The Goldfarb et al. Auto Race FIVES Data

by

Brent Goldfarb

Associate Professor of Entrepreneurship & Strategy
Robert H. Smith School of Business
University of Maryland

Anastasiya (Annie) Zavyalova

Assistant Professor of Strategic Management
Jones Graduate School of Business
Rice University

Sandeep Devanatha Pillai

Assistant Professor
Bocconi University

1. Conditions of Use

The data sets described in this document are part of the FIVES Project on Firm and Industry Evolution and Entrepreneurship, organized by Constance Helfat (Tuck School of Business at Dartmouth) and Steven Klepper (Carnegie-Mellon University). If you download or otherwise obtain any of the data sets described in this document, or obtain modified versions of these data sets, you agree to abide by the conditions of use set forth here.

The organizers of the FIVES Project are not responsible for the content of any of the FIVES Project data sets. The FIVES Project organizers and the FIVES Project data set authors make no representations or warranties of any kind concerning any of the FIVES Project data, including with regard to the absence or presence of errors.

All FIVES Project data sets are protected by copyright. The copyright holders for each data set are indicated on the first page of the documentation for each data set. Permission is granted to reproduce FIVES Project data for non-profit educational and research use only. Requests to reproduce FIVES Project data for other uses should be addressed to: Professor Constance Helfat, Tuck School of Business at Dartmouth, Hanover, NH 03755, constance.helfat@dartmouth.edu.

This document describes the Goldfarb et al. Data, which should be referred to by this name in any derivative works. In any written and published work, users of these data should cite this document, one of the data set author's original articles (see below, Section 4), and the FIVES Project: Data Overview (Helfat & Klepper, 2007).

You may create a new data set by modifying a FIVES Project data set, including but not limited to adding or removing data or merging data from different FIVES Project data sets. If you utilize any FIVES Project data to create new data sets, within six years from the date that you obtain the FIVES Project data, you must provide the FIVES Project with a copy of your modified data set by contacting: Constance Helfat, constance.helfat@dartmouth.edu.

Any written documents or statistical estimates that use FIVES Project data, including but not limited to working papers and publications, must be reported to: Constance Helfat, constance.helfat@dartmouth.edu.

Additional conditions of use (specified by the author of this document):

Users of the Goldfarb et al. Data must not redistribute the data without written permission from one of the study's co-authors: Brent Goldfarb., Anastasiya Zavyalova or Sandeep Devanatha Pillai

2. Dataset Authors

Brent Goldfarb
bgoldfarb@rhsmith.umd.edu +1 202-730-9484
Associate Professor of Entrepreneurship & Strategy
Smith School of Business, University of Maryland

Anastasiya (Annie) Zavyalova
anastasiya.zavyalova@rice.edu +1 713-348-6342
Assistant Professor of Strategic Management
Jones Graduate School of Business, Rice University

Sandeep Devanatha Pillai
sandeep.pillai@unibocconi.it
Assistant Professor
Bocconi University

3. Dataset Summary

This dataset contains the following files:

Data_Goldfarbetal.dta - This file contains data that has been used for the analysis in: Goldfarb, B., Zavyalova, A, and Pillai, S. 2018. Did victories in certification contests affect the survival of organizations in the American automobile industry during 1895-1912? A replication study. *Strategic Management Journal*, 39(8): 2335-2361. Description of the data collection process and of the variables in the dataset are provided in Sections 4 and 5.

Code_Goldfarbetal.do - This file contains the STATA code that was used to analyse the data in Data_Goldfarbetal.dta.

4. Data Description

Data_Goldfarbetal.dta has a Firm-Year data structure and contains the following data:

(a) Racing information collected from *Horseless Age* magazine on 1451 unique car races, held under the auspices of 305 tournaments in which 444 unique car manufacturers participated between 1895 and 1912. For each firm in a given year, the dataset captures information about the number of races that the firm participated in, and the number of first, second and third place finishes that the firm achieved. This information has been further demarcated by the location of the races (in-state vs. out of state), the type of race (hill climbing race, long distance endurance race, and shorter distance speed race), and the spectator attendance.

(b) Information about the firm collected from the *Standard Catalogue of American Cars 1805–1942* covering 707 unique car manufacturers and 2120 manufacturer-years for the period between 1895 and 1912. Firm level information collected from the Standard catalogue includes data on demographics (year of entry, year of exit, location), heredity (i.e., whether the firm was a startup or spinoff), capital that the firm possessed during formation, engine technology (gasoline, steam or electric), and mode of exit (bankruptcy, merger, or sale).

