Documentation for replication for

"The Effect of Information on Market Activity: Evidence from Vehicle Recalls"

RStudio Version 1.4.1717, R Version 4.1.0 (2021-05-18), Windows 10 Operating System

Note: The commands involve large amount of data and therefore need to be run on a sufficiently powerful computer due to the peak memories needed. When this does not occur, the files might not be loaded or might be loaded only partially.

Instructions: Unpack the files in the same folder, then use the file **Shell.R** to execute the commands to go from the raw files to the results. If you are interested to replicate the results from the completed files, after setting the default folder and loading the packages as at the beginning of Shell.R you can run directly file **Analysis_final.R** for the portion of code linked to the table in the paper. See at the end of this document for an explanation of each variable.

Files tagged as "Raw data" are the files used by Shell.R to get the files used in the analysis. Files tagged as "Data" are the files used by Analysis_final.R to get the tables and figures of the paper.

Data on asking prices for vehicles put on sale, and some of their characteristics (zip code, type of seller, km driven) used to generate the results of Appendix G come from the Dutch online vehicle marketplace Autoscout24.nl. While the Dutch database law allows us to collect these data for research purposes, we can share these data to third parties only with the permission of Autoscout24.nl. Researchers interested into obtaining the asking price data for replication purposes should ask permission to Autoscout24 through the contact form at the following address: https://www.autoscout24.nl/bedrijf/contact/

STEP 1: GENERATE COMPACT FILE FOR REGISTRATION AND RESALE DATA

CODE: **1_Clean_registrations_EULER.R**

INPUT: Gekentekende_voertuigen_DATE.RData

OUTPUT: Registration_merged.RData

TEMPORARY FILES CREATED: CleanRegTemp.RData

DESCRIPTION: This file takes the "monthly snapshots" of the NL vehicle fleet downladed from RDW, extract the information on resales, exports, cars out of the sample, recalls, and based on changes from one period to another stores dates of important changes in status (recall, registration).

We consider only: passenger vehicles; the most common bodytypes ("cabriolet", "coupe", "hatchback", "MPV", "Niet geregistreerd", "sedan", "stationwagen"); 4-wheeled vehicles.

STEP 2: INTEGRATE REGISTRATION FILE WITH RECALL INFORMATION

CODE: **2_Clean_recalls_ 2019.R**

INPUT:
Terugroep_actie_status_DATE.RData
Registration_merged.RData
Terugroep_actie_6 april 2019.RData (LAST MONTH OF THE SERIES)

OUTPUT:
Kenteken_recalls.RData (Contains the match between license plate and recall identifier)
Registration_recalls_merged.RData (This file contains the merge of recall information and vehicle data, one obs per vehicle)

DESCRIPTION: This file integrates "Registration_merged.RData" with information on recall codes/number occurred during the vehicle lifetime and with recall-specific dates.

STEP 3: GENERATE COMPACT RECALL STATUS FILE

CODE: **3_Clean_recall_status.R**

INPUT: Terugroep_actie_status_DATE.RData

OUTPUT: Recall_week1.RData

DESCRIPTION: This file takes the "weekly snapshot" of status of each recall for each vehicle affected, and based on changes calculates fixing date and date of new recalls, including anomalies in the data. This is the most reliable recall data, except for the period Oct 2018 – Feb 2019, where updates on fixing were discontinued except for one update at the end of December. (For the case of vehicles with only one recall, which we use in our data, we are able to reconstruct the fixing date using the monthly recall status in Registration_merged.RData, which was not discontinued)

STEP 4: COMPLETE AND CLEAN RECALLS

CODE: **4_Complete_recalls.R**

INPUT:
Recall_week1.RData
Registration_recalls_merged.RData

OUTPUT: Recalls_merged_2019.RData

DESCRIPTION: This file recover recall status information during a disruption in RDW files that occurred in Fall 2018. We recover this info using the data in the main dataset which was not affected by the disruption, and restrict data to cars with one or zero recalls and other cleaning procedures. NOTE: This disruption affected only the date of CLOSING recalls i.e. when a recall was fixed – which we only analyze in one of the Appendix of the paper.

In preparation for the panel transformation, we use some rules to decide which recalls to keep:

1) Use only recalls in the main registration data (i.e. passenger cars)
2) Remove cars exported before Nov 2017 (recall status updates are not meaningful)
3) Keep only cars with zero or one recall (regardless of status)
4) Remove cars starting a recall already closed, not recorded in the main database and not exported. These reflect problems with the exact new recall date (i.e. all cars with a new recall in a given date have initial recall status as fixed) rather than very quick fixing.
5) Remove scrapped/stolen cars (impossible to figure out specific reason why they are out of database)

6) Drop some specific recalls for which RDW was not notified properly.

STEP 5: GENERATE BASIC PANEL OF RECALLS

CODE: **5_Create_Panel_Full.R**

INPUT: Recalls_merged_2019.RData

OUTPUT: Panel_Full.RData

DESCRIPTION: This file takes the merged recall and registration dataset and transforms it into a panel format (Nov 2017 – Mar 2019). This is still not the dataset used for the regressions – as for instance still contains very old vehicles (first registered before 2000) with no type approval.

