Trust and Taxation: Pre-Analysis Plan

Ali Abbas[*], Michael Carlos Best[†], Ali Cheema[‡], Adnan Qadir Khan[§], & Shandana Khan Mohmand[¶]

February 20, 2025

This text represents the intentions of the research team at the time it is filed. We may deviate from this plan if unexpected issues arise, but we will report a "populated PAP" if we materially deviate from it (Duflo et al., 2020).

Abstract

This study aims to trace out the contours of politically feasible property tax reforms. Experiment 1 focuses on eliciting citizens' preferences for the level and progressivity of property taxes and their determinants through a series of vignettes. This Pre-analysis plan presents the details of our experimental design and survey and our plan for how to analyze the resulting data.

[*]International Monetary Fund
[†]Columbia University, NBER, CEPR, BREAD, IFS
[‡]Lahore University of Management Sciences
[§]London School of Economics
[¶]Institute of Development Studies

Contents

1 Introduction

This project focuses on the political economy of progressive property tax reform. Property taxes are beloved by economists and widely derided by politicians and taxpayers. Economists like the tax's efficiency: the base is easy to identify, immobile, and a good signal of taxpayers' wealth and income (Abbas *et al.*, 2023); (Ahmad & Brosio, 2022); (Collier *et al.*, 2018). Politicians dislike the tax's salience (Cabral & Hoxby, 2012); (Nathan *et al.*, 2020) and fear the political consequences of trying to raise taxes on property, partly because wealthy elites could suffer from such reforms and sway politicians' future electoral prospects. Governments thus often struggle politically to reform property taxes. Correspondingly, property taxes are often regressive. The main questions that guide this component of the project are: (i) why do citizens not support progressive property taxes, especially in contexts where inequality is visibly high, and the majority would benefit from public service delivery that is tied to such taxation; and (ii) how can such support be increased? Our project seeks to provide evidence on these questions, which will help governments identify and implement politically feasible reforms by gaining a deeper understanding of preferences over property tax levels and design features expressed by both politicians and citizens.

The impetus for this study comes from provisions of Punjab's Property Tax Act (1958), which require the Government of the Punjab to notify updated property valuations implying a resetting of the tax rate and a change in the policy parameters that determine the distribution of tax liabilities across taxpayers. The Government of Punjab has engaged the research team to conduct a prospective evaluation that allows it to formulate strategies to increase support for higher and more progressive taxes. It has also engaged the team to test the relative efficacy of different models of mass valuation. For this exercise, Punjab's Excise and Taxation Department entered into a MOU with the research team and the Punjab Urban Unit in September 2021. The study context is Punjab, a rapidly urbanizing province with over half of Pakistan's population. Its capital city, Lahore, the site of the study, is home to over 11 million people. Lahore's current effective property tax rate (0.04%) is significantly lower than in comparators (0.5-1% in the US and Europe, 1-2% in China and the Philippines, and 0.65% in Mexico). Property taxes in Lahore generate very low levels of revenue and they are also regressive. These challenges remain despite improvements in tax capacity because of recent initiatives and the digitization of property records.

1.1 Research Questions

We aim to understand citizens' preferences over the level and progressivity of property taxes. We hypothesize that the acceptance of the current regressive tax policy could be driven by

1. A lack of understanding of the current property tax code;
2. Misperception of the importance of property taxes for increasing public good provision;
3. Perceptions that the state's capacity to collect taxes is low;
4. Perceptions that the state's expenditure is inefficient;
5. Perceptions that the policy-making process is captured by elites.

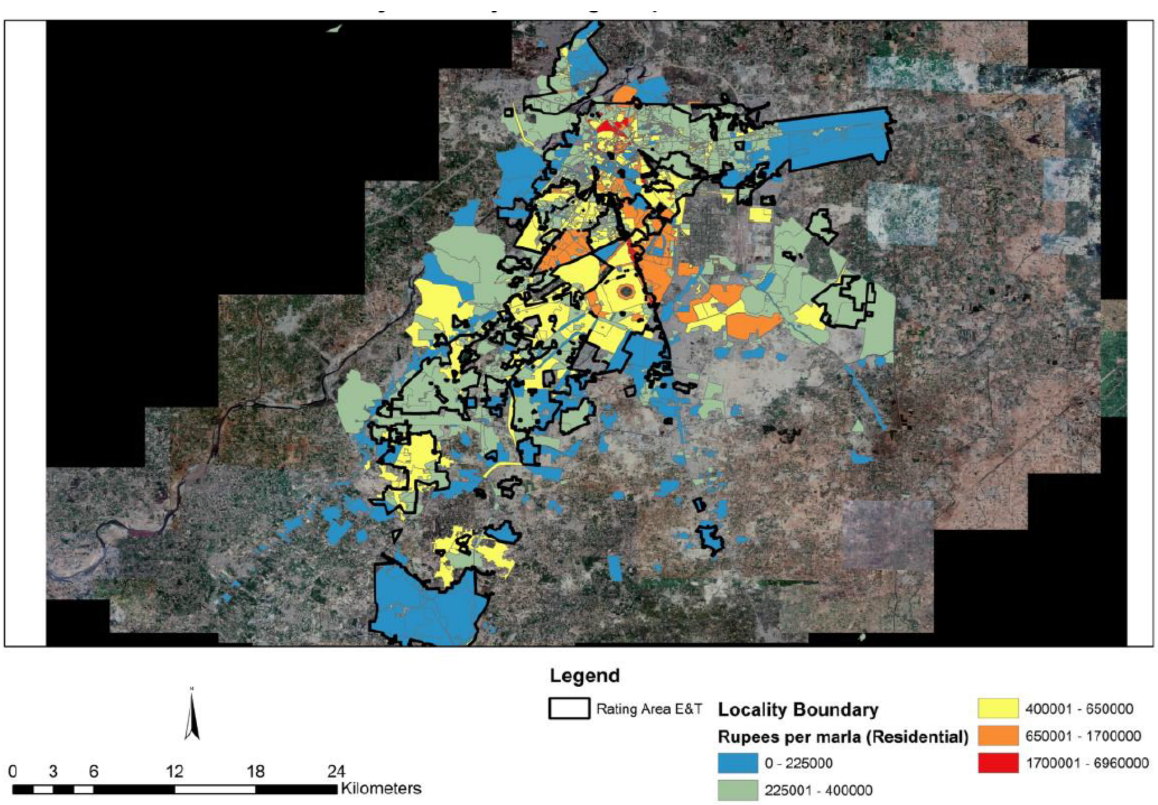
We test for these through a series of vignettes in a survey experiment described in section 3.

2 Context and Sampling

Our study site is Lahore, the provincial capital of the province Punjab, Pakistan. With a population of 110 million, Punjab is the most populous province of Pakistan. Lahore, the provincial capital, is home to 11 million people. Lahore's current effective property tax rate (0.04%) is significantly lower than in comparators (0.5-1% in the US and Europe, 1-2% in China and the Philippines, and 0.65% in Mexico).

We work with the Excise and Taxation (E&T) Department, a provincial government revenue authority that administers the collection and billing of property taxes in metropolitan cities. Excise and Taxation administers tax from almost 1 million properties in Lahore. To administer the property tax, the E&T department has divided its rating area in Lahore into two regions headed by a director. A region is subdivided into zones. A zone is comprised of multiple tax circles. A circle is further divided into multiple localities. Localities vary in value due to a variety of reasons such as amenities etc as shown in Figure 1. Appendix A shows a flow chart of the hierarchy in jurisdictions along with the total number.

Figure 1: Locality boundary coverage of residential areas by capital value in Lahore



The total tax liability from Lahore for the year 2021-2022 was PKR 7.44 billion while the total collection was PKR 5.45 billion (an overall compliance rate of 76%).

2.1 Sampling

We draw our required sample of 7,577 residential properties using multiple data source as given in Table 1. We reach our required sample through a two-stage sampling strategy. In the first stage, a locality-level sample was drawn from the "common list" of localities that appear in both FBR 2022 and DC 2019 public lists, and a property-level sample was drawn in the second stage from the E&T cadaster based on the localities picked in the first stage sample.

Table 1: Description of Auxiliary Data Sources

Term	Details
FBR 2022 list	• Publicly available locality-level list of 1,270 localities of Lahore. • These are capital property rates that were estimated in 2021-22. The list contains: locality, town, residential land rate, commercial land rate.
DC 2019 public list	• Publicly available locality-level list of 1,325 localities of Lahore. • The rates are estimated through a DC-based valuation system that was done in 2018-19. The list contains: locality, town, residential land rate, commercial land rate, residential structure rate, commercial structure rate.
UU's DC mapping list	• A locality-level list, which was exported from ArcMap, of DC areas for which the Urban Unit has digitized maps. The list contains: locality, residential DC land rate, commercial DC land rate.
E&T's GIS data	• A property-level subset of E&T's cadastral that has property geocoordinates and DC locality information entered into it.

FBR: Federal Board of Revenue; DC: Deputy Commissioner's Office; E&T: Excise and Taxation Department.

2.1.1 First-stage

In the first stage we sample neighborhoods (localities) stratifying by property values. However, the Excise & Taxation cadaster only contains assessed values, which deviate strongly from market values. To overcome this, we relied on the Federal Board of Revenue's (The federal government's tax authority) 2022 publicly available list of residential and commercial rates for each locality. At the time of sampling, this was the most recent, and most reliable list of locality-level property values and so this was our primary reference for locality values. One challenge this list posed was that it doesn't contain geographic information on the location of the localities and it was difficult to merge it with property-level data from Excise and Taxation. To overcome this challenge, we used auxiliary locality-level data from the District Administration (DC) along with geo-coded maps provided by Urban Unit (UU) - a semi-private institution aimed to provide Geographic Information System information for key policy reforms (see Table 1).

Figure 2: Data Matching Process for Auxiliary Data Sources

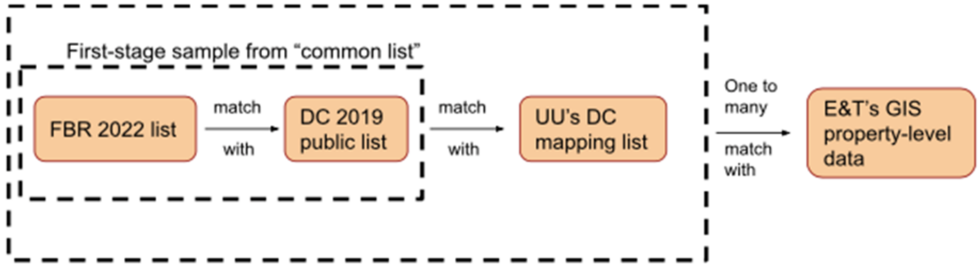


Figure 2 summarizes how the first stage sample was drawn and how it was linked to the second-stage property level data. In order to link FBR 2022 data with Urban Unit's digitized dataset, we first created a "common list" of localities that appear in both FBR and DC lists. To do this, we match the localities by name using STATA's fuzzy matching command to string match locality names across both lists followed by a manual audit to ensure that we match maximum localities across both lists. This common list contained a sampling frame of 1,114 of Lahore as shown Figure 2. This list was further cleaned to drop localities administered by Military and Cantonment Establishment to arrive at a final sampling frame of 1,002 localities that are common across FBR and DC lists.

We use the sampling frame of 1,002 localities from the "common" list to divide the frame into 5 value bins based on the distribution of FBR commercial rates first.[1] All localities within the highest-value bin were randomly ranked and the first 20 localities from this bin were drawn to our sample. Following this, the remaining localities were divided into 5 value bins based on the distribution of residential FBR capital value rates. Within each bin, we intended to draw a random sample of maximum 20 localities from each of the 5 bins to have a total sample of maximum 100 residential localities. All remaining localities, within each bin, were added to our replacement sample as per their respective rank(s). Out of the total sample of 105 localities, we could only map 82 localities from the UU digitized maps list. The remaining 23 localities were replaced by 20 next-in-line mapped localities giving a first-stage sample of 102 FBR/DC localities (see Table 2).

[1] We categorize locality into lowest bin if it was less 20th percentile; low if it was between 20th and 40th percentile; Medium if it was between 40th and 60th percentile; High if it was between 60th and 80th percentile and Highest if it was greater than 80th percentile based on the the FBR commercial/residential capital value rate in the full FBR list (not just the 1,002 localities in the common list).

Table 2: Distribution of Localities by Value Bin and FBR Rate

Distribution of Localities by Value Bin and FBR Rate			
Value Bin	**FBR Rate**	**# in Population**	**# in Sample**
Commercial			
Highest	Commercial	22	19
Total (Commercial)	–	**22**	**19**
Residential			
Lowest	Residential	73	20
Low	Residential	409	20
Medium	Residential	439	20
High	Residential	54	20
Highest	Residential	5	3
Total (Residential)	–	**980**	**83**

These 102 FBR/DC localities with digital maps were then merged with E&T administrative data by overlaying property geo-coordinates from E&T on the digital maps of the localities. Whenever at least 1 property with coordinates in the E&T cadaster fell inside a geo-coded DC locality, we assigned all properties in that E&T locality to that DC locality. Using this method, we merged 66 DC localities with E&T data. The remaining 36 DC localities were merged by showing DC maps to the relevant E&T inspectors who identified the localities manually. 5 localities were subsequently replaced with next-in-line localities as these localities fell out of the E&T's rating area.

2.1.2 Second-stage

The Excise and Taxation Department, Government of Punjab, Pakistan provided us with an anonymized cadaster of 1 million properties and contains information on property use (residential or commercial), ownership status (owned or rented), and property location (main or off-road). In addition, the cadaster contains information on the property's valuation category, which captures the quality of facilities and infrastructure in the property's locality. Each property is assigned a valuation category ranging from A to G. The second stage of our sampling consisted of drawing properties within each locality drawn in the first-stage sample from the E&T property cadaster. Thus, the second-stage sampling frame comprised 179,641 properties corresponding to the 102 DC/FBR localities from the first-stage sample. Only fully residential and fully commercial properties were retained to get this frame. Residential properties were stratified using land area (above and below median). Commercial properties were stratified using a covered area (above and below the median).

The following target sample sizes were set to be drawn from each locality:

- Residential:
  - Lowest-valued: 90 properties (20 localities)
  - Low-valued: 90 properties (20 localities)

- Medium-valued: 120 properties (20 localities)
- High-valued: 150 properties (20 localities)
- Highest-valued: 150 properties (3 localities)

- Commercial:
  - Highest-valued: 150 properties (19 localities)

Since some localities did not have enough properties to meet their sample targets, it was decided to oversample from localities in bins where bin target size $\leq$ bin sample size. Once these bins were determined, localities with at least 30 unsampled properties were identified and picked randomly to meet target sizes. Samples from all 4 strata were drawn from each locality so that their total added up to the target locality sample size and each sample strata size was proportional to actual strata size. A stratified random sample of 12,363 properties was drawn, including 7,577 residential properties and 4,786 commercial properties. The remaining 167,278 properties in the sampling frame make up the replacement sample.

