

Can Mafia Be Good? Evidence from Karachi's Urban Water Market

3.1 Introduction

Clean drinking water is essential. United Nation's Sustainable Development Goal 6 emphasizes universal accessibility and affordability of drinking water, yet many in the developing world lack this basic right. Challenges in the provision of drinking water persist due to supply-side constraints and corruption, especially in low-income communities (Beard & Mitlin, 2021). Often, weak governance and limited institutional capacity lead to inadequate piped infrastructure and intermittent water supplies. With no or restricted access to piped water, people look for other alternative sources of water. Such circumstances pave way for non-state actors to become involved in the distribution and delivery of water resources. Non-state actors procure water legally and illegally. Procuring water illegally has given rise to the water "mafia". The mafia often engages in exploitative behavior by stealing water and selling it at higher prices (Sajjad, 2019). However, they operate in neighborhoods with no other reliable drinking water source.

Economic theory favors government provision of public goods, arguing that consumers are often worse-off when private or illegal actors enter into the provision of such goods. This might be true with the water mafia that charges exorbitantly higher prices to poor and marginalized groups in developing countries. As a result, these groups end up consuming lower quantities of water that further exacerbate the inequalities prevailing in access to drinking water. However, as the burden of unreliable piped water supply often falls disproportionately on these same poor and marginalized groups, the mafia acts as the only reliable source of clean drinking water; in the mafia's absence these communities are forced to switch to lower quality water sources. While the mafia exploits poor and marginalized groups by charging higher prices, it simultaneously enhances their well-being by providing a reliable water supply. However, it remains unclear whether the benefits outweigh the exploitation. Without a thorough understanding of which side has a greater impact, ambitious policies that propose to curtail the activities of the mafia can adversely impact consumers' well-being.

In this paper, I aim to estimate the net social welfare impact of the water mafia by simulating a policy intervention that curtails the operations of large private tankers. Using findings from a novel survey instrument implemented in Karachi, Pakistan, I propose to use a discrete-continuous model (DCC) to estimate residual water demand — demand leftover after piped water supplies are exhausted. The discrete component of the model captures households’ choices among three mutually exclusive alternatives: government tankers, private tankers, and other informal sources. The continuous component models the quantity of water consumed from the selected source. The proposed DCC framework allows for consistent estimation of the factors influencing both the source selection and the quantity demanded. Next, I plan to simulate a policy intervention that eliminates large private tankers (“the water mafia”) from the market. This intervention is expected to (1) return all water previously diverted by these tankers to the piped water system, increasing overall piped supply, and (2) force households to switch to potentially more expensive or inconvenient informal sources for their residual water needs. The overall impact of policy will be quantified by aggregating changes in consumer surplus across all households, relative to the status quo. Through this approach, I will be able to obtain the first-ever, robust welfare estimate of informal water vendors in a developing country context.

Informality and its implications have been widely studied in other sectors. For example, numerous studies have examined the welfare consequences of electricity theft on both, individual households and utility companies (Smith, 2004; Balza et al., 2013). The literature generally finds that policies aimed at reducing theft and formalizing electricity supply tend to increase revenues and profits for utilities. However, these policies can have unintended adverse effects on consumers—particularly poor households—who may be forced to reduce their electricity consumption in response to higher bills (Ahmad et al., 2024). The transition from informal to formal access can therefore result in significant welfare losses for vulnerable groups.

Surprisingly, there is a notable gap in the literature on the welfare implications of informality in water markets, despite the widespread prevalence of informal and illegal water provision in developing countries (Felbab-Brown, 2017). While the critical role of informal vendors in context of state failure is widely recognized, existing research on informal water vendors is largely descriptive, focusing on mapping vendor operations (Klassert, et al., 2023; Raina et al.,

2019), or on empirical estimation of water demand in the presence of multiple water sources (Cheeseman et al., 2008; Coulibaly et al., 2014; Nauges & Berg, 2009).

What remains missing from the literature are: (1) studies that explicitly acknowledge and examine the illegality or informality of alternative water sources, and (2) rigorous analyses of the distributional welfare impacts—both positive and negative—of informality or of policies designed to eliminate informal water provision. Rigorous welfare analysis of informal water vendors is rare because the required micro-data—prices, quantities, timing, reliability, and spatial access to both formal and informal sources— is typically unavailable, unreliable or difficult to obtain.

Moreover, a credible estimation of welfare effects requires a structural modeling framework grounded in microdata. Only a structural approach can account for household heterogeneity in decision-making across multiple water sources and accurately predict both the aggregate and distributional welfare effects of policies targeting informal vendors. The only study that approximates a welfare analysis of policy interventions in informal water markets is Klassert et al. (2015), which employs an agent-based model to simulate the welfare consequences of stricter regulation in the tanker water market via higher tanker water prices. While their results show that higher tanker prices reduce welfare for all households (as tanker demand falls but piped supply remains unchanged), their approach relies on parameters sourced from the literature rather than microdata and does not account for the feedback effects that a reduction in tanker demand may have on piped water supply (since both draw from the same system).

With this paper, I aim to address the gaps in the previous literature. First, to answer the questions raised in this research, I collect a novel dataset that captures the full complexity of households' water purchasing behaviors in a highly informal urban water market. This survey data directly measures informal water purchasing patterns—including source choices, prices, quantities, reliability, and switching behavior—enabling empirical analysis that was not previously possible in the literature. Second, to my knowledge, this is the first study to provide a rigorous, structural estimate of the welfare effects of informal water vendors (“the water mafia”) in a developing country. Through the proposed approach, I move beyond descriptive or agent-based models widely prevalent in the water demand literature when multiple sources are available. Adopting a

structural modeling framework allows me to capture heterogeneity in distributional impacts of policy simulations and provide robust welfare estimates at the household level. The findings from this paper can offer critical new insights for designing policing aimed at improving provision of water in rapidly growing cities facing infrastructure and governance constraints. Lastly, this proposed research is extremely well-timed as climate change is expected to exacerbate water scarcity and intensify shortages in urban areas. Under such circumstances, the role of informal water vendors is likely to become even more significant. To ensure that effective policies are in place to address these evolving realities, it is imperative to rigorously assess the welfare impacts of informal water vendors.

The remainder of this research proposal is organized as follows. Section 3.2 provides background on Karachi's urban water market. Section 3.3 details the sampling and data collection strategy. Section 3.4 presents further descriptive results from the survey data, motivating the core research questions. Section 3.5, I describe the discrete-continuous choice model and econometric model that will be used to estimate residual water demand and outline the proposed policy simulation designed to estimate the welfare impacts of the water mafia, and Section 3.6 concludes.

3.2 Background

There are many challenges present in the distribution of residential water in Karachi. Firstly, the government water board, Karachi Water and Sewerage Board (KWSB) charges fixed lump-sum water tariffs based on the size of the establishment rather than water usage. These tariffs not only inaccurately reflect water usage but are also heavily subsidized as provision of water remains a relatively politicized issue. Municipality water rates in Karachi are reported to be as low as \$0.9 to \$2.3 per month (Khan & Arshad, 2022). Furthermore, bill payment remains relatively low in Karachi. 40% of KWSB customers do not pay their bills and many others use unmetered water (Auditor-General of Pakistan, 2019). The inability of KWSB to charge and collect water bills from its customers has led to inadequate revenue generation for the institution, with revenues being as low as 15-20% (Kunbhar, 2016). Insufficient revenues act as major barrier in reforming the outdated and inadequate water infrastructure in Karachi. Majority of Karachi's water network has passed its designed economic life. The aging water infrastructure results in high water losses, where around 25% of water is lost in leakages before it reaches the residents (Kunbhar, 2016).

These water losses further add to water scarcity that is already present due to natural and climatic factors. Moreover, due to the city's unplanned expansion, only 80% of homes in the city are connected to piped water. Informal settlements, known as "katchi-abadis," are completely disconnected from the piped water infrastructure (Kunbhar, 2016).

As KWSB failed to provide water to all the residents in an adequate and equitable manner, the local government decided to enter into a partnership with the private sector to engage the sector in the supply and distribution of drinking water. The government decided to auction water hydrants, that were placed on KWSB's piped infrastructure, to private companies (Abidi, Ahmed, & Shah, 1998). Currently, there are six KWSB owned hydrant in Karachi.

Winners of the auction would become owners of the hydrants and would be responsible for pumping water flowing through KWSB's pipelines and supplying it in tankers to the residents. The main purpose of entering into this partnership was to ensure supply of drinking water in neighborhoods that are disconnected from the piped infrastructure. Winners of the auction were required to abide by government's terms of supplying water. Under the terms of the contract, the winners were to supply the daily prescribed amount of water at the prices set by KWSB and in neighborhoods stated in the contract (KWSB, 2020).

