Optimal Pricing and Informal Sharing: Evidence from Piped Water in Manila

William Violette*

November, 2017

Abstract

Public utilities subsidize fixed connection fees with high marginal prices to provide access especially for the poor, but this policy can have the opposite effect when many households share connections. Using data on 1.5 million water connections in Manila, I establish that informal sharing networks provide 26% of total access. I structurally estimate household water demand across three sources: purchasing directly from the provider, sharing with a neighbor's tap, or buying from a small-scale vendor. The model predicts that low fixed fees and high marginal prices decrease access by weakening incentives to share water. In contrast, the optimal pricing policy features a high fixed fee and low marginal price, increases shared connections, and ensures nearly universal access to piped water compared to current prices in Manila. This policy improves welfare by up to 156% of consumer surplus or 1.1% of household income.

JEL-Classification: H23, L95, L98, O13, O17, O18

Keywords: Water demand, non-linear pricing, informal sector, developing countries.

*Contact: Federal Trade Commission, Washington, DC: william.j.violette@gmail.com
I am grateful to Andrew Foster, Matthew Turner, Jesse Shapiro, Daniel Björkegren, and Bryce Steinberg for guidance and encouragement. For helpful conversations, I am also grateful to John Friedman, Edward Glaeser, Emily Oster, Brian Knight, and Lint Barrage. In Manila, thank you to the staff of my water company partner, government regulators at MWSS, as well as the USAID Be-Secure project for access to data and excellent conversations. This work was supported through the Watson Institute for International and Public Affairs. Any opinions and conclusions expressed herein are those of the author and do not necessarily represent the views of the the Federal Trade Commission or its Commissioners.

1 Introduction

Despite large investments in piped water throughout the developing world, the share of urban households without piped water has remained stable at 5% for middle income countries and at 20% for low income countries for at least the past decade (World Bank [2015]). Given health, time-savings, and other benefits from piped water, how can water utilities set prices to ensure access while covering costs?[1] A popular approach particularly in developing cities is to subsidize upfront connection fees with high marginal prices.[2] Over 70% of developing cities also increase marginal prices with monthly usage so that large users cross-subsidize access and usage for smaller, poorer users.[3] Recent empirical papers in South Africa and Côte d'Ivoire predict gains in both access and total welfare from increasing marginal prices (Szabó [2015], Diakité et al. [2009]).

Conventional pricing policies often rely on the assumption that each water connection serves a single household; however, many households in developing cities share with neighbors instead of purchasing their own connections. For example, 70% of African cities report frequent water sharing (Keener et al. [2010]) ranging from fetching from neighbors with hoses and buckets in Tangiers, Morocco (Devoto et al. [2012]) to drawing from single taps that each serve compounds in Kumasi, Ghana (Whittington [1992]).[4] In Manila, I establish that 23% of households access water through a neighbor's connection by connecting plumbing directly to the meter (61%), extending plastic pipes (25%), or fetching water with containers (14%). While previous studies have focused on specific suburbs, this paper is the first to my knowledge to measure the role of water sharing in extending access across an urban population.

Water sharing has important implications for the welfare impacts of common pricing policies. Increasing marginal prices weaken incentives for households to extend water access to their neighbors through sharing (Whittington [1992]). Fixed fee subsidies may have limited impacts on water access when sharing households already split connection fees with their neighbors. Since low-income households are more likely to use water from their neighbors,

[1]Gamper-Rabindran et al. [2010], Cutler and Miller [2005], and Galiani et al. [2005] find health improvements in both developed and developing settings while Devoto et al. [2012] find time savings and other utility benefits.

[2]McIntosh [2003] and Komives et al. [2006] from the World Bank and Asian Development Bank argue for low connection fees. Using a randomized controlled trial from Morocco, Devoto et al. [2012] confirm that modest assistance with connection fees can lead to increases in new water connections.

[3]This strategy is also commonly referred to as "Increasing Block Tariffs" or "Convex Pricing" (Borenstein [2012]). Hoque and Wichelns [2013] and Boland and Whittington [2000] provide overviews of the use of increasing marginal prices in the water sector in developing countries.

[4]Nauges and Van den Berg [2006] finds further evidence of widespread water sharing in Sri Lanka.

increasing marginal prices and low fixed fees may combine to act as regressive taxes in the presence of sharing.

This paper analyzes the welfare impacts of pricing policies in the presence of informal water sharing. Welfare impacts depend not only on heterogeneity in demand for water across the population, but also on costs associated with forming and maintaining water sharing relationships. To measure these features, I merge administrative data covering 1.5 million water connections in Manila from 2010 to 2015 with detailed survey data on the number of households and people using each water connection. With comprehensive geographic coverage of half of Manila and a variety of price variation, these data allow for a rich characterization of water demand. Additional variation in water source choices identify the relative fixed and marginal costs associated with sharing water with a neighbor. I find that despite relatively high marginal costs of sharing, shared connections allow the city to save on installation and maintenance costs, providing an efficient means for extending piped water to low-demand households at low marginal prices. At the same time, informal sharing networks serve as resale markets for water, undermining the extent to which cities can redistribute surplus through increasing marginal prices.

To evaluate the impacts of different pricing policies, I build a model of household water demand that includes both initial source choices as well as monthly consumption choices. This model expands on the demand estimation literature for public utilities that focuses almost exclusively on quantity choices (Diakité et al. [2009], McRae [2014], Olmstead [2009], Szabó [2015]). I specify a sequential game where households choose whether to connect individually, share with their neighbors, or purchase from a water vendor. For sharing households, I develop a simple model of the water sharing contract that captures two key features of these relationships: (1) households face challenges in observing each other's demand and (2) households often split the monthly water bill evenly. The model characterizes monthly consumption choices when households face non-linear budget constraints and imperfect control over their exact consumption levels. This approach extends previous models of optimization over non-linear budget constraints by allowing households to anticipate unobserved shocks to their demand (Moffitt [1986], Burtless and Hausman [1978]).

The estimation combines water source and consumption decisions to structurally recover preferences and unobserved costs of water sharing. To include reduced-form approaches to identification within the context of a structural model, I divide the estimation into three steps. The first step uses variation in the non-linear price schedule (both across and within households) to identify water demand for each household using monthly consumption data. To measure the shadow costs of fetching water from a shared connection, the second step analyzes a quasi-experiment where households face leaks in their service pipes, forcing them to substitute to a neighbor's connection. The model maps the extent to which these households are able to offset their consumption to a neighbor's connection into an estimate of the marginal costs of fetching water. This exercise recovers sizable sharing costs on the order of 63.7% of the average marginal water price. This quasi-experimental approach builds on a descriptive literature pointing to potential costs of sharing water (Whittington [1992], Nauges and Van den Berg [2006]). On top of an initial connection fee and monthly service fee, households may face unobserved fixed costs in owning a connection, which may include application costs, plumbing maintenance, as well as land-tenure issues in qualifying for a water connection. The third step estimates these fixed costs by simulating the sequential game to match cross-sectional variation in water source choices. This strategy provides a structural, revealed-preference approach to recovering unobserved fixed costs, complementing previous reduced-form evidence in Devoto et al. [2012].

Combining the model and estimates, I simulate a series of counterfactual policies. First, I consider a connection fee subsidy policy consistent with recommendations from the World Bank, NGOs, and other researchers to improve water access for the poor (Komives [2005], Water and Sanitation for the Urban Poor [2013], Jimenez-Redal et al. [2014], McIntosh [2003]). I discount the fixed fee by 12% while increasing the tariff schedule proportionally to ensuring that the water provider exactly covers costs. This policy reduces access by incentivizing households to substitute toward water vendors, which mainly include deepwells and water refilling stations.

Next, I compute an optimal tariff schedule to maximize consumer surplus while exactly covering production costs. Compared to the current tariff, I find that a simple, two-part tariff characterized by a fixed fee and low, flat marginal price can produce gains on the order of 156% of consumer surplus or 1.1% of household income. These gains are driven both by pricing closer to marginal cost as well as by increasing sharing, which reduces the number of connections to install and maintain.

Finally, I compare the optimal tariff with a social tariff that maximizes consumer surplus only for the bottom 50% of users and is intended to capture goals of ensuring minimum usage and access as well as income redistribution. With sharing, social tariffs are ineffective at reallocating surplus from large to small users; however in a counterfactual without sharing, social tariffs achieve substantial redistribution, over doubling consumer surplus for small users. This evidence suggests that the presence of informal sharing networks constrains non-linear pricing as an effective tool for redistribution.

In focusing on water sharing behavior, this approach abstracts away from several features that may also affect optimal water pricing. First, I do not directly model environmental, health, or other externalities associated with piped water use. Water shortages may lead policymakers to promote conservation through higher marginal prices while positive health externalities may motivate further subsidizing piped water (Timmins [2002], Galiani et al. [2005]).[5] Second, I assume that government regulator is able to fully observe production costs for the water provider, eliminating scope for prices to correct for possible principal-agent frictions as characterized by Laffont and Tirole [1993]. Finally, the counterfactual exercises simulate the public-private partnership regulatory framework in Manila by assuming that revenue must fully cover production costs. Additional data on the costs of raising public funds and risks of regulatory capture would allow for a more flexible analysis taking into account the possibility of using public funding for water infrastructure (Laffont [2005]).

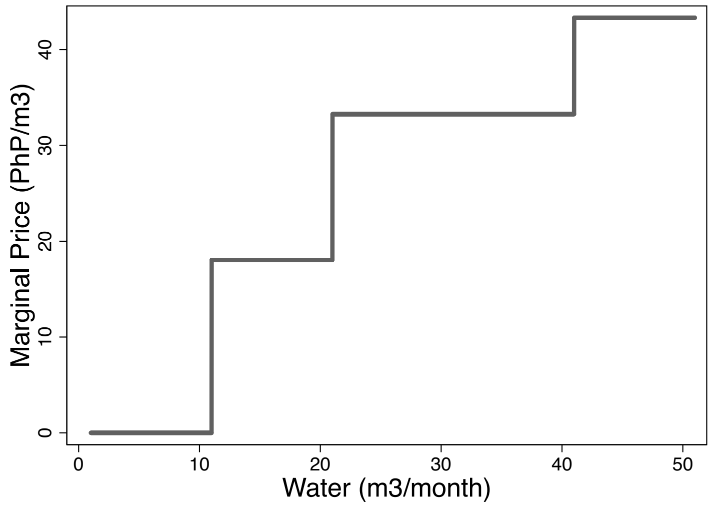
This paper proceeds by first describing the context of water provision in Manila. Section 3 describes the data. Section 4 presents stylized facts about water usage in Manila. Section 5 introduces a model of household water usage and source choices. The estimation is divided into three steps in Section 6. Section 6.1 uses monthly quantity choices and price variation to estimate preferences. Section 6.2 analyzes a quasi-experiment to recover an estimate of the marginal costs of sharing with a neighbor. Given these estimates, Section 6.3 recovers the relative fixed costs of owning a connection or using from a vendor. Section 7 uses these estimates to examine counterfactual pricing policies. Section 8 concludes.

2 Water Provision in Manila

In 1997, citing poor service quality and low coverage of the public water provider, the Philippines conducted the largest water privatization in the developing world. Two private companies evenly divided the city into east and west zones as part of a public-private partnership. Both companies invested heavily in infrastructure, which has resulted in near universal availability of piped water services for Manila. A government regulator oversees these companies and implements rate-of-return regulation: water tariffs are revised periodically in order to ensure that providers recover all operating costs as well as earn a percentage return on any capital investments. Rate-of-return regulation is common in developing countries since high costs of public funds as well as high risks of regulatory capture limit funding of utilities through government revenue (Laffont [2005]).

[5] Bennett [2012] shows that piped water can also lead to negative externalities through worse sanitation, possibly motivating higher prices for piped water in some cases.

Figure 1: Water Tariff in Manila



1 USD = 50 PhP

This paper focuses on one of the providers which serves 1.5 million connections and around 5 million people. Figure 1 shows the regulated tariff in Manila. The increasing marginal price is thought to cross-subsidize consumption among low-demand, poorer households through higher bills from large users. The regulator assigns zero marginal price to the first 10 cubic meters of monthly usage, which is called the "lifeline" rate. This policy is intended to provide a minimum level of consumption for all households. This tariff may also perform efficient price discrimination if low-usage households are more price-sensitive. On top of this tariff schedule, the regulator also charges a fixed fee independent of usage totaling 225 PhP/month. This fee includes both the monthly service charge and the initial connection fee that I have put into monthly terms for the duration of the water connection.[6] The fixed fee allows the provider to recover any operating costs associated with serving the connection. This form of increasing tariff and fixed fee combination is implemented in over 70% of developing cities as well as over 50% of developed cities (Hoque and Wichelns [2013]).

Faced with this fixed fee as well as other costs associated with getting and maintaining a connection such as additional application costs or land tenure insecurity, I find that over 23% of households use a neighbor's tap in Manila. Sharing a connection with a neighbor mechanically increases the total monthly consumption recorded per connection, often doubling the marginal prices faced by these households compared to using a connection alone. Although most households use piped water either through their own or neighboring connections, 6% of households remain unconnected to the network and instead fetch water primarily from

[6]Appendix A.I provides more details about how this monthly fee is calculated and amortized.

nearby deepwells while some use water refilling stations or other vendors.

3 Data

Evaluating different pricing policies requires measuring demand at the household level as well as costs associated with providing piped water. Data was gathered through a partnership with one of the two regulated, private providers in Manila who provided both billing records as well as information about the regulatory structure, pricing, and production costs. Complete billing records for each water connection including the monthly meter reading, the total bill charged, and the precise location for each meter serve as the primary measure of quantity choices. These data span 2010 to 2015, during which connections grew from 900,000 to 1.5 million water connections, generating nearly 100 million transaction records. Connections are divided into residential (90%), semi-business (4%), commercial (5%), and industrial (1%). This paper focuses on residential and semi-business connections.[7] The monthly meter reading is used as the measure of consumption for each connection. Meter readings are rounded to the nearest cubic meter limited to monthly readings below 100 m^3 in order to exclude pipe leaks as well as meter reading errors (readings above 100 m^3 compose less than 0.7% of the sample).

Since the object of interest for welfare is household-level demand, the billing records are merged to survey data recording information on households using each connection as well as demographics for the owners of each connection. To evaluate the performance of the water provider, the regulator contracted with a local university to conduct the Public Assessment of Water Services survey in three waves from 2008 to 2012, surveying around 50,000 water connections.[8] The connection survey records the total number of households and people using each connection, providing the primary measure of sharing behavior. The connection survey also includes basic demographics for the owner of the connection. The demographics used in the analysis include household size, dwelling type (whether apartment, duplex, or single house), number of employed household members, and whether the head of household is employed in a low-skill profession.

A full analysis of water usage in Manila also requires measuring households that are un-

[7]Semi-business connections are largely composed of households that own small roadside stands. These connections provide a useful source of price variation which is described in Appendix A.VIII.iii.

[8]Appendix A.II describes the sampling design and tests the extent to which it may be correlated with demographic characteristics. This exercise finds relatively weak correlations in economic terms with key characteristics.

connected to the bulk water system and instead use water from local vendors. The 2010 Census of Population and Housing records demographics for households that are using water from alternative sources. Households are defined as connected to the piped network if they report using a "faucet community water system" for their cooking, laundry, or bathing needs.[9] Alternatively, households are considered to be using from a water vendor if they report any other category, which includes deep wells, peddlers, or other sources.[10] Census data is merged geographically with the connection survey for 599 wards in Manila.

For additional information on water sharing relationships between households, I fielded a small, non-representative survey of 600 households via mobile phone application. This survey asks about issues associated with using different water sources, basic demographics, as well as how the bill and connection fee are paid for sharing households. Appendix A.IV compares this survey to the connection survey finding roughly similar rates of sharing and demographic characteristics.

The water provider also maintains call records of all service complaints, which can be linked to individual connections. This data is used to measure instances where households face unanticipated shocks to their water source, which are used to identify the marginal costs of using water from a neighbor. Finally, the water provider tracks data on the marginal cost of pumping an additional cubic meter of water as well as the costs associated with installing new connections. These data provide inputs into the provider's budget constraint, which are necessary for computing counterfactual pricing regimes.

With information on water source and monthly expenditures as well as household income, the 2011 Community Based Monitoring System data cover over 70% of households in Pasay City — a large area with a population of 500,000 in downtown metro Manila. These data are used to impute income for households outside of the sample according to the following demographic measures: age of household head and indicators for low-skilled employment of household head, residing in a duplex, residing in a house, as well as all possible household sizes and number of employed household members. The imputation also includes all possible interactions between these measures.

[9] The census was designed largely for rural populations so it does not directly ask whether households are connected to a piped water network.

[10] Since demographic questions in the census and connection surveys are worded differently, Table 12 of Appendix A.III compares measures across the two data sources for households that own connections finding broadly similar patterns.

4 Patterns of Water Use

The welfare consequences of informal sharing networks depend on both the types of households that choose to share with neighbors as well as the costs of these networks. To provide some descriptive evidence, Table 1 merges the connection survey of water connections with billing records associated with each connection. The columns group connections by the number of households they serve: one, two, or at least three. Demographics for connection owners indicate that smaller, slightly younger, and poorer households are more likely to share water with neighbors. Sharing households are over twice as likely to reside in duplexes possibly because transporting water between neighbors requires less effort for duplex residents.

Larger numbers of households sharing a water connection are correlated with more total water usage, but less water per person.[11] This decline may have many possible drivers: (1) sharing households face higher prices by locating at a higher point on the tariff schedule (Figure 1), (2) sharing households may face hassle costs in fetching water from neighboring taps, (3) these price effects may incentivize low-demand households to select into sharing relationships, consistent with the demographic patterns, and (4) low-demand households may have idiosyncratic preferences for sharing relationships. On average, monthly water bills (inclusive of service fees) take up around 3% of household income, which does not include other unobserved costs in accessing water. These findings suggest that policies which affect sharing choices can have significant impacts on water consumption as well as monthly expenditures.

Table 1 also computes the total number of households using from each source. This calculation requires the additional assumption that households sharing with one or two other households are uniquely linked to one connection. This way, the population of households connected to the network can be computed by summing the number of households across all connections. This assumption is consistent with the presence of fixed costs in setting up a connection or a sharing relationship so that households stay committed to one water source. To the extent that households are drawing water from multiple connections, this assumption may lead to an overestimation of the population engaged in sharing relationships. With this assumption, 54% of connected households receive their water from an individual connection, 21% are involved in a sharing relationship with a single neighbor, and 25% are in a sharing relationship with 2 or more other households. These proportions do not take into consideration households that are using from water vendors, which are discussed in more detail below.

[11] Total People Served by a connection includes both the owning household and any other households using a connection.

Table 1: Water Connections by Number of Households Served

	Serving One Household	Serving Two Households	Serving Three or More Households
Owner Demographics			
Age (Head of HH)	46.8	46.4	45.1
HH Size	5.32	4.68	4.49
Total Empl.	1.64	1.43	1.32
Low-Skill Emp. (Head of HH)	0.15	0.19	0.21
Inc. (USD/Mo. Imputed)	405	379	369
Apartment	0.24	0.13	0.13
Single House	0.62	0.52	0.38
Duplex	0.14	0.35	0.49
Connection Attributes			
Water (m3/mo.)	24.66	32.30	42.59
Water Bill (USD/Mo)	10.32	15.06	22.54
Total People Served	5.32	8.94	15.29
Water per Person	5.09	3.86	3.04
Sample			
Connected HHs	42,798	16,930	19,896
Share of Connected HHs	0.54	0.21	0.25
Months per Connection	52	49	45

The data include billing records merged with the connection survey. Income is imputed using location, demographic variables, and their interactions.

Each water connection contributes over four years of billing records on average.

Since the connection survey only includes households connected to piped water, census of population and housing data is required to learn about households using from water vendors. In this data, 6% of the population report using vendors, which is composed of 76% of households using from a "deep well" and 17% using from a "peddler" while the remaining fall into an "other" category. Anecdotal evidence suggests that groundwater from wells often has very low quality. Appendix Table 12 compares means of demographic variables between connection owners from the connection survey and all households from the census. Demographic variables are similar between connection owners in the connection survey and the census data. Households using from vendors have lower imputed incomes, greater probability of low-skill employment, and slightly fewer members than own-use households. Across most dimensions, households using from vendors more closely resemble households that share with their neighbors as opposed to households that use their connections individually.

5 A Model of Household Water Use

This section presents a stylized model of water source choice and monthly quantity decisions at the household level. This model allows for an analysis of the impacts of pricing policies on both source and quantity margins, which are often equally relevant to policymakers (McIntosh [2003], Komives [2005]). Using both margins also enables the estimation to include cross-sectional variation in sources alongside time-series variation in quantities to recover a broad set of demand and cost parameters.

Let the city of Manila be partitioned into neighborhoods, where each neighborhood is defined by a set of households that are physically close enough to share water connections with each other. In the context of Manila, these neighborhoods are often characterized by dense clusters of houses or duplexes. This partition embeds the assumption that neighborhoods are mutually exclusive so that a household located in one neighborhood cannot share water with households in a nearby neighborhood. Restricting the choice set in this way is necessary for computation but may in some cases overestimate the costs that households face in finding a sharing partner.[12]

In the first step, each household makes a one-time decision about their water source between 1) purchasing a connection, 2) using a neighbor's connection, or 3) buying from a local water vendor. These decisions result in a set of water connections for each neighborhood ranging

[12]This limitation is discussed in more detail in Section 6.3.

from an empty set (if all households use from a vendor) to the whole neighborhood (if each household uses an individual connection). Let each connection, C, serve S households that have decided to share the connection. For ease of notation, let the subscript C indicate connection-level attributes, summing over all households using from connection C. This model does not allow for households to choose different water sources over time, which is consistent with large fixed costs involved in initially purchasing a connection or establishing a sharing relationship. Section 5.2.1 extends the model to allow for cases where households may exogenously lose access to their current water source and must substitute toward another nearby source.

In the second step, each household then chooses how much water to use each month from their source. Since water sources provide utility through monthly consumption, it is useful to describe the decision process in reverse order, first characterizing monthly water consumption choices (taking the initial source choice as given) and second, characterizing the initial water source choices.

5.1 Consumption Choices for Households Using Piped Water

In choosing how much water to consume each month, households using piped water face the difficult task of coordinating water consumption among all household members. For example, consider the head of a household trying to manage water consumption across all of their household members each month. This task would involve monitoring the meter located outside of their dwelling and coordinating water use among all members.[13] Consistent with these challenges, mobile survey data indicate that 43% of surveyed household members do not recall how much they paid for water in the last month while 64% do not recall the amount that their household consumed in the previous month.