As explained in Section 7 below, following rigorous piloting, we decided that surveying commercial properties was untenable, so we dropped them from our survey sample and focus exclusively on 7,577 residential properties.

2.2 Weighting

Our survey sample is representative of the areas where it was feasible for us to implement our survey—namely the areas where digital geographic data was either already available or we were able to create it relatively easily (see the dicscussion of the first-stage sampling in section 2.1.1).

Starting from the universe of tax-liable properties in Lahore, and given our sampling protocol, the probability that a particular household is in our sample is the composite of three probabilities:

$$\begin{aligned}\Pr(\text{sample property}) = {} & \Pr(\text{sample property}|\text{sample locality}) \times \\ & \Pr(\text{sample locality}|\text{locality in sampling frame}) \times \\ & \Pr(\text{locality in sampling frame})\end{aligned}$$

We can compute these three probabilities and recover each property's sampling probability. With these we can reweight the survey sample to be representative of the universe of tax-liable properties in Lahore.

3 Survey and Experimental Design

In order to elicit preferences over the level and progressivity of property taxes, and to understand their determinants, we implement a survey with a sample of citizens of Lahore. Section 2.1 above outlines our sampling strategy. At the beginning of the survey, we provide respondents with a brief overview of property taxes, focusing on the concept of the average tax rate (which underlies our measures of progressivity) and its use to assess the tax burden borne by different classes of property holders.

This is followed by a battery of questions that will allow us to assess citizens' comprehension of this concept. Figure B.1 shows the vignette we use to guide respondents through these concepts, and the full survey attached to this PAP contains the knowledge questions (questions s2_q1, s2_q4, s2_q7, s2_q8_1, s2_q8_2, s2_q8_3, s2_q8_4, s2_q9, s2_q11, s2_q12). To anchor beliefs about the overall level of property taxation in Lahore, all respondents were also shown the overall average tax rate of properties in Lahore: 0.04%

3.1 Unit of Analysis

Our unit of analysis is a residential property. Within a household, we survey the person responsible for paying the property tax and other utility bills. We survey 7,577 residential properties which are being used either for self or rental purposes. In case of rented properties, we survey the tenant.

3.2 Interventions

To understand the determinants of citizens' preferences over property taxes, we randomly assign survey respondents to receive combinations of 6 survey-experimental interventions. The implementation of the randomization is described in section 3.5 below.

3.2.1 Correcting Misperceptions

In the Correcting Misperceptions treatment respondents are shown photographs of 5 properties[2] that are representative of Lahore's regressive schedule and we ask citizens to assess the tax liability associated with each property through an interactive dashboard shown in Figure B.2. Respondents' answers are then used to estimate the shape of the average tax rate schedule implied by their answers which is shared with the respondents.

We then provide respondents with the actual average tax rates faced by each of the properties and show them the actual shape of the tax schedule in Lahore. Figure B.3 shows the exhibit used to present this information.

This intervention allows us to (a) measure the inaccuracy of citizen beliefs and (b) correct them by providing respondents with true information about the level of taxes paid by different properties and the distribution of the property tax burden across Lahore's properties.

3.2.2 Placebo

The placebo intervention acts as a placebo for the interventions discussed in sections 3.2.3 – 3.2.6 below. This group receives an informational video message (Please see exhibits from the video in Figure B.4) about the different tiers of government and the assignment of revenue and spending functions to each tier. The placebo message does not contain any information about the fiscal relationship between the citizen

[2]We reduce the number of vignettes to 3 for greater than 7 marlas properties for reasons explained in Section 7

and the state but provides information about general government functions and takes a similar amount of time in the survey as the interventions in sections 3.2.3 – 3.2.6 below.

3.2.3 Public Goods

The Public Goods intervention provides respondents with information on local public good deficits in Lahore and argues that the lack of financing, which is a consequence of low property tax utilization is a big constraint on the government's ability to meet the public service delivery needs of citizens (Please see exhibits from video in Figure B.5).

3.2.4 Revenue Leakage

The Revenue Leakage intervention provides respondents with information on the magnitude of the property tax compliance challenge in Lahore and the potential financing that can become available if there was full compliance. It ends with the message that raising adequate financing for local public good provision in the city will be difficult for government in the absence of improved compliance (Please see exhibits from video in Figure B.6).

3.2.5 Spending Leakage

The Spending Leakage intervention provides respondents with information on citizen perceptions in Lahore about tax reciprocity, i.e. the proportion of taxes that are spent on the provision of public services in the city. It ends with the message that raising adequate financing for local public good provision in the city will be difficult for government in the absence of measures that can strengthen tax reciprocity (Please see exhibits from video in Figure B.7).

3.2.6 Elite Capture

The Elite Capture intervention provides respondents with examples of recent cases where opposition from high value property owners in Lahore successfully delayed the introduction of reforms designed to raise more property taxes from the wealthy. It ends with the message that raising adequate financing for local public good provision in the city will be difficult for government in the absence of cooperation from the wealthy elite of the city (Please see exhibits from video in Figure B.8).

3.3 Preference Elicitation

Following these interventions we collect our main outcome by asking respondents in the experiment about their preferred tax structure. This is done by presenting respondents with information on a series of 9 residential properties, similar to (Fisman *et al.*, 2020). Respondents are then asked what they believe the current average tax rate the property tax is bearing and what they think is the appropriate average tax rate for the property. Rigorous piloting showed that the most effective method for accurately eliciting respondents' preferences involved using the average ATR of Lahore as a benchmark. Respondents were first asked whether the current rate or their preferred rate was above or below this average, followed by questions regarding the magnitude of deviation from the benchmark, either higher or lower.

Figure B.9 shows a screenshot of the Android dashboard developed for the preference elicitation module. Respondents are shown the properties' lot size, built area size, usage, the predicted market value of the property, and the number of stories. The property value predictions come from a random forest algorithm applied to data we gathered from real estate agents on their expert opinions on the values of 12,363 properties (see appendix C for details). Respondents are shown three randomly picked low-value properties (below the 50th percentile of the value distribution), three randomly picked medium-value properties (between the 50th and the 90th percentile of the value distribution), and three randomly picked high-value properties (above the 90th percentile of the value distribution)[3].

We then compute the total revenue from the tax schedule implied by the respondents' answers using the method described in Appendix D by extrapolating from the respondents' answers to the universe of properties in Lahore (see Figure B.10 for a screenshot of the graph generated and follow-up questions). When the respondent's preferred schedule raises more revenue than the current system, we ask the respondents how they would like to spend it (questions spending, budget_support, international_debt, property_tax in the attached survey instrument), choosing between 1) increasing spending on public services, 2) reducing budget support from provincial/federal government, 3) paying back debt owed to international donors, 4) lowering property taxes. Similarly, whenever the respondent's preferred schedule raises less revenue than the current system, we ask respondents how they would like to cover the shortfall, choosing between 1) reducing spending on services, 2) requesting more budget support from provincial/federal government, 3) raising more debt from international donors, 4) raising property taxes.

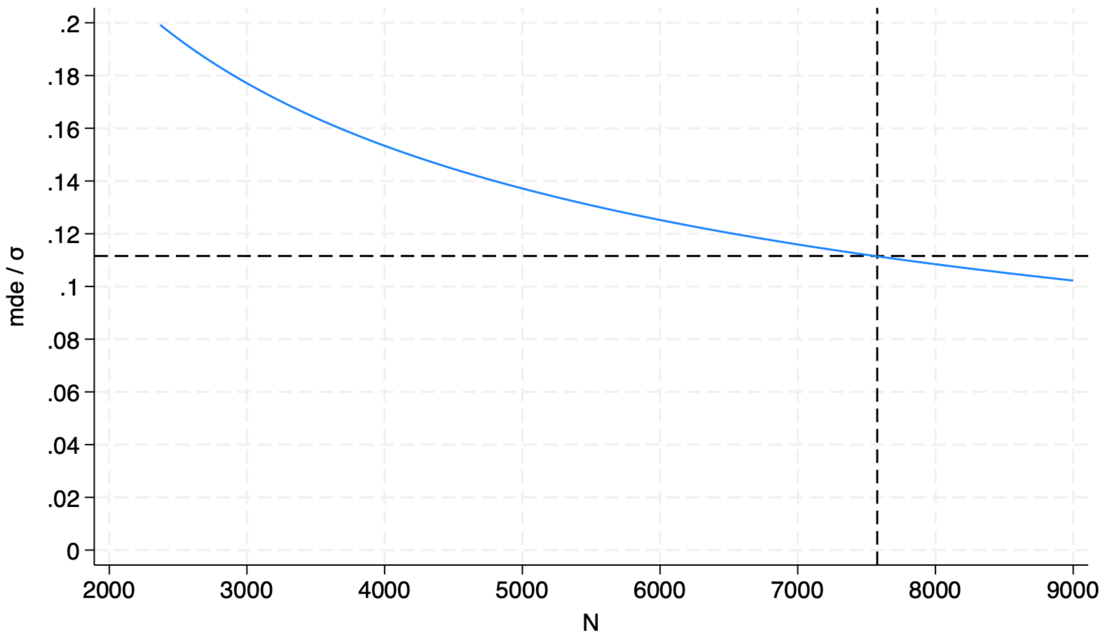
3.4 Statistical Power

Figure 3 shows the minimum detectable effect for our survey experiment as a function of the sample size, showing that we are powered to detect even relatively small effects. As described in detail in section 3.5 below, we assign 1/6th of our sample to each of our treatment arms. This means that the minimal detectable effect comparing any two treaetment arms is given by

$$\frac{\text{MDE}}{\sigma} = \left(t_{\kappa} + t_{\alpha/2}\right) \cdot \frac{2\sqrt{3}}{\sqrt{N}} \tag{3.1}$$

[3]For greater than 7 property sample, we reduced the number of properties to two from each strata and six in total to reduce survey-based fatigue

Figure 3: Power Calculations for Experiment

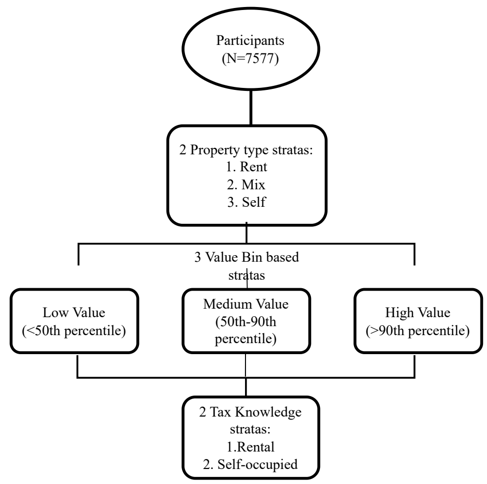


As the figure shows, at our sample size of 7,577, we can expect to detect effects of at least 0.11 (residual) standard deviations, giving us confidence that the experiment is well-powered.

3.5 Treatment Assignment

All participants were interviewed at residential properties. Participants were allocated into distinct groups using a three-tiered stratification process based on property type, property value, and elicitation category. Property types were classified as rented, self-occupied, or mixed. Property values were classified as Low (0 - 50th percentile), Medium (50th - 90th percentile), and High (Above 90th percentile) according to predicted wealth percentiles for each property type (using the predicted property values from the estimation described in appendix C). The third set of strata assigned respondents to view either owner-occupied or rented properties during the preference elicitation module described in section 3.3.

Figure 4: Treatment Assignment Stratification



This resulted in 18 strata formed from the combination of the three property types, three value categories, and two elicitation groups ($3 \times 3 \times 2$) as shown in Figure 4. The residential sample of 7,577 was then evenly distributed across each stratum, ensuring that at least 420 properties were sampled into each of the 18 strata. These 420 properties in each stratum were subsequently randomly assigned to one of six treatment statuses, leading to a treatment status assignment for 7,560 properties. The remaining 17 properties were randomly sorted and assigned a treatment status, resulting in a sample of 1,263 properties assigned to five treatment arms and 1,262 properties to the "pure control" group.

Table 3 summarizes the randomization design, treatment assignment, and assigned sample within each stratum to be surveyed.

Table 3: Sample by type of treatment

Tax Knowledge	**Policy Preferences**	**Sample N**
None	placebo	1,262
Correcting misperceptions	placebo	1,263
Correcting misperceptions	public goods	1,263
Correcting misperceptions	revenue leakage	1,263
Correcting misperceptions	spending leakage	1,263
Correcting misperceptions	elite capture	1,263
	Total N	**7,577**

4 Outcomes

4.1 Primary Outcomes

4.1.1 Tax Progressivity

Our most important primary outcome is survey respondents' desired degree of tax progressivity. We use four measures of progressivity that are commonly used in the literature.[4] Each measure is normalized so that 0 means a proportional tax system, positive numbers mean progressive tax systems, and negative numbers mean regressive tax systems. We also combine the four measures into an index of progressivity since they each capture slightly different aspects of the progressivity of the overall tax schedule, and so that we can use a single measure of progressivity when we explore heterogeneity of the treatment effects.

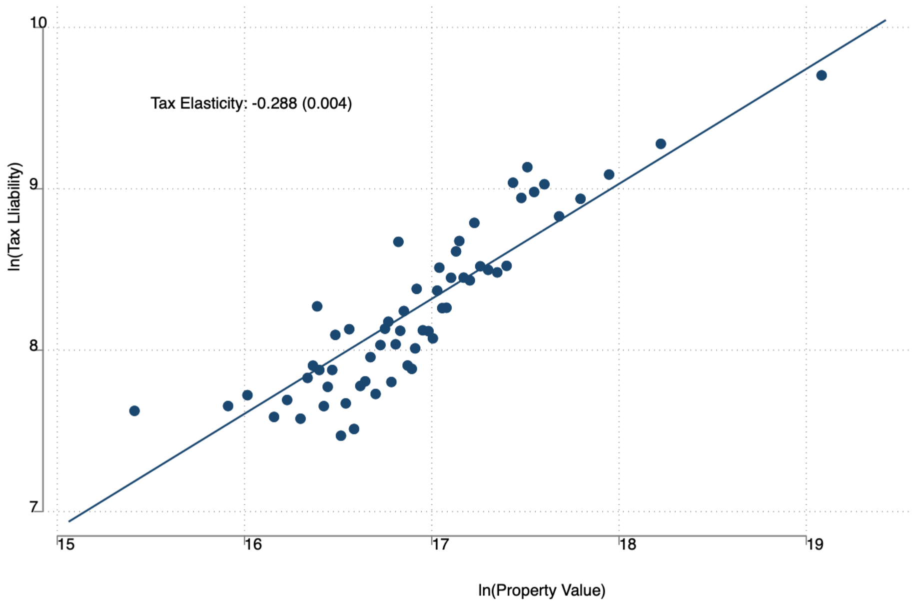
Our progressivity measures are:

1. ***Tax Elasticity:*** The tax elasticity is $\hat{\beta}_1 - 1$ from the regression of log tax liability on log property value shown in equation (4.1). This measure is admirably simple and fits well with the spirit that "A rate structure is progressive where the average rate of tax (i.e., tax liability as a percentage of income) rises when moving up the income scale" (Musgrave & Thin, 1948, p. 498).

$$\ln\left(\text{tax liability}_i\right) = \beta_0 + \beta_1 \ln\left(\text{property value}_i\right) + \varepsilon_i \tag{4.1}$$

Figure 5 shows a binned scatterplot (using the stata implementation of the `binsreg` command (Cattaneo *et al.*, 2024)) of the tax elasticity in the Excise & Taxation cadaster. We find that the tax elasticity is -0.288, indicating that the system is regressive.

