README and guidance

1 Overview

This document describes the replication files for "Congestion pricing with electric vehicle exemptions: Car-ownership effects and other behavioral adjustments" by Elisabeth T. Isaksen and Bjørn G. Johansen for the Journal of Environmental Economics and Management (JEEM).

The replication package constructs the figures and tables in the paper as well as the online appendix from a number of data sources using Stata 17 or 18.

The replication package consists of two parts. Part 1 contains the analysis based on Norwegian administrative data. All the do-files based on these data are group together. This analysis uses proprietary data made available for research, under specific and strict regulations, by Statistics Norway, and cannot be made publicly available.

Part 2 contains the analysis based on traffic and air pollution data. All the do-files based on these data are group together. The analysis on traffic is mainly based on proprietary data collected from a private road toll company (Ferde). These data cannot be made publicly available. Air pollution and weather data are available via public APIs. The raw data downloaded from the APIs are provided in the replication package.

This README-files documents part 1 of the analysis.

2 Data availability and provenance statements

The information used in part 1 of the analysis combines several Norwegian administrative registers (as described in the paper). The data use is subject to the European Union's General Data Protection Regulation (GDPR) per new Norwegian regulations from May 2018. The data used is on loan from Statistics Norway for a limited time-period, located at a secure server at the Frisch Centre and, due to confidentiality and security considerations, the data cannot be transferred outside the secure server or shared with others. Researchers interested in obtaining access to the register data employed in this paper are required to submit a written application to gain approval from Statistics Norway. The application must include a detailed description of the proposed project, its purpose, and its social contribution, as well as a description of the required datasets, variables, and analysis population. Applications can be submitted by researchers who are affiliated with Norwegian institutions accepted by Statistics Norway, or by researchers outside of Norway who collaborate with researchers affiliated with these institutions. The following link contains more information regarding access to data: https://www.ssb.no/en/data-til-forskning. Any researchers interested in replicating the results of this paper are welcome to contact us for further information.

The location of toll gates, fuel prices, consumer price indexes and changes in geographical classifications over time are publicly available and can be downloaded using the links provided under "data references" (see below).

3 Dataset list – part 1

The replication package presupposes the existence of the following datasets in the `STATA\data\in` folder:

Table 1. Data files used in the analysis.

`Filename` ***Data reference***	**Description of data and variables included in analysis**
`Mreg_g[year].sas` *Source: Statistics Norway. (2018o)*	Raw data from the central motor vehicle register. All passenger cars with private individuals registered as the owner at the end of the year. One file per year, for the period 2005-2017; one row per car.
`Mreg__org_g[year].sas` *Source: Statistics Norway. (2018q)*	Raw data from the central motor vehicle register. All passenger cars with enterprises registered as the owner at the end of the year. One file per year, for the period 2005-2017; one row per car.
`kjl[year].sas` *Source: Statistics Norway. (2018n)*	Raw data from periodic vehicle inspections; all passenger cars. One file per year for the time period 2005-2017; one row per car.
`Househ_wp2.dta` *Source: Statistics Norway (2018a, c, d, f-m, p, 2019).*	A set of individual characteristics from various registers. Level of observations: (individual*year). Variables: Person ID, year, household ID, gender, #persons in household, age, basic statistical unit of residence, education code (NUS2000), income net of taxes, net wealth, employment dummy (labor income above one basic amount), retired dummy, home ownership dummy, second home ownership dummy, household type (composition of household members), number of adults in household, number of children below 18, number of children older than 18 not living with their parents. Sample: All individuals above 18 registered in Norway between 2005 and 2017.
`Utvalg_atmlto.dta` *Source: Statistics Norway (2018a, b, e).*	Basic statistical units ('neighborhoods') of workplaces. Level of observations: (individual*year). Variables: Individual ID, year, workplace ID, basic statistical unit of workplace, labor income. Sample: All individuals' workplaces within Norwegian borders between 2005 and 2017.
`Distances.dta`	Characteristics of trips between basic statistical units ('neighborhoods') in Norway along the