As a simple way to capture this coordination task, let each household i choose a goal for water use each month, w_i, knowing that unobserved consumption shocks, ϵ_i, may increase or decrease their realized amount of usage. Consumption shocks, ϵ_i, are assumed to be unobserved to the household and independent across time, bringing the total realized water consumption each month to, $W_i = w_i + \epsilon_i$. Household utility over total water, W_i and a

[13] Given that many developing countries feature steeply increasing prices with monthly usage, these coordination challenges may magnify any distortions caused by non-linear pricing.

numeraire good, x_i is assumed to take the following shape:

$$\begin{aligned} U(x_i, W_i) &= x_i + (1 - H_i)W_i - \frac{1}{2\alpha_i}(W_i - \Gamma_i + \alpha_i)^2 \\ W_i &= w_i + \epsilon_i \\ \epsilon_i &\sim \mathcal{N}(0, \Sigma_\epsilon) \end{aligned} \tag{1}$$

Following from quasi-linear preferences, this quadratic form for utility eliminates income effects and includes a natural satiation point for water consumption, capturing the tendency of households to target specific consumption levels despite low or zero marginal prices as shown in Figure 7.[14] Household preferences for water are characterized by Γ_i, which can be interpreted as a fixed preference for water each month and α_i, which captures the extent to which households are sensitive to price. The exogenous shock to consumption, ϵ_i, is assumed to be drawn from a multivariate normal distribution, Σ_ϵ with zero mean and variance terms, σ_ϵ^2, allowing for correlation with neighboring households i and j through term $\rho_{i,j}^\epsilon$. Although households cannot anticipate the exact realization of ϵ_i, they are assumed to understand the distribution of possible consumption shocks, allowing them to maximize expected utility.

Households that are using water from a neighbor's tap may face an additional utility cost per gallon of water H_i. This hassle cost captures any inconvenience associated with using an additional unit of water from a neighbor's tap.[15] The mobile survey reveals that among households accessing water from a neighbor's connection, 14% fetch water with containers, 25% access water from a single hose/pipe provided by the connection owner, and 61% are able to fully connect their household plumbing near the water meter. These different modes of access suggest that sharing households may face higher time costs fetching water or lower water pressure from piping water additional distances.

To account for idiosyncratic preference shocks for water each month, let household's fixed preference, Γ_i, be decomposed into a static, household-specific component, γ_i, and an idiosyncratic component, η_i, which is assumed to follow a multivariate normal distribution among sharing partners:

$$\begin{aligned} \Gamma_i &= \gamma_i + \eta_i \\ \eta_i &\sim \mathcal{N}(\mathbf{0}, \Sigma_\eta) \end{aligned}$$

[14]This utility form builds on work by Einav et al. [2013] who use a similar form to characterize extensive and intensive choices of health insurance.

[15]Accordingly, the connection owner is assumed to face zero hassle cost.

This preference shock is also assumed to be independent of the consumption shocks, ϵ_C. Aggregating these preferences shocks over all households sharing connection, C, produces a joint shock, η_C, according to the convolution of these random normal variables with a variance equal to $\sum_{i=1}^{S}\sigma_\eta^2 + \sum_{i=1}^{S}\sum_{j>i}^{S}\rho_{i,j}^{\eta}\sigma_\eta^2$.

5.1.1 Determining the Budget Constraint

The household problem is additionally complicated by the presence of an increasing price schedule for water; as a household uses more water in one month, they face higher marginal prices. Figure 1 provides the increasing tariff in Manila. This price schedule produces kinks in household budget constraints leading to non-linear demand functions. Let $P(W_C) = \{p^1, ..., p^K, \overline{W}^0, ..., \overline{W}^K\}$ describe the tariff schedule as a function of total consumption for the water connection, $W_C = \sum_{i=1}^{S} W_i$. The tariff includes K segments where segment k corresponds to price p^k for water consumption between the consumption thresholds (or kink points), $\overline{W}^{k-1}$ and $\overline{W}^k$:[16]

$$P(W_C) = \begin{cases} p^1 & \text{if } W_C \in [\overline{W}^0, \overline{W}^1], \\ p^2 & \text{if } W_C \in [\overline{W}^1, \overline{W}^2], \\ ... & ..., \\ p^K & \text{if } W_C \in [\overline{W}^{K-1}, \overline{W}^K] \end{cases}$$

Given prices and household utility, the final ingredient for characterizing consumption choices is the budget constraint. The budget constraint includes monthly household income, Y_i, and monthly fixed costs of using the water connection, F_i, which are assumed to capture three components: (1) monthly service fees, (2) initial connection fees amortized into monthly terms for the duration of the connection, and (3) unobserved, fixed hassle costs associated with setting up a connection, which are also amortized into monthly terms. Data does not track household water source choices over time in a comprehensive way, which limits the extent to which the model can distinguish between one-time connection costs and monthly service/maintenance costs. This limitation means that the model cannot address the optimal timing of fixed payments or capture potential heterogeneity in monthly fixed costs.

Sharing households must negotiate over how to pay the monthly fixed costs (including the fixed fee paid to the firm and any unobserved costs of using a connection). The model assumes that households commit to dividing any fixed costs at the start of the sharing

[16] In this case in Manila, $\overline{W}^0$ is zero while $\overline{W}^K$ is infinity.

arrangement, allocating F_i to each household. Under quasi-linear utility, any fixed transfers are exogenous to monthly consumption decisions and Section 5.2 later discusses how fixed costs are divided between sharing households.

Water sharing relationships impact monthly consumption choices through the budget constraint.[17] The model assumes that the monthly water bill is split evenly between all sharing households. Descriptive evidence suggests that challenges observing and coordinating usage within members of a household also extend across households. Among owners of the water connection responding to the mobile survey, 28% report splitting the water bill evenly or alternating payment each month, 27% charge sharing households a fixed amount each month regardless of their consumption, and 36% charge other users a price according to their usage.[18] Evenly splitting the bill, alternating payment, or charging a fixed amount provide easy ways to extract compensation without requiring specific information on the monthly consumption of each sharing household. Yet, these methods also introduce free-rider incentives since households do not have to fully pay for their contributions to the total water bill. Since data limitations prevent identifying the specific type of contract associated with each shared connection, the model assumes that all households evenly split the bill and explores the theoretical and empirical implications of alternate models in Appendix A.VI. The model captures bill splitting by evenly dividing the total monthly payment for a water connection across the S sharing households.

Normalizing the price of the numeraire good, x, to one, the budget constraint for household, i, using a connection with households, S, takes the following form given that total consumption for their connection, W_C, falls between thresholds, $\overline{W}^{k-1}$ and $\overline{W}^k$:

$$\begin{gathered}(Y_i - F_i) = x_i + \frac{B_{C,k}}{S_C} \quad \text{if } \; W_C \in [\overline{W}^{k-1}, \overline{W}^k] \\ \text{Where:} \\ B_{C,k} = p^k(W_C^* - \overline{W}^{k-1}) + \sum_{j=1}^{k-1} p^j(\overline{W}^j - \overline{W}^{j-1})\end{gathered} \tag{2}$$

The full budget constraint is a piece-wise expansion of equation (2) over all K price thresholds. $B_{C,k}$ represents the total water bill for connection C given the tariff schedule. This term is calculated by multiplying each price by the volume of consumption exposed to that price.

[17]While the formation of these relationships is discussed in more detail in Section 5.2, this section takes these sharing relationships as exogenously given.

[18]The remaining households report "I don't know" or "other."

5.1.2 Maximizing Expected Utility

When households share a water connection, the consumption choice for each household becomes a strategic decision. Since prices are non-linear, the total consumption level for each connection, W_C, determines the marginal price faced by all households. For example, with an increasing tariff like Figure 1, one household's choice to increase consumption may produce a higher marginal price for other sharing households.

Substituting the budget constraint (equation (2)) into the household utility (equation (1)) for the numeraire good x_i simplifies the household's maximization problem to a single choice variable — targeted consumption, w_i. The household's expected utility problem over consumption shocks ϵ_C then takes the following form:

$$max_{w_i} \sum_{k=1}^{K} \int_{\overline{W}^{k-1}-w_C}^{\overline{W}^{k}-w_C} (Y_i - F_i) - \frac{B_{C,k}}{S_C} + (1-p_s)W_i - \frac{1}{2\alpha}(W_i - \Gamma_i + \alpha_i)^2 d\epsilon_C$$

Where ϵ_C is the convolution of consumption shocks across sharing households distributed as follows:

$$\epsilon_C \sim \mathcal{N}(0, \sigma^2_{\epsilon_C} = \sum_{i=1}^{S} \sigma^2_\epsilon + \sum_{i=1}^{S} \sum_{j>i}^{S} \sigma^2_\epsilon \rho^\epsilon_{i,j})$$

To solve this system, this approach assumes that households follow a perfect information, pure-strategy Nash-Equilibrium in choosing consumption. Using this structure, households compute their best response consumption functions to arrive at a symmetric equilibrium taking consumption for their sharing partners as given.

Taking the derivative of expected utility with respect to w_i (holding sharing partners' choices w_{-i} fixed) while letting $\Phi(.)$ equal the standard normal cumulative distribution function produces the following first-order condition:

$$0 = -w_i + \Gamma_i - \alpha_i \Big[H_i + \frac{MP_C}{S}\Big] \tag{3}$$

Where:

$$MP_C = \sum_{j=1}^{K} p^j \Big[\Phi\Big(\frac{W_C - \overline{W}^j}{\sigma_{\epsilon_C}}\Big) - \Phi\Big(\frac{W_C - \overline{W}^{j-1}}{\sigma_{\epsilon_C}}\Big)\Big] \tag{4}$$

The first order condition (equation (3)) reveals that consumers balance their fixed preference for water, Γ_i, against their price sensitivity, α_i, hassle costs, H_i, and expected marginal price, MP_C (scaled by the number of sharing households, S). Expected marginal price MP_C (equation (4)) intuitively captures the extent to which increasing household consumption, w_i — lifting total consumption, W_C — exposes the entire connection to higher prices in expectation. This first-order condition characterizes the best-response functions for each household.

To arrive at a tractable expression for water demand, it is helpful to simplify equation (4) by assuming that households approximate the normal distribution with a simple uniform distribution. This assumption implies that households give zero likelihood to extremely high or low consumption shocks and even likelihood to moderate consumption shocks. From an empirical perspective, this function allows for an analytical solution to the demand function while ensuring that the estimation can rely on normally distributed errors to estimate rich preference heterogeneity. I use the following piece-wise function form to approximate the normal cumulative distribution:

$$g(z) = \begin{cases} 0 & \text{if } z < -\sqrt{2} \\ \frac{1}{2} + \frac{z}{2\sqrt{2}} & \text{if } z \geq -\sqrt{2} \text{ and } z \leq \sqrt{2} \\ 1 & \text{if } z > \sqrt{2} \end{cases}$$

Appendix A.V graphs this approximation against a standard normal cumulative distribution function and simulates optimal demand under both distributions, finding negligible evidence that this approximation would substantially bias the estimates or affect welfare measures. After substituting the normal cumulative distribution function $\Phi(.)$ for $g(.)$ in equation (4), expected marginal price becomes the sum of potentially overlapping piece-wise functions, meaning that the optimal demand will also be piece-wise. I proceed by describing the different pieces of demand before putting the pieces together to form the full demand schedule.

When households face very small consumption shocks, there is no probability that these shocks expose households to multiple prices. Therefore, solving the best-response function (summing across all households using connection C) yields a simple expression for optimal targeted demand:

$$w_C^j = \gamma_C - \alpha_C[H_C + \frac{p^j}{S}] \tag{5}$$

This case produces a linear demand function where α_C can be readily interpreted as the price sensitivity parameter while γ_C captures the portion of demand that is invariant to prices. In this case, demand is only sensitive to the marginal price, p^j, local to each segment of the tariff schedule. This demand function captures moral hazard from bill-splitting by dividing the price sensitivity, α_C, by the total number of households using the connection, S. Comparing demand among shared connections to individual connections, the extent of over-consumption at an average price of p can be roughly summarized as $\frac{S-1}{S}\alpha_C p$. The price sensitivity, α_C, also attenuates the extent to which the hassle cost, H_C, enters demand.

This expression for demand holds when total demand (including the fixed preference shock η_C) falls sufficiently between two price thresholds:

$$\overline{W}_l + \sqrt{2}\sigma_{\epsilon_C} \ \leq w_C^j + \eta_C \ \leq \ \overline{W}_{l+1} - \sqrt{2}\sigma_{\epsilon_C}$$

These conditions are less likely to be satisfied when connections have high variance in their consumption shocks, σ_{ϵ_C}, and face small distances between price thresholds, $\overline{W}_l$ and $\overline{W}_{l+1}$. In this case, consumption shocks may expose the connection to different prices and households need to consider the probability of facing these prices in choosing their optimal consumption targets. Allowing for households to be exposed to price thresholds j through l, the general solution to the best-response function for the connection C again yields targeted demand as follows:

$$\begin{aligned}
w_C^{j,l} &= \frac{1}{V^{j,l}}\Big(\gamma_C - \alpha_C[H_C - \frac{MP^{j,l}}{S}]\Big) \\
\text{Where :} & \\
MP^{j,l} &= \frac{1}{2}\big[p^j + p^l - \sum_{k=l}^{j}\overline{W}^k(\frac{p^k - p^{k-1}}{\sqrt{2}\sigma_{\epsilon_C}})\big] \\
V^{j,l} &= 1 + \frac{\alpha_C}{2S}(\frac{p_j - p_l}{\sqrt{2}\sigma_{\epsilon_C}})
\end{aligned}$$

The demand function takes a similar form to the linear case, including the fixed preference while weighting the marginal price by the price sensitivity and degree of sharing. Unlike in the linear case, connections face an average of all possible marginal prices, $MP^{j,l}$, that can be realized from their consumption shocks. All terms are also weighted by a variance term, which depends on the gap between the highest and lowest prices j and l. This expression holds for levels of demand between the following thresholds:

$$B_{j,l} \le W_C^{j,l} \le T_{j,l}$$

Where :

$$B_{j,l} = max\{x \in I_{j,l} \,|\, x \le W_C^{j,l}\} \;\;,\;\; T_{j,l} = min\{y \in I_{j,l} \,|\, y \ge D_C^{j,l}\}$$
$$I_{j,l} = \{\overline{W}_l + \sqrt{2}\sigma_{\epsilon_C}, \overline{W}_l - \sqrt{2}\sigma_{\epsilon_C}, \overline{W}_j + \sqrt{2}\sigma_{\epsilon_C}, \overline{W}_j - \sqrt{2}\sigma_{\epsilon_C}\}$$
$$W_C^{j,l} = w_C^{j,l} + \frac{\eta_C}{V^{j,l}}$$

Bottom, $B_{j,l}$, and top, $T_{j,l}$, thresholds sandwich connection demand $W_C^{j,l}$ and are drawn from the full set of possible thresholds, $I_{j,l}$, above and below the tariff kink points, $\overline{W}_l$ and $\overline{W}_j$. In cases with high variance in consumption shocks, kink points j and l may span many other marginal prices, k where $l \le k \le j$. At the same time for small consumption shocks and large distances between price thresholds, this expression simplifies to the linear case in equation (5).

Constructing the full demand schedule for each connection requires simply connecting these demand segments to span the full range of consumption choices. However since targeted demand depends on the variance in consumption shocks, σ_{ϵ_C}, there can be different optimal demand functions for different levels of σ_{ϵ_C}. To determine a single piecewise demand function, I take the approach of (1) assuming ex-ante bounds on σ_{ϵ_C}, (2) constructing the implied optimal demand function given these bounds, and (3) estimating the model to see if the estimated σ_{ϵ_C} rejects these bounds in the data. Appendix A.IX provides the results of this exercise identifying a range of σ_{ϵ_C} that is consistent with the data. Appendix A.VIII.iv describes the demand function given these bounds along with a visual example demand curve in Figure 11.

5.1.3 Deriving the Likelihood Function

Given the optimal consumption choice, W_C, for each connection, observed consumption for a particular connection, Q_C, becomes the sum of optimal consumption and the total consumption shock for that connection, ϵ_C:

$$Q_C = W_C + \epsilon_C \tag{6}$$

The probability of observing any consumption outcome, Q_C, can be expressed as the sum of the probability that the preference shock, η_C, locates the households within a particular price segment multiplied by the joint probability that the consumption shock, ϵ_C, and the preference shock are equal to Q_C conditional on being in this price segment.

These probabilities allow for the derivation the likelihood function for each consumption realization, using some additional notation.[19] For each demand segment, denote the bottom threshold for the preference shock needed to reach that segment as, $\overline{\eta}_{B,j} = B_{j,l} - w^{j,l}$ and the top threshold as $\overline{\eta}_{T,l} = T_{j,l} - w^{j,l}$.[20] Define the variance-adjusted preference shock as $\Psi = \frac{\eta_C}{V_{j,l}}$. Let the sum of the two stochastic terms equal $\zeta = \Psi + \epsilon_C$. For any random variable, x, denote the cumulative distribution function as F_x and the corresponding probability distribution function as f_x. Also, let ϕ and Φ denote the standard normal p.d.f. and c.d.f. respectively. Given equation (13), the contribution to the likelihood function for consumption observation for each connection can be written as:

$$\sum_j f_\zeta(Q - W^{j,l})\Big[F_{\Psi|\zeta=Q-W^{j,l}}(\overline{\eta}_{T,l}) - F_{\Psi|\zeta=Q-W^{j,l}}(\overline{\eta}_{B,j})\Big] \tag{7}$$

This expression takes the sum of the densities of ζ for each consumption level weighted by the probability that the preference shock Ψ locates the desired consumption level for that household between the two thresholds, $\overline{\eta}_{T,l}$ and $\overline{\eta}_{B,j}$ over all thresholds. Given that η_C and ϵ_C are assumed to be independent and normally distributed, the terms in equation (7) can be rewritten using both the convolution and conditional distributions of multivariate normals:

$$\begin{aligned}
f_\zeta(Q - W^{j,l}) &= \frac{1}{\sigma_\zeta}\phi\left(\frac{Q - W^{j,l}}{\sigma_\zeta}\right) \\
F_{\Psi|\zeta=Q-W^{j,l}}(\overline{\eta}_{T,l}) &= \int_{-\infty}^{\overline{\eta}_{T,l}} f_{\Psi|\zeta=Q-W^{j,l}}(x)dx \\
&= \int_{-\infty}^{\overline{\eta}_{T,l}} \frac{1}{\sigma_\zeta\sqrt{1-\rho^2}}\phi\Big(\big(\frac{xV_{j,l}}{\sigma_\eta} - \rho q^{j,l}\big)\frac{1}{\sqrt{1-\rho^2}}\Big)dx \\
&= \Phi\Big(\big(\frac{\overline{\eta}_{T,l}V_{j,l}}{\sigma_\eta} - \rho q^{j,l}\big)\frac{1}{\sqrt{1-\rho^2}}\Big)
\end{aligned}$$

Where:

$$q^{j,l} = \frac{(Q - W^{j,l})}{\sigma_\zeta}, \quad \sigma_\zeta = \sqrt{\frac{\sigma_\eta^2}{V_{j,l}^2} + \sigma_{\epsilon_C}^2}, \quad \rho = \frac{\sigma_\eta}{\sigma_\zeta V_{j,l}}$$

---

[19]By allowing households to anticipate exogenous shocks to their consumption, this approach builds on the standard framework of Burtless and Hausman [1978] and Moffitt [1986] to estimate demand over non-linear budget constraints, which has been recently implemented by Szabó [2015] and McRae [2014].

[20]Conditions on η_C defining each segment are given in more detail in Appendix A.VIII.iv equation (13).

Given these terms, equation (7) can be rewritten in the following way:

$$LL = \sum_{j} \frac{1}{\sigma_\zeta} \phi(q^{j,l}) \Big[\Phi\Big((\frac{\overline{\eta}_{T,l} V_{j,l}}{\sigma_\eta} - \rho q^{j,l}) \frac{1}{\sqrt{1-\rho^2}} \Big) - \Phi\Big((\frac{\overline{\eta}_{B,j} V_{j,l}}{\sigma_\eta} - \rho q^{j,l}) \frac{1}{\sqrt{1-\rho^2}} \Big) \Big] \quad (8)$$

Summing each part of the demand function in equation (13) in the likelihood expression in equation (8) computes the probability of observing consumption outcomes for each connection. In order to include multiple time periods, the model captures variation in consumption over time by assuming that each household receives independent draws of preference shocks, $\eta_{i,t}$, and consumption shocks, $\epsilon_{i,t}$, which aggregate at the connection-level to $\eta_{C,t}$ and $\epsilon_{C,t}$ respectively for each month, t. All other parameters are assumed to remain stable over time. The full log-likelihood is computed by summing the log of the likelihood in equation (8) over all connections and time periods.

5.1.4 Households using from Water Small-Scale Providers

While the vast majority of households use piped water, around 6% of households still source their water from small-scale water vendors. In this model, these small-scale sources are collapsed into a single vendor category, which is modeled as a perfect substitute to piped water and is characterized by a monthly fixed cost, F_V, and a single marginal price, p_V. The monthly fixed cost, F_V, is assumed to capture primarily fixed differences in quality of service between vended water and other sources. Similarly, the marginal price, p_V, includes both the price per unit as well as any differences in water quality or ease of use relative to piped water. Demand for this source becomes a simple linear function of prices given by equation (9).

$$w = \Gamma_i - \alpha_i p_V \quad (9)$$

Since the data do not record monthly consumption information for households using from vendors, these households do not contribute to the likelihood function (equation (8)).

5.2 Choosing a Water Source

This section describes a sequential game where households choose their water sources given their demands for water. This game results in a unique equilibrium of source choices according to the preference parameters for each household as well as the costs of using each water source.

Choosing a water source becomes a strategic decision since each household's water connection can also serve as a water source for other households in a neighborhood. The model assumes that each set of neighboring households plays a full-information, sequential game, deciding whether to purchase their own tap, share with a previously connected household, or purchase from a vendor. Assuming a sequential order of moves acts a refinement to rule out situations with multiple equilibria. Eliminated equilibria mainly include cases that are both strictly dominated and unrealistic such as having all households use from vendors when they would prefer to share a connection.

Let superscripts index households by the order in which they move. Let a^i denote the water source choice for each household i. The choices consist of (1) using from a vendor, v, (2) purchasing a connection, c, or (3) sharing with neighbor, j, given by s^j. The first mover faces a simple choice set defined by $A^1 = \{v, c\}$.

For all other households, define the following function that maps a previous household's choice to connect into a possible sharing opportunity:

$$g(a^j) = \begin{cases} s^j & \text{if } a^j = c \\ \emptyset & \text{else} \end{cases}$$

Given this function, the choice sets for households $i > 1$ consist of $A^i = \{v, c, g(a^1), ..., g(a^{i-1})\}$ where households only have the option to share with neighbors that have previously connected.

