# README

Chanwool Kim*

This version: November 3, 2023

## Replication Instruction

### Archive Structure

### (0) Preliminary

`preliminary.R` loads relevant packages and records paths to directories. It also contains table and graph-generating codes.

### (1) Data Build

`master.R` runs the data build in sequence. For each script, we have:

- `geocoding_api.R` assigns Census tract based on address information in the Portland data. It links the Census tract to each location number.

- `tidy_census_api.R` loads the Census data and saves relevant neighborhood characteristics variables.

- `setup_portland.R` loads raw Portland data and saves in the `.RData` format.

- `delinquency_measure.R` constructs delinquency-related variables: delinquency status and amount, shutoff status.

- `financial_assistance_clean.R` constructs variables for various financial assistance programs: financial assistance and payment arrangement.

- `merge_data.R` merges all data at the account level, saves `account_info_analysis.RData`.

- `bill_usage_clean.R` cleans the detailed financial information data and bill usage data. It summarizes consumption and rates, along with fees, at the bill level.

- `delinquency_construct.R` creates the bill info dataset with delinquency measures and financial assistance information.

---

*Kenneth C. Griffin Department of Economics, University of Chicago.

- `panel_construct.R` creates the panel dataset of all the bills by merging each dataset generated in previous steps at the bill level.

- `panel_estimation_construct.R` aggregates monthly bills into quarterly bills and flags any irregular bills.

## (2) Descriptive Statistics

`master.R` runs the descriptive statistics in sequence. For each script, we have:

- `descriptive_statistics_overview.R` creates tables for basic descriptive statistics and overview of the data (account types).

- `descriptive_statistics_delinquency.R` creates tables and graphs for delinquency measures by the Census tract.

- `descriptive_statistics_graph.R` creates graphs for basic descriptive statistics (pie charts).

- `descriptive_statistics_resmf.R` creates tables for basic descriptive statistics for residential multi-family units only.

- `descriptive_statistics_payment_overview.R` creates tables/graphs for the payment arrangement program.

- `descriptive_statistics_payment.R` creates tables/graphs for the payment arrangement program.

- `descriptive_statistics_linc.R` creates tables/graphs for the financial assistance program.

- `descriptive_statistics_did.R` creates tables for a rudimentary difference-in-differences exercise of delinquency rates on various assistance programs.

## (3) Model

- `static_demand.R` runs a static demand model.

# Data Construction

## Accessing the Data

All of the data used in this study were provided by the Portland Water Bureau.

## Project Overview

The objective of this project is to estimate the effect of personalized pricing for water bills on delinquency. To do so, we first need to estimate the propensity to pay and price elasticity of water usage. While we plan to use experimental data to get the estimates, we want to exploit the existing price variations and assistance programmes. Therefore, we need to construct the following data for each account:

- History of bills and payments each quarter

- History of participation in existing programs

- History of usage each quarter

We create the data we want in the following steps:

1. Clean and merge raw data on the account-quarter bill level

2. Aggregate data to a quarterly level for any monthly plans

3. Flag any anomalies

## Raw Data

All of the data are initially loaded and transformed to `.RData` file in `setup_portland.R`. All are merged to the account-quarter bill level in `panel_construct.R` and `panel_construct_estimation.R`.

