ReadMe1st Companion to: The Short and Long Term Effects of In-Person Performance Feedback:

Evidence from a Large Bus Driver Coaching Program

Journal of Political Economy Microeconomics

Adriaan R. Soetevent* University of Groningen

Gert-Jan Romensen† University of Groningen

May 14, 2025

Contents

1 Restricted access **2**

2 Pre-requisites **2**

3 Code **3**
3.1 Data preparation 3
3.2 Analysis 5

4 Origin Results Presented in Tables, Figures and In-Text **7**
4.1 Tables 7
4.2 Figures 7
4.3 In-Text Numbers 7

5 Origin Results Appendix **8**
5.1 Tables 8
5.2 Figures 8
5.3 In-Text Numbers 8

*Corresponding author. University of Groningen, EEF, P. O. Box 800, 9700 AV Groningen, The Netherlands, a.r.soetevent@rug.nl.

†University of Groningen, Faculty of Economics and Business, Nettelbosje 2, 9747 AE Groningen, The Netherlands, g.j.romensen@rug.nl.

1 Restricted access

The data of this study is subject to restricted access. The agreement between the company (Arriva) and the researchers at the University of Groningen (Romensen and Soetevent) stipulates that in the 10 years following publication, access to the data will be granted to independent researchers and other third parties with a reasonable interest in auditing the data in order to verify whether the empirical analysis and results as reported in the publication can be replicated.

All do-files will be put in the depository, but the contract prevents the deposition of the data in the repository of the journal. Upon request, other researchers who wish to have access to these data can contact the authors. The authors will then submit the data request to the company to discuss whether the "reasonable interest" clause applies. If granted, the researchers will be given access to the anonymized versions of the cleaned data (the names of these data sets end with the letters `DEPO`) that have been used in the empirical analysis. These are sufficient to allow to verify the origins of most of the empirical and statistical findings as they appear in the tables, figures and running text of the paper and online appendix.

With regard to the raw data, the European GDPR does not allow us to provide these either to the Editor or to third-party researchers, because they contain detailed information on individual trips and individual drivers (these are the files classified as "restricted" in Section 3.1, *Data preparation*). For running the main analyses in the paper these data are however not needed and the cleaned data as provided in the `DEPO`-files suffice. All code in the analysis section 3.2 will run using these files. In case access to the raw data is needed, researchers should get into contact with the company directly and establish a separate Data Processing Agreement (DPA).

2 Pre-requisites

The code was run using STATA, version SE 16.1, on a Windows 11 operating system. To run the code, create the folder `"C:/JPEMicReplication/"`[1] and put all files and folders in this folder. Figure 1 shows the tree structure of the data folders. The do-file `X06-BasicRegressions` uses the program `ebayes.ado` developed by Adam Sacarny. It implements the Empirical Bayes estimator of Morris (1983) and is described in Appendix C of Chandra et al. (2016). This program can be downloaded from Adam Sarcarny's web site. Please install it before running the code.

Figure 1: Tree structure Data Folders

Folders			Content	Restricted acccess?
Main	Sub			
C:/JPEMicReplication	/ConstructedData		created data are stored here	No
	/Data	/Dagboeken coaches	diaries coaches	Yes
		/Randomisatie	randomization drivers	Yes
		/Raw data	detailed data files on trips, coaching dates and employees	Yes
		/Using databases	contains the sub-folder "Incidence check" which initally is empty	No
	/DEPO		DEPO[sitory] data files are stored here	Yes
	/Do-files		contains code of all do-files used in data cleaning and analysis	No
	/Logs		initially empty; logs are stored here	No
	/TablesGraphs		initially empty, all created figures and tables are stored here	No

[1] You can use any other name and directory by changing in the master do-files the line `global filepath "C:/JPEMicReplication"` accordingly.

3 Code

This section lists all the STATA do-files that were run in preparing (Section 3.1) and analyzing (Section 3.2) the data. The purpose of each do-file is briefly described as are the data sources used by the code.

3.1 Data preparation

The master do-file `0-Master_do_Preparation.do` calls the do-files `1-Cleaning.do-12-SurveyDataPreparation.do` below that are used to prepare the data for analysis.