The payoffs equal the expected utilities from each source over the duration of usage. Since the data do not measure the exact duration for which households reside in a location using a water source, the model abstracts away from the timing of water use by assuming for simplicity that (1) households do not discount their future utilities and (2) all households demand water for an equal number of months.[21] This approach limits the extent to which this model can examine the timing of any one-time connection fees or other upfront costs of water connections. Instead, the model collapses one-time fees into a single monthly fixed fee. Any heterogeneity in either the duration of demand or discount factors is captured by estimates of this monthly fixed fee.

To compute the payoffs from using vended water, households simply compute their expected utility given by equation (1) according to their demand in equation (9). Let $u^i(v)$ capture this payoff. The payoff from owning a connection depends on whether other households

[21]These assumptions do not have a substantive impact on the results because all model primitives are independent of time.

choose to use the connection as well. Households are assumed to have perfect information about each other's preferences including the distributions of preference and consumption shocks. In order to calculate surplus from sharing, let the set S^i contain all households that have chosen to use household i's connection (including the owner).

In this model, sharing affects the utility of the owner in three ways. First, increasing the number of sharing households splits the total water bill over more households, exacerbating any moral hazard distortions as illustrated in equation (13). Second, under non-linear tariffs, including more consumption on the same connection may alter the marginal prices faced by all sharing households. Third, in this model, owning households ($i = 1$) are assumed to pay the full fixed cost, F, and sharing households commit ex-ante to providing a fixed transfer to the owner each month to compensate the owner. This transfer allows sharing households to help contribute to the fixed costs of owning and maintaining the connection. 60% of sharing households in the mobile survey report splitting some or all of the initial connection fee with their neighbors; for the remaining households, the owner reports paying the full connection fee.

To capture this behavior, the model assumes that sharing households offer transfers to the owner which equally split their surpluses from using the connection. This setup is analogous to a Nash-Bargaining framework with equal bargaining weights. While the division of surplus does not affect the efficiency of sharing relationships, it has normative implications for the types of households that are able to benefit from these arrangements. Three features of the context in Manila support this assumption of a roughly equal division of surplus. First, these sharing arrangements feature long-term relationships between close neighbors who are also often family members. Second, small areas often feature multiple sharing opportunities for each household, creating some competitive pressures on sharing arrangements. Third, in the mobile survey, sharing households appeal to an ethic of fairness as the primary reason for dividing monthly payments instead of other reasons such as matching tariff prices or conserving environmental resources.

Expected utility from using a connection alone is defined as $u^i(c|1)$. This expression can be computed by inputting demand from equation (13) with $S = 1$ into utility given by equation (1). Expected utility from owning and sharing a connection is defined as $u^i(c|S^i)$ and expected utility for household j from using a connection owned by neighbor i as $u^j(s^i|S^i)$.

Sharing surplus for the owner of the connection can be represented as $O^i = u^i(c|S^i) - max\{u^i(c|1), u^i(v)\}$ while surplus for each other sharing household j takes the form $G^j = u^j(s^i|S^i) - max\{u^j(c|1), u^j(v)\}$. Let $t_j(S^i)$ denote the transfer from sharing household j

to the owning household i. Following the outcome of a Nash bargain with equal weights, define $t_j(S^i) = \frac{S^i-1}{S^i}G^j - \frac{1}{S^i}O^i - \sum_{k\in S^i\setminus\{1,j\}} \frac{1}{S^i}G^k$. The total transfer that the owner receives from all sharing partners equals $T(S^i) = \sum_{j\in S^i\setminus\{1\}} t_j(S^i)$. For example, when only two households use the connection, then $T(\{1,2\}) = t_2(\{1,2\}) = \frac{1}{2}(G^2 - G^1)$, which simply divides the surplus in two. While these transfers are usually positive since the owner is assumed to shoulder all of the fixed costs of maintaining a connection, there may be cases where the owner will offer payment to other households to use their connection: for example, if the owner has a much larger demand for water than other users, the owner may benefit substantially from splitting the bill evenly with these other households. This case would be captured through a negative transfer.

From a welfare perspective, these transfers help to ensure that households form efficient sharing relationships. For example, transfers allow neighbors to incentivize others to purchase a connection in cases when owners would not otherwise be willing to own and share a water connection. Transfers also serve to reconcile a tension introduced by the sequential structure of the game where a first mover may be compelled into a sharing relationship by a second mover although the first mover would rather use the connection alone. Transfers eliminate this possibility by calculating surplus from sharing relative to the maximum expected utility from using a vendor or using a water connection alone.

Given these transfers, the payoff for household i choosing to connect takes the form $u^i(c|S^i) + T(S^i)$. Meanwhile, the payoff for household j sharing a connection with household i is $u^j(s^i|S^i) + t_j(S^i)$. Both the past and future decisions of all other households enter these payoffs by determining the resulting set of sharing households, S^i, which affects expected utilities directly as well as through transfers.

The last step in the specification of this game is determining the order in which households choose water sources. The model sequences households according to their expected net utility from owning a connection alone relative to using from a vendor, $u^i(c|1) - u^i(v)$. This arrangement matches the intuition that households with the greatest incentives to own and maintain a water connection would likely be the first to purchase a connection. Table 1 shows that for connections serving two households, the owner's household size is 10% larger than the buyer's household size on average. To the extent that household size serves as a proxy for demand, this finding provides suggestive evidence that larger demand households with greater utility from purchasing a water connection are also more likely to be owners. Given that transfers allow households to efficiently share surplus, the sequence is unlikely to have large impacts on the overall efficiency of the resulting arrangements. However, moving earlier may affect the welfare of particular households. On one hand, early movers can select to join high-value sharing relationships, crowding out these opportunities for later movers. On the other hand, later movers have a broader choice set of already-connected households to establish a sharing relationship.

The sequential game is solved using backward induction and considering only pure-strategy equilibria. First, the last mover, household I, identifies the choices that maximize its payoffs at each possible choice set reached by the game.[22] Given these choices, the next-to-last mover, household $(I-1)$, can calculate its payoffs for each corresponding choice set. Iterating backwards, this process characterizes a unique equilibrium set of source choices for all households. Appendix Figure 14 provides an extensive form for the game that is simulated in this paper.

5.2.1 Exogenous Interruptions to a Water Source

This section describes an extension of the model to allow for exogenous interruptions in a water source that induce households to substitute toward another water source. Exogenous interruptions produce useful variation in the types of water sources used by each household, serving an important role in estimating the hassle costs of transporting water from a neighbor. Due to shifts in the geology, road construction, or aging infrastructure, households sometimes experience underground breaks in the service pipe between the main line and the water meter. These leaks result in repair costs and water bills borne by households that can be large enough to force affected households to switch to an alternative water source.

Descriptive evidence suggests that the decision process to choose an alternative water source in response to a leak is much different from the initial choice households face in deciding on a long-term water source. The mobile survey asks households which source they would substitute towards if they were to be disconnected for at least three months. 30% report that they would "use a neighbor's tap" while the remaining responses are roughly evenly split between using a public or private deep well, fetching from a water truck, or filling bottles at a water refilling shop. Although shared connections vastly outnumber households using from vendors in the population as shown in Table 1, using a neighbor's tap is much less common in response to a sudden disconnection. Since sharing relationships often take the form of long-term contracts, households may face high costs in reaching a temporary sharing agreement with a neighbor. Additional descriptive findings discussed in Section 6.2 suggest that the decision to share with a neighbor is uncorrelated with the features of water demand

[22] In very rare cases where payoffs are equal, the tie-breaking rule prioritizes vendors, then individual connections, and finally shared connections ranked by the sequence of movers.

among both the household experiencing a leak as well as their neighbors. These patterns indicate that substitution may be driven less by minimizing the long-term costs of using each source and more by convenience such as immediate access or relationships between neighbors.

From the perspective of the sequential game, households experiencing leaks are assumed to be last-movers, which implies that other households cannot update their water source choices in response to a neighbor's leak. Additionally, the model assumes that all households do not anticipate any risks of leaks in making their initial connection decisions. Since leaks are rare events as discussed in Section 6.2, this assumption is unlikely to significantly affect predicted water source choices.

Let l index households experiencing a leak while j indexes connections in the same neighborhood (reaching a total of J connections). This model deviates from the initial water source choice by assigning an idiosyncratic preference term, $\xi_{A,l}$, to each alternative water source A — either using from a vendor, v, or sharing with a neighbor, s^j. $\xi_{A,l}$ captures the convenience of temporarily using each source following a sudden loss of water access. For example, $\xi_{A,v}$, may capture idiosyncratic, easy access to a neighbor's private deep well or a water refilling station. Similarly, $\xi_{A,l}$, might measure the relative social capital among potential sharing partners: nearby family members may be more willing to quickly share their connection. The model also includes the term F^S to capture any fixed costs associated with setting up a temporary sharing relationship relative to simply fetching from a vendor. Given these preference shocks, total utility for household l for using alternative A can be parameterized with the following expression:

$$\begin{gathered} U_l(A) = -\mathbb{1}\{A \neq v\} F^S + \xi_{A,l} \\ A \in \{v, s^1, ..., s^J\} \end{gathered} \tag{10}$$

F^S captures the relative substitution to sharing relationships and to water vendors. This model includes the additional assumption that $\xi_{A,l}$ is distributed Type I Extreme Value and is independent of preferences for water usage. Under these distributional assumptions, the model predicts that conditional on choosing a sharing relationship, households evenly distribute between sharing partners. The probabilities of households choosing each alternative source can be written as:

$$Pr[A = v] = \frac{e^{-F^S}}{e^{-F^S} + eJ} = G$$
$$Pr[A = s^j] = \frac{e}{e^{-F^S} + eJ} = \frac{(1-G)}{J}$$
$$\text{for } j = 1..J$$

Let L represent the total households affected by a single connection leak.[23] All L households in the neighborhood of a single leak are assumed to receive the same preference shock $\xi_{A,l}$ so that they follow identical substitution patterns.[24]

These assumptions allow for the specification of a likelihood function that expresses neighbor's consumption following a leak as a function of household preferences and the hassle cost, H. Define $LL(\gamma, \alpha, S)$ as the likelihood function in equation (8) in terms of the preference parameters— γ and α — as well as the number of households using the connection, S. Given the probabilities above, the likelihood for consumption derived from connection j after a nearby leak can be expressed as follows:

$$\begin{aligned} LL_{j,PostLeak} = & \frac{J-(1-G)}{J} \overbrace{LL(\gamma_j, \alpha_j, S_j)}^{L \text{ do not share with } j} + \\ & \frac{(1-G)}{J} \underbrace{LL(\gamma_j + \gamma_L, \alpha_j + \alpha_L, S_j + S_L)}_{L \text{ share with } j} \end{aligned} \tag{11}$$

The first term multiplies the standard likelihood for this connection by the probability that households L do not share with connection j. The second term adds the probability that L households choose to share with connection j following the leak. This probability is multiplied by the likelihood corresponding to joint consumption, which includes both preference terms from connection j as well from L households. This approach recovers the hassle cost, H, by comparing the fixed preferences for households experiencing the leak before and after they start using from a neighbor. Let *pre* take a value of one for months before a leak. Fixed

[23] L is greater than one in cases where connections serving multiple households experience a leak.

[24] This assumption may lead to a slight underestimation of the hassle costs since these households may actually distribute across many different sharing relationships, exposing them to lower points on the tariff schedule than if all households were clustered on the same connection.

preferences for L households take the following form:

$$\gamma_L = \sum_{l=1}^{\overline{L}} (\gamma_l - \alpha_l H_l)$$

$$H_l = \begin{cases} 0 & \text{if } l = 1 \text{ and } pre = 1 \\ H & \text{else} \end{cases}$$

Substituting to a neighbor's connection newly exposes the owner, $l = 1$, to the hassle cost H_i while other households $l \neq 1$ are assumed to be exposed to the hassle cost both before and after the leak. Section 6.2 details potential limitations of this assumption.

6 Specification and Estimation

The estimation strategy takes place in three steps. In the first step, preference parameters — fixed preferences γ, price sensitivities α, as well as variances for the consumption and preference shocks, σ_ϵ and σ_η respectively — are recovered by combining monthly consumption patterns with price variation. The second step analyzes a quasi-experiment where households are exogenously switched from individual connections to shared connections in order to estimate the hassle cost, H, associated with sharing. In the third step, monthly fixed costs associated with an individual connection, F, and using a water vendor, F_V, as well as the marginal price from vendors, p_V, are chosen to best match cross-sectional moments in the data, including the shares of households choosing each source.

6.1 Preference Estimation

Household preferences for water are estimated by maximizing the likelihood function in equation (8) for consumption recorded at each connection and month. This section first details how each set of preferences are parameterized in the estimation:

Fixed Preferences $\hat{\gamma_C} = (\gamma_C - \alpha_C H_C)$: Since consumption data is recorded for connections rather than for households, this approach estimates a separate $\hat{\gamma_C}$ for each connection from the mean observed usage for each connection. $\hat{\gamma_C}$ includes not only the total fixed preference parameters for households sharing a connection γ_C, but also any hassle costs faced by shared households weighted by the price sensitivity, $\alpha_C H_C$. Since hassle costs do not vary over time or with prices, they are indistinguishable from fixed preferences for consumption in this step of the estimation. I later decompose these estimates into household-level preferences and hassle costs as detailed in Section 6.3.

Price Sensitivity α_C : The estimation parameterizes household price-sensitivity terms α_i as a linear function of a vector of observable characteristics Z_i that may affect price sensitivity so that $\alpha_i = \alpha Z_i$. Z_i includes dummy variables for household size, type of dwelling, whether a household member is employed in a low-skilled industry, number of employed household members, and age of the head of household. Descriptive statistics for each variable are presented in Table 16 in Appendix A.VII. For shared connections, the data include information on household size separately for owners and sharing households; however, the other observables are recorded only for owners. For these owner-specific measures, the estimation simply assigns the owner's characteristics to sharing households. While this imputation may be reasonable for spatially correlated characteristics like type of dwelling, this approach may introduce bias if sharing households have systematically different numbers of employed individuals or ages. Estimates of price sensitivities in Table 2 exhibit little heterogeneity across many of these dimensions, which suggests that any bias from mismeasurement may have small effects in evaluating policies. Since household-specific price sensitivities α_i enter linearly into the connection-level demand function in equation (13), connection-level price sensitivity can be written simply as $\alpha_C = \alpha \sum_{i=1}^{S} Z_i$.[25]

The empirical setting in Manila provides a broad range of price variation to identify these price sensitivity terms separately from the fixed preferences. First, the non-linear pricing structure given by Figure 1 forces households to decide whether to consume above or below tariff kink points as they face monthly shocks to their demand. Second, regulated tariff changes over time affecting all households generate time-series variation in the tariff, which can be seen in Appendix A.VIII.ii. Third, I use rate-changes which affect individual connections differently over time described in more detail in Appendix A.VIII.iii. The structural model combines all of these sources of price variation in the estimation. The identification assumption is that the consumption and preference shocks, ϵ, η are normally distributed and uncorrelated with these price changes.

Preference and Consumption Shocks $\sigma_\eta, \rho_\eta, \sigma_\epsilon, \rho_\epsilon$: Heterogeneity in σ_η is separately according to connections serving one, two, or three plus households. The estimation also allows σ_η to vary non-parametrically according to quintiles of mean consumption at the connection-level. Specifically, the estimation includes a separate variance term for the bottom quintile

[25] The model does not control for seasonality or calendar time since the raw data exhibit negligible trends in water usage over time.

of consumption as well as a term for the top two quintiles of consumption.[26] Since larger users naturally have higher variances in their consumption, estimating a single σ_η across all connections produces large bias in the price sensitivity coefficients, α. This bias results from constraining large and small users to receive similarly scaled shocks and therefore enforcing that these users respond to the non-linear tariff schedule in similar ways. While using quintiles of observed consumption may be problematic since consumption choices are themselves endogenous, this approach approximates connection-specific σ_η terms, which would be infeasible given limited months of data available for each connection.

Following a similar strategy, the estimation recovers separate standard deviation terms at the connection-level for consumption shocks σ_ϵ according to the number of households using each connection. Since consumption shocks exhibit little heterogeneity with respect to observed consumption quintiles, I exclude these terms from the estimation to preserve parsimony.

The σ_η and σ_ϵ parameters are separately identified from the extent of consumption clustering around price thresholds relative to the overall variance of consumption. Greater clustering of consumption near price thresholds implies that households face small consumption shocks, allowing them to accurately target price thresholds. Similarly, wide variance in overall consumption is attributed to large preference shocks. Identification for these variance parameters relies heavily on the assumption that these parameters are normally distributed and uncorrelated with each other.

To recover correlations of preference and consumption shocks between neighboring households given by ρ_η, ρ_ϵ, this approach assumes that all household-level shocks are drawn from distributions with the same standard deviations — σ_η and σ_ϵ. Under this assumption, any differences in estimated standard deviations for connections serving different numbers of households can be attributed to correlations in shocks between households in the same neighborhood. Ordering households according to the sequential game in Section 5.2 and assuming small neighborhoods of only three households, let the correlation terms between the first-mover and any other households in the neighborhood, $\rho_{l,1}$, be determined in terms of the following expression where the second subscripts on the σ terms index the number of households using a connection:

$$\rho_{l,1} = \frac{\sigma_{l,2}^2 - 2\sigma_{l,1}^2}{2\sigma_{l,1}^2} \text{ for } l = \eta, \epsilon$$

The next expression extends this framework to identify the correlation between the second

[26]There is too little variation in the data to separately identify a full set of quintiles while also allowing σ_η to vary by sharing status.

mover and all other households in a neighborhood:

$$\rho_{l,2} = \frac{\sigma^2_{l,3} - 3\sigma^2_{l,1} - 4\rho_{l,1}\sigma^2_{l,1}}{2\sigma^2_{l,1}} \text{ for } l = \eta, \epsilon$$

Although sharing relationships and neighborhoods are limited to a maximum of three households, this approach can be used to identify correlations between an arbitrary number of sharing partners.

Sample : The estimation uses a 10% sample of connections from the connection survey (5,000 connections), which serves as the primary sample for the counterfactual exercises. This sample is combined with 2,151 connections from the quasi-experiment examined in the second step of the estimation (Section 6.2), which takes preference estimates from this step as an input. Together, these connections produce 349,265 months of consumption observations used in the preference estimation. All standard errors are bootstrapped using 60 repetitions. This sample does not include households using water from vendors since the data do not include information on consumption for these households. Section 6.3 details the imputation procedure used to infer preferences for these households. Computationally, the estimation minimizes the likelihood function using a trust-region algorithm provided with the analytical gradient vector and Hessian matrix of the likelihood function.

Estimates : Table 2 provides estimates for the different parameters as well as bootstrapped standard errors. The price sensitivity parameter, α, exhibits some statistically significant heterogeneity across demographic measures although the magnitudes of these estimates are relatively small compared to the intercept. Large households have smaller price sensitivities while employment among household members and age of household head tend to increase price sensitivity estimates. These estimates can be difficult to interpret economically since the elasticity of demand depends not only on these α terms, but also on the fixed preferences for water across households.

Estimates of the variance of consumption shocks, σ_ϵ increase slowly as the number of households using a connection increases. This finding provides suggestive evidence that shocks to sharing households may be negatively correlated. The standard deviation of preference shocks, σ_η, increases dramatically with average usage at baseline. After controlling for average usage at baseline, increasing the number of households per connection has a very small effect on the standard deviation of preference shocks. To interpret these estimates in economic terms and to compare to similar studies, I compute the elasticity of demand with respect to price at the connection-level. Given a linear demand structure, the elasticity of demand depends on both the price sensitivity and fixed preferences for water. As a simpli-

Table 2: Demand Estimates

	Estimate	Standard Error
Price Sensitivity : α		
Intercept	0.83	0.09
HHsize 4 to 5	-0.021	0.044
HHsize over 5	-0.098	0.053
Apartment	-0.036	0.052
Single House	0.051	0.043
Low Skill Emp.	0.016	0.053
Over 2 Empl. Members	0.023	0.042
HH Head 36 to 52 years	0.078	0.043
HH Head over 52 years	0.132	0.052
Consumption Shocks : σ_ϵ		
Single HH	4.10	0.07
Two HHs	5.30	0.18
Three HHs	5.78	0.30
Preference Shocks : σ_ψ		
Intercept	13.31	0.45
Below 1st Quintile Usage	-2.99	0.28
Over 3rd Quintile Usage	2.82	0.30
Two HHs	-0.44	0.43
Three HHs	1.13	0.52

Sample includes 349,265 billing months for 7,151 connections.

fied example, consider the calculation for elasticity under a single marginal price and linear demand specification:

$$\frac{\delta Q}{\delta P}\frac{P}{Q} = \alpha_C \frac{P}{\gamma_C - \alpha_C P}$$

According to this expression, increasing the fixed preference for water consumption, γ_C, while holding constant the price sensitivity, α_C, mechanically reduces the elasticity. Many previous studies estimate heterogeneity in γ_C according to demographics alongside a single α_C parameter for the population (McRae [2014], Olmstead [2009], Szabó [2015]). This approach builds on these previous studies by estimating heterogeneity in both terms, which allows for a broader range of possibilities, including that larger users may be relatively more sensitive to prices than smaller users. To compare estimates with previous research under non-linear pricing schedules, this approach follows the same formula for computing elasticities by (1) taking draws of ϵ and η, (2) computing predicted demand under prices that are 1% greater than current prices, and (3) comparing the average percentage change in demand before and after the price change. For tariff blocks equal to zero, prices are increased from 0 to 1 PhP (Szabó [2015]).

Table 3 presents connection-level demand elasticity estimates across terciles of (1) imputed income for the connection owner and (2) average monthly usage for the connection. Elasticities exhibit little heterogeneity across imputed income levels, but substantial variation across usage levels with larger users appearing much less sensitive to prices than smaller users. Since many low-income households share connections with neighbors, there is a weak correlation between income and usage-level, which helps account for the lack of heterogeneity in elasticity.[27] Heterogeneity in elasticities according to monthly usage is consistent with moral hazard when households share a connection: increasing the number of households per connection increases the water use for those connections while also exacerbating free-riding behavior. Different water using activities across the water demand profile may also produce this heterogeneity in elasticities. This heterogeneity provides a potential motivation for usage-based price discrimination. For example, matching larger, less price-sensitive users with higher prices may allow for a more efficient allocation of water while still covering costs of production. The demand estimation yields an average elasticity of 0.78, which can be interpreted as the percentage decrease in quantity due to a one percentage increase in price. This estimate also closely matches reduced-form elasticity estimates from price-change quasi-experiment detailed in Appendix A.VIII.iii. This estimate is in range of the 0.98 estimate from Szabó [2015] as well as similar studies in the developing world that find elasticities

[27] Noise in the income imputation process may also weaken any correlation between elasticity and income.

