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

Montag, Josef

2015

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

Josef Montag†

This paper investigates the presence of odometer fraud in the used-car market in the Czech Republic using a unique dataset of 250,000 car-sale ads. Alternative identification techniques are also discussed. However, selection into the market as well as the practice of rounding odometer readings—possibly strategic yet innocent—render the standard statistical tests unusable. A modification of the last-digit test, which was previously used to detect fraud in election and accounting data, is therefore developed and employed. The results suggest that suspicious patterns are more prevalent in the segment of cars imported from abroad. I also show that this methodology can be used at the firm-level, which may be of interest to authorities and market participants.

Key words: used car market, odometer fraud, digit tests.
JEL classification: K42, R40.

1 Introduction

It is commonly believed that the Czech market for used cars is plagued with odometer fraud. However, this belief is based purely on anecdotal evidence, since no research into the phenomenon has yet been conducted. This paper aims to fill that gap.

Paradoxically, it is much easier to "roll back" the odometer reading on modern computerized vehicles that it was on to older cars with mechanical odometers. In an interview, a car mechanic involved in "clocking" stated that it takes about five minutes to roll back a digital odometer, independent of the extent of the rollback, whereas it takes about an hour

*I thank Pavel Čížek, Petr David, and Marek Litzman for helpful discussions. Any remaining errors should be attributed only to me. This research was funded by the Education for Competitiveness Operational Programme, project no. CZ.1.07/2.3.00/30.0031, co-financed by the European Social Fund and the state budget of the Czech Republic.

†Velký beranov 158, 58821 okr. Jihlava, Czech Republic Phone: +420 604 715 714. Email: josef.montag@gmail.com.

to roll back an analog odometer. This is because analog odometers have to be removed and then put back, whereas digital odometers can be accessed directly from a laptop computer via the car's diagnostic outlet.[1] At the same time it is harder to detect whether a digital odometer has been clocked, because no physical manipulation with the vehicle's inner parts is needed to do so and therefore no physical traces are left behind. Thus, arguably, the marginal cost of committing odometer fraud is in the region of zero.

It is also argued that the problem is worsened by the fact that, unlike in other countries, odometer fraud is not a specific criminal offense in the Czech Republic.[2] However, odometer fraud is punishable under the general Fraud Clause (§ 209 of the Czech Criminal Code) by up to two years of imprisonment (the maximum sentence increases to three years for repeat offenders and rises to eight years if committed by an organized group or for a large benefit). Alternatively it may be punished under the Consumer Detriment Clause (§ 253) by up to one year of imprisonment (rising to five years in the case of organized crime, recidivism, or if the offender received a substantial benefit; the maximum penalty rises further to eight years if a large benefit was involved). In reality, there has only been one conviction so far.[3] This is in stark contrast to the popular perception that altering odometers is a common practice on the used car market.

The purpose of this paper is to explore the ways in which odometer fraud could be detected, using a unique dataset of about 250,000 ads provided by a major Czech website on

[1]Zpověď mechanika: Stáčím tachometry, no a co? [Mechanic's confession: I roll back odometers, so what?] iDnes.cz, Online, August 27, 2007, at http://auto.idnes.cz/zpoved-mechanika-stacim-tachometry-no-a-co-fi6- (last accessed on January 12, 2015).

[2]Odometer fraud is a criminal offense in Belgium, according to § 1 of Article 3 of the Law Punishing Fraud in Vehicle Mileage (Loi réprimant la fraude relative au kilométrage des véhicules) of 2004, and is punishable by a fine and up to one year of imprisonment (Article 8); in Canada, according to section 27 paragraph (1) of the 1985 Weights and Measures Act, it is punishable by a fine and up to two years of imprisonment (section 35); in Germany according to § 22 of the Road Traffic Act (Straßenverkehrsgesetz) 2003, it is punishable by a fine or imprisonment of up to one year; in the United States according to § 32703 of the U.S. Code, Title 49, it is punishable by a fine or imprisonment of up to three years (§ 32709).

[3]A seller was convicted of fraud he committed in 2010 when he sold a Volkswagen Caravelle for 379,000 CZK (approximately 13,600 EUR) with a declared odometer reading of 171,000 km, although he had bought the vehicle ten months earlier for 7,000 EUR with a reading of 340,000 km. The resulting damage was estimated by the court to be at least 159,000 CZK (5,700 EUR) and the seller was sentenced to one year in prison, wholly suspended for 18 months. See the decision of Supreme Court of the Czech Republic no. 8 Tdo 728/2014-22, Online, February 7, 2014, at http://kraken.slv.cz/8Tdo728/2014 (last accessed on January 12, 2015).

which used cars are advertised for sale. More specifically, it looks at whether and how digit-based tests could be used for this purpose. These techniques have been established in the literature as tools for the detection of accounting fraud and election fraud (Beber and Scacco 2012; Debreceny and Gray 2010; Klimek, Yegorov, Hanel, and Thurner 2012; Mebane 2010).[4] However, selection concerns and the practice of rounding odometer readings in ads, which may possibly be strategic yet innocent in terms of the damage, substantially reduce the usability of these tests. In this paper I therefore employ a modification of the last-digit test, which is used on a subset of the data known to be unaffected by rounding. The results suggest that suspicious patterns are more prevalent among used cars imported for sale from abroad, rather than among cars registered domestically. I also show that the same methodology can be employed to reveal fraud at the level of individual firms selling used cars.