Figure 5: Baseline Tax Elasticity

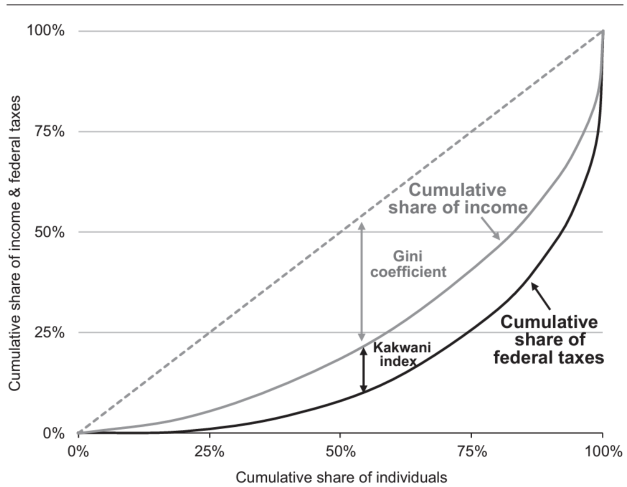


[4]See e.g. Thomas (2023) for a review focusing on low-income countries.

2. ***Kakwani Index:*** This index is based on the Lorenz curves of property wealth and of taxes paid (Kakwani, 1977). The index is easiest to understand visually, as presented by Splinter (2020) for the US income tax:

Figure 3

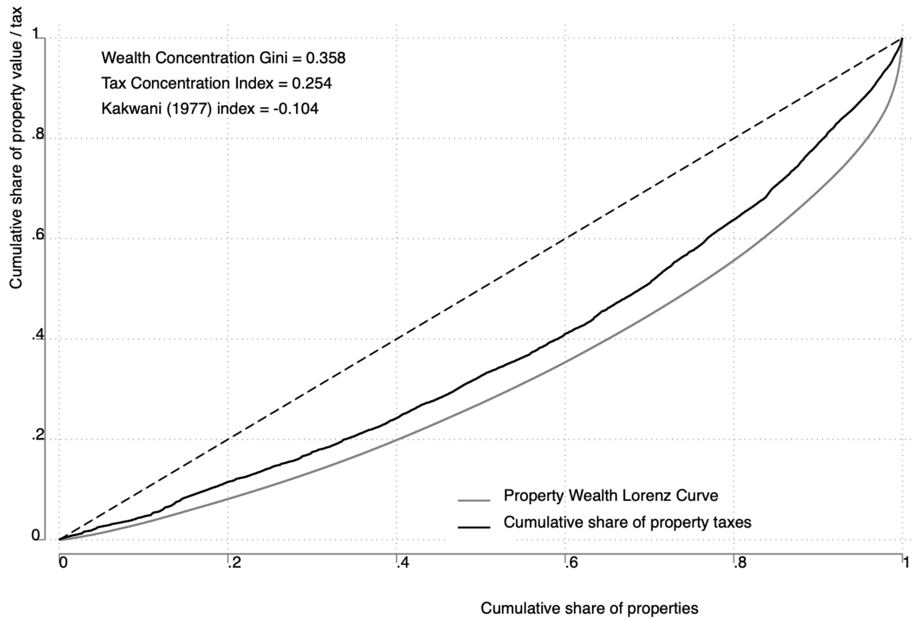
Kakwani Index of Federal Tax Progressivity, 2016



Notes: Income is market income plus social insurance benefits. The Gini coefficient and Kakwani index are two times the area between the curves.
Source: Author's presentation of CBO data.

Figure 6 presents the Kakwani index in Lahore estimated from the Excise & Taxation cadaster. We find that the residential property wealth concentration gini coefficient is 0.358 while the concentration index for taxes is only 0.254, so that the Kakwani index, at -0.104, also indicates that the property tax in Lahore is regressive.

Figure 6: Baseline Kakwani Index

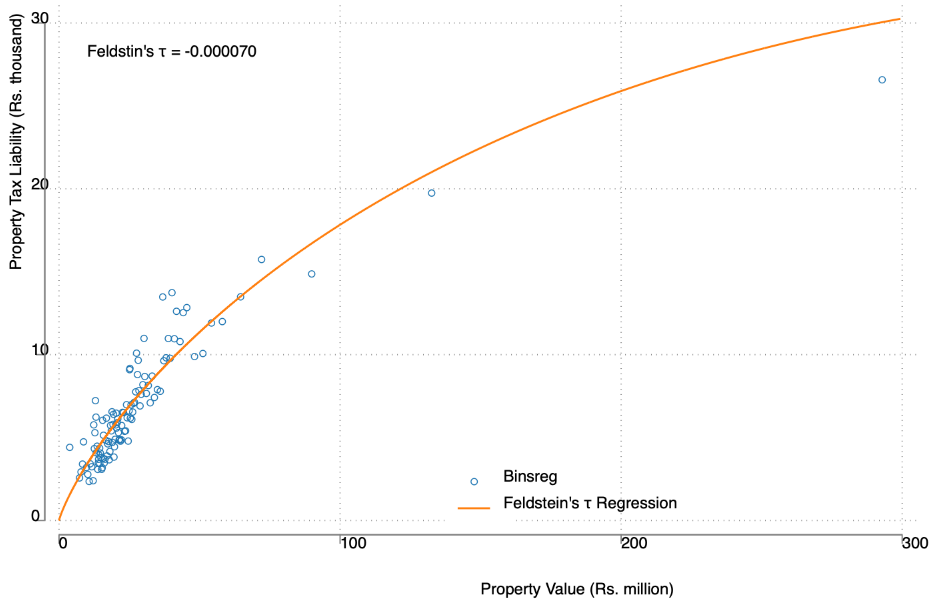


3. ***Feldstein-τ***. This measure derives from Feldstein (1969) and has been widely used in macro public finance, for e.g. Heathcote *et al.* (2017). The measure is the estimated $\hat{\tau}$ from the non-linear regression (4.2):

$$\text{tax liability}_i = \text{property value}_i - \lambda \text{property value}_i^{1-\tau} + \varepsilon_i \tag{4.2}$$

Figure 7 presents our estimation of Feldstein's τ from the Excise & Taxation cadaster. We estimate Feldstein's τ to be equal to -0.00007, again indicating a regressive system.

Figure 7: Baseline Feldstein's τ

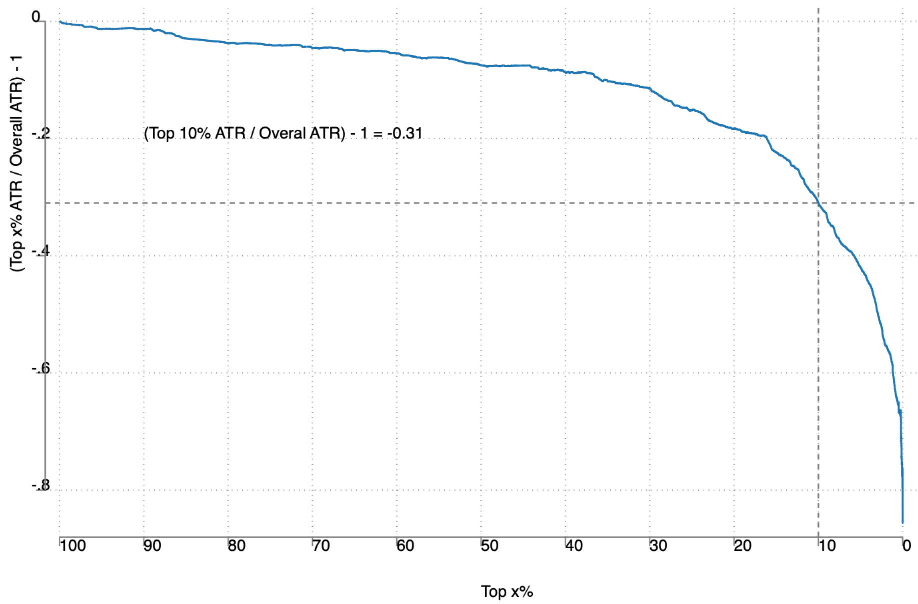


4. ***Top Tax Rate***. Since much of the literature focuses on progressivity at the very top of the distribution (e.g. Piketty & Saez, 2007), our fourth measure focuses on the top 10% of the property value distribution:

$$\text{Top 10\% Progressivity} = \frac{\text{Top 10\% ATR}}{\text{Overall ATR}} - 1 \tag{4.3}$$

Figure 8 presents our estimation of The Average Tax Rate throughout the wealth distribution. For each percentile of the wealth distribution we compute the average tax rate (total tax liabilty / total wealth) of properties above that percentile in the wealth distribution. We normalize this by dividing by the overall ATR and subtracting 1 (so that mechanically our index is equal to zero at the bottom of the distribution). We estimate that the normalized average tax rate of the top 10% is equal to -0.31 again indicating a regressive system.

Figure 8: Baseline Top ATR



5. ***Progressivity Index***. Following Kling *et al.* (2007), our final progressivity measure is an equally-weighted index of our four measures. Denoting the four measures above $Y_1, \ldots, Y_4$, the index is

$$\tilde{Y}_i = \frac{1}{4} \sum_{k=1}^{4} \frac{Y_{ki}}{\mathsf{sd}\,(Y_k)} \tag{4.4}$$

where the standard deviations $\mathsf{sd}\,(Y_k)$ are computed from the control group.

Measures 1. & 3. don't depend on the distribution of property values, but measures 2. & 4. do. To apply these measures to a constant distribution of property values, we take the distribution of residential property values we observe in the baseline data. To apply a respondent's preferences to this full distribution, we estimate a restricted cubic spline using the respondent's 9 responses.[5] With this we have an estimate of the respondent's full tax function to apply to the full property value distribution.

When we study progressivity, we will study all five of these measures separately. However, as mentioned above, when exploring heterogeneity of the treatment effects, we will favor our progressivity index. This way of constructing the index has the virtue of being simple and transparent. But it gives equal weights to all four measures, and some may capture variation across respondents more faithfully than others. If, using our equally-weighted index, it seems that this does not adequately capture the variation in the responses, we will also use the first principal component of the four (normalized) progressivity measures (as in Kling *et al.*, 2007). Additionally, for robustness, we will also study each item individually for robustness.

[5]Specifically, we place knots at the 50th, 75th, and 90th percentiles of the overall residential property value distribution, and then fit a cubic spline restricted to remain between 0 and 100% within the observed range of property values.

4.1.2 Tax Revenue

Our second primary outcome is the level of property taxation preferred by the survey respondent. For this, we take the 9 elicited preferred property tax rates, and use them to estimate a restricted cubic spline.[6] This gives us an estimate of the respondents' full preferred tax schedule.

With the respondents preferred tax schedule we can compute the revenue that would be raised by applying this schedule to the full cadaster of properties in Lahore. Since we only show respondents residential properties, we assume that the respondents' preferred commercial tax rates are such that the proportions of overall revenue raised by residential and commercial properties are the same under the respondents' preferred tax schedule as under the current tax schedule. See Appendix D for technical details.

4.2 Secondary Outcomes

Our secondary outcomes fall into 11 categories.

1. Current property tax' fairness (s8_q18, s8_q20, s8_q21)
2. Tax morale (s8_q13 s8_q14 s8_q15 s8_q17)
3. Representation (s5_q1, a3_q1_1, a3_q1_2, a3_q1_3)
4. Property tax rationale (s8_q5, s8_q7)
5. Behavioral responses to taxing high-value properties (s6_q1, s6_q2, s6_q3, s6_q4, s6_q5)
6. Incidence of taxing high-value properties (b1_q1, b1_q2, b1_q3, b3_q1, b3_q2)
7. Uses of deficit/surplus from preference elicitation. (spending, budget_support, international_debt, property_tax)
8. Behavioral responses to overall tax increases (s6_q6, s6_q7, s6_q8, s6_q9, s6_q10)
9. Incidence of taxing average-valued properties. (b2_q1, b2_q2, b2_q3, b3_q3, b3_q4)
10. Inequality views (s8_q1, s8_q3, s8_q4)
11. Payment behavior in subsequent fiscal year. (s4_q3)

For each outcome, we study each outcome separately, and also create an index of the outcomes in that category to summarize respondents' overall view on that topic.

5 Heterogeneity

We use 7 characteristics of our respondents to explore the heterogeneity of the treatment effects (using the methods described in section 6.3). As our outcomes for the heterogeneity analysis we consider two outcomes: The progressivity index defined in section 4.1.1 and the total revenue from the preference elicitation module.

[6] Specifically, we place knots at the 50th, 75th, and 90th percentiles of the overall residential property value distribution, and then fit a cubic spline restricted to remain between 0 and 100% within the observed range of property values.

1. Prior beliefs and direction of updating
2. Wealth: The questions s20_q10-13 provide detailed information on the types and quantities of assets ownship, while the property condition component reflects the physical state and quality of the respondents' properties. This combined index serves as a measure of wealth. We will use the first component from principal component analysis of s20_q10-s20_q13 and s16_1-s16_q14 to create an index of wealth.
3. Political Participation and partisanship: The questions s13_q1-s13_q4 provide detailed information on various forms of political participation, including engagement with government services and involvement in community, professional, and civic meetings. We will use these questions to create an index of political participation by using the first component from the principal component analysis. This index will serve as a measure of the respondents' level of civic engagement, allowing us to assess how active they are in both formal and informal political processes over the past year. For partisanship, we will create a measure for support for the current and previous governments.
4. Trust (institutional and communal). For institutional trust, the questions s2_q8 and s2_q9 assess the level of institutional trust among respondents, specifically focusing on their trust in property administration and the government.We will use the first component from a Principal Component Analysis (PCA) of these two questions to create an index of institutional trust. This index will capture the underlying dimension of trust in these key public institutions. For communal trust we will use questions s2_q10 and s14_q3 that assess social trust, focusing on trust in fellow citizens of Lahore. We will use the first component from a Principal Component Analysis (PCA) of these questions to create an index of social trust. This index will capture the underlying dimension of trust in the community, enabling analysis of its relationship with civic engagement and public safety perceptions.
5. Knowledge of Tax Policy: The questions s2_q1-s2_q5 assess respondents' knowledge of property tax policies in Lahore, covering topics such as the basis for tax calculation, changes in tax upon renting, prevailing tax rates, exemptions, and the institution responsible for setting tax rates. We will use these questions to construct an index of property tax knowledge, using the first component from a Principal Component Analysis (PCA) to capture the underlying understanding of property tax regulations among respondents. This index will help in analyzing how property tax knowledge influences compliance and attitudes toward taxation policies.
6. Views about Distribution The questions s14_q1 and s14_q2 explore respondents' views on income redistribution through taxation. We will use these questions to analyze attitudes toward redistributive taxation and assess how these views correlate with their preferences for progressivity.
7. SES (demographics) To create an index of Socioeconomic Status (SES) using age, education, we will combine these variables using the following formula:SES Index $=$ $w_1 \times$ Age $+ w_2 \times$ Education In this formula, age is used directly, education is assigned numerical values based on the level attained. The weights w_1, w_2, can be adjusted based on the importance of each factor in determining SES. This index

will provide a composite measure of socioeconomic status, reflecting the combined effects of age, education, and gender on respondents' overall SES.