Table 3: Demand Elasticities for Water Connections

	Estimate	Standard Error
Mean	0.778	0.106
By Income		
1st Tercile	0.797	0.188
2nd Tercile	0.793	0.171
3rd Tercile	0.746	0.170
By Usage		
1st Tercile	1.271	0.213
2nd Tercile	0.717	0.147
3rd Tercile	0.634	0.135

Elasticities are calculated from 7,151 connections making 349,265 connection-months. Income is imputed based on demographics, locations, and their interactions. Standard errors are bootstrapped at the connection level.

between 0.01 and 0.81 (Diakité et al. [2009], Strand and Walker [2005]). Table 4 compares predicted consumption from the model to observed consumption in the data. This exercise finds negligible differences across average consumption and the variance of consumption for connections according to sharing status. Closely matching observed patterns, this model improves on previous approaches to demand estimation likely by recovering a separate fixed preference term for each connection as well as rich heterogeneity in the variance terms (Szabó [2015]).

Table 5 presents correlation estimates decomposed from the consumption and preference

Table 4: Model Performance

	Mean Usage (m3/month)			Standard Deviation		
	Observed	Predicted	Difference	Observed	Predicted	Difference
All	26.55	26.58	-0.03	16.70	16.64	0.06
		(0.18)			(0.15)	
Single HH	23.91	23.98	-0.07	14.78	14.70	0.08
		(0.17)			(0.14)	
Two HHs	30.89	30.83	0.06	16.89	16.99	-0.11
		(0.40)			(0.34)	
Three HHs	41.03	40.88	0.14	21.45	21.40	0.05
		(0.68)			(0.46)	

Table 5: Correlation Estimates

	Consumption Shock	Preference Shock
Owner-Buyer	-0.17	-0.35
	(0.06)	(0.07)
Buyer-Buyer	-0.17	0.21
	(0.16)	(0.12)

Errors are bootstrapped at the connection level.

shocks. Both the consumption shocks, ϵ and preference shocks, η demonstrate negative correlations between neighbors for owners and buyers. This finding suggests that neighboring households may intentionally form sharing relationships with whom their shocks are negatively correlated. This behavior may provide an insurance benefit in the presence of non-linear prices, reducing the risk that joint consumption exceeds price thresholds each month. Alternatively, these negative consumption findings may stem from households that frequently travel ensuring that at least one user is home to take care of the connection. For example, given the high rate of overseas workers from the Philippines, it may be common for households to match with other households with different work schedules.

6.2 Estimating the Marginal Cost of Sharing

To measure the marginal costs of sharing water, this strategy examines how much demand changes when households use their own connection versus when they use from a neighbor's connection. In order to observe the same connection using from both sources, I focus on sudden pipe leaks that often force households to disconnect and use from their neighbors. Specifically, I compare a connection's usage before disconnecting (due to a leak) to the change in usage among neighboring connections after disconnection, identifying the marginal hassle cost from the extent to which a connection may not fully offset its usage by sharing with neighboring connections.

Leveraging substitution due to pipe leaks requires identifying a sample of leaking connections, the timing of leaks/disconnection, and usage for neighboring connections. Over 36 months, the water provider received 106,171 consumer complaints of excessive billing — 34,927 of which can be precisely identified as underground pipe leaks along the service pipe through a text analysis of call transcripts. These complaints do not appear to be spatially correlated with only a handful of leaks occurring among direct neighbors. This approach considers any call identified as an excessive billing complaint as a possible leak. Instances where leaks in main water pipes reduce access for entire neighborhoods at a time are coded separately and excluded from the sample. Leaks are limited to cases with at least 3 months of recorded usage before the leak in order to calculate baseline water demand for these connections. Disconnection is defined as having 0 or missing consumption values for at least 90% of the year following the leak date. Neighboring connections that connect after the leak are dropped since these connections lack the pre-period observations needed to estimate their fixed preferences for water. The estimation includes observations within a year and a half of the leak event to avoid bias from any possible reconnections as well as changes in the composition of neighboring households over time. Since many of these households are not included in the connection survey, demographic characteristics as well as sharing behavior are imputed using averages from small areas surrounding these connections. This imputation introduces noise in the estimation of price sensitivities for these households.

To determine which neighbors to include, Appendix A.X ranks nearby households according to their distance from the leak and examines how the responsiveness of neighbors' consumption levels after a leak dissipates with distance. Results indicate that substitution is concentrated among the nearest four neighbors. Given these findings, the empirical exercise defines neighbors as the four closest connections.

Given the set of leaking households, $J = 4$, as well as their preferences, the final input to the likelihood function in equation (11) is the probability that leaking households choose to use from vendors. This probability is calibrated directly from a hypothetical question in the mobile survey, which finds that 70% of households would use a vendor in response to a sudden disconnection for over three months. Directly including this parameter in the estimation requires the assumption that this hypothetical response provides an accurate measure of true behavior and is generalizable across the full population of households.

To capture heterogeneity in the hassle cost, H is parameterized as a linear function of two summary characteristics of local census wards. First, the share of single houses or apartment buildings in a local area may increase the per-unit cost of sharing since the buildings often require longer informal pipe networks with possibly more maintenance costs as compared with duplexes. Second, the average distance between water meters in a local area provides another useful proxy for the average distance that informal pipe networks may need to extend between households. Importantly, this distance measure does not account for the physical distances between sharing partners or unconnected households. These omissions mean that this average distance measure is likely to overestimate the true distance between all neighbors. Furthermore, areas with smaller physical distances are likely to increase incentives to share as well as the observed distance between meters, which may bias the effect of distance

Table 6: Hassle Cost Estimates

	Sample Mean	Estimate	Standard Error
Intercept	1	-50.98	32.79
% Single Home/Apartment	0.86	39.67	37.99
Neighbor Distance	1.87	17.59	6.24

Estimates include 548 leaking connections and 2,151 neighboring connections, totaling 20,299 connection-months. % single/home and neighbor distance are averages within wards. Neighbor distance is in meters. Standard errors are bootstrapped at the level of the leaking connection (and neighbors).

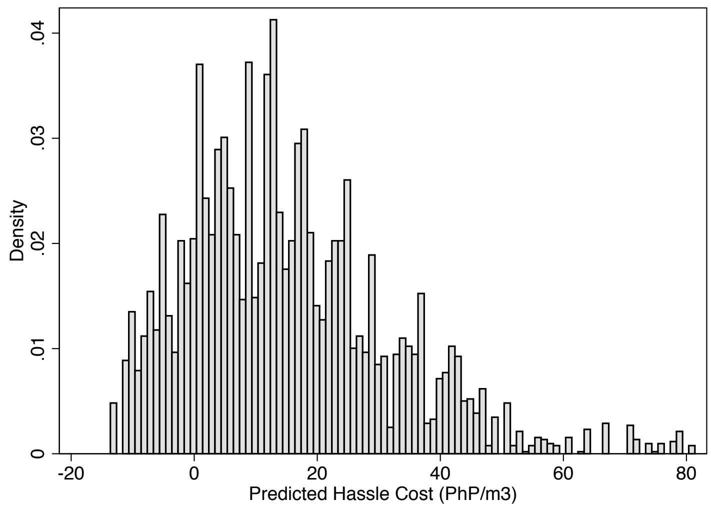
on marginal sharing costs towards zero. Table 16 in Appendix A.VII includes descriptive statistics for these two measures. One additional concern would be that households select into sharing relationships conditional on some unobserved heterogeneity in sharing costs, which is not modeled here. Since households included in this quasi-experiment have selected into purchasing connections at baseline, they may face higher hassle costs compared to households that select into sharing. Therefore, this exercise may provide an upper-bound on hassle costs for the population.

The following additional assumptions are required to ensure that this empirical setting accurately captures the model described in Section 5.2.1. These assumptions are discussed in more detail and tested empirically in Appendix A.XI.

1. *Connections that experience a leak are representative of the population of connections.*
2. *The decision to disconnect conditional on experiencing a leak is uncorrelated with water demand.*
3. *The timing of a leak is uncorrelated with any underlying changes in demand.*
4. *The leaking household's choice of a sharing partner is uncorrelated with that partner's water demand.*
5. *Leaks are uncorrelated with changes in neighborhood demand.*
6. *A leak for one household is uncorrelated with the probability of a neighboring household experiencing a leak.*

Table 6 shows the coefficient estimates for the hassle cost estimation. The hassle cost rises quickly with the share of households living in single homes or apartment buildings (instead of duplexes). Similarly, increasing the average distance between households by one meter raises the predicted hassle cost by 17.59 PhP/m3. For the average household, the hassle cost is estimated to be 15.93 PhP per cubic meter of usage, which is approximately 63.7% of the

Figure 2: Predicted Hassle Cost Distribution



Excludes top and bottom 1% outliers

average tariff.

Figure 2 plots a histogram of predicted hassle costs across the population of households, finding substantial heterogeneity with a long tail of prohibitively expensive hassle costs. Predicted hassle costs also have a correlation of 0.28 with household fixed water preferences.[28] This correlation has substantive implications for the impacts of counterfactual pricing policies; for example, high hassle costs may drive large users to sort away from sharing while small users can happily share water at little additional cost.

With only 548 connections that experience leaks, the estimation recovers wide standard errors for all three parameters.[29] With this in mind, Section 7.2.3 repeats counterfactual policy simulations for varying levels of the hassle cost. Taken together, these results suggest that the time-costs or other physical challenges associated with informally transporting water may play an important role in the extent to which households choose to use water from a neighbor's connection.

[28] Section 6.3.1 later discusses how connection-level fixed preferences γ_C are disaggregated into household-level components, γ_i.

[29] Since Steps 2 and 3 of the estimation rely heavily on estimated preference terms, γ, from Step 1, Appendix A.XIV tests for bias due to noisy estimation of these incidental parameters and finds little evidence of significant bias.

6.3 Fixed Cost Estimation

By matching observed water source choices with predicted source choices (given preference parameters from Section 6.1 and hassle cost estimates from Section 6.2), this step recovers the fixed costs of using a connection as well as prices associated with fetching from a vendor. These fixed costs capture both monthly costs — service fees, plumbing maintenance, etc. — as well as one-time costs put into monthly terms — connection fees, permitting with local governments, etc.

The fixed cost estimation sample includes the random sample of 5,000 connections from the preference estimation.[30] This sample is merged with a subsample of households using water from small-scale vendors drawn from the Census of Population and Housing 2010 as described in Section 3. Neighborhoods of potential sharing partners are limited to groups of three households, which is consistent with the vast majority of households forming sharing relationships of three households or less. Within a three person neighborhood, households can share with a maximum of two other households. Assuming a limited neighborhood size and restricting households to only share within their neighborhoods both work to reduce sharing opportunities for households. Since high fixed costs for new connections incentivize sharing, this approach may underestimate the true fixed costs faced by households.

6.3.1 Disaggregating Connection Preferences to Household Preferences

Simulating water sources choices requires recovering household preferences from the connection preferences estimated in Section 6.1. Household price sensitivity, α_i, can be inferred as a linear function of the vector of demographics specific to each household, Z_i, and estimates of α_C. Since fixed preferences $\hat{\gamma_C}$ are estimated non-parametrically at the connection-level, I specify a parametric method of splitting $\hat{\gamma_C}$ into γ_i for all households using demographics, Z_i:

$$\hat{\gamma_C} = \sum_{i=1}^{S}(\gamma_i - \alpha_i H_i)$$

$$\hat{\gamma_C} + \sum_{i=1}^{S}\alpha_i H_i = \sum_{i=1}^{S}(\lambda_S Z_i + \mu_i)$$

[30] This sample excludes connections used to estimate marginal sharing costs since the majority of these connections use imputed demographics.

Z_i captures demographic characteristics specific to each household. λ_S are coefficients measuring the effect of these demographics on fixed preferences γ_C. These coefficients are allowed to differ across connections serving different numbers of households as well as whether household i is the owner of the connection. μ_i captures any residual variation that is unexplained by demographics. In this case, Z_i only consists of household size since it is the single demographic measure reported separately for connection owners and other users. A linear regression recovers λ_S, which provides predicted levels of γ_i for each household.

Finally, any residual variation in preferences μ_i needs to be allocated across owning and sharing households. Instead of either allocating μ_i entirely to the owner or splitting μ_i evenly across households, this approach allocates $M \times \mu_i$ to the owner and $\frac{(1-M)}{S-1} \times \mu_i$ to neighboring users where M is the ratio of average fixed preferences for single users to shared connections. This approach nests the assumption that connection owners are more likely to resemble individually connected households than their sharing partners. Another limitation is that the data only provide the total number of people using a connection for shared connections. Therefore, in cases where three households use a connection, people are assumed to allocated evenly between the two sharing households.

6.3.2 Imputing Preferences for Households using from Water Vendors

Given the full set of household-level demand parameters for connected households, preferences are then estimated for households using from vendors. Projected α_i can be computed from demographic measures Z_i multiplied by the estimated price sensitivities, α. Since γ_i is estimated non-parametrically for connected households, fixed preferences need to be imputed for unconnected households. Fixed preferences for connected households are regressed on census ward indicators, the full set of demographic indicators used in the demand estimation (Section 6.1), and their full set of interactions. After recovering coefficients for the sample of connected households, predicted fixed preferences can be computed for households using water from vendors. This method requires the assumption any that unobserved variation in preferences is uncorrelated between connected households and households using from vendors. This assumption may not hold if unobservable aspects of preferences drive some households to use from vendors. Since vendors offer high marginal prices possibly attracting households with unobservably low demand, this form of selection may lead this method to overestimate demand among households using from vendors. Disaggregating connection-level preferences as well as including households using from vendors combine to expand the sample from 5,000 connections to 6,363 households.

Table 7: Fixed Cost Estimates

	Estimate	Standard Error
Connection Fixed Cost	352.96	59.13
Vendor Fixed Cost	129.51	27.22
Vendor Price	32.85	3.41
Vendor Price Variance	0.47	3.07

Fixed costs are in PhP/month and prices are in PhP/m3. 6,363 households across 599 wards are included in simulated method of moment estimation using source choices. Standard errors are bootstrapped at the connection level

6.3.3 Simulation and Estimation

With the full set of household preferences, the simulation routine first groups households into random sets of three neighbors within each of the 599 wards. The routine then orders households in the sequential game according to their utilities from using an individual connection. This assumption captures the intuition that households who would have connected in the absence of any sharing are likely to be the first to take-up a new connection. Random draws of the preference and consumption shocks, η and ϵ, are assigned to each household so that households can compute expected utilities over 5 months for all possible water source choices. These shocks are sampled from a multivariate normal distribution taking into account the estimated correlation between sharing partners presented in Table 5. One limitation of the model is that the linear demand structure with normal error terms also predicts occasional negative consumption values. Those values are rounded to zero in calculating household utility.

Households calculate payoffs following Section 5.2 given all potential choice sets. Appendix A.XII maps the full game tree with three players that is simulated in the estimation. This sequential game is repeated 10 times to construct a set of draws to compute simulated moments. The moments used in the estimation include the shares of households choosing different sources — using an individual connection, using water from a neighbor, and using from a vendor — as well as the correlation between each of these source choices and the fixed preference parameters for each household, γ_i. The estimation is designed to place heavier weight on matching the source-choice moments over the correlation moments because these correlations rely on possibly noisy measures of the preference parameters both due to incidental parameter issues in estimation as well as imputation for households using vended water. Table 7 presents the estimation results, finding a total estimated monthly fixed cost of 353 PhP, which is around 50% higher than the 225 PhP monthly fixed fee paid to the provider. This finding suggests that households face sizable barriers to connecting to the water network such as land tenure insecurity or administrative costs. This total fixed fee is also comparable to the total monthly water bill (tariff × volume) paid to the provider each month in Table 1, suggesting that fixed costs compose almost half of the total costs of using piped water.

The estimates also reveal a fixed cost for vendors equivalent to 130 PhP per month. This cost likely captures fixed quality differences between vended water and other water sources. The mean price for vended water of 33 PhP per cubic meter is close to the highest tariff prices in Figure 1. These estimates include a modest variance term for the vendor price, which is consistent with a competitive market for alternative water sources through deep wells, tanker trucks, and water refilling stations.

Since consumer surplus from using piped water is defined relative to using from a vendor, the magnitudes of the welfare calculations may be sensitive to the parameter estimates of the vendor option. As a simple out-of-sample test, Appendix A.XV finds that the estimate of p_v in Table 7 matches closely with the implied p_v from data on total water expenditures for vendor households using additional census data from part of Manila. This exercise provides additional confidence that the vendor price estimates are within a reasonable range. Also, Appendix A.XVI compares the predicted moments to the observe moments used in the estimation. While source moments are matched exactly by design, the estimation is able to roughly match the size and direction of the correlations of fixed preferences with the proportion connected individually and the proportion using from a neighbor.

7 Counterfactual Policies

To reconstruct the economic problem faced by the regulator and investigate how alternative pricing policies might impact welfare, consider the government regulator's maximization problem given by equation (12). In this problem, the regulator chooses a fixed fee, Fee, and a price schedule, P, to maximize consumer surplus while covering the costs of production (holding profits equal to zero). Consumer welfare is defined as the expected sum of all households' utilities subject to welfare weights, δ_i, for different subpopulations, which capture any preferences the regulator may have towards redistribution. The counterfactual exercises consider two sets of these preferences: (1) the regulator weights all households equally and (2) the regulator only gives positive weight to households from the bottom 50th percentile of water demand, summarized by the fixed water preference γ for each household. Giving preference to low-demand households intends to capture dual policy goals of ensuring access as well as minimum monthly usage levels. Since low-demand households are also often low-income households, this approach also provides a proxy for standard preferences toward redistribution.[31] Similarly, this approach may capture efficiency gains if connecting low-users produces positive health externalities (Galiani et al. [2005], Gamper-Rabindran et al. [2010]). For ease of exposition, the utility function below nests the water source choices of each household.

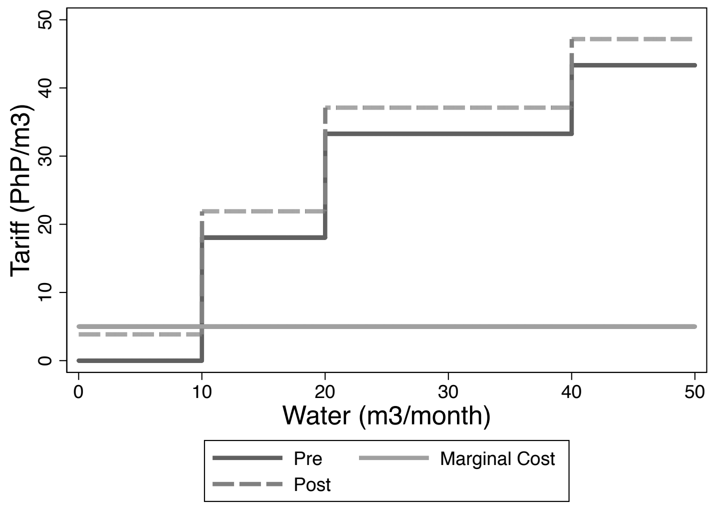
$$
\begin{aligned}
max_{Fee,p_L,p_H} \quad & E\Big[\sum_i \delta_i U_i(Fee, p_L, p_H)\Big] \\
s.t. \quad & \text{Total Connections} \times \big[Fee - \text{Installation/Service Costs}\big] + \\
& \text{Total Water} \times \big[\text{Revenue per Unit}(p_L, p_H) - \text{Marginal Cost}\big] \\
& \qquad\qquad\qquad\qquad \geq \text{Capital Costs}
\end{aligned} \tag{12}
$$

The budget constraint ensures that the fixed monthly fee, Fee, paid to the company for each connection as well as the revenue earned for each unit of water consumption (given the non-linear pricing scheme) outweighs the costs of servicing each connection, the marginal costs of providing a cubic meter of water, as well as any capital costs associated with the pipe-network, pumping stations, and dam maintenance. The water provider calculates that their marginal costs are equal to 5 PhP per m3 and that the fixed fee exactly covers any installation/service costs. To determine the capital costs, this approach subtracts the total marginal and services costs from the total revenue earned by the company, assigning the remainder to capital costs. Capital costs include maintaining the pipes, pumping stations, and billing infrastructure. This calculation provides a close approximation to the capital-expenditures regulatory structure implemented in this context, which exactly compensates the water provider for its costs through increases in the tariff.[32] This regulatory problem forms the basis for evaluating the following counterfactual policies. All counterfactual policies evaluated below hold firm profits equal to zero.

[31] Using income directly is less informative due to noise in the imputation process, especially for households that are sharing with their neighbors.

[32] This approach abstracts away from any principal agent problems between the company and the regulator by assuming that the regulator has perfect information over the firm's capital investments and operating costs (Laffont and Tirole [1993]).

Figure 3: Revenue-Neutral Tariffs
Pre/Post Connection Fee Discount



7.1 Compensated Fixed Fee Discount

First consider a counterfactual policy where the regulator discounts the fixed fee while increasing the tariff to remain revenue neutral. By discounting the fixed fee by 45 PhP, this approach mimics a policy implemented by the water provider; however, while the provider's discount decreased overall revenue, this approach offsets the discount with higher tariffs.[33] This counterfactual also echoes recent recommendations from policymakers and researchers including the World Bank who have advocated for lower connection fees as a way of extending access to piped water among the poor (WSUP [2013], Komives [2005], Jimenez-Redal et al. [2014], Kayaga and Franceys [2007], McIntosh [2003]). Figure 3 provides the current tariff and the compensated tariff, which is shifted upwards by 3.85 PhP/m3 for each tariff block to maintain revenue neutrality.

Table 8 summarizes the impacts of this change on source choices and consumer welfare, which is measured in terms of utility relative to using from a water vendor. While the increase in the tariff drives large users to disconnect and substitute toward vendors, the decrease in the fixed price allows some small users to jointly purchase shared connections, resulting in no change in the proportion of households using from a neighbor. The second row indicates substitution toward water vendors of around 2.31 percentage points in response to the higher tariff. These effects combine to reduce average consumer surplus by 28.7 PhP per month, which is over

[33] Appendix A.XIII analyzes the discount implemented by the provider as an out-of-sample test for the model, matching substitution patterns between water sources induced by the policy change.

Table 8: Fixed Fee Discount Counterfactual

	Pre-Discount	Post-Discount
Fixed Fee (PhP)	225	180
Use from Vendor	0.044	0.067
Use from Neighbor	0.243	0.243
Surplus (PhP)	143.0	114.3
Surplus: Low Users (PhP)	76.4	69.5

Low Users include the bottom 50% of water users.
Surplus and fixed fee are in PhP/month.

half of the initial fixed fee discount of 45 PhP. This decline translates into a 0.31 percent loss of monthly household income. Households in the bottom 50th percentile of water users also experience lower consumer surplus although the magnitude is much less, amounting to 0.07 percent of monthly household income. Taken together, these results suggest that connection fee discount policies in this context can have unintended distributional consequences by reducing both access to piped water and even welfare for the intended beneficiaries.

