IFS

Institute for Fiscal Studies

Justine Knebelmann
Victor Pouliquen
Bassirou Sarr

Working paper

26/34

Discretion versus algorithms: bureaucrats, tax equity and acceptability

Economic and Social Research Council

DISCRETION VERSUS ALGORITHMS: BUREAUCRATS, TAX EQUITY AND ACCEPTABILITY*

Justine Knebelmann
Victor Pouliquen
Bassirou Sarr

Abstract

We study how replacing bureaucrats' discretion with algorithmic assessment affects the accuracy, equity, and public acceptance of tax decisions. Using Senegal's first digital property tax census in the capital, we experimentally vary reliance on bureaucrats versus algorithms to value properties, and compare results to benchmark values provided by real estate assessors. Algorithms significantly improve accuracy and both horizontal and vertical tax equity, with fully rule-based systems outperforming hybrid methods. Performance gaps reflect bureaucrats' limited knowledge rather than collusion. Conditional on liability, taxpayers are indifferent to algorithmic decisions, making algorithms a promising tool to expand the tax base without political backlash.

JEL Codes: D63, D73, H71, O12, O14

Keywords: Algorithm, Bureaucrats, Discretion, Taxation, Equity

*This version: April 24, 2026. Corresponding author: Knebelmann, Sciences Po, Department of Economics justine.knebelmann@sciencespo.fr. Pouliquen: Paris Dauphine University victor.pouliquen@dauphine.psl.eu. Sarr: Ministry of Finance, Senegal basarr@minfinances.sn. We thank the Senegalese Ministry of Finance for excellent collaboration. We are grateful for constructive comments received from seminar participants at MIT, PSE, Sciences Po, NBER Public Economics Program, IIPF, NEUDC, World Bank Tax Conference, Zurich Public Finance and Development Conference and from Pierre Bachas, Augustin Bergeron, Michael Best, Antoine Bozio, Anne Brockmeyer, Denis Cogneau, Esther Duflo, Pascaline Dupas, François Gerard, Marc Gurgand, Jeanne Hagenbach, Anders Jensen, Jonas Hjort, Clément Imbert, Oyebola Okunogbe, Benjamin Olken, Monica Singhal, Tavneet Suri, Jonathan Weigel. This project was funded by EDI, J-PAL DigiFI and FID, with complementary funding by ICTD, J-PAL GI, EUR-PGSE, IGC, the World Bank. We thank Samuel Allain, Aramata Badji, Suzanne Bonmartel, Adrien Ciret, Bilal Choho, Ndeye Rokhaya Diao, Papa Daouda Diene, Oumar Fara Diop, Mor Tacko Hane, Ousseynou Niang, Nicolas Orgeira, Barbara Perez for their contributions to project operations and research assistance, and the *Centre National d'Etudes Spatiales* Geosud program for satellite images. This RCT was registered as AEARCTR-0002277. The views expressed in this paper are our own and do not reflect the views of the Ministry of Finance of Senegal. All errors are our own.

1 Introduction

An effective state is central to economic development (Besley and Persson, 2009), yet development itself helps build state capacity (Finan et al., 2017). This complementarity raises the question of how countries can enter this virtuous cycle. New technologies have sparked enthusiasm among researchers and policymakers as potential leapfrogging tools to overcome deep-rooted constraints (Jack and Suri, 2014; Muralidharan et al., 2016). Among these, algorithmic decision-making is diffusing rapidly: core government functions—ranging from recruiting civil servants to allocating financial assistance—increasingly rely on algorithms, making them a salient example of a new public-sector technology (OECD, 2025).

Algorithms have the potential to improve the accuracy and equity of policy outcomes if they replace unqualified, biased, or self-interested human decision-makers (Angelova et al., 2025; Haseeb and Vyborny, 2022). Conversely, there is a concern that new technologies may not live up to their promise. In particular, algorithms may undermine equity if they lead to discrimination against certain groups (Obermeyer et al., 2019), or if officials' discernment cannot be adequately captured by a systematized process (Duflo et al., 2018). Even when algorithms are more accurate, their introduction may face reluctance from citizens (Burton et al., 2020; Browne et al., 2025).

Despite their rapid global adoption, evidence from settings that allow a comprehensive assessment of algorithmic decision-making along these three key dimensions – accuracy, equity and acceptability – is lacking.

We provide such evidence by studying the introduction of an algorithm to replace bureaucrats' discretion in tax assessment in Dakar, the capital of Senegal. We leveraged the roll-out of the country's first digital property tax census and experimentally varied how the tax base – property value – was determined. Ninety-four cadastral sections (neighborhoods), covering over 41,600 properties, were randomly assigned to two treatment arms. In the Discretion arm, bureaucrats relied on their judgment to assess property values. This is the status quo and a reasonable approach in a context where there is no pre-existing information on real estate values. In the Algorithm arm, valuations were *rule-based*: an algorithm predicted property values using satellite imagery combined with observable characteristics recorded by bureaucrats. Third, we applied an alternative *pure-rule* algorithm to this arm, that predicted values using only remotely collected data, without any bureaucrat input.

We evaluate these alternatives against benchmark market valuations provided by nationally certified real estate assessors for a subset of 2,289 properties. These values are the best available professional benchmark. We use them as our "ground truth", while also verifying that the assessors have prior experience in all areas of the study and that their values correlate with those from alternative sources, which are extremely scarce in this context. These values correspond to what the government aims to obtain – absent resource constraints, it would resort to such assessors, whose training and remuneration exceed the ones of the field bureaucrats.

Our setting is uniquely well suited for the comprehensive evaluation of an algorithm. First, we independently observe the targeted outcome (property values), allowing us to avoid the selective labeling problem that arises in much of the literature (Kleinberg et al., 2018), whereby outcomes are only observed for a selected subset of cases. Second, to compare the algorithm to the status quo, we also need to evaluate bureaucrats. Our ability to observe their valuations individually provides us with an objective measure of their performance, overcoming a well-known challenge in the study of government effectiveness (Best et al., 2023). Finally, determining how equitable an algorithm is usually hinges on a framework that the researcher needs to define (Corbett-Davies et al., 2023). Here instead, we rely on standard (property) tax equity concepts (Avenancio-León and Howard, 2022; Black et al., 2022) making minimal assumptions.[1]

We provide three main results. First, replacing bureaucrats' discretion with algorithms substantially improves the accuracy of tax assessments. In comparison to the discretionary process, the rule-based algorithm reduces the tax base gap – defined as the absolute difference between tax roll value and market value – by 45 percent, and the pure rule algorithm by 56 percent.

Second, switching to algorithms substantially improves both horizontal equity – similar properties facing similar tax bills – and vertical equity – more expensive properties facing higher tax bills. Under discretion, tax bills for properties with similar market values vary widely, with bureaucrat fixed effects explaining 34 percent of the variation in the tax base gap. Bureaucrats also systematically undervalue properties, especially expensive ones, resulting in a regressive tax profile. The effective tax rate is 3.8 percent in the lowest quintile and only 1.7 percent in the top quintile, compared to the expected 4.4 and 8.6 percent implied by the tax code. The rule-based and pure-rule approaches significantly increase progressivity, raising the top-quintile rate to 5.2 and 6.6 percent, respectively.

The pure-rule algorithm outperforms the rule-based process despite weaker model fit statistics

[1]We use the term *equity*, rather than *fairness* which is more frequent in the literature on algorithms. Conceptually, the two notions are the same here. We take the statutory design of the property tax as given and optimal taxation considerations are beyond the scope of this paper.

in the calibration. This is due to distortions introduced by bureaucrats entering erroneous property characteristics. Thus, even limited discretion significantly affects the tax profile.

Third, we examine the political acceptability of algorithmic tax assessment. Bureaucrat survey responses show their support of using an algorithm to determine property values. Using an experimental survey module to document taxpayers' preferences, we find no aversion to the algorithm overall. Leveraging a hypothetical tax bill exercise, which cross-randomizes tax amount with valuation method, we show that algorithmic assessment initially has no significant effect on willingness to pay the tax. Once taxpayers are informed that the algorithm increases equity, a modest preference for it emerges. These findings suggest that scaling-up the use of an algorithm is unlikely to trigger political backlash or reduce compliance.

We investigate the mechanisms behind bureaucrats' valuations and show that lack of knowledge, rather than collusion drives our results, in line with the idea of *passive* rather than *active waste* following Bandiera et al. (2009). In a lab-in-the-field experiment, bureaucrats were shown photographs of properties and asked to estimate their values. They undervalue expensive properties in the lab, where there is no scope for personal gain, just as they do in the field. Consistently, the undervaluation gradient is similar regardless of whether bureaucrats interacted with owners during the field visit. This points to limited knowledge rather than collusion as the key mechanism.

We further examine the implicit algorithm that bureaucrats appear to follow. Their values are poorly explained by objective property features, and only modestly respond to economic characteristics such as household income. We also rule out bureaucrats maximizing subsequent tax payments or the equity of the payment profile. Bureaucrats are biased: in the lab-in-the-field, the same property picture is assigned a 31 percent lower value when the owner is described as retired, associated with being vulnerable in this context (although we verify that retirees do not have lower property values nor incomes).

Finally, we show that neither increased training, learning-by-doing, and screening of bureaucrats, nor incorporating their inputs into the algorithm, would substantially improve performance.

We do observe a trade-off for the lowest-value properties, where bureaucrats retain some advantages. In this market segment, they are less likely than the algorithm to overvalue, and seem to better leverage local information. When they meet the owner or when the property is rented, the gap relative to the algorithm is substantially smaller. Therefore, if the government aims to minimize the risk of overvaluing low-value properties while maximizing overall accuracy, an optimal policy is to apply the discretionary method to the bottom quintile of properties and use the pure rule elsewhere.

The algorithms were calibrated on a sample of 4,916 assessor values and were designed with the administration and international practitioners. We alleviate the concern that our results would mechanically stem from the fact that assessor values are also used as the benchmark in several ways. To avoid any risk of overfitting, we systematically use cross-validation and out-of-sample tests, and provide robustness checks where results are only estimated on observations excluded from the calibration. In addition, we recalibrate the algorithms using property values obtained through our surveys. Although they exhibit lower predictive performance, these algorithms still outperform discretion.

Overall, our results indicate that the pure-rule algorithm outperforms alternative methods for expanding the tax net at scale in an equitable manner. It is highly cost-effective, with assessed amounts rising 68 percent relative to the discretionary method. Moreover, the ratio of paid amounts to assessment costs increases to 35 from 8.5 under the discretionary method.[2]

To the best of our knowledge, this study is the first to evaluate an algorithm across all three dimensions—accuracy, equity, and acceptability—while rigorously identifying the sources of flaws in the discretionary process it is designed to replace.

We first contribute to the literature on the effects of algorithms on economic outcomes (Mullainathan, 2025), with a specific focus on equity.[3] While previous work has investigated what leads a decision-maker to follow an algorithm's prediction (Agarwal et al., 2025), we are not aware of prior research documenting citizens' acceptance of the use of algorithms.[4] We expand the nascent literature on the use of algorithms for public policy (Aiken et al., 2022; Battaglini et al., 2025; Black et al., 2022; Bachas et al., 2025) by providing the first experimental comparison of algorithmic versus discretionary methods in the creation of the tax roll.[5] An advantage in comparison to prior research is our unique ability to fully observe the learning sample and the "ingredients" for an

[2] The risk of dynamic adjustments by taxpayers is also mitigated, given that the pure rule uses only property location and built area as inputs.

[3] Aiken et al. (2022) provide a related application by analyzing the Togolese government's use of an algorithm to target anti-poverty transfers. In comparison to ours, their study is not equipped to fully investigate the bureaucratic sources of the poor performance of the non-algorithmic process.

[4] The existing studies on aversion to algorithms, mostly by scholars in psychology and relying on laboratory or online experiments (Burton et al., 2020; Dietvorst et al., 2015), are not designed to speak to real-life policies.

[5] Existing evidence in the realm of taxation largely focuses on audit case selection. The findings in Bachas et al. (2025), also in partnership with the Senegalese tax administration, are complementary to ours. Highly skilled tax officials outperform a data-driven process for the selection of audit cases. A difference with our setting is that they were not able to observe "ground truth" (tax evasion) *ex ante* nor use it for the calibration of the algorithm. Therefore, the implemented formula is closer to a risk-score based on international best practices.

algorithm that the government actually implemented.[6]

Second, we contribute to the literature on the role of discretion versus rules in organizations (Aghion and Tirole, 1997; Dessein, 2002; Duflo et al., 2018; Bandiera et al., 2021; Kala, 2024; Decarolis et al., 2026), particularly on how bureaucrats shape policy outcomes (Best et al., 2023; Fenizia, 2022; Weigel et al., 2024; Khan et al., 2016, 2019). Our setting displays two key advantages allowing us to make a meaningful contribution. First, we document the influence of bureaucrats taken individually, and enter the "black box" of their discretion by combining administrative and lab-in-the-field survey data. Furthermore, we quantify the impact of varying degrees of discretion.[7]

Third, we contribute to the literature on fiscal capacity, tax equity, and the acceptability of taxation (Besley and Persson, 2009; Okunogbe, 2021; Nathan et al., 2023). While existing work has examined the relation between cadastral updates and tax revenue at a more *macro* level (D'Arcy et al., 2024; Martínez, 2023; Christensen and Garfias, 2021; vom Hau et al., 2023), or the role of technology for "last mile" aspects of taxation, such as GIS-enhanced bill delivery (Dzansi et al., 2025), we focus on a dimension which has received less attention: the granular details of *how* governments can best undertake upstream investments to build fiscal capacity. Our independent observation of the value of the tax base uniquely allows to determine the relative equity of different valuation methods.[8] By documenting for the first time taxpayers' acceptability of an algorithm in tax policy we advance the literature on perceived tax fairness (Stantcheva, 2021; Hvidberg et al., 2023; Hoy, 2025; Ajzenman et al., 2025; Best et al., 2025).

The paper is organized as follows. Section 2 presents the context and the experimental design, and Section 3 a simple conceptual framework. We present the data in Section 4, and the main results on accuracy and equity in Section 5. Section 6 investigates mechanisms, and Section 7 the political acceptability of algorithms. Section 8 lays out considerations for an optimal policy.

[6] Battaglini et al. (2025), Black et al. (2022) and Elzayn et al. (2024) also consider applications in taxation, but they study already realized audits, rather than the introduction of algorithmic selection in comparison to an alternative process. While we fully observe the algorithm used in the policy, they have to back out the algorithm from the realized audit cases, or simulate a non-implemented algorithm.

[7] While we cannot speak to the biases of our certified assessors (Chen and Cohen, 2024), each property having been covered by only one assessor for budgetary reasons, we are able to document potential deviations that arise during real-world implementation of algorithms (Björkegren et al., 2024) – the "rules" in our setting – by testing algorithms with and without bureaucrat inputs.

[8] Brockmeyer et al. (2021) in Mexico and Bergeron et al. (2024) in the Democratic Republic of the Congo (DRC) analyze optimal property tax liabilities. Our results are complementary since their focus is on rates and enforcement, while ours is on the determination of the tax base.

2 Context and Experimental Design

2.1 Background

In Senegal, all owners of a property are subject to an annual tax, and the national tax administration is responsible for generating the tax bills. The tax base is the annual market rental value, the value that is, or could be, obtained from the property rented at the market price.[9] The tax rate is 8.6 percent. There is an abatement for owner-occupied properties: for these, the share of the tax base below 1.5 million F is only taxed at 3.6 percent.[10] The administration conducts field operations to expand the tax net by registering and valuing properties. These operations do not involve any payment of the tax – the distribution of tax bills and recovery of payments is the responsibility of another administration, the national treasury. Before the program, the administration obtained property values by relying on bureaucrats' judgment and interactions with occupants in the field – what we label as discretionary valuations. The administration did not make use of any data-driven strategy to improve the property tax roll, whether it be by using data from satellite images, GIS mapping, or on properties' built area measurement.[11] We provide additional details on the institutional context in Appendix Section C.1.

Indeed, as in other cities of low- and middle-income countries, an important challenge is that there is no systematic information on real estate prices (Behr et al., 2023). Rental agreements are most often informal, channeled through informal brokers, and unreported to the administration.[12] Many properties are inherited within families without formal administrative procedures. When there are formal sales, they are rarely reported to the administration.[13] Furthermore, an additional challenge preventing the effective use of information on property prices is the lack of a systematic addressing system, making it difficult to match entries across sources. For instance, information

[9] Using the annual rental value as the base of the property tax is not unique to Senegal. This is also the case in other African countries (see Franzsen and McCluskey (2017)) and in OECD countries such as France, Italy and Belgium (Almy, 2014).

[10] Senegal's currency is the West Africa CFA Franc that we denote as F.

[11] This situation is not unique to Senegal and many administrations of low- and middle-income countries rely on discretionary or simplistic assessment procedures: a combination of bureaucrats' discretion and owner self-declarations in Monrovia, Liberia (Okunogbe, 2021); value bands based on bureaucrat-collected characteristics in Madina, Ghana (Dzansi et al., 2025); two flat fees in Kananga, DRC (Bergeron et al., 2024).

[12] 30 percent of properties in our baseline survey are rented at least in part. This share is 41 percent in the data from the census, with only 2.6 percent of rentals for which a contract was found to exist.

[13] While we were able to access the full record of property transfers held by the administration, it only includes around 2,000 transactions reported by a dozen of notary offices over the past ten years. Furthermore, the address information is too vague to know in which neighborhood each property is located.

from online sales or rental listings cannot be associated with a precise neighborhood.[14]

The tax performed very poorly in the years preceding the program. The region's property tax revenues were 6 billion F ($10.9 million) in 2022 which is around 7 percent of the total fiscal potential.[15] In our study areas, we estimate that the share of registered (respectively, tax compliant) property owners is 15 (resp., 9) percent. This underperformance of the property tax is a common feature in low- and middle-income countries (Franzsen and McCluskey, 2017; Knebelmann, 2022).[16]

2.2 Experimental Design

Our experiment was embedded in Senegal's first digital property tax census, rolled out by the national tax administration in the Dakar region between June 2021 and April 2023.[17] The census relied on a new property tax management software that we developed in collaboration with the administration. The software is innovative along two dimensions. First, it allows bureaucrats to conduct the census on tablets (rather than on paper), and incorporates pre-loaded property identifiers and GIS coordinates. Second, it enables the semi-automatized valuation of properties using an algorithm. We provide illustrations in Figure A1 and additional details on the program in Appendix Section C.2.[18]

We introduced experimental variation in the degree of reliance on bureaucrats' discretion versus algorithms in the determination of property values in the census. We randomized at the level of cadastral sections – small neighborhoods with 419 properties on average. In 47 sections, valuation followed a fully discretionary method: bureaucrats used their knowledge and judgment to assign values while in the field (Discretion arm). In 47 other sections (Algorithm arm), properties were valued using an algorithm: property values were predicted by combining characteristics collected by bureaucrats and data from satellite images (rule-based valuation). Third, we applied to

[14]Nonetheless, we do use a sample of values recovered from real estate agency listings, simply to verify how their distribution compares to our benchmark values at a more aggregate geographic level than the property, in Section 4.1. Note that volume and coverage of online listings with respect to the Dakar's real estate market is much smaller compared to cities of high-income countries.

[15]Revenue figures are from administrative data from the national treasury. We estimate total potential using the real estate values obtained from the licensed assessors, as described in Section 4, and administrative data on the total number of properties in the region.

[16]Recent micro-estimates of property tax compliance rates highlight a similar problem in other settings – 16 percent in Madina in the Accra region, Ghana (Dzansi et al., 2025), 10 percent in Kampala, Uganda (Manwaring and Regan, 2025), 8.8 percent in Kananga, DRC (Bergeron et al., 2024), 2.2 percent in Monrovia, Liberia (Okunogbe, 2021).

[17]Seven sections (less than three percent of the properties in our sample) were covered in 2019 before an interruption due to the Covid-19 pandemic.

[18]Similar digital applications for property taxation with GIS functionalities are still relatively new on the African continent, but are being increasingly adopted (Dzansi et al., 2025; Knebelmann, 2022).

this arm an alternative algorithm, a pure rule which predicted property values using only remote variables, without any bureaucrat input. More concretely, in the 47 Discretion sections, bureaucrats were responsible for the determination of property values. They tried to speak to the owner and/or the tenants and ask about monthly rental values. If bureaucrats were not able to ask occupants, or if the values provided seemed unrealistic, they used their own judgment to provide the best estimate of the monthly rental value. In the 47 Algorithm sections, bureaucrats started each visit by entering eighteen observable characteristics of the property, visible from the outside (listed in Table A2). These inputs were used to automatically generate predicted property values, which were not visible on the tablet. In both arms, bureaucrats tried to find the name and identification details of the owner by speaking to occupants.

The algorithms are inspired by Computer Assisted Mass Appraisal (CAMA) methods for property valuation. These systems are widespread in some high-income countries such as the USA and Canada, but are firmly established in Sub-Saharan Africa only in some South African cities. We worked hand in hand with practitioners having made suggestions on how to adapt these methods to the African context,[19] and with the administration to select the characteristics. Whether this would yield satisfactory results was initially not clear, since these statistical methods may lack precision in environments where data on property features and prices is scarce. For instance, while models in high-income countries typically include detailed variables on the inside of properties, such as the number of bathrooms, we could only retain property characteristics visible from the outside. Indeed, the administration indicated that accessing the inside of properties could not be a necessary condition for valuation since it was often challenging. We present the algorithms in detail in Section 4.2.

Taken together, the two treatment arms included 41,609 plots. There were an additional 97 pure control sections, in which the property tax census did not take place.[20] The following variables were used for stratification: the size of the section measured by the number of properties, the tax office the section depends on, and the share of properties eligible for the property tax according to the baseline survey.[21] Figure A2 shows the geographical scope of the study and the experimental

[19] Franzsen and McCluskey (2017); Davis et al. (2012); McCluskey et al. (2013); Guan et al. (2011); Moore (2005); Ali et al. (2018); Fish (2018); International Association of Assessing Officers (2022).

[20] Initially, 96 sections were assigned to be covered by the census. The 193 sections (96 for the census and 97 pure control) were selected by the administration based on their tax potential, excluding informal settlements and industrial areas. In total, they include 83,360 properties and correspond to around a third of the Dakar region. Census operations were interrupted in two sections, that we drop from the analysis. We provide more detail in Sections 4 and D.2.

[21] There are eight tax offices covering the whole region. A property is *not* eligible for the tax if it is a vacant plot of land, if it is a building totally under construction, if it belongs to the state, if it is a public school or religious institution.

design.

The bureaucrats were 268 temporary employees fully managed by the administration.[22] Bureaucrats' assignment to sections was planned by the tax office and was orthogonal to the randomization (72.5 percent of bureaucrats worked in both arms). No specific information was used for these assignments. The program being rolled out for the first time, there was no prior information on bureaucrat performance that the tax office could have used. The valuation method for a given section was programmed directly in the application, and neither the tax office managers nor the bureaucrats were aware of it before starting a given section. Once a section was started, it needed to be completed before the bureaucrats were deployed to the next. A given property was visited by only one bureaucrat whose identification number was recorded in the application. Supervisors divided properties of a given section among bureaucrats using the map on the tablets. Overall, we can credibly consider that bureaucrats were assigned to properties in a quasi-random way. In Section 4.3, we verify that bureaucrat characteristics are balanced across arms, and do not correlate with underlying property values.

3 Conceptual Framework

In this section, we lay out the respective advantages of relying on an algorithm versus bureaucrat discretion to create the tax roll. This requires clarifying what the objective of the administration is. The administration seeks to register the maximum number of properties with a tax base accurately reflecting market values, in order to tax to the full potential. At the same time, it seeks to ensure tax equity, more precisely vertical equity – more expensive properties should face higher tax bills – and horizontal equity – two properties of similar market value should face similar tax bills – as opposed to dispersion. Hence, while determining how equitable an algorithm is usually hinges on a framework that the researcher needs to define (Corbett-Davies et al., 2023), we instead rely on standard property tax equity concepts (Avenancio-León and Howard, 2022; Black et al., 2022) – allowing us to make minimal assumptions.

An algorithm provides a systematic and predictable valuation process. It reduces the risk for collusion between bureaucrats and owners – under the rule-based process, the core components of the prediction (built area and location) are not entered by bureaucrats, and under the pure rule,

[22] 32 percent had already worked with the administration in previous instances. They received a four day training, in addition to daily supervision by the tax offices. It is frequent for administrations in low and middle-income countries to resort to additional staff for large field operations, see for instance Bergeron et al. (2024); Guzmán Lizardo et al. (2025); Duflo et al. (2025).

there is no delegation to bureaucrats at all. However, the switch to an algorithm comes at the cost of prediction errors. By construction, the algorithm cannot take into account the idiosyncrasies of each property. This could result in very expensive properties being undervalued and the lowest-value properties being overvalued, worsening tax equity compared to a situation with bureaucrat judgment in the field. Anecdotally, insufficient consideration for the specificities of a given property is precisely what led the Spanish Supreme Court to issue a 2020 decision allowing the administration's formula-based determination of a property's value to be overridden using individual inspections.[23]

To evaluate the algorithms along these dimensions, an important advantage of our setting comes from our independent observation of property values for the whole analysis sample. We thus avoid the selective labeling problem which is prevalent in the literature (Kleinberg et al., 2018). Under this condition, the object the algorithm seeks to predict is only observable for a subset of cases that have been selected in the decision-process and are therefore not representative of the full sample.[24] Another advantage is that we have complete information on the algorithms' inner workings, as opposed to settings where it has to be backed out from the data (Elzayn et al., 2024).

A key feature of bureaucrats' discretion is that they have access to local information in the field. This can improve outcomes if this private information is relevant for valuations, for instance if they learn about rents paid by tenants, see details of the properties, or have local knowledge. They could also adjust values in a way that is welfare enhancing, even if this departs from the tax code, if they adjust based on taxpayers' income for instance. But conversely, another way bureaucrats might use local information is if they are biased towards or against certain types of occupants. If bureaucrats are inaccurate, this could stem from lack of expertise (*passive* channel following Bandiera et al. (2009)), or from rent-seeking if they collude with taxpayers and collect bribes in exchange for lower liabilities (*active* channel). Again, our independent observation of property values is key to enable us to test these different channels, combined with our ability to observe which bureaucrat valued which property in a setting where these assignments were quasi-random.

Finally, even if the algorithm proves superior, its adoption is feasible if bureaucrats and taxpayers accept it. Aversion to algorithms has been widely documented in laboratory or online

[23]Spanish Supreme Court (Contentious-Administrative Chamber). (2020). Judgment No. 843/2020, 23 January 2020, ECLI:ES:TS:2020:843 and https://laadministracionaldia.inap.es/noticia.asp?id=1178333

[24]A well-known example is the risk of recidivism in the study of algorithms used in the judiciary, only observed for defendants that have been released (Ludwig and Mullainathan, 2024; Angelova et al., 2025). Similarly, Bachas et al. (2025) only observe tax evasion for firms that have been audited.

settings. Bureaucrats' discretion could be more acceptable to taxpayers in a context where its use is widespread in the implementation of public policies.

4 Data

4.1 Data Sources

GIS mapping of properties. We compiled a geocoded dataset with property identifiers and ground built area measurement. The mapping was provided by the administration,[25] built area was recovered by GIS experts using high resolution satellite images.[26] There are 193 sections (83,360 properties), 96 of which were randomly assigned to be covered by the census, and 94 effectively covered.

Market property values. We hired licensed real estate assessors to provide market values for a representative sample of 4,916 properties spanning the 193 census *and* pure control sections. The assessors were members of the real estate section of the Senegalese National Order of Experts, a government-supervised professional body that licenses and regulates certified experts. Its members are officially recognized and highly qualified, comparable to chartered professional organizations in many countries. Assessors went in the field to see each property from the outside, provided their estimate of the market value and collected property characteristics. This procedure was very similar to the one a chartered surveyor in the United Kingdom, or a licensed real-estate appraiser in the USA, would follow.[27] In a *first best* situation, the Senegalese government would send these assessors to each property and use their valuations as the tax base. In other words, these values are what the government is after: they are the best available professional benchmark. However, resource constraints make it impossible for the government to resort to these highly paid professionals for all properties.[28] In our study, we use these assessor values as our "ground truth" and we refer to them as *market values*. Figure A3 provides illustrations of the range of property values.

We provide evidence supporting the reliability of the assessor values in Table A1. Assessors had

[25] We rely on the administration's national cadastral identifier which is at the level of the plot of land. In Dakar, there being multiple property owners on one plot of land is extremely rare (0.7 percent of census observations). Hence we refer to our unit of analysis as the "property" throughout the paper.

[26] We use Pléiades satellite images (50cm resolution) made available by the French Space Agency *Centre National d'Etudes Spatiales*.

[27] More detail on the Order and the data collection procedure can be found on their website (`https://www.experts-ones.com/`) and in Appendix Section D.1.

[28] The daily fee of the assessors is ten times the daily fee paid by the government to field bureaucrats.

previous experience working in 97 percent of the sections, whether of high or low value. As explained in Section 2, there is virtually no available pre-existing data that we could use to check correlations at the property level. Nonetheless, we check how the values compare with other data collected throughout the project, and with external data from rental listings at a more aggregate geographical level. It is reassuring that assessor values display a 0.74 correlation with values for fully rented properties picked up in the census (Panel (B) of Table A1).[29] Among a small subset of online rental listings that we were able to manually assign to a section, we find that 96 percent of values fall within the bounds of the assessors' values for the same section. At the municipality level, 99 percent of observations of values per square meter from real estate agency listings fall within the bounds of assessor values ($N = 2,020$, Panel (C) of Table A1).[30]

Property tax census data from the digital tool. For each property covered in the census, we have the bureaucrat's identification number and all variables collected by the bureaucrat: owner and tenant information, usage and rental status of the property, number of floors and rooms, property value, the observable characteristics used for the algorithm if the section is in the Algorithm arm. 94 sections out of the 96 targeted ones were covered, including 38,227 properties.[31] Henceforth, we refer to these property values as the *census values* or the *tax roll values*.[32]

Analysis sample. For our analysis, we restrict the sample to properties in sections covered by the census, for which we have a market value and built area measurement.[33] We drop properties classified as non-eligible by bureaucrats. In Table A5, we show that implementation did not differ across arms.[34] The analysis sample includes 2,289 properties – 1,166 in the Algorithm arm and 1,123 in the Discretion arm.

Bureaucrat surveys. Bureaucrats completed a self-administered baseline survey during their train-

[29]We also find a 0.69 correlation with rental values collected in our survey.

[30]Table A1 also reports the Kolmogorov-Smirnov distance between the two distributions. We collected the online listings for which we manually determine the section from Senegalese real estate websites. The address information is too imprecise to link these to specific properties. The larger sample of listings observed at the level of the municipality are rental ads compiled by a Senegalese real estate intelligence firm. We provide additional details in Table A1 and in Appendix D.3.

[31]Census operations were interrupted in two sections (one in each arm) after only a handful of properties had been covered. We drop these sections from the analysis. In all other sections, 91.9 percent of properties were successfully registered. We provide additional details in Appendix D.2.

[32]We observe whether a payment was subsequently made for each property. We use this information to test additional mechanisms in Section 6.3, and for cost-effectiveness considerations in Section 8. We study compliance more thoroughly in follow-up work (Bonmartel et al., 2026).

[33]In Panel (A) of Table 1 we show that the availability of built area is balanced across arms. Missing built area is most often due to cases of vacant land.