6 Statistical Procedures

In this section we describe in detail the statistical procedures we will use to estimate treatment effects in our data. For convenience, we will number the treatment arms so that we can use their numbers as a shorthand reference. Table 4 shows our numbering

Table 4: Treatment Arm Numbering

	Intervention	
No.	Tax Knowledge	Policy Preferences
0	Correcting Misperceptions	Placebo
1	None	Placebo
2	Correcting Misperceptions	Public Goods
3	Correcting Misperceptions	Revenue Leakage
4	Correcting Misperceptions	Spending Leakage
5	Correcting Misperceptions	Elite Capture

6.1 First Stage Estimation

We begin our analysis by studying the impacts of the treatments on beliefs. We do this in two ways. First, we compare posterior beliefs between the treatment and control groups. Second, for the treatment groups, we also elicit priors, and we present descriptive statistics on the changes in beliefs.

6.1.1 Treatment–Control Comparisons

Policy Preferences Interventions We pool the pure control group (group 1) and the correcting misperceptions only group (group 0) into a joint control group. Our main estimating equation is:

$$Y_i = \beta_0 + \mathbf{X}_i\boldsymbol{\beta}_1 + \sum_{k=2}^{5} \delta_k D_{ki} + \gamma_{\text{Stratum},i} + \varepsilon_i \tag{6.1}$$

where the D_{ki} are indicators for the four policy preferences treatment groups, $\gamma_{\text{Stratum},i}$ are randomization stratum fixed effects, and we use heteroskedasticity-robust standard errors. The $\mathbf{X}_i$ are post-lasso controls for precision (Wager *et al.*, 2016). For likert-style questions, if we have fewer than 5% of respondents replying *"Strongly Disagree"* or *"Disagree to some extent"*, we will pool the two together; and similarly for replies of *"Agree to some extent"* and *"Strongly Agree"*.

In auxiliary analysis we also analyze quantile treatment effects for continuous outcomes and to see which part of the distribution the movement in beliefs is in.

Correcting Misperceptions Intervention For this, we use the respondents' answers in the preference elicitation module to the question *"What do you think the average tax rate of this property currently is?"* for the nine properties they are shown. For each property j that the respondent sees, we compute their absolute misperception as the absolute differnece between their response and the true tax rate the property faces:

$$M_{ij} = |ATR_{ij} - ATR_j|$$

We then aggregate these into four measures of the respondent's misperception. The first three average misperception of low-value / medium-value / high-value properties (exploiting the fact that each respondent sees three properties in each value range) and the last averages all nine properties. We then use equation (6.1) to estimate the effect of the correcting misperceptions treatment using only groups 0 and 1. We also compute our five progressivity measures described in section 4.1.1 and compute the absolute discrepancy between the elicited beliefs and the true values.

6.1.2 Within-Respondent Changes in Beliefs

We will also present descriptive evidence on how beliefs changed, and then use this as a dimension of heterogeneity of the treatment effects as described in section 6.3. For each treatment we do this as follows:

- **Public Goods**: we compare `s10_q6` to `s6_q1` and compute the differnce between the two responses. We then classify respondents into those whose beliefs became more optimistic (`s10_q6` > `s6_q1`), and those whose beliefs became more pessimistic (`s10_q6` < `s6_q1`).
- **Revenue Leakage**: We compare `s10_q3` to `s7_q1`. For each we compute the share of revenue the respondent expects to be collected and then compute the difference between the two responses. We then classify respondents into those whose beliefs became more optimistic and those whose beliefs became more pessimistic.
- **Spending Leakage**: We compare response to `s10_q1` to `s8_q1` and `s10_q2` to `s8_q1`. For each pair, we compute the difference and then average across the two items to compute respondent i's change in beliefs. We then classify respondents into those whose beliefs became more optimistic (increased shares of spending) or pessimistic (decreased shares of spending).
- **Elite Capture**: We compare responses to `s10_q4` to `s9_q1` and `s10_q5` to `s9_q2`. For each comparison, we check whether the respondent moved up a category in the 5-point strongly disagree – strongly agree scale. If the respondent increased in at least one comparison and didn't decrease in either comparison, they are classified as more pessimistic. If the respondent decreased in at least one comparison and didn't increase either comparison, they are classified as more optimistic.
- **Correcting Misperceptions**: The correcting misperceptions module elicits prior beliefs for 5 properties[7]. From these responses we can compute our tax elasticity

[7] We reduced this to 3 for second-phase

and "Feldstein-τ" measures of progressivity descibed in section 4.1.1 (the remaining measures require the full property value distribution and cannot be implemented with only 5 property responses per respondent) and compute changes between priors and posteriors. From this we can classify respondents into those who became more optimistic (posteriors are more progressive than priors) and those who became more pessimistic (posteriors are less progressive than priors).

These groups can then be used to estimate heterogeneous treatment effects by direction of updating, as described further in section 6.3. For all interventions, if we see that 15% or more of respondents did not materially change their beliefs, we will create a third group of non-updaters.

6.2 Reduced-Form Treatment Effect Estimation

Our basic analysis of the reduced-form effects of the interventions will use the following regression specification:

$$Y_i = \beta_0 + \mathbf{X}_i\boldsymbol{\beta}_1 + \sum_{k=1}^{5} \delta_k D_{ki} + \gamma_{\text{Stratum},i} + \varepsilon_i \tag{6.2}$$

where the D_{ki} are indicators for the five treatment groups (using the correcting misperceptions + placebo group as the omitted group), $\gamma_{\text{Stratum},i}$ are randomization stratum fixed effects, and we use heteroskedasticity-robust standard errors. The $\mathbf{X}_i$ are post-lasso controls for precision (Wager *et al.*, 2016).

For the lasso variable selection we will include: . For any categorical variables, we will include all levels of the variable as a group in a grouped-lasso procedure. For robustness, we will also estimate a saturated model as suggested by Lin (2013).

Since our sample is not perfectly representative of taxpayers in Lahore, we will reweight the data to estimate Population Average Treatment Effects (PATEs) as described in section 2.2. We will also present unweighted estimates of the Sample Average Treatment Effects (SATEs).

6.3 Heterogeneous Treatment Effects

We proceed sequentially. We will do everything we say here and ppublish the completed pre-analysis plan. If we don't make it very far down the list, we will put the parts we don't get to only in the completed pre-analysis plan.

Beliefs is an important HTE. 1. priors (split into 2-3 groups using controls' distribution) 2. direction of updating.

Heterogeneity Buckets Michael Ali Abbas Ali Cheema Adnan Shandana Total Rank Wealth (Measured by house values and housing quality) 1 2 1 1 1 6 1 SES (Demographics) 6 6 4 4 5 25 6 Housing quality 0 0 Political Participation and Partisanship 2 4 2 2 2 12 2 Knowledge of tax policy 3 1 6 5 6 21 4 Trust (Institutional and Communal) 4 5 3 3 4 19 3 Views about distribution 5 3 5 6 3 22 5

We follow the approaches in Guess *et al.* (2023) that builds on Semenova & Chernozhukov (2020); Kennedy (2023) to estimate treatment effect heterogeneity. This can be implemented using the `tidyhte` library in R. We use a non nonparametric two-stage regression procedure that has been proven to be statistically optimal under weak conditions (Kennedy 2020) – and under essentially no conditions in an experiment. The DR-Learner first constructs a doubly robust pseudo-outcome and then regresses this pseudo-outcome on moderators of interest; cross-fit sample splitting is used to prevent overfitting and reduce bias.

There are three main pre-specification choices:

1. how to estimate the pseudo-outcome
2. how to estimate the second-stage regression of the pseudo-outcome
3. how to do sample splitting

The definition of the pseudo-outcome is:

$$\hat{\varphi}(Z) = \frac{A - \pi(X)}{\pi(X)[1 - (X)]}[Y - \hat{\mu}_A(X)] + \hat{\mu}_1(X) - \hat{\mu}_0(X)$$

where X denotes a covariate, A is a binary treatment (we will break the data into subsets with binary treatments, or extend this model to allow for multiple treatments), Y is the outcome of interest, $\pi(X)$ are the (known) propensity scores, $\hat{\mu}_a(x)$ is a regression prediction of the outcome $\mathbb{E}\left[\hat{Y}|X = x, A = a\right]$. This pseudo-outcome is a double-robust mimic of the difference in the potential outcomes.

The regression predictions will be built using standard machine-learning models (for example those available in the `superlearner` package in R)

The second-stage regression of the pseudo-outcome on the covariate of interest will depend on the type of the covariate:

1. For discrete covariates (with five or fewer levels) we will use within-group averages.
2. for continuous covariates we will use local polynomial regression.

We will use 10-fold cross-validation for tuning parameter selection, both for regression predictions and second-stage pseudo-outcome regression, with 12-fold cross-fitting to separate building of regression predictions from the second-stage regression and second-stage model selection. This ensures efficient use of the data, as each observation will separately contribute to regressions in both first and second stages.

6.4 Adjustment for multiple comparisons

We will implement a sharpened FDR adjustment (Benjamini *et al.*, 2006) in a hierarchical manner building on Anderson & Magruder (2022); Guess *et al.* (2023). Our analysis proceeds sequentially in descending order of the priority of our analysis. Each step tests a set of hypotheses, and only the treatments for which we do not reject the null of no treatment effect proceed to the next step, economizing maximally on power.

1. F first stage outcomes. Continue to the next step only for treatments rejecting null of no effect on the index first-stage outcome.
2. Study P primary outcome tests. Adjust FDR for $F + P$ comparisons. Continue to the next step only for treatments rejecting the null of no effect on the index of progressivity.
3. Add S_1 step-1 secondary outcome comparisons and H_1 step-1 heterogeneity comparisons. Adjust FDR for $F + P + S_1 + H_1$ comparisons.

4. Add S_2 step-2 secondary outcome comparisons and H_2 step-2 heterogeneity comparisons. Adjust FDR for $F + P + S_1 + H_1 + S_2 + H_2$ comparisons.
5. Add S_3 step-3 secondary outcome comparisons and H_3 step-3 heterogeneity comparisons. Adjust FDR for $F + P + S_1 + H_1 + S_2 + H_2 + S_3 + H_3$ comparisons.
6. Add S_4 step-4 secondary outcome comparisons and H_4 step-4 heterogeneity comparisons. Adjust FDR for $F + P + S_1 + H_1 + S_2 + H_2 + S_3 + H_3 + S_4 + H_4$ comparisons.
7. Add S_5 step-5 secondary outcome comparisons and H_5 step-5 heterogeneity comparisons. Adjust FDR for $F+P+S_1+H_1+S_2+H_2+S_3+H_3+S_4+H_4+S_5+H_5$ comparisons.
8. Add S_6 step-6 secondary outcome comparisons and H_6 step-6 heterogeneity comparisons. Adjust FDR for $F + P + S_1 + H_1 + S_2 + H_2 + S_3 + H_3 + S_4 + H_4 + S_5 + H_5 + S_6 + H_6$ comparisons.
9. Add S_7 step-7 secondary outcome comparisons and H_7 step-7 heterogeneity comparisons. Adjust FDR for $F + P + S_1 + H_1 + S_2 + H_2 + S_3 + H_3 + S_4 + H_4 + S_5 + H_5 + S_6 + H_6 + S_7 + H_7$ comparisons.

We group our secondary outcomes and heterogeneity dimensions into the seven sequential groups as shown in table 5.

Table 5: Sequential Groups of Secondary Outcomes and Heterogeneity Dimensions

Group	Secondary Outcomes	Heterogeneity
1		Prior beliefs and direction of updating
2	Current property tax's fairness Tax morale	Wealth (property value)
3	Representation Property tax rationale	Political Participation and partisanship
4	Behavioral response to taxing high-value properties Incidence of taxing high-value properties	Trust (institutional and communal)
5	Uses of deficit/surplus	Knowledge of Tax Policy
6	Behavioral responses to overall tax Incidence of taxing average-value properties Inequality views Payment behavior in subsequent fiscal year	Views about distribution
7		SES (demographics)

7 Survey Rollout

Over time, the study faced several design challenges due to external shocks and a tense political climate. Our initial pilots showed notably low response rates, especially among commercial units. Consequently, we dropped the 4,786 sample commercial properties from the study, as respondents from these units were unwilling to engage with our surveyors.

For residential properties, we divided the sample by land area, as the first two pilots indicated considerable differences in response rates between properties smaller than 7 marlas and those larger (1,575 sq. ft.). The response rate for properties under 7 marlas was 44%, while those over 7 marlas had a response rate of only 10%. We therefore decided to split the survey into two phases. In the first phase we surveyed the 4,897 sample properties below 7 marlas, and in phase 2 we surveyed the remaining 2,680 larger properties.