Table 1: Raw Data List

| Data Name [Original Name] | Dimensions | Processing Script | Relevant Raw Variables | Identifier |
|---|---|---|---|---|
| account_info [UM00200M] | 700006×67 | merge_data.R | account info[1], LAST_BILL_AMT | account, person |
| bill_info [UM00260T] | 5105378×78 | delinquency_measure.R | dates[2], bill character[3], bill amounts[4] | account, person |
| financial_info [AR00200T] | 20135250×63 | bill_usage_clean.R | ADJUSTED_BILL_AMT, ITEM_SUMMARY | account, location |
| usage_info [UM00262T] | 39238143×50 | bill_usage_clean.R | consumption character[5], consumption amount[6] | account, location |
| address_info [UM00100M] | 198090×58 | geocoding_api.R | address[7] | location |
| location_relation [UM00120T] | 598021×30 | setup_portland.R | - | account, person, location |
| financial_assist [UM00232T] | 55717×13 | financial_assistance_clean.R | - | account |
| financial_assist_detail [Linc Data] | 182134×28 | financial_assistance_clean.R | dates[8], consumption[9], fees[10], bill[11], LINC_TIER_TYPE | account |
| payment_arrangement [CO00200M] | 850272×32 | financial_assistance_clean.R | STATUS_CD, START_DT, END_DT, ARRANGEMENT_AMT | account |
| payment_arrangement_info [CO00210T] | 1619030×14 | financial_assistance_clean.R | AMOUNT_DUE, OUTSTANDING_AMT | account[12] |
| cutoff_info [RS00200M_CUTOF] | 11340×63 | delinquency_measure.R | EFFECTIVE_DT | account, person, location |
| reconnect_info [RS00200M_RCNCT] | 12132×63 | delinquency_measure.R | EFFECTIVE_DT | account, person, location |
| collection_info [CO00400T] | 79582×12 | panel_construct.R | - | account |
| collection_amount [CO00450T] | 67684×53 | panel_construct.R | AMT_DUE, ACT_COL_AMT | account |
| code_info [AR50100C] | 162×50 | setup_portland.R | address | - |

## Water Bill

This section provides a brief discussion on Portland Water Bureau's billing rules and how we deal with different prices.

The city adjusts water rates annually on July 1. Water use is measured in ccf or centum (100) cubic feet. One ccf is 748 gallons. The meter rounds down to the nearest whole number.

## Water Prices

A typical bill that households receive contains the following components: water volume, sewer volume, stormwater off-site, stormwater on-site, Portland Harbor Superfund, and base charge, along with clean river rewards, discounts, and late fees for those applicable. For the purposes of our analysis, we consolidate the prices for the bill into water variable price, sewer variable price, and fixed price. To categorize fixed and variable prices, we note the following:

- Water variable price is the price applied for the consumption of water.

- Sewer price includes the price applied for the consumption of sewer, along with Portland Harbor Superfund, BOD, TSS, and cleanriver discount that depends on the sewer consumption level.

  - For residential accounts, the water used in the winter months (reads from $2/1 - 4/30$ for quarterly accounts, $12/1 - 4/30$ for monthly accounts) is used to set what is called the winter average. If the water usage is lower for a given quarter, then the lower value is used to bill for sewer.

  - Customers can fill out a form that then determines the amount of the on-site stormwater that is discountable. It could be up to 100%, depending on how stormwater is managed at a property.

- Fixed price includes all other prices, including base price, fixed portion of the Harbor Superfund, etc., that only depend on the area designated by Portland and not on the water/sewer consumption level.

---

[1] ACCOUNT_CLASS_DFLT, CYCLE_CD

[2] BILL_RUN_DT, PERIOD_FROM_DT, PERIOD_TO_DT, DUE_DT

[3] CANCELED_BILL_YN, ERROR_YN, AUDIT_OR_LIVE, CORRECTED_BILL_YN, OFF_CYCLE_YN, BILL_TP, SOURCE_CD

[4] PREV_BILL_AMT, TOTAL_PAYMENTS, AR_DUE_BEFORE_BILL, AR_DUE_AFTER_BILL

[5] BILL_PRINT_CD, REPORT_CONTEXT, BC_DETAIL_PRORATED_YN, BC_ACTIVE_DAYS, BC_STANDARD_DAYS

[6] BC_DETAIL_AMT, CONS_LEVEL_AMT

[7] HOUSE_NO, STREET_PFX_DIR, STREET_NM, STREET_NM_SFX, CITY, PROVINCE_CD, POSTAL_CODE

[8] BILL_DT, LINC_EFFECTIVE_DATE, LINC_EXPIRY_DATE

[9] WATER_CONS, SEWER_CONS

[10] PENALTY_FEES, PENALTY_FEES_REVERSED

[11] NET_BILL_AMT, BILLED_AMT_BEFORE_DIS, LINC_DISCOUNT_AMT, CRISIS_VOUCHER_AMT

[12] Match to payment_arrangement using PAY_ARRANGEMENT_REF

– There are two components to the Portland Harbor Superfund total: sewer volume and stormwater. Sewer volume is ccf billed for sewer times $.12. On the stormwater, it is $.36 per 1,000 square feet of impervious area (IA). Standard residential IA is 2,400 square feet.