[34]There is no significant difference between Discretion and Algorithm sections in the number of properties covered per day (column (1) of Table A5), the probability for a property to be covered (column (2)), the probability for a property to be classified eligible (column (3)), the probability that the bureaucrat met the owner (column (7)). This serves as evidence that bureaucrat effort did not differ across arms.

ing, including information on their background and views about taxation. After the census, they completed a longer endline survey which included modules on job satisfaction and the measurement of cognitive and non-cognitive skills. We also included an experimental lab-in-the-field property valuation task that we use to test mechanisms. In Appendix Section F.1, we define the variables of the bureaucrat surveys. 287 (respectively 180) out of 293 bureaucrats completed the baseline (resp., endline) survey.[35]

Property owner surveys. 2,474 property owners (taxpayers) were surveyed at baseline in 2018, 1,238 of which are in sections covered by the property tax census. 4,342 were surveyed at endline in 2025, with 1,987 in census sections. The surveys included information on the socio-economic conditions of the household, the property's usage and characteristics, property value and rents. At endline, we included an experimental module eliciting taxpayers' preferences regarding valuation methods. We also surveyed 1,665 tenants. See Appendix F.2 for a detailed description of the variables from the surveys.

4.2 The Property Valuation Algorithm

We calibrated the property valuation algorithm on the sample of 4,916 assessor valuations using an elastic-net regression and five-fold cross-validation. We followed best practices from the property valuation literature and international practitioners.[36] The functional form is:

$$Ln(MarketValue_{ij}) = \alpha + \beta Ln(BuiltArea_{ij}) + \gamma floors_{ij} + \sum_k \theta_k X_{k,ij} + Sec_j + \epsilon_{ij} \qquad (1)$$

where $MarketValue_{ij}$ is the annual rental value of property i in section j, $BuiltArea$ is ground built area multiplied by the number of floors, $floors$ is the number of floors, and the X_k variables are the property characteristics collected by assessors.[37] Sec_j is a section fixed effect. There are two versions of the algorithm. The rule-based algorithm includes all covariates, while for the pure rule, we re-calibrate after dropping the X_k variables.[38] Table A2 displays the resulting coefficients.

[35] 98.8 percent (respectively, 81.4) of the properties covered by the census were covered by a bureaucrat for which we have a baseline (resp., endline) survey. Some bureaucrats were difficult to track down because they did not resume working on the program after the Covid-19 pandemic.

[36] See Appendix E for additional details and references.

[37] The X_k covariates are: usage, type of fence, state of the fence, type of cladding, state of the cladding, cement wall, presence of decorative tiles, quality of doors and windows, landscape improvement, architectural improvement, presence and type of garage, balcony, location with respect to main road, type of road, presence of sidewalk, whether the property is at a street corner (angle), presence of street lights. All variables and response categories are listed in Table A2.

[38] In the current version of the pure rule, we still rely on the number of floors reported by assessors to compute total

Table A3 summarizes the performance of each algorithm. We estimate performance on test samples. The adjusted R^2 estimated through cross-validation is 0.90 for the rule-based algorithm, and 0.86 with the pure rule (rows (1) and (2) in Table A3). For the rule-based version, the out-of-sample mean absolute percentage error (MAPE) is 30.6, and 62.5 percent of predicted values fall within 30 percent of the market value. These figures are 34.8 and 60.9 percent respectively for the pure rule. The limited decrease in algorithm performance when switching to the pure rule reveals that section fixed effects and built area explain a very large share of the variation in property values.

Our performance indicators lie at the top end of those found in similar settings. Franklin (2019) finds an R^2 of 0.85 in Addis Ababa, Ethiopia, Ali et al. (2018) an R^2 of 0.56 in Rwanda. Behr et al. (2023) find a MAPE ranging between 30 and 64 percent in South Africa. Bergeron et al. (2024) find a MAPE of 69.9 percent in the DRC, and that 53.5 percent of predictions are within 50 percent of the true value.

We calibrated alternative algorithm versions that we use to assess the robustness of our results in Section 5.4. In Appendix G, we provide details on each different version, and the performance statistics are summarized in Table A3.

4.3 Balance Checks

In Table 1, we verify that section and bureaucrat characteristics are balanced across the Algorithm and Discretion arms. Section characteristics are balanced by construction due to our randomization, while the balance of bureaucrat characteristics results from their quasi-random assignment to areas. We report the coefficient on an indicator for discretionary sections in a property-level regression in Panel (A), and a regression at the bureaucrat X section level in Panel (B). In Panel (C), we regress bureaucrat characteristics on market values. In each Panel, we can rule out the joint significance of all characteristics. None of the p-values for the coefficient of interest are below 0.05, except for one out of twelve characteristics in Panel (C) but with a very small magnitude.[39] These verifications confirm that we can draw causal interpretations of the effects on the tax roll of: (i) reliance on bureaucrat discretion versus algorithms; (ii) bureaucrats' assignment to properties.

built area. However, it is technically feasible to recover building height from more sophisticated satellite images than the ones we worked with. The Senegalese administration has acquired such images, but has not processed them to date. We include a robustness check using an algorithm calibration without building height and our main results hold, as explained in Appendix Section G.1.

[39] A property with a one percent higher market value is 0.04 percentage points (0.09 percent) less likely to be visited by a bureaucrat with three years or more of higher education.

TABLE 1
BALANCE AND SUMMARY STATISTICS

Panel A: Section characteristics across treatment arms	Mean (SD)	$\hat{\beta}_{Discretion}$	P-value	N
Source: Cadastral data				
Number of plots	419 (177.61)	71.5037 (37.40)	0.06	94
Built area (0/1)	0.93 (0.25)	0.0132 (0.02)	0.51	41,609
Built area (m2)	282 (855.39)	−10.3793 (27.44)	0.71	41,609
Source: Assessors				
Eligible (0/1)	0.87 (0.33)	0.0153 (0.02)	0.39	2,844
Value (million F)	10.71 (21.93)	1.9127 (2.39)	0.43	2,469
Value per m2 (million F)	0.02 (0.02)	0.0026 (0.00)	0.22	2,409
Source: Baseline				
Non-response (0/1)	0.53 (0.50)	−0.0298 (0.03)	0.32	2,537
Value (million F)	4.10 (4.93)	−0.1074 (0.41)	0.79	879
Rented (0/1)	0.30 (0.46)	0.0285 (0.03)	0.30	1,238
Owner-occupied (0/1)	0.60 (0.49)	0.0148 (0.03)	0.64	1,238
High household income (0/1)	0.22 (0.41)	−0.0020 (0.03)	0.94	1,238
In tax net (0/1)	0.18 (0.38)	0.0025 (0.03)	0.93	1,238
Joint significance *Cadastral* and *Assessor* variables			0.36	
Joint significance *Baseline* variables			0.65	

Panel B: Bureaucrat characteristics across treatment arms	Mean (SD)	$\hat{\beta}_{Discretion}$	P-value	N
Age	31.51 (5.80)	0.1483 (0.34)	0.66	1,288
Female	0.28 (0.45)	−0.0129 (0.03)	0.61	1,288
Ever worked with tax adm.	0.18 (0.38)	0.0266 (0.02)	0.20	1,288
From Dakar	0.48 (0.50)	−0.0125 (0.03)	0.66	1,288
Municipality of residence	0.06 (0.24)	0.0010 (0.01)	0.94	1,266
Any higher education	0.98 (0.14)	−0.0139 (0.01)	0.12	1,288
Long higher education	0.40 (0.49)	−0.0124 (0.03)	0.66	1,288
Ethnic group: Wolof (majority)	0.31 (0.46)	0.0122 (0.03)	0.65	1,288
Religion: Tidjane (majority)	0.56 (0.50)	0.0303 (0.03)	0.28	1,288
Public service motivation (index)	0.06 (0.85)	−0.0170 (0.05)	0.73	1,288
In favor of government's role (index)	0.13 (0.97)	−0.0700 (0.06)	0.21	1,268
In favor of widespread taxation (index)	0.01 (1.00)	−0.0402 (0.06)	0.48	1,288
Joint significance			0.71	

Panel C: Bureaucrat characteristics and property values	Mean (SD)	$\hat{\beta}_{Ln(Value)}$	P-value	N
Age	31.53 (5.64)	0.0327 (0.18)	0.86	2,236
Female	0.25 (0.43)	0.0066 (0.01)	0.49	2,236
Ever worked with tax adm.	0.18 (0.39)	−0.0009 (0.01)	0.94	2,236
From Dakar	0.48 (0.50)	0.0019 (0.02)	0.90	2,236
Municipality of residence	0.09 (0.29)	−0.0042 (0.007)	0.56	2,187
Any higher education	0.99 (0.10)	0.0007 (0.002)	0.79	2,236
Long higher education	0.41 (0.49)	−0.0368 (0.01)	0.01	2,236
Ethnic group: Wolof (majority)	0.32 (0.47)	0.0012 (0.01)	0.92	2,236
Religion: Tidjane (majority)	0.57 (0.50)	−0.0193 (0.02)	0.21	2,236
Public service motivation (index)	0.08 (0.80)	0.0033 (0.02)	0.86	2,236
In favor of government's role (index)	0.08 (0.93)	0.0166 (0.03)	0.55	2,214
In favor of widespread taxation (index)	−0.02 (0.99)	0.0068 (0.03)	0.81	2,236
Joint significance			0.41	

Notes: This Table verifies that section and bureaucrat characteristics are balanced across treatment arms, and that bureaucrat characteristics do not correlate with property values. In Panel (A), we regress section characteristics on an indicator for Discretion sections. Observations are at the property level, except for the outcome *Number of plots* where observations are at the section level. In Panel (B), we regress bureaucrat characteristics on an indicator for Discretion sections. Observations are at the bureaucrat X section level. In Panel (C), we regress bureaucrat characteristics on the logarithm of market values from licensed assessors. Observations are at the property level. In all regressions, we control for strata fixed effects and we cluster errors at the section level. The second column reports the coefficient of interest with its standard error in parentheses. At the bottom of each Panel, we report the p-value for an F-test of the joint significance of all variables in a regression where the Discretion indicator is the outcome. See Appendix Section F for a detailed description of survey variables.

5 Results

We compare discretionary and algorithmic valuations along several dimensions. To assess accuracy, we use the tax base gap – the difference between tax roll value and market value – and the assessment ratio – a property's tax roll value divided by its market value.[40] We assess horizontal equity by studying how dispersed tax roll values are for properties of the same market value, and by measuring the extent to which individual bureaucrats shape tax liabilities. Finally, we assess vertical equity by considering the resulting profile of effective tax rates and by studying how the assessment ratio varies through the distribution of property values.

5.1 Algorithms outperform Discretion in Accuracy

We compare the accuracy of discretion and the algorithms using the following regression:

$$Gap_{ijk} = \alpha + \beta D_{jk} + S_k + \epsilon_{ijk} \tag{2}$$

where Gap_{ijk} is the tax base gap for property i of section j and strata k, computed as $Gap_{ijk} = TaxRollValue_{ijk} - MarketValue_{ijk}$. We estimate the regression twice. In the Discretion arm, the tax roll value is bureaucrats' discretionary value. In the Algorithm arm, the tax roll value is the rule-based value in the first estimation, and the pure rule value in the second estimation. Amounts are in millions of F, and Gap is winsorized at the one percent level. The sign and value of the tax base gap indicate whether there is over- or undervaluation and by how much, in monetary amounts. D_{jk} is an indicator for Discretion sections, and S_k is a strata fixed effect. Standard errors are clustered at the section level. Results are shown in column (1) of Table 2. Column (2) displays results from a quantile regression where the outcome is the median gap. In column (3), we use the absolute value of the tax base gap - this specification prevents positive and negative values from compensating each other, capturing inaccuracies in any direction.[41]

We find that bureaucrats' discretion leads to strong undervaluation and inaccuracies, that the algorithms mitigate. Panel (A) compares discretion to the rule-based method. The estimated $\hat{\beta}$ shows that the gap (respectively, median gap) is -4.19 million F (resp., -1.67 million F) wider under discretion. The absolute tax base gap is 3.84 million F, or 82 percent, wider under discretion.

[40] The assessment ratio is a common metric in the property taxation literature (Avenancio-León and Howard, 2022; McMillen and Singh, 2020; Dray et al., 2025), adapted to our context. We provide details in Appendix D.4.

[41] The tax base gap is similar to the numerator of the tax gap used in Khan et al. (2016), a seminal study on bureaucrats' influence on property tax liabilities.

The coefficient is significant at the one percent level in all specifications. Panel (B) shows that the pure rule outperforms discretion to an even greater extent. The median tax base gap is -2.08 million F wider, and the absolute tax base gap 4.75 million F or 130 percent wider, under discretion. Additionally, in Table A4, we report median assessment ratios for each method. They are in the vicinity of one for the rule-based and pure rule valuations and significantly below for discretion.

TABLE 2
EFFECTS ON ACCURACY: TAX BASE GAP

	Gap	Gap (median)	|Gap|
	All amounts in millions of F		
	(1)	(2)	(3)
Panel A: Rule-based versus Discretion			
$\hat{\beta}_{Discretion}$	-4.19^{***}	-1.67^{***}	3.84^{***}
	(1.27)	(0.36)	(1.38)
Mean (sd) in Rule-based	−2.77 (12.72)	−0.34	4.67 (12.15)
Panel B: Pure Rule versus Discretion			
$\hat{\beta}_{Discretion}$	-4.93^{***}	-2.08^{***}	4.75^{***}
	(1.14)	(0.47)	(1.22)
Mean (sd) in Pure rule	−1.89 (10.41)	−0.20	3.65 (9.93)
N# properties		2289	
N# sections		94	
Mean (sd) market value		15.80 (77.00)	
Median market value		5.60	

Notes: This Table compares the accuracy of discretionary and algorithmic valuations, using regression 2: $Gap_{ijk} = \alpha + \beta D_{jk} + S_k + \epsilon_{ijk}$ where Gap_{ijk} is the tax base gap for property i of section j and strata k, D is an indicator for Discretion sections and S_k is a strata fixed effect. The tax base gap is the tax roll value minus market value, where tax roll value is the value from the census and market value is the value from licensed assessors. In Panel (A), values for the Algorithm arm are the rule-based valuations incorporating bureaucrats' inputs. In Panel (B), values for the Algorithm arm are the pure rule valuations. Column (2) relies on a quantile regression, the outcome is the median gap, and *Mean in Rule-based* reports the median. In column (3), the outcome is the absolute value of the tax base gap. All amounts are in millions of F and winsorized at the 1% level. *,** and *** indicate statistical significance at the 10, 5 and 1% level respectively. Errors are clustered at the section level. Sample: properties in census sections for which we have market values ($N = 2,289$).

These results suggest that in field implementation, the pure rule slightly outperforms the rule-based method, even though the opposite holds in the calibration (see Section 4.2). We quantify these effects by estimating a similar regression on the Algorithm arm only. We first directly compare the two algorithms, and find that the rule-based method increases the absolute tax base gap by 1.03 million F compared to the pure rule (Table A7, Panel (B)). Second, we compare the rule implemented by bureaucrats to the rule with assessors' inputs that were used in the calibration. The absolute tax base gap is 1.94 million F wider when the rule is delegated to bureaucrats (Panel (A) of Table A7). This suggests that even a limited degree of discretion – in this case, delegating the entry of property characteristics to bureaucrats – is detrimental for policy outcomes. We investigate which property characteristics contribute the most to these differences in Table A8.[42]

Taken together, these results show that the algorithms improve accuracy, and the pure rule with no delegation to bureaucrats performs best. Bureaucrats substantially undervalue properties. The implications for tax liabilities are large: total liabilities amount to 8 billion F in the Discretion arm compared to 10.7 billion F with the rule-based process, and 13.8 billion F with the pure rule.

5.2 Algorithms outperform Discretion in Horizontal Tax Equity

In Figure 1, we show the scatter plot of properties' tax roll values over market values. Under discretion (Panel (A)), a striking observation is the strong dispersion of valuations, with properties of similar market value facing very different tax roll values. This is mitigated by the rule-based process (Panel (B)) and the pure rule (Panel (C)). We measure these horizontal inequities using the coefficient of dispersion of the resulting tax rates, capturing the average deviation of the tax rate from its median.[43] Average dispersion is 143 percent of the median tax rate under discretion, 60 percent under the rule-based process, and 46 percent with the pure rule. Hence, removing discretion increases horizontal equity.

Discretion gives full leeway to individual bureaucrats, and bureaucrat-induced variability in the tax roll may be one driver of these horizontal inequities. We estimate bureaucrat fixed effects in the following specification:

$$|Gap|_{ijb} = \alpha_b + Val_j + \epsilon_{ijb} \tag{3}$$

where $|Gap|_{ijb}$ is the absolute value of the tax base gap (in millions of F and winsorized at the one percent level) for property i of section j covered by bureaucrat b, α_b is the bureaucrat fixed effect, and Val_j is a section-level control for market values.[44] Errors are clustered at the section level. We estimate equation 3 separately for the Discretion arm (1,054 properties and 197 bureaucrats) and

[42]In Panel (B) of Table A8, we report the contribution of each characteristic X_k to the aggregate difference between the algorithm implemented by bureaucrats versus the algorithm with calibration inputs. We find that "type of cladding" is playing the largest role – this is likely to be due to the fact that it has six modalities, and that it requires technical expertise to differentiate them from each other.

[43]We compute effective tax rates as a property's tax bill amount divided by its market value. The coefficient of dispersion is widely used in the property assessment literature, see Appendix D.4 for details.

[44]Val_j is a categorical variable indicating in which decile of market value per square meter section j is. In an alternative specification, we replace this control by an indicator for high-value properties and find similar results – see Appendix G.2.

the Algorithm arm using rule-based values (1,063 properties and 190 bureaucrats).[45] Thanks to the quasi-random assignment of bureaucrats to properties, the estimated α_b is an unbiased measure of the quality of a bureaucrat's valuations: a larger α_b means bureaucrat b drives values further away from market values. We shrink each estimated fixed effect proportionally to the noise with which it is estimated using an empirical Bayes procedure, this yields the vector of adjusted fixed effects $\alpha_{b,EB}$.[46] Summary statistics are reported in Table A9. The share of variance in the tax base gap explained by bureaucrat fixed effects is 34 percent under discretion, against 9 percent under the rule.[47] This means that under discretion, a taxpayer's tax bill strongly depends on which bureaucrat visited the property, contributing to these horizontal inequities.[48]

5.3 Algorithms outperform Discretion in Vertical Tax Equity

We compare the vertical equity achieved with each method. We find that discretion generates a strongly regressive tax profile, and that this is mitigated with the algorithms. Table A4, reporting assessment ratios by quintile of market value, reveals a strong undervaluation gradient under discretion. While the median assessment ratio is 0.83 in the first quintile, suggesting that bureaucrats value low-value properties relatively accurately, it decreases down to 0.23 in the top quintile (Panel (A) of Table A4).[49] Under the algorithms, the first quintile of properties are slightly overvalued (assessment ratios of 1.23 and 1.24 with the rule and pure rule respectively). The assessment ratios are close to one in the other quintiles.

We show the tax profiles obtained with each method in Figure 2, displaying the median effective tax rate by quintile. The gray dashed line shows the legal tax profile calculated by applying the statutory tax rate to the benchmark market values.[50] For the lowest quintile, the expected tax rate is 4.4 percent. The observed tax rates are 3.8 percent in the Discretion arm (black line), 6.7 percent

[45] Properties that are not covered by the census are dropped since no bureaucrat is associated to them. As shown in column (2) of Table A5, the probability of being covered does not differ across arms.

[46] Our estimates may suffer from noise due to the limited number of observations for each bureaucrat. For the shrinkage procedure, we follow the methodology developed in Chandra et al. (2016), in line with Kane and Staiger (2008); Morris (1983). The kernel density estimates of the distribution of the fixed effects are pictured in Figure A9.

[47] We can compute the share of variance as $Var(\alpha_{b,EB})/Var(|Gap|)$ due the orthogonality of bureaucrats' assignment to property features.

[48] Our results are in a similar range as those in Weigel et al. (2024) reporting that bureaucrats explain 21 to 34 percent of variance in tax compliance in the DRC, and Best et al. (2023) who find that bureaucrats explain 39 percent of variation in quality-adjusted prices in the Russian public procurement system.

[49] Table A4 also reports the results from a regression of the assessment ratio on indicators for each quintile, separately for each method, to confirm that the differences across quintiles are significant. See Table notes for more details.

[50] The increase in the legal tax rate over quintiles is due to the abatement applied to main residence properties, as explained in Section 2.1. We use the main residence variable from the census. Panel (B) of Table A5 confirms that there is no difference across arms in the probability to be classified as main residence in our analysis sample.

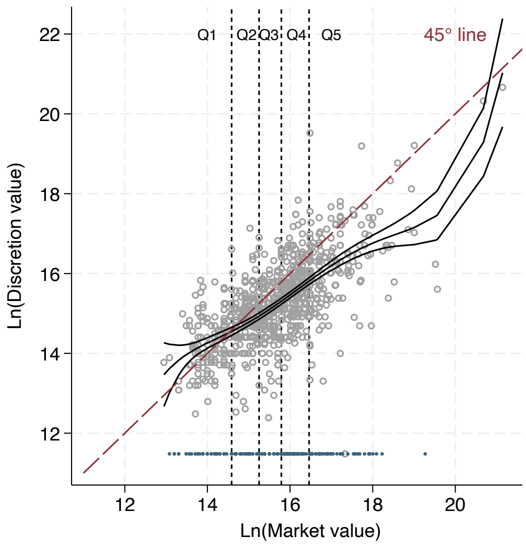
with the rule-based method (in blue), and 5.4 percent with the pure rule (in red). For properties of the top quintile, the tax rate is 1.7 percent instead of the expected 8.6 percent under discretion. The top quintile rate is 5.2 percent under the rule and 6.6 percent under the pure rule. The slope indicated in the legend is estimated as the regression coefficient of $TaxRate$ on $Ln(MarketValue)$. It is negative for discretion (-0.011) and this is reversed with the pure rule (0.001).[51] There are large implications for the way the tax burden is shared: the share of tax liabilities due by the top 10 percent of properties is 49.9 under discretion, 63.3 with the rule-based process, and 66.9 percent under the pure rule. Overall, the pure rule performs best in terms of vertical equity.[52]

While the algorithms offer substantial improvements in comparison to discretion, the tax profiles in Figure 2 and the assessment ratios in Table A4 also show that they tend to overvalue low-value properties and undervalue the most expensive ones. This pattern is not unique to our context: some degree of regressivity has been shown to be inherent to property valuation models because of unobserved variable bias (McMillen and Singh, 2020; Berry, 2021; Amornsiripanitch, 2024). This might come from property features assessors see in the field but that are not captured in the characteristics. Although bureaucrats' valuations are relatively less inaccurate for low-value properties, we still find that the algorithms significantly outperform discretion for this segment, as shown in Panel (A) of Table A13. But this results in a trade-off: retaining discretion could be preferred to minimize the risk of overvaluing low-value properties, even if it comes at the cost of lower accuracy and horizontal equity. We return to this question in Section 8.

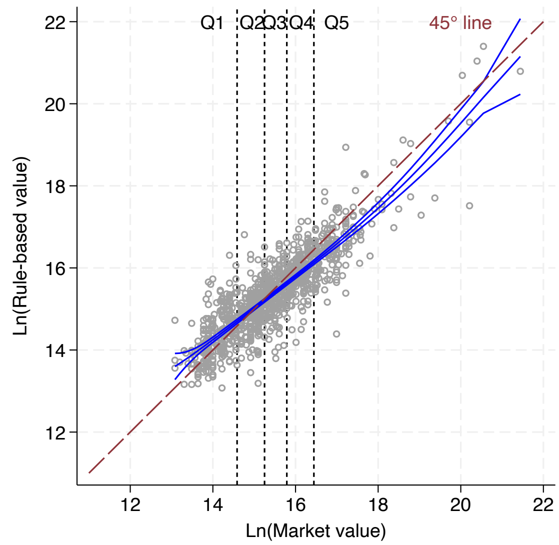
[51] In Figure A7, we verify that the same patterns hold when considering properties' rank instead of value. Methodologically, we follow the social mobility literature (Chetty et al., 2019, 2020) and compute rank-preservation slopes, where a slope closer to one is suggestive of stronger vertical tax equity. The rank-preservation slope is 0.28 under discretion, 0.69 under the rule, and 0.94 with the pure rule.

[52] The lower performance of the rule implemented by bureaucrats compared to the pure rule is due to the erroneous characteristics entered by bureaucrats, similarly to what we describe in Section 5.1. To visualize this, in green in Panel (A) of Figure A4, we show the tax profile obtained with the algorithm if implemented perfectly, i.e. using the inputs from the assessors. The profile appears much closer to the legal tax profile and the slope is 0.004.

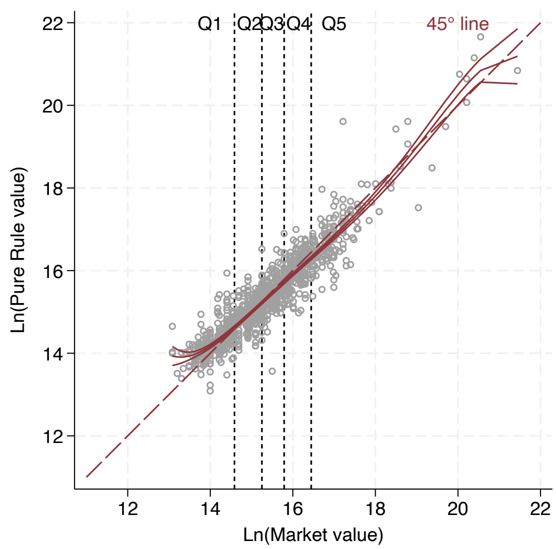
FIGURE 1
HORIZONTAL EQUITY: VALUATIONS AND DISPERSION



(A) DISCRETION



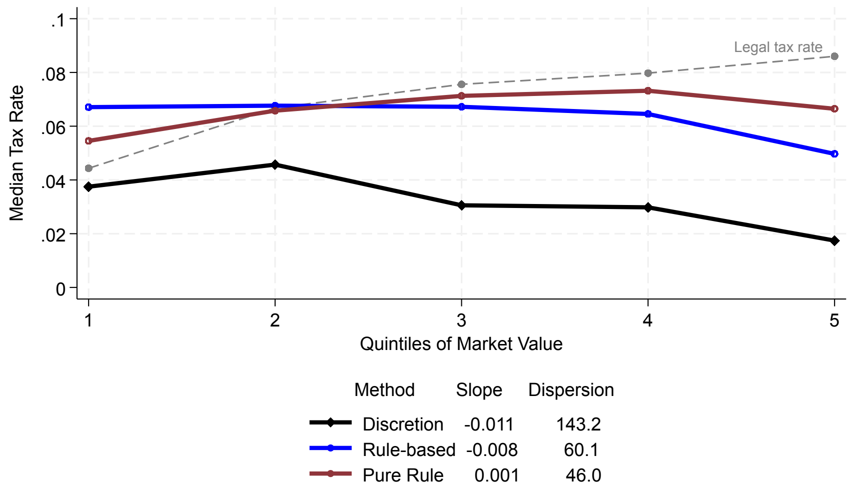
(B) RULE-BASED



(C) PURE RULE

Notes: This Figure shows property valuations under each method. In each Panel, the y-axis is the logarithm of the tax roll value generated by each method, and the x-axis is the logarithm of market values from licensed assessors. The vertical dashed lines *Q1-Q5* indicate the quintiles of market value. We display the sixth-degree polynomial fit between the two values with its 95% confidence interval, in black (Panel (A)), blue (Panel (B)) and red (Panel (C)). In Panel (A), the blue dots are observations with a tax roll value equal to zero. Sample: properties in census sections for which we have market values (N = 1,123 in the Discretion arm and N = 1,166 in the Algorithm arm).

FIGURE 2
VERTICAL EQUITY: TAX PROFILES



Notes: This Figure shows the tax profile for each method by displaying the median effective tax rate by quintile of market value. A property's effective tax rate is computed as the tax liability based on the tax roll value divided by its market value. The gray dashed line is the benchmark legal tax profile obtained by applying the statutory tax rate to market values. In the legend, we report the slope, capturing vertical equity, and the dispersion, capturing horizontal inequity. The slope is estimated as the regression coefficient of $TaxRate$ on $Ln(MarketValue)$, while the dispersion is the average percentage dispersion of the effective tax rate around its median (see Appendix D.4 for details). Sample: properties in census sections for which we have market values (N = 1,123 in the Discretion arm and N = 1,166 in the Algorithm arm).

5.4 Robustness Checks

We conduct several robustness checks to support the finding that the algorithms outperform discretion in terms of accuracy and tax equity. In Appendix Section G.1, we provide details on the specifications used for each robustness checks. A first concern could come from the fact that assessor values are used for the calibration of the algorithms and as our benchmark. Our procedure relying on cross-validation already prevents the results from being driven by overfitting. As an additional verification, we reproduce the results with a full separation of observations used for the calibration and for the analysis. Results are virtually unchanged when restricting the analysis sample to half of the Algorithm arm of which no observation is used in any step of the algorithm calibration (Panels (A) and (B) of Table A6 and Panel (B) of Figure A4).[53] Furthermore, we provide results where property values from a different data generating process are used to calibrate the algorithms, namely values collected through our surveys. In spite of the lower performance of

[53]We also find similar results when calibrating the algorithms on pure control areas, where the census did not take place at all (Panels (C) and (D) of Table A6 and Panel (C) of Figure A4). However, these algorithms cannot include section fixed effects or else they cannot be applied to our analysis sample. Instead, we use strata fixed effects, making these algorithms much less precise and policy-relevant.

these algorithms (R^2 ranging between 0.33 and 0.58, see Table A3), due to lower data quality,[54] we still find that these algorithms are more accurate than discretion (Panels (E) to (H) of Table A6).[55]

Second, we loosen our definition of what the benchmark value is for each property. Instead of a single market value, we use the lower and upper bounds provided by the assessors for each property. We set $Gap_{ijk} = 0$ for any observation where the tax roll value is within these bounds. We still find that discretion substantially and significantly widens the tax base gap compared to the algorithms, as shown in Panels (I) and (J) of Table A7 and Panel (F) of Figure A4.

A third concern could be that although bureaucrats operating in Discretion and Algorithm areas have similar observable characteristics, they differ along unobservable margins, contributing to the aggregate differences we find. We alleviate this concern by estimating our main regression including bureaucrat fixed effects, and find that for a given bureaucrat, discretion widens the absolute tax base gap by 4.5 million F compared to the rule-based method (Panel (C) of Table A7).[56]

6 Mechanisms

We investigate whether bureaucrats' inaccuracies stem from a lack of knowledge (*passive* channel) or collusion with taxpayers (*active* channel). Second, we test whether local information influences their valuations in ways that reveal that their objective function differs from that of the government, in ways that remain policy-relevant, or through biases. Finally, we test whether bureaucrats' valuations could be improved or incorporated into an algorithm.

6.1 Bureaucrats' Lack of Knowledge

Lab-in-the-field valuation exercise: We use an experimental module of the bureaucrat endline survey to isolate bureaucrats' knowledge about property values from whatever may happen during

[54]While one could hope that the tenant survey provides higher quality data, one limitation is that most tenants only rent a subset of rooms, which means that we need to extrapolate rents to the whole property using some assumptions (see Appendix F.2.3).

[55]Only the coefficients for the median tax gap are significant, but all coefficients retain similar sign and magnitude as our main results. The tax profiles plotted in Panel (D) of Figure A4 suggest that the algorithms calibrated on the surveys are more regressive than our main algorithms. This is first because the lower precision of the prediction model generates more regressivity. Second, in the raw survey data, we observe that respondents of the first quintile overvalue their properties. When considering quintiles two to four, the slopes are between -0.016 and -0.012, close to the slope found with discretion. Overall, the algorithms calibrated on the surveys display more accuracy and more horizontal equity than discretion, and similar vertical equity.