- `1-Cleaning.do` Cleans the raw monthly trip-level data set and identifies driver-specific extreme observations of the outcome variables.

  Uses: `rawtripdata_1501 tm 1701` and `merged_ZHNDAV_rawtripdata.dta`: restricted access.

  Saves: `cleaned_tripdata_1501 tm 1701.dta` and `cleaned_ZHDAV_1507 tm 1712.dta`: restricted access.

- `2-Coachlogs.do` Cleans the coach logs and prepares a file that can be attached to the trip-level data.

  Uses: `RandomisatieFinal.csv`.
  Uses: `Ecobox Coaches V2 - RUG v01.xlsx`: restricted access.

  Saves: `Using databases/coachlogs.dta`: restricted access.

- `3-Weather.do` Retrieves data on daily weather in the region for the relevant sample period.

  Uses: `KNMI_20170227(1).txt`.

  Saves: `Using databases/weather.dta`

- `4-Driverinformation.do` Constructs a data set with background information on drivers.

  Uses: `RandomisatieFinal.csv`.
  Uses: `Medewerkers.xlsx`: restricted access.

  Saves: `Using databases/Medewerkers.dta`: restricted access

- `5-Overallfeedback.do` Shows per driver and feedback round if and which targeted messages were part of the feedback report. Uses: `overallfeedbackarriva.xls`.

  Saves: `Using databases/Overallfeedback.dta`

- `6-Preparingforanalysis.do` Determines driver-specific post-feedback and post-experimental period. It also constructs various variables that will be used in the empirical evaluation of the feedback programs.

  Uses: `feedback_‘i’.xlsx`.
  Uses: `cleaned_tripdata_1501 tm 1701.dta`: restricted access.
  Uses: `cleaned_ZHDAV_1507 tm 1712.dta`: restricted access.
  Uses: `Using databases/Overallfeedback.dta`
  Uses: `Using databases/weather.dta`
  Uses: `Using databases/Medewerkers.dta`: restricted access.

Saves: ConstructedData/analysis_tripdata_1501 tm 1701.dta: restricted access.
Saves: ConstructedData/analysis_ZHDAV_1507 tm 1712.dta: restricted access.

- 7-Feedbackincidencecomposition.do This do-file shows the incidence of the targeted peer-comparison messages (share of drivers receiving a certain message combination) as well as the composition of these messages (%A;%B:%C).

  Uses: RandomisatieFinal.csv.
  Uses: feedback_'i'.xlsx: restricted access.
  Uses: Overallfeedback.dta.

  Saves: Using databases/randomization.dta: restricted access.
  Saves: Using databases/FeedBackReceived.dta.

- 8-Peereffort.do Constructs figures that highlights how often treated drivers were eligible for targeted peer-comparison messages. Uses: feedback_'i'.xlsx: restricted access.

  Saves: Using databases/Incidence check/feed_incidenceRANK_'i'.dta: restricted access.
  Saves: Using databases/Incidence check/feed_incidence_'i'.dta: restricted access.
  Saves: Using databases/Incidence check/FullDataRelPerformanceAndFeedback.dta: restricted access.

- 9-Coachingtimes.do Constructs a figure that illustrates when drivers received their first coaching. Uses: Using databases/coachlogs.dta: restricted access.
  Uses: analysis_tripdata_1501 tm 1701.dta: restricted access.
  Uses: ConstructedData/analysis_ZHDAV_1507 tm 1712.dta: restricted access.

  Saves: Using databases/coachyearweeks.dta: restricted access.
  Saves: Using databases/coachyearweeksZH.dta: restricted access.

- 10-PreparationForAnalysis.do Creates additional variables used in basic regressions and data set DataMainAnalysis.dta.

  Uses: ConstructedData/analysis_tripdata_1501 tm 1701.dta: restricted access.
  Uses: ConstructedData/analysis_ZHDAV_1507 tm 1712.dta: restricted access.

  Saves: ConstructedData/DataMainAnalysisFR.dta: restricted access.
  Saves: ConstructedData/DataMainAnalysisZH.dta: restricted access.
  Saves: ConstructedData/DataMainAnalysisALL.dta: restricted access.