## Penalty for Delinquency

In addition, a 1% penalty is only charged on the current unpaid bill, not a cumulative total of the overall balance in arrears. Pre-pandemic, a residential single-family account was eligible for shutoff if they owed $115 or more and the amount was 56 days past due on their bill. The timelines are shorter for returned items and broken payment arrangements. The threshold balance was set at $1,000, and moved down to $500 in the summer of 2023. The accounts enrolled in financial assistance were not eligible for the shutoff, which changed in 2023 summer when the city lowered the threshold to $500. Payments are applied to the oldest aged receivables first and split proportionately across the different charges (water, sewer storm).

## Frequency of Billing

Portland allows quarterly, monthly, and bi-monthly billing frequencies. Most residential single-family households are on a quarterly basis, with less than 1% on monthly or bi-monthly bills.

## Merging and Filtering

This section describes how we merge and filter the raw data.

We begin with the bill info dataset, which consists of bills generated by the city from 2019 to 2023. We focus on residential accounts that use both water and sewer and are eligible for shutoff in case of delinquency. This reduces to residential single-family accounts, as multi-family units such as condos are not eligible for shutoff.

As presented in Table 1, most of the data have an account as the identifier. We primarily use the account number to merge the datasets and use person and location numbers to complement the merge. Any data that follow bills (`bill_info`, `financial_info`, `usage_info`, etc.) contain the bill generation date, which is used to merge into the account-bill level.

## Valid Bills

We only keep bills that are not canceled, sent in error, audited, or corrected. We also only keep those that have valid date variables that can be matched to relevant usage and financial data using the bill-generated date. We then drop off any anomalies (about 0.5% of the monthly bills that do not occur in a natural subsequent order).

**Census Data**

Portland Water Bureau serves Multnomah, Washington, and Clackamas counties in Oregon. For the Census tracts in these counties, we obtain total number of households, household size, unemployment rate, average income, average earnings, average retirement income, average SSI, average cash assistance, average food stamp amount, share of those under the poverty line, and share of races from 2021 American Community Survey. We link the location number to each account using `location_relation` data and merge in Census tract information using the geocoded Census tract number from `address_info` that link location numbers to the Census tract. We drop bills that do not have a matching location number.

After merging the location data, we add in various program participation information (payment arrangements, financial assistance) and flag whether the bill took place when they were enrolled in the program. Similarly, we add in cutoff/reconnect status.

The merged dataset is saved as `portland_panel.RData`.

## Aggregation

Some households signed up for a payment plan, which allowed them to pay a quarterly bill over roughly three equally split bills each month. As we want our panel to be at the account-quarter level, we need to aggregate these monthly bills into quarterly bills again.

A typical quarterly bill in the data looks like the following:

Table 2: Typical Bill

| previous_bill | total_payments | leftover_debt | current_bill | BILL_TP | SOURCE_CD |
|---|---|---|---|---|---|
| PREV | PREV_PAYMENTS | PREV−PREV_PAYMENTS≡PREV_DEBT | TOTAL_BILL+PREV_DEBT | REGLR | |
| TOTAL_BILL+PREV_DEBT | TOTAL_PAYMENTS | TOTAL_BILL+PREV_DEBT−TOTAL_PAYMENTS≡TOTAL_DEBT | NEW_BILL+TOTAL_DEBT | REGLR | |

When an account signs up for a monthly payment plan, their bills in the data would look like the following:

