How data can improve property tax implementation in Rwanda

Jonathan Bower, Kaspar Kundert, Patrick McSharry, and Paul Brimble

- Rwanda introduced a new property tax law which entered into force on 1 January 2019. This property tax sought to increase municipal revenues from a low base and empower local government as part of a wider decentralisation process.
- IGC and the consulting firm GOPA, funded by GIZ, generated a prototype Computer Assisted Mass Appraisal (CAMA) for Kigali applying machine learning to remote sensing and GIS data.
- This policy brief draws out some lessons and unanticipated benefits of using a CAMA for administering Rwanda's property tax on buildings.
- Having a relatively simple model boosts transparency, ease of replicability and accuracy, and additional revenue due to a CAMA could greatly exceed its cost.
- It is possible to check which properties have a building, check for missing building declarations and for illegal claims of building tax exemption, as well as using the model to assess whether the current land tax rates are consistent as a percentage of land value.

POLICY BRIEF RWA-20141

MARCH 2022

This project was funded by IGC Rwanda

www.theigc.org

DIRECTED BY

FUNDED BY

Introduction

In 2018, Rwanda introduced a new property tax law, No. 75/2018, which entered into force on 1 January 2019. This property tax sought to increase municipal revenues from a low base and empower local government as part of a wider decentralisation process. The new law amended the previous law and now also taxes buildings. Furthermore, unlike land, this tax is based on a percentage of the building value. To apply this building tax, a measure of the value of buildings is necessary. The law states that:

> **"The valuation of immovable property is done either by a certified valuer or by a computerised mass valuation system. However, for self-assessment of tax on immovable property, the acquisition value and the construction value of a building remain acceptable until valuation by a certified valuer or by computerised mass appraisal system is effective"[1] (Article 15)**

In this context, this policy brief draws out some lessons learned and some unanticipated benefits of using a computerised mass appraisal system, otherwise known as a Computer Assisted Mass Appraisal or CAMA, for administering the tax on buildings under law No. 75/2018.

Throughout 2020 and 2021, a team of collaborators commissioned by GIZ's Decentralization and Good Governance Programme and IGC, including Kaspar Kundert working with the consulting firm GOPA on behalf of GIZ, Paul Brimble from the University of Oxford and Patrick McSharry from Carnegie Mellon University produced a property valuation model that predicts building values in Kigali, with the goal of it being useful for the Rwanda Revenue Authority (RRA). This exercise resulted in a CAMA prototype and a report submitted to RRA containing many valuable technical observations and recommendations which, if implemented, could lead to improved and more equitable building taxation in Rwanda. The report also contains a CAMA how-to guide – a kind of step-by-step standard operating procedure for applying CAMA in Rwanda. The development of the CAMA prototype built on an earlier IGC-funded detailed property valuation exercise that attempted to predict 2015 property values in Kigali, resulting in a paper by Brimble et al. (2020) called "Using machine learning and remote sensing to value property in Kigali".

We present insights from both exercises in this policy brief. We offer reflections in two broad categories: lessons learned from generating the model behind a CAMA for Rwanda, and how a CAMA could be used for monitoring and compliance. We do not discuss the important areas of data governance, ICT systems or institutional capacity in RRA that need to be addressed to implement a CAMA efficiently.

1 https://www.rra.gov.rw/fileadmin/user_upload/decentralized_taxes_law.pdf

Lessons learned

1. Generating a CAMA model in Rwanda

We present selected insights here, drawn from our experience generating property valuation models for Kigali, which may also be useful in other country contexts.

1.1: Having a relatively simple model boosts transparency, ease of replicability and accuracy

It is important to have a property valuation model with a sufficiently low number of variables and a relatively simple functional format, for three reasons.

- First, for **transparency**, if any government official, or possibly a tax-paying member of the public, asks what variables are taken into account when modelling property and building values, a simpler model with a lower number of variables is easier to explain.[2] A model with a functional form that cannot be easily explained to a non-statistician is too complex to be transparent; thus we recommend a format that is readily understandable such as a linear Ordinary Least Squares (OLS) model: $Y = B_0 + B_1.X_1 + B_2.X_2 + ... + B_n.X_n$. The more stakeholders understand and trust the model, the greater the confidence in it.

- Second, for **replicability**, if a model is to be reproduced on a regular basis, it is important to keep the effort and cost of replicability down. Having too many variables would require more data acquisition and processing, which takes time, capacity, and money. However, if variables important for property values are missing, this could decrease accuracy unacceptably, so a balance needs to be struck. Whilst the final model in the updated property valuation exercise has 50 variables, they come from just 9 unique data sources.[3]

- Thirdly, for **accuracy**, two things are important: i) selecting the right variables in the model, as noted above, and ii) a simple functional form such as a linear OLS model, which turns out to perform better than complex approaches for the purpose of a CAMA, because the task of a CAMA is to predict values out-of-sample – see footnote for explanation.[4] Occam's razor, also known as the principle of parsimony applies here: it states that "among competing hypotheses, the one with the fewest assumptions should be selected".