Auctioning hydrants to private sector has introduced an informality in the urban water market in Karachi. The owners started selling water informally by going over the amount allocated to them for water supplies. The informal water is sold at rates exorbitantly higher than the water rates set by KWSB in the auction contract. It is reported that almost 10 million gallons of water is stolen on a daily basis, that is worth \$540 million annually (Kunbhar, 2016). In fact, 20% of water in Karachi is supplied informally. Many other small private players have also entered the informal urban water market and started stealing water from underground water pipelines owned by KWSB. However, due to the characteristics of the owners of the hydrants, these are large companies with at least 60 tankers (KWSB, 2020), the major share of informal sales is supplied by the hydrant owners. Moreover, evidence indicates that many smaller tankers/trucks/carts owners purchase water from these hydrant owners and sell it in neighborhoods (Ahmed N. , 2008).

Households participating in Karachi's urban water market face three main alternatives of drinking water supply, piped water, formal tanker water, and informal tanker water. The preferred source of water for most households remains piped water as it is drastically cheaper than any other source of drinking water. However, piped water remains unreliable, irregular, and not accessible to all. Under such circumstances, households either rely entirely on tanker water or augment piped water with tanker water supply to meet their daily water requirements. The choice between formal and informal tanker water depends on the household's access to the sources. Due to the location, some households do not have any access to formal tankers which mainly operate in areas near the hydrant.

3.3 Sampling Strategy & Data

The household decision-making process can be summarized as follows: households receiving sufficient quantities of piped water decide only on how much piped water to consume. In contrast, households with insufficient or no piped water must decide on the alternative water source to purchase from and the quantity to purchase from that source. The available alternative sources include government tankers, private tankers, and smaller informal sources such as pick-up trucks and drums. These sources vary significantly in price, with government tankers being

the least expensive and informal sources generally being the costliest. Consequently, the aggregate water demand curve for households in Karachi exhibits a step-function pattern.

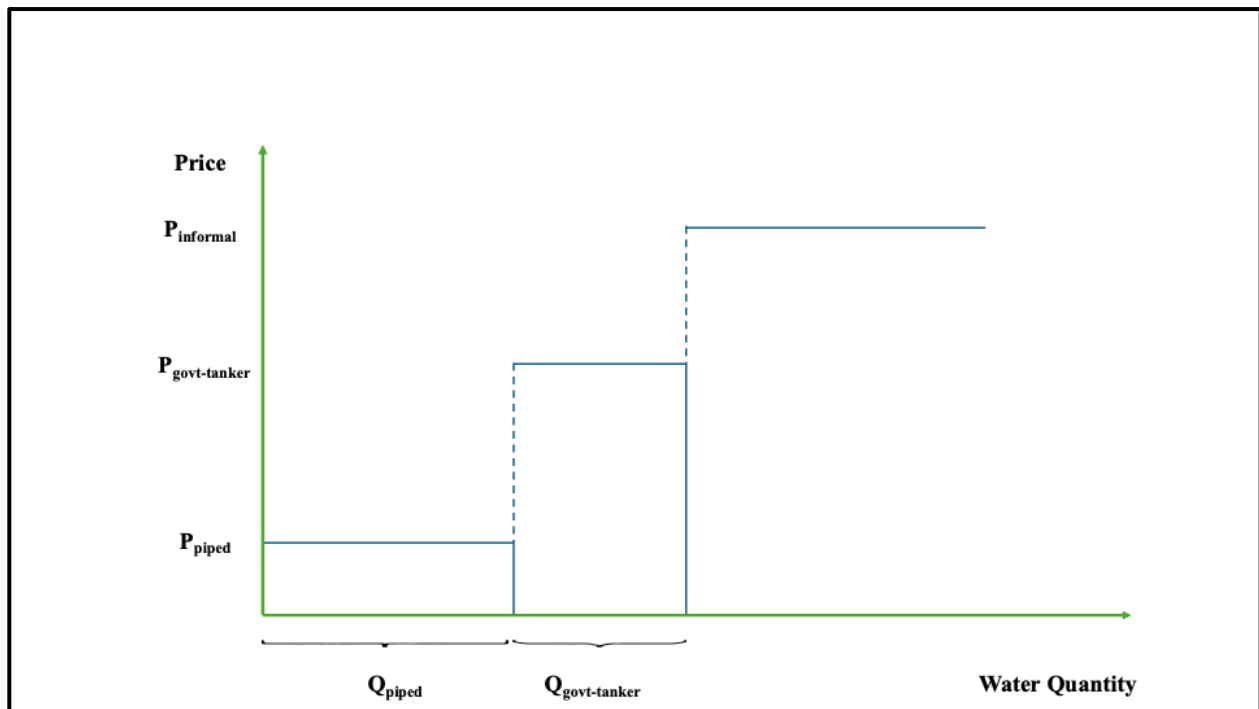


Figure 1. Aggregate Step-Function Water Demand Curve. This figure illustrates the step function nature of aggregate demand curve in Karachi’s urban water market. The cheapest water source in the market is piped water. However, it is irregular and in limited supplies. Households consume all the amount of piped water they receive and then meet the residual demand with one of the available alternatives. Government tankers are the cheapest but in limited supply. Private tanker and other informal sources are much more expensive.

To address the research questions raised in this paper, the data sample must include three types of households:

1. Households relying solely on tanker water,
2. Households relying solely on piped water,
3. Households that augment piped water with tanker water.

Additionally, representation of all three types of alternative water sources is required in the collected data sample.

Since no prior research in Karachi has identified neighborhoods with reliable access to piped water or those fully dependent on tanker water and its different sources, I adopt a mixed sampling strategy to achieve the desired sample composition.

Due to timing and budget constraints, I decided to focus on implementing the survey effort in one district in Karachi. Using purposive sampling, I chose Korangi district served by Landhi hydrant—a major filling station for both government and private tankers. I specifically chose to work in Korangi district because a previous study on willingness to pay for improved piped water services highlighted that this district exhibits considerable variation in water access—some households have piped water, while others rely solely on tanker water (Khalil et al., 2023). This diversity made Korangi an ideal location to study the full spectrum of water sourcing behaviors relevant to this research.

Within the areas served by the Landhi hydrant, I employed a two-stage sampling strategy. First, I randomly selected two towns—Landhi and Korangi Town—using simple random sampling. Each town in Karachi is divided into smaller union councils. I used mobilizer-assisted purposive sampling to select 3 union councils in each town and select neighborhoods to interview in each union council. Within each selected union council, neighborhoods were carefully chosen to provide comprehensive geographic coverage. Figure 2 below shows the spatial distribution of the sample collected.

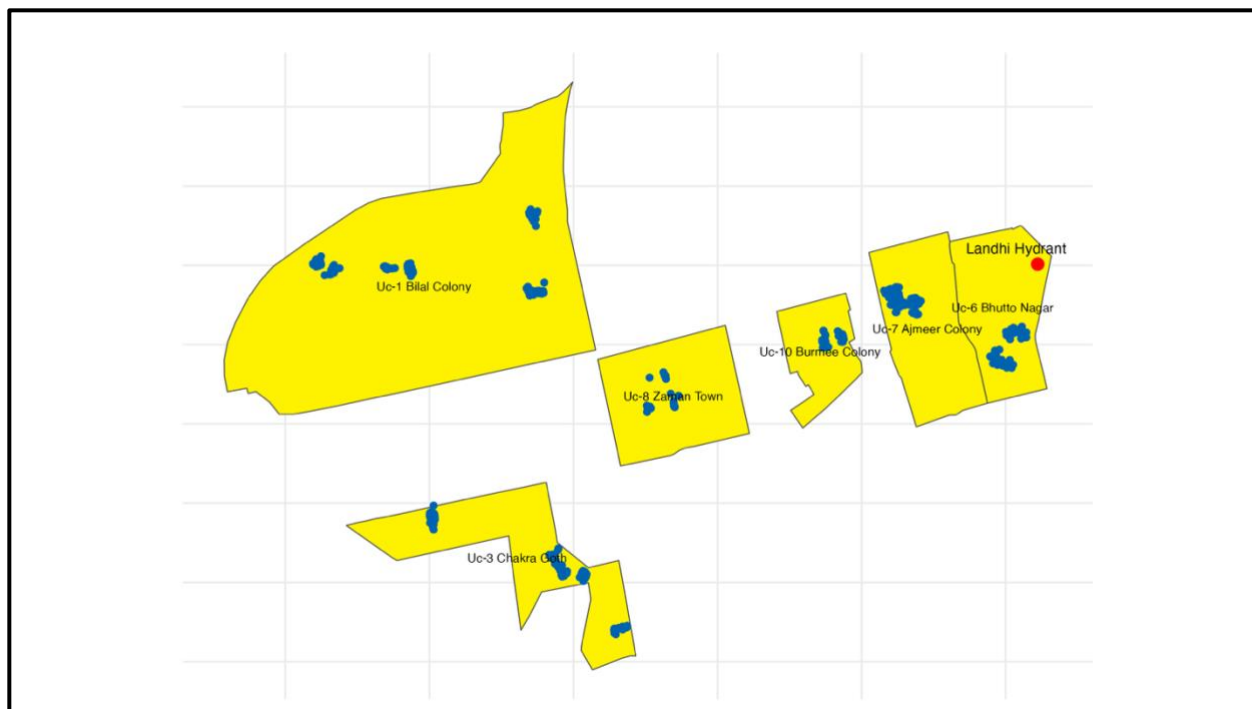


Figure 2. Spatial Distribution of Study Area and Interview Locations

Using this sampling approach, data were collected from a total of 459 households. Descriptive statistics are provided in the table below.