[56]Under discretion, bureaucrats could leave the value empty when they lacked information. We verify that the results hold on the intensive margin and using a Lee bounds exercise, as shown in Panels (E) to (H) of Table A7 and Panel (E) of Figure A4 .

field visits. Bureaucrats were shown the picture of a property with an indication of its neighborhood, and asked to provide their best estimate of the monthly rental value. Each bureaucrat completed the exercise twice: for a property from the lowest quintile and a property from the top quintile.[57] We show the distribution of bureaucrats' responses and the true market value for each property in Figure 3. While some bureaucrats overvalue the low-value property (top panel), there is a mass of responses in the vicinity of the true market value, and bureaucrats' mean value is still in the lowest quintile of market values. On the other hand, the distribution of responses for the high-value property (bottom panel) is completely to the left of its true value. The average value given by bureaucrats places the property in the third instead of fifth quintile. The average (resp. median) assessment ratio is 0.31 (resp. 0.26). This is very close to the assessment ratio observed in the field for the highest quintile (0.23, see Table A4). We conclude that the main driver of the undervaluation gradient found under discretion is bureaucrats' lack of knowledge about the values of high-end properties.[58] While previous research has investigated the relationship between individuals' misperceptions about the income or wealth distributions and their support towards various tax policies (Stantcheva, 2021; Hvidberg et al., 2023; Hoy and Mager, 2021; Hoy, 2025), we shed light here on how these misperceptions may strongly affect outcomes when they are held by the *individuals implementing* a given policy.

Bureaucrats' implicit algorithm: We estimate bureaucrats' "implicit algorithm", by using the same methodology as for the calibration of our main algorithm, but with bureaucrat discretionary values as the outcome variable to be predicted.[59] Results are shown in Table A12. The R^2 is 0.49 (column (2)) against 0.85 for the algorithm with assessor values (column (1)).[60] The coefficient on built area is 0.36 against 0.52, and the standard deviation of the estimated coefficient for section fixed effects is 0.35 against 0.45, revealing that bureaucrats put too low weight on these two major determinants of property values. These results show that bureaucrats' values are poorly explained by objective

[57] See details on the survey module in Appendix F.1. While the questions were not incentivized, they were presented as a means for the research team to better understand how the program was being implemented. There are no reasons to believe that bureaucrats would strategically provide wrong answers: they were not aware that an analysis of their values had been conducted, nor that benchmark market values had been collected.

[58] In Table A11, we use a regression with bureaucrat fixed effects to confirm that expensive properties are less accurately valued for a given bureaucrat.

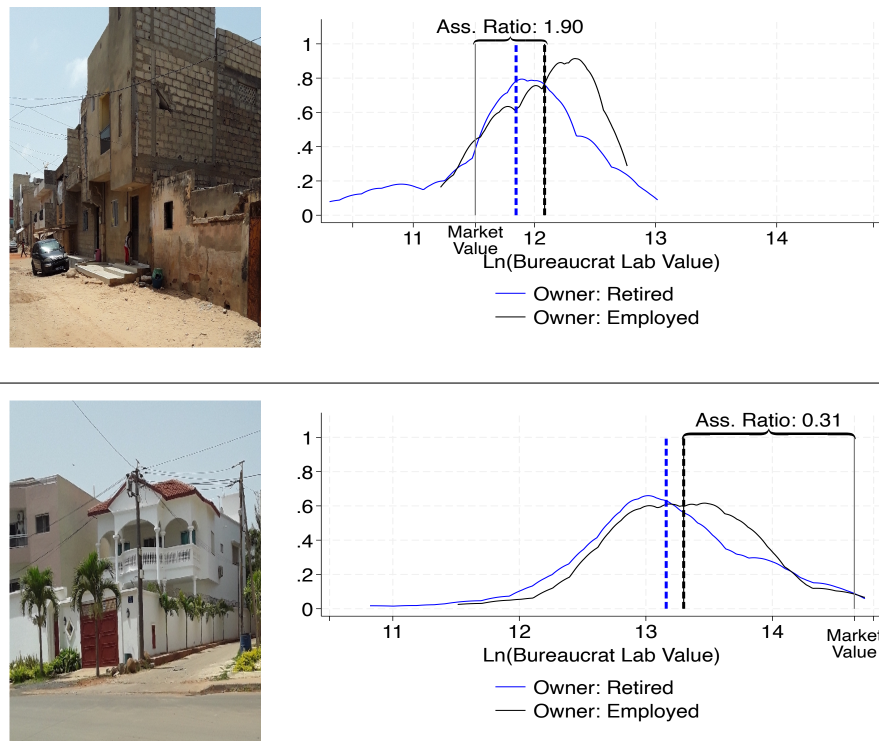
[59] This exercise is similar in spirit to the study of judges' jail-or-release decisions and their comparison with the prediction models used in the USA' judicial system (Kleinberg et al., 2018; Ludwig and Mullainathan, 2024; Angelova et al., 2025). More similar to what is done in those papers, we use bureaucrats' implicit algorithm to investigate whether they have a different objective function than the government, in Section 6.3.

[60] The R^2 is slightly lower than the one from Table A3 because we re-calibrate the algorithm on this smaller sample for which we have bureaucrats' values.

property characteristics, which is indicative of low expertise in property valuation.[61]

Taken together, these results provide evidence that bureaucrats' lack of knowledge is a first-order factor leading to the tax profile found under discretion.

FIGURE 3
LAB-IN-THE-FIELD: BUREAUCRATS' LACK OF KNOWLEDGE AND BIASES



(A) PROPERTY PICTURES **(B) BUREAUCRATS' LAB-IN-THE-FIELD VALUES**

Notes: This Figure shows results from a lab-in-the-field valuation exercise with bureaucrats conducted for a low-value property from the first quintile (Top) and a high-value property from the fifth quintile (Bottom). We show the distribution of bureaucrats' estimated value for each property picture. The gray vertical line is the true market value of each property obtained from licensed assessors. The black (resp., blue) curve is the kernel density of values for bureaucrats randomly assigned to be told the owner was employed (resp., retired). The vertical dashed lines show the mean of each distribution. We report the mean assessment ratio $Ass.Ratio$, computed as the mean value for *owner=employed* over market value. We report regression results in Table A11. See Appendix F.1 for details on the survey module and variables. Sample: 166 bureaucrats surveyed at endline.

[61] We also test whether bureaucrats' algorithm performs well in predicting the values from the census for *fully rented* properties. They offer a reliable benchmark which does not depend on assessor values. There is a 0.64 correlation between these values and their predictions using the bureaucrat algorithm. This is lower than the 0.74 correlation these values display with assessor values (Table A1). The sample is small and hence the results should be treated with caution, but they nonetheless provide reassurance that there is not a superior ability of bureaucrats to recover market values that we are missing with our methodology.

6.2 Ruling out Collusion

An alternative mechanism for the undervaluation gradient could be collusion (*active* channel), if owners of expensive properties offer bribes in exchange for lower tax liabilities. It is important to keep in mind that there was no (official) direct exchange of money between bureaucrats and owners. Taxpayers received their tax bills from bureaucrats of another administration later on, and payments were made in person in treasury offices. This makes opportunities for corruption less direct than in "tax farming" contexts where the bureaucrats directly collect tax payments or deliver tax bills (Khan et al., 2016; Dzansi et al., 2025; Weigel et al., 2024). Nevertheless, we use three pieces of evidence to rule out collusion as a driver of the undervaluation gradient.

First, the findings from lab-in-the-field valuations prove that undervaluation is strong even when there are no possible gains from corruption. The hypothetical median effective tax rate based on bureaucrats' answers is 1.8 percent for properties of the top quintile – strikingly close to the 1.7 percent found in the census.

Second, we compare the undervaluation gradient obtained in cases where the owner was met versus not met during the field visit, with the assumption that collusion would occur when the owner is met.[62] Results are shown in Figure 4: the undervaluation gradient appears similar in both cases. The median assessment ratio in the top quintile is 0.31 when the owner is met, and 0.20 when the owner is not met.[63]

Third, survey responses are consistent with these results, even if they should be treated with caution. When asked whether some colleagues have been "offered arrangements by owners" (without any reference to whether or not they accepted), 84 percent of bureaucrats reply "none" or "almost none".[64] Finally, in the endline owner survey, only 1 percent of respondents report having been solicited a bribe and 7 percent think informal payments are frequent. These figures are low compared to similar programs in other contexts.[65]

[62]This variable being reported by bureaucrats, one could worry that it is manipulated. However, we expect this to be unlikely: (i) bureaucrats were not aware that we were making comparisons with market values; (ii) bureaucrats were incentivized to meet owners and recover their identification details, their monthly bonus takes into account the share of owners for which they recover this information; (iii) the supervisors were in the field with the bureaucrats all day, and also validated all forms being submitted. It is likely supervisors would have detected if some bureaucrats spent time talking with owners, but then reported not having met them.

[63]We verify that the difference in gradients is not significant by checking that the γ_n coefficients are not significant in the following regression: $AR_i = \sum_{n=1}^{5} \beta_n Q(n)_i + \sum_{n=1}^{5} \gamma_n M_i \cdot Q(n)_i + \epsilon_i$, where AR_i is the assessment ratio for property i, the $Q(n)$ are indicators for each quintile of the distribution of market values, M_i is an indicator taking value one if the bureaucrat met the owner.

[64]When asked directly whether this happened to them, 79 percent of bureaucrats answer "never" and 20 percent "once or twice".

[65]These figures are upper bounds since the questions referred to bribes *at any step* in the taxation process. As a

6.3 Bureaucrats' Biases and Fairness Perceptions

The key feature of discretion is that bureaucrats may adjust assessments based on local information. In this section and the next, we investigate how local information shapes bureaucrats' values.

Lab-in-the-field valuation exercise: To test for biases, we included in the experimental module presented in Section 6.1 randomized information on the social status of the owner. For each property, half of the respondents were told the owner was employed, and half that he was retired.[66] In this context, being retired is associated with vulnerability. The distribution of values for retired owners is shown with the blue curve in Figure 3: whether it be for the low- or high-value property, bureaucrats provide lower values for retired owners when looking at the exact same property picture. We quantify this more precisely using a regression in Table A11 and find that bureaucrats assign a 31.5 percent lower value to a given property if told that the owner is retired, suggesting that their valuations are subject to biases.[67]

Survey responses: Bureaucrats' biases could originate either from what they believe the relationship between a certain characteristic and a property's value to be, or from adjustments they make to a property's value as a whole to align with what they perceive as fair. Bureaucrats' survey responses suggest that these biases stem from their perceptions of what is fair. 42 percent disagree with the statement that it is fair for a retired owner to pay the tax. When asked which types of owners can benefit from tax rebates, 11 percent select single mothers, although there is no specific provision in the tax code nor in their training for this category of owners. When asked what types of errors are worse, 23 percent select overvaluation, while only 8 percent select undervaluation, suggesting that there is slight reluctance to "overtax" (although too weak to explain the observed undervaluation gradient).

We next attempt to back out bureaucrats' objective function, using their implicit algorithm in which we include owner characteristics.

Bureaucrats' implicit algorithm: In columns (3) to (5) of Table A12, we add as predictors of bureaucrats' values information on the owner. The variables used in column (3) are entered by the bureaucrats – while they assign slightly lower values if the owner is deceased, receives a pension,

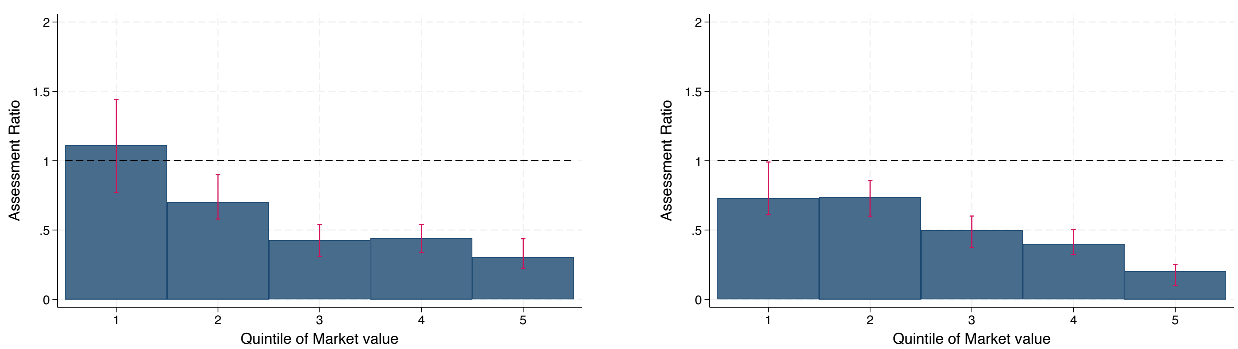
comparison, 13.9 to 24 percent of households think tax collectors solicit bribes in Madina, Ghana (Dzansi et al., 2025), and 76 percent of households in Punjab, Pakistan (Khan et al., 2016). Balan et al. (2022) in the DRC report levels of corruption similar to those we find.

[66] See details on the experimental module in Appendix F.1 and a summary of the design in Figure A13.

[67] To alleviate the concern that this could result from some form of statistical discrimination, for instance if the inside of properties owned by retirees tend to be of lower standing, or if they are poorer, we test these correlations in our survey data. We find that, if anything, being retired correlates with slightly higher household income, and does not correlate with liquidity constraints nor property values.

if the property is the main residence, or has multiple owners, the explained variation does not increase substantially ($R^2 = 0.50$) compared to the model with property characteristics only. In columns (4) and (5), we add variables from the owner surveys.[68] Bureaucrat valuations respond to households' economic situation, but this does not appear as a main driver: a one standard deviation increase in the household income index (which corresponds for instance to switching from the 10th to 60th, or from the 50th to 80th percentiles) increases bureaucrats values by only 3 percent (column (4)). In a model where missing values are imputed by the mean and controlled for using indicators, the sign on the liquidity index even suggests higher values for more constrained households (column (5)). Overall, only 49 percent of variation in bureaucrat values is explained by this enriched implicit algorithm.

FIGURE 4
THE UNDERVALUATION GRADIENT BY WHETHER THE OWNER WAS MET



(A) BUREAUCRAT MET OWNER **(B) BUREAUCRAT DID NOT MEET OWNER**

Notes: This Figure shows the median assessment ratio by quintile of market value, under discretion, separately for properties where the bureaucrat met the owner (Panel (A)), and properties where the bureaucrat did not meet the owner (Panel (B)). The assessment ratio is the tax roll value of a property divided by its market value obtained from licensed assessors. Sample: properties in the Discretion arm for which we have market values ($N = 1,123$).

Ruling out that bureaucrat adjust values to other tax-relevant characteristics Another way bureaucrats could be using local information is by adjusting their values to taxpayers' subsequent compliance with the tax.[69] We verify that this is not the case by studying subsequent tax payments. As shown in Panel (D) of Figure A5, bureaucrats generate lower effectively paid tax rates than the

[68] See details on the survey variables in Appendix F.

[69] Contextually, it is important to keep in mind that these bureaucrats were not in charge of collecting payments, and had no prior knowledge on owners' compliance, as opposed to other contexts where bureaucrats' ability to predict payments has been shown to be an important channel (Dzansi et al., 2025; Balan et al., 2022).

pure rule.[70] The tax payment profile is also more regressive under discretion.[71] The observed top-quintile rate is 0.50 percent with the pure rule against 0.21 percent with discretion. We also monitor taxpayer visits, complaints and appeals, which remain rare across the board (4 percent of properties). While there is a 1.3 percentage point higher probability of any of these occurring in the Algorithm arm, this difference disappears when controlling for tax bill amount.

Taken together, these results show that bureaucrat valuations are subject to biases. It seems that these stem partly from what they perceive as fair, beyond what the government mandates them to do – which is simply to get accurate market values. However, we can rule out that this is welfare improving with respect to the implemented tax policy.

6.4 Local Information only helps for Low-value Properties

While the previous section explored the influence of any form of information available in the field, we now test whether bureaucrats' valuations are improved in two instances where we know from the data that the field visit gave access to information that is relevant for the estimation of property values.

The first situation is one where the owner was met. The pattern from Figure 4 suggests that for low-value properties, bureaucrats' valuations are more accurate if they met the owner. This could come from information provided by the owner in the conversation, or from visual inspection made possible if the owner let the bureaucrat in the property. We investigate this more systematically by estimating regression 2 restricting the sample to properties where the bureaucrat met the owner (Panel (B) of Table A13). We further split the sample into low-value and high-value properties.[72] The results suggest that the difference in accuracy between discretion and the rule-based method is indeed smaller when the owner is met, with a non-significant coefficient of 1.37 million F on the absolute tax gap (column (3) of Panel (B.1)) against our main estimate of 3.84 million F significant at the 1 percent level in Table 2. But this does not hold for high-value properties: even when the owner is met, the tax base gap is strongly and significantly wider with discretion.[73]

[70]The tax bills were printed out with discretionary values in the Discretion sections, and pure-rule values in the Algorithm sections. Total payments were 15 percent higher under the pure rule – we return to these figures in the cost-benefit analysis in Section 8.

[71]We report mean effective tax rates, the median being 0 due to the low overall compliance rate (25 percent). For the same reason, we cannot compute the dispersion statistic for which the median is the denominator.

[72]We control for bureaucrat fixed effects to account for bureaucrat heterogeneity in ability to meet the owner. Hence, the assumption we are making to interpret these results is that for a given bureaucrat, the properties for which the owner is met do not differ across arms. Table A5 supports this assumption by confirming that the the probability to meet the owner does not differ in Algorithm versus Discretion sections.

[73]One could be worried that owners strategically provided information that deteriorates bureaucrats' judgment,

Another instance in which a bureaucrat was more likely to access valuable local information is when the property is rented. We restrict the sample to rented properties and re-estimate specification 2.[74] The results show that for high-value rented properties, the tax base gap is substantially wider under discretion (Panel (C.3)). For low-value rented properties, the difference in accuracy between rule and discretion is less concerning (only significant at the 10 percent level when considering the absolute tax base gap).

Taken together, these results show that bureaucrats are able to leverage some valuable local information to improve the accuracy of valuations, but this at best helps for low-value properties only, hence strongly limiting the advantages of discretion in this setting.

6.5 Limited Gains from Screening, Training or including Bureaucrat inputs in Algorithm

Given the strong influence of individual bureaucrats on discretionary assessments, a natural question is whether discretion could outperform algorithms if the administration were able to screen bureaucrats, better train them, or combine their inputs with the algorithm.

Screening bureaucrats. We first define top bureaucrats, building on the fixed effects estimated in Section 5.2. In Panel (A) of Figure 5, we sort bureaucrats by their estimated $\alpha_{b,EB}$. We then run specification 2, starting by the sample with all bureaucrats, and iterating by removing the worst bureaucrats one by one. The red line indicates when the coefficient on Discretion sections stops being significant. This occurs after removing the 77 (39 percent) worst bureaucrats. We label the remaining bureaucrats as top bureaucrats – those whose discretionary valuations are not significantly outperformed by the rule-based method. The coefficient remains non-significant throughout almost all iterations once the worst bureaucrats are removed. This suggests that bureaucrat screening is *at best* as accurate as the rule-based method.[75] We show the tax profile of top versus bottom bureaucrats in Panel (A) of Figure A6 – top bureaucrats' tax profile is slightly less dispersed and regressive, but still outperformed by the rule.

Next, we investigate whether there are observable characteristics that predict being a top bu-

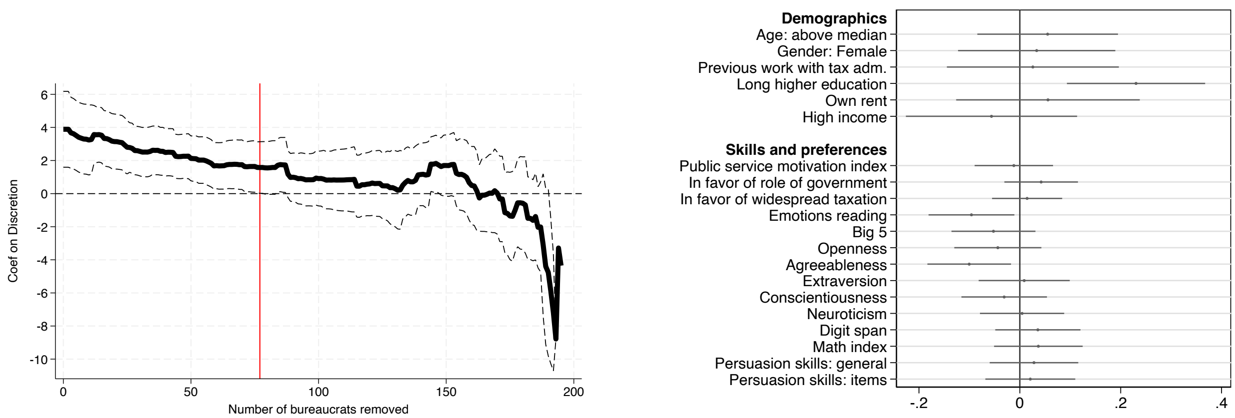
which could lead to more inaccuracies when the owner is met. To verify that bureaucrats are not doing better *without* owner inputs, in Panel (D) of Table A13, we restrict the sample to cases where the bureaucrat explicitly indicates having made the assessment herself without relying on any information from occupants. Discretion remains strongly and significantly dominated by the rule-based method.

[74] In Table A5, we verify that the probability of a property to be listed as rented does not differ across arms. In Figure A8, we show the probability of meeting the owner and of the property being rented by quintile of market value.

[75] The sign is reverted – discretion outperforms the algorithm – only when left with the 29 best bureaucrats – but the remaining sample size is small (N=120) hence this result needs to be treated with caution.

reaucrat, that the government could screen on. The results are shown in Panel (B) of Figure 5. Long higher education – defined as having three years or more of higher education – is significantly associated with a 23 pp (43 percent) higher probability of being a top bureaucrat.[76] Other bureaucrat characteristics do not correlate significantly with the probability of being a top bureaucrat.[77] The tax profile of bureaucrats with long higher education is slightly less dispersed and regressive than that of bureaucrats without long higher education (Panel (B) of Figure A6) but this is marginal when compared to the improvements offered by the algorithms.

FIGURE 5
TOP BUREAUCRATS



(A) IDENTIFYING TOP BUREAUCRATS **(B) CHARACTERISTICS OF TOP BUREAUCRATS**

Notes: In this Figure, we use the bureaucrat fixed effects estimated in Section 5.2 to identify and characterize top bureaucrats. In Panel (A), we classify top bureaucrats as those for which the difference in tax base gap between discretionary and rule-based valuations is no longer significant. We rank bureaucrats using their estimated $\alpha_{b,EB}$, and estimate regression 2 195 times, removing bureaucrats one by one starting by the worst one. The number of bureaucrats removed is indicated on the x-axis. The solid black line shows the estimated $\hat{\beta}$ coefficient on the indicator for Discretion sections and the dashed line its 95% confidence interval. The red line shows the split between bureaucrats still in the sample when the coefficient is no longer significant – the top bureaucrats – and the rest. In Panel (B), we assess whether observable characteristics correlate with being a top bureaucrat. We report coefficients from bivariate regressions of an indicator for top bureaucrat on variables captured in the bureaucrat surveys. The variables are defined in detail in Appendix F.1.

Overall, these results suggest that while screening on long higher education is a feasible way

[76] 98 percent of bureaucrats have some higher education, 42 percent have long higher education. Table 1 shows a slight imbalance in higher education across arms, as noted in footnote 39. For this reason, in Panel (D) of Figure A8, we verify that there is a meaningful share of bureaucrats with long high education in all quintiles. Furthermore, we test the robustness of the significant correlation between long higher education and being a top bureaucrat using alternative specifications in Appendix Section G.2.

[77] While very few bureaucrats are property owners (4.8 percent), those who are tenants (49 percent) report paying an average rent that is in the bottom five percent of market values. This lack of exposure to expensive properties could contribute to bureaucrats' poor valuations. The fact that our variables for "own rent" and "high income" do not significantly correlate with the probability of being a top bureaucrat may be due to their small variation in the sample of bureaucrats. In Appendix Table A16, we test whether top bureaucrats have different outcomes in their field visits: while they are significantly more likely successfully recover a property value, this is not reflected in supervisors' evaluation of their performance.

to retain top bureaucrats, their tax profile is still much less accurate and equitable than the one provided by the algorithms, especially the pure rule.

Training and learning. Second, we investigate whether bureaucrats' lack of knowledge can be corrected by training or learning-by-doing. To test the effectiveness of an easily implementable addition to bureaucrats' training, we incorporated a randomized information treatment in the lab-in-the-field valuation exercise. Half of the bureaucrats were shown a fact sheet with key figures on the distribution of property values in the Dakar region. The sheet is shown in Figure A14.[78] We do not find that the valuations of treated bureaucrats are more accurate, as reported in column (2) of Table A11. Furthermore, we test for learning in the field by assessing whether the tax base gap is reduced with the number of days worked, the number of properties covered, or after being exposed to the rule, and find that this is not the case (see Table A14).

Algorithm with bureaucrat inputs. A third way the government could leverage the information collected by bureaucrats is to incorporate their inputs in an augmented version of the algorithm. While we test this *ex post* since bureaucrats were not shown algorithm predictions during implementation, this test speaks to recent findings on the complementarities between algorithmic and human-based decisions (Agarwal et al., 2025; Sadka et al., 2024). We re-calibrate the algorithm on the Discretion arm and add variables collected by bureaucrats to see the extent to which it improves the algorithm's performance. Results are shown in Table A15. In column (2), bureaucrat's total value for the property is added to the model, as well as an indicator for whether the property is rented. In column (3), the bureaucrat's value is broken down into its two components, rent value and owner-occupied value. We repeat the exercise in columns (4) to (6) restricting to positive values. Overall, these additions only marginally improve the performance of the algorithm. The R^2 increases to 0.87 against 0.86 for the initial algorithm (column (1)), and the share of predictions within 30% of the underlying value increase to 63 percent against 58 percent.

Taken together, these results show that there is no straightforward channel through which incorporating bureaucrats' discretionary valuations leads to substantial accuracy or equity gains.

[78] See details on the experimental module in Appendix F.1 and a summary of the design in Figure A13. The fact sheet draws on Stantcheva (2021) and Hoy (2025). We verified bureaucrats' comprehension of the fact sheet using two simple interpretation questions that needed to be answered correctly before moving to the valuation.

7 Political Acceptability of the Algorithms

The previous sections have shown that the algorithms are more accurate and equitable than discretion. A final consideration to determine whether their adoption is the preferred policy is whether they are accepted by bureaucrats and taxpayers. Indeed, aversion to algorithms, and more generally backlash faced by the automation of policy-making, have been documented in other contexts (Burton et al., 2020; Browne et al., 2025).

7.1 Bureaucrat Acceptability

We use responses from the surveys to document bureaucrats' positive views about the adoption of algorithmic valuation methods. When bureaucrats were asked to rank different strategies to obtain accurate property values, they ranked "Using an automatic formula" second on average, just after "Value declared by tenant". Among the 45 bureaucrats who only worked in rule-based sections, 83 percent felt they had autonomy in their work, alleviating a common concern regarding automation in the workplace.[79] Qualitative discussions with the heads of divisions in the tax administration throughout the lifecycle of the project also point towards general optimism about the adoption of digital methods.[80]

7.2 Taxpayer Acceptability

To examine taxpayers' acceptability of algorithm-based tax assessment, we included in our property owner survey an experimental module designed to elicit taxpayers' preferences around valuation methods. Figure A15 summarizes the experimental design. Respondents were first shown two vignettes reproduced in Figure A16: one explaining the discretionary valuation by a bureaucrat and the other explaining valuation using an algorithm. They were then asked to indicate which method they considered best according to eight performance and fairness features, before indicating which method they would prefer for their own property tax. When selecting which method was best, it was always possible to answer "neither one nor the other".[81] We find that 50.3 percent of respondents prefer the algorithm over discretion for the region – revealing that there is

[79]Indeed, they were still in charge of registering properties and finding the identification details of owners and tenants. See Appendix F.1 for details on bureaucrat survey variables.

[80]This is consistent with Bachas et al. (2025) who report that 80 percent of tax inspectors, also from the Senegalese tax administration, are in favor of data-driven methods for the selection of audit cases.

[81]The questions used in the module are laid out in Appendix F.2.

no strong algorithm aversion nor appreciation in this setting. The share declines to 44.5 percent when asked about preferences for their own property.

In Panel (A) of Figure 6, we show that these preferences vary by property value. The wealthiest taxpayers are the ones with the strongest preference for the algorithm. In the lowest quintile, 37 percent of respondents prefer the algorithm and 63 percent discretion, while in the top quintile, 58 percent prefer the algorithm and 42 percent discretion.[82] Paradoxically, owners at the top of the distribution are those we have shown to "lose" the most from the adoption of the algorithm, due to higher tax liabilities.[83]

To better understand the drivers of taxpayers' preferences, we estimate correlations between preferring the algorithm (versus discretion) for oneself and considering it is best to achieve each of the eight performance and fairness dimensions. Results are shown in Panel (B) of Figure 6. The feature displaying the strongest correlation is whether taxpayers' think the algorithm will yield the lowest tax bill (correlation of 0.83). Whether respondents consider the algorithm provides more fairness, more transparency, more horizontal equity, more vertical equity, and is best for the region also all correlate strongly with it being their preferred method, but to a lesser extent (0.72 to 0.76). While almost all respondents (76 percent) view the algorithm as less prone to corruption, there is still a large share not selecting it as their preferred method – hence the low displayed correlation (0.47).

Documenting taxpayers' preferences is informative in itself, but the pressing question for policy-makers is whether these preferences matter for tax compliance decisions. We use a hypothetical tax bill exercise to test two hypotheses. First, we assess whether the valuation method applied to an individual's tax bill matters for compliance, controlling for tax bill amount. Second, we assess whether learning that one method (the algorithm) is more equitable matters for tax compliance.[84] Respondents were told about a hypothetical tax bill for next year and were asked how likely it was that they would pay this tax. Let us denote each tax bill as $T(A, M)$, where A is its amount

[82] We verify that the probability to prefer the algorithm in the fifth quintile is statistically significantly larger than that of the first quintile. The pattern does not stem from differences in education, since the results are similar when controlling for whether the respondent completed tertiary education (36 percent). The results do not either stem for general preferences around digitization – 97 percent of respondents are favorable to the government increasing the use of digitization, and there is no gradient by property values.

[83] No reference was made to respondents' own tax bills. Enumerators were blind to the randomization of the census, and we do not find a significant difference in the preference for the algorithm across the two arms.

[84] To our knowledge, our study is the first to document taxpayers' preferences regarding tax assessment methods, and in particular the use of algorithms. The second hypothesis resonates with existing work testing whether taxpayers' acceptance of a tax responds to increased knowledge about how equitable or progressive it is (Stantcheva, 2021; Ajzenman et al., 2025; Hoy, 2025; Best et al., 2025).

and M the assessment method. These two dimensions were crossed-randomized in a two-by-two design. Each respondent faced an amount A with $A \in \{Low, High\}$ and a method M with $M \in \{Algorithm, Bureaucrat\}$.[85] We estimate the effect of A and M on taxpayers' willingness to pay the tax. Results are shown in Panel (C) of Figure 6, with the two bars to the left. The willingness to pay the tax is 10.2 percentage points or 15.9 percent higher for the low tax bill – which is expected and provides reassurance that we can meaningfully learn from these survey responses.[86] The coefficient on $Algorithm$ is positive but small and insignificant – suggesting that algorithmic assessment in itself does not matter for compliance.[87]

In a second step, respondents received an information treatment explaining that the algorithm provides stronger horizontal tax equity, and asked again about their willingness to pay the tax bill.[88] We find that this leads to a strong increase in the coefficient on $Algorithm$, now associated with a 2.9 percentage points or 4.6 percent higher willingness to pay the tax, significant at the 10 percent level. This is equivalent to the estimated effect of a 14.5 percent lower tax rate, or to switching from the legal tax rate from 8.6 to 7.3 percent. Nonetheless, the effect remains small in size, confirming that respondents' compliance decisions are most sensitive to their own tax liability.