Table 3: Bill on Monthly Payment Plan

| previous_bill | total_payments | leftover_debt | current_bill | BILL_TP | SOURCE_CD |
|---|---|---|---|---|---|
| PREV | PREV_PAYMENTS | PREV−PREV_PAYMENTS≡PREV_DEBT | QB1_BILL+PREV_DEBT | MSTMT | QB1 |
| QB1_BILL+PREV_DEBT+QB2_BILL+QB3_BILL | QB1_PAYMENTS | QB1_BILL+PREV_DEBT−QB1_PAYMENTS≡QB1_DEBT | QB2_BILL+QB1_DEBT | MSTMT | QB2 |
| QB2_BILL+QB1_DEBT+QB3_BILL | QB2_PAYMENTS | QB2_BILL+QB1_DEBT−QB2_PAYMENTS≡QB2_DEBT | QB3_BILL+QB2_DEBT | MSTMT | QB3 |
| QB3_BILL+QB2_DEBT | QB3_PAYMENTS | QB3_BILL+QB2_DEBT−QB3_PAYMENTS≡QB3_DEBT | NEW_BILL+QB3_DEBT | REGLR | |

where `BILL_TP` of `MSTMT` refers to the monthly bills and `SOURCE_CD` specifies the corresponding "month" of the split bills.

Therefore, the goal is to aggregate this back to the typical quarterly bill as shown above. A typical "aggregated" bill should look like the typical bill above.

We do so by noticing that `previous_bill`, `total_payments`, and `leftover_debt` remain the same for `QB1` and that `QB3_DEBT` effectively summarizes all the debt accumulated throughout

6

the monthly bills. Let us begin with the first row of the typical quarterly bill. If we replace `current_bill` for `QB1` by `previous_bill` for `QB2`, we now have the row for `QB1` equivalent to the first row in the typical quarterly bill.

Now, to the second row. We first replace the `previous_bill` for the `REGLR` bill with the `previous_bill` for `QB2`. We then replace `total_payments` with:

$$\texttt{QB1\_PAYMENTS} + \texttt{QB2\_PAYMENS} + \texttt{QB3\_PAYMENTS} \equiv \texttt{TOTAL\_PAYMENTS}$$

Notice now that:

$$\texttt{TOTAL\_BILL} \equiv \texttt{QB1\_BILL} + \texttt{QB2\_BILL} + \texttt{QB3\_BILL}$$
$$\Rightarrow \texttt{QB3\_DEBT} = \texttt{QB3\_BILL} + \texttt{QB2\_DEBT} - \texttt{QB3\_PAYMENTS}$$
$$= \texttt{QB3\_BILL} + (\texttt{QB2\_BILL} + \texttt{QB1\_DEBT} - \texttt{QB2\_PAYMENTS}) - \texttt{QB3\_PAYMENTS}$$
$$= \texttt{QB3\_BILL} + \texttt{QB2\_BILL}$$
$$\quad + (\texttt{QB1\_BILL} + \texttt{PREV\_DEBT} - \texttt{QB1\_PAYMENTS})$$
$$\quad - \texttt{QB2\_PAYMENTS} - \texttt{QB3\_PAYMENTS}$$
$$= \texttt{QB1\_BILL} + \texttt{PREV\_DEBT} + \texttt{QB2\_BILL} + \texttt{QB3\_BILL}$$
$$\quad - (\texttt{QB1\_PAYMENTS} + \texttt{QB2\_PAYMENTS} + \texttt{QB2\_PAYMENTS})$$
$$= \texttt{TOTAL\_BILL} + \texttt{PREV\_DEBT} - \texttt{TOTAL\_PAYMENTS}$$

So we do not have to change `leftover_debt` and `current_bill` for the last row. We then remove `QB2` and `QB3` bills and have completed the aggregation.

Sometimes, final bills occur before the complete cycle of monthly bills occurs (e.g., `FINAL` occurs immediately after `QB2`). In some cases, another monthly cycle occurs immediately after a monthly cycle (e.g., QB1 happens immediately after QB3). These cases are aggregated similarly to the typical case mentioned above.