7.2 Characterizing Optimal Tariff Schedules

This section considers the welfare impacts of six tariff structures in order to inform regulators facing a variety of institutional settings and redistributional priorities: while some regulators may be restricted to imposing simple two-part tariffs, others may allow for complex, non-linear price schedules like the current tariff in Manila (Hoque and Wichelns [2013]). While encouraging efficient use, many developing countries also include distributional motivations for their tariff structures such as ensuring that most households reach a minimum level of usage (Szabó [2015]).

As a benchmark, this exercise examines a first-best tariff where (1) the marginal price equals the marginal cost, (2) every household pays a fixed fee equal to the fixed costs of installing and maintaining a connection, and (3) the regulator is able to perfectly price discriminate according to household demand, extracting additional transfers from high-demand households to exactly fund the fixed capital costs. According to standard industrial organization theory, this tariff is designed to perfectly maximize welfare in a world without sharing (Auerbach and Pellechio [1978], Oi [1971]). I then compare how the current tariff in Manila performs relative to this first-best tariff structure.

The next tariff imposes restrictions on the regulator by only allowing for a simple two-part tariff — Part (1) a monthly fixed fee and; Part (2) a marginal price. First, I identify the two-part tariff that maximizes overall consumer surplus. Second, I define a social two-part tariff that maximizes surplus only for low users, capturing regulator preferences toward redistribution and ensuring basic levels of access. Low users are defined as households with fixed preferences for water usage, γ, below the average level. Since households exhibit little heterogeneity in their price sensitivities, this approach provides a simple way of summarizing water demand across the population.

Finally, this exercise calculates optimal three-part tariffs — Part (1) a monthly fixed fee; Part (2) a marginal price for any usage below 20 m3 per month, and; Part (3) a marginal price for any usage above 20 m3 per month. This structure more closely mirrors the current tariff implemented in Manila. Optimal three-part tariffs are computed both to maximize surplus overall and to maximize surplus only for low-demand households.[34]

Does the presence of informal water sharing affect optimal prices in this context? To test this hypothesis, I compare optimal tariffs in a counterfactual setting where sharing is prohibitively expensive (Section 7.2.1) to optimal tariffs in the current setting in Manila (Section 7.2.2).

7.2.1 Optimal Tariffs Without Sharing

Identifying optimal tariffs without sharing requires an additional assumption about the costs to the water provider in this counterfactual setting. When households are unable to share, the provider raises less revenue at current prices. To account for this lost revenue, capital costs are recalculated to match the revenue raised from implementing the current tariff when sharing is not possible. This procedure allows for direct welfare comparisons between the current tariff and counterfactual tariffs; however, the magnitudes of these results are not directly comparable to counterfactuals with sharing explored later in Section 7.2.2.

Figure 4 provides the optimal tariffs without sharing, and Table 9 includes the corresponding optimal monthly fixed fees as well as the impacts of these tariff structures on the share of households using vendors and using neighboring connections. The last two rows include consumer surplus measures for the average household as well as the bottom 50% of users. Since producer surplus is always held to be zero in these exercises, these consumer surplus measures fully capture welfare. For ease of comparison, Table 9 also includes the average tariff level in the second row.

[34] I calculate optimal tariffs for a 5% subset of the sample since non-linear budget constraints make this process computationally expensive. Preliminary tests suggest that the results are robust to this sampling technique.

Figure 4: Optimal Tariffs Without Sharing

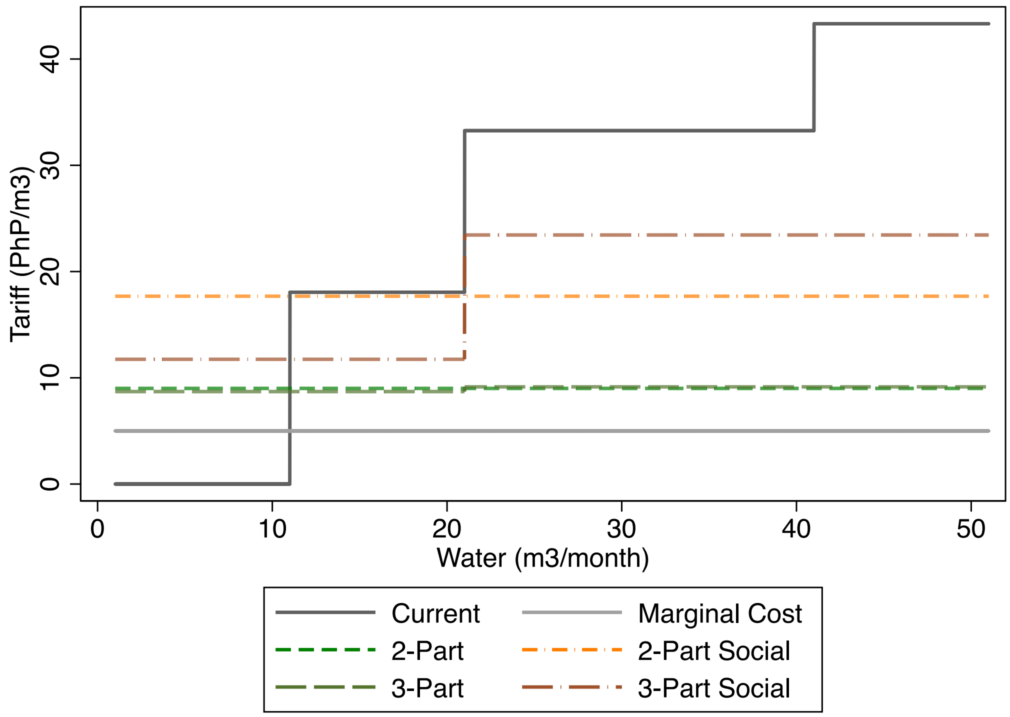


Table 9: Optimal Tariffs Without Sharing

	First-Best	Current	2-Part	2-Part Social	3-Part	3-Part Social
Fixed Fee (PhP)	225	225	290	31	293	108
Avg. Tariff (PhP/m3)	5.0	22.5	9.0	17.7	8.9	17.6
Use from Vendor	0.21	0.25	0.33	0.19	0.33	0.21
Use from Neighbor	0	0	0	0	0	0
Surplus (PhP)	324.6	124.1	310.5	265.6	310.6	239.8
Surplus: Low Users (PhP)		48.5	35.5	66.8	36.3	75.6
Usage (m3)	33.8	17.3	28.7	24.9	28.6	23.0

The first-best fee does not include individual transfers imposed by the regulator.
Social tariffs place all welfare weight on the bottom 50% of water users. All terms are monthly.

The first-best tariff schedule produces the highest level of consumer surplus reaching 324.6 PhP/month or 1.64% of household income. This tariff also produces the minimum share of households using vended water at 21%. With costless transfers, the regulator could theoretically reallocate all surplus to the lowest users providing an upper bound on redistribution through pricing at one and a half times the average surplus.

Relative to the first-best, implementing the current tariff without sharing reduces average surplus dramatically and increases the share of households using vended water by raising the average marginal price. The optimal two-part tariff (third column) features a higher fixed fee which affords a lower marginal price at all consumption levels. Compared to the current tariff, the optimal two-part tariff brings consumer surplus up to 96% of the first-best average surplus. These welfare gains are primarily driven by increased consumption for large users since surplus for low-users is cut in half and the share of households using from vendors increases by 8 percentage points.

The optimal social tariff (fourth column) maximizes surplus for households with below average demand. This social tariff improves welfare for low-users by reducing the fixed fee dramatically and almost doubling the marginal price. Low fixed fees incentivize small users to connect to the network, reducing the share of households using from vendors by around 14 percentage points. Average surplus declines since high marginal prices reduce consumption for large users; however, surplus for low users more than doubles. This result points to the attractiveness of using non-linear pricing to achieve normative goals of redistribution when shared connections are infeasible.

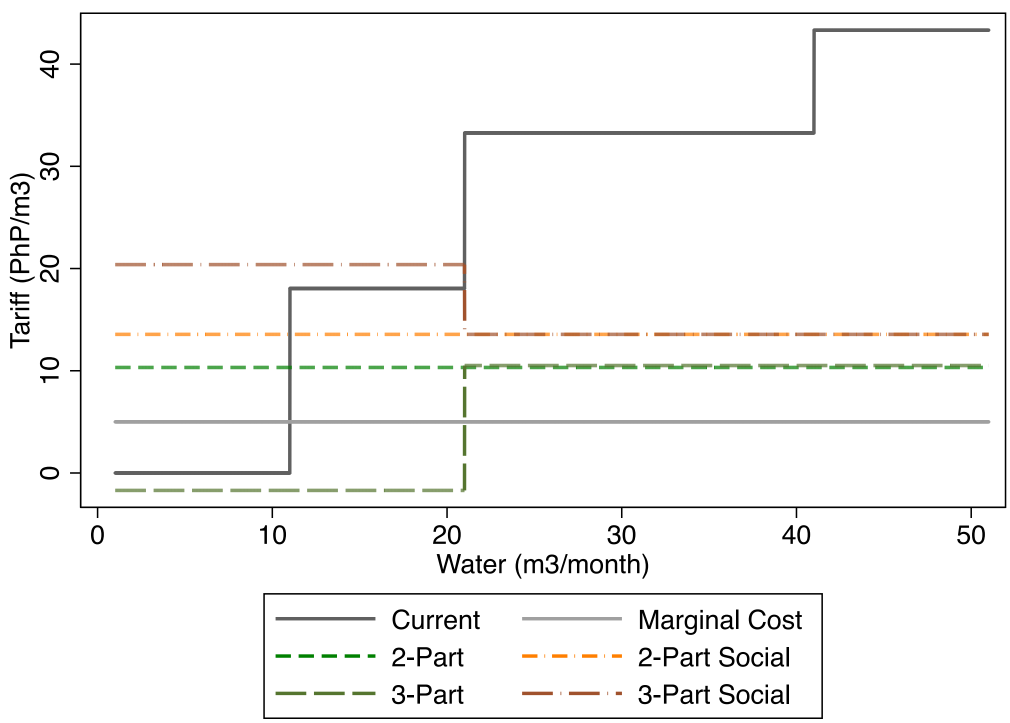
By adding an extra segment to the tariff schedule, the optimal three-part tariff (fifth column) tests the extent to which more complex pricing schedules can improve efficiency. Figure 4 graphs this three-part tariff relative to the current tariff. The shape of the optimal three-part tariff resembles the current tariff structure with a low first segment followed by a higher second segment (although the optimal three-part tariff maximizes total surplus while the current tariff is likely motivated by distributional concerns). The intuition for this finding is that a low first price segment combined with a high fixed fee provides an additional mechanism for the regulator to price discriminate on the extensive margin; under this pricing regime, only relatively large users connect and their high usage levels allow the regulator to further lower prices (since the marginal price well exceeds marginal cost). Despite this attempt to improve welfare, surplus improvements are negligible relative to the optimal two-part tariff.

Compared to the social two-part tariff, the social three-part tariff (sixth column) uses a higher fixed fee to achieve a slightly lower marginal price for water usage up to 20 m3. The higher fixed fee drives another 2% of households to use from vendors; however, the lower marginal price benefits low-users who remain connected. Relative to the social two-part tariff, the social three-part tariff achieves a surplus gain of over 8 PhP/month for low-users while lowering average surplus by around 25 PhP/month.

These results provide empirical support for the widespread practice of using non-linear pricing schedules to achieve goals of redistribution (Hoque and Wichelns [2013]). When sharing is prohibitively expensive, non-linear pricing is able to over double surplus for the bottom half of water users (relative to optimal pricing that weights all users equally).

7.2.2 Optimal Tariffs With Sharing

Figure 5: Optimal Tariffs With Sharing



This section calculates optimal tariffs in the current environment in Manila (where households may choose to share). Figure 5 plots the marginal prices associated with each counterfactual tariff. Table 10 computes the impacts of these tariffs on source choices as well as welfare. Moving from the current tariff schedule (second column) to an optimal two-part tariff (third column), I find large welfare improvements on the order of 156% of consumer surplus or 1.1% of household income. These gains are mainly enjoyed by large users benefiting from a lower marginal price relative to the current tariff since average consumption nearly doubles. At the same time, flattening the tariff reduces the implicit "tax" on pooling consumption on a shared connection, allowing low-users to enjoy lower marginal prices from their neighbors' connections. These lower marginal prices compensate for the higher hassle and free-rider costs from sharing, boosting surplus for low-users by around 50% and reducing the share of households using from vendors from 4% to 1%. Low marginal prices combined with higher fixed fees both work to increase the share of households using from a neighbor's connection from 25% to 44%.

A social two-part tariff (fourth column) reallocates surplus to low-users by subsidizing fixed connection fees with a higher marginal price. Compared to an optimal tariff, this social tariff brings the percentage using from vendors to zero while also decreasing the share using from neighbors by 7 percentage points. These improvements in access come at a small cost to average consumption, but only increase surplus for low-users by a modest 4.4 PhP/month. By contrast, social tariffs almost doubled low-user surplus in a world without sharing (Section 7.2.1). These findings suggest that informal sharing networks may constrain the effectiveness of non-linear pricing in redistributing surplus.

An optimal three-part tariff with sharing (fifth column) follows the same pattern as in the case without sharing: increasing marginal prices select for high-usage households to connect, which provide greater revenue to cover the fixed costs. As before in Section 7.2.1, this tariff produces negligible changes in welfare. Figure 5 indicates an optimal three-part tariff with an almost zero first price segment followed by a much higher second price segment, which again broadly aligns with the current tariff structure. This tariff structure supports the "lifeline" tariff as a way of efficiently maximizing surplus (rather than achieving distributional goals). Column six (along with Figure 5) finds that a social three-part tariff subsidizes entry

Table 10: Optimal Tariffs With Sharing

	First-Best	Current	2-Part	2-Part Social	3-Part	3-Part Social
Fixed Fee (PhP/month)	225	225	259	108	490	-12
Avg. Tariff (PhP/m3)	5.0	22.5	10.3	13.6	4.4	17.0
Source Vendor	0.01	0.04	0.01	0	0.01	0
Source Neighbor	0.39	0.25	0.44	0.37	0.44	0.35
Surplus	372.1	142.5	365.4	357.9	367.2	349.2
Surplus: Low Users		78.5	119.3	123.7	121.4	123.8
Consumption (m3/month)	35.7	17.9	31.9	30.6	32.0	30.2

The first-best fee does not include individual transfers imposed by the regulator.
Social tariffs place all welfare weight on the bottom 50% of water users.
Surplus is measured in terms of PhP/month.

with a negative fixed price, which is offset by a high marginal price for usage up to 20 m3. This tariff reduces average surplus by over 8 PhP/month to produce only 0.1 PhP/month of additional surplus for low-users. This result underscores the extent to which sharing networks constrain the ability of cities to achieve redistribution through complex, non-linear pricing strategies.

7.2.3 Optimal Two-Part Tariffs and Marginal Sharing Costs

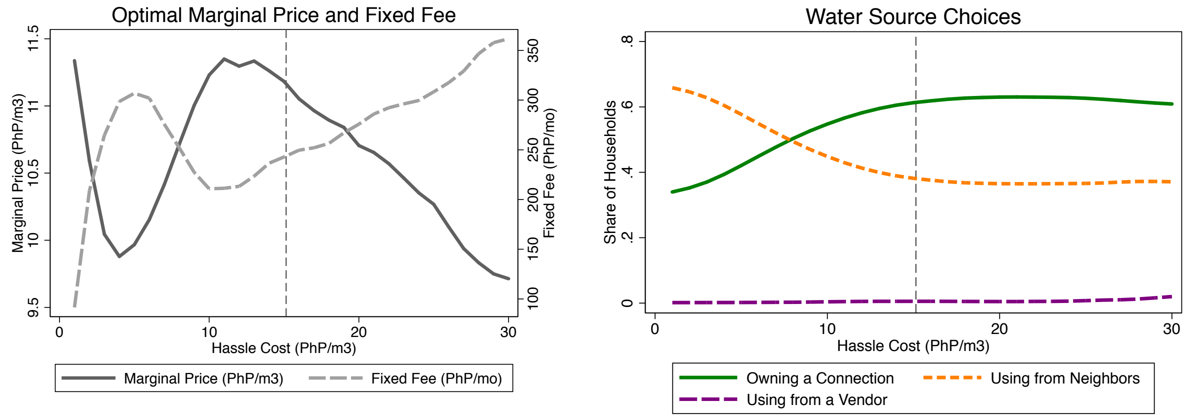
To investigate mechanisms driving these optimal pricing schedules, this section considers how the optimal two-part tariff varies continuously with counterfactual levels of the hassle cost, H. I follow the same routine as in Section 7.2.1 by (1) calculating total revenue raised under the current tariff (in Figure 1) given the counterfactual hassle cost, (2) recalculating the implied capital costs from total revenue under this hassle cost, and (3) determining the optimal two-part tariff maximizing total consumer surplus while just covering costs.[35]

Figure 6 plots the optimal marginal price, fixed fee, and share of the population choosing each water source as a function of the hassle cost. The vertical gray lines show the average estimated hassle cost of 15.93 from Section 6.2. When hassle costs are very low, the regulator chooses a relatively high marginal price and low fixed fee. This equilibrium maximizes sharing so that one third of households purchase connections and each share their connection with two neighbors. The high marginal price serves as an way to mitigate overconsumption from free-riding on these shared connections: when three households share a tap, they respond to one third of the marginal price through spitting the bill. Therefore, setting a price around 11.5 PhP/m^3 ensures that households actually respond closer to the marginal cost of 5 PhP/m^3. This high marginal price also raises greater revenue allowing the regulator to lower the fixed fee. As the hassle cost increases to 5 PhP/m^3, the optimal marginal price decreases as more households substitute away from sharing arrangements. This substitution reduces the extent of free-rider distortions, driving the marginal price closer to the marginal cost.

As the hassle cost increases from 5 PhP/m^3 to around 13 PhP/m^3, it becomes optimal for the regulator to increase the marginal price as a way of taxing shared connections and subsidizing individual connections. In this range, households are not able to internalize the extent to which costly sharing relationships reduce their own contributions to the capital costs. Therefore, a high marginal price serves as an additional disincentive for sharing.

[35] As the hassle cost rises, the current tariff raises less revenue, lowering the capital costs. The patterns below are robust to instead holding the capital costs fixed while increasing the hassle costs.

Figure 6: Two-Part Tariffs and Sharing Hassle Costs



The vertical gray lines show the estimated hassle cost of 15.14 PhP/m^3 from Section 6.2.

Above 13 PhP/m^3, the hassle cost becomes large enough to disincentivize sharing on the margin, allowing the regulator to begin lowering the marginal price again. Yet, at these high levels of hassle cost, some households begin substituting toward vended water, providing a new incentive for subsidizing access by keeping prices above the marginal cost (consistent with theoretical work by Auerbach and Pellechio [1978]). The slow decline in marginal price over this range helps to offset the increasing hassle cost, reducing the rate of substitution away from shared connections.

This exercise demonstrates how the presence of water sharing can have additional implications for optimal water pricing. Free-rider behavior among sharing households creates an incentive for the regulator to maintain high marginal prices, and large hassle costs similarly motivate high marginal prices as an efficient tax on shared connections.

8 Conclusion

This paper demonstrates that pricing policies intended to increase water access especially among the poor can have unintended consequences when low-income households can simply access the piped water network through their neighbors. The counterfactual exercises find that a simple two-part tariff structure outperforms the increasing block tariff currently implemented in Manila in terms of overall access to water, total welfare, and even welfare among the bottom half of water users. In this counterfactual, high fixed costs of accessing an individual connection allow for sharing networks to play an important role in extending water access despite sizable hassle costs. I also find that complex non-linear price schedules are largely ineffective in terms of achieving both efficiency and equity goals when sharing networks are readily available.

Taken together, these results suggest that households may act as efficient subcontractors in redistributing water to their neighbors. Instead of sending trained engineers to install and maintain connections in dense neighborhoods, providers can instead rely on informal networks to extend water distribution. From this perspective, cities may benefit from considering how prices not only might encourage access and redistribute consumer surplus, but also may affect the endogenous formation of water sharing networks. For example, price discriminating according to neighborhood income instead of by monthly usage levels may achieve similar goals of redistribution without distorting sharing networks. While this policy may not be appropriate in Manila since neighborhoods are often very heterogeneous in income, this approach may work well in other contexts where neighborhoods are more segregated, such as in Colombia where electricity rates vary according to the social category of each neighborhood (McRae [2014]).

Since sharing networks constrain the menu of pricing policies available to policymakers, cities may benefit from considering a broader range of policies to improve welfare from water use. For example, cities may be able to lower the frictions associated with these sharing networks by providing technical resources for managing plumbing networks or distributing extra water meters to assist households in accurately splitting the bill. The empirical results also confirm previous findings that substantial barriers exist for households to connect to the network (Devoto et al. [2012]). Cities may be able to similarly achieve substantial welfare gains by streamlining administrative requirements for new connections and by allowing households to pay connection fees in smaller increments.

Appendices

A.I Connection Fee

This section constructs a total fee paid to the water provider each month as the sum of the monthly service fee, 150 PhP/month on average, and the connection fee of 6,000 PhP, amortized over the duration of the water connection. To amortize the connection fee without information on full durations of residence for all households, I assume that households move to Manila, use a water connection for 80 months, then leave the city. This assumption is roughly consistent with the hazard rate of disconnection for water connections. Dividing the connection fee by 80 months and adding it to the monthly service fee results in a total monthly fee of 225 PhP. Any credit constraints, discount rates, or other barriers to accessing a water connection are captured in monthly terms in the fixed cost estimates in Section 6.3.

A.II Assessing the Sampling of the Connection Survey

This section assesses the extent to which the connection survey constitutes a representative sample of connected households. The documentation for the connection survey indicates that sampling was stratified based on geographic areas, assigning a target number of households owning connection to be survey according to the Census population for each area. Households owning connections were randomly interviewed until these targets were reached. While ensuring a random sample of connections within areas, this method may oversample areas with few connections. To test this hypothesis, the following equation predicts the share of connections surveyed in each Meter Reading Unit (MRU) — the smallest geographical unit used by the water provider including a couple hundred water connections on average — according to demographics of households owning connections, $Z_{MRU,i}$, in each of these MRUs:

$$ShareSurveyed_{MRU} = \beta Z_{MRU,i} + \epsilon_{MRU,i}$$

Table 11 provides the results for this regression. Although many coefficients are statistically significant consistent with large sample sizes for this estimation, the magnitudes of the effects are small in economic terms across most measures. For example, increasing the average number of households per connection by a standard deviation (0.65) results in a 0.0021 percentage point decrease in the probability of being surveyed; given a mean survey probability of 6 percentage points, this change represents only a 3.3 percent decrease. Dwelling types are the strongest predictor of coverage, indicating that surveyors may have oversampled apartments possibly due to easier availability to survey. Taken together, these results suggest that to the extent that non-random sampling may affect results, it will lead to an underestimation of the importance of sharing networks in Manila.