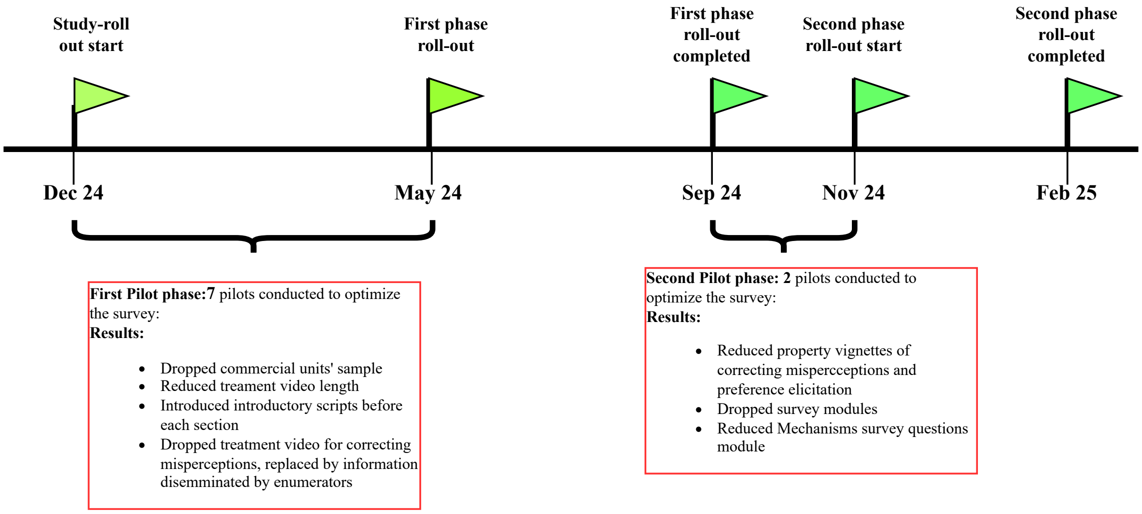
Before rolling-out the phase 1 survey, we performed rigorous piloting and incorporated feedback to revise survey structure, shorten the treatment videos, add introductory scripts before each survey section, and streamlined information to enhance clarity and participation. The first phase of the survey for properties under 7 marlas was rolled-out in the last week of May 2024, and data collection for 4,897 households concluded by the first week of September 2024. The response rate with replacements for this first phase was 53.5%.[8] The second phase started in November 2024 and will be completed in mid-Feb 2025.

When moving to the second phase, our pilots encountered critically low response rates for the larger properties over 7 marlas, prompting us to implement further strategies to improve engagement. Subsequent pilots revealed that to achieve a satisfactory response rate, the total survey length, inclusive of treatment videos and the interactive dashboard, should not exceed 30 minutes. To meet this goal, we reduced the number of property vignettes for correcting misconceptions module from 5 to 3, and for preference elicitation, from 9 to 6. Additionally, we dropped sections on tax knowledge, political participation, and general preferences, and significantly condensed the mechanisms section. Both the longer and the shorter versions of the survey are included with our experiment's registration at https://www.socialscienceregistry.org/trials/15393.

Furthermore, since a substantial portion of the second phase sample was from gated communities, we secured access through the management committees of these societies. The management leadership aided our surveys by connecting us with respondents. The timeline below outlines the main changes.

[8]The response rate is defined as total surveys completed including main and replacement properties expressed as a percentage of total properties attempted.

Figure 9: Timeline of survey roll-out



7.1 Replacement Strategy

When respondents were unwilling to participate in the survey, they were replaced using the following two methods:

1. **“One to One” on field replacement**: If refusals occur, enumerators were directed to identify a comparable property within the neighborhood, ensuring it has the same characteristics regarding land area, built area, and occupancy status. Priority was given first to land area, then built area, and lastly occupancy status, with rented properties being replaced only by other rented properties. Enumerators were also directed to record coordinates in case the property needs to be revisited later.
2. **Using replacement properties from the property cadaster**: In case the survey wing was unable to complete the target number of properties from a locality, the remaining properties were completed by providing them additional sample randomly drawn from the second-stage replacement sample.

References

ABBAS, ALI, AHSAN, NAVEED, AMIN, MUHAMMAD AHMAD, BEST, MICHAEL CARLOS, CALLEN, MICHAEL, CHEEMA, ALI, FAROOQUI, AHSAN, MASUD, OMAR, & MOHMAND, SHANDANA KHAN. 2023. *Property tax utilisation and equity in Punjab: Policy challenges and reform options*. technical report.

AHMAD, EHTISHAM, & BROSIO, GIORGIO. 2022. *Beneficial Property Taxation for Emerging Market Countries: Addressing Climate Change and Post-pandemic Recovery*. Springer Nature.

ANDERSON, MICHAEL L., & MAGRUDER, JEREMY. 2022. *Highly Powered Analysis Plans*. NBER Working Paper Number 29843.

BENJAMINI, YOAV, KRIEGER, ABBA M., & YEKUTIELI, DANIEL. 2006. Adaptive Linear Step-up Procedures That Control the False Discovery Rate. *Biometrika*, **93**(3), 491–507.

CABRAL, MARIKA, & HOXBY, CAROLINE. 2012. *The hated property tax: salience, tax rates, and tax revolts*. Tech. rept. National Bureau of Economic Research.

CATTANEO, MATIAS D., CRUMP, RICHARD K., FARRELL, MAX H., & FENG, YINGJIE. 2024. On Binscatter. *American Economic Review*, **114**, 1488–1514.

COLLIER, PAUL, GLAESER, EDWARD, VENABLES, TONY, MANWARING, PRIYA, & BLAKE, MICHAEL. 2018. Land and property taxes for municipal finance. *Cities that Work. London: International Growth Centre*.

DUFLO, ESTHER, BANERJEE, ABHIJIT, FINKELSTEIN, AMY, KATZ, LAWRENCE F., OLKEN, BENJAMIN A., & SAUTMANN, ANJA. 2020. *In Praise of Moderation: Suggestions for the Scope and Use of Pre-Analysis Plans for RCTs in Economics*. NBER Working Paper No. 26993.

FELDSTEIN, MARTIN S. 1969. The Effects of Taxation on Risk Taking. *Journal of Political Economy*, **77**, 755–764.

FISMAN, RAYMOND, GLADSTONE, KEITH, KUZIEMKO, ILYANA, & NAIDU, SURESH. 2020. Do Americans want to tax wealth? Evidence from online surveys. *Journal of Public Economics*, **188**, 104207.

GUESS, ANDREW M., MALHOTRA, NEIL, PAN, JENNIFER, BARBERÁ, PABLO, ALLCOTT, HUNT, BROWN, TAYLOR, CRESPO-TENORIO, ADRIANA, DIMMERY, DREW, FREELON, DEEN, GENTZKOW, MATTHEW, GONZÁLEZ-BAILÓN, SANDRA, KENNEDY, EDWARD, KIM, YOUNG MIE, LAZER, DAVID, MOEHLER, DEVRA, NYHAN, BRENDAN, RIVERA, CARLOS VELASCO, SETTLE, JAIME, THOMAS, DANIEL ROBERT, THORSON, EMILY, TROMBLE, REBEKAH, WILKINS, ARJUN, WOJCIESZAK, MAGDALENA, XIONG, BEIXIAN, DE JONGE, CHAD KIEWIET, FRANCO, ANNIE, MASON, WINTER, STROUD, NATALIE JOMINI, & TUCKER, JOSHUA A. 2023. Reshares on social media amplify political news but do not detectably affect beliefs or opinions. *Science*, **381**(6656), 404–408.

HEATHCOTE, JONATHAN, STORESLETTEN, KJETIL, & VIOLANTE, GIOVANNI L. 2017. Optimal Tax Progressivity: An Analytical Framework. *Quarterly Journal of Economics*, **132**, 1693–1754.

KAKWANI, NANAK C. 1977. Measurement of Tax Progressivity: An International Comparison. *Economic Journal*, **87**, 71–80.

KENNEDY, EDWARD H. 2023. Towards optimal doubly robust estimation of heterogeneous causal effects. *Electronic Journal of Statistics*, **17**(2), 3008 – 3049.

KLING, JEFFREY R, LIEBMAN, JEFFREY B, & KATZ, LAWRENCE F. 2007. Experimental Analysis of Neighborhood Effects. *Econometrica*, **75**, 83–119.

LIN, WINSTON. 2013. Agnostic Notes on Regression Adjustments to Experimental Data: Reexamining Freedman's Critique. *Annals of Applied Statistics*, **7**, 295–318.

MUSGRAVE, RICHARD A., & THIN, TUN. 1948. Income Tax Progression, 1929–48. *Journal of Political Economy*, **56**, 498–514.

NATHAN, BRAD C, PEREZ-TRUGLIA, RICARDO, & ZENTNER, ALEJANDRO. 2020. *My Taxes are Too Darn High: Why Do Households Protest their Taxes?* Tech. rept. National Bureau of Economic Research.

PIKETTY, THOMAS, & SAEZ, EMMANUEL. 2007. How Progressive is the U.S. Federal Tax System? A Historical and International Perspective. *Journal of Economic Perspectives*, **21**, 3–24.

SEMENOVA, VIRA, & CHERNOZHUKOV, VICTOR. 2020. Debiased machine learning of conditional average treatment effects and other causal functions. *The Econometrics Journal*, **24**(2), 264–289.

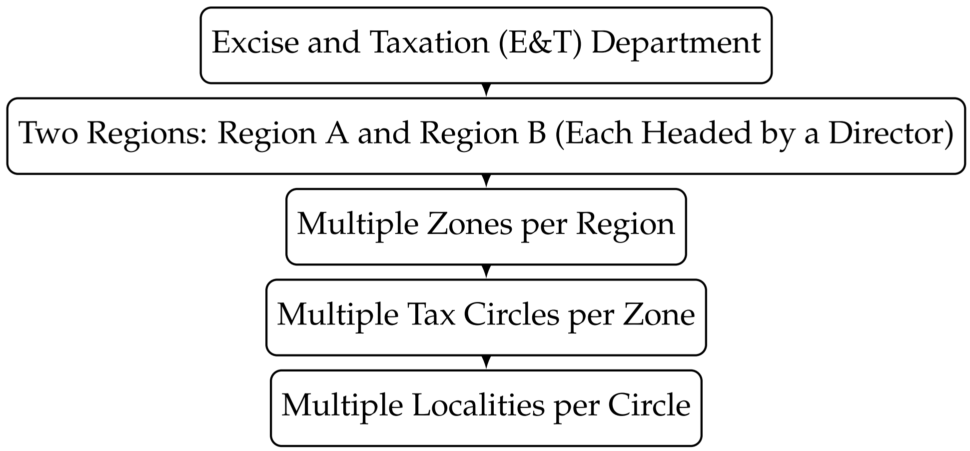
SPLINTER, DAVID. 2020. U.S. Tax Progressivity and Redistribution. *National Tax Journal*, **73**, 1005–1024.

THOMAS, ALASTAIR. 2023. *Measuring Tax Progressivity in Low-Income Countries*. World Bank Policy Research Working Paper #10460.

WAGER, STEFAN, DU, WENFEI, TAYLOR, JONATHAN, & TIBSHIRANI, ROBERT J. 2016. High-Dimensional Regression Adjustments in Randomized Experiments. *Proceedings of the National Academy of Sciences*, **113**, 12673–12678.

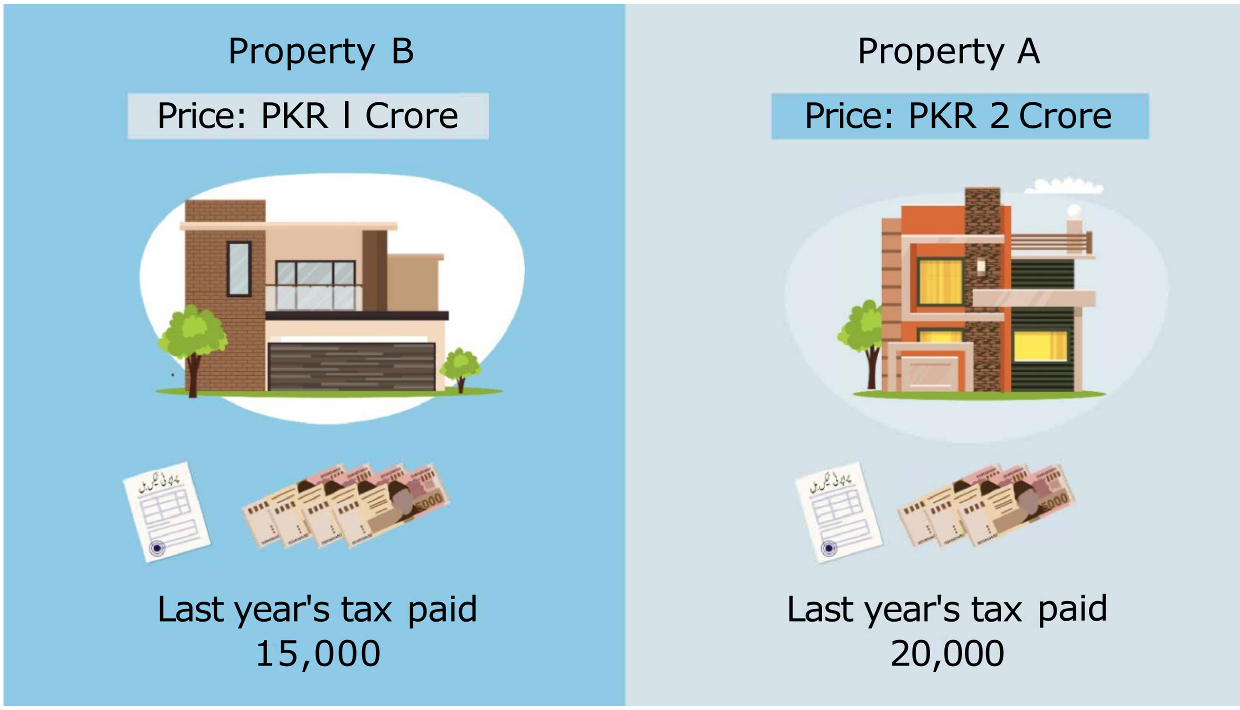
A Excise and Taxation Hierarchy

Figure A.1: Flowchart of Property Tax Administration in Lahore



B Treatment Vignettes and Exhibits

Figure B.1: Manipulation checks vignette for ATR comprehension



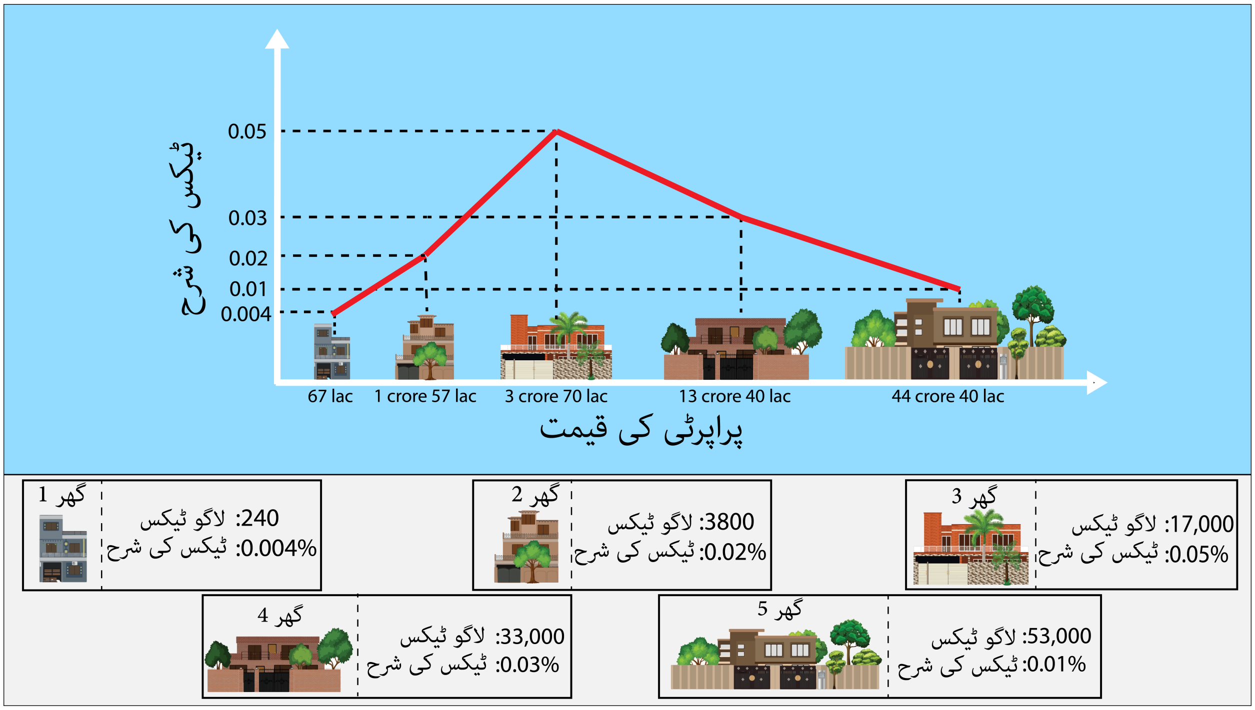
Source: IDEAS-LUMS Property Valuation Survey 2024-2025
Notes: Figure shows vignette to check respondents' comprehension of Average Tax Rate. The respondents were shown a video before this vignette and were then asked to calculate ATRs for both properties. The correct answer was for Property A was 0.1% and Property B was 0.15%. The purpose of the vignette was to show that even though Property A is paying a higher tax in rupees, but its tax burden (tax rate) was lower than that of Property B.