Table 1 Descriptive statistics from Survey Data

Variable Name	
% in Korangi Town	53
Average Age (years)	41
% of Female Respondents	51
% illiterate	28
% Daily Wager	22
% of Lowest Income	22
% of High Income	15
UC Based Sample Distribution (%)	
UC-1 Bilal Colony	26
UC-10 Burmee Colony	9
UC-3 Chakra Goth	23
UC-6 Bhutto Nagar	15
UC-7 Ajmeer Colony	24
UC-8 Zaman Town	5

Table 1 shows the descriptive statistics from the survey data. The sample collected is well-balanced with good representation of households across towns and UCs. There is also decent representation of households across income groups.

There are several limitations of my sampling strategy. First, the focus of my study is only on Korangi district in Karachi. The results from this study may not be generalizable to other districts or for whole of Karachi. Second, there may be potential bias in selecting union councils as the selection may be based on mobilizer's prior knowledge. Some referral bias may persist due to reliance on field contacts and community networks during household identification. Last, cross-sectional data does not capture seasonal or temporal trends in purchasing or water use behaviors.

It is important to acknowledge that the sampling strategy does not produce a representative sample of Karachi's entire population. However, given the novelty of this research and the absence of prior detailed data on neighborhoods' access to piped and tanker water sources, this purposive and mixed sampling approach was necessary to capture the diverse household water sourcing behaviors relevant to this study. This approach prioritizes capturing purchasing behaviors across different formal and informal water sources, enabling me to answer critical research questions that have, to date, remained unexplored in this context. Future research can build on these findings with more representative samples as baseline information becomes available.

For this study, I designed a comprehensive and novel household survey questionnaire specifically tailored to the research questions posed in this paper. The questionnaire captures detailed information on all sources of household water consumption understanding households' decision-making processes when faced with formal and informal water sources.

First, the survey includes an extensive module dedicated to piped water usage, covering aspects such as availability, supply frequency, and billing information. Collecting this detailed data is critical, as it helps establish thresholds of piped water insufficiency that drive households to participate in the tanker water market. Moreover, this information enables me to investigate potential systematic inequities in piped water supply across different households and neighborhoods.

Second, the questionnaire features a detailed module on tanker water consumption. This module documents the frequency and quantity of tanker water purchases, prices paid, specific sources from which water is purchased, and key challenges related to accessibility and reliability. Understanding these dimensions is crucial for accurately characterizing household purchasing behavior in the alternative water market.

Finally, the survey incorporates a detailed section on socio-economic and demographics, enabling analysis of consumption patterns and disparities across different household groups.

3.4 Descriptive Results

I have undertaken preliminary analysis of the data collected from the survey. The descriptive results presented serve two main purposes: (1) to demonstrate the presence of meaningful variation in water access, supply, and prices — variation required for identification in the structural model; and (2) to provide empirical evidence that motivates the core research question. At this stage, only descriptive and exploratory analyses are presented; the main structural results will be developed after the completion of model estimation.

Overall, the data collected has good representation of all three types of households, with 26% relying only on piped water, 63% on tanker water and 10% using both, piped and tanker water. Access to the piped water network is highly variable, with 38% of households disconnected from the piped network. Disconnection often reflects geographic factors—households are located in neighborhoods without piped infrastructure or in areas where pipes have not supplied water for years. Households must also pay a substantial one-time capital cost to connect (averaging PKR 3,000; Khalil et al., 2023), which serves as an additional barrier. As a result, many disconnected households are unlikely to gain access to piped water, even if supply were to increase, because they are located at the tail end of the distribution system. In the structural analysis, these disconnected areas are assumed to remain excluded from piped water under all policy scenarios.

Among the 286 households that are connected to piped network, only 168 households receive piped water. Within the households receiving piped water, there is tremendous variations in piped water supply as shown in Figure 3 below.

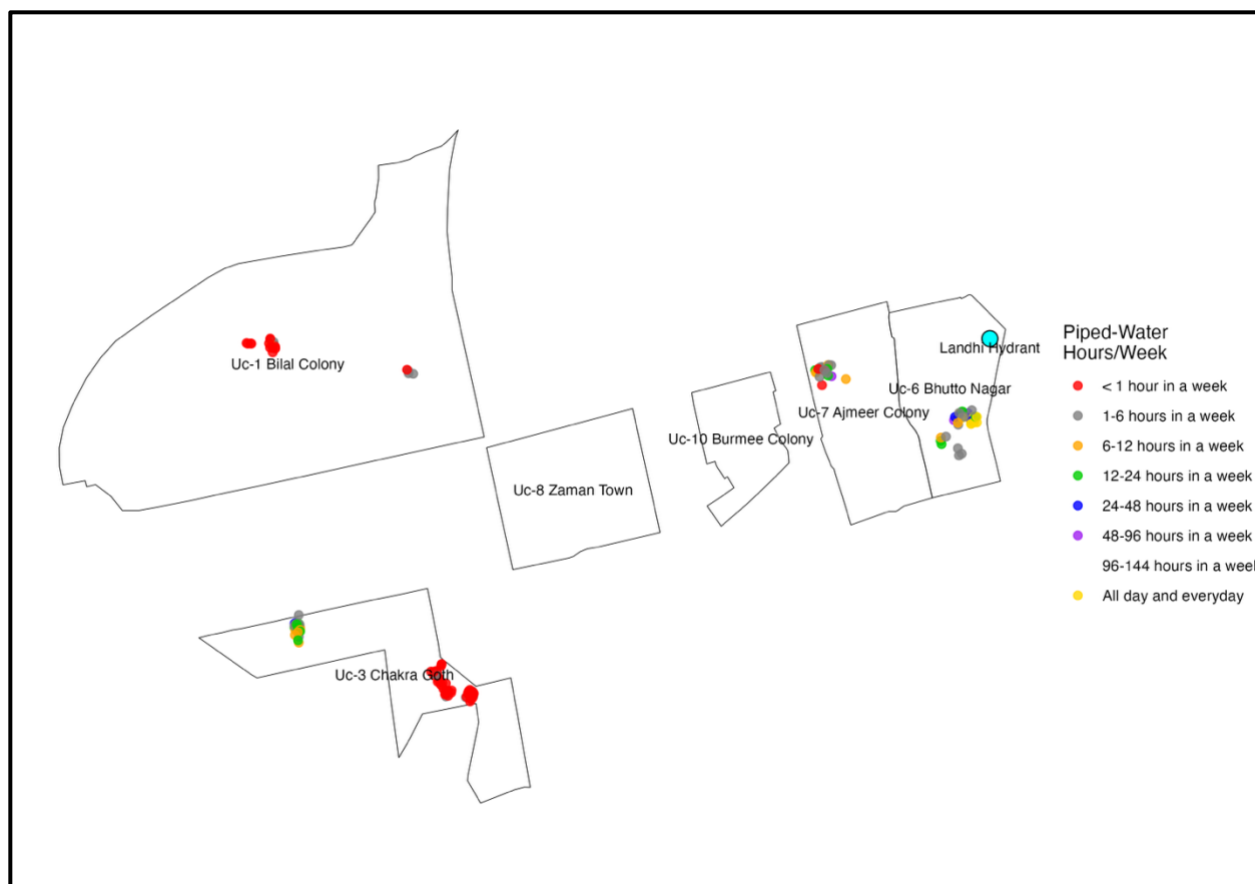


Figure 3. Variation in Piped Water Frequency. This figure illustrates the spatial heterogeneity in piped water frequency in a week. Frequency is measured as the total number of hours households receive piped water in a typical week, based on both the number of supply days and the duration per day. The map highlights substantial variation: households located closer to the main hydrant tend to receive piped water more frequently, while those farther away have much more limited access, sometimes receiving water for only a few hours each week.