Source: Halse et al (2022).	road network. Level of observations: (basic statistical unit*basic statistical unit). Variables: basic statistical unit of origin, basic statistical unit of destination, distance, distance with bus lane, time, road toll by year. Sample: all origin-destination pairs in Norway where at least one individual commutes.
`Househ_select.dta` *Source: Statistics Norway (2018l).*	A set of all individual IDs and household IDs by year. Used to merge household ID to car ownership data, which has car ownership recorded by individual ID.
`Kollektivtransport_grk` *Source: NPRA (2020).*	A dataset containing level of service data for public transport from the Norwegian transport models. Level of observations: (origin basic statistical unit*destination basic statistical unit). Variables: origin, destination, time spent walking to/from stations, onboard time, waiting time, number of boardings, average cost per trip. Sample: All origin-destination pairs in Norway.
`Grk_link_to_grk_number_in_2014.dta` *Source: Statistics Norway (2020).*	A mapping between basic statistical unit IDs in 2014 and IDs used during subsequent years, in case the basic statistical unit ID has changed.
`Bergen_center.csv` *Source: located at* `\additionaldata\bergen_center.csv`	A list of the basic statistical unit IDs that were inside the toll cordon in Bergen in 2014. The list is created manually, and included as part of the replication package.
`Stavanger_center.csv` *Source: located at* `\additionaldata\stavanger_center.csv`	A list of the basic statistical unit IDs that were inside the toll cordon in Stavanger in 2014. The list is created manually, and included as part of the replication package.
`Haugesund_center.csv` *Source: located at* `\additionaldata\haugesund_center.csv`	A list of the basic statistical unit IDs that were inside the toll cordon in Haugesund in 2014. The list is created manually, and included as part of the replication package.
`Kristiansand_center.csv` *Source: located at* `\additionaldata\kristiansand_center.csv`	A list of the basic statistical unit IDs that were inside the toll cordon in Kristiansand in 2014. The list is created manually, and included as part of the replication package.

See in-line comments in the script `\dofiles\00_masterfile.do` for more detailed information regarding each data file. Sources for each file are listed under “Data references” below.

Data references

Halse, Askill Harkjerr; Johansen, Bjørn Gjerde; Sand, Øystein (2022). Road tolls in Norway. Can be accessed at DOI: https://doi.org/10.18710/M2D2XP (last accessed 2022).

NPRA (2020). Level of service data for public transport from the regional and national Norwegian transport models (`kollektivtransport_grk`). Norwegian Public Roads Administration, Oslo (last accessed: 2020).

Statistics Norway. (2018a). Wage information until 2015 (`atmlto2005- atmlto2005`; 10 separate datasets). Statistics Norway, Oslo/Kongsvinger (last accessed: 2018)

Statistics Norway. (2018b). Firm locations (`bed_for_grk`). Statistics Norway, Oslo/Kongsvinger (last accessed: 2018)

Statistics Norway. (2018c). Constant person characteristics (`befolkning`). Statistics Norway, Oslo/Kongsvinger (last accessed: 2018)

Statistics Norway. (2018d). Variable person characteristics (`befolkning_tp`). Statistics Norway, Oslo/Kongsvinger (last accessed: 2018)

Statistics Norway. (2018e). Firms and enterprises (`bof2005-bof2015`; 11 separate datasets). Statistics Norway, Oslo/Kongsvinger (last accessed: 2018)

Statistics Norway. (2018f). Living conditions (`bofor2015-bofor2016`; 2 separate datasets). Statistics Norway, Oslo/Kongsvinger (last accessed: 2018)

Statistics Norway. (2018g). NUS code for highest ongoing education (`bu_igang`). Statistics Norway, Oslo/Kongsvinger (last accessed: 2018)

Statistics Norway. (2018h). Information regarding completed education (`f_utd_demografi`). Statistics Norway, Oslo/Kongsvinger (last accessed: 2018)

Statistics Norway. (2018i). Marital status (`famsam`). Statistics Norway, Oslo/Kongsvinger (last accessed: 2018)

Statistics Norway. (2018j). Family type (`famtyp`). Statistics Norway, Oslo/Kongsvinger (last accessed: 2018)

Statistics Norway. (2018k). Basic statistical unit of residence (`grunnkrets`). Statistics Norway, Oslo/Kongsvinger (last accessed: 2018)

Statistics Norway. (2018l). Household specific demographics (`hush_fam2005-hush_fam2017`; 13 separate datasets). Statistics Norway, Oslo/Kongsvinger (last accessed: 2018)