Figure B.2: Correcting Misperceptions Dashboard

Source: IDEAS-LUMS Property Valuation Survey 2024-2025

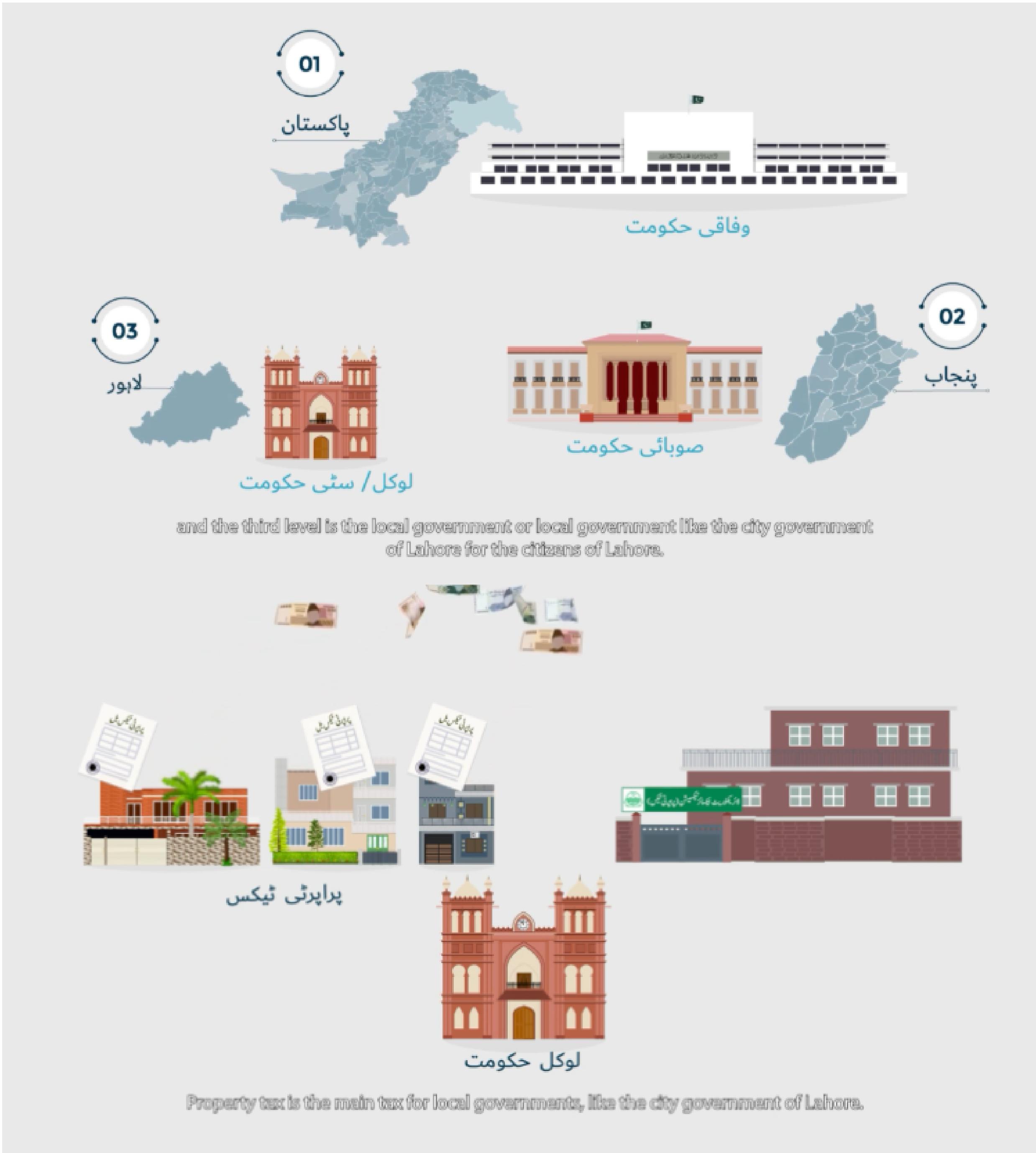
Notes: Figure shows screenshot from Android Application developed to capture respondents' reported tax liabilities before revealing actual tax liabilities of the 5 properties. For greater than 7 marlas properties' sample, we reduce the number of properties shown to 3 keeping property 3, 4 and 5 only.

Figure B.3: Correcting Misperceptions Treatment Exhibit



Source: IDEAS-LUMS Property Valuation Survey 2024-2025
Notes: Figure shows the exhibit which was shown to the respondents by enumerators. The figure shows property tax rates decreasing after an initial peak at mid-range property values, indicating a regressive tax pattern where higher-value properties have lower effective tax rates.

Figure B.4: Placebo Video Information Exhibit



Source: IDEAS-LUMS Property Valuation Survey 2024-2025
Notes: Figure shows an exhibit from the placebo informational video. The exhibit shows tiers of the government and property tax being the main revenue source for the third tier i.e. local government.

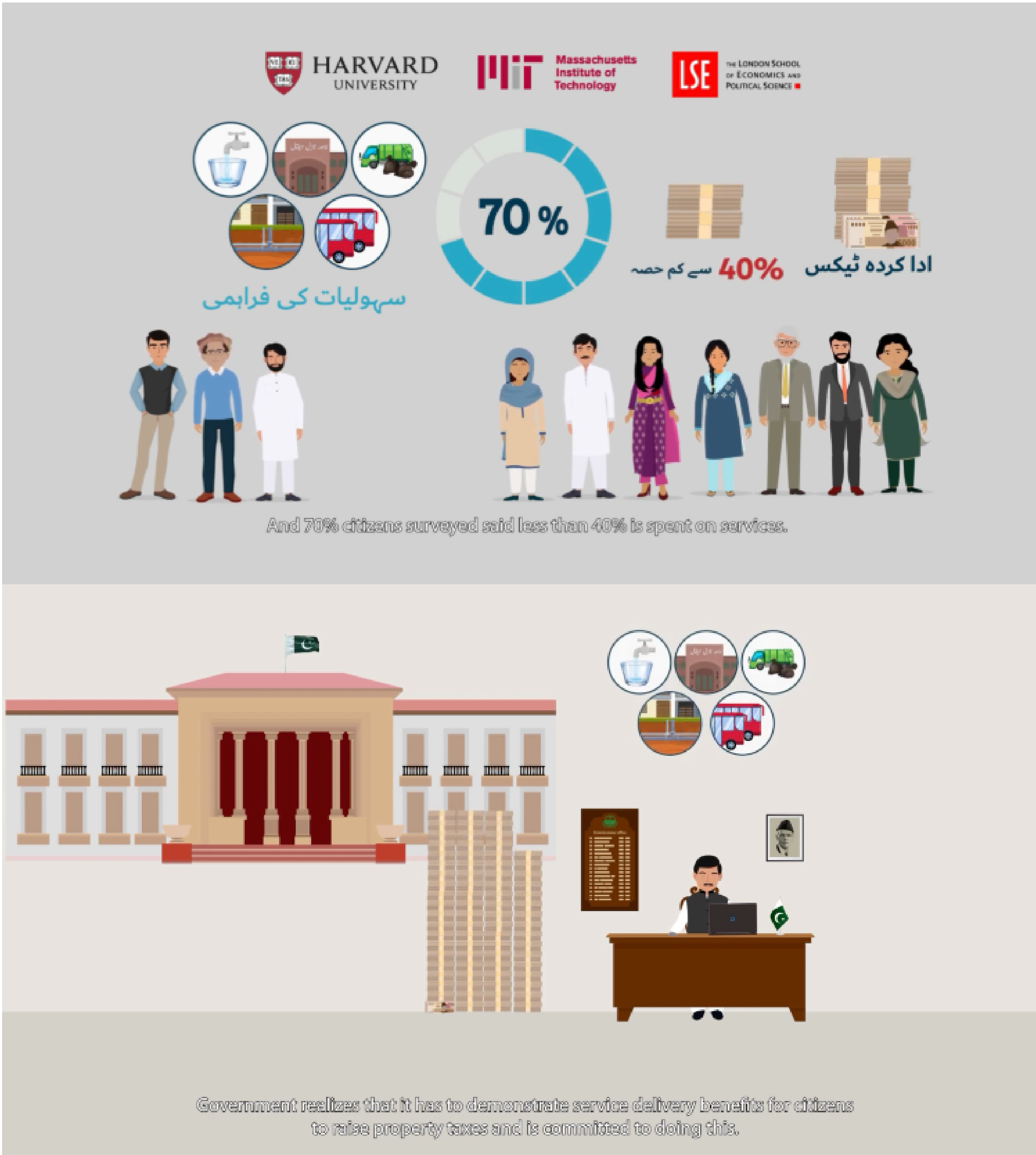
Figure B.5: Public Goods Treatment Video Exhibit

Source: IDEAS-LUMS Property Valuation Survey 2024-2025
Notes: Figure shows an exhibit from the Public Goods Treatment video. The exhibit shows low property tax utilization is a big constraint on the government's ability to meet the public service delivery needs of citizens.

Figure B.6: Revenue Leakages Treatment Video Exhibit

Source: IDEAS-LUMS Property Valuation Survey 2024-2025
Notes: Figure shows an exhibit from the Revenue Leakages Treatment video. The exhibit shows implications of low revenue collection on public goods service provision and ends with the message that raising adequate financing for local public good provision in the city will be difficult for government in the absence of improved compliance

Figure B.7: Spending Leakages Treatment Video Exhibit



Source: IDEAS-LUMS Property Valuation Survey 2024-2025
Notes: Figure shows an exhibit from the Spending Leakages Treatment video. The exhibit shows citizens' perception of utilization of revenue collected from property taxes and ends with a message reinforcing the message that local public good provision in the city will be difficult for government in the absence of measures that can strengthen tax reciprocity

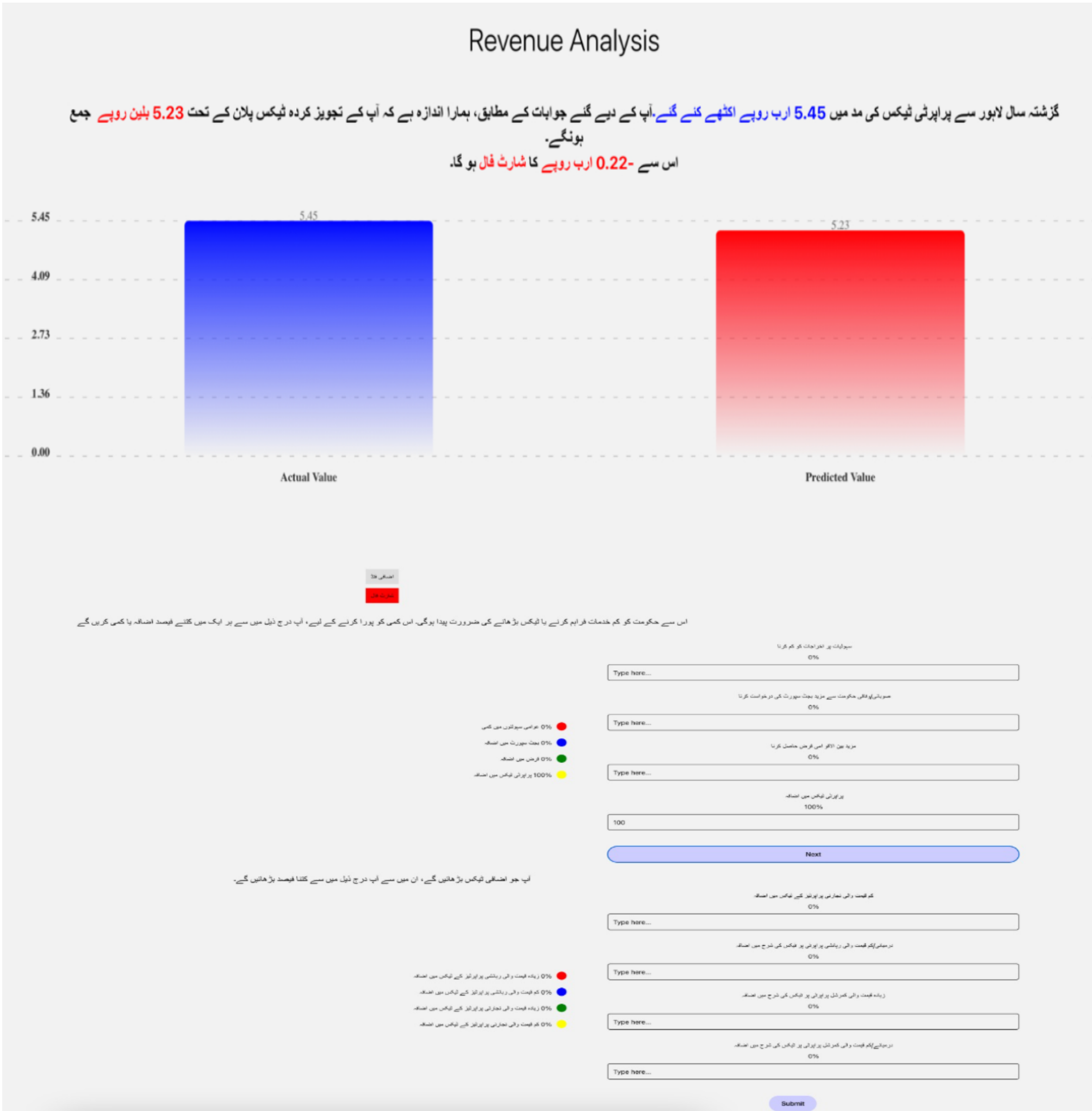
Figure B.8: Elite Capture Treatment Video Exhibit

Source: IDEAS-LUMS Property Valuation Survey 2024-2025
Notes: Figure shows an exhibit from the Elite Capture Treatment video. The exhibit shows recent cases where opposition from high value property owners in Lahore successfully delayed the introduction of reforms designed to raise more property taxes from the wealthy. It ends with the message that raising adequate financing for local public good provision in the city will be difficult for government in the absence of cooperation from the wealthy elite of the city.