STEP 6: GENERATE LINK WITH VEHICLE CHARACTERISTICS

CODE: **6_Create_Link_Char.R**

INPUT: Gekentekende_voertuigen_DATE.RData

OUTPUT: Link_Char.RData

DESCRIPTION: This file takes the raw data and extracts information on characteristics of vehicles + license plates. This is used to avoid a too large size of the dataset used for the analysis, the information is taken only when needed.

STEP 7: GENERATE LINK WITH MAKE AND MODEL CHARACTERISTICS

CODE: **7_Create_Link_Make.R**

INPUT: Gekentekende_voertuigen_DATE.RData

OUTPUT: Link_Makes.RData

DESCRIPTION: This file takes the raw data and extracts information on the make and the model of the vehicles. This is used to avoid a too large size of the dataset used for the analysis, the information is taken only when needed.

STEP 8: GENERATE APK INSPECTION DATA

CODE: **8_Clean_inspection.R**

INPUT: Geconstateerde_Gebreken_DATE.RData

INTERMEDIATE FILES: APK_1.RData, APK_2.RData,…

OUTPUT: Merged_APK (Clean information on APK inspection dates and outcomes)

APKtot_recent.RData (Information on the most recent inspection date and outcome for each month of the panel, used in the APK analysis)

DESCRIPTION: This file put together the APK vehicle inspection data and extract information on the most recent APK check .

STEP 9: GENERATE LINK WITH RECALL DATE

CODE: **9_Create_Link_Date.R**

INPUT:

Terugroep_actie_DATE.RData
Recall_week1.RData

OUTPUT: Link_Recalls_Date.RData

DESCRIPTION: This file generates a link to the license plate and the recall for the whole dataset (not just to the panel used in the main analysis), and it is used for the descriptive statistics of recall occurrences over time.

STEP 10: GENERATE PANEL FOR THE MAIN ANALYSIS

CODE: 10_Complete_Panel.R

INPUT:

Panel_Full.RData

HCPI.RData

APKtot_recent.RData

Link_Char.RData

OUTPUT: Panel_all_complete.RData

DESCRIPTION: This file adds the information on real price, inspection, brand rating, plus other variables used to the analysis to the main panel. The outcome is the file used in most of the main analysis.

STEP 11: GENERATE PANEL WITH ONLY WITHIN VERSION RECALL VARIATION

CODE: 11_Create_Panel_Ver.R

INPUT: Recalls_merged_2019.RData

OUTPUT:Panel.RData

DESCRIPTION: This file generates a panel that contains only versions with variation in recalls, used in the event study graph of the main results.

STEP 12: GENERATE ADDITIONAL RECALL INFORMATION

CODE: 12_Recall_info.R

INPUT: Terugroep_actie_DATE.RData

OUTPUT: Recall_info_2019.RData

DESCRIPTION: This file generates additional information on recalls, such as the specific vehicle part involved.

STEP 13: GENERATE PANEL FOR VARIANT ANALYSIS

CODE: 13_Create_Panel_Variant.R

INPUT: Recalls_merged_2019.RData

OUTPUT: Panel_var.RData

DESCRIPTION: This file generates a panel for the variant event study graph of Appendix C

STEP 14: GENERATE PANEL FOR MODEL ANALYSIS

CODE: 14_Create_Panel_Model.R

INPUT:

Recalls_merged_2019.RData

Link_Char.RData

OUTPUT: Panel_model.RData

DESCRIPTION: This file generates a panel for the model event study graph of Appendix C

VARIABLE DESCRIPTION

Variable name	Description
aantalcilinders	Number of cylinders
aantaldeuren	Number of doors
aantalzitplaatsen	Number of seats
APKtot	Number of vehicle defects in the last vehicle inspection (NA are either no defects or cars recently bought with no inspection yet)
brutobpm	Sales tax amount
catalogusprijs	Catalog price in EUR at the time of first registration
cilinderinhoud	Engine displacement
date	Month-year of observation
datumaankondigingproducent	Same as variable "rec_DAP"
datumeersteafgiftnederland	The date a car arrived in NL for the first time
datumeerstetoelating	The date a car was first registered anywhere in the world
datumeigenarengeu00efnformeerd	Same as variable "rec_DEG"
datuminformereneigenaar	Same as variable "rec_DIE"
datumtenaamstelling_all	All recorded dates of registration changes in NL, separated by ";". The last date is always at the end. We have all changes starting from November 2017 + the most recent change happened before that date.
difflastsale	Number of months from the last registration change, including new registrations
difflastsaleyr_d	Number of years (in factor format) from the last registration change, including new registrations
exported_last	Last date the car switched status to exported. We have all changes starting from November 2017 + some changes happened from November 2015 (exported cars disappear from the dataset after exactly two years they got exported, so we can calculate when the status changed). In a very limited number of cases (1000-2000 cars over 700,000 exported cars) a car is exported and then comes back, in which case we still keep the export date.
exportindicator	The export status for the last month of the dataset
handelsbenaming	Vehicle model
hcpi	Dutch monthly consumer price index (Base month-year: Jul 2013)
inrichting	Vehicle type
kenteken	Vehicle license plate
massaledigvoertuig	Unladen vehicle mass
massarijklaar	Mass in running order

