

This document provides a description of the codes and data used to generate the results in “The Value of Free Water: Analyzing South Africa’s Free Basic Water Policy” by Andrea Szabo

I. Data

All files referred to below are in the Data folder.

Survey data

survey.dta: contains the household data from the 2010 survey. Variables are labeled in the Stata file and variable names correspond to the survey question numbers in Mabopane_Questionnaire.pdf

The survey data was collected as part of the project “Sustainable Water Pricing,” approved by the University of Houston IRB under protocol number 10275-02 on May 10, 2010.

surveymatlab.do creates the final demographic variables used by the maximum likelihood estimation. Run this file on survey.dta

Weather data

weather.dta: the variable maxt records average maximum daily temperature in Fahrenheit. Observation weighted averages are reported in Table 1.

Source: <http://www.wunderground.com/history/airport/FAJS/2001/4/1/MonthlyHistory.html>

Tariff data

tariff.dta contains the 20 different price schedules (in 2008 Rand), see footnote 9 in the paper.

Table 4 in the Online Appendix shows the observation weighted averages.

Variable names:

tyear: tariff year. The first tariff year is from 2001 July to 2002 June. The next tariff year always starts in July.

rich: 0: no sanitation, 1: sanitation

ind: 0: non-indigent, 1: indigent

p1-p8: corresponding marginal prices

w1-w7: corresponding kink points

Consumption billing data

The billing data is confidential and not provided. I obtained access to it by personally visiting the office of the Chief Financial Officer of Odi Water (Odi Water Services, Lucas Mangope Highway, Opposite Morula Sun, Mabopane 0190, Gauteng, South Africa). The data was downloaded from Odi’s servers in January 2009.

II. Estimation

All files referred to below are in the Programs folder.

Summary statistics

Table 1, Panel B variables and Table 8 results are generated by `surveymatlab.do`. Run on `survey.dta`.

Regression analysis

Table 2 results are generated by `regression.do`

Maximum likelihood estimation

`main.m` produces the parameter estimates and standard errors in Appendix Table 9 using the following files

- `likelihood.m`, the likelihood function, see Appendix 9.3, which in turn uses the following files in creating the bounds for η (Proposition 1):

`compare1.m`

`compare2.m`

`compare3.m`

`condition.m`

`difference.m`

`indirect.m`

`kinksegment.m`

`segmentkink.m`

`uiij.m`

- `standarderrors.m`, used in computing the standard errors using the score and Hessian of the likelihood function (see end of Section 5.2).

Expected consumption

`expected_consumption.m` computes expected consumption predicted by the estimated model, as described in Appendix 9.4

Marginal effects

`marginale.m` computes the marginal effects and corresponding standard errors in Table 9, column 3 and 4, after the parameters and the expected consumption are computed.

Price elasticities

`priceelasticity.m` computes the price elasticities in Table 3 after the parameters and the expected consumption are computed.

Model performance

expected_consumption.m computes the average error of the model

modelperformance.m computes the average error for consumer groups in Table 4.

III. Counterfactual analysis without free water

counterfactual1.m computes the consumption after removing the 0 prices. The positive prices are based on the administrative data as explained in the paper. Run the file after the parameters and the expected consumption are computed. This file creates the consumption results in Table 5.

counterfactual2.m computes several other versions of this counterfactual as discussed at the end of Section 7.1. Each counterfactual generates a data file which can be used to compute expenditure and other summary statistics.

fbw.do computes summary statistics of counterfactual consumption. It also computes expenditure under the counterfactuals. Run on the xls files generated by counterfactual1.m and counterfactual2.m (running this on a.xls creates the expenditure results in Table 5).

IV. Social planner problem, optimal tariffs

Optimal tariff

optimalprice.m is the main file which reads in the data and starts the optimization routine for all of the following cases.

Results in paper	Files used	Sample
Figure 5, optimal	constraint7.m sum7.m	All (indigent and non-indigent) for tariff year 2006/07
Figure 5, optimal with free water	constraint7i.m sum7i.m	All (indigent and non-indigent) for tariff year 2006/07
Figure 6, left panel, optimal and Figure 6, right panel, optimal	constraint7.m sum7.m	All (indigent and non-indigent) for tariff year 2007/08
Figure 6, left panel, optimal with free water and Figure 6, right panel, optimal with free water	constraint7i.m sum7i.m	All (indigent and non-indigent) for tariff year 2007/08
Figure 7, left panel, optimal	constraint7.m sum7.m	Only non-indigent for tariff year 2007/08
Figure 7, right panel, optimal	constraint7.m sum7.m	Only indigent for tariff year 2007/08
Figure 8, left panel, optimal and Figure 8, right panel optimal	sum7.m constraint.m	All (indigent and non-indigent) for tariff year 2007/08
Figure 8, left panel, optimal with free water and Figure 8, right panel optimal with free water	sum7i.m constrainti.m	All (indigent and non-indigent) for tariff year 2007/08

All corresponding optimal tariffs in the Online Appendix can be computed based on the above table by changing the marginal cost from 8.2 to 4 as indicated in the code.

The above codes use 2 additional inputs computed elsewhere.

1. Total revenue for the sample described in column 3: use current.m
2. Total consumption for the sample described in column 3: directly from the data

Compensating variation

compensating_variation.m is the main file which reads in the data and starts the optimization routine for all of the following cases which appear in Table 6 and 7 in the paper.

Results in paper	Files used	Sample
Table 6, compensating variation, column 1	compensated2.m	All (indigent and non-indigent) for tariff year 2006/07
Table 6, compensating variation, column 2	compensated2i.m	All (indigent and non-indigent) for tariff year 2006/07
Table 6, compensating variation, column 3	compensated2.m	All (indigent and non-indigent) for tariff year 2007/08
Table 6, compensating variation, column 4	compensated2i.m	All (indigent and non-indigent) for tariff year 2007/08
Table 7, compensating variation, column 1, row 2	compensated2.m	Only non-indigent for tariff year 2007/08
Table 7, compensating variation, column 1, row 3	compensated2.m	Only indigent for tariff year 2007/08
Table 7, compensating variation, column 2	compensated2.m	All (indigent and non-indigent) for tariff year 2007/08
Table 7, compensating variation, column 3	compensated2i.m	All (indigent and non-indigent) for tariff year 2007/08

The above codes use the baseline utility (with observed consumption) as an input. Use baseline_utility.m to get this. Variable H (baseline utility) is read in by compensating_variation.m. Do this separately for the samples listed in the third column in the table above.