A.III Comparing Demographics in the Census to the Connection Survey

Census households are divided into three groups according to their water source for cooking, laundry, or bathing needs: the first group includes households that select "own use, faucet community water system" defined as own use, the second group includes "shared use, faucet community water system" defined as shared use, and the third group includes all other categories defined as vendor. I find a lower percentage of households sharing connections in Table 1 than in Table 12. One reason is that all households sharing a connection may identify as owning the connection while the water provider only registers one household as the owner. Since the empirical strategy combines data from the connection survey with the census, I examine the first two columns for comparability across samples and find that demographics appear broadly similar.

Comparing columns two and three, I find that shared-use households have lower imputed incomes, greater probability of low-skill employment, and slightly smaller households than own-use households. These trends mirror the findings from connection owners in Table 1. Vendor households captured in the final column of Table 12 closely resemble shared-use households across age, employment, and income measures. These households are especially likely to reside in single houses, which may lead to easier access to vendors as well as more difficult access to neighboring connections.

A.IV Details on Mobile Survey

The mobile survey was conducted using Pollfish— an organization that advertises the opportunity to complete a survey for the chance to win an Amazon gift certificate. Pollfish

Table 11: Predicting Share of Connections Surveyed with Connection Owner Demographics

VARIABLES	(1) Percent Surveyed
Mean Consumption (m3)	0.000286***
	(3.94e-05)
HHs per Connection	-0.00334***
	(0.000775)
HH Size	-0.000285
	(0.000190)
Age HoH	0.000138***
	(2.27e-05)
Total Empl.	4.04e-05
	(0.000326)
Low Skill Emp.	0.00158*
	(0.000909)
Apartment	0.00704***
	(0.00157)
Single House	-0.0114***
	(0.00117)
Constant	0.117***
	(0.00250)
Observations	45,907
R-squared	0.018
Mean Coverage	.06
Std. Dev. Coverage	.06
Cluster MRU	Yes

Robust standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$
2,947 Meter Reading Units (MRUs)

Table 12: Household Demographics across
Survey and Census Data by Water Source

	PAWS	Census		
		Own Use	Shared Use	Vendor
Age (Head of HH)	46.5	46.6	42.4	44.2
HH Size	5.14	4.22	4.14	4.22
Employed HH Members	1.57	1.63	1.44	1.50
Low Skill Emp. (Head of HH)	16%	29%	35%	36%
Inc. (USD/Mo. Imputed)	397	387	372	380
Apartment	21%	27%	30%	23%
Single House	58%	60%	57%	65%
Duplex	21%	12%	13%	12%
Households	56,944	280,321	53,184	20,841
Share of Pop.		79%	15%	6%

The data include connection survey data on households owning water connections as well as 2010 Census data on households. This sample includes overlapping wards between surveys (67%)

Table 13: Mobile and PAWS Water Source Choices

	(1)	
	Mobile	PAWS
Single User	0.62	0.54
Share with 1 HH	0.13	0.21
Share with 2 or more HHs	0.17	0.25
Obs.	594.00	63,997.00

The PAWS data have been adjusted to reflect that shared connections serve multiple households.

reaches potential respondents through ads on mobile apps and Facebook. The survey was fielded in 6 batches of 100 respondents in July, 2017, which allowed for the addition and modification of questions across survey rounds.

Since this survey is not designed to be representative of the population of households in Manila, the following tables compare descriptive statistics between the mobile survey and the connection survey (Public Assessment of Water Services (PAWS)). Table 13 compares the shares of households choosing each water source — individual, sharing with 1 other household, or sharing with 2 other households — across the two datasets, adjusting for the fact that the connection survey was conducted at the water connection-level. I find similar shares of households choosing these three water sources across these datasets with more households choosing to connect as single users in the Mobile survey data. Since the

Table 14: Mobile and PAWS Descriptives by Water Sharing Status

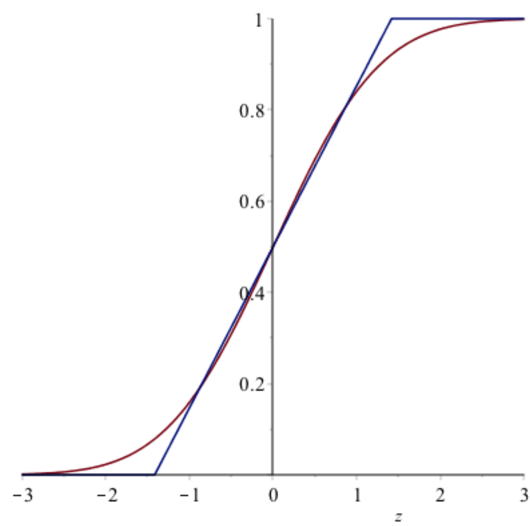
	Single User		Share: 1 HH		Share: 2 or more HHs	
	Mobile	PAWS	Mobile	PAWS	Mobile	PAWS
Age	31.71	47.16	29.33	46.52	29.15	45.44
Apartment/Duplex	0.41	0.39	0.44	0.49	0.53	0.63
Single House	0.59	0.61	0.56	0.51	0.47	0.37
HHsize: 1-2	0.09	0.06	0.08	0.11	0.12	0.14
HHsize: 3-5	0.48	0.52	0.43	0.62	0.40	0.60
HHsize: 6-10	0.36	0.39	0.36	0.27	0.32	0.25
HHsize: >10	0.07	0.02	0.12	0.01	0.16	0.01
Other Users: 1-6	.	.	0.56	0.91	0.40	0.18
Other Users: 7-10	.	.	0.25	0.08	0.27	0.42
Other Users: 11-15	.	.	0.11	0.01	0.16	0.24
Other Users: >15	.	.	0.08	0.00	0.17	0.16
Obs.	244.00	47,327.00	113.00	9,257.00	130.00	7,413.00

mobile survey was conducted over 8 years after the connection survey, this increase in single connections is consistent with a general substitution toward purchasing own connections (with increasing income and declining vendor opportunities).

Table 14 compares demographics within each of these categories of water source. The first row indicates that Pollfish respondents tend to be much younger than connection survey respondents, which is consistent with higher mobile phone use among younger people as well as the fact that the connection survey largely targeted heads of household while any member could respond to Pollfish. Dwelling types and household size appear largely similar for single users with differences increasing as the number of households using each water connection increases. Since the connection survey only captures demographics for water connection owners while the mobile survey covers all users, this trend may capture differences between these groups. These findings suggest that the mobile and connection surveys are broadly consistent in capturing the population of water users despite differences in survey methods and timing.

A.V Linear Approximation

This section now investigates the extent to which the linear approximation might impact coefficient estimates as well as welfare calculations. The following figure compares the linear approximation to the standard normal CDF function over the range $(-3, 3)$:



To ballpark the effects of this approximation, I simulate a dataset with two prices and one kink point as well as simulated parameters that are close to estimated values. Using this dataset, demand can be calculated in two ways: (1) according to the approximated function in equation (13) and (2) according to the exact demand function given by the numerical solution to the first order condition in equation (3). Table 15 provides the parameters used for simulation as well as the simulation results. The second and third columns provide estimates from the likelihood function in equation (8). The second column shows estimates when consumption is generated through the exact numerical solution to the first order conditions. In this case, the likelihood function would be misspecified since the likelihood relies on the linear approximation. The third column shows estimates for when consumption is generated using the linear approximation, ensuring that the likelihood function is properly specified in this case. The estimates rely on a small sample of households in order to speed performance. Sampling error can help explain why the estimates in column three do not exactly match the true parameters. The last column finds small differences between estimates in columns two and three, providing some evidence that the linear approximation may not introduce substantial bias in the estimation. Welfare differences between following approximated versus actual versus are negligible, totaling -0.055 PhP per month.

A.VI Alternative Sharing Contracts

Two alternative sharing contracts include (1) alternating bill payment each month and (2) charging a fixed price to sharing households. Alternating bill payments can be thought of as a fixed price contract where each household faces a payment of zero each month. To

Table 15: Simulation Estimates for Linear Approximation

Parameters	Value	Est: Exact DGP	Est: Approx. DGP	Diff.
σ_ϵ	5	5.18	5.01	3.27%
σ_η	10	9.76	9.92	-1.59%
α	1	0.49	0.49	-1.47%
γ	[10,60]	34.72 mean	34.87 mean	-0.62%
Simulated Data				
P_L	[0,20]			
P_H	[20,45]			
Kink Point	20			
Households	2000			
Months per HH	80			

The prices as well as the γ terms are sampled from a uniform distribution on the given interval. Approx. DGP stands for approximated data generating process, which generates consumption values using the linear approximation. Exact DGP calculates consumption using the true demand function.

model fixed price contract, consider a simple setting with two households and a linear budget constraint. Let household 1 pay the full bill while household 2 pays a fixed amount Z each month regardless of their consumption:.

The household optimization problems and solutions are as follows:

$$\begin{aligned}
&max_{x_1,w_1} \;\; E\big[U(x_1,w_1)\big] \\
&\quad \text{s.t.} \quad Y_1 = x_1 + p(W_1 + W_2) \\
&\qquad\qquad w_1 = \Gamma_1 - \alpha_1 p \\
&max_{x_2,w_2} \;\; E\big[U(x_2,w_2)\big] \\
&\quad \text{s.t.} \quad Y_2 = x_1 + Z \\
&\qquad\qquad w_2 = \Gamma_2
\end{aligned}$$

Joint consumption takes the following form:

$$(w_1 + w_2) = (\Gamma_1 + \Gamma_2) - \alpha_1 p$$

In the case where price sensitivities are equal between households, $\alpha_1 = \alpha_2 = \alpha$, joint consumption under a fixed price contract is empirically identical to joint consumption under even bill-splitting. Therefore, assuming bill-splitting when households use fixed price contracts may bias price sensitivity estimates when these price sensitivities are very different from each other. This bias becomes less of a concern if households alternate paying the bill because the joint estimate of α will capture some weighted average of α_1 and α_2. Fixed price contracts may lead to a different allocation of surplus from sharing contracts depending on the size of the fixed price Z. Yet, since the model assumes that households are able to split expected surplus from the sharing relationship through fixed monthly transfers in Section 5.2, the type of contract may have little effect on empirical or welfare estimates.

A.VII Descriptive Statistics for Estimation Sample

Table 16 provides descriptive statistics for the connection-level sample used in the demand estimation (Section 6.1) and simulated method of moments estimation (Section 6.3). Demographic variables are expressed in terms of dummy variables capturing various features that might drive demand for water. Ward-level attributes include features that determine hassle-costs of sharing with a neighbor. The sample includes a random sample of 5,000 connections combined with a sample of unconnected households in order to ensure a representative sample of the population of households.

A.VIII Price Variation

A.VIII.i Targeting the Non-linear Tariff

Figure 7 provides a histogram of monthly consumption choices across all water connections and months in the data alongside the tariff schedule (censored below 60 m3). Usage hovers around 20 m3 per month; by comparison, households in the United States use closer to 50 m3 per month. Intuitively, households may be less willing to consume an additional cubic meter on the margin if the price for that unit is much higher. The histogram provides some evidence of bunching at 10 and 20 m3, although the magnitudes are small. Many households are also observed to consume below 10 units where the marginal price is actually zero.

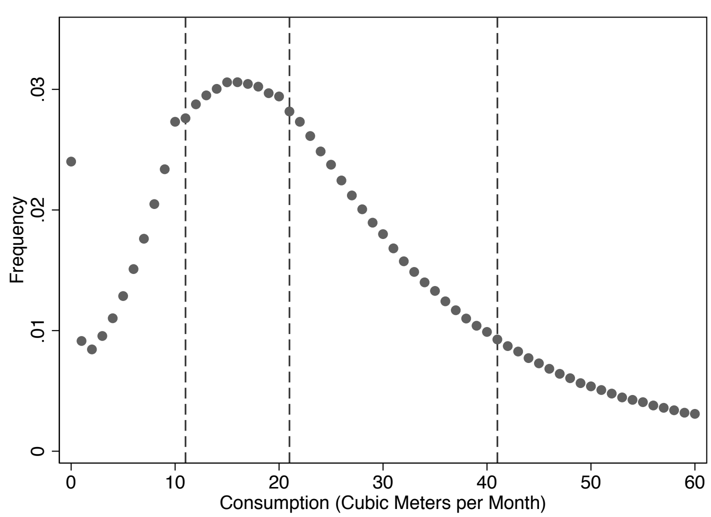
A.VIII.ii Time Series Variation in Tariffs

This figure indicates time series variation in prices as a result of the government regulator raising prices incrementally to ensure production costs are covered.

Table 16: Descriptives for Estimation Sample

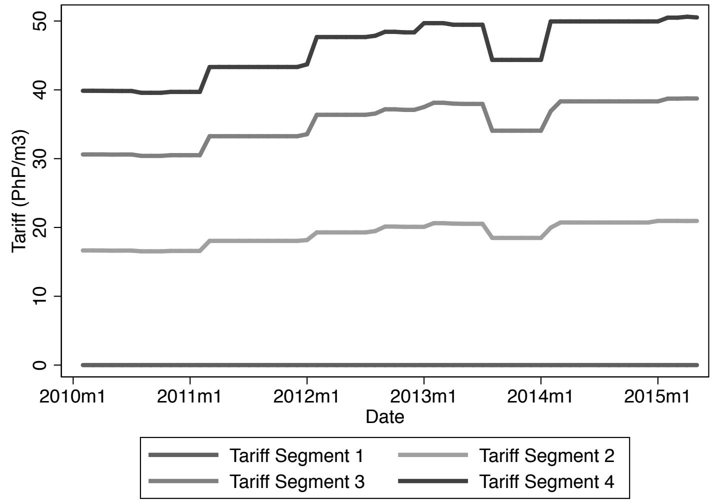
	Mean	Min	Max	Standard Deviation
Owner Demographics				
Water (m3/mo.)	27.07	0.00	100.00	17.67
Shared 2 HHs	0.14	0.00	1.00	0.34
Shared Over 3 HHs	0.09	0.00	1.00	0.29
4 or 5 HH members	0.41	0.00	1.00	0.49
Over 5 HH members	0.36	0.00	1.00	0.48
Apartment	0.23	0.00	1.00	0.42
Single House	0.59	0.00	1.00	0.49
Low-Skill Emp. (Head of HH)	0.17	0.00	1.00	0.37
Two or more Empl. HH members	0.45	0.00	1.00	0.50
HoH Age Between 36 and 52 Yrs	0.38	0.00	1.00	0.49
HoH Age above 52 Yrs	0.36	0.00	1.00	0.48
Ward-Level Characteristics				
Apartment or Single House	0.86	0.00	1.00	0.08
Dist. Between Neighbors (meters)	1.87	0.00	27.63	1.16
Sample Characteristics				
Total Connections		5,000		
Unconnected HHs		286		
Number of Wards		599		

Figure 7: Consumption Histogram



Consumption values are truncated at 60 to highlight bunching at kink points.

Figure 8: Regulated Tariff Changes



A.VIII.iii Within Connection Price Variation

This setting also provides quasi-experimental variation in the tariff over time at the connection-level, which serves as a useful setting to examine household demand response to price changes over time. Previous work estimating demand for public utilities has relied on time-series variation from regulated changes in the tariff (Szabó [2015], McRae [2014]). While the estimation strategy in this paper also includes this variation, regulated tariff changes are often relatively small, affect all households equally during the same month, and are anticipated well in advance. Through a quirk in the regulatory agreement, the water provider is able to upgrade households that are discovered engaging small businesses to a higher tariff class called the semi-business rate. The vast majority of these businesses are roadside stands, locally known as Sari-Sari stores. The connection survey covers 771 connections that are upgraded from residential to semi-business status used in the analysis.

Figure 9 compares the tariff structure between residential and semi-business rates. The tariffs mainly deviate between 10 and 20 cubic meters where the semi-business rate is about 30% higher.

Table 17 provides summary statistics and t-tests comparing residential connections to semi-business connections, and semi-business connections to connections that are upgraded from semi-business to residential. Semi-business connections use more water, while having smaller households sizes, greater low-skilled employment, and higher likelihoods of living in a single house. Despite using more water, those upgraded to semi-business rates are broadly similar to the population of semi-business connections.

Figure 9: Residential and Semi-Business Tariffs

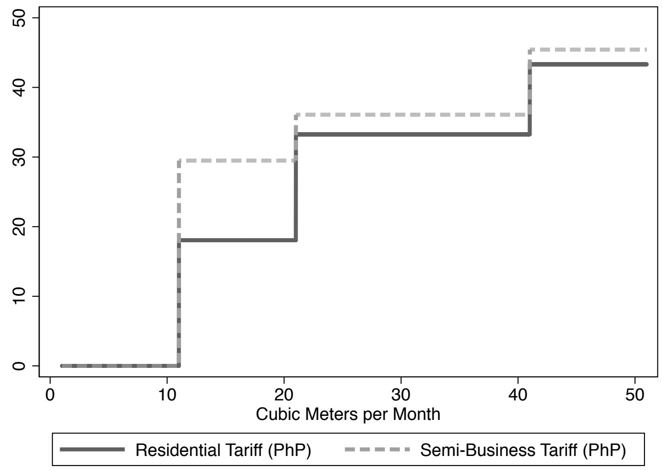
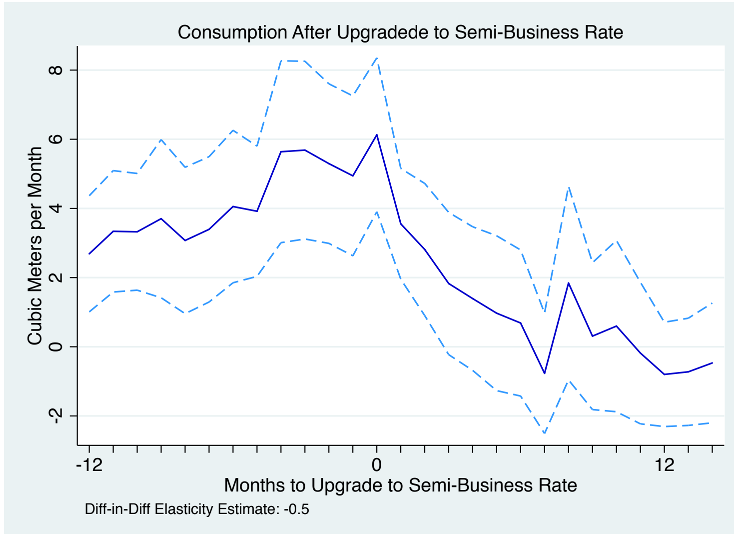


Table 17: Summary Statistics and Comparison of Means for Residential, Semi-Business, and Upgraded Connections

	Residential Residential	Semi-Bus Semi-Bus	Upgraded to Semi-Bus	T-Test: Semi-Bus to Res	T-Test: Semi-Bus to Semi-Upgrade
Cons. (avg.)	29.42	30.80	35.20	-1.37***	-5.78***
Age HoH	46.26	49.28	46.25	-3.01	0.01
HH Size	5.15	4.97	5.20	0.18	-0.05
HH Size Empl.	1.58	1.51	1.50	0.07*	0.08*
Low Skill Emp.	0.16	0.22	0.25	-0.06***	-0.09***
Apartment	0.23	0.14	0.18	0.09***	0.05***
Single House	0.56	0.68	0.62	-0.12***	-0.06***
Obs.	47578	5568	771	42010	46807

Data are from the connection survey.

Figure 10: Demand Response to Time Series Price Variation



The following equations estimate both event study and differences-in-differences strategies used to determine consumption response to changes in the rate class. For connections i in month t, these equations predict consumption as a function of individual, θ_i, and time, δ_t fixed effects. The event study in the first equation includes β_r dummy variables for months to the rate change while the differences-in-differences approach in the second equation captures the β effect on consumption of being after the rate-change, $Post_{i,t}$.

$$
\begin{aligned}
w_{i,t} &= \sum_{r=-12}^{12} \beta_r \mathbb{1}\{t = r\} + \theta_i + \delta_t + \epsilon_{t,i} \\
w_{i,t} &= \beta Post_{i,t} + \lambda_1 r_{i,t} + \lambda_2 Post_{i,t} \times r_{i,t} + \theta_i + \delta_t + \epsilon_{t,i} \\
r &: \text{ month to upgrade to semi-business rate}
\end{aligned}
$$

Figure 10 plots average consumption for these connections relative to the month that they are upgraded to the semi-business tariff level. Before the upgrade consumption trends slightly upward, before dropping immediately following the tariff increase. A simple differences-in-differences estimate calculates a sizable price elasticity of 0.5 from this change, indicating significant demand responsiveness to time-series variation in the tariff. This elasticity is smaller than the structural estimates in Section 6.1.