Access to piped water remains the main determinant for participation in the alternative water market. Among the 339 households that participate in the alternative water market, 86% receive no piped water at all. The 14% households that do receive piped water, receive it very less (on average 4 hours a week) compared to average of 25 hours a week for households receiving piped water and not participating in the alternative water market. I also find significant variation in access to piped water across income status. After controlling for household characteristics and geographic factors, I find that poor households have lower likelihood of receiving piped water and these differences are statistically significant for lower-middle and upper-middle income groups but not for rich households.

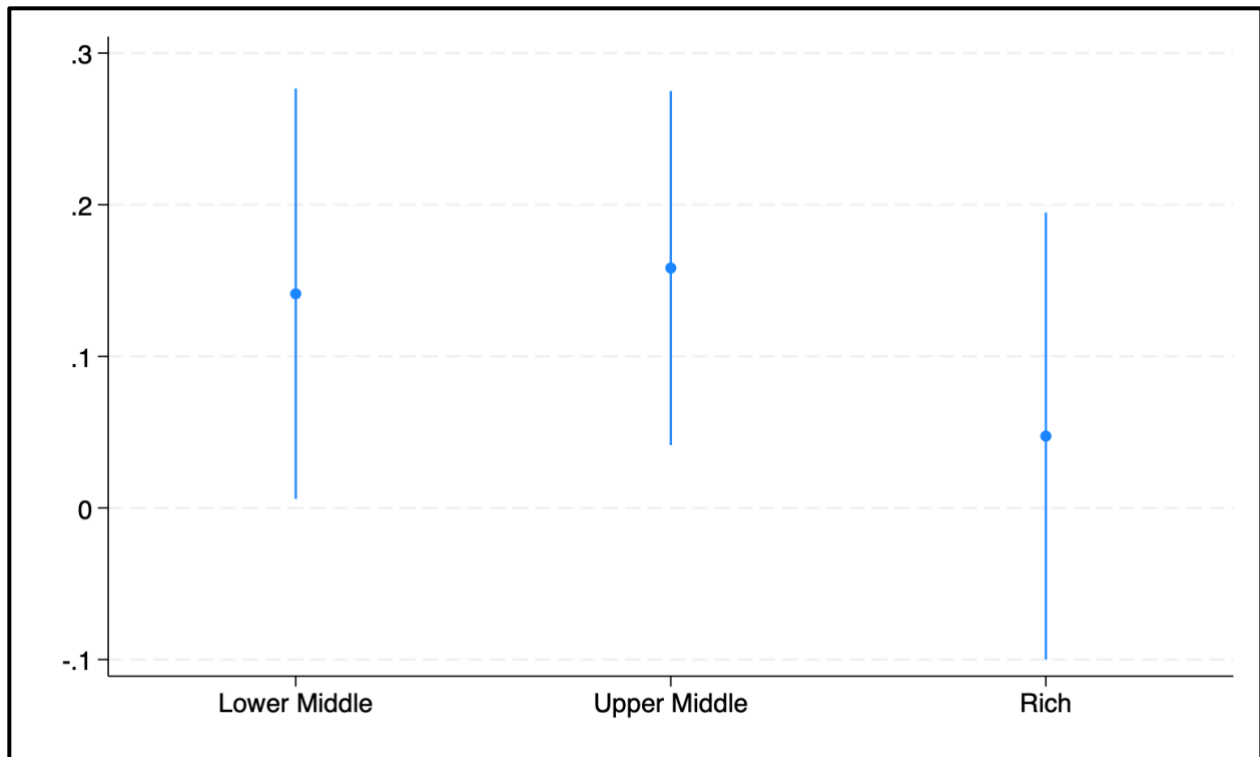


Figure 4. Predicted Probability of Receiving Piped Water by Income Group Relative to Poorest Households. This figure displays the predicted probability of receiving piped water for households in different income groups, relative to the poorest group (not shown, serves as reference). The results are from a linear probability model where the probability of receiving piped water is estimated on geographic and household factors including distance to the nearest hydrant, distance to pipeline, proximity to industrial areas, household wealth group, education, household size, and other relevant controls. Geographic coordinates and spatial cutoffs are included to correct for potential spatial dependence in the error terms. The dots represent point estimates, while the vertical lines show 95% confidence intervals. Both lower middle and upper middle-income households are significantly more likely to receive piped water compared to the poorest, while rich households also have a higher probability, though the confidence interval is wider and includes zero.

In addition to these disparities in access, I find that poor households systematically receive piped water much less frequently than upper-middle or rich households. Specifically, poor households receive a median of just 0.6 hours of piped water per week, compared to 5 hours per week for richer households—a difference that is statistically significant. This disparity is primarily due to the geographic factors: poor households are disproportionately located at the tail-end of the water distribution system, farther from both the Landhi Water Hydrant and the main water pipelines.

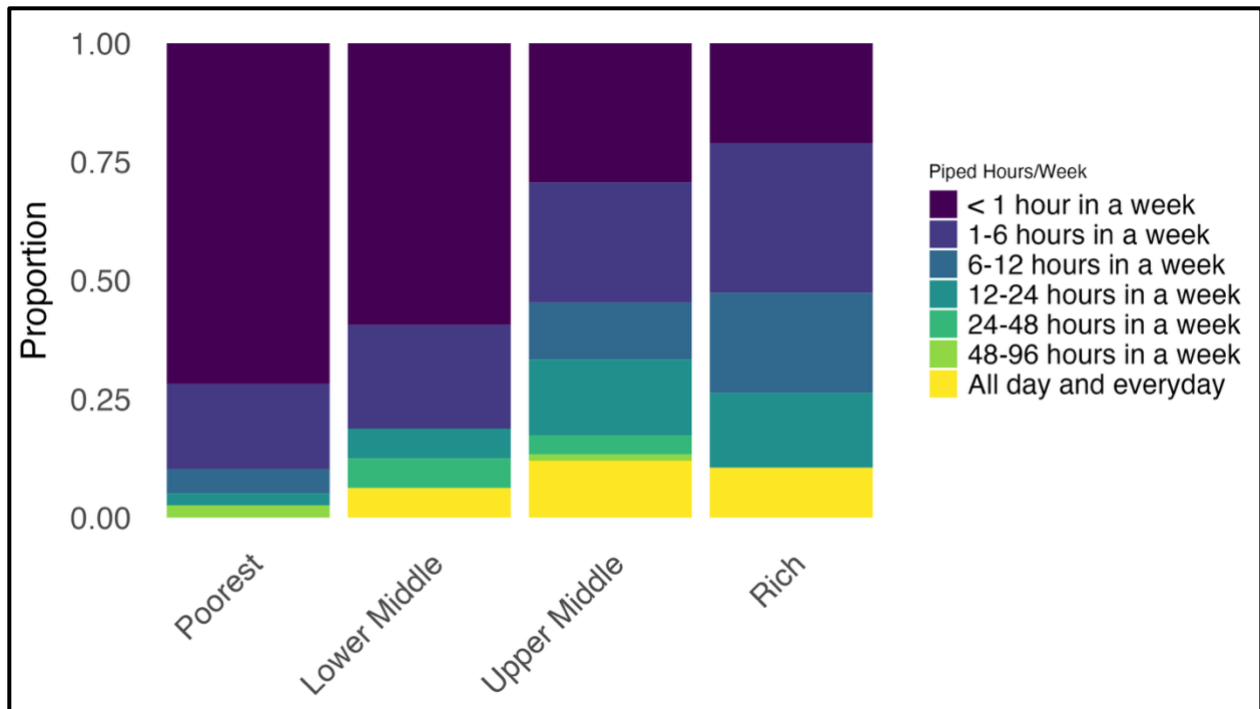


Figure 5. Proportion of Piped Water Hours per Week by Income Group. This stacked bar chart illustrates the distribution of weekly piped water hours received by households across different income groups. The figure reveals stark differences: poor and lower-middle-income households are overwhelmingly concentrated in the lowest availability categories, with the vast majority receiving less than 6 hours of piped water per week. In contrast, upper-middle and rich households are more likely to receive substantially more piped water, with a greater share enjoying at least 24 hours or even continuous supply

These findings demonstrate that systematic inequalities in piped water infrastructure disproportionately burden poorer households, forcing them to look for alternatives. However, I find evidence that households are not able to access all available water alternatives. Poor households are significantly less likely than wealthier groups to purchase from the cheapest option—government tankers—due to both geographic and institutional barriers. Government tankers primarily operate in areas near the hydrant, require customers to have large storage capacity, and are accessible only through an online booking system, all of which pose challenges for poorer households, who typically live farther from hydrant, lack large storage tanks, and face digital access constraints. Consequently, poor households are much more reliant on smaller, informal sources, such as pick-up trucks and drums, which supply water in small quantities but are more accessible to them.