Taken together, these results are reassuring. The government should not expect resistance to the adoption of algorithmic methods, even more so if efforts are made to communicate about their benefits in terms of tax equity.

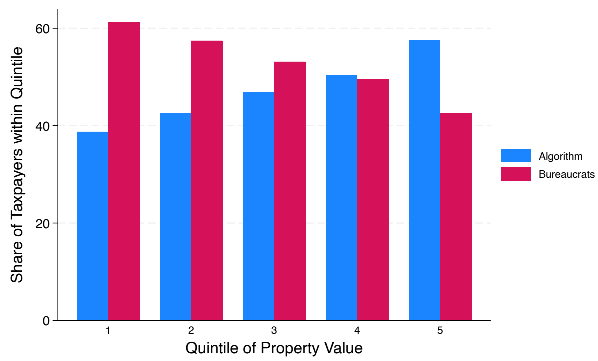
[85] See Appendix F.2 for additional details. The tax bill amounts are personalized for each taxpayer based on predicted property values, with the high tax bill amount being in the range of the true tax liabilities. The low tax bill corresponds to a 50 percent discount relative to the high tax bill.

[86] Our result is in a similar range to that estimated in the DRC where a 50% decrease in the tax rate increases compliance by 7.4 percentage points (Bergeron et al., 2024). Converting our result to an elasticity estimate yields that a 1 percent increase in the tax rate decreases compliance by 0.32 percent.

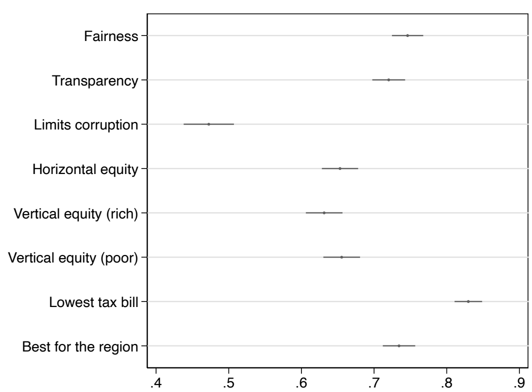
[87] We verify that the coefficient on $Algorithm$ remains non-significant when we do the estimation separately for low and high tax bills.

[88] The exact phrasing was "With the latter (a bureaucrat in the field), two similar properties of the same neighborhood very often have different tax bill amounts. With the first method (the formula) two similar properties of the same neighborhood very often have the same tax bill." See Appendix F.2 for additional details.

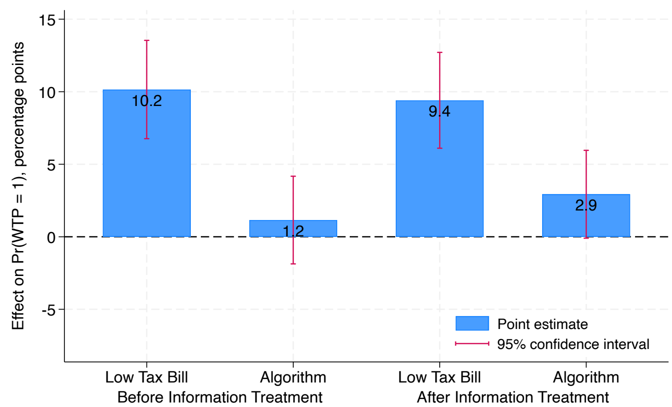
FIGURE 6
TAXPAYERS' PREFERENCES OVER VALUATION METHODS



(A) TAXPAYERS' PREFERRED METHOD BY QUINTILE OF PROPERTY VALUE



(B) CORRELATES OF TAXPAYERS' PREFERRED METHOD



(C) EFFECT OF TAX BILL AMOUNT AND ALGORITHM ON WILLINGNESS TO PAY TAX

Notes: This Figure shows taxpayers' preferences over valuations methods – *Bureaucrat discretion* versus *Algorithm* – elicited using the vignettes pictured in Appendix F.2 and a hypothetical tax bill exercise. Panel (A) shows the share of taxpayers preferring each method for their own property by quintile of property value (property values are predicted using the pure rule algorithm). In Panel (B), we report pairwise correlations between preferring the *Algorithm* and selecting the *Algorithm* as the best method to achieve a given feature, for eight different features. Panel (C) reports results from a regression where the outcome variable is the willingness-to-pay (WTP) a hypothetical tax bill. Respondents were presented with a tax bill for next year with either a *High* or *Low* amount, assessed using the *Algorithm* or *Bureaucrat discretion*, following a 2x2 randomization design. The bars to the left show coefficients for the first round. The bars to the right show coefficients for the second round after an information treatment explaining that the algorithm provides stronger horizontal tax equity. We provide details on the survey module in Appendix F.2. Sample: N = 3,958 respondents from the property owner endline survey.

8 Towards an Optimal Policy

Building on the accuracy and equity results, and on the finding that algorithms *per se* do not affect compliance, we lay out the optimal assessment policy, before turning to cost-benefit considerations.

8.1 Optimal Policy

The pure rule is the method providing the most accuracy, horizontal equity, and vertical equity. Hence, given these objectives, it is the optimal policy for the government. However, there is a trade-off if the government additionally wishes to limit the risk of overvaluing low-value properties. While the pure rule remains more accurate than discretion as measured by the tax base gap, it tends to overvalue low-value properties. These are precisely the properties for which bureaucrats perform relatively better. Therefore, an alternative optimal policy is to use discretion in the first quintile, and the pure rule in quintiles two to five. In Panel (E) of Figure A5, we plot the resulting tax profile, which proves satisfactory when compared to the legal tax profile.[89]

8.2 Cost-benefit Analysis

We assess the costs and benefits of each method in Table A17. We consider a scenario in which the government plans to update the tax roll a few years from now. There are two types of costs, field-work costs and costs specific to the calibration of an algorithm. We first consider assessed tax amounts. The ratio of total assessed amounts to costs is the lowest with discretion (133), the highest with the pure rule (798), and 463 with the optimal policy. Next, we consider paid amounts. The ratio of payments to costs (respectively, the net gain per property) is 35 (resp., 35,600 F) with the pure rule versus 8.5 (resp., 26,900 F) for discretion. Hence, the adoption of the pure rule is the most cost-effective strategy.[90]

[89] Two caveats apply. First, the prediction of belonging to the lowest quintile yields 17 percent false negatives – these properties will be overtaxed. Second, using discretion will necessarily entail less horizontal equity in this segment of the market.

[90] We provide details on the calculations in Table A17.

9 Conclusion

We study the introduction of an algorithm to replace bureaucrats' discretion in the assessment of the tax base in the first digital property tax census in Senegal, spanning over 41,600 properties in the capital city. Our setting uniquely allows us to evaluate the accuracy, equity and acceptability of algorithmic decision-making in comparison to the status quo.

We find that the algorithms outperform discretion in accuracy, horizontal equity – properties of similar values are more likely to face similar tax bills – and vertical equity – the tax profile is less regressive. A pure-rule algorithm, using only variables that can be collected remotely, is more accurate than a rule-based algorithm incorporating inputs from the bureaucrats. Under full discretion, bureaucrats' property valuations are significantly below market values, especially for expensive properties, which generates a regressive tax profile. Finally, we show that taxpayers are indifferent to the use of an algorithm conditional on tax liability, hinting to limited risk of political backlash.

Overall, the pure-rule algorithm is the most promising strategy for an equitable expansion of the tax net. There is a trade-off for the lowest-value properties only, where even if discretion provides lower accuracy and horizontal equity, it reduces the risk of overvaluation.

We show that bureaucrats' lack of knowledge about high-value properties plays a fundamental role, while collusion is ruled out as a main driver of the regressive profile they generate. This aligns with the notion of *passive* rather than *active* waste. We also rule out the possibility that bureaucrats successfully predict tax compliance. In fact, the pure rule led to more tax collections and a more equitable payment profile compared to discretion.

Our results have already had a strong policy impact: the administration has adopted the digital tool and has started working on the country-wide expansion of algorithm-based valuation methods, while soliciting the research team for continued collaboration. At a time when algorithms are being increasingly used for public policies, our findings also point to directions for future research, notably the importance of documenting the equity and acceptability of algorithms in comparison to the status quo decision process.

References

Agarwal, Nikhil, Alex Moehring, Pranav Rajpurkar, and Tobias Salz, "Combining Human Expertise with Artificial Intelligence: Experimental Evidence from Radiology," *working paper*, November 2025. 1, 6.5

Aghion, Philippe and Jean Tirole, "Formal and Real Authority in Organizations," *Journal of Political Economy*, 1997, *105* (1), 1–29. 1

Aiken, Emily, Suzanne Bellue, Dean Karlan, Chris Udry, and Joshua Blumenstock, "Machine learning and phone data can improve targeting of humanitarian aid," *Nature*, 2022, *603*, 864–870. 1, 3

Ajzenman, Nicolas, Guillermo Cruces, Ricardo Perez-Truglia, Darío Tortarolo, and Gonzalo Vazquez-Bare, "From Flat to Fair? The Effects of a Progressive Tax Reform," Working Paper 33286, National Bureau of Economic Research December 2025. 1, 84

Ali, Daniel, Klaus Deininger, and Michael Wild, "Using Satellite Imagery to Revolutionize Creation of Tax Maps and Local Revenue Collection," 2018, *8437.* Wold Bank Policy Research Working Paper. 19, 4.2, E, 100

Almy, Richard, "Valuation and Assessment of Immovable Property," *OECD Working Papers on Fiscal Federalism*, 2014, *19*. 9

Amornsiripanitch, Natee, "Why are Residential Property Tax Rates Regressive?," *Working Paper*, 2024. 5.3, E

Angelova, Victoria, Will Dobbie, and Crystal S Yang, "Algorithmic Recommendations and Human Discretion," *The Review of Economic Studies*, 09 2025, p. rdaf084. 1, 24, 59

Avenancio-León, Carlos F and Troup Howard, "The Assessment Gap: Racial Inequalities in Property Taxation," *The Quarterly Journal of Economics*, 02 2022, *137* (3), 1383–1434. 1, 3, 40, D.4

Bachas, Pierre, Anne Brockmeyer, Alipio Ferreira, and Bassirou Sarr, "Algorithms and Bureaucrats: Evidence from Tax Audit Selection in Senegal," *World Bank Research Paper*, 2025. 1, 5, 24, 80

Balan, Pablo, Augustin Bergeron, Gabriel Tourek, and Jonathan Weigel, "Local Elites as State Capacity: How City Chiefs use Local Information to increase Tax Compliance in the Democratic Republic of the Congo," *American Economic Review*, 2022, *112*, 762–97. 65, 69

Bandiera, Oriana, Andrea Prat, and Tommaso Valletti, "Active and Passive Waste in Government Spending: Evidence from a Policy Experiment," *American Economic Review*, September 2009, *99* (4), 1278–1308. 1, 3

__ , Michael Best, Adnan Khan, and Andrea Prat, "The Allocation of Authority in Organizations: A Field Experiment with Bureaucrats," *The Quarterly Journal of Economics*, 2021, *136* (4), 2195–2242. 1

Battaglini, Marco, Luigi Guiso, Chiara Lacava, Douglas L. Miller, and Eleonora Patacchini, "Refining public policies with machine learning: The case of tax auditing," *Journal of Econometrics*, 2025, *249*, 105847. 1, 6

Behr, Daniela, Lixue Chen, Ankita Goel, Khondoker Haider, Sandeep Singh, and Asad Zaman, "Estimating House Prices in Emerging Markets and Developing Economies: A Big Data Approach," *World Bank Working Paper*, 2023. 2.1, 4.2, 100

Bergeron, Augustin, Gabriel Tourek, and Jonathan L. Weigel, "The State Capacity Ceiling on Tax Rates: Evidence from Randomized Tax Abatements in the DRC," *Econometrica*, 2024, *92* (4), 1163–1193. 8, 11, 16, 22, 4.2, 86, 100

Berry, Christopher, "Reassessing the Property Tax," *Working Paper*, 2021. 5.3, E

Besley, Timothy and Torsten Persson, "The Origins of State Capacity: Property Rights, Taxation, and Politics," *American Economic Review*, September 2009, *99* (4), 1218–44. 1

Best, Michael Carlos, Jonas Hjort, and David Szakonyi, "Individuals and Organizations as Sources of State Effectiveness," *American Economic Review*, August 2023, *113* (8), 2121–67. 1, 48

Best, Michael, Luigi Caloi, François Gerard, Evan Kresch, Joana Naritomi, and Laura Zoratto, "Greener on the Other Side: Inequity and Tax Compliance," Technical Report CEPR Discussion Paper No. 20510 July 2025. 1, 84

Björkegren, Daniel, Joshua Blumenstock, and Samsun Knight, "Manipulation-Robust Prediction," *mimeo*, 2024. 7

Black, Emily, Hadi Elzayn, Alexandra Chouldechova, Jacob Goldin, and Daniel Ho, "Algorithmic Fairness and Vertical Equity: Income Fairness with IRS Tax Audit Models," *Working Paper in ACM Conference on Fairness, Accountability and Transparency*, June 2022. 1, 6, 3

Bonmartel, Suzanne, Bilal Choho, Denis Cogneau, Marc Gurgand, Victor Pouliquen, and Bassirou Sarr, "Scaling Fiscal Capacity: Evidence from Senegal's Capital City. Final Research Report," *mimeo*, 2026. 32

Brockmeyer, Anne, Alejandro Estefan, Juan Carlos Suárez Serrato, and Karina Ramirez Arras, "Taxing Property in Developing Countries: Theory and Evidence from Mexico," *NBER Working Paper*, 2021, (28637). 8

Browne, Oliver, Ludovica Gazze, Michael Greenstone, and Olga Rostapshova, "Man vs. Machine: Technological Promise and Political Limits of Automated Regulation Enforcement," *The Review of Economics and Statistics*, July 2025, (4), 1136–1148. 1, 7

Burton, Jason W., Mari-Klara Stein, and Tina Blegind Jensen, "A systematic review of algorithm aversion in augmented decision making," *Journal of Behavioral Decision Making*, 2020, *33* (2), 220–239. 1, 4, 7

Chandra, Amitabh, Amy Finkelstein, Adam Sacarny, and Chad Syverson, "Health Care Exceptionalism? Performance and Allocation in the US Health Care Sector," *American Economic Review*, August 2016, *106* (8), 2110–44. 46

Chen, Huaizhi and Lauren Cohen, "Assessing Assessors," Working Paper 33238, National Bureau of Economic Research December 2024. 7

Chetty, Raj, John N Friedman, Emmanuel Saez, Nicholas Turner, and Danny Yagan, "Income Segregation and Intergenerational Mobility Across Colleges in the United States*," *The Quarterly Journal of Economics*, 02 2020, *135* (3), 1567–1633. 51

_ **, Nathaniel Hendren, Maggie R Jones, and Sonya R Porter**, "Race and Economic Opportunity in the United States: an Intergenerational Perspective*," *The Quarterly Journal of Economics*, 12 2019, *135* (2), 711–783. 51

Chioda, Laura, David Contreras-Loya, Paul Gertler, and Dana Carney, "Making Entrepreneurs: Returns to Training Youth in Hard versus Soft Business Skills," *NBER Working Paper 28845*, May 2021. F.1.2.1

Christensen, Darin and Francisco Garfias, "The Politics of Property Taxation: Fiscal Infrastructure and Electoral Incentives in Brazil," *The Journal of Politics*, 2021, *83* (4), 1399–1416. 1

Corbett-Davies, Sam, Johann Gaebler, Hamed Nilforoshan, Ravi Shroff, and Sharad Goel, "The Measure and Mismeasure of Fairness," *Journal of Machine Learning Research*, 2023, *24.* 1, 3

Davis, Peadar, William McCluskey, Terry Grissom, and Michael McCord, "An empirical analysis of simplified valuation approaches for residential property tax purposes," *Property Management*, 2012, *30* (3), 232–254. 19, E

Decarolis, Francesco, Raymond Fisman, Paolo Pinotti, and Silvia Vannutelli, "Rules, Discretion, and Corruption in Procurement: Evidence from Italian Government Contracting," *Journal of Political Economy: Microeconomics*, 2026. 1

Dessein, Wouter, "Authority and Communication in Organizations," *Review of Economic Studies*, 2002, *69*, 811–838. 1

Dietvorst, Berkeley, Joseph Simmons, and Cade Massey, "Algorithm aversion: People erroneously avoid algorithms after seeing them err," *A Journal of Experimental Psychology: General*, 2015, *144*, 114–126. 4

Dodell-Feder, David, Kerry Ressler, and Laura Germine, "Social cognition or social class and culture? On the interpretation of differences in social cognitive performance," *Psychol Med*, January 2020, *50* (1), 133–145. F.1.2.1

Dray, Sacha, Camille Landais, and Stefanie Stantcheva, "Wealth and Property Taxation in the United States," *mimeo*, 2025. 40, D.4

Duflo, Esther, Michael Greenstone, Rohini Pande, and Nicholas Ryan, "The Value of Discretion in the Enforcement of Regulation: Experimental Evidence and Structural Estimates from Environmental Inspections in India," *Econometrica*, 2018, *86* (6), 2123–2160. 1

_, Rema Hanna, Benjamin Olken, and Shreya Tandon, "The Limits of Government Outsourcing: Property Tax Re-Assessments in India," *working paper*, 2025. 22

Dzansi, James, Anders Jensen, David Lagakos, and Henry Telli, "Technology and Local State Capacity: Evidence from Ghana," *working paper*, 2025. 1, 11, 16, 18, 6.2, 65, 69

D'Arcy, Michelle, Marina Nistotskaya, and Ola Olsson, "Cadasters and Economic Growth: A Long-Run Cross-Country Panel," *Journal of Political Economy*, 2024, *132* (11), 3785–3826. 1

Elzayn, Hadi, Evelyn Smith, Thomas Hertz, Cameron Guage, Arun Ramesh, Robin Fisher, Daniel E Ho, and Jacob Goldin, "Measuring and Mitigating Racial Disparities in Tax Audits*," *The Quarterly Journal of Economics*, 09 2024, *140* (1), 113–163. 6, 3

Fenizia, Alessandra, "Managers and Productivity in the Public Sector," *Econometrica*, 2022, *90* (3), 1063–1084. 1

Finan, Frederico, Benjamin Olken, and Rohini Pande, "Chapter 6 - The Personnel Economics of the Developing State," in Abhijit Vinayak Banerjee and Esther Duflo, eds., *Handbook of Economic Field Experiments*, Vol. 2 of *Handbook of Economic Field Experiments*, North-Holland, 2017, pp. 467–514. 1

Fish, Paul, "Practical Guidance Note: Training Manual for Implementing Property Tax Reform with a Points-Based Valuation," 2018. ICTD African Tax Administration Paper 2. 19, E

Franklin, Simon, "The demand for government housing: evidence from lotteries for 200,000 homes in Ethiopia," *mimeo*, 2019. 4.2, 100

Franzsen, Riel and William McCluskey, *Property Tax in Africa: Status, Challenges and Prospects*, Lincoln Institute of Land Policy, Cambridge, Massachusetts, 2017. 9, 2.1, 19, E

Guan, Jian, Jozef Zurada, and Alan Levitan, "A comparison of regression and artificial intelligence methods in a mass appraisal context," *Journal of Real Estate Research*, 2011, *33* (3), 349–387. 19, E

Haseeb, Muhammad and Kate Vyborny, "Data, discretion and institutional capacity: Evidence from cash transfers in Pakistan," *Journal of Public Economics*, 2022, *206*, 104535. 1

Hoy, Christopher, "How does progressivity impact tax morale? Experimental evidence across developing countries," *Journal of Development Economics*, 2025, *172*, 103398. 1, 6.1, 78, 84, F.1.2.2

— **and Franziska Mager**, "Why Are Relatively Poor People Not More Supportive of Redistribution? Evidence from a Randomized Survey Experiment across Ten Countries," *American Economic Journal: Economic Policy*, November 2021, *13* (4), 299–328. 6.1

Hvidberg, Kristoffer B, Claus T Kreiner, and Stefanie Stantcheva, "Social Positions and Fairness Views on Inequality," *The Review of Economic Studies*, 03 2023, p. rdad019. 1, 6.1

International Association of Assessing Officers, "A Review of the Methods, Applications and Challenges of Adopting Artificial Intelligence in the Property Assessment Office," 2022. 19, D.4, E

Jack, William and Tavneet Suri, "Risk Sharing and Transactions Costs: Evidence from Kenya's Mobile Money Revolution," *American Economic Review*, January 2014, *104* (1), 183–223. 1

Kala, Namrata, "The Impacts of Managerial Autonomy on Firm Outcomes," *Econometrica*, 2024, *92* (6), 1777–1800. 1

Kane, Thomas J and Douglas O Staiger, "Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation," *National Bureau of Economic Research Working Paper 14607*, December 2008, (14607). 46

Khan, Adnan Q., Asim I. Khwaja, and Benjamin A. Olken, "Tax Farming Redux: Experimental Evidence on Performance Pay for Tax Collectors," *The Quarterly Journal of Economics*, 2016, *131* (1), 219–271. 1, 41, 6.2, 65

—, —, and —, "Making Moves Matter: Experimental Evidence on Incentivizing Bureaucrats through Performance-Based Postings," *American Economic Review*, 2019, *109* (1), 237–70. 1

Kleinberg, Jon, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig, and Sendhil Mullainathan, "Human Decisions and Machine Predictions*," *The Quarterly Journal of Economics*, 08 2018, *133* (1), 237–293. 1, 3, 59

Knebelmann, Justine, "Digitalisation of property taxation in developing countries: Recent advances and remaining challenges," *ODI Report*, 2022. 2.1, 18

Lee, David S., "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects," *The Review of Economic Studies*, 07 2009, *76* (3), 1071–1102. G.1

Lizardo, Pablo Guzmán, Niccolò Meriggi, and Wilson Prichard, "Unlocking Tax Compliance: The Role of Enforcement and Religious Appeals," *working paper*, 2025. 22

Ludwig, Jens and Sendhil Mullainathan, "Machine Learning as a Tool for Hypothesis Generation*," *The Quarterly Journal of Economics*, 01 2024, *139* (2), 751–827. 24, 59

Manwaring, Priya and Tanner Regan, "No shame in my name: Public disclosure and tax compliance in a low-capacity state," 2025. Working Paper. 16

Martínez, Luis R., "Natural Resource Rents, Local Taxes, and Government Performance: Evidence from Colombia," *The Review of Economics and Statistics*, 04 2023, pp. 1–28. 1

McCluskey, William, M. McCord, P. Davis, M. Haran, and D. McIlhatton, "Prediction accuracy in mass appraisal: a comparison of modern approaches," *Journal of Property Research*, 2013, *30* (4), 239–265. 19, E

McMillen, Daniel and Ruchi Singh, "Assessment Regressivity and Property Taxation," *Journal of Real Estate Finance and Economics*, 2020, *60* (1), 155–169. 40, 5.3, E

Moore, Wayne, "Performance Comparison of Automated Valuation Models," *Journal of Property Tax Assessment and Administration*, 2005, *3* (1), 43–60. 19, E

Morris, Carl N., "Parametric Empirical Bayes Inference: Theory and Applications," *Journal of the American Statistical Association*, 1983, *78* (381), 47–55. 46

Mullainathan, Sendhil, "Economics in the Age of Algorithms," *AEA Papers and Proceedings*, May 2025, *115*, 1–23. 1

Muralidharan, Karthik, Paul Niehaus, and Sandip Sukhtankar, "Building State Capacity: Evidence from Biometric Smartcards in India," *American Economic Review*, October 2016, *106* (10), 2895–2929. 1

Nathan, Brad, Ricardo Perez-Truglia, and Alejandro Zentner, "Paying Your Fair Share: Perceived Fairness and Tax Compliance," *NBER Working Paper No. 32588*, 2023. 1

Obermeyer, Ziad, Brian Powers, Christine Vogeli, and Sendhil Mullainathan, "Dissecting racial bias in an algorithm used to manage the health of populations," *Science*, 2019, *366*, 447–453. 1

OECD, "Governing with Artificial Intelligence: The State of Play and Way Forward in Core Government Functions," *OECD Publishing, Paris*, 2025. 1

Okunogbe, Oyebola, "Becoming Legible to the State. The Role of Detection and Enforcement Capacity in Tax Compliance," *World Bank Policy Research Working Paper 9852*, 2021. 1, 11, 16

Plaisant, Odile, "Validation par Analyse Factorielle du Big Five Inventory français (BFI-Fr) Analyse Convergente avec le NEO-PI-R," *Annales medio-psychologiques*, 2008. F.1.2.1

Sadka, Joyce, Enrique Seira, and Christopher Woodruff, "Information and Bargaining through Agents: Experimental Evidence from Mexico's Labour Courts," *The Review of Economic Studies*, 02 2024, *91* (6), 3677–3711. 6.5

Stantcheva, Stefanie, "Understanding Tax Policy: How do People Reason?," *The Quarterly Journal of Economics*, 09 2021, *136* (4), 2309–2369. 1, 6.1, 78, 84, F.1.2.2

vom Hau, Matthias, José Alejandro Peres-Cajías, and Hillel David Soifer, "No taxation without informational foundation: on the role of legibility in tax state development," *Journal of Institutional Economics*, 2023, *19* (3), 426–443. 1

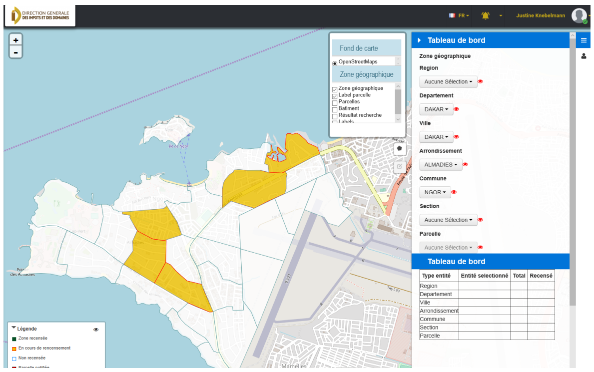
Weidmann, Ben and David Deming, "Team Players: How Social Skills Improve Team Performance," *Econometrica*, November 2021, *89* (6), 2637–2657. F.1.2.1

Weigel, Jonathan, Pedro Bessone, Augustin Bergeron, Gabriel Tourek, and John Kabeya Kabeya, "Supermodular Bureaucrats: Evidence from Randomly Assigned Tax Collectors in the DRC," 2024. 1, 48, 6.2

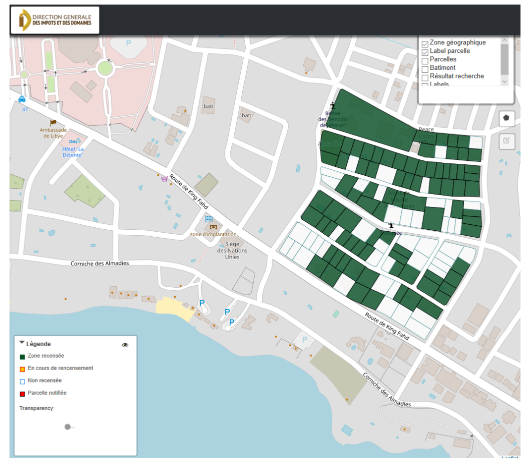
Supplemental Appendix

A Additional Figures

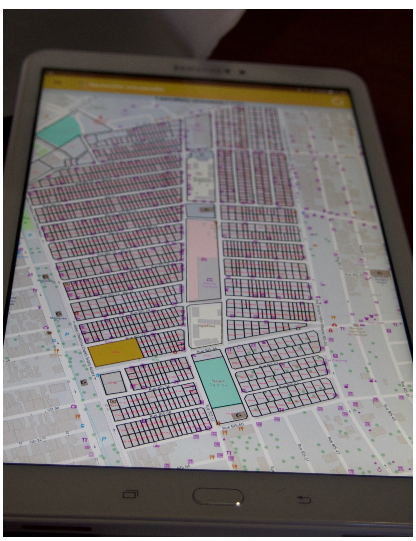
FIGURE A1
PROPERTY TAX CENSUS OPERATIONS WITH THE NEW DIGITAL TOOL



(A) WEB INTERFACE: SECTIONS



(B) WEB INTERFACE: PROPERTIES WITHIN A SECTION

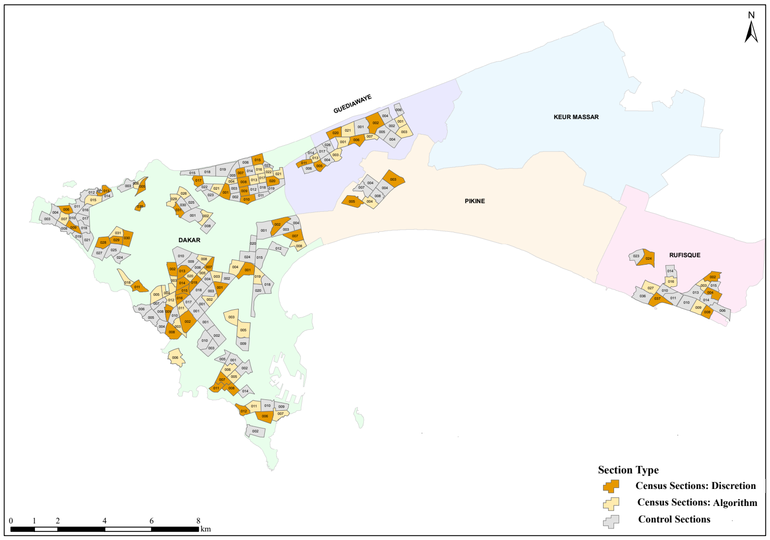


(C) TABLET INTERFACE WITH PRE-LOADED PROPERTY IDENTIFIERS

(D) BUREAUCRATS

Notes: This Figure illustrates the property tax census operations using the new digital tool. The digital tool has a Web interface for tax office managers, and a tablet interface used by bureaucrats in the field.

Figure A2
Experimental Design in the Dakar Region



Notes: This map shows the design of our experiment in the region of Dakar, which comprises the cities of Dakar, Guediawaye, Pikine, Rufisque and Keur Massar. 48 cadastral sections (in orange) were randomly assigned to be covered by the property tax census with the discretionary valuation method, while another 48 sections (in yellow) were assigned to be covered by the census with algorithmic valuation. An additional 97 sections (in gray) were randomly assigned to be a pure control group with no census. We stratified by tax office, number of properties, share eligible for the tax. The 96 sections assigned to treatment span 42,423 properties. 94 of them were effectively covered by the census.

Figure A3
Examples of Low, Medium and High Value Properties in Study Areas

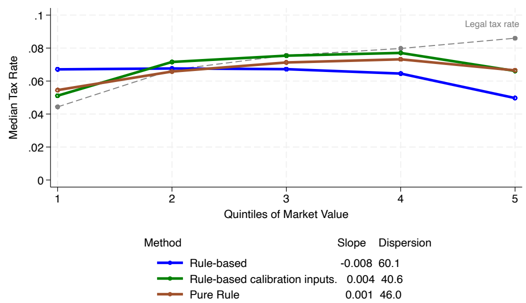
(A) Low-value property

(B) Medium-value property

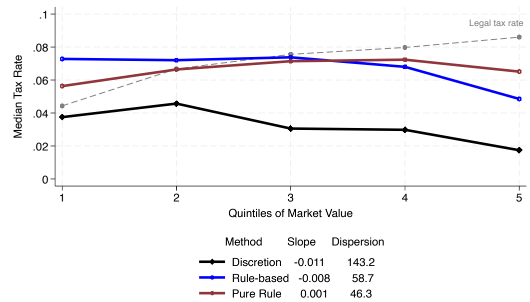
(C) High-value property

Notes: This Figure shows properties from the study areas in the region of Dakar. Picture (A) shows a property from the bottom 10% of the property value distribution (monthly rental value of 100,000 F or $163), Picture (B) shows a property with a value around the median of the distribution (520,000 F or $845), and Picture (C) shows a property from the top 10% of the distribution (2.3 million F or $3,740). Source: market values by licensed real estate assessors.

