Time-Varying Risk Aversion? Evidence from Near-Miss Accidents

Data and Code Description

Matthew Shum and Yi Xin (Caltech)

October 21, 2020

The data we use for this paper are provided by OK Chexian, which is an insurance technology firm in China. The data are proprietary, so they cannot be made available online without the permission of the company. Researchers who are also interested in obtaining the data need to submit research proposals to OK Chexian. The contact information can be found on the company's website: https://www.okchexian.com/indexEng.html. Reasonable applications will be considered by the company. We can also help put researchers in touch with the company directly upon their request.

The code we developed for this paper was written in Stata (MP 15) and Matlab (R2018a). We list all the replication files with a brief description of each below.

1. Cleaning Raw Data

1) **clean_address.do**: clean raw longitude and latitude location information
   Input: geo_data/location_start 4.csv
   Output: geo_data/clean_addr.dta

2) **clean_weather.do**: clean weather info collected for 363 Chinese major cities
   Input: weather_info/weather-csv/`city_name'.csv
   Output: working_data/weather_all.dta

3) **clean_quote.do**: clean quotes info submitted by users via the app
   Input: raw_data/quote_order_all.dta
   Output: working_data/quote_day_level_clean_042719.dta

4) **clean_claim.do**: clean raw claims filed by users, first-step cleaning
   Input: raw_data/claim_details.dta
   Output: working_data/claim_clean_070219.dta

5) **clean_claim_2.do**: further clean claim data
   Input: working_data/claim_clean_070219.dta
   Output: working_data/claim_clean_further_07032019.dta

6) **clean_repair.do**: clean repair data
   Input: raw_data/repair_raw.dta
   Output: working_data/repair_clean_further_070319.dta

7) **clean_main.do**: clean raw trip data, merge geographic and weather info, generate trip and day level clean data
   Input: raw_data/trip_event_all.dta
   Merge: working_data/clean_addr.dta, get geo information
   Merge: working_data/weather_all.dta, get weather information
   First Output: working_data/clean_trip_level_geo_weather_081519.dta
   Second Output: working_data/clean_day_level_geo_weather_081519.dta

8) **clean_quote_driving.do**: merge quote data with driving data
   Input: working_data/clean_day_level_geo_weather_081519.dta,
   Merge: working_data/quote_day_level_clean_042719.dta
   Output: working_data/merge_quote_drive_081619.dta

9) **clean_fam.do**: construct near-misses that occurred on familiar/non-familiar trips
   Input: working_data/clean_trip_level_geo_weather_081519.dta
   Output: working_data/clean_day_level_geo_weather_familiar_012020.dta

2. Generating Main Results

1) **summary.do**: produce summary statistics
   use working_data/clean_trip_level_geo_weather_081519.dta
   — generate trip-level summary statistics
   use working_data/clean_day_level_geo_weather_081519.dta
   — generate day-level summary statistics
   use working_data/merge_quote_drive_081619.dta
   — generate summary statistics for quotes and orders
   use working_data/claim_clean_further_07032019.dta
   — generate summary statistics for claims

2) **plot.do**: generate Figures 1 and 2 in the paper
   use working_data/clean_day_level_geo_weather_081519.dta
   — plot "before/after a NM, change in risky behavior"
   — plot long-run effect of near-misses on risky behavior

3) **gen_nm_diff.do**: use this file to generate measures of nm and differenced variables

4) **reg_day_level.do**: produce main results reported in the paper
   use working_data/clean_day_level_geo_weather_081519.dta
   — generate regression results for baseline and iv (full sample)
   — robustness check: using alternative definitions for NM
   — adding additional lags
   — robustness check: regular drivers
   — robustness check: experience using the app
   use working_data/merge_quote_drive_081619.dta
   — robustness check: experienced vs. inexperienced drivers

— robustness check: 30+ drivers

use working_data/clean_day_level_geo_weather_familiar_012020.dta

— robustness check: near-misses occurred on familiar/non-familiar roads

5) **reg_day_level_add.do**: produce additional robustness check results

use working_data/merge_quote_drive_081619.dta

— robustness check: male users

use working_data/clean_day_level_geo_weather_081519.dta

— robustness check: rural vs. urban area

— robustness check: economically developed vs. less developed area

6) **collect_reg_coef.do:** use this file to generate latex tables automatically

7) **permutation.do:** shuffle near-miss measures, run IV specification

use working_data/clean_trip_level_geo_weather_081519.dta

— shuffle within individual, generate 20 random draws

— shuffle all, generate 20 random draws

8) **placebo.do:** replace near-misses with other time-varying variables, run IV specification

use working_data/clean_trip_level_geo_weather_081519.dta

— switch the values for risky behavior and near-misses

— use other time varying variables: driving scores, traffic, weather, etc.

9) **calibration.do:** validate near-miss measures using real accidents, run logistic regression of accident on risky measures

use working_data/merge_quote_drive_081619.dta

merge working_data/claim_clean_further_07032019.dta

merge working_data/repair_clean_further_070319.dta

— summarize the relationship between NM and real accidents

— predict real accidents using risky behavior

10) **bootstrap.do**: prepare data for bootstrap

use working_data/clean_trip_level_geo_weather_081519.dta

randomly select 80% of the observations, run the main regression for nm de1 and def2

obtain 100 subsamples and estimates based on each

generate bootstrap_data_072020.mat (including estimated coefficients, levels of risky behavior and covariates for each sample).

11) **bootstrap.m (Matlab):** estimate changes in risk-aversion and other parameters, compute standard errors.

use data bootstrap_data_072020.mat

— model 1: allow both changing risk aversion and a salience parameter (reported in online appendix)

— model 2: allow only changing risk aversion (reported in main text)