- 11-DataToDepository.do Scrambles observations in "DataMainAnalysisALL.dta", "FullDataRelPerformanceAndFeedback.dta", "coachlogs.dta", and "FeedBackReceived.dta" and removes all driver-related information not used in the analysis. The resulting data sets with the added word "DEPO" can be made available upon request, see Section 1.

  Uses: Using databases/Scrambling drivers and dates for data depository - ALL.txt: restricted access.
  Uses: ConstructedData/DataMainAnalysisALL.dta: restricted access.

Uses: `Using databases/Incidence check/FullDataRelPerformanceAndFeedback.dta`: restricted access.
Uses: `Using databases/coachlogs.dta`: restricted access.
Uses: `ConstructedData/FeedBackReceived.dta`: restricted access.

Saves: `ConstructedData/DataMainAnalysisDEPO.dta`
Saves: `ConstructedData/FullDataRelPerformanceAndFeedbackDEPO.dta`
Saves: `ConstructedData/coachlogsDEPO.dta`
Saves: `ConstructedData/FeedBackReceivedDEPO.dta`

- `12-SurveyDataPreparation.do` Prepares survey data from employees for analysis.

  Uses: `Arrivasurveyresponses01042021.xlsx`: restricted access.

  Saves: `SurveyDataDEPO.dta`.

3.2 Analysis

The master do-file `X0-Master_do_Analysis.do` calls the do-files `X01-Coachvsnoncoachdescriptives.do`-`X09-AnalysisSurvey.do` below that are used in the empirical data analysis.

- `X01-Coachvsnoncoachdescriptives.do` Assesses whether the phase-in of coaching was done in a quasi-random manner. [WARNING: Takes long to run]

  Uses: `DataMainAnalysisDEPO.dta`.
  Uses: `driversLWcityruralDEPO.dta`.

- `X02-BasicPictures.do` Creates some descriptive figures of the data.

  Uses: `DataMainAnalysisDEPO.dta`.

- `X03-RegrPeerComp.do` Analysis on tailored performance feedback.

  Uses: `DataMainAnalysisDEPO.dta`.

- `X04-RegressionsCoachingFR.do` and `X04-RegressionsCoachingZH.do` Analysis on the effects of in person feedback, first file for Treatment region, second for Control region. `X04b-RegressionsCoachingRobustRushFR.do` is identical except that rush hours are excluded as control variables to perform a robustness check.

  Use: `DataMainAnalysisDEPO.dta`.

- `X04-RegressionsCoachingCATTstatic.do` and `X04b-RegressionsCoachingRobustRushFR.do` analyze for the treatment region the effects of in person feedback using Sun and Abrahams (2021) CATT analysis. The first estimates the time invariant effects, the second the dynamic treatment effects.

  Use: `DataMainAnalysisDEPO.dta`.

- `05-CostBenefitAnalysisCoaching.do` Short cost-benefit analysis of coaching program.

- `X06-BasicRegressions.do` Some basic regressions on the determinants of fuel economy and comfort and estimation (EB-adjusted) driver quality metrics.

  Uses: `DataMainAnalysisDEPO.dta`

  Saves: `EstimatedFixedEffectsDriverLevelDEPO.dta`

- `07-HetEffectsCoaching.do` Tests for heterogeneity in treatment effect coaching related to salience of effort and fatigue.

  Uses: `DataMainAnalysisDEPO.dta`

- `X08-HetEffectCoachingProductivity` Tests for heterogeneous treatment effects conditional on pre-coaching productivity.

  Uses: `DataMainAnalysisDEPO.dta`

- `X09-AnalysisSurvey.do` Analysis Eco Coach survey data.

  Uses: `SurveyDataDEPO.dta`

4 Origin Results Presented in Tables, Figures and In-Text

4.1 Tables

Table	Source
Table 1	`X01-Coachvsnocoachdescriptives.do`, line 127-165.
Table 2	`X06-BasicRegressions.do`, lines 289-308, 310-344, 686 .
Table 3	`X04-RegressionsCoachingFR.do` [line 109-234] and `X04-RegressionsCoachingCATTstatic.do` [line 124-134] .
Table 4	`X08-HetEffectCoachingProductivity.do`, line 389-412.
Table 5	`X07-HetEffectsCoaching.do`, line 224-287.
Table 6	`X04-RegressionsCoachingZH.do`, line 99-161.
Table 7	`X04-RegressionsCoachingZH.do`, line 99-161.