Figure B.9: Preference Elicitation Dashboard

Source: IDEAS-LUMS Property Valuation Survey 2024-2025
Notes: Figure shows screenshot from Android Application developed for eliciting respondents' preferred tax schedule. Respondents were shown property characteristics along with a scale with a default value at Lahore's average tax rate of 0.04%. Respondents were then asked whether they preferred a higher or a lower tax rate than Lahore's average on this property and they the scale was adjusted based on respondents response (Panel A). Once all 9 properties were asked, the respondent was show a summary table (Panel B), followed by the shape of the curve based on their preferred schedule and whether their response generated a progressive, neutral or a regressive schedule (Panel C). Respondents were then given a chance to revise their preferred tax rate the updated graph was shown to them as well (Panel D).

Figure B.10: Revenue Calculations and Posterior Questions



Source: IDEAS-LUMS Property Valuation Survey 2024-2025

Notes: Figure shows screenshot from Android Application developed for eliciting respondents' preferred tax schedule. Based on the final responses for preferred tax schedule, respondents were shown how much their preferred tax schedule will raise in revenue. If the revenue was greater than the current revenue, the respondents were asked where they wanted to spend and if it was less than the current revenue, they were asked from where they will source the deficit from. These questions were not asked for greater than 7 marlas properties to reduce survey time as explained in Section 7

C Property Value Prediction

This appendix describes the procedures we followed to estimate predicted property values for all 802,000 properties in the Excise & Taxation cadaster. Section C.1 describes our survey with real estate agents to create the training data used for our estimation. Section C.2 describes how we impute localities for the parts of the cadaster missing locality information. Section C.3 describes the random forest algorithm and its performance.

C.1 Training Data: Real Estate Agent Valuation

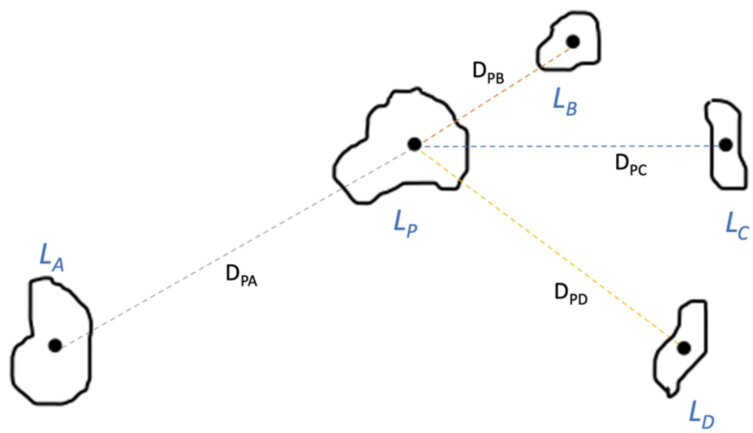
We get 2023 market value data from real-estate experts. For every property assessed, participating dealers were requested to provide estimates on the capital value, potential value as an open plot, and the rental value. Dealers were also asked about their own confidence levels in the reported values and their observations regarding property trends over the past six months, as well as their expectations for the next six months.

C.2 Imputing Missing Locality Data

One of the key inputs into our property valuation algorithm is the neighborhood a property is located in. However, the Excise & Taxation cadaster only contains the names of localities, not geocoded data on their location. To impute this missing data we follow the following procedue. Localities were categorized into one of four distinct groups.

- **Type I (TI):** This category included localities that were present in the valuation sample. Property geocodes obtained during the valuation exercise were utilized for these localities to determine a quasi-centroid in cases where the cadastre lacks geocodes. Once the quasi-centroids were obtained for all TI localities, the localities where the rates were missing were assigned DC residential and commercial rates from nearest possible locality using Mahalanobis distances

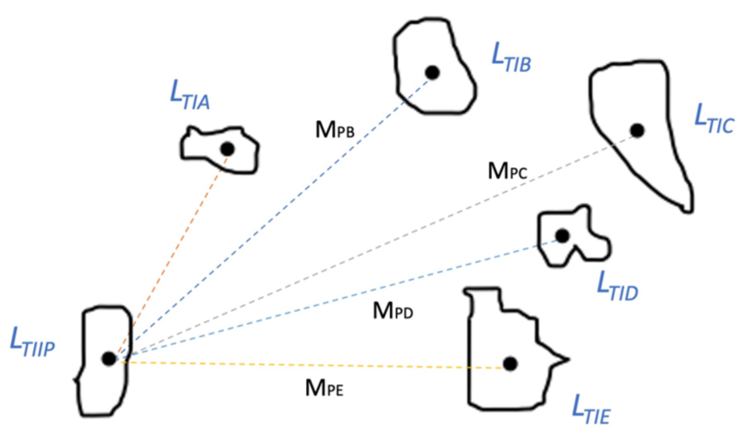
Figure C.1: Assigning DC values to a TI locality using location attributes



Source: IDEAS-LUMS Property Valuation Survey 2023
Notes: Figure shows if rates were missing from locality L_p, commercial and residential locality rates were assigned from locality L_B as its nearest to L_P. The distance was determined using Mahalanobis distance.

- **Type II (TII):** This category included localities which were not drawn in the main sample but had geo-codes and commercial and residential locality-level rates from the DC 2018-19 list. For each TII locality, Mahalanobis distances were computed for their proximity with all TI localities using location (i.e. longitude and latitude) and fanciness (i.e. DC residential and commercial rates) attributes. They were then linked to the closest TI locality (see Figure C.2) for the prediction model.

Figure C.2: Linking a TII locality to a TI locality using location and fanciness attributes

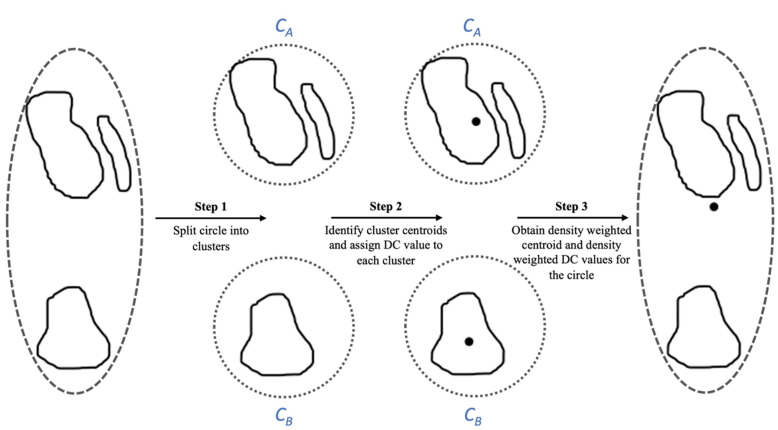


Source: IDEAS-LUMS Property Valuation Survey 2023
Notes: Figure shows out-of-sample L_{TIIP} locality was matched to L_{TIA} due to its proximity in terms of Mahalanobis distance.

- **Type III (TIII):** These types of localities were not drawn into the sample but had location geo-codes, and at least one of the residential and commercial DC rates was missing. For each TIII locality Mahalanobis distances were computed to all TI and TII localities using location attributes (i.e. longitude and latitude). Each TIII locality was then assigned DC rates of the closest TI or TII locality. If the TIII locality was assigned DC rates of a TI locality, then it was also linked to the same TI locality for the prediction model. Otherwise, location and "estimated" fanciness measures were used to link this TIII locality to a TI locality.

- **Type IV (TIV):** These localities were not drawn in the sample and did not have geocoded location information or DC rates. TIV localities were first split into two subtypes: a) locality lies in E&T defined circle that has geocodes and DC rates; and b) locality lies in a E&T defined circle that has no geocodes property and no DC rates. For a), missing information was filled using the strategy employed for TII localities. The only change was that circle centroids and average DC rates at the circle level were assigned to type a) localities. For localities from sub-sample b), 37 circle boundaries were plotted on QGIS. 14 of these circles were scattered around different parts of the city. It was decided that these circles would be (manually) split into clusters and cluster centroids and densities were computed using AsiaPop data. Each cluster was assigned DC rates of the closest TI, TII or TIII locality. These values were then computed using a density-weighted centroid and density-weighted DC residential and commercial rates for each circle (see Figure C.3). Mahalanobis distances were computed in the final step to link each type b) locality to a TI locality using density-weighted centroid and density-weighted DC rates.

Figure C.3: Dealing with a circle that has clusters in different parts of the city



Source: IDEAS-LUMS Property Valuation Survey 2023

C.3 Random forest property value data

This section details the procedures adopted to generate the random forest data. One source of estimating baseline levels of progressivity is administrative data. For this purpose, rental and capital market values for 2023 were obtained by surveying a sample of 12,363 commercial and residential properties from real estate experts. This 12,363 sample was then expanded to 802,592 properties using random forest. The random forest data was then merged with the property tax collection data obtained from E&T to create a unique dataset that contains information on 2022-2023 capital and rental market values and actual tax liabilities from FY 2021-2022.

The E&T property cadastre has 2,069 localities, of which only 407 were sampled. Predicting the property values for 1,662 out-of-sample localities was crucial because the neighbourhood is one of the key determinants of property value.

For this purpose, the location and average value of the locality were used to predict the property values. For location, an average of the property geocodes in the locality was taken to get a quasi-centroid (with latitude and longitude). Secondly, DC 2018-19 land rates (and not structure rates) served as a measure of fanciness for that locality. Both residential and commercial DC land rates were used to link the localities as they significantly differ even within a locality.

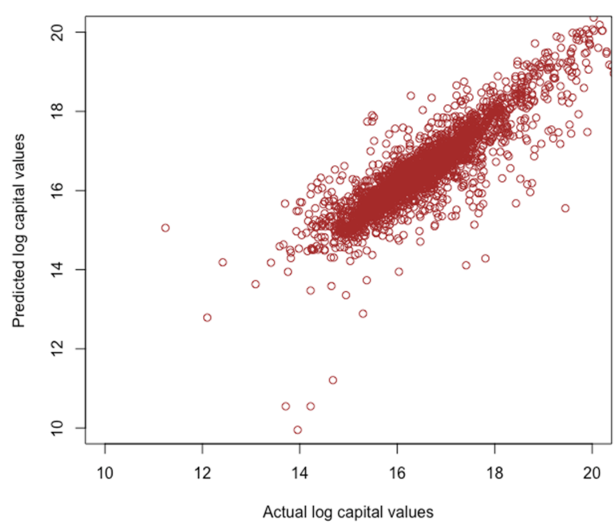
For the current context, property value is a function of land area (L), built area (B), residential use dummy (R), and a vector of cluster dummies (C) such that:

$$V = f(L, B, R, C)$$

To predict V, random forest model was set up where 75% of the data was used to predicted log of V using the logs of L, B, R and C. The remaining 25% of the sample was reserved for cross-validation, a technique used to assess the model's predictive performance on unseen data, thus providing insights into its generalizability.

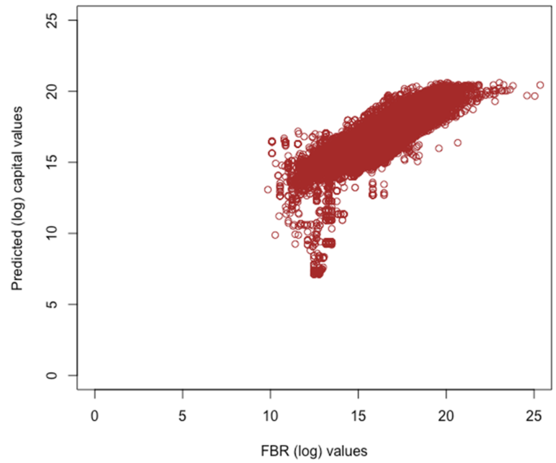
The results of this cross-validation (Figures C.4) show a high correlation between the predicted and actual capital values within the cross-validation sample. The results are robust with actual values as well. (See Figure B.5).

Figure C.4: Relationship between predicted and actual capital values



The entire valuation sample was then used to train the random forest model and predictions were made for the full valuation sampling frame where we had FBR values. Figure C.5 shows that the correlation between predicted values and FBR values was positive but not as strong as with the cross-validation sample in Figure C.4.

Figure C.5: Relationship between predicted and FBR capital values for the valuation sampling frame

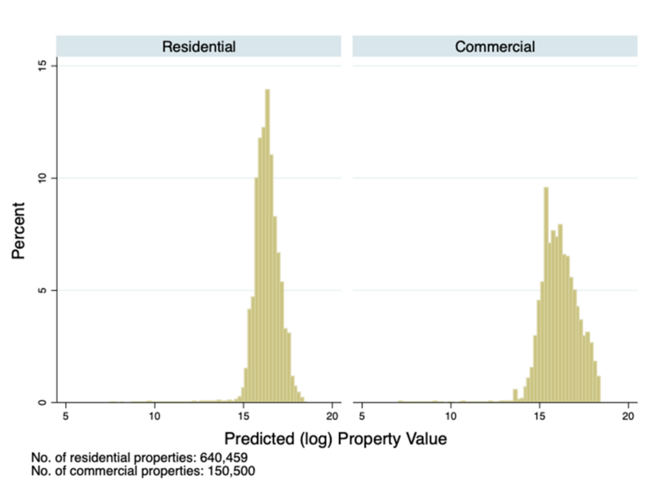


Source: IDEAS-LUMS Property Valuation Survey 2023

The final step in this process was to set up a random forest model and predict values for all residential and commercial properties in the cadastre. This was done by fixing the number of trees to 100 in the final specification and the number of variables used at each split to 2.