A.VIII.iv Characterizing the Optimal Demand Function

The optimal demand function takes a piecewise form depending on the proximity of different price segments.[36] The aggregate demand function is as follows (where L, M, H index low, medium, and high prices relative to each segment):

$$D_C^* = \begin{cases} W_C^{1,1,1} & \text{if} \quad \frac{\eta_C}{V^{1,1}} \leq \overline{\eta}_{1,L}^1 \\ W_C^{1,2,2} & \text{if} \quad \overline{\eta}_{1,L}^1 < \frac{\eta_C}{V^{1,2}} \leq \overline{\eta}_{2,L}^2 \\ W_C^{1,2,3} & \text{if} \quad \overline{\eta}_{2,L}^2 < \frac{\eta_C}{V^{1,3}} \leq \overline{\eta}_{1,U}^2 \\ W_C^{2,3,3} & \text{if} \quad \overline{\eta}_{1,U}^2 < \frac{\eta_C}{V^{2,3}} \leq \overline{\eta}_{2,U}^3 \\ W_C^{3,3,3} & \text{if} \quad \overline{\eta}_{2,U}^3 < \frac{\eta_C}{V^{3,3}} \leq \overline{\eta}_{3,L}^3 \\ W_C^{3,4,4} & \text{if} \quad \overline{\eta}_{3,L}^3 < \frac{\eta_C}{V^{3,4}} \leq \overline{\eta}_{3,U}^4 \\ W_C^{4,4,4} & \text{if} \quad \overline{\eta}_{3,U}^4 < \frac{\eta_C}{V^{4,4}} \end{cases} \tag{13}$$

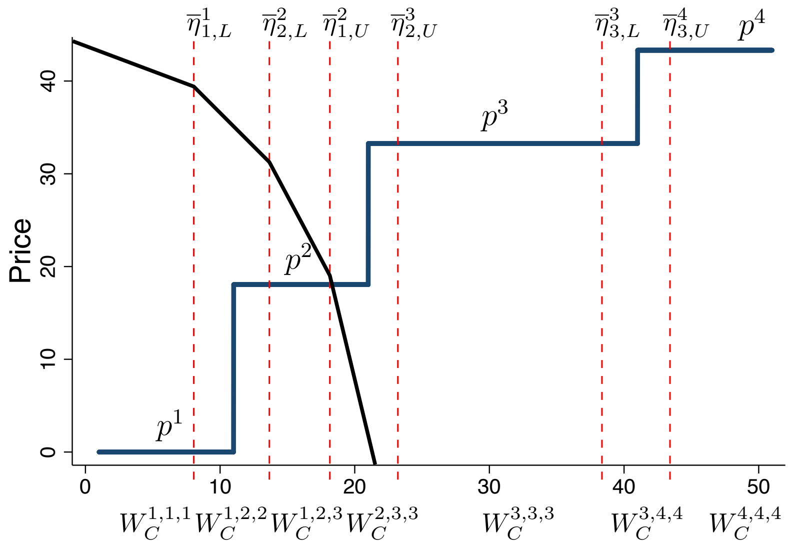
Where:

$$\begin{aligned} W_C^{L,M,H} &= \frac{w(P^{L,M,H})}{V^{L,H}} + \frac{\eta_C}{V^{L,H}} \\ w(p) &= \gamma_C - \alpha_C[H_C + \frac{p}{S}] \\ P^{L,M,H} &= \frac{1}{2}[p^H + p^L - \overline{W}^H(\frac{p^H - p^M}{\sqrt{2}\sigma_{\epsilon_C}}) - \overline{W}^M(\frac{p^M - p^L}{\sqrt{2}\sigma_{\epsilon_C}})] \\ V^{L,H} &= 1 + \frac{\alpha_C}{2S}(\frac{p^H - p^L}{\sqrt{2}\sigma_{\epsilon_C}}) \\ \overline{\eta}_{l,L}^j &= \overline{W}_l - w(p^j) - \sqrt{2}\sigma_{\epsilon_C} \ , \ \overline{\eta}_{l,U}^j = \overline{W}_l - w(p^j) + \sqrt{2}\sigma_{\epsilon_C} \end{aligned}$$

In this expression, $w_C^{L,M,H}$ captures the deterministic component of demand while $W_C^{L,M,H}$ adds the stochastic component of fixed preferences. Connection-level fixed preferences γ_C, price sensitivities α_C, and hassle costs H_C sum across their respective values for all households using a connection. A series of thresholds in the level of unobserved fixed preference, η_C, given by $\overline{\eta}_{l,L}^j$ and $\overline{\eta}_{l,U}^j$ work to determine which prices influence demand at each point in the schedule. Crossing a threshold means that the consumption shock now exposes the connection to a different set of prices. Figure 11 graphs an example demand curve where

[36] This approach is only relevant for strictly increasing block tariffs like the one in Manila. Tariffs that include both increasing and decreasing blocks require comparing indirect utilities across all segments as in Szabó [2015]. I extend the model to allow for decreasing marginal prices in the counterfactual exercises in Section 7.

Figure 11: Example Demand Curve



the slope is determined by the functions $W_C^{L,M,H}$ and the changes in slope are given by the preference thresholds.

This section now discusses the separate parts of the demand function in more detail.

Demand over a Single Price: The first, third, and fourth price segments of the tariff schedule in Manila given in Figure 11 exhibit large gaps of 20 m^3 between price changes. Within these segments, the consumption shocks are not large enough to ever bounce the households across a tariff kink point. Therefore, households are exposed only to the current price for that segment. Setting the low, medium, and high prices all equal to k in expression $W_C^{L,M,H}$, demand arrives at the following expression:

$$W_C = \gamma_C - \alpha_C[H_C + \frac{p^k}{S}] + \eta_C$$

This case produces a linear demand function where α_C can be readily interpreted as the price sensitivity parameter while γ_C captures the portion of demand that is invariant to prices. This demand function captures moral hazard from bill-splitting by dividing the price sensitivity, α_C, by the total number of households using the connection, S. Comparing demand among shared connections to individual connections, the extent of over-consumption at an average price of p can be roughly summarized as $\frac{S-1}{S}\alpha_C p$.

Demand over Many Prices: When households choose to consume close to price thresholds points, they become exposed to price segments both above and below the thresholds.

Demand becomes a weighted average of these two price segments. In this case, demand takes the following form, setting low and medium prices equal to k, and the high price equal to $k+1$:

$$W_C = \frac{1}{V^{k,k+1}}\Big(\gamma_C - \alpha_C H_C - \frac{\alpha_C}{2S}[p^k + p^{k+1} - \overline{W}^{k+1}(\frac{p^{k+1}-p^k}{\sqrt{2}\sigma_{\epsilon_C}})]\Big) + \frac{\eta_C}{V_{k+1,k}}$$

Setting $L = M = k$ simplifies the demand expression so that only the single nearby kink point, $\overline{W}^{k+1}$, is considered. Under increasing prices, a decrease in the kink point, $\overline{W}^{k+1}$, lowers demand by increasing the probability that the household faces a higher price.

When the distance between price segments is relatively small compared to the size of the consumption shocks, then households are exposed to three prices when consuming on a particular segment. Despite locating far from the neighboring kink points, households still face some probability of crossing into another price segment either above or below their current price. This demand function applies to households locating between the second and third tariff blocks because this gap only spans 10 m^3 (as shown in Figure 11). The demand function corresponds to the full expression for demand, setting indices for the low price equal to 1, the middle price equal to 2, and the high price equal to 3.

A.IX Determining the Demand Function with Respect to Variance of Consumption Shocks

The specification of the demand function is sensitive to ex-ante assumptions on the standard deviation of the consumption shocks, σ_ϵ. To identify the appropriate demand function in this empirical setting, I maximize the likelihood function in equation (8) over three demand specifications with different assumptions on σ_ϵ before selecting equation (13) as the primary specification.

I first construct and estimate demand under the limiting assumption that σ_ϵ falls below the minimum threshold, $min_{\forall k}\ \{\frac{\overline{W}_{k+1}-\overline{W}_k}{2\sqrt{2}}\} = \frac{10}{2\sqrt{2}} = 3.53$. Under this assumption, households only consider one price at all segments and two prices near thresholds. This assumption produces the following demand function:

$$D^*_C = \begin{cases} W_C^{1,1,1} & \text{if} \quad \frac{\eta_C}{V^{1,1}} \leq \overline{\eta}^1_{1,L} \\ W_C^{1,2,2} & \text{if} \quad \overline{\eta}^1_{1,L} < \frac{\eta_C}{V^{1,2}} \leq \overline{\eta}^2_{1,U} \\ W_C^{2,2,2} & \text{if} \quad \overline{\eta}^2_{1,U} < \frac{\eta_C}{V^{2,2}} \leq \overline{\eta}^2_{2,L} \\ W_C^{2,3,3} & \text{if} \quad \overline{\eta}^2_{2,L} < \frac{\eta_C}{V^{2,3}} \leq \overline{\eta}^3_{2,U} \\ W_C^{3,3,3} & \text{if} \quad \overline{\eta}^3_{2,U} < \frac{\eta_C}{V^{3,3}} \leq \overline{\eta}^3_{3,L} \\ W_C^{3,4,4} & \text{if} \quad \overline{\eta}^3_{3,L} < \frac{\eta_C}{V^{3,4}} \leq \overline{\eta}^4_{3,U} \\ W_C^{4,4,4} & \text{if} \quad \overline{\eta}^4_{3,U} < \frac{\eta_C}{V^{4,4}} \end{cases} \tag{14}$$

Assuming:

$$\sigma_{\epsilon_C} \leq \frac{\overline{W}^2 - \overline{W}^1}{2\sqrt{2}} = \frac{10}{2\sqrt{2}} = 3.53$$

Estimates according to this demand function in the first column of Table 18 find that the standard deviation for consumption shocks for connections serving single households through connections serving three or more households exceeds the 3.53 threshold in all cases. Next, I estimate the primary specification in equation (13), which holds under the hypothesis that σ_ϵ falls between 3.53 and $\frac{\overline{W}_2 - \overline{W}_1}{2\sqrt{2}} = \frac{20}{2\sqrt{2}} = 7.06$. Estimates from the second column of Table 18 fail to reject the hypothesis.

Finally, I allow for greater variance of σ_ϵ exceeding 7.06 and ensuring that the household responds to three prices at both interior price segments in the tariff schedule. This specification

Table 18: Alternative Demand Specification Estimates

	$[\sigma_\nu < 3.53]$	$[3.53 < \sigma_\nu < 7.06]$	$[7.06 < \sigma_\nu < 10.59]$
Price Sensitivity : α			
Intercept	0.84	0.83	0.89
HHsize 4 to 5	-0.02	-0.02	-0.02
HHsize over 5	-0.10	-0.10	-0.10
Apartment	-0.04	-0.04	-0.04
Single House	0.05	0.05	0.05
Low Skill Emp.	0.02	0.02	0.02
Over 2 Empl. Members	0.02	0.02	0.03
HH Head 36 to 52 years	0.08	0.08	0.08
HH Head Over 52	0.13	0.13	0.14
Consumption Shock : σ_ν			
Single HH	4.10	4.10	4.44
Two HHs	5.31	5.30	5.47
Three plus HHs	5.76	5.78	5.92
Preference Shock : σ_ψ			
Intercept	13.34	13.31	13.38
Below 1st Quintile Usage	-3.00	-2.99	-3.16
Over 3rd Quintile Usage	2.82	2.82	2.81
Two HHs	-0.48	-0.44	-0.33
Three HHs	1.12	1.13	1.27

takes the following functional form:

$$D_C^* = \begin{cases} W_C^{1,1,1} & \text{if} \quad \frac{\eta_C}{V^{1,1}} \leq \overline{\eta}_{1,L}^1 \\ W_C^{1,2,2} & \text{if} \quad \overline{\eta}_{1,L}^1 < \frac{\eta_C}{V^{1,2}} \leq \overline{\eta}_{2,L}^2 \\ W_C^{1,2,3} & \text{if} \quad \overline{\eta}_{2,L}^2 < \frac{\eta_C}{V^{1,3}} \leq \overline{\eta}_{1,U}^2 \\ W_C^{2,3,3} & \text{if} \quad \overline{\eta}_{1,U}^2 < \frac{\eta_C}{V^{2,3}} \leq \overline{\eta}_{3,L}^3 \\ W_C^{2,3,4} & \text{if} \quad \overline{\eta}_{3,L}^3 < \frac{\eta_C}{V^{2,4}} \leq \overline{\eta}_{2,U}^3 \\ W_C^{3,4,4} & \text{if} \quad \overline{\eta}_{2,U}^3 < \frac{\eta_C}{V^{3,4}} \leq \overline{\eta}_{3,U}^4 \\ W_C^{4,4,4} & \text{if} \quad \overline{\eta}_{3,U}^4 < \frac{\eta_C}{V^{4,4}} \end{cases} \tag{15}$$

Assuming:

$$\sigma_{\epsilon_C} \geq \frac{\overline{W}^2 - \overline{W}^1}{\sqrt{2}} = \frac{10}{\sqrt{2}} = 7.06$$
$$\sigma_{\epsilon_C} \leq \frac{\overline{W}^3 - \overline{W}^2}{\sqrt{2}} = \frac{20}{\sqrt{2}} = 14.12$$

The final column in Table 18 again rejects this hypothesis finding estimates of σ_ϵ that fall below the 7.06 threshold. Comparing all of the parameter estimates across specifications, I find very minor differences according to the different demand functions, providing suggestive evidence that the estimation strategy is robust to ex-ante assumptions on σ_ϵ.

A.X Hassle Cost: Neighbor Radius

The following equation predicts usage for neighboring households, i, relative to a nearby leak according to their distance rank, DR, relative to the leaking household with rank 1 being closest and rank D being furthest from the leaking household. To capture any spatial changes in consumption over time, the regression also allows for differential time trends before and after the leak according to distance rank, DR, captured by λ_1^d and λ_2^d:

$$w_{t,i} = \sum_{d=1}^{D} \mathbb{1}\{DR_i = d\} \Big[\beta^d Post_{i,t} + \lambda_1^d r_{i,t} + \lambda_2^d Post_{i,t} \times r_{i,t}\Big] + \theta_i + \epsilon_{t,i}$$

$r_{i,t}$: Months to leak

DR_i : Distance Rank Relative to Leaking Connection

$Post_{i,t}$: After leak

Table 19: Sharing and Distance to Leaking Connection

VARIABLES	(1) Q (m3)
Distance Rank 1	0.972*
	(0.504)
Distance Rank 2	0.981**
	(0.444)
Distance Rank 3	-0.760
	(0.479)
Distance Rank 4	0.253
	(0.512)
Distance Rank 5-5	0.216
	(0.517)
Constant	22.68***
	(0.124)
Observations	59,225
R-squared	0.678
Pre Post Trends by Rank	Yes

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The coefficients of interest, β^d, are presented in Table 19. I find that the closest neighbor receives the most consumption from the leaking household, closely followed by the second and fourth neighbors. Substitution drops off quickly past the fourth neighbor, reaching statistically insignificant and economically smaller magnitudes for the fifth through eighth neighbors.

A.XI Hassle Cost Assumptions

The following additional assumptions are required to ensure that this empirical setting accurately captures the model described in Section 5.2.1:

1. *Connections that experience a leak are representative of the population of connections.* This assumption is supported by small economic differences in means of observable measures between connections with leaks and without in Table 20 although some differences are statistically significant.

2. *The decision to disconnect conditional on experiencing a leak is uncorrelated with*

Table 20: Descriptives and Balance Test for Non-Leaking and Leaking Connections

	Mean No Leak	Mean Leak	Diff.	Std. Error
Cons. (avg.)	29.43	32.42	-2.98***	0.28
Age HoH	46.56	46.36	0.20	0.24
HH Size	5.14	5.12	0.01	0.03
HH Size Empl.	1.58	1.62	-0.04**	0.02
Low Skill Emp.	0.16	0.14	0.02***	0.01
Apartment	0.21	0.23	-0.02***	0.01
Single House	0.58	0.57	0.01	0.01
Obs.	56853.00	3976.00	52877.00	0.00

Drink is a dummy for drinking tap water. Water flow is a dummy for low water pressure. Water storage is a dummy for owning a water storage container. Std. errors result from a T-test of differences in means.

Table 21: Balance Test for Disconnection among Leaking Connections

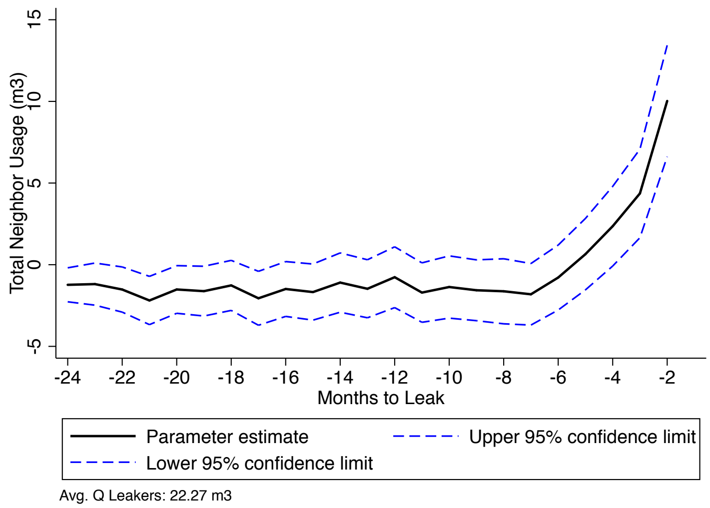
	Mean No DC	Mean DC	Diff.	Std. Error
Cons. (avg.)	28.026	30.063	-2.037***	0.402
Leak Size	42.304	59.398	-17.094***	1.216
Age HoH	46.400	46.209	0.191	0.319
HH Size	5.094	4.924	0.170***	0.047
HH Size Empl.	1.538	1.500	0.037	0.024
Low Skill Emp.	0.154	0.180	-0.026***	0.008
Apartment	0.154	0.166	-0.012	0.008
Single House	0.672	0.626	0.046***	0.010
Obs.	32246.000	2270.000	29976.000	0.000

Drink is a dummy for drinking tap water. Water flow is a dummy for low water pressure. Water storage is a dummy for owning a water storage container. Std. errors result from a T-test of differences in means.

water demand. Table 21 provides partial support by showing few large differences in characteristics between those that disconnect and those that remain connected other than the size of the leak. This evidence suggests that the disconnection decision is driven primarily by variation in the magnitude of the leak.

3. *The timing of a leak is uncorrelated with any underlying changes in demand.* I test this assumption with a regression predicting a connection i's consumption in the 30

Figure 12: Pre-Trend in Leaker's Consumption



months before the leak event controlling for a connection fixed effect, θ_i:

$$w_{t,i} = \sum_{r=-30}^{0} \beta_r \mathbb{1}\{t = r\} + \theta_i + \epsilon_{t,i}$$

$$r : \text{Month to leak}$$

$$\theta_i : \text{Connection-level fixed effect}$$

I include a dummy variable for each month before the leak. Figure 12 plots demeaned monthly consumption relative to the month of the leak. This figure shows stable a consumption pattern interrupted by a large jump at the exact time of the leak.

4. *Leaks are uncorrelated with changes in neighborhood demand.* A simple regression tests this assumption by predicting total consumption among four closest neighbors in neighborhood N, before and after a nearby leak, controlling for a neighborhood fixed effect, θ_N. I include a dummy variable for each month relative to the month that the neighbor leaks:

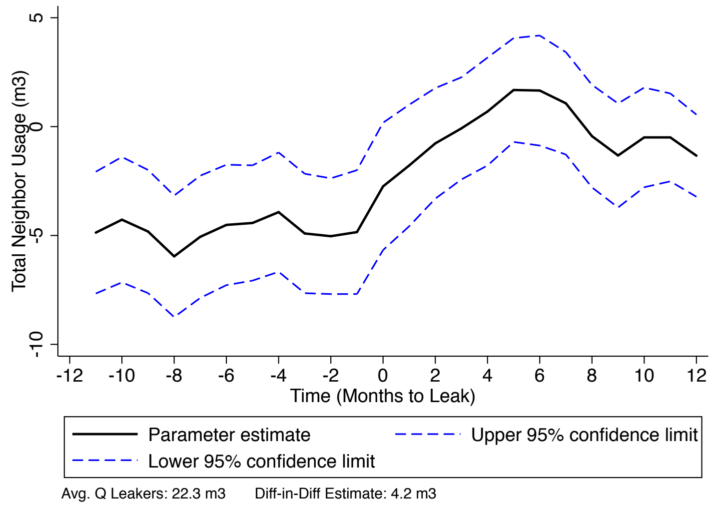
$$w_{N,t} = \sum_{r=-14}^{14} \beta_r \mathbb{1}\{t = r\} + \theta_N + \epsilon_{N,t}$$

$$r : \text{Month to leak}$$

$$N : \text{Neighborhood}$$

Figure 13 plots the results of this regression indicating average total neighborhood

Figure 13: Neighbors' Water Usage following a Leak Disconnection



consumption in the months just preceding and following the date that the leakis reported. I find a stable trend in neighbors' consumption before the leak followed by a quick leap as the leaking household starts to share with their neighbors. This sharing behavior remains persistent for at least a year after the initial pipe leak, providing a useful window of time to estimate the hassle cost.

5. *The leaking household's choice of a sharing partner is uncorrelated with that partner's water demand.* This assumption holds when the ω term in equation (10) is equal to zero. If ω is nonzero allowing leaking households to weigh the utility from using each source in this decision, households may choose to share with low demand neighbors to enjoy lower points on the tariff in Figure 1. To test this hypothesis, I predict the heterogeneous increase in neighbor's consumption following a leak according to that neighbor's water demand, which is proxied by average consumption prior to the leak. The following regression equation tests this relationship by predicting changes in usage according to the rank of average consumption at baseline among all neighbors:

$$w_{t,i} = \sum_{d=1}^{4} \mathbb{1}\{Rank_i = d\}\beta^q Post_{i,t} + \theta_i + \epsilon_{t,i}$$

$$r : \text{ Month to leak}$$

$$Rank : \text{ Rank of avg. usage in the neighborhood at baseline}$$

$$\theta_i : \text{ Connection-level fixed effect}$$

One issue with this analysis is that predicting the usage response of a neighbor as

Table 22: Leak Demand Response and Rank of Baseline Usage in Neighborhood

VARIABLES	(1) Q (m3)	(2) Q (m3)	(3) Difference
Demand Rank 1	1.480***	1.334***	0.145
	(0.247)	(0.292)	
Demand Rank 2	0.918**	-0.0839	1.002
	(0.361)	(0.398)	
Demand Rank 3	0.157	1.275*	-1.117
	(0.541)	(0.663)	
Demand Rank 4	-0.122	-0.813	0.691
	(1.013)	(0.929)	
Demand Rank 5	1.769	4.117	-2.348
	(1.915)	(3.283)	
Constant	22.06***	21.99***	
	(0.0799)	(0.0409)	
Observations	59,225	42,108	
R-squared	0.678	0.710	
Connection FE	Yes	Yes	
Simulated Leak	No	Yes	

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

a function of their previous consumption is likely to produce mean reversion bias: connections with low usage at baseline demonstrate larger increases in consumption than connections with high usage at baseline due to natural variation in preference shocks. To test for mean reversion, I create a random treatment date using only data before the leak has occurred. Table 22 presents the β^d estimates of the true leak-date specification in the first column and the mean randomized leak-date estimates in the second column. The third column takes the difference of the first and second columns, which provides a coarse method of netting out any mean reversion bias. The first column predicts strong increases in consumption for smaller users (ranks 1 and 2) while large users (ranks 3 and 4) exhibit small and even negative demand responses. These differential responses are amplified in the second column (with placebo treatment dates), suggesting that mean reversion may play a large role in biasing estimates. The third column shows that after adjusting for mean reversion bias, the demand response of neighboring connections is roughly equal as the relative demand increases for these connections. This finding provides suggestive evidence that households are not selecting sharing partners conditional on their neighbors' water demands.

6. *In response to a leak, the choice to share with a neighbor or use a vendor is uncorrelated with water demand.* This assumption also holds when the ω term in equation (10) is equal to zero. If water demand were correlated with substitution across these sources, then I would expect to find that the water demand of the leaking household predicts a differential rise in usage among neighbors following the leak. Since vended water is characterized by high prices and low fees, a simple model would predict that households with low water demand would substitute toward vendors at a higher rate and would therefore offset less consumption to neighboring households following a leak. To test this assumption, I use average consumption for a leaking household prior to the leak as a proxy for their water demand.[37] The following equation predicts total usage in neighborhood, N, before and after a household in that neighborhood experiences a leak. The regression allows for separate impacts of a neighbor's leak, β^d depending on that neighbor's quintile of average consumption (measured before the leak).

$$w_{t,N} = \sum_{d=1}^{5} \mathbb{1}\{Quintile_N = d\}\beta^d Post_{N,t} + \theta_N + \epsilon_{N,t}$$

$$r : \text{Month to leak}$$

$$Quintile : \text{Quintile of avg. usage for leaking connections at baseline}$$

$$\theta_N : \text{Neighborhood-level fixed effect}$$

Table 23 presents the results from this regression in the first column. The size of the coefficients increase as the quintile of leaker demand increases, which is consistent with larger users at baseline using greater amounts of water from their neighbors. To compare the relative demand response for large and small users, column three divides this coefficient estimate by the average usage for each quintile at baseline to arrive at the percentage of water consumption that is offset among neighbors after a leak occurs. Column three finds no strong trends in the share of consumption that is offset to neighbors across water demand quintiles. This finding provides suggestive evidence that household's are left to use whichever source is most immediately available (neighbor or vendor) following a leak.