Statistics Norway. (2018m). Income from all sources (`inntekt`). Statistics Norway, Oslo/Kongsvinger (last accessed: 2018)

Statistics Norway. (2018n). Odometer readings from EU controls (`kjl2005-kjl2017`; 13 separate datasets). Statistics Norway, Oslo/Kongsvinger (last accessed: 2018)

Statistics Norway. (2018o). The central motor vehicle register, cars registered at private individuals (`mreg_g2005-mreg_g2017`; 13 separate datasets). Statistics Norway, Oslo/Kongsvinger (last accessed: 2018)

Statistics Norway. (2018p). Wealth and debt (`skattestat`). Statistics Norway, Oslo/Kongsvinger (last accessed: 2018)

Statistics Norway. (2018q). The central motor vehicle register, cars registered at enterprises (`mreg_org_g2005-mreg_org_g2017`; 13 separate datasets). Statistics Norway, Oslo/Kongsvinger (last accessed: 2018)

Statistics Norway. (2019). Ownership of secondary home (`sek_fritid_g2012-sek_fritid_g2012`; 5 separate datasets). Statistics Norway, Oslo/Kongsvinger (last accessed: 2018)

Statistics Norway (2020). Changes in regional classifications over time. Downloaded from Statistics Norway (available at: https://www.ssb.no/metadata/alle-endringer-i-de-regionale-inndelingene). Statistics Norway, Oslo/Kongsvinger (last accessed: 2020)

4 Computational requirements

4.1 Hardware requirements

Estimations were done on a Windows 10 virtual machine.

- CPU speed: 2.10 GHz, 40 virtual processors
- Installed memory (RAM): 160 GB

We do not recommend attempting to replicate the results without access to a relatively fast computer/server, due to the computationally intensive regressions.

The largest .dta dataset that is read into memory is approximately 10 GB. The full list of data under "data references" will take up approximately 100 GB of hard drive space.

4.2 Software requirements

All code is run using Stata (code was last run using Stata 18, but will work using Stata 17 as well).

The following .ado-programs from SSC are used:

- estout
- reghdfe
- spmaps

These packages can be installed by uncommenting the lines 15-17 in the script `\dofiles\00_masterfile.do`.

4.3 Runtime requirements

Given the aforementioned hardware, total run time (including results for the online appendix) is approximately 12 hours.

5 Description of programs/code

The replication package requires the user to set the path to the following folders (this is done in `00_masterfile.do` at lines 5-13):

- `Rawdata`
- `Additionaldata`

- Newdata
- Dofiles
- Figures
- Tables
- Ster

All code is located in the folder \dofiles. This code assumes that the folders \rawdata and \additionaldata contains all datasets as described in Table 1. These are the only folders that are required to have any content before the code is run – all other folders are used for storing results created by the programs. The folder \newdata will contain new datasets created by the code. The folder \ster will contain .ster-files that store regression results. The folder \figures will store figures in .pdf-format, and the folder \tables will store summary statistics and regression results in the form of .tex-files (to be read in by a Latex compiler). Note that the most detailed documentation is in the form of in-line comments in the do-files themselves.

5.1 Description of code

DATA PREPARATION:

All tables and figures are constructed using a set of STATA programs. The results are reproduced by running the master do-file 00_masterfile.do located in the \dofiles folder. Provided that it is set up correctly (see the section "Instructions" below) it calls the other scripts to generate data sets and replicate tables and figures.

- The first step of the master do-file sets paths (lines 5-13).
- The second part of the master do-file merges the various datasets located in the folders \rawdata and \additionaldata together (lines 25-135). This is achieved by calling the do-files \dofiles\01_formatting_car_ownership.do, \dofiles\02_creating_car_events.do, \dofiles\03_creating_end_of_year_stock.do and \dofiles\04_creating_data_hh_endofyear.do. Output datasets from these do-files are stored in the folder \newdata. Input datasets read and output datasets generated from each of these do-files are described in more detail in 00_masterfile.do. As the do-files use output datasets from previous do-files as input, they must be run in order.
- At line 183 in 00_masterfile.do, the file \dofiles\05_creating_project_specific_data.do is called to create the main dataset used in all subsequent analyses. The main role of this do-file is to implement the data selection criteria as defined in the paper to define the population, construct outcome and control variables, and define treatment status based on households' location and toll payments in 2014.
- At lines 191-205, we define global parameters for subsequent analyses. These globals define the time period of analysis and which explanatory/control variables that are used in the various regressions.