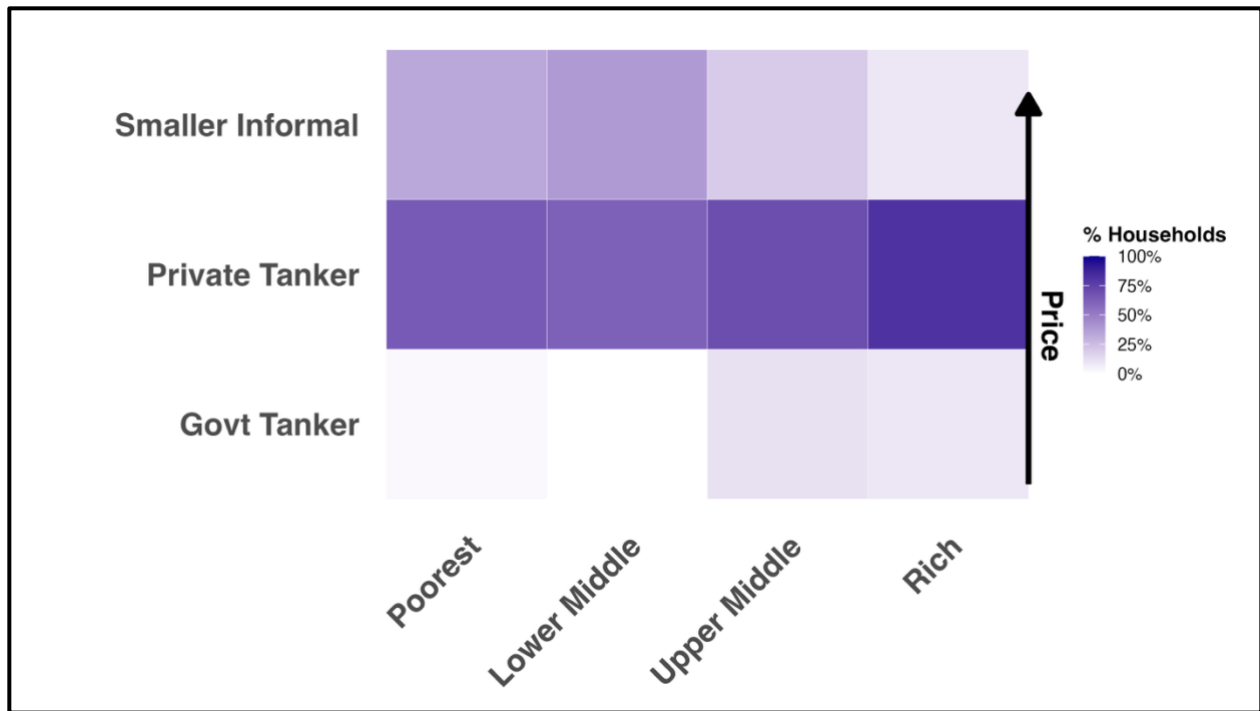


Figure 6. Household Reliance on Water Sources by Income Group. This heatmap illustrates the distribution of water source choices across household income groups. Each column corresponds to an income group (Poorest, Lower Middle, Upper Middle, Rich), and each row represents a type of tanker water source (Government Tanker, Private Tanker, Smaller Informal sources). The color intensity within each cell indicates the percentage of households in a given income group relying on that water source. The map shows poorer households rely proportionally more on smaller informal sources, which, despite higher per-unit costs. Richer households are more reliant on cheaper, government tankers and private tankers.

A key challenge associated with smaller sources is their significantly higher cost relative to larger tankers. As shown in Figure 7, informal sources—particularly drums—not only charge higher prices, but also exhibit considerable price volatility compared to formal government tankers. This underscores both the cost premium and the unpredictability that households face

when relying on these smaller, informal sources.

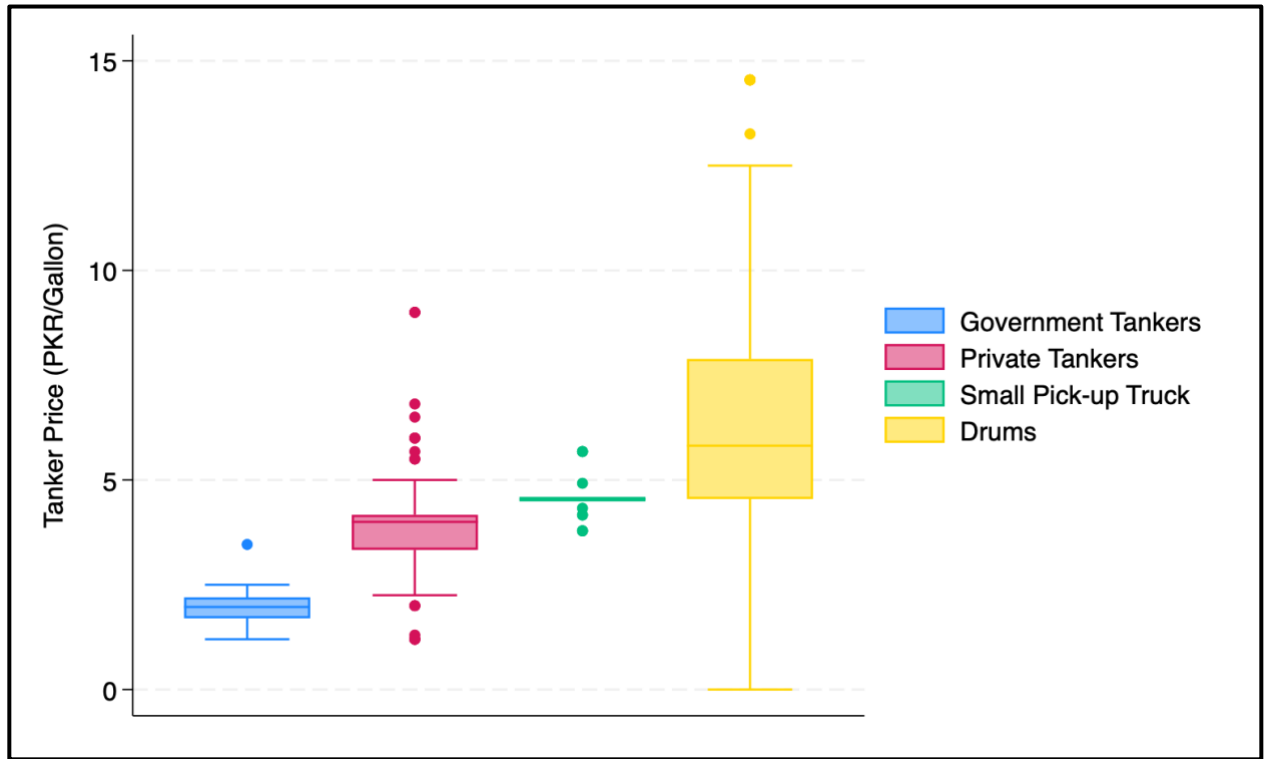


Figure 7. Distribution of Alternative Water Source Prices. This boxplot displays the distribution of per-gallon tanker water prices by source type: government tankers, private tankers, small pick-up trucks, and drums. Government tanker prices are the lowest and most consistent, with minimal dispersion. Private tanker and small pick-up truck prices exhibit greater variation, while prices for drums are both the highest and most variable, with a wide interquartile range and several outliers.

Even within the same water source, I find evidence that poorer households pay higher markups—particularly for private tankers and drums. This pattern likely stems from the fact that poor households are often located in areas lacking alternative reliable water sources, which creates opportunities for the water mafia to exploit their market power and charge higher prices. Overall, poor households face substantially higher per-gallon water prices—paying an average of PKR 6 per gallon for tanker water compared to only PKR 4 per gallon paid by richer households.

