Replication Material for article: **Learning from Law Enforcement**
Authors: **Libor Dusek and Christian Traxler**

1. DATA

There were two primary sources of data:

- Data from the speed cameras, containing the record of every ride through the speed camera zones (zone id, anonymous car id, precise time of entry and time of exit, measured speed).

- Data from the speeding ticket administration, containing the record of every offense (ride above the enforcement cutoff) that was issued a ticket (information about the offense, dates of administrative steps, etc) as well as the specific day the ticket was delivered.

Both datasets were parsed to Stata, cleaned, and merged according to time of the day of entry to the zone, measured average speed and location. Based on the time of the day, weather condition controls (from a separate source) were merged in.

1.A Data for RD Design

In the next step, individual outcomes for each car were computed. For the RD analysis, we computed the maximum speed during an assignment period of a = {3, 4, 5, 6} months as well as indicators for outcome (or follow-up) periods of f = {3, 4, 5, 6} months. (This step is described in Section 4.1 of the paper and detailed in Online Appendix C.1.) The resulting output is stored in the following datasets:

Dusek_Traxler_RD_a3.dta Dusek_Traxler_RD_a4.dta
Dusek_Traxler_RD_a5.dta Dusek_Traxler_RD_a6.dta

To keep file size at reasonable levels, we reduced sensitive information (such as the precise millisecond of each camera zone entry/) from these datasets. The data file *_a4.dta (which is the starting point for our main RD analysis, which uses a = 4) contains more detailed information as compared to the other files (*_a3.dta, *_a5.dta, *_a6.dta; which are only used in a series of sensitivity checks).

In two complementary analyses, we used supplementary data from the initial five speed camera zones (covering a longer time period) and from additional camera zones that were set-up more recently (see Sections 2 and 4.4 of the paper). These data were prepared exactly as described above (but only for a = f = 4). The resulting data are stored in

Dusek_Traxler_RD_supplementary.dta

For another complementary analysis, we went back to the raw data to exploit the exact time from exiting one and entering another camera zone. Based on the observed time gap, one can compute cars' speed in the unmonitored road between the two speed camera zones. The resulting data (again, computed for parameters a = f = 4) can be found in:

Dusek_Traxler_RD_interim_road.dta

1.B Data for Event Study Design

Going back to the micro data, we re-compiled the data in a different structure that is suitable for the event study. We defined individual driving outcomes histories around the times when a car received a first (or second) speeding ticket. The resulting data are stored in the following files:

Dusek_Traxler_EVENT_weekeffects
Dusek_Traxler_EVENT_montheffects

In all datasets, the unit of observation is a single ride. In the event study dataset, rides resulting in a speeding ticket contain non-missing values of variables related to the ticket administration. The datasets differ in the time unit by which individual outcomes, histories, and time-related dummies are defined: either by the weeks or by the months that precede or follow an individual ticket.

The event-study samples focus on cars with at least one speeding ticket. The weekly dataset sample was then narrowed down to cars with at least one observation (other than the ride that triggered the ticket) during 12 weeks before receiving the ticket and at least one observation during 20 weeks after the ticket. The monthly dataset was likewise narrowed to cars with at least one observation (other than the trigger observation) during 6 months before the ticket and at least one observation in 24 months after the ticket.

2. CODE

2.A Code replicating RD Results

The key results of the RD analyses at the (01) enforcement cutoff and the (02) high-fine cutoff (i.e., the main figures and tables from the paper as well as most figures and tables from the online appendix) are produced with the Stata do files

RD_01.do and RD_02.do .

Results for the sensitivity analyses w.r.t. the a- and f- parameters (assignment and follow-up period, see above) at the (03) enforcement cutoff and the (04) high-fine cutoff are produced by the do files:

RD_03.do and RD_04.do .

The code for the bunching analysis (at both cutoffs) is in file

RD_05.do

The two analyses that make use of the supplementary data are replicated with

RD_06.do

The RD analysis that examines responses at unmonitored roads (between camera zones) is in

RD_07.do

2.B Code replicating Event Study Results

The code for producing all figures and tables from the event study analysis is provided in the do file:

EVENT_01.do

The file calls another do-file that contains support programs for plotting of coefficients, which is called

supportprograms.do

2.C Running the code

Download the do-files and place them in your project folder. (In the do-files, this is set as C:/yourdir).

Download the data and place them in a subfolder /datasets in your project folder (ie, C:/yourdir/datasets).

The code needs Stata 14 or higher. (We worked with a server version but also a (local) MP2 Stata 14.)

Several parts require a working RAM of up to (at least) 12GB; some steps are quite time consuming.

The code in RD_05.do makes use of two small ado files – minor modification of McCrary (2008)'s code for his manipulation test (which make use of different ranges of the running variable) - which are available in the repository for this project:

DCdensityThree.ado
DCdensityFour.ado

The code also makes use of several publicly available Stata packages:

RDD: coefplot , cmogram, grstyle , gtools, rdrobust, parmest

Event analysis: estout/eststo, ftools, labutil, outreg2, panelstat, parmest, ranktest, reghdfe

Note 1: The paper reports robust estimates with bias-corrected SEs – which is stored in the third output 'block' (stored in tex- and txt-files) obtained from the rdrobust command.

Note 2: We worked with the 2018-09 version of rdrobust (Calonico et al., 2017). We noticed minor differences – mainly regarding the optimal bandwidth selection – between different (in particular: earlier) versions of the rdrobust package. (Ideally, a replication should use a version after the Fall 2016 and before the Winter 2020 update of the package.)

3. OVERVIEW: Code and output

Tables	Do-file
2, Panel A	RD_01
2, Panel B	RD_01
2, Panel C	RD_02
2, Panel D	RD_02
3, Panel A	RD_01
3, Panel B	RD_01
3, Panel C	RD_02
3, Panel D	RD_02
A1	RD_01
A2	RD_01
A3	RD_02
A4, Panel A	RD_01
A4, Panel B	RD_02
A5, Col1+2	RD_01
A5, Col3+4	RD_02
A6	RD_06
A7, Panel A&B	RD_01
A7, Panel C&D	RD_02
A8	RD_03
A9	RD_03
A10	EVENT_01
A11	EVENT_01
A12	EVENT_01
A13	EVENT_01

Tables	Do-file
C1, Panel A+B	RD_01
C1, Panel C+D	RD_02
C2	RD_06
C3	RD_07
C4, Panel A + B	RD_01
C4, Panel C + D	RD_02
D1	EVENT_01
D2	EVENT_01

Figure	Do-file
2a	RD_01
3a	RD_01
3b	RD_02
3c	RD_01
3d	RD_02
4a	RD_01
4b	RD_02
5a/b	EVENT_01
6a/b	EVENT_01
7a/b	EVENT_01
A2, a + c	RD_05
A2, b + d	RD_05
A3	RD_05
A4	RD_05
A5	RD_01
A6	RD_02
A7	RD_01
A8, a + c	RD_01
A8, b + d	RD_02
A9, a + c	RD_03
A9, b + d	RD_04
A10, a + c	RD_01
A10, b + d	RD_02