

Identifying Odometer Fraud: Evidence from the Used Car Market in the Czech Republic

Josef Montag

International School of Economics
Kazakh-British Technical University

Motivation

- ▶ Odometer fraud is believed to be omnipresent in the Czech used car market, but is also a problem elsewhere.
- ▶ But the evidence is only anecdotal.
- ▶ No research.
- ▶ This paper points out and attempts to fill this gap.

Why should we care?

The two markets: new and used cars

	Czech Republic 2014	France 2013	Germany 2013	United Kingdom 2013	United States 2015
New cars sold (millions)	0.2	1.8	3.0	2.3	16.8
Used cars sold (millions)	0.6	5.3	7.1	7.4	38.3
Used/new cars sold ratio	3.3	3.0	2.4	3.3	2.3
New car average price				£15,960	\$34,430
Used car average price				£5,770	\$18,500
Used/new price ratio				0.36	0.54
New car sales (billions)				£367	\$579.8
Used car sales (billions)				£427	\$708.1
Used/new sales ratio				1.16	1.22

Legal setting

Odometer fraud

- ▶ It is a specific crime in Belgium, Canada, Germany, and the US.
- ▶ In the Czech Republic it is punishable under the general Fraud Clause or the Consumer Detriment Clause, both with sentence up to 8 years.
- ▶ No enforcement.
- ▶ Only one case so far
 - ▶ A sale of VW Caravelle for 13,600 euro with
 - ▶ declared odometer reading of 171,000 km.
 - ▶ The seller bought the vehicle a year earlier for 7,000 euro with
 - ▶ a reading of 340,000 km.
 - ▶ One year prison sentence suspended for 18 months.

How is odometer fraud done?

It used to be done mechanically.

- ▶ Costly, as takes about an hour to roll back analog odometer.
- ▶ Leaves mechanic traces, easy to detect.

Nowadays it is done electronically.

- ▶ Cheap as it takes about 5 minutes.
- ▶ The mechanic plugs in a laptop and types in the new mileage.
- ▶ Leaves only electronic traces, not easy to detect.

This paper

- ▶ The aim was to explore the ways in which odometer fraud could be detected in data on used car for sale.
- ▶ I employ methods developed in election and accounting fraud literature.

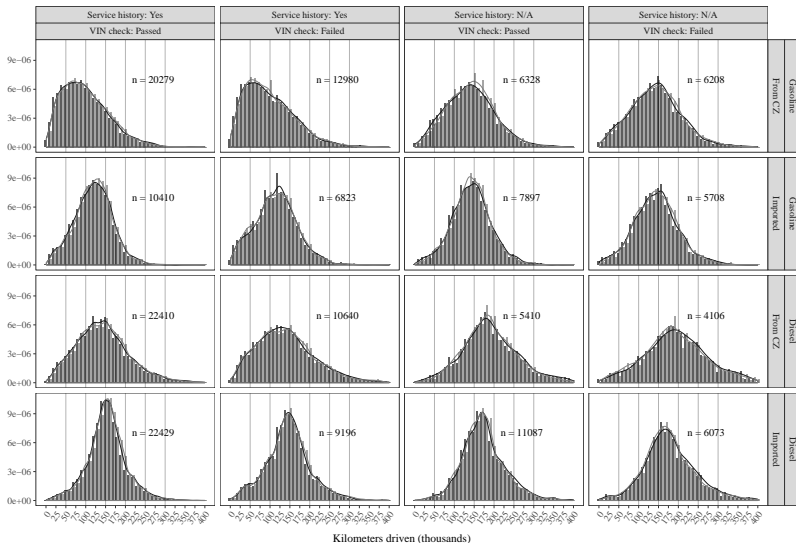
Data

- ▶ I have obtained a complete database of ads from a major Czech web site for used cars for sale in the Czech Republic.
- ▶ Data cover 22 months period between January 1, 2012 and November 5, 2013.
- ▶ The raw dataset contains 440,052 records with 38 columns
- ▶ There are 249,996 unique ads in total, identified by
 - ▶ VIN code, where available and valid (156,190),
 - ▶ a test based on duplicated values for 19 variables (93,806).
- ▶ I focus on cars 2 to 30 years old with mileage between 1000 and 400,000 km (213,090 observations).