(c) Firm's annual production from Raff-Trajtenberg (1996) dataset. This dataset is based upon numbers reported in the *Standard Catalogue* and other sources for leading firms. While production numbers for leading firms are well known (Smith, 1968), the information for smaller firms is spotty. Sometimes only the total quantity the firm ever produced is known, but how those quantities are distributed across years is not. At other times, even this number is not known with certainty. In the former case, a distribution with a single production quantity peak is assumed across the years of known activity. In the latter case, the same technique was applied, but to a "guesstimated" number based on the qualitative description in the *Standard Catalogue*. While this implies a lack of precision, to our knowledge, this is the most comprehensive database of production quantities from the period.

Notes:

I. We defined an organization as an entity that manufactured cars for sale and did not include in our sample lone inventors who made experimental cars. To ensure that our sample is made up of entrants who were producing and offering products sold in the market, we followed Raff and Trajtenberg (1996) and cross-checked with two additional periodicals, *Automotive Industries* and *Motor*, which listed manufacturers displaying cars in the National Automobile Show in New York.

II. To ensure that we have a comprehensive dataset and that our result is not dependent on our criteria for inclusion, we complemented our dataset with data generously shared by scholars who have studied the early U.S. automobile industry. In a separate effort, Ioannis Ioannou (2014) constructed a dataset, similar to that of Steven Klepper, that tracked spinoffs in U.S. automobile industry. We refer to these datasets as the "Dobrev" dataset and the "Ioannou" dataset, respectively. We cross referenced our data with these two datasets and found 469 firms that were in either the Dobrev or Ioannou datasets but not in our sample of verified manufacturers. The unified dataset *in the paper*, which includes the Dobrev and Ioannou datasets and relaxes the restrictions of entry into the New York Automobile Show, is comprised of 1176 unique manufacturers and 3191 manufacturer-years. The *disclosed* data omits firms that are uniquely in the Dobrev dataset because those authors did not give us permission to disclose any parts of their data, and hence will have fewer than 1176 unique manufacturers reported in the paper. Note that per the discussion above, data set drops to 2120 manufacturer-years once we screen for actual evidence of manufacturing a car in the Standard Catalogue.

III. A firm is considered to be a verified firm if the *Standard Catalogue of American Cars* confirms that the firm manufactured cars for sale regardless of its inclusion in other datasets. Unless otherwise specified in the variable description, only verified firms are used in the construction of variables.

IV. This dataset only includes races that were reported in the *Horseless Age* magazine. It is possible that some races were conducted that was not reported in the magazine. However given the prominence of the magazine in the industry during the era, we are confident that the vast majority of the races, and certainly most of the important races, have been reported by the magazine and as a result included in the dataset.

5. Variables

Racing Data Variables	**Description**
appearances	Aggregate # of races focal firm participated through year t
attendance	Total attendance at races that the focal firm participated in year t
everraced	Set to 1 if the manufacturer has ever raced (as per coverage in *Horseless Age*) through year t
lap07	ln(1+ cumulative # of race appearances by the focal firm before 1907)
lattend	ln(Total attendance at races that the firm participated in year t)
lfp07	ln(1+ cumulative # of first place race finishes by the focal firm before 1907)
lnaccucontest	ln(1+ cumulative # of contests through year t)
lnappear	ln(1+ cumulative # of firm race appearances through year t)
lnappvics	ln(1+ cumulative # of first place race finishes by the focal firm through year t)
lnclimbwin	ln(1+ cumulative # of first place finishes in climbing contests through year t)
lnendwin	ln(1+ cumulative # of first place finishes in endurance contests through year t)
lnfirst	ln(1+ cumulative # of first place finishes through year t)
lnrelwin	ln(1+ Cumulative # of first place finishes in reliability contests through year t)
lnsecond	ln(1+ cumulative # of second place finishes through year t)
lnspeedwin	ln(1+ cumulative # of first place finishes in speed races through year t)
lnstartcap	ln(capitalization of the firm in the beginning)
lnstrengthwin	ln(1+ cumulative victories in endurance, climbing, reliability contests through year t)
lnthird	ln(1+ cumulative # of third place finishes through year t)
ltot	ln(1+ # of races in the firm's state of origin in year t)

missing_attendance	Number of races firm participated in year t without attendance data
neverwon	Set to 1 if manufacturer never won a race through year t
onedecile	# of races firms participated in year t that was in the top decile of attendance
onedecilewin	# of first place finishes in year t that was in the top decile of attendance
participationcount	Number of races firm participated in year t
race_firm_ratio	# of races in firm's state of origin in year t / # of in-state manufacturers in year t
raceperf	Set to 0 if the manufacturer never raced, 1 if the manufacturer raced but never won, and 2 if the manufacturer won at least once
startupvics	Ln (1+ cumulative # of first place finishes through year t by start up) = lnfirst*startup
startupvics1year	Cumulative # of first place finishes in year t by start up
totallocalraces	# of races in the firm's state of origin in year t
victories	Aggregate first place finishes by the firm through year t
year	Year of car race (between 1895 - 1912)