All the other usage or components of the bills (financial assistance amount, etc.) could simply be aggregated and filled into the first month because that is when they were applied.

We aggregate any flags (participation in payment arrangement, financial assistance, cutoff) and indicate if any of the bills were flagged.

Once the aggregation is completed, we add in flag `AGG` to indicate that the bill has been aggregated and `CHOP` on any remaining monthly bills that could not be aggregated as they are still in place at the end of the data scraping period.

The resulting dataset is saved as `portland_panel_estimation.RData` and used for the static demand model estimation.

## Flags and How to Use Them

This section discusses various flags in the final panel data where they are relevant for the potential dropping of the observations.

### Bill Type (BILL_TP)

While the final bill is already flagged in the raw data (BILL_TP is FINAL), we do not have a raw indicator of the first bill. Therefore, we flag the first bill when the bill a) first appeared in the raw data, b) no previous bill associated with it, and c) no payment was made.

Sometimes, the owner reoccupies the property after leasing it for a while, or a household returns to the property after living in another place. This comes up in the data as a bill reappearing after the final bill. We flag these as RESUME. Similarly, when the first appeared in the bill data with no previous bill but made non-zero total payments, we flag it as RESUME, assuming that the payment must have been from a previous bill[13].

### Seniors and Disabled (senior_disabilities)

When households sign up for financial assistance programs, they may provide evidence that they belong to the seniors or disabled group. While the regular financial assistance recipients have an expiry date and have to re-enroll after expiration, seniors and disabled do not have an expiry date for their discounts. In data, this is done by assigning the expiry date in 2065 or beyond.

### Proposal for Dropping Observations

From the estimation sample, we propose the following additional "drops" for the bills:

- First bill (BILL_TP is FIRST)

  - First bills do not have any previous bills, and they are off-cycle.

- Last bill (BILL_TP is FINAL)

  - Final bills are off-cycle, and they often do not get sent to any collection activities.
  - There is an incentive to keep the final bill below the threshold to avoid collection agencies. However, the shutoff threat does not work for the final bills.

- Resumed bill (BILL_TP is RESUME

  - These are essentially FIRST bills, except they have previous bill amounts (which was the previous FINAL bill). The issue is that the previous bill was off-cycle. Now, a major complication is that there is a large gap before resumption (regardless of whether we drop RESUME or not).

---

[13]This may not always be the case. For instance, an owner may assume a tenant's unpaid bill. This is up as a remaining task.

- Incomplete monthly payment cycle (`agg` is `CHOP`)

    - These bills are where aggregation failed because the data scraping period ended before a quarter had passed. They do not have any usage information attached as they were not aggregated back to the quarterly level.

- Monthly or bi-monthly billing cycle (`CYCLE_CD` is not `QUARTER`)

    - Note that this is different from the monthly payment plan. Those who had a monthly billing cycle pay for the past month's consumption, whereas those on a monthly payment plan pay (roughly) a third of the past quarter's consumption.

    - There was a very small subset of non-quarterly accounts: 1253 monthly and 246 bi-monthly bills in the dataset (0.04% and $\leq 0.01\%$, respectively).

Additionally, we propose the following flags to pay close attention but not use them as drop observations:

- Seniors and disabled (`senior_disabilities` is `TRUE`)

    - They are not part of the price experiments but represent an important subgroup.

## Resulting Panel Data

The table below lists the variables in the final estimation panel dataset:

