons

## Results

- 87.4% of the total out-of-sample observations available in 1998 (3828 transactions) have at least one spatiotemporal neighbor to comparables for prediction and comparison between CSA and HPM.
- Compared to the typical HPM, CSA provides superior out-of-sample prediction performance, but similar performance to the HPM based on the gamma transformation for the continuous variables (see **Tab. 1**).
- MSE, RMSE and MAD show CSA to perform better than the typical HPM but worse than the gamma transformation HPM. Both methods underestimate the sales prices at the high end of the price spectrum, but HPM greatly overestimates the prices at the lower end of the spectrum as well. In all cases, the gap for the prediction of high prices has a more pronounced impact in the HPM (see **Fig. 2**).
- Even though the global performance of the models (Pseudo-R2) is similar for both approaches, CSA provides predictions closer to the actual price.

**Table 1.** Performance and out-of-sample statistics by method

| Statistics | Method | | |
|---|---|---|---|
| | HPM | | CSA |
| (Dependent variable: $\ln(y_{i,t})$) | Typical | Gamma | |
| Number of better predictions | 1574 | 1882 | |
| Mean square error (MSE) | 0.2303 | 0.1549 | 0.1868 |
| Root mean square error (RMSE) | 0.4799 | 0.3936 | 0.4322 |
| Mean absolute deviation (MAD) | 0.3450 | 0.2883 | 0.2965 |
| Mean standard deviation (MSD) | − 0.0208 | − 0.0185 | 0.0155 |
| Coefficient of dispersion (COD) | 0.0384 | 0.0468 | 0.0441 |
| Pseudo-$R^2$ | 0.6144 | 0.7344 | 0.6814 |

N=3828 transactions with at least one neighbor

## Discussion

- With the very limited number of independent variables, we do not necessarily show that HPM is inadequate, but that information constraints do have a serious impact on HPM predictions.
- CSA proves to be a useful tool for predicting prices under information constraints, when only a few characteristics are available:
  - Potentially due to the implicit spatial and temporal patterns captured in the comparables, accounting for characteristics with a spatial structure;
  - Underlining a link between the CSA, SAR and STAR in terms of spatial or spatiotemporal structure.[4,8]
- One of the weaknesses of the CSA is that prediction is possible only if there is a comparable available to make a projection.
- CSA may fail to identify an adequate number of comparables for some observations; thus, not all observations can be included in such analysis:
  - This can be partly improved by relaxing some of the similarity constraints in defining comparables.
- CSA needs to deal with the main HPM disadvantage, the need to capture the implicit price of all characteristics from unbiased regression coefficients:
  - This illustrates the trade-off in CSA regarding the precision in identifying comparables, including the spatial and temporal dimensions that we argue should take center stage in the process.

## Conclusions

- A computationally simple and useful forecasting tool based on spatiotemporal nearest neighbors and a matching estimator approach.[7]
- CSA, based on the limited information of only a few independent variables, outperforms the "a-spatial" HPM.
- From the perspective of the real estate industry, the CSA proposed in this paper:
  - Can potentially provide a valuable tool as it does not require statistical analysis;
  - Works well under informational constraints;
  - Improves the comparable selection process by reducing arbitrariness and subjectivity.
- CSA can also be potentially combined with other HPM-based methods, such as the difference-in-differences (DID) approach, to evaluate the impact of extrinsic amenities on house prices.

## Contact

Sonia Yousfi
Laboratoire d'Économie de Dijon
Email: sonia.yousfi@u-bourgogne.fr
Website: ledi.u-bourgogne.fr

## References

1. Des Rosiers F, Lagana A, Thériault M, Beaudoin M (1996) Shopping centers and house values: an empirical investigation. J Prop Valuat Invest 14(4):41–63
2. Des Rosiers F, Lagana A, Thériault M (2001) Size and proximity effects of primary schools on surrounding house values. J Prop Res 18(2):1–20
3. Dubé J, Legros D (2013) Dealing with spatial data pooled over time in statistical models. Lett Spat Resour Sci 6(1):1–18
4. Dubé J, Legros D, Thanos S (2018) Past price "memory" in the housing market: testing the performance of different spatio-temporal specifications. Spat Econ Anal 13(1):118–138
5. Krause A, Kummerow M (2011) An iterative approach to minimizing valuation errors using an automated comparable sales model. J Prop Tax Assess Adm 8:39–52
6. Pagourtzi E, Assimakopoulos V, Hatzichristos T, French N (2003) Real estate appraisal: a review of valuation methods. J Prop Invest Finance 21(4):383–401
7. Rosenbaum PR, Rubin DB (1983) The central role of the propensity score in observational studies for causal effects. Biometrika 70:41–55
8. Thanos S, Dubé J, Legros D (2016) Putting time into space: the temporal coherence of spatial applications in the housing market. Reg Sci Urb Econ 58:78–88
9. Yousfi, S., Legros, D., Dubé, J., Thanos, S. (2020) Mass appraisal without statistical estimation: a simplified comparable sales approach based on a spatiotemporal matrix. The Annals of Regional Science, 64: 349–365.