Descriptive statistics

Statistic	Mean	St. Dev.	Min	Median	Max
Mileage (Km)	140,695.00	63,843.00	1,000	139,908.0	400,000
Year of production	2004.40	4.27	1982	2005	2011
Service history available (in %)	66.16	47.32	0	100	100
VIN check passed (in %)	62.48	48.42	0	100	100
From (in %)					
Czech Republic (=1)	42.18	49.38	0	0	100
Germany	21.49	41.07	0	0	100
Elsewhere	17.08	37.63	0	0	100
N/A	19.26	39.43	0	0	100
First digits					
Digit no. 1	2.49	2.48	1	1	9
Digit no. 2	4.23	2.76	0	4	9
Digit no. 3	4.13	3.07	0	4	9
Digit no. 4	2.39	2.99	0	0	9
Last digits					
Digit no. 1	1.78	2.82	0	0	9
Digit no. 2	2.00	2.90	0	0	9
Digit no. 3	2.59	3.03	0	1	9
Digit no. 4	4.62	2.92	0	5	9
Digit no. 5	4.56	2.76	0	5	9
Number of observations	213,090				

Distributions of odometer readings



Empirical approach

Digit tests used to detect fraud

- ▶ Two commonly used tests:
 - ▶ First Digit Test and
 - ▶ Last Digit Test.
- ▶ The distribution of digits in true data is known.
- ▶ At the same time people are bad randomizers.
- ▶ Deviations of digit distributions in reported data from the expected distributions suggest fraud.
- ▶ The First Digit Test is not valid in our setting.

Empirical approach

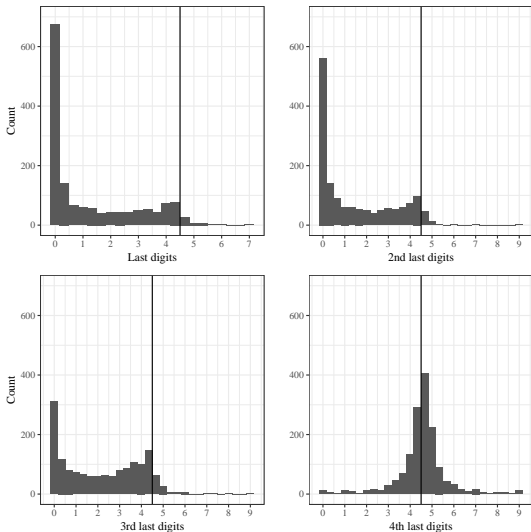
I use the Last Digit Test

- ▶ The last digits in count data should be uniformly distributed.
- ▶ Beber and Scacco (2012): almost no assumptions needed.
- ▶ Their Monte Carlo simulations show that last digit test works if:
 1. Standard deviation > 10 , and
 2. Mean $>$ standard deviation.
- ▶ Satisfied in our data for digits up to the level of thousands.

Problems

- ▶ Car may be driven between the rollback and sale (false negative).
- ▶ Odometer readings in ads are often rounded at various levels.
- ▶ Rounding results in deviation of last digits from random distribution (false positive).

Rounding patterns by sellers



A fix

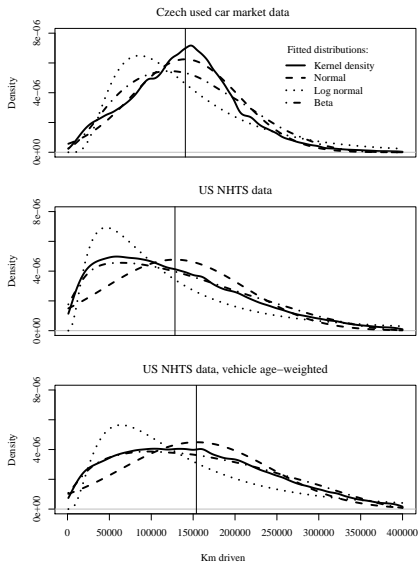
- ▶ Note that if a digit in a number is followed by a non-zero we are not looking at a rounded number.
- ▶ And so, except for the final digit itself, we can separate numbers with non-rounded digits.
 - ▶ Take **160,000**: Its fourth-last digit is zero, however it is not clear whether it is a true reading or whether or a result of rounding (and the following three digits are rounded as well).
 - ▶ Now take **160,405**: the fourth-last digit cannot be the result of rounding.
- ▶ Thus digit test are valid for a subset of data in with the digits to be tested are followed by a non-zero digits.

The validity of last digit test

Examining Beber and Sacco's conditions for last digit test validity

	Czech data		US data (NHTS)			
	Mean	St. Dev.	Mean	St. Dev.	Re-weighted	
					Mean	St. Dev.
Mileage (Km)	140694.8	63843.0	128312.6	83600.4	153866.2	88809.4
Last digit trimmed	13629.4	6224.1	9800.6	9645.8	12369.4	10806.1
Last two digits trimmed	1357.7	622.8	978.6	964.3	1235.2	1080.2
Last three digits trimmed	134.2	62.5	96.8	96.2	122.4	107.9
Last four digits trimmed	13.5	6.3	9.8	9.6	12.3	10.7