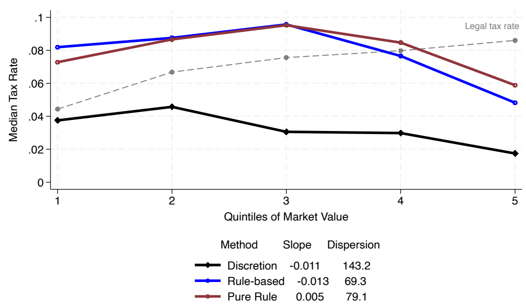
Figure A4
Tax Profiles: Robustness and Additional Results (1/2)



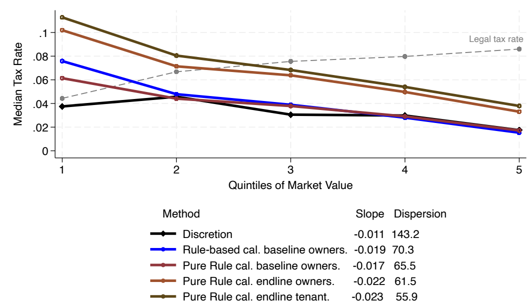
(a) Rule-Based with calibration Inputs



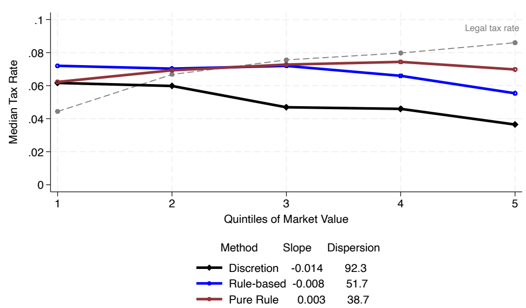
(b) Calibration Excludes Half of Algorithm Arm



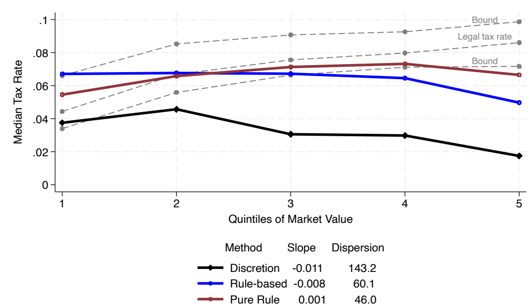
(c) Calibration Excludes Census Areas



(d) Calibration on Survey Values



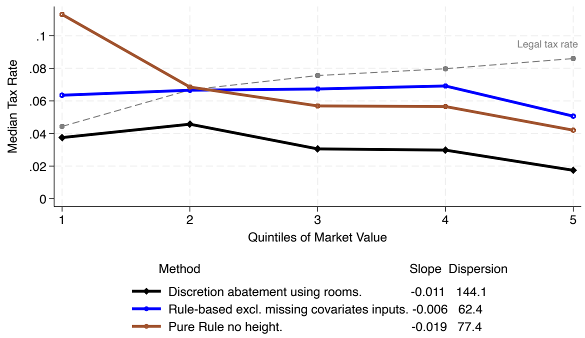
(e) Intensive Margin



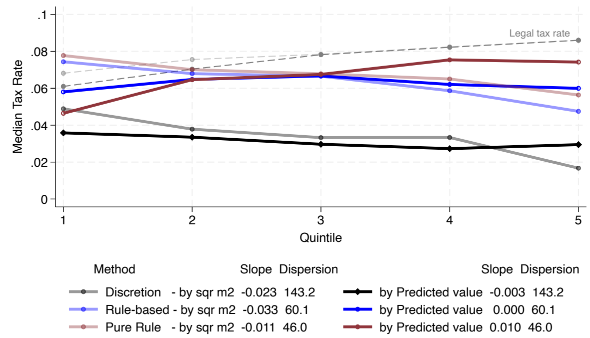
(f) Bounds on Market Values

Notes: This Figure shows the tax profile for each method, using alternative computations as complements to the results shown in Figure 2. We display the median effective tax rate by quintile of market value. A property's effective tax rate is computed as the tax liability based on the tax roll value divided by its market value. In Appendix G, we provide details on the values used in each Panel.

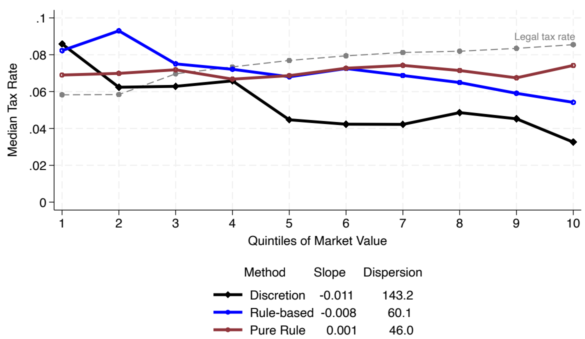
Figure A5
Tax Profiles: Robustness and Additional Results (2/2)



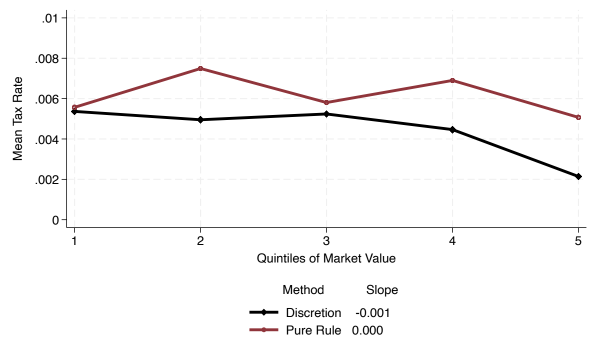
(a) Alternative calculations



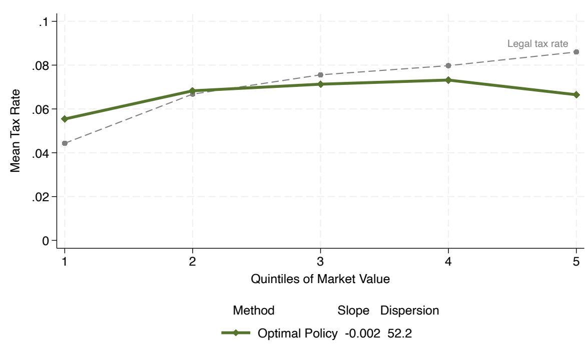
(b) Alternative quintiles



(c) Median Tax Rates by Decile



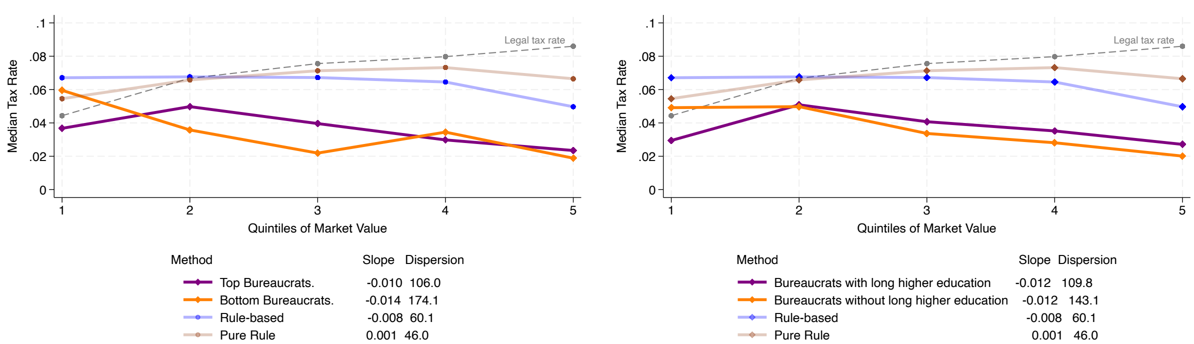
(d) Mean Tax Rates in Payments



(e) Optimal Policy

Notes: This Figure shows the tax profile for each method, using alternative computations as complements to the results shown in Figure 2. We display the median effective tax rate by quintile of market value. A property's effective tax rate is computed as the tax liability based on the tax roll value divided by its market value. In Appendix G, we provide details on the values used in each Panel.

Figure A6
Tax Profiles when Screening Bureaucrats

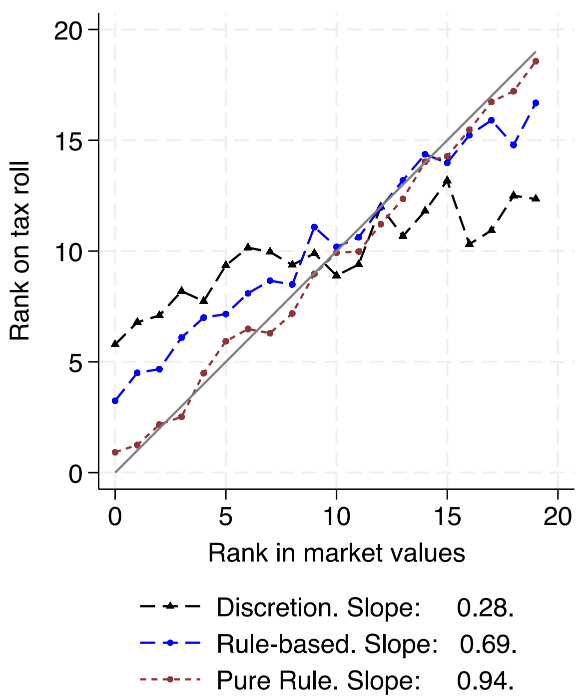


(A) Top vs Bottom Bureaucrats

(B) By Bureaucrat's Education Level

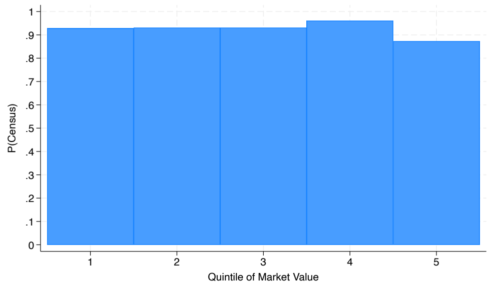
This Figure shows the tax profiles when splitting the discretionary valuations depending on the bureaucrat who covered each property. In Panel (A), we split by top versus bottom bureaucrats, where top bureaucrats are defined using bureaucrat fixed effects, following Section 6.5 and Figure 5. In Panel (B), we split by whether or not the bureaucrat completed long higher education (three years or more).

FIGURE A7
VERTICAL EQUITY: RANK-RANK CORRELATIONS

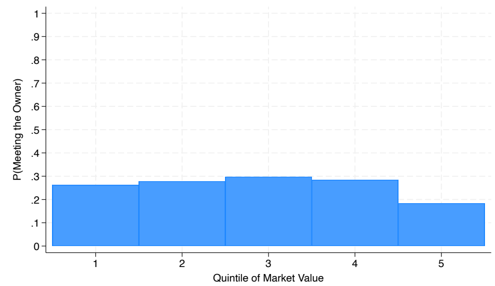


Notes: This Figure shows the rank-rank correlation between tax roll values and market values, separately for the discretionary valuations (black line), rule-based valuations (blue line) and pure rule valuations (red line). The market value is the value obtained from licensed assessors. Tax roll value is the value from the census that ends up on the tax roll. The x-axis shows a property's rank in market values, grouped in 20 bins. Ties are assigned the same rank. The y-axis shows the mean rank for the bin. We estimate the slope by regressing the binned tax roll rank on the binned market value rank. The gray line is the 45-degree identity line. Sample: properties in census sections for which we have market values ($N = 1,123$ in the Discretion arm and $N = 1,166$ in the Algorithm arm).

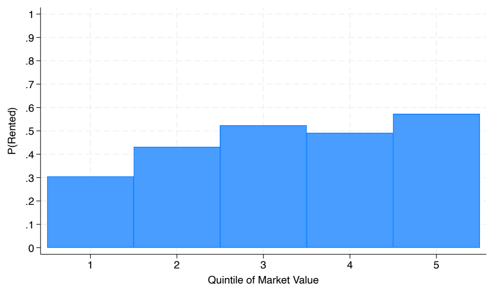
FIGURE A8
CHARACTERISTICS OF CENSUS VISITS BY QUINTILE OF MARKET VALUE



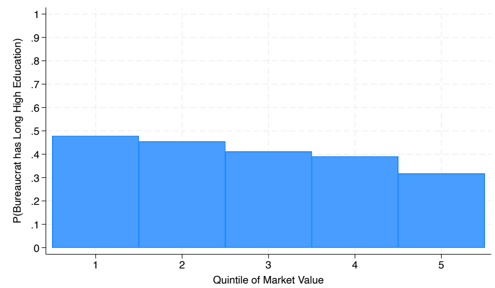
(A) PROBABILITY OF PROPERTY BEING COVERED IN CENSUS



(B) PROBABILITY OF OWNER BEING MET



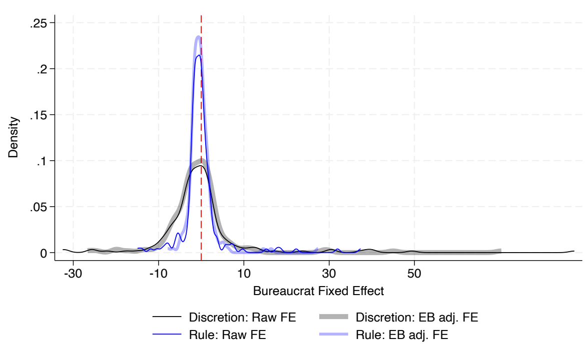
(C) PROBABILITY OF PROPERTY BEING RENTED



(D) PROBABILITY OF BUREAUCRAT HAVING LONG HIGHER EDUCATION

Notes: This Figure reports descriptive statistics on property visits during the census by quintile of market value, where market values are obtained from licensed assessors. The variables used in Panels (A), (B), (C) are from the property tax census data, Panel (D) additionally uses a variable from the bureaucrat survey. Sample: properties in census sections for which we have market values ($N = 2,289$).

FIGURE A9
DISTRIBUTION OF BUREAUCRATS FIXED EFFECTS



Notes: This Figure shows the distribution of bureaucrat fixed effects. The α_b are estimated in specification 3 which we run separately for each arm: $|Gap|_{ijb} = \alpha_b + Val_j + \epsilon_{ijb}$. $|Gap|_{ijb}$ is the absolute value of the tax base gap, measured as tax roll value minus market value for property i of section j covered by bureaucrat b. The market value is the value obtained from licensed assessors. Tax roll value is the value from the census that ends up on the tax roll. We then apply an empirical Bayes adjustment to recover $\alpha_{b,EB}$. We plot the kernel density estimate of the distribution of α_b and $\alpha_{b,EB}$ under discretion (in gray and black) and under the rule-based method (in light blue and blue).

B Additional Tables

TABLE A1
VALUATIONS BY LICENSED REAL ESTATE ASSESSORS

	All	Low value	High value
Panel A: Assessors' field work			
N sections	193	92	101
- per assessor	24.1	12.8	13.4
Info. from Office (%)	97.4	97.8	97.0
Info. from Agencies (%)	55.4	52.2	58.4
Info. from Occupants (%)	67.4	67.4	67.3
Panel B: Correlations with census and survey (property level)			
Census (fully rented)	0.74 (N=217)	0.48 (N=70)	0.66 (N=145)
Owner endline survey (rented)	0.69 (N=275)	0.47 (N=97)	0.58 (N=178)
Census (fully rented, contract)	0.64 (N=49)	0.59 (N=13)	0.30 (N=35)
Tenant survey	0.61 (N=390)	0.42 (N=144)	0.40 (N=246)
Owner baseline survey (rented)	0.62 (N=394)	0.41 (N=160)	0.46 (N=234)
Panel C: Correlations with real estate agency listings (section or municipality level)			
Share of online values within assessor bounds for a given section	0.96 (N=84)		
Kolmogorov–Smirnov distance to agency values by m2 in municipality	0.46 (N=16)		
Share of agency values per m2 within assessor bounds for a given municipality	0.99 (N=2020)		
Kolmogorov–Smirnov distance to agency values by room in municipality	0.38 (N=16)		
Share of agency values per room within assessor bounds for a given municipality	0.99 (N=11847)		

Notes: This Table shows descriptive statistics on the work of the licensed real estate assessors. The first column refers to the full sample, the second (respectively third) column shows results for *Low value* (resp. *High value*) sections - depending on whether the mean section-level value is below or above median. In Panel (A), the variables *Info. from [Office][Agencies][Occupants]* indicate the share of sections for which assessors report having used information from their own Office, from real estate Agencies, from Occupants met during field work. Panel (B) reports correlations at the property level between assessor values and values from other sources. *Census* refers to the property tax census, *rented* indicates that at least one room is rented and *fully rented* indicates that the whole property is rented. The value from the *Tenant survey* is an extrapolation of tenant-reported rent using the property's total number of rooms, as explained in Appendix F.2. *Owner endline* and *Tenant survey* observations are restricted to pure control areas where the census did not take place. Panel (C) reports comparisons with values from real estate agency listings. These can only be assigned to a given section ($N = 84$), or to a given municipality ($N = 2,020$ for which we have the built area measurement and $N = 11,847$ for which we have the number of rooms, spanning 16 municipalities). We provide details on the work of the assessors in Appendix D.1, on survey variables in Appendix F, and on the values from agency listings in Appendix D.3.

TABLE A2
PROPERTY VALUATION ALGORITHMS: COEFFICIENTS

Ln(BuiltArea)	0.56
Floors	.179
Residential	
Commercial	.242
Mixed	.127
Doors and Windows	
Very Good	.118
Average	
Bad	-.218
Landscape	-.078
Architecture	.03
Garage	
Simple	.067
Double	.138
None	
Balcony	.18
On Main Road	.023
Near Main Road	.016
Off Main Road	
Road Type	
Tarmac	.016
Pavements	.037
Gravel	.105
Sand	
None	-.031
Sidewalk	.055
Angle	.124
Street Lights	0

(A) RULE-BASED (1/2)

Fence Type	
None	
Metal	-.216
Wall	.028
Wall w. wrought iron	.017
Fence State	
Very Good	.045
Average	
Bad	-.079
Cement	.198
Cladding Type	
Wis	-.057
Plain	-.12
Paint	
Tiles	.013
Stone	.072
None	-.226
Cladding State	
Very Good	.064
Average	
Bad	-.117
Tiles	.05
Constant	11.651
Section FEs	
N zero	19/193
mean	.139
max	2.341
sd	.517

(B) RULE-BASED (2/2)

Ln(BuiltArea)	0.56
Floors	.179
Constant	11.651
Section FEs	
N zero	15/193
mean	.161
max	2.204

(C) PURE RULE

Notes: This Table reports the coefficients from the property valuation algorithms presented in Section 4.2. Panels (A) and (B) report results for the rule-based algorithm and Panel (C) for the pure rule. The algorithms are calibrated on the sample of market values from licensed assessors ($N = 4,916$) following specification 1, using an elastic-net regression and cross-validation. More details are available in Appendix E and performance statistics are shown in Table A3.

TABLE A3
PROPERTY VALUATION ALGORITHMS: PERFORMANCE

	Calibration sample	Algorithm	Location FE	R2	Adjusted R2	RMSE	MAPE	Freddie Mac 30%
(1)	Market values	Rule-based	Section	0.91	0.90	0.36	30.6	62.5
(2)	Market values	Pure rule	Section	0.87	0.86	0.43	34.8	60.9
(3)	Market values excl. half rule arm	Rule-based	Section	0.90	0.90	0.37	28.7	67.4
(4)	Market values excl. half rule arm	Pure rule	Section	0.87	0.86	0.43	33.1	60.3
(5)	Market values excl. census areas	Rule-based	Strata	0.88	0.87	0.43	48.3	48.2
(6)	Market values excl. census areas	Pure rule	Strata	0.83	0.82	0.51	56.8	42.4
(7)	Owner baseline survey	Rule-based	Section	0.44	0.33	0.63	46.6	39.1
(8)	Owner baseline survey	Pure rule	Section	0.43	0.31	0.64	46.7	44.4
(9)	Owner endline survey excl. census areas	Pure rule	Strata	0.60	0.58	0.55	51.1	45.2
(10)	Tenant endline survey excl. census areas	Pure rule	Strata	0.46	0.41	0.81	77.4	37.2
(11)	Market values excl. missing covariates	Rule-based	Section	0.90	0.90	0.37	30.7	64.8
(12)	Market values	Pure rule no height	Section	0.63	0.61	0.72	67.0	39.1

Notes: This Table reports performance statistics for the different versions of the property valuation algorithms presented in Section 4.2. *Algorithm* indicates whether the model is rule-based (with all property characteristics) or a pure rule (only built area and location); *Location FE* indicates the type of location fixed effect included in the model (where section is more granular than strata); *RMSE* is the root mean squared error and *MAPE* is the mean absolute percentage error – computed using monetary amounts (not the $ln()$ transformation) as the outcome variable, to be more conservative. *Freddie Mac 30%* refers to the share of predictions that fall within 30% of the true value. R^2, *Adjusted* R^2 and *RMSE* are estimated out-of-sample in the cross-validation procedure; *MAPE* and *Freddie Mac 30%* are computed on a test sample not used in the calibration. Appendix E provides details on the calibration of the algorithms used in the main analysis (rows (1) and (2)) and Appendix G on the different additional versions (rows (3) to (12)).

TABLE A4
ASSESSMENT RATIOS BY QUINTILE UNDER EACH VALUATION METHOD

	Q1 (1)	Q2 (2)	Q3 (3)	Q4 (4)	Q5 (5)
Panel A: Discretion					
Median Ass. Ratio	0.83	0.72	0.50	0.44	0.23
$\hat{\beta}_n$	Ref.	-0.10	-0.33	-0.40	-0.61
P-value		0.31	0.00	0.00	0.00
P-value $\hat{\beta}_n \neq \hat{\beta}_{n+1}$		0.00	0.27	0.00	
Panel B: Rule-based					
Median Ass. Ratio	1.23	0.97	0.89	0.79	0.60
$\hat{\beta}_n$	Ref.	-0.27	-0.34	-0.45	-0.64
P-value		0.00	0.00	0.00	0.00
P-value $\hat{\beta}_n \neq \hat{\beta}_{n+1}$		0.10	0.03	0.00	
Panel C: Pure Rule					
Median Ass. Ratio	1.24	0.99	0.96	0.94	0.87
$\hat{\beta}_n$	Ref.	-0.25	-0.27	-0.30	-0.37
P-value		0.00	0.00	0.00	0.00
P-value $\hat{\beta}_n \neq \hat{\beta}_{n+1}$		0.57	0.44	0.06	

Notes: This Table reports assessment ratios by quintile under each method, discretion in Panel (A), rule-based method in Panel (B) and pure rule in Panel (C). The first row of each Panel shows the median assessment ratio, computed as a property's tax roll value divided by its market value obtained from licensed assessors. We test whether the median assessment ratio changes significantly across quintiles of market values using the following quantile regression at the median: $AR_i = \sum_{n=1}^{5} \beta_n Q(n)_i + \epsilon_i$ where AR_i is the assessment ratio for property i, $Q(n)$ an indicator for quintile n. Standard errors are clustered at the section level. The second row of each Panel reports the $\hat{\beta}_n$ coefficients. Below, we report the P-value for a test of whether $\hat{\beta}_n$ is significantly different from the reference $Q(1)$. The final row of each Panel reports the P-value for a test of whether the coefficients for two subsequent quintiles are significantly different from each other. Sample: properties in census sections for which we have market values, $N = 1,123$ in the Discretion arm and $N = 1,166$ in the Algorithm arm.

TABLE A5
EFFECTS ON EXTENSIVE MARGIN OUTCOMES

Dependent Variable	Properties per day (1)	Covered (2)	Eligible (3)	Valued (4)	Rented (5)	Main Res. (6)	Owner Met (7)	Conflict (8)
Panel A: Full Sample								
$\hat{\beta}_{Discretion}$	1.156	0.002	-0.031	-0.228***	-0.006	-0.051**	0.002	-0.014*
	(1.306)	(0.019)	(0.021)	(0.023)	(0.017)	(0.021)	(0.014)	(0.007)
Strata FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	2309	41609	41609	41609	41609	41609	41609	41609
N sections	94	94	94	94	94	94	94	94
Adj R2	0.05	0.03	0.06	0.12	0.05	0.04	0.02	0.02
Mean of dep.	16.56	0.92	0.79	0.71	0.38	0.46	0.24	0.05
Panel B: Sample with Market Values								
$\hat{\beta}_{Discretion}$	-0.008	0.027	0.031	-0.178***	0.018	-0.020	0.002	-0.009
	(0.087)	(0.021)	(0.022)	(0.030)	(0.023)	(0.028)	(0.025)	(0.010)
Strata FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	1103	2408	2408	2289	2289	2289	2289	2289
N sections	94	94	94	94	94	94	94	94
Adj R2	0.03	0.03	0.03	0.10	0.05	0.04	0.02	0.02
Mean of dep.	2.03	0.93	0.88	0.82	0.43	0.54	0.26	0.05

Notes: This Table compares extensive margin outcomes of the property tax census for Discretion versus Algorithm sections. Panel (A) shows results for the whole sample while Panel (B) is restricted to properties for which we have market values. We report coefficient $\hat{\beta}_{Discretion}$ for an indicator for Discretion sections. In column (1) the outcome is the number of properties covered in a day and observations are at the day X section level. In columns (2) to (8) regressions are at the property level and the outcome variables are indicators taking value one, respectively, if: the property is covered in the census, it is classified as eligible for the tax, there is a positive property value, it is classified as rented (at least in part), it is classified as main residence, the bureaucrat reports having met the owner, the bureaucrat reports that there were tensions or conflict (scraped from text comments). In all regressions, we control for strata fixed effects and errors are clustered at the section level. Source: data from the property tax census.

TABLE A6
EFFECTS ON ACCURACY: TAX BASE GAP ROBUSTNESS RESULTS (1/2)

	Gap	Gap (median)	∣Gap∣
	All amounts in millions of F		
	(1)	(2)	(3)
Panel A: Rule-based excluding half of Algorithm arm (N = 1706)			
$\hat{\beta}_{Discretion}$	-4.72^{***}	-1.84^{***}	3.93^{***}
	(1.22)	(0.40)	(1.30)
Mean in Rule-based	−2.22 (11.88)	−0.28	4.59 (11.18)
Panel B: Pure Rule excluding half of Algorithm arm (N = 1706)			
$\hat{\beta}_{Discretion}$	-4.94^{***}	-1.85^{***}	4.69^{***}
	(1.15)	(0.41)	(1.22)
Mean in Pure Rule	−1.84 (10.14)	−0.15	3.68 (9.62)
Panel C: Rule-based excluding census areas (N = 2289)			
$\hat{\beta}_{Discretion}$	-4.97^{**}	-2.52^{***}	2.42
	(2.07)	(0.59)	(1.79)
Mean in Rule-based	−2.16 (17.32)	0.24	6.24 (16.30)
Panel D: Pure Rule excluding census areas (N = 2289)			
$\hat{\beta}_{Discretion}$	-6.06^{***}	-2.70^{***}	2.46
	(2.19)	(0.73)	(1.60)
Mean in Pure Rule	−1.07 (16.66)	0.30	6.15 (15.52)
Panel E: Rule-based calibrated on owner baseline survey (N = 2289)			
$\hat{\beta}_{Discretion}$	−0.18	−0.54	1.22
	(1.92)	(0.33)	(1.91)
Mean in Rule-based	−6.87 (18.09)	−1.92	7.19 (17.96)
Panel F: Pure Rule calibrated on owner baseline survey (N = 2289)			
$\hat{\beta}_{Discretion}$	−0.28	−0.53	1.27
	(1.93)	(0.33)	(1.92)
Mean in Pure Rule	−6.78 (18.03)	−1.94	7.16 (17.88)
Panel G: Pure Rule calibrated on owner endline survey (N = 2289)			
$\hat{\beta}_{Discretion}$	−2.40	-1.64^{***}	2.48
	(1.87)	(0.49)	(1.83)
Mean in Pure Rule	−4.72 (17.25)	−0.51	6.02 (16.84)
Panel H: Pure Rule calibrated on tenant endline survey (N = 2289)			
$\hat{\beta}_{Discretion}$	−2.94	-2.04^{***}	2.74
	(1.81)	(0.37)	(1.79)
Mean in Pure Rule	−4.09 (17.02)	−0.01	5.86 (16.50)
Panel I: Rule-based excluding missing covariates (N = 2289)			
$\hat{\beta}_{Discretion}$	-4.53^{***}	-1.76^{***}	3.76^{***}
	(1.27)	(0.40)	(1.38)
Mean in Rule-based	−2.43 (12.86)	−0.33	4.77 (12.18)
Panel J: Pure Rule no height (N = 2289)			
$\hat{\beta}_{Discretion}$	-3.84^{***}	-1.64^{***}	2.71^{***}
	(1.43)	(0.37)	(0.84)
Mean in Pure Rule	−3.24 (15.53)	−0.41	3.69 (5.92)

Notes: This Table compares the accuracy of discretionary and algorithmic valuations, using alternative specifications as complements to the main results shown in Table 2. In Appendix G, we provide details on the values used in each Panel.