4.2 Figures

Figure	Source
Figure 1	DOES NOT CONTAIN EMPIRICAL DATA.
Figure 2	`X04-RegressionsCoachingFR`, line 174-186.
Figure 3	`X04-RegressionsCoachingCATT.do`, line 161-176.

4.3 In-Text Numbers

p. 13 "At the end of this period, 110 drivers (27% of the full sample) was never coached and 21 drivers (5%) participated in additional coaching sessions."

- `X04-RegressionsCoachingCATT.do`, line 144-151.

p.26 "In the weeks following coaching, 17.52 liter of fuel is saved per coached driver, which amounts to €19.27, or €60 per day of coaching."

- `X05-CostBenefitAnalysisCoaching.do`.

5 Origin Results Appendix

5.1 Tables

Table	Source
Table C.1-C.4	`X06-BasicRegressions.do`, lines 136-220.
Table F.5	This tables summarizes the different reasons used in `1-Cleaning.do` to drop observations to arrive at the final sample.
Table F.6	`1-Cleaning.do`, line 277-279.
Table F.7	DOES NOT CONTAIN EMPIRICAL DATA, DATA RETRIEVED FROM THE COMPANY.
Table G.8	`X03-RegrPeerComp.do`, line 310-343.
Table G.9	`X03-RegrPeerComp.do`, line 345-394.
Table G.10	`X03-RegrPeerComp.do`, line 238-301.
Table H.11	`X04-RegressionsCoachingFR.do` [line 107-233] and `X04-RegressionsCoachingCATT.do` [line 194-205].
Table M.13	Summary employees invited for survey.
Table M.14	`X09-AnalysisSurvey.do`, line 118-128.
Table M.15	`X09-AnalysisSurvey.do`, line 630-639.
Table N.16	`X04b-RegressionsCoachingRobustRushFR.do`, line 110-234.

5.2 Figures

Figure	Source
Figure B.1	`X06-BasicRegressions.do`, lines 401-436.
Figure D.2	`10-Coachingtimes.do`, line 92.
Figure D.3	`X06-BasicRegressions.do`, lines 563-590.
Figure E.4	DOES NOT USE EMPIRICAL DATA.
Figure G.5	`X03-RegrPeerComp.do`, line 31-33.
Figure I.6	`X02-BasicPictures.do`, line 36-74.
Figure J.7-J.10	`X04-RegressionsCoachingFR.do`, line 236-297.
Figure K.11	`X08-HetEffectCoachingProductivity.do`, line 331-385.
Figure L.12	`X04-RegressionsCoachingZH.do`, line 163-173.
Figure M.13	`X09-AnalysisSurvey.do`, line 105-116.
Figure M.14	`X09-AnalysisSurvey.do`, line 428-464.
Figure M.15	`X09-AnalysisSurvey.do`, line 568-604.
Figure N.16	`X04-RegressionsCoachingFR.do`, line 300-343.

5.3 In-Text Numbers

A.18 "Similarly, when we replace PF_{it} in (A.9) by PA_{it} (a variable that equals one if the observation is in the post-announcement period), γ estimates the aggregate effect of the EcoManager launch. We find $\hat{\gamma} = -0.021$ (s.e. 0.055)."

- `X03-RegrPeerComp.do`, line 232-263, Table 6E.

fn. G.6 – `X03-RegrPeerComp.do`, line 437-446.

References

Chandra, Amitabh, Amy Finkelstein, Adam Sacarny, and Chad Syverson, "Health Care Exceptionalism? Performance and Allocation in the US Health Care Sector," *American Economic Review*, August 2016, *106* (8), 2110–2144.

Morris, Carl N., "Parametric Empirical Bayes Inference: Theory and Applications," *Journal of the American Statistical Association*, March 1983, *78* (381), 47–55.