

# Data and code addendum

## Estimating Willingness to Pay for E85 in the United States Using an Intercept Survey of Flex Motorists

*Sebastien Pouliot*

*Kenneth Liao*

*Bruce Babcock*

*April 15, 2018*

This document explain the code to replicate the results in *Estimating Willingness to Pay for E85 in the United States Using an Intercept Survey of Flex Motorists* published in the *American Journal of Agricultural Economics*. The article provide the necessary information about data collection and thus we will not describe them in this document.

### Software requirements and folder structure

We generated figures and estimated regression models using R, which is available to download for free at <https://cran.r-project.org/>. We recommend using RStudio as an IDE. RStudio is available to download for free at <https://www.rstudio.com/> and one of its advantage is its ability to use an *R Project* which facilitates setting working directories and organizing files. Using RStudio, we recommend to first open the project file `Data and code.RProj`. Opening this file sets the working directory so that all code should run without issues related to file location.

The R files to replicate the results in the paper are in the main folder. There are four subfolders to help organize other files.

- **Data:** This subfolder contains the data. The file `SP-off-RP data.rds` contains the data in R format, and is the file data that is called by other files. The subfolder also contain the excel `SP-off-RP data.xlsx` for users interested in reading the data in a different format.
- **Figures:** Figures are saved as `.png` files in this subfolder.
- **Results:** Estimation results are saved in this subfolder as excel file. Results are also saved as `.RData` file for later use.
- **Code:** This subfolder contains R files that are sourced by other files. These R files contain functions to estimate models and produce figures. We describe these files in a section below.

The R code requires several packages. All files that require a package have at their beginning a line that checks whether the package *pacman* is already installed and if not, then the package is automatically installed. Then, other packages are loaded using the function `p_load` from the *pacman* package, which will install a package, if necessary, and then load it. The

first time, you might have to run the code twice for the packages to be installed and loaded.

It is possible to run code by making use of the `.Rmd` file that produces this text. To do so, in the R code chunks, set `eval=TRUE`. Setting `include=TRUE` and `echo=TRUE` will include the R code and the R output in the pdf compiled using the `.Rmd` file. You can compile the `.Rmd` file, click the *Knit* button in RStudio. However, this will produce only partial results because the R files require selecting either the sample with 479 observations or the sample with 670 observations as explained below.

Estimating the regression models takes several hours and even several days on most machine if running files one by one. One way to speed up computation is to open several instances of R or RStudio and running multiple files at the same time. The code does not use parallel processing as it did not lead to major speed improvement. The part of the code that takes the most time is the numerical calculation of the Jacobian to calculate robust standard errors. This could be improved by specifying the Jacobian analytically but we chose to use numerical calculations that are less prone to coding errors.

You should not expect to obtain the exact same results as those reported in the paper, but they should be closed. We found that slightly different results depending on which computer we estimated the models.

## **R files to replicate results in the paper**

We explain below the files necessary to replicate the results in the paper. The data are anonymized so that the name and the address of the fuel stations are not included. We do not provide code to show how the data are assembled as much of that work was done using Microsoft Excel.

### **File 1) Simulations.R**

This file performs simulations and produces Figure 1 that shows the self-selection self-selection in our survey. The figure is saved in the Figures subfolder. The simulations do not use the data we collected.

### **File 2) Summary.R**

This file prepares data summary that we show in Tables 1, 2, and 3. The data summaries are not saved into excel files but rather displayed in the console.

The file also produces Figures D.1 and D.2 in appendix, saved in the Figures subfolder.

### **File 3) Figures.R**

This file generates Figures 2 and 3, saved in the Figures subfolder. It also includes additional code to generate alternative figures for the premium for E85 (i.e. price difference between E85 and E10).

#### **File 4) Regressions ~ ratio.R**

This is the file that generates the main results in the paper. It produces the results we show in Tables 4, 5, and 6. The file must be run twice, selecting either for 479 or 670 observations on lines 50 and 51. The file sources functions from the files `func_logit_solver_deltamethod.R` and `Func_SP_WMLE_logit.R` in the Code subfolder. The tables of results are saved in the Results subfolder.

#### **File 5) Graph demand curves.R**

This file uses regression results to produce Figure 4, saved in the Figures subfolder. The R code sources functions from the file `Demand_graph_func.R` in the Code subfolder.

### **R files to replicate results in appendix**

The appendix contains additional information about the survey and the data, alternative regression models for the E85 premium and marginal effects for regressions where we used station fixed effects. We list the files to replicate these results below.

#### **File B) Summary.R**

This file produces the data summaries in appendix Tables B.1, B.2 and B.3.

#### **File C) Regressions ~ premium.R**

This file produces appendix Tables C.1 and C.2 that consider alternative regression models where motorists make their fueling decisions based on the premium for E85. The file must be run twice, selecting for 479 or 670 observations on lines 50 and 51.

#### **File C) Graphs of demand curve ~ premium.R**

This file uses regression results for the premium model to produce Figure C.a, saved in the Figures subfolder. The R code sources functions from the file `Demand_graph_func.R` in the Code subfolder.

#### **File E) Regressions ~ station fixed effect.R**

This file estimate regression models where we use station fixed effects instead of state fixed effects. It produces appendix Tables E.1 and E.2. The file also produces tables for the estimated coefficients that are not shown in the appendix. The file must be run twice, selecting for 479 or 670 observations on lines 50 and 51.

### **R files in the *Code* subfolder**

Several of the files above use code in the *Code* subfolder. This subfolder contains four files that we briefly describe below.

**File Demand graph func.R**

The file contains functions to make Figure 4. In particular, it contains code to plot the standard errors around the demand curves in Figure 4.

**File func\_logit\_solver\_deltamethod.R**

The file contains functions to summarize WTP based on estimated regression coefficients. For instance, it is used to compute estimated WTP distribution parameters that we report in Table 6. The standard errors are calculated using the delta method.

**File Func\_SP\_WMLE\_logit.R**

The file contains custom functions to estimate models that use RP data only and functions to estimate models that use SP-off-RP data. The code calculates robust standard errors and that is what significantly slows down the estimation.

**File Graph parameters.R**

This files contains parameters (i.e. theme) for making plots using the package ggplot2. We make inconsistent use of this file when generating figures and sometimes overwrite some of the the parameters. Some of the font choices in this file may cause incompatibilities across computers that can lead to warning messages when generating figures.