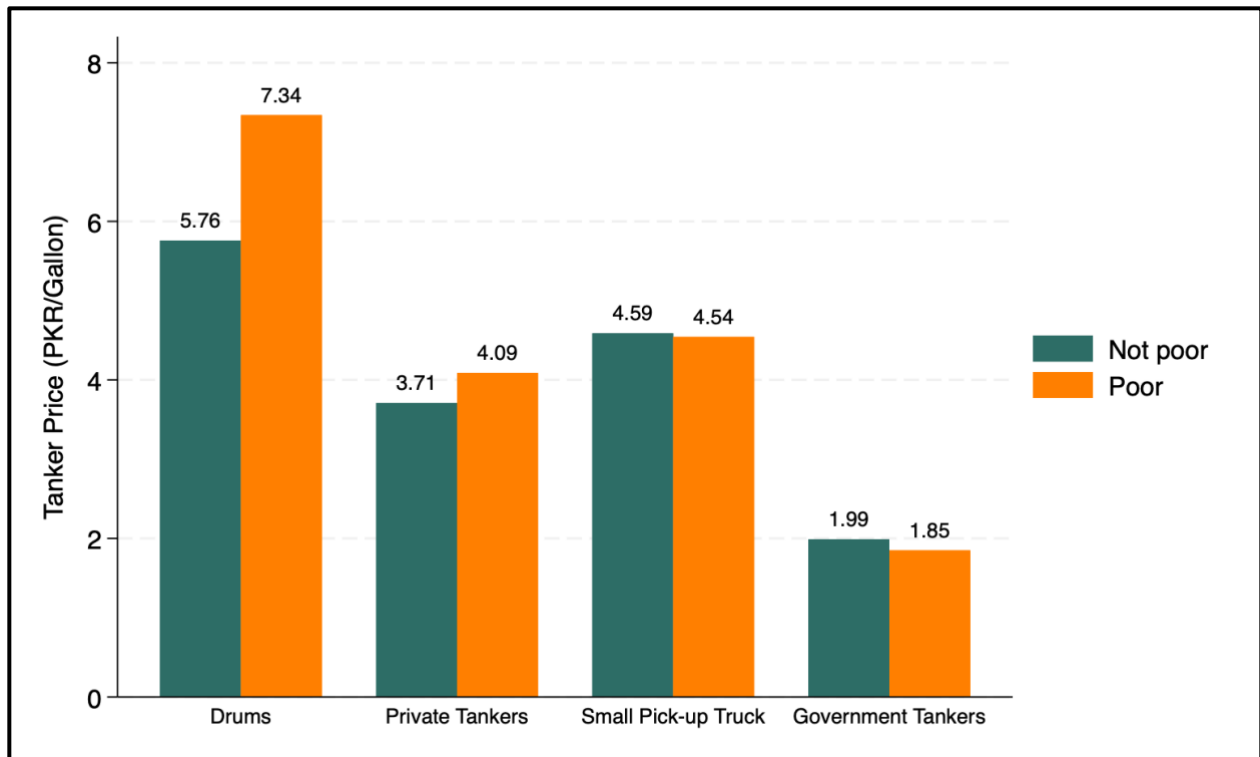


Figure 8. Price Paid by Source and Income. This bar chart displays the average price (in PKR per gallon) that poor and non-poor households pay for water from different tanker sources. Poor households consistently pay more than non-poor households for most water sources, with the largest disparity observed for drums and private tankers.

Consequently, poorer households are forced to consume significantly lower quantities of water, as each rupee buys less water—as illustrated in Figure 9 below.

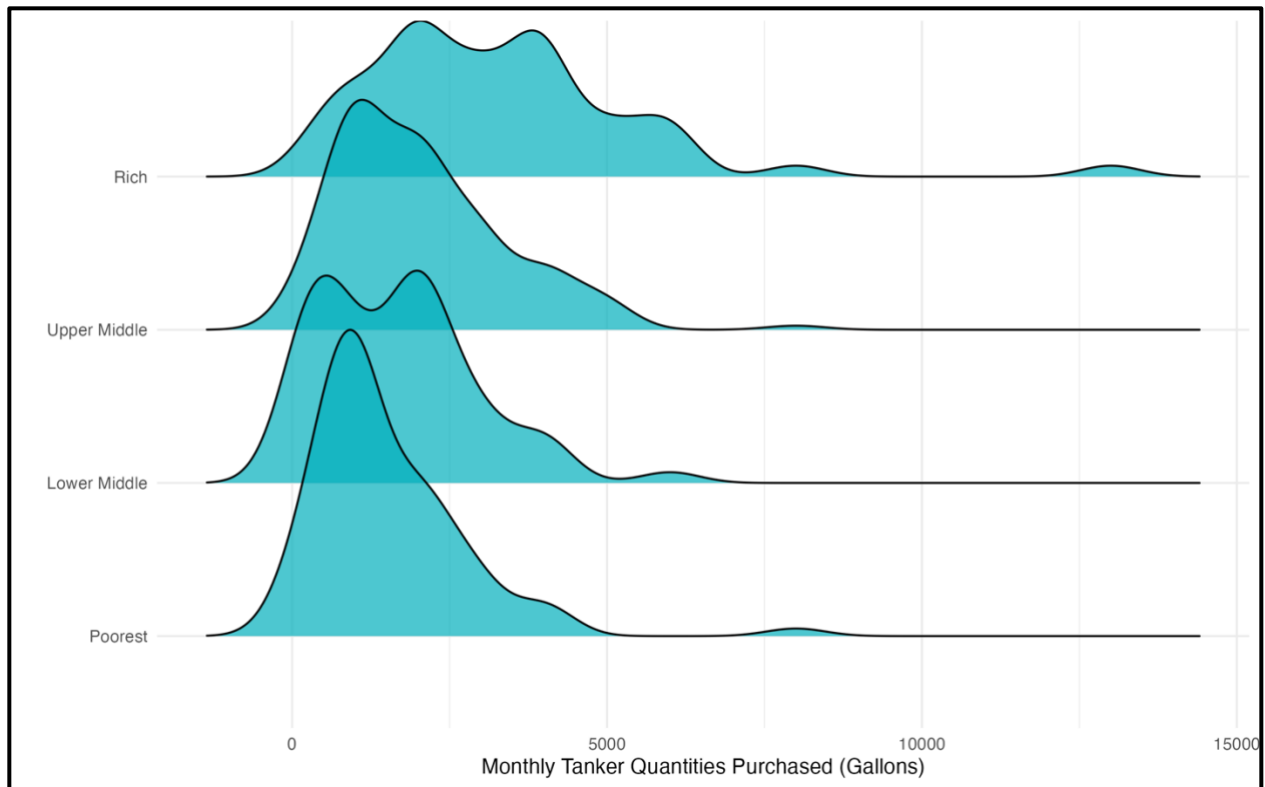


Figure 9. Distribution of Tanker Water Quantities Purchased by Income. This figure illustrates the spatial heterogeneity in tanker quantities purchased in a month across income status. Tanker quantities purchased in the month are calculated by multiplying one-time tanker quantity with times a tanker is purchased in a month. The density plots highlight substantial variation in consumption as richer households purchase much larger quantities of water compared to poorest households.

To summarize, the data finds that (1) disparities in piped water supply exist and compel poorer households to obtain water from the mafia and, (2) mafia uses this as opportunities to charge markups to poorer households making water unaffordable. These patterns provide direct empirical motivation for the core research question: what is the net social welfare impact (positive or negative) of the water mafia? To answer this, the next section outlines the DCC framework, and a policy simulation designed to quantify these welfare effects of the mafia.

3.5 Proposed Discrete Continuous Model

I plan to implement a discrete-continuous choice (DCC) model to estimate residual water demand and conduct welfare analyses under both the status quo and counterfactual scenarios. This model is particularly suitable for my context, where households not only select among mutually exclusive water sources to satisfy their residual water needs but also determine the quantity of water to consume from the chosen source.

In the DCC framework, the discrete component represents household's choice of water source, while the continuous component represents the water consumption decision conditional on the chosen alternative. The DCC framework assumes that households' source and consumption decisions are correlated, implying that observed variables jointly influence both the choice of water alternative and water consumption. Moreover, Dubin and McFadden (1984) reject the null hypothesis that unobserved factors affecting source choice are independent from those influencing consumption intensity. By employing a DCC framework, I can jointly model both decisions in a way that accommodates this correlation in unobservables, yielding consistent and unbiased parameter estimates.

Discrete-continuous choice models are widely adopted in the water literature. Studies focused on estimating water demand under tiered pricing structures have long recognized the advantages that DCC framework offers over traditional methods such as OLS and instrumental variable approaches. Because households under tiered pricing simultaneously decide which price block to use and how much water to consume, OLS estimates suffer from bias due to price endogeneity, while IV estimates often lack robustness due to violations of the exclusion restriction (Szabó, 2015). The DCC framework effectively addresses these issues and is therefore advocated for estimating water demand and conducting welfare analyses under tiered pricing schemes.

While the use of DCC framework to model water demand in presence of multiple water sources has been widely advocated, its implementation has been rare. Many studies adopt two-step estimation approaches which corrects for selection bias as mentioned by Heckman (1979). For instance, Cheeseman et al. (2008) investigates demand for water across households using (1) municipal water only and (2) municipal water with well water and correcting for the selection

bias by modeling households' decision to obtain a well. While these reduced-form approaches offer simpler analysis that is easy to implement and relies on fewer modeling assumptions, structural models are far more grounded in utility theory and better equipped to model complex behaviors and conduct plausible welfare analyses (Olmstead, 2009).