TABLE A7
EFFECTS ON ACCURACY: TAX BASE GAP ROBUSTNESS RESULTS (2/2)

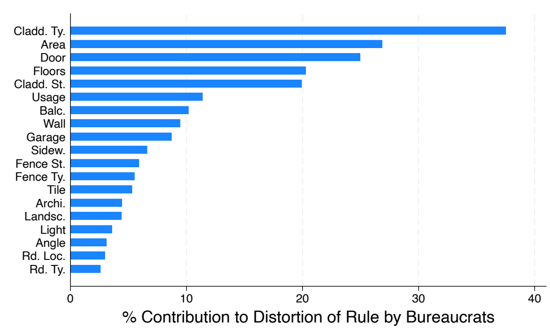
	Gap	Gap (median)	|Gap|
	All amounts in millions of F		
	(1)	(2)	(3)
Panel A: Rule-based calibration inputs vs Rule-based Bureaucrat inputs (N = 2331)			
$\hat{\beta}_{Rule\_Bur}$	-1.90^{***}	-0.33^{***}	1.94^{***}
	(0.49)	(0.12)	(0.40)
Mean in Rule-based	−0.87 (8.18)	0.01	2.74 (7.76)
Panel B: Pure Rule versus Rule-based Bureaucrat inputs (N = 2332)			
$\hat{\beta}_{Rule\_Bur}$	-0.88^{***}	−0.15	1.03^{***}
	(0.32)	(0.10)	(0.26)
Mean in Pure Rule	−1.89 (10.41)	−0.20	3.65 (9.93)
Panel C: Rule-based with Bureaucrat fixed effects (N = 2117)			
$\hat{\beta}_{Discretion}$	-5.03^{***}	-2.15^{***}	4.47^{***}
	(1.26)	(0.49)	(1.25)
Mean in Rule-based	−1.73 (10.77)	−0.18	3.81 (10.21)
Panel D: Pure Rule with Bureaucrat fixed effects (N = 2117)			
$\hat{\beta}_{Discretion}$	-5.94^{***}	-2.22^{***}	5.60^{***}
	(1.21)	(0.51)	(1.12)
Mean in Pure Rule	−0.79 (7.88)	−0.05	2.72 (7.44)
Panel E: Rule-based intensive margin (N = 1884)			
$\hat{\beta}_{Discretion}$	-2.52^{**}	-0.92^{***}	2.60^{**}
	(1.06)	(0.26)	(1.13)
Mean in Rule-based	−1.73 (10.77)	−0.18	3.81 (10.21)
Panel F: Rule-based Lee bounds (N = 2289)			
Lower bound	−5.37		2.14
Upper bound	−1.22		5.61
CI for $\hat{\beta}_{Discretion}$	[−6.29;−0.12]		[1.08;6.45]
Panel G: Pure Rule intensive margin (N = 1884)			
$\hat{\beta}_{Discretion}$	-3.33^{***}	-1.06^{***}	3.68^{***}
	(0.96)	(0.28)	(0.98)
Mean in Pure Rule	−0.79 (7.88)	−0.05	2.72 (7.44)
Panel H: Pure Rule Lee bounds (N = 2289)			
Lower bound	−5.79		3.19
Upper bound	−3.22		5.29
CI for $\hat{\beta}_{Discretion}$	[−6.69;−2.21]		[2.25;6.11]
Panel I: Rule-based using assessor bounds (N = 2278)			
$\hat{\beta}_{Discretion}$	-3.57^{***}	-1.25^{***}	3.40^{***}
	(1.09)	(0.32)	(1.16)
Mean in Rule-based	−2.14 (10.86)	0.00	3.47 (10.51)
Panel J: Pure Rule using assessor bounds (N = 2278)			
$\hat{\beta}_{Discretion}$	-4.31^{***}	-1.50^{***}	4.31^{***}
	(0.98)	(0.37)	(1.02)
Mean in Pure Rule	−1.33 (8.60)	0.00	2.49 (8.34)

Notes: This Table compares the accuracy of discretionary and algorithmic valuations, using alternative specifications as complements to the main results shown in Table 2. In Appendix G, we provide details on the values used in each Panel.

TABLE A8
PROPERTY CHARACTERISTICS REPORTED BY BUREAUCRATS VERSUS ASSESSORS

Characteristic	Share Identical
Angle (dummy)	0.92
Landscape (dummy)	0.81
Street lights (dummy)	0.79
Wall (dummy)	0.79
Balcony (dummy)	0.78
Usage	0.75
Road type	0.71
Floors	0.71
Correcting for Ground floor	0.87
Architecture (dummy)	0.66
Sidewalk (dummy)	0.65
Tiles (dummy)	0.65
Garage	0.62
Fence type	0.55
Fence state	0.51
Road (main)	0.50
Quality doors and windows	0.49
Cladding state	0.48
Cladding type	0.14

(A) SHARE OF MATCHING CHARACTERISTICS



(B) CONTRIBUTION TO AGGREGATE DIFFERENCE

Notes: This Table compares property characteristics reported by the bureaucrats in the rule-based valuation with those reported by the assessors that were used to calibrate the algorithm. In Panel (A), we report the share of observations for which each characteristic matches. For floors, we additionally report the share of matches after correcting for cases where the bureaucrat reported one floor more than the assessor since counting the ground floor was a common mistake. In Panel (B), we report the relative contribution of each characteristic to the aggregate differences in values between the rule implemented by bureaucrats and the rule with assessor inputs. The contribution of characteristic X_k is calculated as the sum of the absolute difference in predicted values due to this characteristic: $\sum_i |\theta_k X_{k,RuleBur,i} - \theta_k X_{k,RuleCalibration,i}|$ where the X_k are the 18 observable characteristics, subscript $RuleBur$ indicates the value taken by X_k when entered by bureaucrat and $RuleCalibration$ the value taken by X_k when entered by assessors for property i, θ_k is the coefficient for X_k in the algorithm (shown in Table A2). Sample: properties in census sections for which we have market values, restricted to the Algorithm arm ($N = 1,166$).

TABLE A9
BUREAUCRAT FIXED EFFECTS: DESCRIPTIVE STATISTICS

	Discretion (1)	Rule-based (2)
N# observations	1,054	1,063
N# Bur FE	197	190
Mean of outcome	7.97	3.81
Variance of outcome	216.70	104.33
R2 without Bur FE	0.40	0.28
R2 with Bur FE	0.52	0.37
P-value of F test on Bur FE	0.00	0.00
Var(Bur FE)	129.87	29.57
Var(Shrinked Bur FE)	73.40	9.91
Share Variance	**0.34**	**0.09**

Notes: This Table summarizes results from the estimation of bureaucrat fixed effects α_b presented in Section 5.2 using specification 3. The outcome $|Gap|_{ijb}$ is the absolute value of the tax base gap measured as tax roll value minus market value for property i of section j covered by bureaucrat b, in millions of F and winsorized at the 1% level. We apply an empirical Bayes adjustment to recover the shrinked fixed effects $\alpha_{b,EB}$. The *P-value of F-test on Bur FE* confirms that the α_b are jointly significant. *Share Variance* refers to the share of variance in the tax base gap explained by bureaucrat fixed effects.

TABLE A10
BUREAUCRAT FIXED EFFECTS: ROBUSTNESS

	(1) No shrinkage	(2) Control for high value	(3) Top bureaucrats: $\alpha_{b,EB} < 0$
Bureaucrats fixed effects			
N# observations	1,054	1,054	1,054
N# Bur FE	197	197	197
Share variance (Discretion)	0.60	0.42	0.34
Share variance (Rule-based)	0.28	0.16	0.09
Correlate top bureaucrat			
N# Top Bur	121	144	116
Coef. for Long Higher Education	0.23**	0.13**	0.21**
	(0.07)	(0.06)	(0.07)

Notes: This Table reports additional results for bureaucrat fixed effects, discussed in Sections 5.2, 6.5 and Table A9. In column (1) we report results for the estimated α_b with no shrinkage. In column (2), the fixed effects are recovered from an alternative version of specification 3 in which we replace the section-level control Val_j by an indicator for high-value properties. In column (3), we define top bureaucrats as those with a negative fixed effect instead of applying the procedure shown in Figure 5. The last row reports the coefficient obtained when regressing an indicator for top bureaucrat on an indicator for long higher education with the standard error in parentheses. We provide details in Appendix G.2.

TABLE A11
LAB-IN-THE-FIELD: BUREAUCRATS' LACK OF KNOWLEDGE AND BIASES

Dependent Variable	Ass. Ratio (1)	Ass. Ratio (2)	Ln(Value) (3)	Ln(Value) (4)
High value	-1.419***	-1.491***		1.232***
	(0.112)	(0.109)		(0.091)
Information		-0.085		
		(0.150)		
Information X High value		0.127		
		(0.150)		
Retired owner			-0.378***	-0.241***
			(0.091)	(0.091)
Retired X High value				0.090
				(0.133)
Strata FE	No	Yes	Yes	Yes
Bureaucrat FE	Yes	No	No	No
N	280	280	280	280
R2	0.83	0.62	0.07	0.57
Adj R2	0.60	0.61	0.01	0.54
Mean in reference	1.74	1.78	12.84	12.08

Notes: This Table shows results from a lab-in-the-field valuation exercise with bureaucrats, presented in Sections 6.1 and 6.5 and Figure 3. Column (1) displays results from the following regression: $AR_{ib} = \alpha_b + \beta \textit{High value}_{ib} + \epsilon_{ib}$, where AR_{ib} is the assessment ratio (bureaucrat b's value over market value) for property i, $\textit{High value}_{ib}$ is an indicator for the high value property and α_b a bureaucrat fixed effect. In column (2), we add the indicator $\textit{Information}_b$ taking value one if the bureaucrat was randomly assigned to receive the information on the distribution of market values (shown in Figure A14), and its interaction with $\textit{High value}_{ib}$. Column (3) shows results from the following regression: $Ln(Value_{ib}) = \alpha + \beta_1 Retired_{ib} + \epsilon_{ib}$. $Retired_{ib}$ takes value one if bureaucrat b was randomly assigned to receive the information that the owner of property i was retired (versus employed). Observations are at the bureaucrat X property level. *,** and *** indicate statistical significance at the 10, 5 and 1% level respectively. Standard errors are clustered at the bureaucrat level. See Appendix Section F.1 for details on the survey module and variables. Sample: 166 bureaucrats surveyed at endline.

TABLE A12
BUREAUCRATS' IMPLICIT ALGORITHM

	Assessor value (1)	Bur. value (2)	Bur. value Owner status (3)	Bur. value Owner status, income and social char. (4)	Bur. value Owner status, income and social char. (5)
Property characteristics					
Ln(BuiltArea)	0.52	0.36	0.36	0.37	0.34
Floors	0.24	0.18	0.16	0.13	0.17
N# positive section FE	28/47	32/47	31/47	12/46	29/47
SD section FE	0.45	0.35	0.36	0.27	0.35
Owner status					
Male			(.)	(.)	(.)
Female			0.000	0.000	0.000
Deceased			−0.100	−0.172	−0.086
Unknown			0.000	0.000	0.047
Multiple			−0.112	0.000	−0.111
Main residence			−0.182	0.000	−0.170
Pension			−0.020	−0.031	0.000
Owner X Bureaucrat					
Same commune				0.000	0.095
Same ethnic group				−0.009	−0.075
Income					
Liquidity index				0.000	0.009
Income index				0.032	0.018
Social characteristics					
Owner retired				0.000	0.000
Foreign				0.000	0.000
Political connect. index				0.000	0.000
Performance					
R2	0.87	0.56	0.58	0.47	0.58
Adj R2	0.85	0.49	0.50	0.17	0.49
RMSE	0.40	0.69	0.68	0.69	0.68
MAPE	36.30	42.90	43.00	28.40	44.80
Freddie Mac 30%	67.60	43.20	40.50	55.60	37.80
N	821	821	821	339	821

Notes: This Table reports results from the calibration of bureaucrats' implicit algorithm. We follow specification 1 and use an elastic-net regression and cross-validation, on the sample of properties for which we have bureaucrats' discretionary values and property characteristics by the assessors. In column (1), the predicted outcome is the assessor value. In columns (2) to (5), the predicted outcome is bureaucrats' discretionary value. In all columns, the model includes the property characteristics listed in Panels (A) and (B) of Table A2. In column (3), we add information on owner status collected by bureaucrats. In columns (4) and (5), we add information on owner status, income and social characteristics as well as owner X bureaucrat similarities, relying on variables from the property owner endline survey. In column (5), we impute missing values using the mean and add indicators for whether a value is missing. All variables are binary, except the indices that have a mean of 0 and a standard deviation of 1. *Liquidity index* captures the household's liquidity constraints, *Income index* captures household income, *Political connect. index* captures the household's political connections. R^2, *Adjusted* R^2 and *RMSE* are estimated out-of-sample in the cross-validation procedure; *MAPE* and *Freddie Mac 30%* are computed on a test sample not used in the calibration. See Appendix E for details on the algorithm calibration and Appendix F.2 for details on the survey variables. Sample: properties in census sections for which we have market values, restricted to the Discretion arm.

TABLE A13
DOES LOCAL INFORMATION IMPROVE DISCRETIONARY VALUATIONS?

	Gap	Gap (median)	$\mid$Gap$\mid$
	All amounts in millions of F		
	(1)	(2)	(3)
Panel A: Full sample			
Panel A.1: Low-value properties			
$\hat{\beta}_{Discretion}$	-0.52^{***}	-0.54^{***}	0.42^{***}
	(0.18)	(0.11)	(0.14)
Mean in Rule-based	0.50 (1.76)	0.20	1.07 (1.49)
N	884	884	884
Bur. FE	No	No	No
Panel A.2: High-value properties			
$\hat{\beta}_{Discretion}$	-6.18^{***}	-4.11^{***}	5.60^{***}
	(1.54)	(0.55)	(1.62)
Mean in Rule-based	−3.75 (14.49)	−1.67	6.31 (13.57)
N	1,233	1,233	1,233
Bur. FE	No	No	No
Panel B: Owner Met			
Panel B.1: All properties			
$\hat{\beta}_{Discretion}$	-2.62^{**}	-2.02^{***}	1.37
	(1.09)	(0.78)	(1.28)
Mean in Rule-based	−2.12 (13.68)	−0.13 (13.68)	4.01 (13.25)
N	597	597	597
Bur. FE	Yes	Yes	Yes
Panel B.2: Low-value properties			
$\hat{\beta}_{Discretion}$	0.34	-0.06^{***}	0.98^{*}
	(0.52)	(0.00)	(0.49)
Mean in Rule-based	0.40 (1.39)	0.68 (1.04)	1.00 (1.04)
Bur. FE	Yes	Yes	Yes
Panel B.3: High-value properties			
$\hat{\beta}_{Discretion}$	-5.99^{***}	-3.51^{*}	4.19^{***}
	(1.53)	(2.08)	(1.47)
Mean in Rule-based	−4.27 (18.34)	−1.29 (19.67)	6.58 (17.64)
Bur. FE	Yes	Yes	Yes
Panel C: Rented			
Panel C.1: All properties			
$\hat{\beta}_{Discretion}$	-4.96^{**}	−1.03	4.72^{**}
	(2.18)	(1.01)	(1.86)
Mean in Rule-based	−1.89 (11.19)	−0.21 (11.19)	4.34 (10.49)
N	976	976	976
Bur. FE	Yes	Yes	Yes
Panel C.2: Low-value properties			
$\hat{\beta}_{Discretion}$	1.09	0.45^{***}	1.26^{*}
	(0.79)	(0.00)	(0.74)
Mean in Rule-based	0.80 (2.21)	0.34 (2.21)	1.29 (1.96)
Bur. FE	Yes	Yes	Yes
Panel C.3: High-value properties			
$\hat{\beta}_{Discretion}$	-7.54^{**}	−2.18	6.64^{**}
	(3.68)	(1.43)	(2.99)
Mean in Rule-based	−3.48 (13.79)	−1.48 (13.79)	6.14 (12.82)
Bur. FE	Yes	Yes	Yes
Panel D: Bureaucrats' estimate			
$\hat{\beta}_{Discretion}$	-5.11^{***}	-2.76^{**}	4.26^{***}
Mean in Rule-based	−1.73 (10.77)	−0.18 (10.77)	3.81 (10.21)
N	1,195	1,195	1,195
Bur. FE	Yes	Yes	Yes

Notes: This Table reports results from regression 2 comparing the accuracy of discretionary and rule-based valuations, restricting the sample to specific subsets of properties. In Panel (A) we split the sample into properties of quintiles 1 and 2 (Panel (A.1)) and 3-4-5 (Panel (A.2)) of market values. In Panel (B), observations are restricted to cases where the bureaucrat met the owner. In Panel (C), observations are restricted to cases where the property is rented at least in part. In Panel (D), observations from the Discretion arm are restricted to cases where the bureaucrat indicated having made the estimation without relying on any information from occupants. All amounts are in millions of F and winsorized at the 1% level. *,** and *** indicate statistical significance at the 10, 5 and 1% level respectively. Errors are clustered at the section level.

TABLE A14
TESTING WHETHER BUREAUCRATS LEARN IN THE FIELD

	Discretion						Rule-based			
	Gap	|Gap|	Gap	|Gap|	Gap	|Gap|	Gap	|Gap|	Gap	|Gap|
			All amounts in million F					*All amounts in million F*		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Numb. properties	-0.873***	0.770**					0.122	-0.388		
	(0.307)	(0.299)					(0.222)	(0.285)		
(Numb. properties)^2	0.018	-0.021*					-0.000	0.009		
	(0.012)	(0.012)					(0.005)	(0.007)		
Numb. days			-1.379***	1.210**					0.152	-0.486
			(0.471)	(0.452)					(0.293)	(0.369)
(Numb. days)^2			0.045**	-0.048**					0.001	0.013
			(0.022)	(0.022)					(0.007)	(0.010)
Exposed to rule					-4.231	5.055				
					(3.176)	(3.145)				
Section control for Market Value	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bureaucrat FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	1054	1054	1054	1054	1054	1054	1063	1063	1063	1063
Mean	-6.14	7.99	-6.14	7.99	-6.14	7.99	-1.62	3.81	-1.62	3.81
R2	0.48	0.51	0.48	0.51	0.47	0.51	0.26	0.42	0.26	0.42
Adj R2	0.35	0.39	0.35	0.39	0.34	0.39	0.09	0.29	0.09	0.29

Notes: In this Table, we assess whether there is any learning by bureaucrats over the course of the property tax census. Columns (1) to (6) show results for discretionary valuations while columns (7) to (10) show results for rule-based valuations. The outcome is the tax base gap defined as tax roll value minus market value, where market value is the value obtained from licensed assessors. In even-numbered columns the outcome is the absolute value of the gap. All amounts are in millions of F and winsorized at the 1% level. The regressors of interest are the number of properties visited by a given bureaucrat and its squared value (columns (1), (2), (7) and (8)); the number of days worked by a given bureaucrat and its squared value (columns (3), (4), (9) and (10)), and an indicator taking value once a bureaucrat has been exposed to the rule (columns (5) and (6)). We control for bureaucrat fixed effects and a section-level decile of market values. Errors are clustered at the section level. *,** and *** indicate statistical significance at the 10, 5 and 1% level respectively. Sample: properties in census sections for which we have market values ($N = 2,289$).

TABLE A15
ALGORITHM WITH BUREAUCRAT INPUTS

	Assessor characteristics only	With Bur. Value	With Bur. Value Components	Assessor characteristics only	With Ln() Bur. Value	With Ln() Bur. Value Components
	(1)	(2)	(3)	(4)	(5)	(6)
Coefficients						
Ln(BuiltArea)	0.579	0.576	0.574	0.543	0.523	0.413
Floors	0.204	0.198	0.204	0.218	0.177	0.179
Bur. Value Total		$-1.4 \cdot 10^{-9}$			$7.7 \cdot 10^{-2}$	
Bur. Value Rented			$-1.6 \cdot 10^{-9}$			$1.1 \cdot 10^{-1}$
Bur. Value Owner			$1.5 \cdot 10^{-10}$			$2.9 \cdot 10^{-2}$
Rented		0.068	0.080		0.063	
Performance						
R2	0.88	0.88	0.88	0.88	0.87	0.84
Adj R2	0.86	0.87	0.87	0.86	0.85	0.65
RMSE	0.39	0.39	0.40	0.38	0.37	0.42
MAPE	31.80	33.40	30.90	27.30	28.60	40.80
Freddie Mac 30%	58.40	60.80	63.20	68.90	68.90	50.00
N	1210	1210	1210	821	821	212

Notes: This Table reports results from the calibration of the algorithm, adding bureaucrat inputs as predictors. The algorithm is calibrated following specification 1, using an elastic-net regression and cross-validation. In column (1), we include the property characteristics listed in Panels (A) and (B) of Table A2. In column (2), we add bureaucrats' value for the whole property and an indicator taking value one if the bureaucrat reports the property as rented at least in part. In column (3), we include bureaucrats' values in two separate components, value for rent and value of owner-occupied parts. Columns (4)-(5)-(6) repeat the same specifications but restricting to observations for which bureaucrats entered positive values. R^2, *Adjusted* R^2 and *RMSE* are estimated out-of-sample in the cross-validation procedure; *MAPE* and *Freddie Mac 30%* are computed on a test sample not used in the calibration. More details on the algorithm calibration are available in Appendix E. Sample: properties in census sections for which we have market values, restricted to the Discretion arm.

TABLE A16
TOP BUREAUCRATS' CENSUS OUTCOMES

Dependent variable	Positive Value	Owner Met	Owner Details	Contract	Comment	Conflict	Bureaucrat Estimate	Supervisor Evaluation
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Top bureaucrat	0.060***	-0.006	-0.021	0.004	0.034	-0.015	0.023	-0.066
	(0.021)	(0.016)	(0.014)	(0.005)	(0.029)	(0.012)	(0.027)	(0.172)
Strata FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	19231	19231	19231	19231	19231	10541	10541	11212
R2	0.11	0.03	0.09	0.02	0.33	0.02	0.18	0.20
Adj. R2	0.11	0.03	0.08	0.01	0.33	0.02	0.18	0.20
Mean of dependent	0.65	0.26	0.35	0.03	0.55	0.09	0.18	0.09

Notes: This Table reports results from regressions of property-level outcomes on an indicator taking value one if the property was covered by a top bureaucrat. Top bureaucrats are defined in Section 6.5 and Figure 5, as those for which the difference in tax base gap between discretionary and rule-based valuation is not significant. Outcomes in columns (1)–(7) are binary variables, while *Supervisor Evaluation* (column (8)) is a standardized score (z-score). *Owner details* takes value one if the bureaucrat recovered the name and/or ID number of the owner. *Contract* takes value one if there is at least one rental contract reported for the property. *Comment* takes value one if the bureaucrat left a comment. *Conflict* and *Bureaucrat Estimate* are conditional on leaving a comment: the former takes value one if the bureaucrat mentioned tensions or conflict with occupants, the latter takes value one if the bureaucrat explicitly states that she made an estimation herself. *Supervisor Evaluation* is a general performance grade assigned by supervisors to each bureaucrat. *,** and *** indicate statistical significance at the 10, 5 and 1% level respectively. We control by strata fixed effects and errors are clustered at the section level. Sample: properties of the Discretion arm covered by the census ($N = 19,231$).

TABLE A17
COST-BENEFIT ANALYSIS

	N# properties	Cost *(million F)*	Assessed amounts *(million F)*	Ratio *Ass/Cost*	Paid amounts *(million F)*	Ratio *Paid/Cost*	Net gain per property *(F)*
(1) Assessors	5,025	15.1	4,993	331	–	–	–
(2) Discretion	16,651	60.1	7,989	133	508.7	8.5	26,938
(3) Rule-based	16,026	75.4	10,311	137	–	–	–
(4) Pure Rule	16,026	16.6	13,251	798	587.5	35.4	35,623
(5) Optimal Policy	16,026	28.4	13,144	463	–	–	–

Notes: This Table summarizes costs and benefits for different valuation methods. *N# properties* indicates the number of properties: those covered by assessors in row (1), number of eligible properties in the Discretion arm in row (2), number of eligible properties in the Algorithm arm in rows (3) to (5). We report costs of assessors' field work in row (1). Under discretion, costs include bureaucrats' and their supervisors' field work. Assessors' daily fee is 10 times that of bureaucrats, but this difference is dampened by the existence of high supervision costs for bureaucrat field work, and by the larger number of properties covered per day by the assessors. Rule-based costs additionally include algorithm calibration costs – assessor field work and GIS work. Algorithm calibration costs are the only costs for the pure rule. The costs under the optimal policy are algorithm calibration costs and discretion costs for properties predicted to be in the lowest quintile. *Net gain per property* is the property-level average of *Paid amount* minus *Cost*.

C Details on the Institutional Context and Program

C.1 Institutional Context

In Senegal, property taxation is managed by the national government. The national tax administration, *Direction Générale des Impôts et Domaines du Sénégal* or *DGID* is in charge of registrations, valuations, and the creation of tax bills. The national treasury is in charge of distribution, collecting payments and enforcement. Distribution is done physically door-to-door. Taxpayers must visit a treasury office to pay in person. Property tax revenues accrue to cities at the end of the year. The region includes five cities with eight tax offices: Dakar (Dakar Plateau, Grand Dakar, Dakar Liberté, Ngor Almadies, Parcelles Assainies tax offices), Guediawaye (Guediawaye tax office), Pikine and Keur Massar (Pikine tax office), Rufisque (Rufisque tax office).

In our paper, we use the term *property tax* to refer to the combination of the property tax *per se* and the garbage tax. Indeed, both are managed through the same tax bills and have the same base. Their exact denominations are *Contribution Foncière des Propriétés Bâties* or *CFPB* and *Taxe d'Enlèvement des Ordures Ménagères* or *TEOM*. The garbage tax has a rate of 3.6 percent. The property tax has a 5 percent rate, and if the property is the owner's main residence, there is a reduction of the tax base by 1.5 million F. The tax code also stipulates that owners who receive a formal state pension may be exempt from the property tax component but are still liable for the garbage tax – according to our survey estimates this situation applies to 20 percent of owners.

The discretionary valuation method which is the status quo before the program is called the *comparative method* in the tax code (meaning that bureaucrats should compare the property with similar ones for which they know values). The tax code also provides for a *cadastral valuation method* where experts from the cadastral division of the administration conduct in-depth visits and measurements to value a property. Due to its length and cost, in practice, these inspections barely ever occur. In theory, owners are supposed to come to the tax administration office once a year to declare their property's value, but only a very small minority respect this obligation. As such, the values on the pre-program tax roll originate either from self-declarations by owners, or from discretionary valuations by bureaucrats. A study of the administrative data between 2015 and 2023 shows that once a property was registered, its value was never updated within that time span.

Resource constraints are the main reason why property census operations were extremely rare before the digitization program. The staff available to conduct this field work is limited and also works on other taxes. Expanding the tax net through a property tax census was extremely time consuming before digitization: all the information was collected on paper and needed to be typed into the system once back into the office. The administration had no mapping of which properties were registered or not. The new digital tool allows to overcome these constraints.

C.2 Program

The program was part of broader efforts of the government to expand the tax net and leverage technology to increase its tax-to-GDP ratio. As an illustration, in 2021, the tax administration launched both its *Yaatal* (or "expansion") program, aiming to double the number of registered taxpayers, and a national property census project (*Recensement National des Propriétés Bâties*), with the objective for all properties of the national territory to be covered by property tax census operations in the near future. Two dimensions made this program a first-of-a-kind for the Senegalese government: the deployment of a new digital tool at scale, and the fact that this enables census operations to be comprehensive (to cover all properties) in treated areas.

We started working on the application in 2017, in collaboration with the administration and a private Senegalese company. It has a Web components (to assign tasks, visualize and validate information, monitor advancement on maps, create tax bills), and an android component allowing to conduct the census on tablets in the field with pre-loaded maps. The property identification system and mapping integrated in the software were already used by the tax administration and especially the cadaster department. However, this was the first time they were incorporated in a digital application and systematically used for taxation. Each city is divided in municipalities or *Communes*, further divided into cadastral sections. Properties are identified using a unique code *Numéro d'Identification Cadastrale* or *NICAD*, these identifiers and property boundaries are pre-loaded in the application.

The bureaucrats working on the census were hired through several channels: some had already done similar tasks for the administration in the past, some had been suggested by the municipalities, finally, many were recruited through an online job advertisement on a public employment platform. They received a four day training delivered by the tax administration, covering local public finance concepts, the utilization of the application, reading maps, property characteristics and valuation, and interactions with occupants.

In the field, bureaucrats were equipped with caps, shirts and badges showing their affiliation to the tax administration. They were paid a monthly fee and a bonus based on the number of properties they covered, and the share for which they recovered the name and identification details of the owner. These incentives were exactly the same across treatment arms. There was a supervisor for every fourteen bureaucrats on average.

The field operations first started in 2019, seven sections (less than three percent of the properties in our sample) were covered, before an interruption due to the Covid-19 pandemic. The operations resumed in June 2021 and lasted until April 2023.

D Market Values and Property Tax Census Data: Additional Details

D.1 Market Values by Assessors

The assessors we hired are licensed real estate assessors, affiliated to the Senegalese National Order of Experts (https://www.experts-ones.com/), a government-supervised professional body that licenses and regulates certified experts. Its members are officially recognized and highly qualified, comparable to chartered professional organizations in many countries. The usual job of these assessors is to provide certified market valuations of properties for insurance purposes, before a sale or a inheritance, in relation to construction projects, etc. We worked with eight assessors from four different offices, between June and August 2022. They are highly educated with typically a background in real estate or engineering. Their daily fee is ten times the daily fee paid to field bureaucrats hired by the government for the program.[91]

To design the data collection procedure, we relied on discussions with practitioners to build on methodologies used in more established property tax systems such as the United Kingdom and South Africa. Before starting the field work in a given area, the assessors were asked to gather location-specific information from their office as well as from real estate agencies and brokers they are in close contact with. Once in the field, they filled in a section-level form with their estimation of a standard property's market rental value for that section, and reported whether they had relied on external information to work in this section (we display these responses in Panel (A) of Table A1). Then, they went to see each property that had been sampled for valuation and assigned to them. All valuations were done from the outside to avoid biases due to non-response. We provided assessors with properties' ground built area pre-loaded in the questionnaire to assist in valuations. Assessors reported their best estimate of the property's market rental value, as well as lower and upper bounds.

The sampling of properties to be visited by assessors was done in way to allow partial overlap between assessor valuations and (i) the baseline property owner survey; (ii) properties covered by the census for which a rental contract is noted (this is very rare, 2.6 percent of census observations). The rationale for doing so is to be able to check correlations between different sources of rental values. We first draw 26 properties randomly in each section. If less than 13 were also included in the owner baseline survey, we do some replacements to reach 13/26 overlapping properties (or the maximum number possible if there are fewer baseline observations in the section). If less than 2 properties were also covered by the census and with a rental contract, we do some replacements to reach 2/26 of these (or the maximum number possible). Then, we add a random draw of replacement properties for each section. As a result, among the 5,806 properties sampled for assessor valuation, 1,383 were covered in the baseline owner survey, and 138 were covered by the tax census with a rental contract being reported. In Panel (B) of Table A1, we report correlations with such values from other sources.

D.2 Data from the Property Tax Census

Data extractions. The information recorded in the field on tablets by the bureaucrats is automatically sent to a data server hosted by the tax administration. We received regular data extractions from the server and use this to compile our dataset for the analysis. We mainly use two extractions, one recovered on January 31st, 2023 (for census operations carried out between 2019 and January 2023), and one recovered on May 5th, 2023 (for census operations between January and May 2023).

[91] For budgetary reasons, it was not feasible to hire more assessors, nor to have multiple assessors value the same property.

A very small number of properties were covered twice (0.7 percent of observations) – we keep the observation from the most recent visit. Each bureaucrat has a unique identifier allowing us to know which bureaucrat covered each property.

Sections. Initially, 97 sections were randomly assigned to treatment. However, before the census started, one section was "dissolved" by the cadaster, with its properties being all allocated to either one of two other existing sections.[92] These two sections which gained extra properties are in our control arm hence this change does not affect our empirical strategy, and simply brings the number of sections assigned to treatment to 96. Out of these 96 sections, the census was interrupted in two of them in the very first days, because of pre-existing tensions between the local population and the tax administration around property titles. Only 11 out of 89 and 16 out of 404 properties were covered before the interruption, and the information was not used to generate tax bills. We drop these two sections – one in each arm– from the analysis.[93] Hence the number of treated sections in our analysis is 94, with 47 in each arm.

Creating the property-level dataset and tax liabilities. The forms on the tablets are structured so that bureaucrats enter information at the owner level, and at the tenant level if there are tenants. For our analysis, we aggregate information at the property level. For 99.32 percent of the 38,417 properties covered in the census, there is only one owner per property – even if the property is a building, it is most frequently owned by a single individual. To generate tax liabilities, we need information on whether the property is occupied by the owner as her main residence. Bureaucrats tick a box indicating whether the property is used for main residence. If this is the case, and no parts of the property are for rent, we apply the 1.5 million F abatement to the tax base. If the property is both used as main residence and rented in part (this is the case in 18.7 percent of observations in the census), we use additional information, since the abatement is only applied to the component of the tax base corresponding to the main residence. To assign a share of the value to owner-occupied parts, we rely on (i) the number of rooms used as main residence in the Algorithm arm (see Appendix E),(ii) the the value to owner-occupied parts directly entered by bureaucrats in the Discretion. In a robustness check, we apply the procedure using number of rooms to both arms (see Appendix G.1).

