

Data description and programs for “Can Self-Set Goals Encourage Resource Conservation? Field Experimental Evidence from a Smartphone App”

Data access: The data of our paper comes from a field experiment with a large utility in Germany (<https://www.stadtwerke-muenster.de>). In cooperation with the utility and an IT company, we developed the ‘energy app.’ The energy app was integrated into the pre-existing ‘münster:app,’ which is owned by the utility and hosted by the IT company. We collected our data via the energy app.

A non-disclosure agreement with the utility does not allow us to make the data available to the general public. However, researchers may also send us their STATA code and receive the generated log file within one week. A copy of the data set is permanently stored at our institutions. Alternatively, researchers may reach out to our cooperation partner and ask for the data.

Data sets:

app_data.dta: We observe an unbalanced panel data set. The unit of observation are electricity meter readings of app users. Each user could submit up to five meter readings.

Explanation of main variables:

- *userid*: ID of app user
- *submittedday*: Date of when the user was conducting the meter reading.
- *number*: Numbering of the meter readings, sorted by their submitted date. I.e., *number*=1 refers to the first meter reading provided by the user. *number* can take values from 1 to 5.
- *scan*: Numbering of the meter readings, organized by the experimental periods. I.e., *scan*=3 refers to the meter reading 3 starting experimental period 3 (cf. Figure 2). *scan* deviates from *number* if a user does not provide one of the meter readings. E.g., if a user does not submit the meter reading 2 according to Figure 2, meter reading 3 will be recognized by *scan*=3 while *number*=2. *scan* can take values from 1 to 5.
- *E1*: Indicator for the first experimental period (i.e., the baseline period). We likewise defined the indicators *E2*, *E3* and *E4* for experimental periods 2-4 (cf. Figure 2). We match experimental periods to meter readings (i.e., the rows of our data) by assigning each meter reading the scan value of the prior scan as experimental period. E.g., *E1*=1 is assigned to meter reading 2 (since the value of the prior scan is *scan*=1). If meter reading 2 is missing (i.e., there is no *scan*=2), *E1*=1 is assigned to meter reading 3. The indicators *E1*-*E4* are the outcome variables of the extensive margin analysis.
- *treatment*: Indicator for the user being in the treatment/goal prompt group. Equals 0 for the control group.
- *incentive*: Indicator for the user receiving the savings subsidy treatment in experimental period 4. Equals 0 if the user was not randomized into receiving the savings subsidy.
- *treatment_inc*: Interaction of being in the goal prompt and the savings subsidy treatment.
- *checked*: Contains the meter value as verified by our RAs. Checked is missing if the meter value could not be verified.
- *normconsumption_checked*: Difference in verified meter values between the subsequent rows. The difference between meter values give the electricity consumption. Since the time distance between meter readings might vary across participants (e.g., in case of missed meter readings) we normalize consumption to reflect a 28-day period.
- *normconsumption_baseline_checked*: Difference in checked meter values between the second (i.e., *number*=2) and first (i.e., *number*=1) meter reading. This difference always reflects the baseline period. To adjust for slight time differences between the meter readings we again normalize consumption to a 28-day period.

- *log_kwh*: The logarithm of *normconsumption_checked*. This is the outcome variable for the intensive margin analysis.
- *goal*: The goal set by the user. *goal* is missing for control group users. *goal* is assigned to the row of the experimental period it applies to. It thereby differs from *goal_help*, which shows each goal in the row of the experimental period at which the goal is set.
- *relative_goal*: The goal relative to baseline consumption. I.e., *goal* divided by *normconsumption_baseline_checked*.

survey_data.dta: After meter reading 5 users are redirected to the survey. The unit of observation is the app user.

Explanation of main variables:

- *beta*: Present focus parameter. Calculated as described in Section 3.3. Based on responses to questions 59 and 60 of the survey.
- *loss_aversion*: Loss aversion parameter. Calculated as described in Section 3.3. Based on responses to questions 28-58 of the survey.
- *var*: Difference between the maximum and minimum electricity price (in EUR cents) respondents think they pay. I.e., the difference between responses to survey questions 27 and 26.
- *aim_to_save*: Response to first item of survey question 11 ("I regularly intent to use less energy"). Responses are coded from 0 "never" to 3 "always."
- *do_not_save*: Response to second item of survey question 11 ("But then, I frequently consume more electricity than intended."). Responses are coded from 0 "never" to 3 "always."
- *loss_feeling*: Response to survey question 12 ("How do you feel in the following situation: My electricity costs are 100 EUR higher than expected."). Responses are coded from -3 "very bad" to 3 "very good."
- *gain_feeling*: Response to survey question 13 ("How do you feel in the following situation: My electricity costs are 100 EUR lower than expected."). Responses are coded from -3 "very bad" to 3 "very good."
- *female*: Indicator for being female. Response to survey question 19.
- *age*: Respondents' age. Response to survey question 20 (open-end response).
- *abitur*: Indicator for having obtained university qualification. Equals 1 if survey question 22 was answered with "Graduated from high school or otherwise obtained a university qualification" and 0 otherwise.
- *working*: Indicator for the German job categories "Auszubildende/-r", "Arbeiter/-in", "Angestellte/-r", "Beamter/Beamtin" or "Selbstständiger" as surveyed in question 24.
- *income_new*: Respondents' personal monthly net income. *income_new* alters the response to the income categories of survey question 21 (saved in variable *income*) by taking the midpoint of the selected income bracket.
- *dwelling_size*: Square meters of the dwelling the participant lives in. Response to survey question 5 (open-end response).
- *household_size*: Number of persons in the household, including the participant. Response to survey question 2, can take values from 1-7.

Programs:

Replication files.do: We used STATA for all computations and provide the do-file used for analysis for replication. After preparing the data and naming the variables according to the names used in this script, the do-file can be run to reproduce all results.