## Table 4: Variable List in `portland_panel_estimation.RData`

| Variable Name | Description | Type | Values |
|---|---|---|---|
| ACCOUNT_NO | account number | character | - |
| PERSON_NO | person number | numeric | - |
| LOCATION_NO | location number | numeric | - |
| OCCUPANCY | occupancy status[14] | character | OWNER, TENANT, NA |
| tract | Census tract code[15] | character | - |
| BILL_RUN_DT | date when bill was generated | date | - |
| bill_year | year for BILL_RUN_DT | integer | - |
| PERIOD_FROM_DT | start date of the billing period | date | - |
| PERIOD_TO_DT | end date of the billing period | date | - |
| DUE_DT | bill due date | date | - |
| SOURCE_CD | month number for monthly payment plans | character | QB1, QB2, QB3 |
| BILL_TP | bill type (first, regular, final, resume, monthly) | character | FIRST, REGLR, FINAL, RESUME, MSTMT |
| OFF_CYCLE_YN | indicator if bill is off the regular billing cycle | logical | - |
| CYCLE_CD | billing cycle code | character | QUARTER, MONTH, BIMONTH |
| previous_bill | amount of previous bill | numeric | - |
| total_payments | total payments made | numeric | - |
| leftover_debt | leftover debt (= previous_bill-total_payments) | numeric | - |
| current_bill | current bill | numeric | - |
| writeoff_amount | amount written off (officially) | numeric | - |
| final_writeoff | amount implicitly written off from uncollected final bill | numeric | - |
| collection_sent_amount | amount sent to collection agency | numeric | - |
| collection_collected_amount | amount collected by collection agency | numeric | - |
| delinquent | whether previous bill went delinquent (= leftover_debt $\geq 0$) | logical | - |
| usage_bill_amount | newly generated usage bill (absent arrears) | numeric | - |
| usage_bill_water_cons | water portion of the usage bill | numeric | - |
| usage_bill_sewer_cons | sewer portion of the usage bill | numeric | - |
| bill_penalty | fee for penalty | numeric | - |
| bill_donate | amount of donation | numeric | - |
| bill_bankrupt | amount bankrupted | numeric | - |
| bill_leaf | special fee for leaf removal | numeric | - |
| price_water | rates for water (variable) | numeric | - |
| price_sewer | rates for sewer[16] (variable) | numeric | - |
| price_fixed | fixed price | numeric | - |
| price_donation | rates for donation | numeric | - |
| price_discount | rates applied for discount (financial assistance) | numeric | - |
| water_cons | water consumption (in ccf) | numeric | - |
| sewer_cons | sewer consumption (in ccf) | numeric | - |
| payment_arrange | indicator if in payment arrangement | logical | - |
| payment_arrange_status | indicator if the payment arrangement is in good standing[17] | logical | - |
| financial_assist | indicator if receiving financial assistance | logical | - |
| cutoff | indicator if currently experiencing water cutoff | logical | - |
| LINC_TIER_TYPE | financial assistance (LINC) tier[18] | character | Tier 1, Tier 2 |
| net_after_assistance | net bill after discount is applied | numeric | - |
| bill_before_assistance | bill before discount is applied | numeric | - |
| discount_assistance | amount of discount applied | numeric | - |
| crisis_voucher_amount | amount of crisis voucher (separate from discount) | numeric | - |
| senior_disabilities | indicator if seniors or disabled[19] | logical | - |
| agg | status of the aggregation[20] | character | AGG, CHOP |

**Remaining Work**

- Track "transferred" bills (either by PERSON_NO or LOCATION_NO)

- Update panel dataset with the new financial_info dataset

- Add any payments to the final bill from the financial_info dataset

---

[14]Not all accounts have this information. We cannot simply assume that those missing this information are either owners or tenants, so we leave it blank.

[15]Census tract code is an 11-digit code consisting of 2-digit state, 3-digit county, and 6-digit tract code. Even though the Portland Water Bureau's service area covers 3 counties, the last 6 digits uniquely identify the tracts, and so the raw data reports those.

[16]While this is a variable price, it does not directly get multiplied to the amount of sewer used in that particular quarter. The consumption level used to charge the sewer portion of the bill is from the consumption level during the winter quarter.

[17]It should be noted that the majority of the payment arrangements were terminated for broken terms.

[18]Tier 1 is eligible for those below 60% of the average monthly income and Tier 2 for 30%.

[19]Seniors and disabled were part of the financial assistance program. Because the program was opt-in, we do not observe whether someone belongs to the seniors/disabled group unless they signed up for the program.

[20]CHOP refers to those who could not be aggregated because the data scraping period ended before the completion of the full monthly payment cycle.