Monte Carlo simulations: fitted distributions

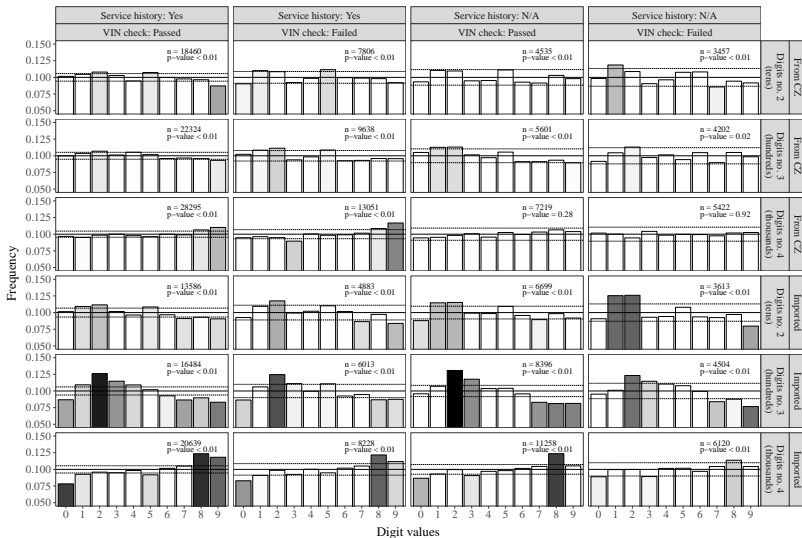


Monte Carlo simulations: p -values under rejection criteria (in %)

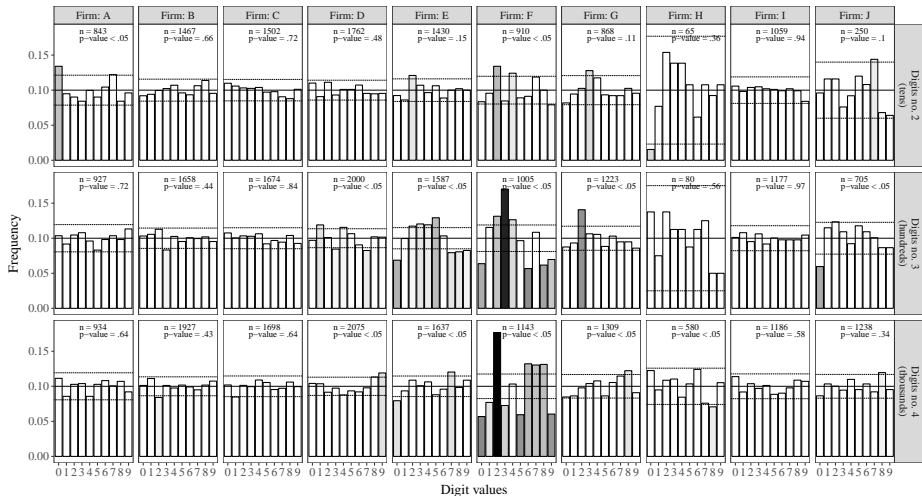
	Last digit (units)		Second-last digit (tens)		Third-last digit (hundreds)		Fourth-last digit (thousands)		Fifth-last digit (tens of thousands)	
	$p < 0.05$	$p < 0.01$	$p < 0.05$	$p < 0.01$	$p < 0.05$	$p < 0.01$	$p < 0.05$	$p < 0.01$	$p < 0.05$	$p < 0.01$
Sampling distribution:	Czech used car market data									
Normal distribution	4.74	1.08	4.74	0.88	4.96	1.10	5.36	1.08	100.00	100.00
Log-normal distribution	4.62	0.88	4.78	0.80	4.66	0.92	5.18	1.10	100.00	100.00
Beta distribution	4.96	1.06	5.18	1.02	5.44	1.16	5.04	1.00	100.00	100.00
Density distribution	4.88	0.98	4.86	1.00	4.68	1.12	5.54	1.00	100.00	100.00
	US NHTS data									
Normal distribution	5.24	0.88	5.62	1.06	4.60	1.06	6.64	1.02	100.00	100.00
Log-normal distribution	5.04	1.04	4.90	1.04	4.86	0.90	5.08	1.06	100.00	100.00
Beta distribution	5.26	1.14	5.12	1.00	5.04	0.88	8.12	1.84	100.00	100.00
Density distribution	5.44	1.18	5.34	1.18	4.92	1.04	5.48	1.40	100.00	100.00
	US NHTS data, re-weighted									
Normal distribution	5.14	1.18	4.72	1.04	5.30	1.00	5.62	1.18	100.00	100.00
Log-normal distribution	4.74	1.16	5.36	1.26	5.30	1.00	4.68	0.98	100.00	100.00
Beta distribution	5.74	1.00	5.18	0.94	5.10	0.76	5.06	1.12	100.00	100.00
Density distribution	5.06	0.96	5.26	1.02	5.58	1.00	6.00	1.32	100.00	100.00

Results

Last digits, given any of the next digits is different from zero



Last digits, given any of the next digits is different from zero for the ten sellers with the highest numbers of ads



Conclusions and questions for further research

- ▶ Suspicious patterns more frequent among imported vehicles.
- ▶ No other data cut (e.g. VIN code or service history availability) yields significant differences.
- ▶ Longitudinal vehicle data would be valuable in further research.
- ▶ Possibility to test Akerlof (1970).

Thank you.