2 The GOPA report suggests an explanation to the public might be "Your property value was calculated by CAMA on the basis of the city Master Plan, the buildings on your plot, the size of your land, the accessibility of the plot and your neighbourhood."

3 The Master Plan zoning regulations, Ecopia building footprints, Land Administration Information System at RLMUA, Open Street Map, Normalised Difference Vegetation Index, Bus stops (TAS), Markets (MINICOM), Nightlights and Schools

4 The role of a CAMA is to use property sales values to accurately predict the values of properties for which sales data or valuation data on their values are unavailable. This means a CAMA has to accurately predict values out-of-sample, which means that it cannot "overfit" the quirks of the data in-sample, but must generalize well to all properties to which it will be applied. Brimble and McSharry find that a range of complex spatial econometric models are outperformed or matched on accuracy by simple OLS models when modelling out-of-sample values. Thus a simple OLS model – although with the variables selected by machine learning and cross-validation, is advised. The methodology is fully explained in Brimble et. al. (2020) "Using machine learning and remote sensing to value property in Kigali".

1.2: Certain types of characteristics reliably predict property values

From both valuation exercises, Brimble and McSharry found that for total property value the famous adage "location, location, location" is confirmed. Locational variables, especially distance to a road, a bus stop and a bus route, transformations of these variables (such as logs or quadratic transformations), are consistently important, which underlines the importance of the interplay between urban connectivity and property prices. Other land use, land cover and land zoning variables, especially relating to nature or vegetation cover, were consistently significant. Nightlights data were also predictive - they correlate to overall economic activity, so it makes sense that an area with more economic activity would have more expensive properties.

Structural building variables - especially building area and volume, taken from the IGC-funded building footprint data, are also consistently important for total property value; they are, of course, vital for a property valuation exercise that seeks to include buildings. Building footprints are enormously valuable, but do not tell the whole story of the value of a building: the most obvious gap in the otherwise rich dataset drawn upon by Brimble and McSharry during their modelling exercise is a set of more detailed building characteristics such as wall material, roof material, number of windows, and number of rooms. This has been collected in partial form by RRA as part of the annual property tax declaration exercise and by Rwanda's Institute of Real Property Valuers, but it does not have the degree of completeness and reliability necessary to use it in a comprehensive valuation exercise.

1.3: On average, CAMA building value predictions in the updated model are slightly higher than self-declared building values

The GIZ report published by GOPA found that for the properties in the study area which had buildings and declarations of value for those buildings, and when 109 buildings with implausibly high values predicted by the model (above 5 billion RWF) were eliminated, the model predicated that the total value of the buildings was 721 billion RWF, whereas the self-declared value of the buildings was 507 billion, a difference of 42%. Excluding properties valued at less than 10,000 RWF (10 USD) reduced the difference to just 9%. This implies that on average, the model is not far away from self-declared values that are not set at implausibly low levels and suggests that data cleaning could further improve results.

1.4: The revenue generated due to a CAMA could greatly exceed its cost

Implementing a CAMA requires an up-front investment, as well as institutional technical capacity and an operating budget in RRA to continue to run it. From the Government's perspective, the revenue that could be generated from the CAMA greatly exceeds the cost in direct proportion to how thoroughly it is implemented and in proportion to how much it improves compliance. According to the report, should RRA succeed in recouping the building tax on 20%[5] of the properties that

5 If those percentages can increase, the benefit cost ratio will increase, but this would require RRA to follow up with the individual owners to ask them to declare their buildings

have buildings but have not declared them, CAMA would generate 4 – 7 times the amount of money it costs to run it. By recouping on top of that the building tax on 20% of the obviously under-declared properties, CAMA helps to recoup almost 10 times the amount it costs to sustain the system. If a stricter definition of under-declared properties – that is, properties that are under-valued in a self-declaration – is applied rather than the 10,000 RWF threshold applied in the GOPA report, the benefit cost ratio could rise considerably. The benefit cost ratio would also increase if a CAMA was applied over a larger geographical area, either in the entire Kigali Province or both Kigali and the secondary cities.

1.5: The technical difficulty of estimating building values suggests a review of the way a CAMA is used during the next round of property tax reform in Rwanda

Rwanda's property tax law does not levy a single tax on each entire property but specifies that land taxes and building taxes are determined separately in different ways: land taxes are taxed at a rate per square metre whereas buildings are taxed at a percentage of the building value. Thus it is necessary to value a building independently from the land upon which it is built.

However, when generating a model to predict building values, sales data distinguishing between building and land values for calibrating the model were unavailable. This means that the accuracy of any model predicting building values cannot be verified. To get around this difficulty, Brimble & McSharry extracted building values by estimating whole property values first, and then subtracting predicted land values based on the variables in the model relating to land only. However, whilst this approach is systematic and reasonable, its accuracy is not verifiable as noted.

The next round of property tax reform in Rwanda – which may be some years from now – might consider applying the property tax as a percentage of the value of the entire property rather than being split between buildings and land. This would be technically much easier to execute with a CAMA, and accuracy (in the form of R-squared) would be observable. The issue of a building tax exemption for the first property with a building on it would be politically necessary to address, but could be dealt with by applying a decreased percentage of the total property value (land and building) for "first house owners".