7. *A leak for one household is uncorrelated with probability of a neighboring household experiencing a leak.* While the data show little evidence that leaks are spatially correlated, I cannot rule out unreported leaks affecting nearby households. The presence of unobserved spatial correlation in leaks would lead this method to overestimate the

[37]Using the same test as before, I detect no evidence of mean reversion given this demand proxy.

Table 23: Neighborhood Usage Response and Quintile of Usage for Leaking Connection

VARIABLES	(1) Q (m3)	(2) Avg. Q (m3)	(3) Share Offset
Quintile 1	-0.397	9.704	-0.0409
	(1.612)		
Quintile 2	5.290***	16.69	0.317
	(1.587)		
Quintile 3	4.311**	22.66	0.190
	(1.818)		
Quintile 4	4.560**	29.96	0.152
	(1.927)		
Quintile 5	3.838**	47.79	0.0803
	(1.918)		
Constant	65.40***		
	(0.331)		
Observations	18,763		
R-squared	0.788		
Neighborhood FE	Yes		

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

hassle costs associated with sharing a connection.

A.XII Game Tree with Three Players

This section maps the game tree for three players, 1, 2, and 3 in Figure 14.

A.XIII Out-of-Sample Test: A Connection Fee Discount Program

A connection fee discount program implemented by the water company provides a useful setting to test the out-of-sample predictions of the economic model: do predicted water source choices match observed choices in response to a discount in the fixed fee? According to the model, reducing the fixed fee would make individual connections more attractive relative to shared connections or water vendors. This section compares the predicted and observed increase in connections in response to the program.

In 2009, the government regulator issued new regulations reducing the connection fee from 6,000 to 2,400 PhP for low-income areas partially in an effort to extend access and affordability to low-income households. Under the assumption that households reside in Manila for 80 months and face negligible credit constraints, this discount lowered the monthly fixed fee faced by households by 45 PhP on average. In the presence of credit constraints, this estimate can be interpreted as a lower bound on the policy's impact on the monthly fixed price faced by households. Figure 15 indicates the share of new connections receiving discounted fees. The sudden rise by 37 percentage points coincides with the change in regulations and will be used to identify the timing of the discount policy. A small fraction of households are observed to receive connection fee discounts prior to the program as part of other, smaller pro-poor strategies implemented by the water provider.

Since the data do not identify precisely which local areas or specific households were eligible for the discount, additional assumptions are necessary to measure eligibility. First, I define treated areas as any meter reading unit (MRU) that receives at least one discounted connection over the full duration while control areas include any MRUs that never receive discounts. MRUs are common geographical designations used by the water provider and include between 100 and 500 meters. There are 3,757 MRUs in the service area. I include only MRUs that were established long prior to the discount fee program to ensure that these areas

Figure 14: Sequential Game Choosing Water Sources

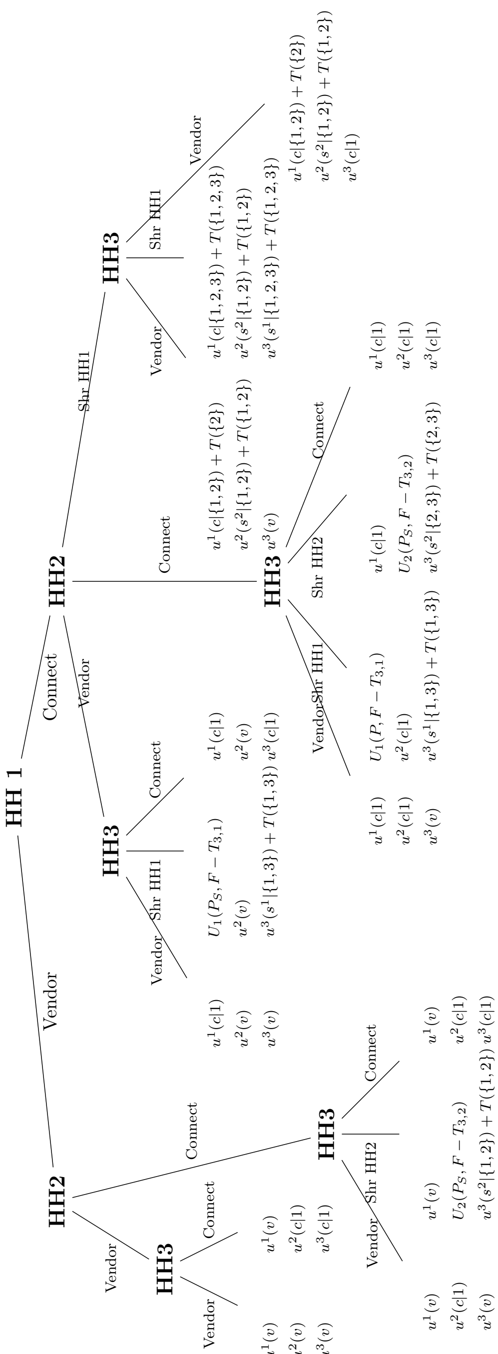
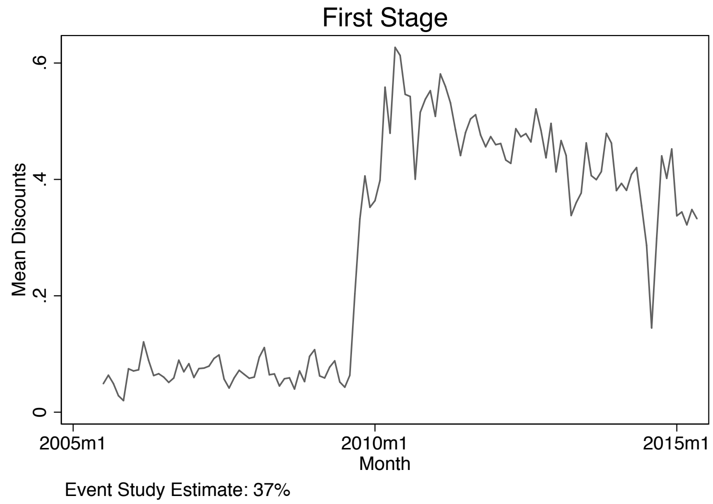


Figure 15: Share of New Connections Receiving a Discount



are likely to be at saturation in terms of water connections prior to the program. Second, to identify the share of households in treated areas that were eligible, I conceptually divide new connections into two groups — (1) households substituting away from shared connections or water vendors in response to the program and (2) new households migrating to the city. I then measure eligibility as the share of new migrants receiving discounted connections fees after the policy.

This approach requires the following assumptions:

1. Before the discount program, households that have already moved to the city have reached an equilibrium at the current prices; therefore, all new full-price connections in the pre-period are driven by migration to the city. This assumption is consistent with the large fixed costs incurred in setting up a new connection.

2. The rate of migration is unaffected by the policy, which is supported by the stable rate of new connections for untreated areas in Figure 16 below.

3. After the discount policy is implemented, all new migrant households that are eligible take up the discounted connections.

4. Rates of eligibility are similar across migrant and non-migrant groups.

Under these assumptions, I compute the impact of the policy on new full-price connections in the treated areas and interpret any reduction in the rate of new full-price connections as driven by new migrants receiving discounted connections. I estimate the following simple difference-in-differences equation across all treated and untreated Meter Reading Units, m,

Table 24: Difference-in-Differences Estimate for Full Priced Connections

	(1)
VARIABLES	Diff-in-Diff Estimate
Post X Treated	-0.184***
	(0.0168)
Post	0.205***
	(0.00976)
Treated	0.343***
	(0.0156)
Constant	0.0901***
	(0.00677)
Observations	382,462
R-squared	0.004
Avg. New Conn.	.39

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Clustered at the MRU Level: 3,757 MRUs

and months, t, pre and post policy implementation:

$$\text{New Full-Price Connections}_{t,m} = \beta_1 \text{Post X Treated}_{t,m} + \beta_2 \text{Post}_{t,m} + \beta_3 \text{Treated}_{t,m} + \epsilon_{t,m}$$

Table 24 presents the results from this estimation.[38] Adding the coefficients in Table 24 shows that in the absence of the policy, treated areas were expected to have added 0.64 new full-price connections each month; yet, as a result of the policy, these areas actually experienced 0.46 new full-price connections with the remaining households receiving discounted connections. This discrepancy implies that 28% of households were eligible for the program.

I next examine the reduced-form impacts of the policy using the following event-study regression equation, which computes the average number of new connections in bi-monthly intervals relative to the discount fee program for both treated and untreated areas:

$$\text{New Connections}_{t,m} = \sum_{r=0}^{T} \beta_r^{\text{Treated}} \mathbb{1}\{t = r\} + \sum_{r=0}^{T} \beta_r^{\text{Untreated}} \mathbb{1}\{t = r\} + \theta_m + \epsilon_{t,m}$$
$$t = 2 \text{ Month Intervals}$$
$$m = \text{Meter Reading Unit}$$

[38]These results do not exactly match reduced-form results from Table 25 because the data only have information on the price per connection for about 80% of new connections.

Figure 16: Reduced Form: Connection Fee Discount

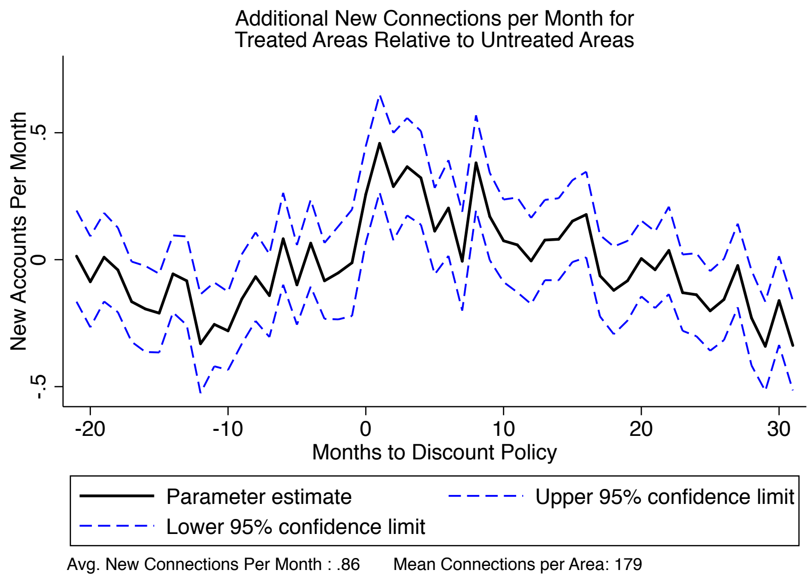


Figure 16 provides the graphical estimates from this equation. The differences-in-differences estimate shows a relatively stable pre-trend interrupted by a substantial increase in the number of connections just following the discount program implementation. This increase lasts for a couple years before returning to pre-program levels. This trend provides evidence that households who are sharing water or purchasing from vendors are gradually learning about the price change and choosing to purchase their own connections. After a couple years, this additional demand for connections has been satisfied, letting migration take over as the sole driver of new connections.

A simple difference-in-differences estimate provided by Table 25 calculates an increase of 0.13 new connections per month for the treated areas on a base average of 0.79 new connections per month. Aggregated over the full 69 months of the program, this program led to a total of 8.74 new connections per MRU. Given that the average MRU has 179 connections over the full time period, this program generated about a 3.5% increase in the number of new connections.

To compare results from this policy to counterfactual estimates from the model, I now convert the change in number of connections induced by the policy to a shift from shared and vendor connections to individual connections. Given that 14% of connections serve two households and 10% of connections serve three or more households, and 6% of households access water from vendors, then 179 connections per MRU are calculated to serve an average population of around 250 households. This policy results in a 3.5 percentage point increase in the share

Table 25: Difference-in-Differences Estimate

VARIABLES	(1) Diff-in-Diff Estimate
Post X Treated	0.127***
	(0.0215)
Post	0.0163
	(0.0122)
Constant	0.794***
	(0.00855)
Observations	382,462
R-squared	0.062
Avg. New Conn.	.86
Area FE	Yes

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Clustered at the MRU Level: 3,757 MRUs

of households owning a connection, which can be directly compared to predictions from the counterfactual exercise.

In order to relate the quasi-experimental results from the policy to the model outputs, I simulate the impacts of the connection-fee discount policy by (1) randomly selecting 28% of households, (2) lowering the fixed fee by 45 PhP for these households, and (3) recomputing their water source choices. As a result of this policy, the share of households owning a water connection is calculated to rise by 1.34 percentage points. This finding is smaller, but comparable to the reduced form results indicating a 3.5 percentage point increase. There may be many reasons that the counterfactual results lag behind the reduced-form estimates. First, the assumptions used to amortize the connection fee discount into monthly terms may work to understate the impact of the discount, especially in cases where households face credit-constraints in funding large, lump-sum connection fees. Second, in practice, the discount policy targeted a low-income segment of the population, which may have different substitution patterns across water sources.

Despite these caveats, the fact that the reduced-form and model-based estimates are similar in magnitude lends support to the counterfactual exercises implemented in Section 7. Table 26 calculates the impacts of this connection fee discount on water source choices as well as consumer welfare. Consumer welfare is computed as the difference between current utility and the utility that each household would receive from using vended water. Since

Table 26: Connection Fee Discount: Out-of-Sample Test Results

	Current	Discount
Fixed Fee (PhP)	225	180
Source Vendor	0.044	0.039
Source Neighbor	0.243	0.235
Surplus (PhP)	143.0	153.1
Surplus: Low Users (PhP)	76.4	84.0

Fixed fee and consumer surplus are in PhP/month.

utility is assumed to be quasi-linear, consumer welfare is measured in PhP/month. Lowering the fixed fee leads to substitution away from both vendor and neighbor water sources, which improves consumer welfare especially for the bottom 50% of users.

A.XIV Test for Incidental Parameter Bias

The estimation strategy relies heavily on connection-specific fixed preference terms estimated in the Section 6.1 to recover parameters in Section 6.2 and Section 6.3. Since the data contain a limited number of months of consumption for each connection (averaging around 50 months), noise in the estimation of these preference terms may produce bias in latter estimates. I test for this type of bias by simulating data given true parameters close to the estimates and reestimating the model on this simulated data. Table 27 provides results for this test.

The first column presents the true assumed parameter values, the second column provides the estimation results for the simulated data using the full sample, and the third column includes estimates for a subsample of the data where I have dropped half of the observations for each connection. Examining the second column and percent deviations from the true parameters in the third column, I find that the incidental parameters do not seem to lead to a large degree of bias in any of the coefficients with the maximum occurring for the variance estimate for the vendor price. Dropping half of the time periods, the fourth and fifth columns indicate that the hassle cost and connection fee parameters would be the most sensitive to any possible bias from incidental parameters.

Table 27: Simulation Test for Incidental Parameter Bias

Parameter	True Value	Est: Full T	(% Diff)	Est: 50% T	(% Diff)
σ_ϵ	5.00	5.03	0.01	5.01	0.00
σ_η	12.00	11.91	-0.01	11.79	-0.02
α	0.50	0.51	0.02	0.51	0.03
γ (avg.)	34.32	34.60	0.01	33.98	-0.01
H	15.00	14.17	-0.06	11.74	-0.22
F	400.00	393.72	-0.02	378.30	-0.05
F_V	100.00	99.87	-0.00	98.70	-0.01
p_V	40.00	41.00	0.02	40.48	0.01
σ_v	5.00	5.69	0.14	5.88	0.18

T refers to months observed for each connections. The first sample includes half of the time periods per connection (around 35 months), while the second sample includes all months. Both samples include 5,000 accounts.

A.XV Out of Sample Test for Vendor Price Estimates

As a simple out-of-sample test, I compare the estimate of p_V in Table 7 to the implied p_V from data on total water expenditures for vendor households using additional census data from a section of Manila. The 2011 Community Based Monitoring System data cover over 70% of households in Pasay City — a large area with a population of 500,000 in downtown Manila — and includes information on income (used to compute imputed income) as well as water source and monthly expenditures. In this data, households using from vendors report expenditures of 365 PhP/month while households using individual connections report expenditures of 470 PhP/month and households sharing connections report expenditures of 365 PhP/month. I can express average marginal expenditures, E_v, by multiplying average vendor demand by the vendor price and adding the vendor fixed cost, $E_v = (\gamma - \alpha p_v)p_v + F_v$. Plugging in expenditures, $E_v = 365$ PhP/m^3, average fixed preferences for vendors, $\gamma = 37$, and price sensitivity, $\alpha = .91$, as well as solving the quadratic expression, I find that this level of expenditure is consistent with vendor prices of either 7.9 PhP/m^3 or 32.8 PhP/m^3. The latter price almost perfectly matches the estimated price of 32.85 PhP/m^3, providing additional support for the structural estimates from Section 6.3.

A.XVI Simulated Method of Moments Model Fit

By weighting the first three moments (water source choices) highly by design, the model almost exactly matches moments shown in Table 28. The model is also able to match the direction of correlations between fixed preferences for water β_i and percent owning connections

Table 28: Simulated Method of Moments Model Fit

Moments	%Owning	%Use Neighbor	%Vendor	Correlation (β_i, %Own.)	Correlation (β_i, % Neigh.)	Correlation (β_i, %Vendor)
True	0.714	0.245	0.041	0.377	-0.375	-0.046
Estimated	0.714	0.243	0.042	0.525	-0.554	0.003

as well as percent using from a neighbor; however, the model predicts stronger correlations in both cases. One interpretation is that there may be other factors besides simply water preferences and heterogeneity in hassle costs that are driving water source decisions. The model has difficulty matching correlation between fixed preferences and share using from vendors, which may be partially due to noise generated in the process of imputing fixed preferences for households using from vendors.

References

A. J. Auerbach and A. J. Pellechio. The two-part tariff and voluntary market participation. *The Quarterly Journal of Economics*, pages 571–587, 1978.

D. Bennett. Does clean water make you dirty? water supply and sanitation in the philippines. *Journal of Human Resources*, 47(1):146–173, 2012.

J. J. Boland and D. Whittington. The political economy of water tariff design in developing countries: increasing block tariffs versus uniform price with rebate. *The political economy of water pricing reforms*, 215235, 2000.

S. Borenstein. The redistributional impact of nonlinear electricity pricing. *American Economic Journal: Economic Policy*, 4(3):56–90, 2012.

G. Burtless and J. A. Hausman. The effect of taxation on labor supply: Evaluating the gary negative income tax experiment. *Journal of political Economy*, 86(6):1103–1130, 1978.

D. Cutler and G. Miller. The role of public health improvements in health advances: the twentieth-century united states. *Demography*, 42(1):1–22, 2005.

F. Devoto, E. Duflo, P. Dupas, W. Parienté, and V. Pons. Happiness on tap: piped water adoption in urban morocco. *American Economic Journal: Economic Policy*, 4(4):68–99, 2012.

D. Diakité, A. Semenov, and A. Thomas. A proposal for social pricing of water supply in côte d'ivoire. *Journal of Development Economics*, 88(2):258–268, 2009.

L. Einav, A. Finkelstein, S. P. Ryan, P. Schrimpf, and M. R. Cullen. Selection on moral hazard in health insurance. *American Economic Review*, 103(1):178–219, 2013.

M. S. Feldstein. Equity and efficiency in public sector pricing: the optimal two-part tariff. *The Quarterly Journal of Economics*, pages 176–187, 1972.

S. Galiani, P. Gertler, and E. Schargrodsky. Water for life: The impact of the privatization of water services on child mortality. *Journal of political economy*, 113(1):83–120, 2005.

S. Gamper-Rabindran, S. Khan, and C. Timmins. The impact of piped water provision on infant mortality in brazil: A quantile panel data approach. *Journal of Development Economics*, 92(2): 188–200, 2010.

S. F. Hoque and D. Wichelns. State-of-the-art review: designing urban water tariffs to recover costs and promote wise use. *International journal of water resources development*, 29(3):472–491, 2013.

R. Jimenez-Redal, A. Parker, and P. Jeffrey. Factors influencing the uptake of household water connections in peri-urban maputo, mozambique. *Utilities Policy*, 28:22–27, 2014.

S. Kayaga and R. Franceys. Costs of urban utility water connections: Excessive burden to the poor. *Utilities Policy*, 15(4):270–277, 2007.

S. Keener, M. Luengo, and S. G. Banerjee. Provision of water to the poor in africa: Experience with water standposts and the informal water sector. 2010.

K. Komives. *Water, electricity, and the poor: Who benefits from utility subsidies?* World Bank Publications, 2005.

K. Komives, J. Halpern, V. Foster, Q. T. Wodon, and R. Abdullah. The distributional incidence of residential water and electricity subsidies. 2006.

J.-J. Laffont. *Regulation and development.* Cambridge University Press, 2005.

J.-J. Laffont and J. Tirole. *A theory of incentives in procurement and regulation.* MIT press, 1993.

A. McIntosh. *Asian water supplies: reaching the urban poor.* Asian Development Bank, 2003.

S. McRae. Infrastructure quality and the subsidy trap. *The American Economic Review*, 105(1): 35–66, 2014.

R. Moffitt. The econometrics of piecewise-linear budget constraints: a survey and exposition of the maximum likelihood method. *Journal of Business & Economic Statistics*, 4(3):317–328, 1986.

C. Nauges and C. Van den Berg. Water markets, demand, and cost recovery for piped water supply services: evidence from southwest sri lanka. 2006.

W. Y. Oi. A disneyland dilemma: Two-part tariffs for a mickey mouse monopoly. *The Quarterly Journal of Economics*, 85(1):77–96, 1971.

S. M. Olmstead. Reduced-form versus structural models of water demand under nonlinear prices. *Journal of Business & Economic Statistics*, 27(1):84–94, 2009.

J. Strand and I. Walker. Water markets and demand in central american cities. *Environment and Development Economics*, 10(03):313–335, 2005.

A. Szabó. The value of free water: Analyzing south africa's free basic water policy. *Econometrica*, 83(5):1913–1961, 2015.

C. Timmins. Measuring the dynamic efficiency costs of regulators' preferences: Municipal water utilities in the arid west. *Econometrica*, 70(2):603–629, 2002.

Water and Sanitation for the Urban Poor. Getting to scale in urban water supply. 2013.

D. Whittington. Possible adverse effects of increasing block water tariffs in developing countries. *Economic Development and Cultural Change*, 41(1):75–87, 1992.

WHO/UNICEF. Piped water coverage, 2015. data retrieved from World Development Indicators, `https://data.worldbank.org/indicator/SH.H2O.BASW.UR.ZS`.