This document describes the replication files for the paper, "Regulating Untaxable Externalities: Are Vehicle Air Pollution Standards Effective and Efficient?" by Mark Jacobsen, James Sallee, Joseph Shapiro, and Arthur van Benthem.

Two sets of files are not publicly available and hence not included here. One is a Vehicle Identification Number (VIN) decoder, purchased from DataOne. Other researchers may purchase these data directly from DataOne.

The other includes smog check and remote sensing records and variables from the Colorado Department of Public Health and the Environment (CDPHE). Other researchers may obtain these data directly from CDPHE.

File naming is fairly straightforward. The Stata programs 1_f1_newcar_trend.do through 1_f8_propTaxes.do produce the figures, and the programs 2_t2_tier0.do through 2_t4_tier2.do produce the tables. Excluding records derived from DataOne and CDPHE, the data used in these programs are in the /dataRAW and /dataSTATA subdirectories.

Quantitative model

Our code and inputs for the quantitative model are divided into subdirectories containing the C source code, inputs to control the model scenarios, and inputs containing base data.

Raw outputs from the C code are written to the directory: "pollutionmodel/output". The figures and summary data appearing in the paper are compiled from these outputs using the stata program "compile_quantitative_model_output.do" in the directory "pollutionmodel/stata_code".

The main quantitative model program is contained in "pollutionmodel/code/main.c" and each run of the program considers one scenario, as defined using the inputs specified in "pollutionmodel/carsim/controlfile.txt." The file "pollutionmodel/carsim/controlfiles-main.csv" contains columns of text, each of which defines a single scenario and produces one model output when copied into controlfile.txt and executing the main program. Running the program once for each of these columns produces all of the main scenarios considered in the paper and appendix.

Detail on the contents of the directories and running the model follow:

pollutionmodel/code/

File name	Contents
main.c	Main program code for quantitative model
broydn.c	Numerical methods
broydn2.c	Numerical methods
fdjac.c	Numerical methods
fminsq.c	Numerical methods
fminsq2.c	Numerical methods
lnsrch.c	Numerical methods
mjlib.c	Utility subroutines

mjlib.h	Header file (see above)
nrutil.c	Utility subroutines
nrutil.h	Header file (see above)
qrdcmp.c	Numerical methods
qrupdt.c	Numerical methods
rotate.c	Numerical methods
rsolv.c	Numerical methods

pollutionmodel/carsim

This is the default working directory for the main code. The code looks for the column of input in "controlfile.txt" to define and start a given run. Each welfare result in the paper relates to two runs of the model: first to create a baseline and then to consider a particular policy (such that welfare effects can be defined as the difference between the two).

For example, running the model using the inputs in column "b" of "controlfiles-main.csv" creates our main baseline scenario. Running it again using the inputs in column "ptpd2" of "controlfiles-main.csv" produces the "Delay Tier 2 by 4 years" scenario in Row 1 of Table V. The scenario identifier (e.g. "b" or "ptpd2") is used as the name of the corresponding output file written to "pollutionmodel/output". A list of all the scenario identifiers and corresponding results in the paper follows:

Run identifier	**Paper location**
b	Main baseline
ptpd2	Table V, row 1
ptpd4	Table V, row 2
ptpe2	Table V, row 3
ptpe4	Table V, row 4
ptp10	Table V, row 5
p100	Table V, row 6
p100rn	Table V, row 7
pn100	Table V, row 8
pf100	Table V, row 9
p010	Table A10, row 24

pa100	Table A10, row 25
b_ic	Baseline with imperfect competition
p100_ic	Table A10, row 20
ptpd4_ic	Table A10, row 10
b_loscrap	Baseline with lower scrap elasticity
p100_loscrap	Table A10, row 13
ptpd4_loscrap	Table A10, row 3
b_hiscrap	Baseline with higher scrap elasticity
p100_hiscrap	Table A10, row 14
ptpd4_hiscrap	Table A10, row 4
b_loage	Baseline with lower vintage substitution
p100_loage	Table A10, row 15
ptpd4_loage	Table A10, row 5
b_hiage	Baseline with higher vintage substitution
p100_hiage	Table A10, row 16
ptpd4_hiage	Table A10, row 6
b_hipgas	Baseline with higher gasoline price
p100_hipgas	Table A10, row 21
ptpd4_hipgas	Table A10, row 11
b_morecafe	Baseline with more stringent CAFE standards
p100_morecafe	Table A10, row 17
ptpd4_morecafe	Table A10, row 7
b_hiinc	Baseline with faster income growth
p100_hiinc	Table A10, row 18
ptpd4_hiinc	Table A10, row 8
b_hidisc	Baseline with higher internal discount rate
p100_hidisc	Table A10, row 22
ptpd4_hidisc	Table A10, row 12
b_altvmt	Baseline with alternative VMT schedule
p100_altvmt	Table A10, row 19
ptpd4_altvmt	Table A10, row 9
b_hitpcost	Baseline with higher cost exhaust improvement
ptp10_hitpcost	Table A10, row 23
b68	Baseline with 0.68% property tax
pf10068	Table A10, row 26

pollutionmodel/data

The model uses a set of plain text input files containing baseline data, pollution parameters, and policy specifications. Appendix F.1 of the paper provides detail on each of the data sources for our baseline. A short description of each of the input files in the directory follows:

Input file name	**Contents**
cafetarget_central.txt	Central case fuel economy policy targets
cafetarget_morecafe.txt	More rapidly increasing fuel economy targets
damages_msa.txt	Damages per ton of pollution, average when emitted in MSAs
damages_non_msa.txt	Damages per ton of pollution, average when emitted outside of MSAs
damages.txt	Damages per ton of pollution, national average
ddafile.txt	Table of ownership by vehicle age and income group from NHTS
elasticity_data_markup25_high_age.txt	Vehicle demand elasticities, higher age substitution case
elasticity_data_markup25_low_age.txt	Vehicle demand elasticities, lower age substitution case
elasticity_data_markup25.txt	Vehicle demand elasticities, central case
lambda_calib.txt	Fuel economy standard shadow costs
newcar_pollution_improvement.txt	Rate of improvement in new-vehicle emissions, applies post-2014
pollution_age_factors.txt	Rate of deterioration in emissions by age, applies post-2014
pollution_2000_through_2014_fleet.txt	Vehicle emissions data, 2000 through 2014
tailpipe_cost_central.txt	Tailpipe control costs, central
tailpipe_params_central.txt	Calibrated parameters of tailpipe control function, central
tailpipe_params_high.txt	Calibrated parameters of tailpipe control function, higher cost
tailpipe_policy_10.txt	Tailpipe policy specification: 10% emissions reduction
tailpipe_policy_d2.txt	Tailpipe policy specification: delay 2 time steps (4 years)
tailpipe_policy_d4.txt	Tailpipe policy specification: delay 4 time steps (8 years)
tailpipe_policy_e2.txt	Tailpipe policy specification: accelerate by 2 time steps (4 years)
tailpipe_policy_e4.txt	Tailpipe policy specification: accelerate by 4 time steps (8 years)
tailpipe_policy_zero.txt	Tailpipe policy specification: no change in tailpipe emissions
vehicle_data.txt	Baseline quantity, price, fuel economy data
vmt_schedule_alt.txt	Alternative VMT schedule by vehicle age
vmt_schedule.txt	VMT schedule by vehicle age

pollutionmodel/stata_code

The Stata code in "compile_quantitative_model_output.do" is in two main parts. The first section reads in the text outputs from the C program, located in "pollutionmodel/output." The second section compiles the results into tables and produces the graphics summarizing quantitative model outputs.