Report: Review of Automated Valuation Model (AVM) Methodology for Wales Revaluation (Report 1 of 2)

For: Valuation Office Agency, United Kingdom

By: Paul Bidanset, Founder & Research Scientist, Center for Appraisal Research and Technology (CART)

Project Background

The United Kingdom's Valuation Office Agency (VOA) are developing automated valuation models (AVMs) to revalue domestic hereditaments of Wales. In support of their research and development of the AVMs, VOA sought academic critical friends with relevant expertise to provide support and assurance through a quality control review. Specifically, the critical friends were expected to:

1. Advise on statistical and modeling techniques employed by VOA to ensure appropriate use, specifically, of geographically weighted regression (GWR) and its extension for comparability weighting (GCWR), with regard to published academic research and international best practices among government property tax/valuation authorities
2. Advise on alignment of VOA modeling procedures with international standards, such as those published by the International Association of Assessing Officers (IAAO)
3. Recommend any additional statistical test and analyses VOA should incorporate in their research and modelling procedures

Paul Bidanset and Peadar Davis, PhD were asked to serve as critical friends based on their authorship of the paper "*[a]ccounting for locational, temporal, and physical similarity of residential sales in mass appraisal modeling: the development and application of geographically, temporally, and characteristically weighted regression*"[1]. Project deliverables include a 2-page report by each summarizing project findings and recommendations.

Summary of Findings:

1. **The VOA modelling team have a strong comprehension of the mathematical foundation of AVMs used for property taxation, including GWR and its extensions, as outlined in academic literature.** The team understand the theoretical and applied limitations and constraints of the algorithm - and devise solutions for more appropriate configurations of the model. They anticipate issues that may negatively impact model performance and implementation "in the real world" and adapt to avoid them. They identify appropriate algorithmic extensions (e.g.MX-GWR) and substitutes (e.g., SRF) that more adequately handle shortcomings of GWR.
2. **VOA modeling methodology, including specification, calibration/optimization, and quality control review, strictly follows industry best practices and standards.** Modeling datasets are appropriately split into training and testing samples to protect against overfitting. Ratio studies are performed to test valuation accuracy and uniformity. Ratio studies are appropriately breaking out by location and other classifications (e.g., house type).
3. **The AVMs created by VOA yield overall reliable valuations:** The overall performance of valuations produced by VOA AVMs is quite good and in-line with models of comparable data and markets. The ratio study tests indicate a large degree of accuracy and uniformity across the valuations produced by the AVMs. The team has successfully identified areas that the models may not be appropriate for and will require additional consideration.
4. **AVM methodology is well-documented.** Reporting language is sophisticated enough to not omit essential complex components (e.g., mathematical or statistical concepts), but explains simply, in a straightforward

[1] Bidanset, P., McCord, M., Lombard, J. A., Davis, P., & McCluskey, W. (2018). Accounting for locational, temporal, and physical similarity of residential sales in mass appraisal modeling: the development and application of geographically, temporally, and characteristically weighted regression. *Journal of Property Tax Assessment and Administration*, 14(2), 5-13.

manner for stakeholders of all professional backgrounds. Visualizations are clear and helpful, and rationale is explained to provide an intuitive understanding. Such documentation will help ensure continued progress through office growth and succession and will promote defensibility and acceptance among stakeholders.

Recommendations:

1. When performing ratio studies, include confidence intervals for the median ratio, COD, PRD, and PRB. According to the IAAO standard on ratio studies, confidence intervals may be calculated around the median ratio, COD (using bootstrapping), PRD (using bootstrapping), and PRB, and any value within that range cannot be ruled out as the "true" value. Confidence intervals are used in the industry to cast doubt on results that indicate noncompliance. While confidence intervals may be programmed from scratch, several packages exist in R including DescTools, which has a MedianCI() function, as well as assessr and taxr, which offer confidence interval functions for COD, PRD, and PRB.
2. When performing ratio studies, include a test for the statistical significance (t-value or p-value) of the PRB coefficient. This should be reported along with the PRB coefficient to protect against false positive indication of vertical inequity.
3. When performing ratio studies, include a test for a Gini coefficient to test for vertical inequity. While the PRD and PRB are useful indicators of vertical equity, the IAAO has recognized that they suffer from shortcomings that can sometimes distort their reliability (e.g., in the presence of extreme outliers or sampling issues). Recent research by Quintos[2] (2020) presents an alternative metric called a Gini coefficient that is suggested to be a more reliable indicator.
4. The criteria for the weighting allocation used for the comparable sales selection similarity scoring should be elaborated. VOA should anticipate having to explain why certain thresholds/values were chosen (e.g., why is distance weighted by .33 and not .32, .34, or even 50?). Perhaps it was the case that these thresholds were based on expert analysis, or that they these values yielded the most accurate results; this should be addressed to promote defensibility and acceptance of AVM methodology by stakeholders.
5. Evaluate model results when applied to universe of properties (sold and unsold). As a next step (if not already done or planned), models should be applied to the universe of properties. Explore values in the tails in the high and low ends of different variables (price, size, age, etc.), as well as rare or unique properties, and areas that do not transact often. As these properties are usually not represented in modeling datasets, additional calibration may be required to assign them reasonable valuations.
6. Continue to explore other algorithms for optimization. Many algorithms exist and new ones are continually being developed. VOA should continue to compare performance of models against variations of GWR (MX-GWR), as well as new modifications (e.g. fast/scalable GWR), spatial interpolation models (e.g., co-kriging), and others. More simplistic baseline regression (e.g., log-level, level-level) models should also be run, perhaps even at local (e.g., township, market area) level. Sometimes these models, despite their comparative computational simplicity, can still yield reasonable, and even sometimes better, ratio study scores. Their transparency of value contributions can also be very favorable to governments for reasons of increases explain-ability to stakeholders.

[2] Quintos Ph D, C. (2020). A Gini measure for vertical equity in property assessments. Journal of Property Tax Assessment & Administration, 17(2), 2.