In Karachi's urban water market, households choose among several mutually exclusive alternatives to satisfy their residual water demand—demand remaining after piped water supplies are exhausted. For households lacking piped water access, this residual demand corresponds to their total water consumption. This context presents unique challenges compared to traditional discrete-continuous choice (DCC) models commonly used in water demand studies under tiered pricing structures. Specifically:

- Restricted choice sets: Households do not have access to all alternative water sources.
- Endogenous market participation: Participation in the alternative water market depends on the sufficiency of piped water supply, meaning some households do not engage in this market at all.

Traditional DCC frameworks assume universal availability of piped water and participation in that market. These assumptions do not hold here, precluding the straightforward application of standard DCC models. To address these challenges, I plan to adopt DCC framework implemented in literature investigating households' fuel choice. Specifically, I adapt the joint maximization approach proposed by Nesbakken (2001).

Households derive utility from consumption of water and other goods. The households maximize utility with respect to water consumption subject to the budget constraint and choice of water alternative.

Following Nesbakken (2001), I assume households' indirect utility function to take the following linear form for household i and source j :

$$V_{ij} = Z'_{1i}\gamma_j + \beta Y - \alpha P_j + \delta S_i + \varphi C_i + Z'_{2i}\lambda + \frac{H_j\alpha}{\beta} + \eta_i + \varepsilon_j \quad (1)$$

where

- Z'_{1i} is a vector of household and source characteristics affecting source choice. These include household location (distance to the Landhi hydrant and to the main roads), house size, family size, source reliability and convenience.
- Y is household's income proxied by asset index (details are provided in the supplementary text).
- P_j are source-specific prices.
- S_i represents piped water hours in a month.
- C_i is binned variable on storage capacity
- Z'_{2i} is a vector of household characteristics affecting water consumption including having additional water sources such as groundwater, family size, number of bathrooms, having a garden.
- $\frac{H_j\alpha}{\beta}$ is a technical term included to obtain the demand equation in a linear form through Roy's identity.
- η_i captures unobserved heterogeneity water demand for the household.
- ε_j captures unobserved characteristics/preferences for alternative j .

Parameters to be estimated are $\alpha, \beta, \gamma_j, \delta, \varphi$ and λ .

Households choose the water alternative that maximizes their utility. So, household i chooses source j if

$$V_{ij} = \max_k V_{ik} \quad \forall k \neq j \quad (2)$$

The consumption of water with respect to chosen alternative j is determined by Roy's identity:

$$X_j = - \frac{\frac{\partial V_j}{\partial P_j}}{\frac{\partial V_j}{\partial Y}} \quad (3)$$

3.5.a Proposed Econometric Model

The discrete choice is a two-stage process. Household i first decides whether to participate in the tanker market (i.e., not rely solely on piped water). Let

$$P_i = \begin{cases} 1 & \text{if household } i \text{ participates in tanker water market} \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

Let the latent utility difference of participating be

$$U_i = \gamma_1 S_i + \gamma_2 PS_i + \zeta \mathbf{Dist}'_i + \eta UC_i + \varepsilon_i, \quad (5)$$

where

S_i represents piped water hours in a month.

PS_i is an indicator variable of piped water sufficient which takes the value of 1 if $S_i > 10$

\mathbf{Dist}'_i is a vector of distance variables that affect piped water availability.

UC are UC-level fixed effects.

The first-stage doesn't include prices as piped participation depends almost entirely on physical availability, not on household preferences or prices.

The probability that household i participates in the tanker market is determined by the latent utility U_i defined in Equation (5):

$$\pi_i^{(1)} = \Pr (D_i = 1) = F(U_i), \quad (6)$$

where $F(\cdot)$ is the cumulative distribution function of the error term ε_i (assumed standard normal for the probit case). Households with $U_i > 0$ participate in the tanker market, while those with $U_i \leq 0$ rely exclusively on piped water.

Conditional on participating, the household chooses a tanker source $j \in \mathcal{J} = \{\text{Govt, Private, Other}\}$. The household chooses the water alternative that maximizes the indirect utility, V_{ij} . So the probability that source j is chosen is given by:

$$\pi_{ij}^{(2)} = \Pr (V_{ij} = \max_k V_{ik}) \quad \forall k \neq j \quad (7)$$

Assuming ε_j are independently and identically extreme value distributed, Equation (7) yields (McFadden, 1974):

$$\pi_{ij}^{(2)} = \frac{\exp(V_{ij})}{\sum_{k=1}^K \exp(V_{ik})} \quad (8)$$

As all water sources are available or accessible to every household, I define an accessibility indicator, A_{ij} , which equals 1 if source j is accessible to household i , and 0 otherwise.

Accessibility, A_{ij} , is determined by household self-reports regarding the availability of each source. However, to account for potential misreporting or lack of awareness, this information is corroborated by ensuring no household in the neighborhood (defined as households within 200-m radius) purchases from the source deemed unavailable.

The choice probability adjusts accordingly:

$$\pi_{ij}^{(2)} = \frac{A_{ij} * \exp(V_{ij})}{\sum_k A_{ik} * \exp(V_{ik})} \quad (9)$$

To define the probabilities, I am only concerned with differences in utility across alternatives for a given household.

$$\pi_{ij}^{(2)} = \frac{A_{ij} * \exp(W_{ij})}{\sum_k A_{ik} * \exp(W_{ik})} \quad (10)$$

where

$$W_{ij} = Z'_{1i}\gamma_j + \beta Y - \alpha P_j + \delta S_i + \varphi C_i \quad (11)$$

Z'_{1i} includes source and household characteristics. A binary indicator on perceptions of government tankers' service is included. The variable takes a value of 1 if households perceive government tankers as unreliable. A storage insufficiency indicator is also included to penalize government and private tanker alternatives as these cannot be purchased by households with insufficient storage. The choice probabilities are given by a generalized multinomial logit model.

From equations (1) and (3), water consumption conditional on chosen water alternative is given by

$$X_{ij} = Z'_{1i}\gamma_j + \beta Y - \alpha P_j + \delta S_i + \varphi C_i + Z'_{2i}\lambda + \eta_i \quad (12)$$

It is important to note that X_{ij} is observed only for households that participate in tanker market and choose alternative j . Therefore, water consumption is not independent of the household's source choice. This creates potential selection bias, as the expected value of the error term conditional on the chosen alternative, $E[\eta_i|j] \neq 0$. To address this bias, a correction term is computed for each alternative based on the estimated choice probabilities and incorporated into the demand equation.

$$X_{ij} = Z'_{1i}\gamma_j + \beta Y - \alpha P_j + \delta S_i + \varphi C_i + Z'_{2i}\lambda - \left[\sigma_j \log \pi_{ij}^{(2)} - \sum_{k \neq j} \sigma_k m_{ik}^{(2)} \right] + \mu_j + \psi_j \xi_i \quad (13)$$

where

ξ_i is correction term for participation and $\xi_i = \frac{\varphi(\bar{U}_i)}{\Phi(U_i)}$

$\sigma_k = \sigma \rho_k$ where ρ_j is the correlation between η_i and ε_j .

$$m_{ik}^{(2)} = \pi_{ik}^{(2)} \log \pi_{ik}^{(2)} / (1 - \pi_{ik}^{(2)})$$

$$\mu_j \sim N(0, \tau_j)$$

The purpose of including the selection correction term, $\left[\sigma_j \log \pi_{ij}^{(2)} - \sum_{k \neq j} \sigma_k m_{ik}^{(2)}\right] + \mu_j + \psi \xi_i$ is to allow correlation between unobserved factors influencing source choice and those affecting consumption.

In Equation (13), the discrete (choice) and continuous (consumption) decisions are linked via common parameters such as prices, income, piped water supply and storage capacity. By estimating these two components simultaneously, I obtain consistent parameter estimates using a joint log likelihood function. Following Nesbakken (2001), I define

$$D_{ij} = \begin{cases} 1 & \text{if household } i \text{ chooses water alternative } j \\ 0 & \text{otherwise} \end{cases} \quad (14)$$

Then, the joint log likelihood is defined as

$$L(\theta) = \sum_{i=1}^N \sum_{j=1}^J \left[D_{ij} \log \pi_i^{(1)}(\theta) + D_{ij} \log \pi_{ij}^{(2)}(\theta) + D_{ij} \log f_{ij}(X_{ij}) \right] \quad (15)$$

Where $\pi_i^{(1)}(\theta)$ is the participation probability given by equation (6), $\pi_{ij}^{(2)}(\theta)$ is the choice probability given by Equation (10), $f_{ij}(X_{ij})$ is conditional probability density function from Equation (13) where error term is normally distributed and θ are parameters to be estimated.