Corrections: values. We correct two types of entries which we assume are due to typos during the field work. First, we identify cases were property values are negative (2 observations). We replace these by the absolute value. Second, we identify cases where a monthly value is too small to be realistic, below 10,000 F (16 USD). We replace these by 0. There are 610 of these cases out of 20,079 properties in the Discretion arm; 30 out of 1,166 in the analysis sample.

Corrections: observable characteristics. There were several occurrences of a technical bug leading to the temporary absence of some or all of entries for observable characteristics on the tablets in the field. 192 properties out of the 18,148 of the Algorithm arm have all observable characteristics missing, and 8,744 have one characteristic or more missing, most often these four: Architecture, Sidewalk, Quality of Doors and Windows, Presence of a Cement Wall. This affects 569 out of 1,166 observations of the analysis sample. Our replacement strategy is as follows: we replace each missing characteristic by its mean value in a given section. If there are no occurrences of the characteristic in the section, we replace the missing characteristic by its mean value overall. In one of

[92] This section was in the commune of Rufisque Ouest, covered by the Rufisque tax office.
[93] These sections are in the commune of Yoff, covered by the Ngor-Almadies tax office.

our robustness checks, we re-calibrate the algorithm dropping the characteristics that were often missing, and our main results are unchanged (see Appendix G.1).

Other variables. Other variables recorded on the tablets that we use for the analysis are: whether the bureaucrat met the owner during the visit; the status of the owner (Male, Female, Deceased, Other); whether the property is the main residence of the owner; whether the property is rented; whether the owner receives a government pension. The bureaucrat enters this information for each owner in the rare instances of multiple owners. Finally, bureaucrats could leave comments (text entries) for each property, there is a comment for 54 percent of observations. We process the text from the comments by screening for key words and identifying two types of instances: (i) the bureaucrat reports the occurrence of conflict or tensions between her and the occupants; (ii) the bureaucrat reports having estimated the value without relying on any information from occupants.

D.3 Data from Real Estate Agency Listings

One major challenge limiting the utilization of property values recovered from real estate listings is the imprecision of addresses. Most often, each listing can only be associated with a municipality.[94] Furthermore, online listings are much less prevalent than in cities of middle- or high-income countries, and only cover small segments of the market. Nonetheless, we use two sources of data as sanity checks for our assessor values.

First, we use data on real estate agency listings between 2019 and 2025 that were compiled by *AML Estate Intelligence*, a Senegalese real estate market intelligence firm.[95] Out of the approximately 13,000 listings, we exploit the 2,020 for which we have the built area measurement, and the 11,847 for which we have the number of rooms. These span 16 of the 19 municipalities covered by our sample. In Figure A10, we plot the distribution of values per square meter in each municipality, for our assessor values and the real estate listings. The results provide reassurance that assessor values very strongly overlap with the posted listings. We provide statistics on the comparison of assessor values and these agency listings in Table A1. We compute the Kolmogorov–Smirnov distance between the two distributions in each municipality. This distance has a mean of 0.46 when considering prices per square meter, and 0.38 when considering prices per room suggesting strong overlap. Next, we compute the share of listings for which the value is within the bounds of assessor values for the same municipality and find that this is the case for 99 percent of listings.

Second, we collected data from online rental listings from the most populated real estate websites in Dakar, and manually assigned a cadastral section to listings by reading the description of the neighborhood included in the ads.[96] This could only be achieved for 84 listings spanning 18 sections. Next, we verify whether the values fall within the bounds of assessor values for the same section. This is the case for 96 percent of them, as reported in Table A1.

D.4 Property Tax Equity Metrics

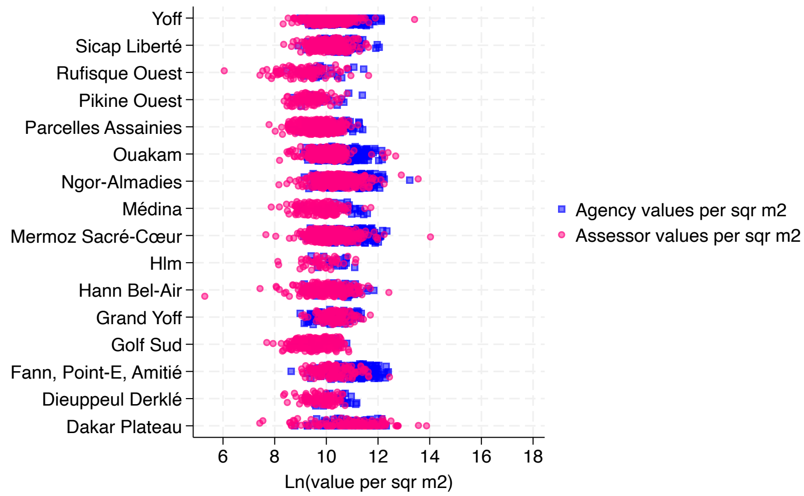
We draw on two metrics from the property valuation literature throughout our analysis. The assessment ratio is computed as the tax roll value – obtained through the property tax census,

[94] Each city is divided into municipalities, which are themselves divided into cadastral sections. Our study spans four cities as shown in Figure A2, and nineteen municipalities.

[95] `https://aml-services.org/`. We thank Tim Ferber, Paris School of Economics, for sharing this dataset.

[96] The websites are `https://keur-immo.com`, `https://www.expat-dakar.com`, and senegalcity.com (which has closed mid-2025).

FIGURE A10
DISTRIBUTION OF VALUES FROM ASSESSORS AND FROM AGENCY LISTINGS BY MUNICIPALITY



Notes: This Figure shows the distribution of property values, by municipality, separately for assessor values (pink dots) and agency listings (blue squares). The x-axis is the logarithm of rental value per square meter. Assessor values were obtained from the licensed assessors we hired and agency values were compiled by a real estate intelligence firm *AML Estate Intelligence*. Sample: 4,061 properties in the assessor sample and 2,020 properties in the agency sample.

using the algorithms in Algorithm sections and bureaucrats' value in Discretion sections – divided by the market value obtained from assessors:

$$AR_i = \frac{TaxRoll_i}{MarketValue_i} \tag{4}$$

Hence, while in the USA property taxation literature (Avenancio-León and Howard, 2022; Dray et al., 2025) for instance, this ratio is computed as assessor valuations divided by realized market values, we adapt this measure to our context where realized "market values" are not observed but are obtained from assessors. As a measure of horizontal (in)equity, we use the coefficient of dispersion, labeled *dispersion* in the tax profile graphs shown in Figure 2, A4 and A5. It is computed as the average percentage deviation of the effective tax rate from its median (the average difference between an observation and the median, in absolute value, divided by the median and multiplied by 100). A higher *dispersion* value is indicative of less horizontal equity. We adapt the coefficient of dispersion (COD) metric from the property valuation literature where it is most commonly used to measure variability (horizontal inequity) in assessment ratios (International Association of Assessing Officers, 2022).

E Property Valuation Algorithms

Selection and coding of characteristics. We selected the observable characteristics to be used in the rule-based property valuation algorithm by drawing from existing methodologies of the cadaster department of the tax administration, and through work sessions bringing together members of the administration, the research team, and international practitioners. The retained characteristics are: number of floors, usage (residential, commercial or mixed), type of fence (four options), state of the fence (very good, average, bad), type of cladding (six options), state of the cladding (very good, average, bad), cement ('hard') wall (yes or no), presence of decorative tiles (yes or no), quality of doors and windows (very good, average, bad), landscape improvement (yes or no), architectural improvement (yes or no), garage (simple, double or none), balcony (yes or no), location with respect to main road (on, near, off), type of road (five options), presence of sidewalk (yes or no), whether the property is at an angle/street corner (yes or no), presence of street lights (yes or no). As much as possible, the phrasing of the characteristics and their different modalities were preserved from pre-existing forms used by the cadaster. The characteristics that have 'yes' or 'no' answers are coded as binary variables. The characteristics that have multiple choice answers are all coded as categorical variables, including characteristics related to state or quality. Both the rule-based and the pure rule algorithms include as predictors total built area, number of floors and section fixed effect.

Calibration. For the functional form as well as the calibration details, we followed recommendations from the property valuation literature (Davis et al., 2012; McCluskey et al., 2013; Franzsen and McCluskey, 2017; Ali et al., 2018; Fish, 2018; Guan et al., 2011; Moore, 2005; International Association of Assessing Officers, 2022).[97] Out of our sample of 4,916 properties with market values, we randomly assign 10 percent of observations ($N = 517$) to the test sample. The test sample is constructed by sampling from all sections to ensure that we can estimate section fixed effects. We run an elastic-net regression following specification 1. The elastic net is retained because we find that it performs better out-of-sample than the simple OLS and a Lasso regression, and a random forest model does not increase precision much but strongly increases complexity.[98] We use cross-validation, by estimating the model five times and recovering the median value of each coefficient. Next, we apply the resulting coefficients to the dataset to obtain predicted values, and identify outliers. Outliers are defined as predictions for which the residual is more than three standard deviations away from the mean value of residuals. These observations (59 for the rule and 51 for the pure rule) are dropped, following McCluskey et al. (2013). We repeat the calibration of the elastic-net regression with five-fold cross-validation. We recover the median value for each coefficient, these are our final coefficients.

Computation of predicted values in monetary amounts. The algorithm predicts $\widehat{Ln(Value)}$. To compute predicted property value $\widehat{Value}$, a correction term should be applied to $exp(\widehat{Ln(Value)})$.[99] The adjusted predicted value can be written as $\widehat{Value} = \alpha_c \cdot exp(\widehat{Ln(Value)})$ where α_c is a correction term. If it is assumed that the error term in the prediction model is nor-

[97] We had the opportunity to interact directly with some of these practitioners and experts between 2017 and 2022.

[98] We find a R^2 of 0.83 with a random forest model. An elastic-net-regression with quadratic and cubic terms for built area and number of floors, as well as interactions between all section fixed effects and built area achieves 65.6 percent of predictions within 30 percent of market value) – these precision gains are limited with respect to the increased complexity.

[99] See Woolridge (2012) *Introductory Econometrics: A Modern Approach 5th edition*, Chapter 6 Section 4.

mally distributed, it can be shown that predicted values should be computed with $\alpha_c = exp(\frac{\hat{\sigma^2}}{2})$ and thus $\widehat{Value} = exp(\frac{\hat{\sigma^2}}{2})exp(Ln(\widehat{Value}))$ where $\hat{\sigma^2}$ is the estimator of the variance of the error term. Based on the distribution of residuals (Panels (C) and (D) of Figures A11 and A12), we can assume that the error term is normally distributed. We compute $\hat{\sigma^2}$ using the RMSE and find $\alpha_c = 1.09$ for the rule-based algorithm and $\alpha_c = 1.12$ for the pure rule. Hence, using the correction term α_c implies multiplying all predicted values by a uniform adjustment factor. However, empirically, we find that our algorithm predictions tend to slightly over-estimate market values throughout most of the distribution – we therefore decide to use $\alpha_c = 1.00$ and report all results accordingly. Our main results are unchanged when using $\alpha_c = 1.09$ (resp., $\alpha_c = 1.12$).

Performance measures. We use performance measures that are standard in the property valuation literature (International Association of Assessing Officers, 2022). Figures A11 and A12 display the diagnostic for normality of residuals. Table A3 reports comprehensive performance statistics for all versions of the algorithm (see Appendix G for a description of each robustness version). The displayed value for the R^2, Adjusted R^2 and RMSE is the mean of each statistic over the five iterations in the cross-validation process. The MAPE and Freddie Mac 30% statistics are estimated on the test sample, and are computed using the property values in monetary amounts, not the $ln()$ transformation, making us more conservative in assessing model performance. Overall, the performance of our algorithms (adjusted R^2 of 0.90 and Freddie Mac 30% of 62.5 for the rule-based algorithm, 0.86 and 60.9 percent respectively for the pure rule) is very satisfactory when compared to similar contexts.[100] The linear fit of predictions over market values (blue dashed line in Panels (A) of Figures A11 and A12) show that the algorithm performs relatively less well at the very bottom of the distribution of property value, where it tends to over-estimate values, and at the very top, where it tends to under-estimate values. This is in line with results from other contexts – some degree of regressivity is commonly found in property valuation models due to unobserved variable bias (such as property features the assessors saw in the field but that are not captured in the variables). This has been shown to be an inherent drawback of property valuation models (McMillen and Singh, 2020; Amornsiripanitch, 2024; Berry, 2021).

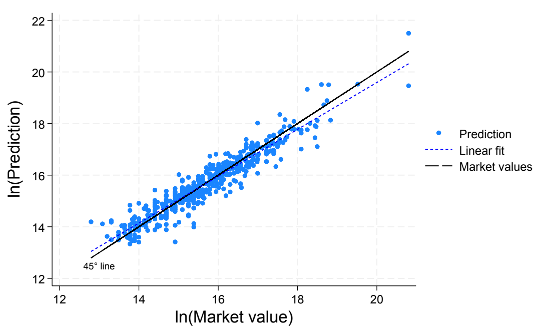
Computation of tax liability. The algorithm yields a predicted value at the level of the whole property (house or building). Indeed, this is the level at which we have built area and assessor values. To generate tax liabilities, we need to divide this value into different sub-components in two instances: (i) when there are multiple owners within a same house or building – this is a rare feature in the Dakar real estate, less than 0.4 percent of observations in the census data – since each owner needs to receive a tax bill; (ii) when the property is partly rented and partly occupied by the owner (18.7 percent of observations in the census data) since in these cases the abatement only applies to owner-occupied part of the tax base. We rely on the number of rooms collected in the census to divide property value into its different components. If the property is partly rented and partly occupied by the owner: the corresponding values are computed as a share of total value, based on the number of rooms allocated to each usage. If there are multiple owners, each owner's value is computed based on her number of rooms out of total number of rooms. This is the policy-relevant methodology since the administration does not have any information on built area measurement at a more granular level than the property as a whole (for instance at the level of apartments within the same building). The resulting tax profile using the rule-based algorithm is shown in green in

[100] Ali et al. (2018) find and R^2 around 0.56 in Rwanda, Franklin (2019) finds an R^2 of 0.85 in Addis Ababa, Ethiopia. Behr et al. (2023) find an MAPE ranging between 0.30 and 0.64 in South Africa, and Bergeron et al. (2024) a MAPE of 69.9 and 53.5 percent of predictions within 30 percent of the true value in the Democratic Republic of Congo.

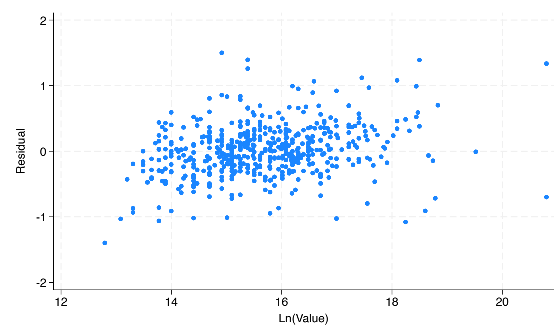
Panel (A) of Figure A4 (Rule-based calibration inputs). The same procedure relying on the number of rooms to recover tax liabilities is applied to assessor valuations to generate the legal tax profile shown in gray in Figures 2, A4 and A5.

Implementation in the new digital tool. The implementation of the algorithm is integrated into the software. The integration of geocoded property details and cadastral data into the software allows to automatically recover built area measurement and location fixed effects. For this experimental roll-out phase, the calculation was done by the research team, in order to include the most recent updates in the algorithm. Going forward, the calculation will be integrated in the software. The software administrator will be able to modify the list of characteristics and their coefficients. This flexibility is key for the sustainable adoption of the digital tool by the administration.

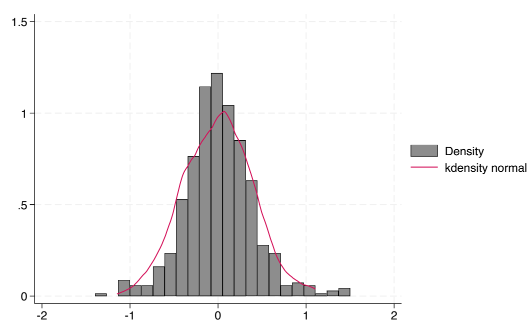
FIGURE A11
RESIDUALS IN THE PROPERTY VALUATION ALGORITHM: RULE-BASED



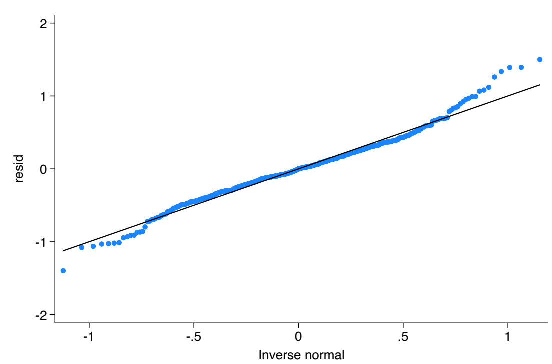
(A) PREDICTIONS OVER MARKET VALUES



(B) RESIDUALS OVER MARKET VALUES



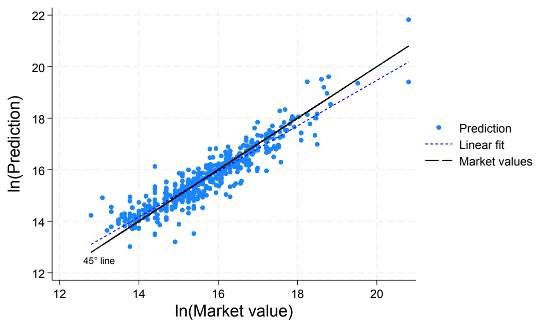
(C) DISTRIBUTION OF RESIDUALS



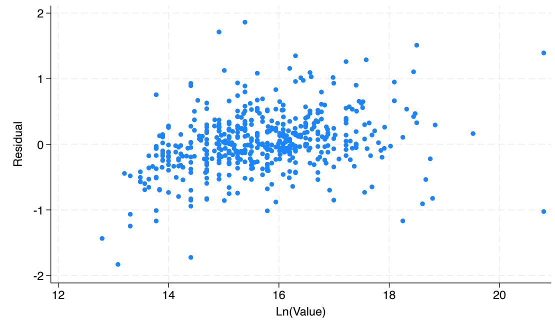
(D) DIAGNOSTIC: NORMALITY OF RESIDUALS

Notes: This Figure shows graphical results from the property valuation algorithm applied to the test sample, using the rule-based version following specification 1. In Panel (A) we plot predicted values $Ln(\widehat{Value})$ for the test sample over market values $Ln(Value)$. In Panel (B) we plot residuals $(Ln(\widehat{Value}) - Ln(Value))$ over $Ln(Value)$. In Panel (C) we plot the histogram of residuals (in gray) and the kernel density of a normal distribution with a mean of zero and with the same standard deviation as the distribution of residuals (in red). Panel (D) is a Q-Q diagnostic plot, where quantiles of the residual are plotted over the expected quantiles for a normal distribution.

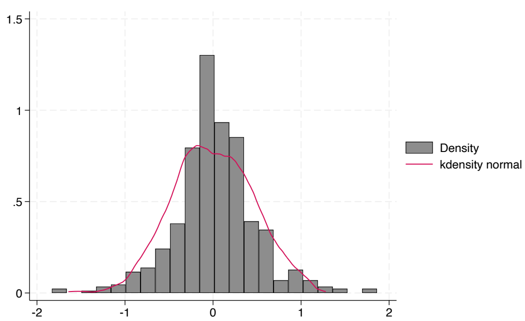
FIGURE A12
RESIDUALS IN THE PROPERTY VALUATION ALGORITHM: PURE RULE



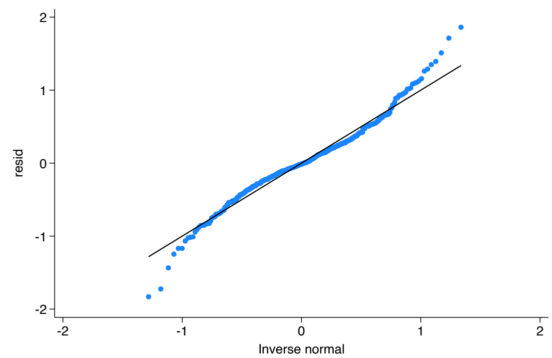
(A) PREDICTIONS OVER MARKET VALUES



(B) RESIDUALS OVER MARKET VALUES



(C) DISTRIBUTION OF RESIDUALS



(D) DIAGNOSTIC: NORMALITY OF RESIDUALS

Notes: This Figure shows graphical results from the property valuation algorithm applied to the test sample, using the pure rule version following specification 1. In Panel (A) we plot predicted values $Ln(\widehat{Value})$ for the test sample over market values $Ln(Value)$. In Panel (B) we plot residuals $(Ln(\widehat{Value}) - Ln(Value))$ over $Ln(Value)$. In Panel (C) we plot the histogram of residuals (in gray) and the kernel density of a normal distribution with a mean of zero and with the same standard deviation as the distribution of residuals (in red). Panel (D) is a Q-Q diagnostic plot, where quantiles of the residual are plotted over the expected quantiles for a normal distribution.

F Surveys: Additional Details

F.1 Bureaucrat Surveys

F.1.1 Baseline Bureaucrat Survey

Bureaucrats completed a self-administered survey at baseline ($N = 247$). The variables we use are:

Age: above median. Takes value one if the bureaucrat is older than the median age, 30 years old.

Gender: Female. Takes value one if the bureaucrat is female.

Previous work with the tax administration: Takes value one if the bureaucrat did any work with the tax administration before the program, either as a civil servant or a temporary employee.

Any higher education: Takes value one if the bureaucrat completed high school and studied in a higher education institution (including vocational training).

Long higher education: Takes value one if the bureaucrat completed a three year or more degree in a higher education institution (including vocational training).

Ethnic group: Self-reported with a possibility not to answer: Wolof or Lebou, Poular, Serere, Diola, Other. Wolof is the majority group in the sample.

Religion: Self-reported with a possibility not to answer: Islam Tidjane, Islam Mouride, Islam Other, Christian. Tidjane and Mouride are orders within the broader muslim community. Tidjane is the majority group in the sample.

Public service motivation: Standardized score computed from the sum of Likert-scale responses to the following questions: "It is important for me to work in the public sector"; "I would not mind doing the same job in the private sector"; "It is not necessarily important for me that my work is useful for the community"; "I do not hesitate to devote all my energy to work". Answers for the second and third statement are reverse coded.

In favor of role of government: Standardized score computed from the sum of Likert-scale responses to the following questions: "According to me, the government can do a lot to make society more fair"; "the government should have the responsibility to satisfy everyone’s basic needs (versus individuals taking care of their own needs)".

In favor of widespread taxation: Standardized score computed from the sum of Likert-scale responses to the following questions: "It is fair for a retired person to pay taxes if (s)he owns property"; "Only the richest people should pay taxes".

F.1.2 Endline Bureaucrat Survey

At the end of the census, bureaucrats completed an endline survey, administered by the research team ($N = 180$). We describe the variables we use and the lab-in-the-field valuation exercise included in the survey.

F.1.2.1 Variables

The variables we use are:

Own rent: Amount of rent paid by the bureaucrat if (s)he is a tenant.

High income: Takes value one if the bureaucrat's monthly income (before the program) is above above 100,000 F, which is the case for 40 percent of respondents.

Emotions reading: This measure is borrowed from the psychology literature and informs on the ability of an individual to understand the mental state of others. It is captured through the Read the Mind in the Eyes Test (RMET), of which we implemented a multi-racial version (Weidmann and Deming, 2021; Dodell-Feder et al., 2020), in which bureaucrats had to label 24 photographs of actors' faces with an emotion. We then compute a standardized score for each bureaucrat based on these responses.

Big five score: Standardized score from a francophone version of the big five personality traits test (Plaisant, 2008). We also use separately the standardized score for each subcomponent: *Openness, Agreeableness, Extraversion, Conscientiousness, Neuroticism.*

Digit span: Standardized score for two digit span tests – in which the respondent is asked to repeat a series of digits of increasing length – the first being a forward test (asked to repeat in the same order as the digits were enumerated) and the second being a backward test.

Math index: Standardized score obtained from six separate math questions: four basic math questions and two with calculations frequently encountered when discussing rental values: $10 + 5$; $27 - 4$; $32 - 13$; 7×6; 150000×4; 70000×12.

Persuasion skills - general: Standardized score from a verbal exercise where the bureaucrat was evaluated on her ability to persuade a (hypothetical) property owner to cooperate by providing information. The evaluation was done by the research team, and the bureaucrat was asked: "Consider you are doing the census, and the owner of a property refuses to cooperate and provide information. What would you tell them?" The research team assigned a general grade based on how persuasive the bureaucrats' discourse was. This exercise builds on Chioda et al. (2021).

Persuasion skills - items: Standardized sum of eight possible persuasion strategies included in bureaucrats' reply in the verbal exercise. The research team ticked the arguments if mentioned by the respondent. The items are: giving more explanations on how the information will be used; suggesting to change languages or make polite salutations adapted to the owners' profile; suggesting to break the ice with some small talk or jokes; mentioning the public services the tax revenues will fund; reassuring the owner by saying that all owners of the neighborhood are getting the census; suggesting to have the supervisor of the tax office intervene; threatening the owner by mentioning possible prosecutions; offering to plan a meeting for later.

Supervisor evaluation: Standardized score given by supervisors when asked: "How would you grade the overall performance of bureaucrat [Name]?" for each bureaucrat they supervise.

Corruption: We asked about corruption using two questions: "During the property tax census, did an owner ever offer you an arrangement for you to change the value you were entering?"; "According to you, did it ever happen to another bureaucrat that an owner offered an arrangement to change the value they were entering?" where in both cases the modalities were "[This happened very often; This happened often; This happened once or twice; This never happened]".

Autonomy: We captured the sense of autonomy at work using a Likert-scale answer to the following question: "In my recent work with the DGID, I have the leeway to propose ideas and solutions to the problems encountered".

Best strategy to obtain accurate property values: We captured bureaucrats' views on this using the following question "According to you, what is the best strategy for the DGID to recover high quality data on rental values?" where the respondent could choose among "Bureaucrats collect values reported by owners"; "Bureaucrats collect values reported by tenants"; "Bureaucrats estimate values using their knowledge and judgment"; "A formula automatically calculates values based on built area and property characteristics".

F.1.2.2 Lab-in-the-Field Valuation Exercise

We designed a lab-in-the-field valuation exercise serving two purposes. First, it allows to measure bureaucrats' accuracy independently from whatever happens during the field visits, and holding constant the information they have on a given property. Second, the module allows to causally estimate how bureaucrats' valuations respond to information on owner status (to test the presence of biases) and to information on true market values (to test the potential of increased training).

The exercise consisted in providing values for properties shown on pictures (displayed in Figure 3). Bureaucrats were shown the picture of a high-value property (in the top 10% of market values) and asked "Consider you are covering this property (located in neighborhood [NAME]) during the census. There is no occupant to talk to, a neighbor tells you the owner is M. Ba, an employed man. What is your best estimate of the monthly rental value you would write down on your tablet?" The exercise was then repeated for a low-value property (in the bottom 10% of market values). The sequence of questions and the experimental design are summarized in Figure A13. We included two randomized treatments.

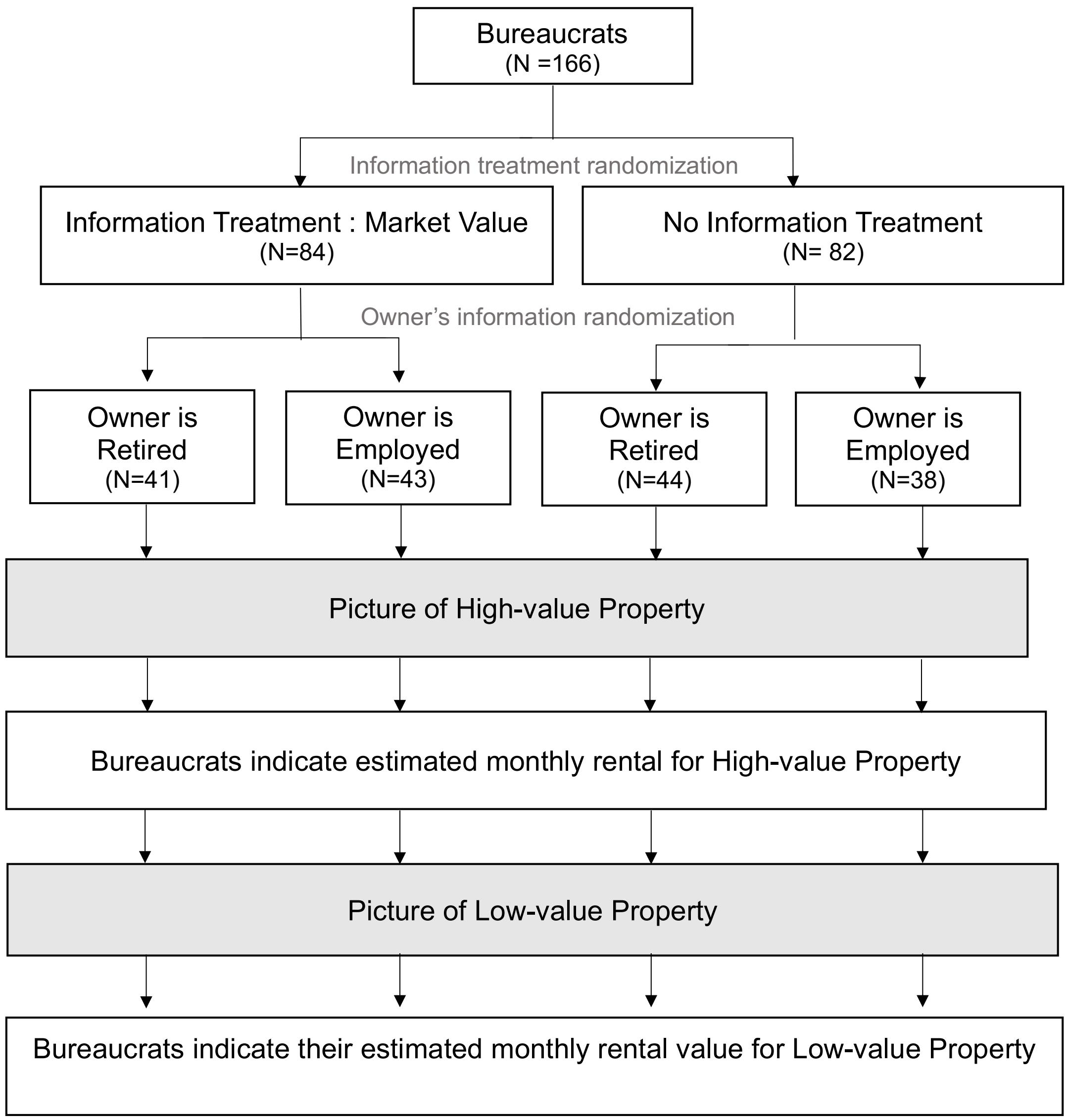
Randomized information treatment on distribution of market values: At the very start of the exercise, half of the bureaucrats received an information treatment: a pedagogical chart explaining the distribution of property values in the region shown in Figure A14 (we build on Stantcheva (2021) and Hoy (2025)). This corresponds to a simple and cheap extension of bureaucrats' training that could easily be integrated by the administration. There were two comprehension questions that needed to be answered correctly before moving to the next questions. This ensured that respondents read and understood the chart.

Retired owner randomization: For each property, the message was modified for half of the bureaucrats, instead of being told "the owner is M. Ba, an employed man", they were told "the owner is M. Ba, a retired man". We chose this feature because in this context, being retired is associated with vulnerability and deserving support. This modification allows to test whether bureaucrats are biased, if they adjust their valuations to owner status, while any information they have on the property is held constant.