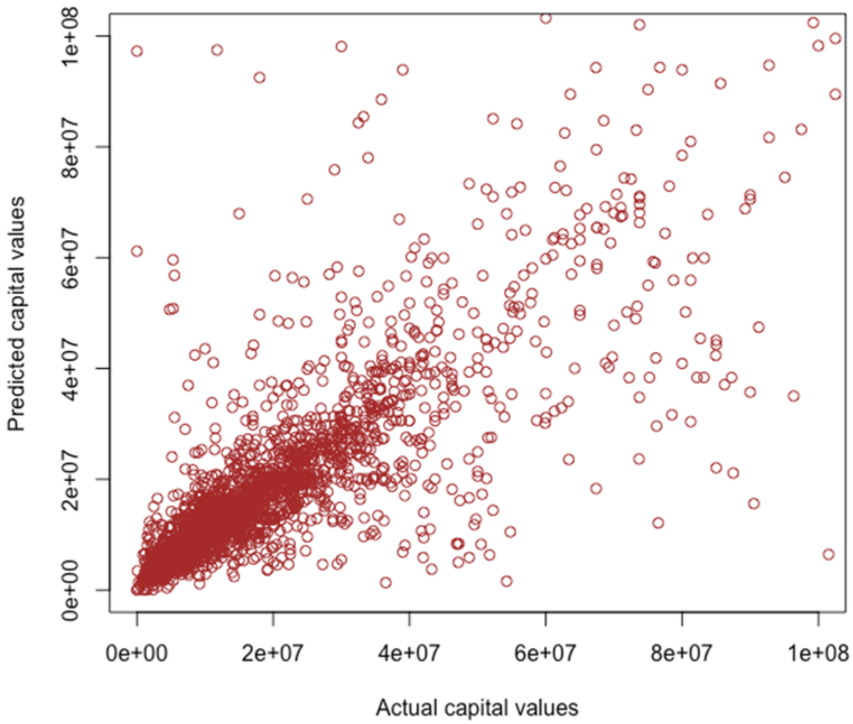
As expected, both residential and commercial property value distributions are right-skewed with few very highly valued properties (see Figure C.6).

Figure C.6: Distributions of predicted property values by property use



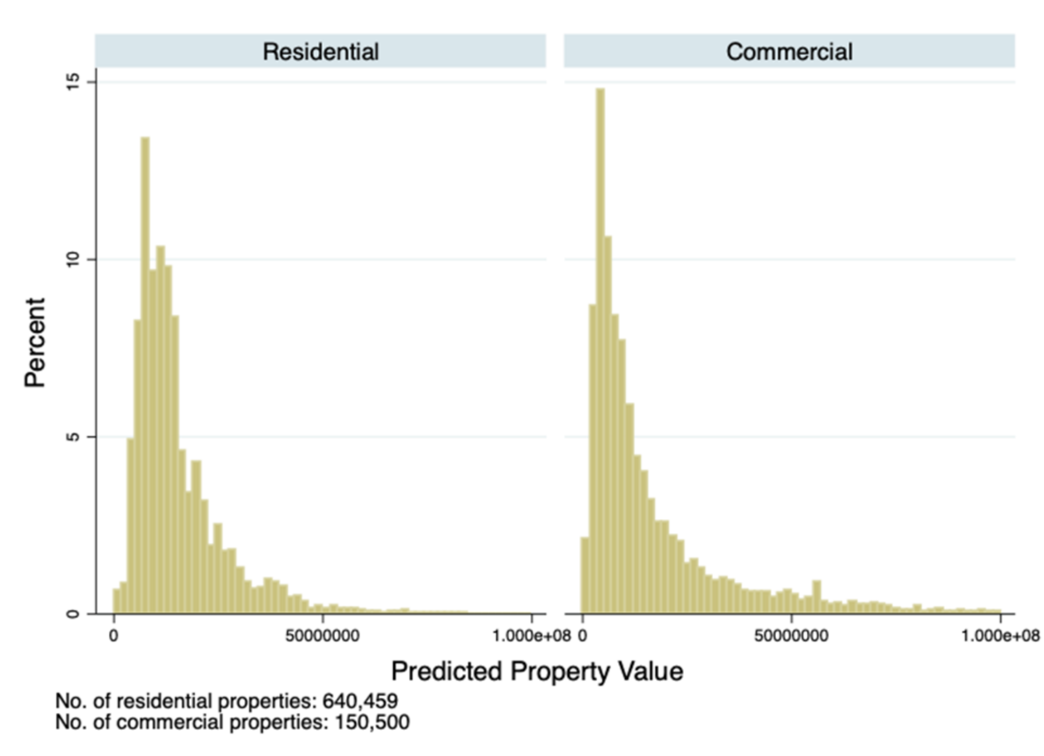
Source: IDEAS-LUMS Property Valuation Survey 2023
Notes: Values restricted to PKR 800 million for better visualization

Figure C.7: Relationship between predicted and actual (log) capital values for the cross-validation sample



Source: IDEAS-LUMS Property Valuation Survey 2023

Figure C.8: Distributions of predicted (log) property values by property use



Source: IDEAS-LUMS Property Valuation Survey 2023
Notes: Values restricted to PKR 800 million for better visualization

D Total Revenue from Survey Responses

This section show how we calculated the revenue value from preferred ATR from survey responses. The calculated revenue value caters for the tax demand from the cadaster as well as the tax complaince.

D.1 Setup

We aim to take the 9 survey responses on desired ATRs for 9 properties and return an estimate of the total revenue raised by that tax schedule using the following setup.

Each property i is of one of 2 property types:

1. Residential, self-occupied: $t(i) = (r, s)$
2. Residential, rented: $t(i) = (r, r)$

The respondents j are randomly assigned one of these property types: $t(j) \in \{(r, s), (r, r)\}$. They see 9 properties from this property type (and none of the other type).

Properties also fall into 3 type-specific value strata:

1. $\mathcal{S}_1^t$: The property's log value $v_i \leq k_1^t$. Where k_1^t is the 50th percentile of the distribution of residential self-occupied log property values.
2. $\mathcal{S}_2^t$: $k_1^t < v_i \leq v_2^t$. where k_2^t is the 90th percentile of the distribution of residential self-occupied log property values.
3. $\mathcal{S}_3^t$: $k_2^t < v_i$.

D.2 Restricted Spline

Using respondent j's survey responses, we ran a 3-piece linear spline of the ATR y_i^j on the log of the property value v_i. This can be done in two steps.

1. Create 3 variables:
   (a) $v_{1i}^{t(j)} = \min\{v_i, k_1^{t(j)}\}$
   (b) $v_{2i}^{t(j)} = \max\{\min\{v_i, k_2^{t(j)}\}, k_1^{t(j)}\} - k_1^{t(j)}$
   (c) $v_{3i}^{t(j)} = \max\{v_i, k_2^{t(j)}\} - k_2^{t(j)}$

   where we
2. Run a restricted OLS regression of y_i^j on a constant, $v_{1i}^{t(j)}$, $v_{2i}^{t(j)}$, and $v_{3i}^{t(j)}$:

$$y_i^j = \beta_0^j + \beta_1^j v_{1i}^{t(j)} + \beta_2^j v_{2i}^{t(j)} + \beta_3^j v_{3i}^{t(j)} + \varepsilon_i^j$$

   with the restrictions that it should never predict an ATR lower than 0 or higher than 100

With the $\hat{\boldsymbol{\beta}}^j$ in hand, we compute predicted ATRs for any property:

$$\hat{y}_i^j\left(v_i\right) = \hat{\beta}_0^j + \hat{\beta}_1^j v_{1,i}^{t(j)} + \hat{\beta}_2^j v_{2,i}^{t(j)} + \hat{\beta}_3^j v_{3,i}^{t(j)} \tag{D.1}$$

Since the spline is piecewise linear, to impose the restrictions that it never predicts less than 0 or more than 100, we imposed the restrictions at each of the knots and at the extremes. Specifically, there are 8 constraints:

1. $\hat{y}(\underline{v}) = \hat{\beta}_0 + \hat{\beta}_1 \underline{v} \geq 0$
2. $\hat{y}(\underline{v}) = \hat{\beta}_0 + \hat{\beta}_1 \underline{v} \leq 100$
3. $\hat{y}(k_1) = \hat{\beta}_0 + \hat{\beta}_1 k_1 \geq 0$
4. $\hat{y}(k_1) = \hat{\beta}_0 + \hat{\beta}_1 k_1 \leq 100$
5. $\hat{y}(k_2) = \hat{\beta}_0 + \hat{\beta}_1 k_1 + \hat{\beta}_2 (k_2 - k_1) \geq 0$
6. $\hat{y}(k_2) = \hat{\beta}_0 + \hat{\beta}_1 k_1 + \hat{\beta}_2 (k_2 - k_1) \leq 100$
7. $\hat{y}(\bar{v}) = \hat{\beta}_0 + \hat{\beta}_1 k_1 + \hat{\beta}_2 (k_2 - k_1) + \hat{\beta} (\bar{v} - k_2) \geq 0$
8. $\hat{y}(\bar{v}) = \hat{\beta}_0 + \hat{\beta}_1 k_1 + \hat{\beta}_2 (k_2 - k_1) + \hat{\beta} (\bar{v} - k_2) \leq 100$

where $\underline{v}$ is the lowest (log) property value for which we need to predict and $\bar{v}$ is the highest (log) property value we need to predict.

The `restriktor` package in `R` lets us implement this. But it wants the constraints in the syntax $\mathbf{R}\hat{\boldsymbol{\beta}} \geq rhs$ which in our case is

$$\begin{pmatrix} 1 & \underline{v} & 0 & 0 \\ -1 & -\underline{v} & 0 & 0 \\ 1 & k_1 & 0 & 0 \\ -1 & -k_1 & 0 & 0 \\ 1 & k_1 & (k_2 - k_1) & 0 \\ -1 & -k_1 & -(k_2 - k_1) & 0 \\ 1 & k_1 & (k_2 - k_1) & (\bar{v} - k_2) \\ -1 & -k_1 & -(k_2 - k_1) & 0 \end{pmatrix} \hat{\boldsymbol{\beta}} \geq \begin{pmatrix} 0 \\ -100 \\ 0 \\ -100 \\ 0 \\ -100 \\ 0 \\ -100 \end{pmatrix} \tag{D.2}$$

D.3 Total Revenue Estimates

We estimate how much total revenue would be raised by the respondent's preferred tax schedule. To do this we have to overcome three challenges:

1. We only ask the respondents about 9 properties, so we need to extrapolate to all other properties.
2. We only ask the respondents about one type of property (either residential self-occupied or residential rented).
3. Not all of the tax demanded from households is actually paid i.e. less than 100% tax payment compliance.

We will go through these one by one:

D.3.1 Tax demand for "my" property type

Each respondent is either type $t(j) = (r, s)$ or type $t(j) = (r, r)$ and their type determines the cutoffs of the value bins (low/middle/high) for them. In general, the tax demand for the respondent's property type is

$$R_1^j = \sum_{t(i)=t(j)} \hat{y}_i^j v_i \tag{D.3}$$

That is, we sum across all properties whose type $t(i)$ is the same as that assigned to the respondent ($t(j)$). For each property, we use the spline predictions of the ATR $\hat{y}_i^j$ and the log value of the property v_i to predict the tax demand for that property.

We compute this for all the properties in the cadaster and then add them all up to get total revenue via a shortcut as follows: note that in equation (D.3), we can break open the predicted ATR:

$$\hat{y}_i^j v_i = \left(\hat{\beta}_0^j + \hat{\beta}_1^j v_{1,i}^{t(j)} + \hat{\beta}_2^j v_{2,i}^{t(j)} + \hat{\beta}_3^j v_{3,i}^{t(j)}\right) v_i \tag{D.4}$$

and so,

$$\begin{aligned} R_1^j =& \hat{\beta}_0^j \sum_{t(i)=t(j)} v_i + \hat{\beta}_1^j \sum_{t(i)=t(j)} v_{1,i}^{t(j)} v_i + \hat{\beta}_2^j \sum_{t(i)=t(j)} v_{2,i}^{t(j)} v_i + \hat{\beta}_3^j \sum_{t(i)=t(j)} v_{3,i}^{t(j)} v_i && \text{(D.5)} \\ =& \left[\hat{\beta}_0^j V^{t(j)} + \hat{\beta}_1^j V_1^{t(j)} + \hat{\beta}_2^j V_2^{t(j)} + \hat{\beta}_3^j V_3^{t(j)}\right] && \text{(D.6)} \end{aligned}$$

where

$$\begin{aligned} V^{t(j)} &= \sum_{t(i)=t(j)} v_i && \text{(D.7)} \\ V_1^{t(j)} &= \sum_{t(i)=t(j)} v_{1,i}^{t(j)} v_i && \text{(D.8)} \\ V_2^{t(j)} &= \sum_{t(i)=t(j)} v_{2,i}^{t(j)} v_i && \text{(D.9)} \\ V_3^{t(j)} &= \sum_{t(i)=t(j)} v_{3,i}^{t(j)} v_i && \text{(D.10)} \end{aligned}$$

Note that we pre-compute all of the sums in (D.7)–(D.10). There are two types of properties and four numbers, so this is eight "V" numbers. Finally, we multiply the "V" numbers by our coefficients and add them to get $R_1^{j,t(j)}$.

D.3.2 Other property types

We also need to deal with the fact that we only asked people about one property type $t(j) \in \{(r,r), r(s)\}$. Other property types also have a share in the total revenue which should be catered in our revenue estimation. This means we will have to scale up our revenue estimate from the proportion of revenue calculated from two property types used in survey questions. To cater for other property types, we use the cadaster.

We compute the total tax demand of the two property types we ask respondents about:

$$D_1^\tau = \sum_{t(i)=\tau} d_i \tag{D.11}$$

where d_i is the tax demanded from property i and $\tau \in \{(r,r),(r,s)\}$ are the two property types we are interested in. Then, also calculate total tax demanded from the entire cadaster

$$D = \sum_{i=1}^{N} d_i \tag{D.12}$$

and use these to scale our revenue estimate up:

$$R_2^j = R_1^j \frac{D}{D_1^{t(j)}}. \tag{D.13}$$

D.3.3 Noncompliance

Finally, to deal with the fact that not all of the tax demanded is actually paid, we again use the aggregate amounts to scale the revenue estimate. For this get the total amount paid by everyone from the cadaster:

$$T = \sum_{i=1}^{N} t_i \tag{D.14}$$

and we use this to scale our revenue estimate down:

$$R^j = \frac{T}{D} R_2^j = \frac{T}{D_1^{t(j)}} R_1^j \tag{D.15}$$

D.4 Dealing with Non-response of 9 "main" properties

In what's outlined above, we are assuming that each respondent answers 9 preference elicitation questions. But they may refuse/not know how to answer for all 9. If there is a non-response, we continue until we have asked all 9 properties. At the end of the 9 properties, we check how many responses we have for the various types of properties the respondent saw.

The respondent is randomly assigned a type $t(j) \in \{(r,s),(r,r)\}$. For that type, they saw 9 properties: 3 low-value, 3 medium-value, and 3 high-value. If

1. They gave a response for ≥ 1 low-value property; AND
2. They gave a response for ≥ 1 medium-value property; AND
3. They gave a response for ≥ 1 high-value property

Then

1. fill in the missing responses with the average of the obtained responses in that category (low/medium/high-value).
2. Continue as normal

If they do not satisfy all 3 criteria above, abort the revenue estimation.