Maximization of the log-likelihood function will yield estimates of parameters allowing me to estimate baseline utility.

3.6 Conclusion

Accessing clean drinking water remains a challenge for a significant proportion of the population in the developing world. Often, non-state actors get involved in the distribution of drinking water to bridge the gap between supply and demand. These actors cater specifically to the vulnerable population groups that are disproportionately burdened by inadequate piped water infrastructure.

This paper is the first rigorous evidence on potential distributional impact of the informal water vendors and the significance of these players in the delivery of water resources.

The empirical findings of this research offer valuable insights for the design of policies aimed at improving water access and equity for low-income communities, who disproportionately bear the costs of inefficient water resource management. Furthermore, the results have broader implications for the regulation of informal providers in other essential service markets—such as electricity, health, and education—where the balance between public provision and private or informal sector involvement is actively debated.

Despite its contributions, this study has several limitations. First, the analysis relies on cross-sectional survey data, which limits the ability to make causal inferences about the dynamics of water access and household responses over time. Second, while the analysis addresses distributional impacts and welfare, it does not capture all potential indirect effects—such as health outcomes. In particular, eliminating the water mafia could prompt some households to switch to riskier water sources not observed in the dataset, potentially leading to an underestimation of negative welfare impacts. Finally, the data collection process limits the generalizability to other contexts with different institutional arrangements or market structures.

Supplementary Information “Can Mafia Be Good? Evidence from Karachi’s Urban Water Market”

S1. Data Construction of Income Variable

In the survey questionnaire, households are asked to report their average monthly income. However, there is significant measurement error in the variable as household were not truthfully reporting their income. Income was underreported whereas, expenditures were overreported. To obtain a reliable measure of income, an index using principal components analysis on detailed information on ownership assets from the survey data was undertaken.

While PCA typically uses binary indicators of asset ownership, the survey data include richer information on the quantity of assets owned. Therefore, multiple component analysis (MCA)

was employed to create categorical variables such as “Does not own,” “Owns one,” “Owns two,” “Owns three,” and “Owns four or more” for each asset (Filmer & Scott, 2012).

The asset index constructed using MCA was then utilized to define income categories. Drawing on World Bank poverty thresholds for Pakistan (World Bank, 2023) —21.9% (national poverty rate), 39.8% (lower middle income), and 84.5% (upper middle income)—households are categorized as poor (≤ 21.9 th percentile), lower middle income (>21.9 th and ≤ 39.8 th), upper middle income (>39.8 th and ≤ 84.5 th), and rich (>84.5 th percentile).

To validate the asset index as a proxy for income, a comparison of wealth classification based on asset index and an index using expenditure data from the survey was conducted. However, as the expenditure data is missing for many households, this comparison was only undertaken for a subset of households that reported their expenditures. The resulting correlation between asset-based and expenditure-based wealth classifications was high (0.6), indicating that the asset index provides a credible proxy for income status.

Bibliography

- Felbab-Brown, V. (2017, March 23). *Stealing water*. Retrieved August 2025, from Brookings: <https://www.brookings.edu/articles/stealing-water/#:~:text=In%20Karachi%20and%20Delhi%2C%20such,list%20goes%20on%20and%20on.>
- Abidi, Y. A., Ahmed, M., & Shah, M. H. (1998). Implementation of private sector participation in Karachi. *24th WEDC Conference* (pp. 173-177). Islamabad: WEDC, Loughborough University.
- Ahmad, H. F., Ali, A., Meeks, R., Wang, Z., & Younas, J. (2024). Down to the wire: leveraging technology to improve electric utility cost recovery. *SSRN Working Paper Series*.
- Ahmed, N. (2008). *Water Supply in Karachi: Issue and Prospects*. Karachi, Sindh, Pakistan: Oxford University Press.
- Ahmed, S., Ali, S. H., Mirza, M., & Lotia, H. (2017). The limits of water pricing in a developing country metropolis: empirical lessons from an industrial city of Pakistan. *Water*, 9(7).

- Auditor-General of Pakistan. (2019). *Special Study Report on Mechanism of Revenue Generation by Revenue Resource Generation Department, KW&SB*. Islamabad: Auditor-General of Pakistan.
- Balza, L., Jimenez, R., & Mercado, J. (2013). *Privatization, institutional reform, and performance in the Latin American electricity sector*. Washington, D.C: Inter-American Development Bank.
- Beard, V. A., & Mitlin, D. (2021). Water access in global South cities: The challenges of intermittency and affordability. *World Development*, 147.
- Cheeseman, J., Bennett, J., & Hung Son, T. V. (2008). Estimating household water demand using revealed and contingent behaviors: evidence from Vietnam. *Water Resources Research*, 44(11).
- Coulibaly, L., Jakus, P. M., & Keith, J. E. (2014). Modeling water demand when households have multiple sources of water. *Water Resources Research*, 50(7), 6002-6014.
- Dubin, J. A., & McFadden, D. L. (1984). An econometric analysis of residential electric appliance holdings and consumption. *Econometrica*, 52(2), 345-362.
- Filmer, D., & Scott, K. (2021). Assessing asset indices. *Demography*, 49(1), 359-92.
- Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica*, 47(1), 153-161.
- Janjua, S., Hassan, I., Ali, M., Ibrahim, M., Zafar, A., & Kim, S. (2021). Addressing social inequality and improper water distribution in cities: a case study of Karachi, Pakistan. *Land*, 10(11).
- Khalil, S., Sheikh, S. B., Jamal, U., & Khan, H. F. (2023). Household water insecurities and willingness to pay in Karachi. *Water Policy*, 25(12), 1139-1160.
- Khan, H. F., & Arshad, S. (2022). Beyond water scarcity: Water (in)security and social justice in Karachi. *Journal of Hydrology: Regional Studies*, 42(101140).
- Klassert, C., Sigel, K., Gawel, E., & Klauer, B. (2015). Modeling residential water consumption in Amman: the role of intermittency, storage, and pricing for piped and tanker water. *Water*, 7(7).
- Klassert, C., Yoon, J., Sigel, K., Klauer, B., Talozzi, S., Lachaut, T., . . . Gawel. (2023). Unexpected growth of an illegal water market. *Nature Sustainability*, 6, 1406-1417.
- Kunbhar, Z. (2016, November 16). *Karachi's burgeoning water mafia*. Retrieved November 2023, from Eco-Business: <https://www.eco-business.com/news/karachis-burgeoning-water-mafia/>
- KWSB. (2020, July 05). *Auction Document: Supply of Potable Water from Nipa Hydrant– KWSB Through Tanker Services to District East Karachi*. Retrieved November 2023, from KWSB: <https://www.kwsb.gos.pk/wp-content/uploads/2020/07/05.-NIPA-Financial.pdf>
- McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In P. Zarembka, *Frontiers in Econometrics* (pp. 105–142). Academic Press, New York.

- Nauges, C., & Berg, C. v. (2009). Demand for piped and non-piped water supply services; evidence from southwest Sri Lanka. *Environmental and Resource Economics*, 535-549.
- Nesbakken, R. (2001). Energy consumption for space heating: a discrete-continuous approach. *The Scandinavian Journal of Economics*, 103(1), 165-184.
- Olmstead, S. M. (2009). Reduced-form versus structural models of water demand under nonlinear prices. *Journal of Business & Economic Statistics*, 27(1), 84-94.
- Persson, T. (2000). Household choice of drinking water source in the Philippines. *Asian Economic Journal*, 16(4), 303-316.
- Raina, A., Zhao, J., Wu, X., Kunwar, L., & Whittington, D. (2019). The structure of water vending markets in Kathmandu, Nepal. *Water Policy*, 21(S1), 50-75.
- Sajjad, Y. (2019). *The water tanker mafia*. Institute of Business Administration Karachi, Journalism . Karachi: Institute of Business Administration.
- Siddiqui, N., & Haider, E. (2023, March 23). *Political Links to the Water Mafia in Karachi*. Retrieved August 2025, from UC Institute on Global Conflict and Cooperation: <https://ucigcc.org/blog/political-links-to-the-water-mafia-in-karachi/>
- Smith, T. B. (2004). Electricity theft: a comparative analysis. *Energy Policy*, 32(18), 2067-2076.
- Szabó, A. (2015). The value of free water: analyzing South Africa's free basic water policy. *Econometrica*, 83(5), 1913-1961.
- World Bank. (2023, April). *Pakistan*. Retrieved August 2025, from Poverty & Equity Brief South Asia: https://databankfiles.worldbank.org/public/ddpext_download/poverty/987B9C90-CB9F-4D93-AE8C-750588BF00QA/current/Global_POVEQ_PAK.pdf