Firm Data Variables	**Description**
age	Firm age
annualnowinner	# of races that the firm participated which did not yield any winners (Note: Variable not used for analysis)
carmake	Name of automobile manufacturing firm
cohort	Entry cohorts used by Klepper (2007): 1 if entry <1905, 2 if entry between 1905 and 1909, 3 if entry between 1910 and 1912
cohort1	Set to 1 if firm entered the automobile industry between 1896-1904
cohort2	Set to 1 if firm entered the automobile industry between 1905-1909
cohort3	Set to 1 if firm entered the automobile industry between 1910-1912

dyear_YYYY	Set to 1 if the observation year is YYYY
exmerger	Set to 1 if firm exited through merger
exsale	Set to 1 if firm exited through sale
failure	Set to 1 if firm exited and shut down in year t+1
firmid	Unique numerical identifier assigned to each manufacturer
gastech	Set to 1 if the manufacturer offered gasoline powered car for sale
ioannisunmatched	Firm in Ioannis but did not produce according to the Standard Catalogue
iaonnis	Firm in Ioannis data
IN	Set to 1 if firm founded in Indiana
last	Set to 1 in the final year of appearance in the dataset.
lentries	ln(1+average number of firms that entered each race in a given year)
lnautosales	ln(1+ cumulative # of cars produced through year t)
lnfirst1year	
lnmodelct	ln(number of models launched by the firm in year t)
lnprodYM1	ln(1+lagged production)
lnprodYP1	ln(1+ # of cars produced in year t)
localeventsentered	# of in-state contest events that the firm participated in a year t (Note: Variable not used for analysis)
localracesentered	# of in-state races that the firm participated in a year t (Note: Each contest event can have multiple races; Variable not used for analysis)
lodwin	ln(1+Aggregate appearances in the top decile of races as measured by number of entries)
MI	Set to 1 if firm founded in Michigan
missstartcap	Set to 1 if startup capital information unavailable

modelcount	# of models firm launched in year t
multientryraces perfirm	# of races in which multiple cars of the same manufacturer competed in a given year
multientryperc	# of races in which multiple cars of the same manufacturer competed / total number raced the manufacturer participated in a given year
multiplenowinn erentry	# of races in which multiple cars of the same manufacturer competed that did not yield any winners (Note: Variable not used for analysis)
nowinnerperc	# of races that the firm participated which did not yield any winners / total # of races the firm participated in a given year
NY	Set to 1 if firm founded in New York
OH	Set to 1 if firm founded in Ohio
parenttopten	Set to 1 if atleast one firm founder worked for a top ten automobile manufacturer the year before firm founding.
pre1907	Set to 1 if year t <=1907
spinoff	Set to 1 if at least one of the firm's founders worked for an automobile firm prior to founding of current firm
startup	Set to 1 if the firm is a startup
state	Firm state
verifiedmanufa cturer	Set to 1 if *Standard Catalogue of American Cars* shows that the manufacturer actually produced cars for sale
young	Set to 1 if firm founded within the previous 2 years

Industry Data Variables	**Description**
acquisitions	Cumulative # of automobile firm acquisitions in the industry in year t
acquisitions2	acquisitions^2
deaths	Population deaths of automobile manufacturers in year t
deaths2	deaths^2
exits	# of automobile manufacturer exits in year t

exits2	exits^2
gdppercap	GDP per Capita (annual)
instatemanu	Total # of manufacturers in the same state as focal firm in year t
instateunverimanu	# of manufacturers in state in a given year that is in Iannou dataset but not in Standard Catalogue of American Cars
instateverimanu	# of manufacturers in state in a given year as per Standard Catalogue of American Cars
lnacquisitions	ln(1 + # of automobile firm acquisitions in year t)
lncumacquisitions	ln (1 + # of acquisitions of verified and unverified firms in year t)
lncumexit	ln (1+ # of verified and unverified automobile firm exits in year t)
lncummergers	ln (1 + # of verified and unverified automobile firm mergers in year t)
lnexit	ln(# of verified automobile firm exits in year t)
lnmergers	ln(1+ # of verified automobile firm mergers in year t)
mergers	# of verified mergers in year t
mergers2	mergers^2
popcumdensity	Number automobile firms active in year t
popden	Number of firms active in year t
popden2	popden^2