merk	Vehicle manufacturer
missing	Number of times a car disappears from the database to reappear again. Used to track stolen cars or other anomalies
nationaalopgegevenaantalvoertuigenterugroepactie	Number of vehicles affected by the recall in the Netherlands
openstaandeterugroepactieindicator	The “has any open recall” status in the last month of dataset. Empty if car disappeared from the dataset
publicatiedatumrdw	Date of publication of recall by RDW i.e. when the recall becomes available to the public. All recalls from 2012
rating_merk	Reliability rating for brands/manufacturers as provided by the Dutch consumer association
realprice	Real catalog price in EUR
Rec_codes	All recalls codes for the vehicle (open and fixed), separated by “;”. All recalls from 2012.
rec_DAP	Date a recall has been recognized by the manufacturer, separated by “;” in case of multiple recalls. All dates from 2012 are available.
rec_DEG	Planned date of vehicle owner notified, separated by “;” in case of multiple recalls. Not individual based. All dates from 2012 are available.
rec_DIE	Planned date of notification of vehicle owner, separated by “;” in case of multiple recalls. Not individual based. All dates from 2012 are available.
rec_DMBR	Date of recall notification by RDW, separated by “;” in case of multiple recalls. All dates from 2012 are available.
rec_fixed	Dummy = 1 if recall was fixed (switching from open to close) that month
rec_fixed_t	Dummy = 1 if recall was fixed (switching from open to close) that month or in any previous month
rec_new	Dummy = 1 if recall started that month
rec_new_t	Dummy = 1 if recall started that month or in any previous month
rec_new_dist	Distance in month from the start of the recall, only after the start of the recall. 1= starting month of the recall
rec_new_dist_both	Distance in month from the start of the recall. 1= starting month of the recall. 0 and negative numbers are the months before the recall
rec_new_dist_cat	Distance in 3 months-groups from the start of the recall, only after the start of the recall. 1= starting three months of the recall
rec_part	Type of vehicle part recalled

rec_PDR	Date a recall has been published by RDW in their recall database (not vehicle specific), separated by ";" in case of multiple recalls. All dates from 2012 are available.
rec_precat1	Dummy = 1 if date corresponds to the 1-3 months before the start of the recall
rec_precat2	Dummy = 1 if date corresponds to the 4-6 months before the start of the recall
rec_precat3	Dummy = 1 if date corresponds to the 7-9 months before the start of the recall
rec_precat4	Dummy = 1 if date is earlier than 9 months before the start of the recall
rec_postcat1	Dummy = 1 if date corresponds to the 1-3 months after the start of the recall
rec_postcat2	Dummy = 1 if date corresponds to the 4-6 months after the start of the recall
rec_postcat3	Dummy = 1 if date corresponds to the 7-9 months after the start of the recall
rec_postcat4	Dummy = 1 if date is later than 9 months after the start of the recall
recall_fixed	The dates in which a car switches from "with at least one open recall" to "without any open recall, separated by ";". In case of vehicles with multiple recalls, this indicator cannot be 100% reliable to identify all fixes. This is used only to integrate and check our main sources of recalls. We have all changes starting from November 2017.
recall_new	The dates in which a car switches from "without any open recall" to "with at least one open recall", separated by ";". In case of vehicles with multiple recalls, this indicator cannot be 100% reliable to identify all new recalls. This is used only to integrate and check our main sources of recalls. We have all changes starting from November 2017.
recall_no	Months a vehicle is classified as "without any open recall", separated by ";". We have all months starting from November 2017
recall_out	Last date in which a car disappears from the database (probably because scrapped or exported)
recall_yes	Months a vehicle is classified as "with at least one open recall", separated by ";". We have all months starting from November 2017
referentiecoderdw	Recall identifier
resale	Dummy = 1 if car got resold (registration change) that month
risicobeoordelingrdw	Risk category of the recall
scrappage_date	The date a car disappears from the dataset for the last time without being exported. Despite the name of the variable, a car can disappear

	also if it gets stolen. Some cars can potentially reappear in the dataset at a later time
status_first	Status of the recall (open/fixed) when the recall appears for the first time. Generally cars with a fixed recall status are older cars that are exported in NL
status_last	Last known status of the recall. Ignores when recall is fixed and then reopened again (generally those reopened recalls are fixed again within a short period of time).
time_NL	Number of years the car is registered in the Netherlands
toegestanemaximummassavoertuig	Maximum mass allowed
totaalaantalvoertuigenterugroepactie	Number of vehicles affected by the recall worldwide
typegoedkeuringsnummer	The type-approval of the vehicle, the 1st tier EU vehicle classification.
uitvoering	The version of the vehicle, the 3rd tier EU vehicle classification.
variant	The variant of the vehicle, the 2nd tier EU vehicle classification.
vervaldatumapk_all	All recorded dates of last scheduled mandatory check, separated by ";". The last date is always at the end.
zuinigheidslabel	Fuel economy label