1.6: Making data-driven property tax modelling repeatable

Preparing the input data for, and then predicting property values in a CAMA remains a complex, technical endeavour requiring adequately skilled staff, either within RRA itself, or within an organisation to which such computations could be outsourced.

With the law calling for properties values to be re-assessed whenever they change by more than 20% (e.g. after the renovation of a house) or at least every 5 years, it must be possible to run the CAMA model regularly. To ensure reliable property predications which are comparable over time, the operators of CAMA must be enabled to adhere to the same methods and procedures from year to year. With repeatability in mind, GIZ with its implementing partner GOPA complemented the development of the CAMA prototype with a 55-page document containing detailed descriptions for each and every step executed in the form of standard operating procedures (SOP) for operating CAMA at RRA.

2. How CAMA could be used to monitor and enforce compliance with the building tax

The process of developing a CAMA prototype led the team to understand the different practical and effective types of analysis that a CAMA enables. We describe here the ways in which the data can be used to assist with monitoring and enforcement of property tax compliance and implementation in a fair way.

2.1: It is possible to check which properties have a building, and thus find undeclared buildings

The team used building footprints as an input into the CAMA model, which allows for the possibility to easily check, for every parcel in the study area, whether it has a building or not. It is readily possible to cross-check this with the data on self-declarations that was collected by RRA in 2020, and see whether buildings exist (or existed in 2019 when the building footprint data were collected) which are not declared. It is then possible to calculate, using the CAMA building value predictions, how much building tax has been lost each year through failure to declare buildings.

2.2: It is possible to check for illegal claims of building tax exemption

The property tax law allows owners of just one property with a building on it to be exempt from building tax, but they must pay tax on buildings on any additional properties they own. For the properties that are declared to have building tax exemptions, it is possible to check whether these claims are permitted by checking whether there is just one building tax exemption per TIN number. This could lead to a text message and email being sent to owners with more than one exemption claim, to ask them to pick just one, and pay tax on the other. This would be likely to increase compliance, and hence, tax revenue.

2.3: It is possible to identify properties with under-declarations of building value

As implied in sections 1.3 and 1.4, using the CAMA model predictions, it is possible to compare whether the self-declarations are roughly in line with the model predictions. Of the 89,000 properties in the study area, the implementing team of GIZ through GOPA found some 20,000 had buildings which were valued at 10,000 RWF or less. However, it would be possible to experiment with different thresholds or methods of how to identify under-valued buildings. The property tax law states that:

"if the value difference between the taxpayer's self-assessment and the tax administration's counter-valuation is more than twenty percent (20%), the value from counter-valuation will be regarded as the final market value" (Article 14)

Whilst this may make a threshold of 80% seem logical to determine whether self-declarations of building value are undervalued relative to the CAMA valuation, we argue otherwise. The CAMA model is not driven by sufficiently detailed data on buildings to understand whether it is accurate to within 20% of a building's "true" value. Also, as noted, its accuracy on building valuations is not observable anyway. This is why we recommend using a more conservative method to identify the buildings that are under-valued: either an absolute cutoff that leaves no question of whether a building is under-valued such as 10,000 RWF (or even a much higher figure such as 200,000 RWF), or a relatively low percentage of the CAMA value such as 10%. It follows that raising the threshold from 10,000 RWF to, say, 200,000 RWF, or by moving from 10,000 RWF to 10% of the CAMA building value, would vastly increase the number of properties determined to have under-valued buildings.

After a building is determined to be under-valued, RRA could then send a text message to its owner asking them to log into the Rwanda Automated Local Government Taxes Management System and re-value their building, possibly with a suggested value from RRA.

2.4: It is possible to use land sales data and/or predicted land values to check whether the current land rates are consistent as a percentage of land value

If property tax on land was in the form of a percentage rate on land value, this would ensure that taxes are applied in a fair way: an owner with land half as valuable as land held by another owner would pay half as much tax, rather than – for example – the same amount. However, the current system is that for each land use type, a range of permissible values is specified by a Ministerial Order and then the rate is chosen at the District level. Whilst the current system certainly does try to apply higher land taxes per square metre to more valuable land, it does not explicitly pin the tax rate to the land value but to the land type. Moreover, the CAMA was not used to guide the range of tax rates specified in the Ministerial order. This does not guarantee that in some parts of the country the land tax regime will not be potentially unfair in terms of a widely varying percentage of land value being applied. The CAMA model could identify whether this is the case, and could potentially be used to inform the next Ministerial Order on land tax rates.

Conclusion

If a transparent, replicable and accurate CAMA, and the data behind it, are used to help implement Rwanda's property tax on buildings, this could significantly improve the fairness, consistency and revenue from the tax, especially by detecting implausible self-declarations. A CAMA requires institutional and financial capacity to implement, but would be likely to bring in several times as much revenue as it costs to implement. Moreover, an updated CAMA could inform further property tax reforms whenever they eventually happen.