As explained above, we are not at liberty to share any of the data. Researchers wanting to replicate the results must apply for these data themselves, as explained in the "Data availability and provenance statements" above. It is unlikely that researchers will be able to obtain the exact same files as we have utilized, since Statistics Norway's process of creating datasets and naming variables might differ across commissions and caseworkers. Therefore, the abovementioned scripts for creating the datasets will likely have to be modified. However, the scripts should contain enough information in the form of in-line comments for this process to be straight-forward. We encourage

researchers interested in replicating the results to contact us if they require any assistance in putting these datasets together.

ANALYSIS:

- Next, the master do-file runs the base regressions for car ownership at lines 210-213, and stores the results in the ster folder. Finally it produces figures (lines 215-231) and tables (lines 233-242) for the main text, by calling separate do-files. Appendix figures and tables are replicated in the same manner by calling separate do-files at lines 252-287.

The files that are called to create figures and tables consist of the following steps, documented by in-line code: (1) open one of the main datasets from `\newdata`; (2) do the necessary sample selections and creation of variables (in some cases this includes merging additional data to the main datasets from `\additionaldata`); (3) run any necessary regressions, (4) store the results of these regressions as .ster-files in the folder `\ster_files`; (5) create tables and figures by accessing (a) the edited dataset and/or (b) importing estimation results from `\ster_files`; and (6) storing these tables and figures in the folders `\tables` and `\figures`.

OUTPUT GENERATION:

The tables and figures are named as in the published article. As an example, Figure 3 in the article consists of panels (a) and (b). These files are created by the script `fig3.do`, and saved as `fig3a.pdf` and `fig3b.pdf` in the `\figures` folder. Figures are stored as .pdf, while tables are stored in .tex format.

By running `00_masterfile.do` from start to finish, the following tables and figures will be produced:

Table 2. List of output files produced by running the script 0_Masterfile.do.

Main text		Online Appendix	
Tables:	**Figures:**	**Tables:**	**Figures:**
Table 2	Figure 2	Table D1	Figure C1
Table 3	Figure 3	Table D2	Figure E1
Table 4	Figure 4	Table D3	Figure E2
	Figure 5	Table D4	Figure E3
		Table D5	
		Table E1	

5.2 Instructions

To re-run the code, do the following steps:

Step 1: Obtain access to register data from Statistics Norway (see elsewhere in this document and https://www.ssb.no/en/data-til-forskning/utlan-av-data-til-forskere for more information). Provided that the request is granted (i.e. is in accordance with the Statistics Act of Norway), obtaining data will take approximately 0.5-1 year and likely cost between 50,000 and 300,000 NOK.

Step 2: Save the folders at your preferred location, including subfolders. This location must be on a secure server that satisfies requirements for storage of indirectly identifiable personal data.

Step 3: Ensure that the folders from `\rawdata` and `\additionaldata` contain all files described in Table 1, obtained from Step 1.

Step 4: Copy the filepaths to where the folders are stored.

Step 5: Open the master script `\dofiles\00_masterfile.do` and change the globals on lines 5-12 so that it matches your folder path. Also ensure that the command at line 13 points to the PLUS folder where .ado-packages are installed.

Step 6: In case your PLUS folder does not contain the packages listed at lines 15-17, uncomment and run these lines to install the required packages.

Step 7 (optional): In case the data you have obtained from Statistics Norway have different formats or use different naming conventions, make the necessary changes to the files `\dofiles\01_[…].do` through `\dofiles\05_[…].do`.

Step 8: Run the master script `STATA\dofiles\00_masterfile.do`. When everything is finished, you will find all figures and table inputs (as presented in Table 2) in the folders `\figures` and `\tables`.

Optional: As long as the globals from the master file have been set correctly and the datasets are created (i.e. lines 1-213 of the master file is ran), it is possible to run subsequent input files one by one, since each input file starts by loading a fresh version of the dataset(s).