Treatment assignment was randomized using a stratification on bureaucrats' observed accuracy in the property tax census, gender, and education level. The two randomizations were independent. 166 bureaucrats completed this part of the survey.[101]

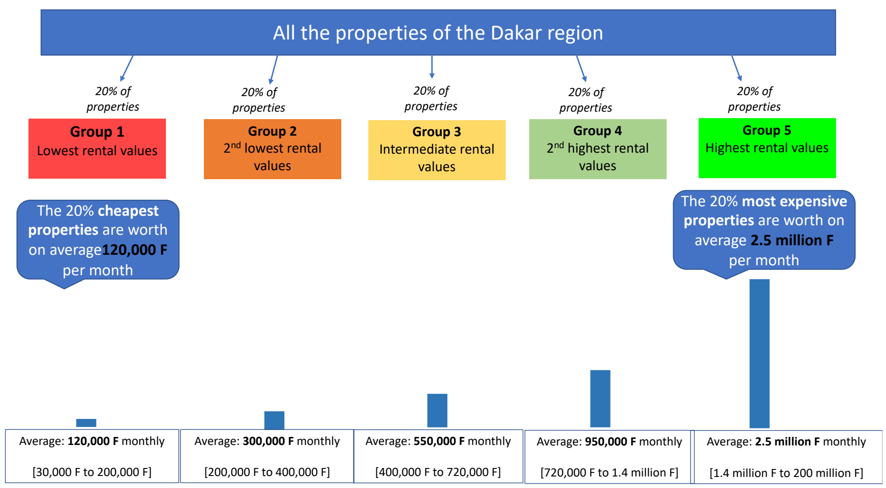
[101] Due to a technical error in the survey implementation, some bureaucrats only completed the exercise for one property. This is why the number of observations in Table A11 is $N = 280$.

Figure A13
Experimental Design of the Lab-in-the-Field Valuation Exercise



Notes: This Figure summarizes the experimental design of the lab-in-the-field property valuation exercise included in the bureaucrat endline survey and described in Section F.1.2.2.

FIGURE A14
INFORMATION TREATMENT ON PROPERTY VALUES USED IN LAB-IN-THE-FIELD VALUATION EXERCISE



Notes: This Figure shows the information treatment that was presented to the randomly selected respondents of the bureaucrat endline survey, as per the experimental design illustrated in Figure A13. The original version was in French.

F.2 Owner and Tenant Surveys

F.2.1 Baseline Property Owner Survey

2,474 property owners were surveyed at baseline in 2018, 1,238 of which in sections covered by the tax census, spanning both Discretion and Algorithm sections.

Property value: Owners are asked to provide their best estimate of their property's value – the value that could be obtained if rented at market prices. Respondents who were not able to provide a value were asked to choose a bracket – for these cases we use the midpoint of the bracket. If part of the property is rented, we ask separately for the amount of rents collected each month, and the value of non-rented parts.

Rented: Takes value one if the owner declares that at least part of the property is rented out.

Owner-occupied: Takes value one the owner reports that this property is her main residence.

High household income: Takes value one if household monthly income is above 200,000 F.

In tax net: Takes value one the owner reports having ever received a tax bill for the property tax, or having ever paid the property tax.

Retired: Takes value one if the owner reports being retired (whether or not reporting receiving a pension).

Ethnic group: Self-reported with a possibility not to answer: Wolof or Lebou, Poular, Serere, Diola, Other. This variable is used to generate an indicator for whether the bureaucrat and the owner are from the same ethnic group.

F.2.2 Endline Property Owner Survey

4,342 property owners were surveyed in the endline survey in 2025, with approximately half (2,186) in sections covered by the census. The survey collected owner characteristics, elicited preferences regarding valuation methods, and implemented a hypothetical tax bill experiment designed to measure owners' willingness to pay taxes and its sensitivity to the valuation method.

For the variables used in bureaucrats' implicit algorithm in Table A12: we first verify that there is no treatment effect of the census on these variable.

F.2.2.1 Variables

Income index: Standardized index computed using respondent's own income, whether the owner/respondent has a formal job, amount of rent collected from the property.

Liquidity index: Standardized index computed using answers to the question "How difficult would it be for your household to face an additional expense of [AMOUNT] F in the coming month?", asked for four different amounts between 20,000 F and 400 000 F.

Retired: Takes value one if the owner reports being retired (whether or not reporting receiving a pension).

Foreign: Takes value one if the owner is not a Senegalese national.

Political connections index: Standardized index computed using responses to "Have you, or someone close, ever worked for the government, an administration, or as an elected official?"; "Was it you or someone close?"; "Do you hold a role as: Neighborhood delegate / Neighborhood councilor/ Municipal councilor / Religious leader".

Corruption: We ask two questions to capture perceptions of corruption: "Some people carry out informal transactions or gifts to agents from the administration to pay lower taxes. According to you, how frequent is this in your neighborhood?" to which respondents could answer "Very frequent/Frequent/Relatively rare/Very rare"; "We have mentioned several activities carried out by the tax administration. During one of these, did it ever happen that an agent from the administration suggested that you make an informal transaction or a gift?"

Property value: Owners are asked to provide their best estimate of their property's value – the value that could be obtained if rented at market prices. An incentive was added to this question to induce respondents to make the effort to provide a response. Respondents are told that they can increase their thank you payment if they provide a market rental value that is in the vicinity of values provided by real estate professionals for their neighborhood and property type. At the very end of the survey, respondent's value is compared to a predicted market value established by the research team using the algorithm (based on the property's location and built area). They could get up to 500 F on top of the standard 2500 F thank you payment if the values were close.

F.2.2.2 Experimental Module: Taxpayers' Preferences over Valuation Methods

We used vignettes to elicit taxpayers' preferences over valuation methods (bureaucrat discretion versus algorithm). We used a hypothetical tax bill exercise to capture the elasticity of tax compliance to valuation methods. The sequence of questions and the experimental design are summarized in Figure A15.

Vignettes. First, the respondents were shown vignettes explaining each valuation method. The vignettes are shown in Figure A16. The enumerator were asked to read the different steps of each method, hence the full script was as follows:

"I will present two methods used by the administration to determine the amount of tax that each property owner must pay. I will then ask you for your views on each method.

A first possible method for the administration to determine a property's rental value is to use a computer or machine, and to rely on a formula or algorithm.[102] Let me explain with an image. [Read caption on vignette]. First, utilization of information on the property, such as built area, lo-

[102] In this context, the word "machine" is commonly used to refer to a computer.

cation and characteristics (number of floors, etc). Second, a formula is used to calculate the rental value. Third, this value is used to calculate the tax amount.

A second possible method for the administration to determine a property's rental value is that an agent judges of the rental value while being in the field. Let me explain with an image. [Read caption on vignette]. First, the agent observes the property. Second, he talks with the owner to obtain more information. Third, he uses his judgment to assess the rental value. Fourth, this value is used to calculate the tax amount."

Importantly, enumerators made no reference to the respondent's own potential tax bill. Rather, the script remained general, by saying that the administration had two concurrent methods used to establish tax liabilities for property taxation.[103]

After the explanation of the two vignettes, respondents were asked which method – the one relying on bureaucrats' discretion or the one relying on an algorithm – was best along the following dimensions:

1. "Which method is most fair?"
2. "Which method is most transparent?"
3. "Which method limits corruption the most?"
4. "Which method is most likely to guarantee that the owners of two similar properties in the same neighborhood face the same tax amount?" [horizontal equity]
5. "Which method is most likely to guarantee that rich owners face the highest tax amounts?" [vertical equity (rich)]
6. "Which method is most likely to guarantee that poor owners face the lowest tax amounts?" [vertical equity (poor)]
7. "Which method do you consider best overall best for the Dakar region?"
8. **"Which method would you prefer be applied to your property?"**
9. "Which method applied to your property would according to you yield the lowest tax amount?"

In all cases, there were three response categories: "The first method, the machine or formula" – "The second method: an agent in the field" – "Neither one nor the other". Question (8) is our measure of taxpayers' preferred method. To investigate drivers of taxpayers' preferences, we generate a variable $PrefAlgorithm_i = 1$ if respondent i answers the algorithm to question (8). For each dimension k, we generate $BestAlgorithm_{ki}$ taking value 1 if respondent i selects the algorithm for dimension k. In Panel (B) of Figure 6, we report correlation coefficients between $PrefAlgorithm_i$ and $BestAlgorithm_{ki}$ for each feature k. A strong correlation means respondents who consider the algorithm best for a given feature are also likely to cite the algorithm as their preferred method. A low correlation means that respondents are likely to consider the algorithm best for a given feature while not choosing the algorithm as their preferred method (or vice versa).

Hypothetical tax bill and elasticity of tax compliance. Next, to assess the sensitivity of respondents' tax compliance to valuation methods, they were told about a hypothetical tax bill for next year and asked about their willingness-to-pay (WTP). The tax bills varied along two dimensions

[103] The enumerators were blind to the section-level randomization and to whether or not the respondents had received a tax bill.

that were cross-randomized in a 2 x 2 design: *High* versus *Low* tax bill amount, and valuation following *Bureaucrat's discretion* versus *Algorithm*. Tax bill amounts were personalized for each property owner to be realistic. We predicted the value of the tax base using our pure rule algorithm. In the *High* tax bill arm, we applied to the predicted value the main residence abatement, and a 30 percent discount.[104] In the *Low* tax bill arm, we further reduced this tax bill amount by 50 percent. The treatments were cross-randomized within cadastral sections, with the following sample sizes by cell: Algorithm X High (N = 977), Algorithm X Low (N = 954), Bureaucrat discretion X High (N = 988), and Bureaucrat X Low (N = 1,039). The exact script was:

"Imagine you receive a tax bill for this property next year, and that the method used is [METHOD], and the amount to be paid for the year is [AMOUNT]. To what extent is it likely or not likely that you will pay this tax bill?"

To which respondents could answer "Very likely";"Likely"; "Unlikely"; "Very unlikely".

In a second stage, all respondents received an information treatment in which they were told that the algorithm-based method improves tax equity. More precisely, the information was shared as follows:

"Some researchers conducted a study to compare these two methods. They showed that with the second method, the agent in the field, two similar properties from the same neighborhood very often have different tax amounts. With the first method, the formula, two similar properties from the same neighborhood very often have the same tax amounts."

They were then asked again about their likeliness to pay the same hypothetical tax bill.
We measure the sensitivity of tax compliance with the following regression:

$$WTP_{ij} = \alpha + \beta Low_{ij} + \gamma Algorithm_{ij} + \epsilon_{ij} \tag{5}$$

where $WTP_{ij} = 1$ if respondent i of section j answered "Very likely" or "Likely" to pay the tax bill, $Low_{ij} = 1$ if the respondent was randomly assigned to the low tax bill treatment and $Algorithm_{ij} = 1$ if the respondent was randomly assigned to the algorithm treatment. We estimate the same regression twice, once using responses provided before the information treatment, and once using responses provided after the information treatment. The results are presented in Panel (C) of Figure 6. The β coefficient measures the how the willingness to pay the tax responds to tax liability. γ measures the sensitivity of tax compliance to the valuation method. There is aversion to algorithms if γ is negative and significant. The change in γ in between the first (before information treatment) and second (after information treatment) estimation measures the sensitivity of tax compliance to stronger equity of the valuation system.

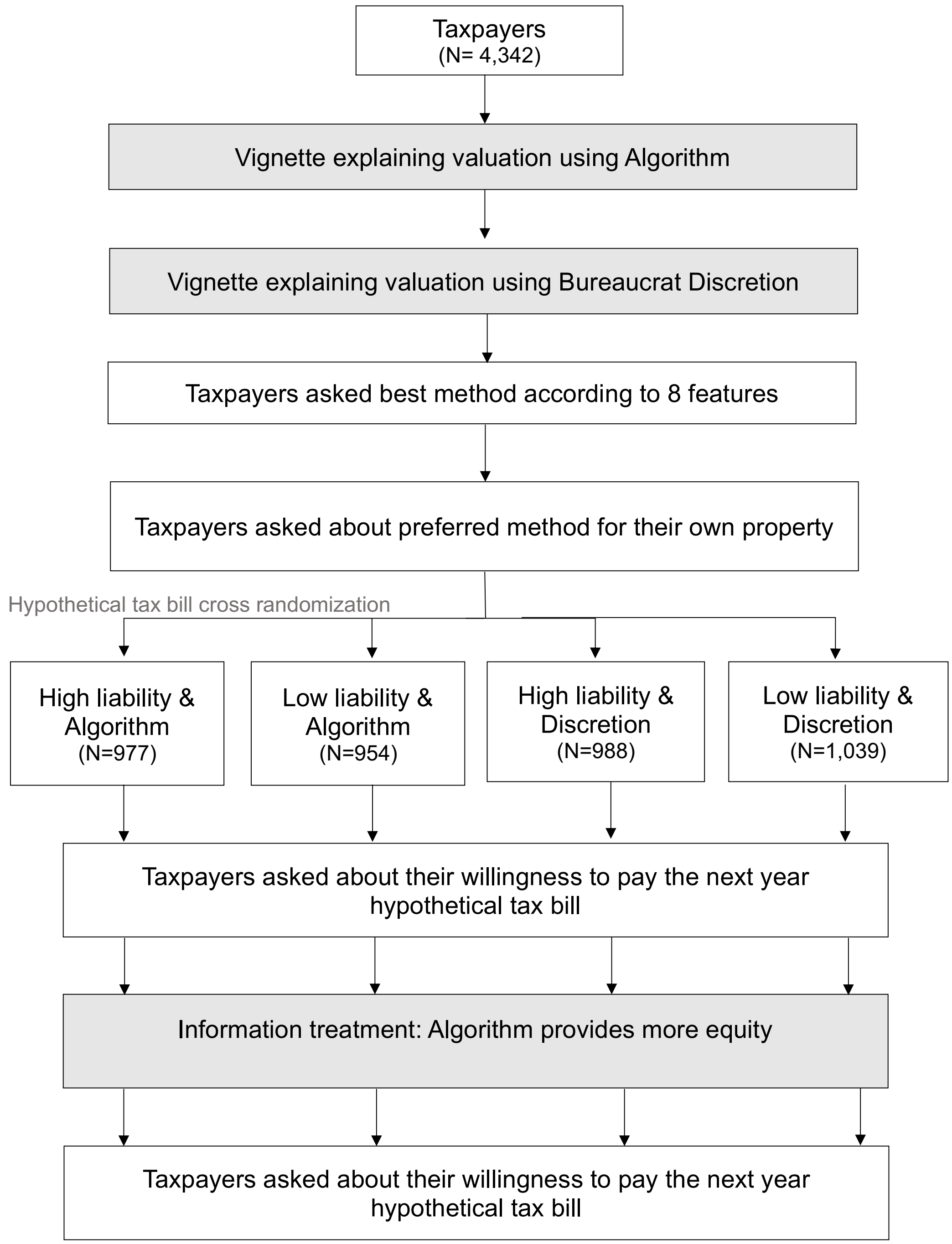
F.2.3 Endline Tenant Survey

We surveyed 1,665 tenants at endline. Of these, 695 were in properties that also appear in the owner endline survey, while 970 were in other properties.

[104] This is in line with the 30 percent discount that was applied by the administration to all the tax bills that were printed and distributed following the census.

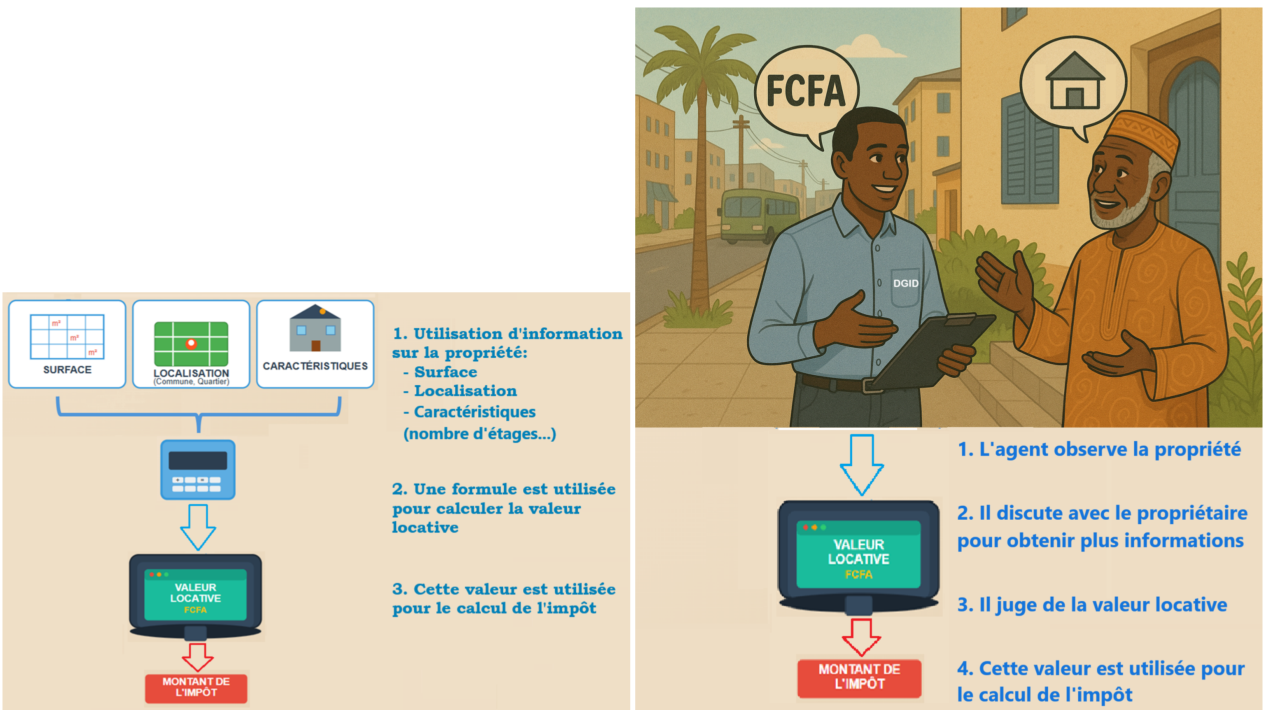
Rent: We asked tenants the amount of monthly rent they pay. To use this variable in our analysis, we need to adjust it to the property as a whole. Indeed, only 14.2 percent of surveyed tenants rent the whole property. For the remaining cases, we estimate rents at the property level using the following calculation: $Rent_i = Rent_t \cdot \frac{R_i}{R_t}$ where $Rent_i$ is our variable of interest, rent for property i, $Rent_t$ is the rent amount paid by tenant t, R_i is the total number of rooms of the property and R_t is the number of rooms rented by tenant t. We observe R_i in the tenant survey, but it is missing in 57.6 percent of cases. We impute where possible. First, for observations where tenants report the number of apartments in the building (17.7 percent of cases), we estimate the total number of rooms in the property by multiplying the number of apartments reported by tenants by the average number of rooms per apartment. Second, we recover R_i from the owner survey for the same property (10.4 percent of cases). Overall, this procedure yields a sample of 1,500 property-level rent observations for our analysis.

FIGURE A15
EXPERIMENTAL DESIGN OF MODULE ON TAXPAYERS' PREFERENCES OVER VALUATION METHODS AND HYPOTHETICAL TAX BILL



Notes: This Figure summarizes the experimental design of the module on taxpayers' preferences over valuation methods and the hypothetical tax bill exercise included in the property owner endline survey and described in Section F.2.2.2.

Figure A16
Vignettes used to Elicit Taxpayers' Preferences over Valuation Methods



(A) Algorithm valuation (B) Discretionary valuation

Notes: This Figure shows the vignettes presented to taxpayers in the module of the property owner endline survey devoted to preferences over valuation methods. Enumerators read the text in blue to respondents while showing the vignette. The text in Panel (A) translates to: "1. Use of information on the property: built area, location, characteristics (number of floors..); 2. A formula is used to calculate the rental value; 4. This value is used to calculate the tax amount". The text in Panel (B) translates to: "1. The agent observes the property; 2. He talks with the owner to obtain more information. 3. He uses his judgment to assess the rental value. 4. This value is used to calculate the tax amount".

G Robustness Checks

G.1 Robustness Checks for the Comparison of Algorithm and Discretion Valuations

We conduct a series of robustness checks for our main results discussed in Section 5.4, on the superior accuracy and equity obtained with the Algorithms. In this section, we provide details on the specifications used for these robustness checks. Some of the robustness checks imply calibrating alternative versions of the algorithms – we report their summary statistics and performance in Table A3. Our verifications then proceed as follows. We estimate regressions similar to 2, and show results in Tables A6 and A7 (as complements to Table 2). Second, we display tax profiles in Figures A4 and A5 (as complements to Figure 2).

Specification using calibration inputs: In this version, we use the inputs (property characteristics) from the assessor dataset to generate rule-based predictions. It serves as a benchmark – indicating how well the algorithm performs when shutting down any errors due to the entry of regressors. By construction, the performance statistics are the ones shown for our main rule-based algorithm (row (1) in Table A3). The resulting tax profile is shown in green in Panel (A) of Figure A4. We quantify the differences between the rule implemented by bureaucrats and this rule with calibration inputs in Panel (A) of Table A7. The sample includes observations from the Algorithm arm only, with two observations per property, and the reported coefficient is the one on an indicator for predictions with bureaucrat inputs. We find that even this limited degree of discretion substantially increases the tax base gap: when the rule is implemented by bureaucrats (rather than using calibration inputs), the median tax base gap increases by 0.33 million F, in the direction of under-valuation, and the absolute gap increases by 1.94 million F or 71 percent.

Our procedure relying on cross-validation already prevents the results from being driven by overfitting, but as an additional verification, we resort to specifications where the analysis is conducted on observations that were not used at any point in the calibration of the algorithm. Second, we re-calibrate the algorithms using values from a different data generating process.

Calibration of algorithms excluding half of the Algorithm arm: In the calibration of the algorithm, we use as the test sample half of the observations from the Algorithm arm of our analysis sample. Then, we conduct our analysis using this test sample only. The drawback is that the sample size for the analysis is smaller ($N = 1706$ instead of $N = 2289$). The performance statistics are shown in rows (3) and (4) of Table A3. Regression results are shown in Panels (A) and (B) of Table A6, and the tax profile in Panel (B) of Figure A4. The results are almost indistinguishable from our main results. For the pure rule for instance, the estimated coefficients are -4.94, -1.85 and 4.69 (in columns (1), (2) and (3) respectively) against -4.93, -2.08 and 4.75 in Table 2.

Calibration of algorithms excluding census areas: In this version, we only use the subset of assessor observations from the pure control areas, where the census did not occur, to calibrate the algorithm. The drawback is that we cannot use section-fixed effects, since by definition the sections for the calibration differ from the ones used for the analysis. Hence, we use strata fixed effects instead – we have 50 strata and 193 sections. The performance statistics are shown in rows (5) and (6) of Table A3. Regression results are shown in Panels (C) and (D) of Table A6, and the tax profiles in Panel (C) of Figure A4. Although the adjusted R^2 of the algorithms drops to 0.87 (ruled-based) and 0.82 (pure rule), we still find that discretion substantially and significantly increases the tax base gap.

Calibration on owner baseline survey values: In this version, the algorithms are calibrated using property values recovered in the owner baseline survey. The sample size for the calibration is $N = 1,680$ including 133 observations in the test sample. In the survey, owners who were not able to provide a value directly were asked to choose a bracket – for these cases ($N = 571$) we take the midpoint of the bracket as the corresponding value. We use the property characteristics from the assessor datasets, but when they are missing ($N = 270$), we resort to the characteristics collected by enumerators during the owner baseline survey. Performance statistics are shown in rows (7) and (8) of Table A3, the R^2 are low – 0.33 for the rule-based version and 0.31 for the pure rule – which is likely due to the low quality of this property value variable. Regression results are shown in Panels (E) and (F) of Table A6. In spite of the poor performance of the algorithms, we still find the same signs on the coefficients, indicating that discretion increases the tax base gap, even though the coefficients are not significant. The tax profiles plotted in Panel (D) of Figure A4 suggest that the algorithms calibrated on the surveys are more regressive than our main algorithms. The reasons are twofold, first, the lower precision of the prediction model generates more regressivity, and second, in the raw survey data we observe that respondents of the first quintile overvalue their properties. The slopes considering quintiles two to four are between -0.016 and -0.012, close to the slope found with discretion. Overall, the algorithms calibrated on the surveys display more accuracy and more horizontal equity than discretion, and similar vertical equity.

Calibration on owner endline survey values: In this version, the pure rule algorithm is calibrated using property values recovered in the owner endline survey (see Section F.2). We only use observations from pure control areas where the census did not occur. For this reason we rely on strata fixed effects instead of section fixed effects. The sample size for the calibration is $N = 3,286$ including the test sample. The performance statistics are shown in row (9) of Table A3, the R^2 is 0.58. Panel (G) of Table A6 reports the regression results and the tax profile is shown in Panel (D) of Figure A4. In spite of the poorer performance of this algorithm compared to our main one, we still find that discretion substantially increases the tax base gap. The coefficients are of the same order of magnitude, and the one for the median tax base gap is significant (column (2)).

Calibration on tenant endline survey values: The pure rule algorithm is calibrated using property values recovered in the tenant endline survey. We only use observations from pure control areas where the census did not occur. For this reason, we rely on strata fixed effects instead of section fixed effects. Only one tenant per property was surveyed. Most often, tenants occupy one or two rooms, it is rare that they rent the whole property. Hence we extrapolate the rent reported by the tenant to the whole property relying on the ratio of rooms occupied by the tenant over total number of rooms – see details in Appendix F.2.3. The sample size for the calibration is $N = 1,472$ including 730 observations in the test sample. Performance statistics are shown in row (10) of Table A3. The regression results are displayed in Panel (H) of Table A6 and the tax profile is shown in Panel (D) of Figure A4. In spite of the poorer performance of this algorithm compared to our main one, we still find that discretion increases the tax base gap. The coefficient is significant when considering the median tax base gap (column (2)), and displays similar sign and magnitudes as our main results otherwise.

Intensive margin specifications: We verify that our results hold when considering the intensive margin only. First, we drop observations for which the tax roll value is 0. Regression results in Panels (E) and (G) of Table A7 and the tax profile in Panel (E) of Figure A4 show that the results hold. Because there is a difference across arms in the probability of a positive value (as shown in Column

(4) of Table A5), we also implement a Lee bounds exercise (Lee, 2009). Results are shown in Panels (F) and (H) of Table A7: since 0 is excluded from the confidence interval found for $\hat{\beta}_{Discretion}$, we conclude that discretion significantly widens the tax base gap even with extreme assumptions on the nature of selection into being assigned a positive value.

Specifications using bounds for assessor values: In this exercise, we relax our assumption on the existence of a single "true value" for each property, and instead rely on the upper and lower bounds provided by assessors (see Appendix D.1). Let LB and UB be respectively the lower and upper bounds for $MarketValue$. While in our main specification 2,

$$Gap = TaxRollValue - MarketValue,$$

here, we set:

$$Gap = 0 \text{ if } LB \leq TaxRollValue \leq UB;$$
$$Gap = TaxRollValue - UB \text{ if } TaxRollValue > UB \text{ and}$$
$$Gap = TaxRollValue - LB \text{ if } TaxRollValue < LB.$$

Hence, this is a conservative approach since the gap is always computed as the distance to the closest bound. As shown in Panels (I) and (J) of Table A7, we still find that discretion substantially and significantly increases the tax base gap: the absolute value of the gap is 3.40 million F or 98 percent larger than with the rule-based process, and 4.31 or 173 percent larger than with the pure rule. In Panel (F) of Figure A5, we add these bounds to the tax profile.

Calibration excluding missing characteristics: In our analysis sample, there are 595 out of 1,166 observations of the Algorithm arm for which at least one of the observable property characteristics is missing in the census data (see Section D.2). In the main analysis, we carry out some replacements as explained in Appendix D.2. As an alternative, we re-calibrate the algorithm excluding the four characteristics that were often missing. The performance is unchanged (row (11) of Table A3). Regression results are shown in Panel (I) of Table A6 and the tax profile in Panel (A) of Figure A5. All results are extremely similar to what we find using our main rule-based algorithm.

Pure rule without building height: In this version of the pure rule, we only use section fixed effects and ground built area as predictors, and ignore the height/number of floors. The performance declines with an R^2 of 0.61 as shown in row (12) of Table A3. Regression results are shown in Panel (J) of Table A6 and the tax profile in Panel (A) of Figure A5. In spite of the lower performance of this algorithm, it still largely and significantly outperforms discretion.

Alternative calculation of abatement: For cases where the property is partly rented and partly occupied by the owner (18.8 percent of observations in the census and 21.16 percent in our analysis sample), we need to determine the share of the value which corresponds to the main residence. This is the share of the value that is subject to the abatement. For our main analysis, as described in Section D.2, we rely on bureaucrats' inputs in the Discretion arm, and on the number of rooms in the Algorithm arm. As a robustness check, we extend the procedure relying on the number of rooms to the Discretion sections to ensure that this difference is not driving the results. We show the resulting tax profile in Panel (A) of Figure A5 – it appears indistinguishable from the main tax profile reported in Figure 2.

Alternative categorization into quintiles: In Panel (B) of Figure A5, we show the tax profiles obtained when using alternative quintiles of property values. First, we compute quintiles of market values per square meter. Note that this exercise is slightly different than the one proposed in our main analysis since these quintiles do not inform us on the final total value of the tax base. Rather, it shows how the tax rates vary by "standing" in the real estate market. Second, we use quintiles of predicted values, where we predict values using our pure-rule algorithm. In both cases, we confirm the finding that discretion is the most regressive, and offers the lowest horizontal equity.

Specifications with bureaucrat fixed effects: In Panels (B) and (C), of Table A7, we re-estimate regression 2, adding bureaucrat fixed effects. The sample is slightly smaller because we drop properties that were not covered at all by the census, for which we do not have a bureaucrat identifier. We find that our main results hold within bureaucrat: for a given bureaucrat, the tax base gap is substantially larger under discretion. It is 4.47 million F or 117 percent (respectively 5.60 million F or 206 percent) wider with discretion compared to the rule-based (resp. pure rule) valuation. To achieve convergence in the quantile regression when the outcome is the median gap, we winsorize at the 5% level.

G.2 Robustness Checks for the Analysis of Bureaucrat Fixed Effects

We assess the robustness of the findings on bureaucrats' influence on the tax roll presented in Sections 5.2 and 6.5 and Table A9 in several ways. The results are displayed in Table A10. We first verify that results remains similar using the non-shrinked $\alpha_{b,EB}$: we find that bureaucrat fixed effects explain 60 percent of variance in the tax base gap under discretion, against 28 percent under the rule, and it remains the case that long higher education correlates significantly with the probability of being a top bureaucrat. Long higher education is significantly associated with a 23 pp higher probability of being a top bureaucrat. Second, we use an alternative version of specification 3, in which we replace the section-level control by an indicator for high-value properties. Bureaucrat fixed effects account for 42 percent of the variance in the tax base gap. It is still the case that long higher education remains a significant predictor of being a top bureaucrat, long higher education is significantly associated with a 13 pp higher probability of being a top bureaucrat.[105] Finally, if we alternatively define top bureaucrats as all those with a negative fixed effect ($\alpha_{b,EB} < 0$), we also find that long higher education is the only significant predictor of being a top bureaucrat, long higher education is significantly associated with a 21 pp higher probability of being a top bureaucrat.

[105] Another reason for conducting this verification is to alleviate the concern that our results are driven by the small imbalance in bureaucrat assignment, leading properties of the highest quintile to have a slightly lower probability of being covered by a bureaucrat with long higher education as shown in Table 1 and Figure A8.