2 Data and Summary Statistics

This paper uses a unique dataset I have obtained from a company that runs one of the main Internet web services for advertising used cars for sale in the Czech Republic. At the beginning of this research, I approached all the main Czech companies running such websites with a data request for the purposes of a research into odometer fraud, and offered them in return the possibility of using this cooperation in their public relations or marketing. Only one company responded. Subsequently, they provided me with their complete and fully documented database of ads covering the 22-month period between January 1, 2012 and November 5, 2013.[5]

The raw dataset contains 440,052 records with 38 columns. Entries include an identifier for the seller, the region where the car is located, the dates when the ad was placed and when it ended, the car's characteristics (type of vehicle, producer, condition, engine displacement

[4]See also Benford (1938); Cho and Gaines (2007); Diekmann (2007); Jacob and Levitt (2003); Judge and Schechter (2009); Nye and Moul (2007); Rauch, Göttsche, Brähler, and Engel (2011); Varian (1972).

[5]The data and code producing the results reported in this paper are available upon request.

and power, type of chassis, number of seats and doors, color, year of production, fuel type, and a set of codes for car equipment), the car's mileage, the date of validity of its last technical check-up, the sale price, identifiers for whether the car has been crashed, whether the current owner is the first owner, whether the vehicle's service history is available, and the car's Vehicle Identification Number, if provided by the seller. Individual ads are placed on the website for a default period of 30 days, and can be renewed thereafter.[6] As a result, an ad may occur in multiple records, depending on whether it was placed only once or whether it was subsequently renewed. The data does not contain any information as to whether the seller is a private person or a company, nor does it contain a unique identifier for each ad.

My first task was therefore to clean up the dataset by removing duplicate records. The natural way to do so is to look for duplicates of the Vehicle Identification Number (VIN), which is a 17-character vehicle identifier carrying information about the manufacturer (digits 1–3), type (automobile platform, model, and body style, digits 4–9), and eight digits (digits 10–17) uniquely identifying the particular vehicle.[7] Because the VIN uniquely identifies each car, it can be used by buyers to check the information provided by the seller against public sources. However, the website does not require sellers to provide the VIN code, and as a result the field is empty in almost 168,000 records (38 percent of all records). In addition, some entries with the VIN "provided" in fact include nonsensical strings of digits, such as zeros, ones, or Xs, in place of a valid VIN code. In order to separate invalid VIN codes from valid ones, I designed a code that tests the structure and content of each VIN code in the data and returns an indicator that is equal to one if the item passed the test and zero if not.[8] The result is 200,447 records with a valid VIN code. After removing

[6]At the time of writing of this paper (January 2015), the fee for a standard ad was 562 CZK, which is about 20 EUR. The fee for an ad 'highlighted' with a photo of the car was 2,000 CZK, or about 72 EUR.

[7]The VIN is defined in the norm "ISO 3779:2009 Road Vehicles – Vehicle Identification Number (VIN)," by the International Organization for Standardization, Geneva, Online, October 15, 2009, at http://www.iso.org/iso/catalogue_detail?csnumber=52200 (last accessed November 26, 2014). ISO 3779:2009 specifies the content and structure of a VIN in order to establish a uniform identification numbering system for road vehicles worldwide. It applies to motor vehicles, towed vehicles, motorcycles and mopeds.

[8]The test consists of nine checks: check 1 identifies whether the item is nonempty; check 2 checks whether the length of the string is equal to 17; check 3 checks for the presence of letters I, O, and Q, which

duplicates, by keeping only the first record for each car identified by a valid VIN code, I obtain a dataset containing 156,190 unique records.

The rest of the dataset consists of 238,363 records for which the VIN is either incorrect or empty. Unique ads were identified in this subset in two steps: (i) I used the subset of ads with valid VIN codes to obtain a set of variables with identical values across duplicated records. Then, I selected variables for which the values were identical across duplicated records in 95 percent of cases; this resulted in 19 variables altogether. (ii) In the second step, I use those 19 variables to identify duplicate records for ads without a correct VIN code. The result is that 93,806 records without VIN code are identified as unique across the 19 variables.[9] This procedure results in a final dataset consisting of 249,996 unique used car ads in total.

The summary statistics of the final dataset are reported in Table 1. The average car in the dataset has an odometer reading of 123,000 km and was produced in 2005. Over two thirds of the ads state that the service history is available and almost two thirds of sellers provide a valid VIN code. Czech-produced Skoda vehicles are the most often offered brand, followed by Ford and Volkswagen. Almost half of the ads are for cars of Czech origin, that is cars that were registered in the Czech Republic under their last owner. About one fifth of the cars are imported from Germany, 15 percent are imported from other countries, and almost one fifth of ads do not state the car's origin.

The bottom two sections of Table 1 report descriptive statistics for the first four and last five digits of the odometer reading. The first digits should generally follow the Benford distribution, which has a mean equal to 3.44, however the mean in the data is 2.56.

do not exist in valid VIN codes; check 4 checks that only alphanumeric symbols are used; tests 5 and 6 check the World Manufacturer Indicator (WMI), that is the first three characters, against the two available WMI lists; check 7 checks whether the overall pattern of the VIN item fits the given structure; for manufacturers that use the check digit in their VIN codes (which is the ninth letter in each VIN code), check 8 computes the check digit and compares it with the check digit given in the item; and check 9 checks whether the WMI from the VIN code matches the manufacturer name given by the seller. A VIN code provided by the seller is designated as valid if it passes all nine checks.

[9]These variables are: type of vehicle, manufacturer, manufacturer and model code, manufacturer code (three letters) used by the website, World Manufacturer Indicator, VIN code, body style, model, year of production, power, fuel, maximum weight, all wheel drive, odometer units, tuning, number of remaining repayments, value of repayments, price including VAT, and price currency.

Similarly, the last digits should have a uniform distribution with a mean of 4.5, but the mean in the data is 2.84. However, these deviations may be explained by factors other than fraud, as discussed below. To further illustrate irregularities in the data, distributions of odometer readings are plotted in Figure 1. The data are split by service history, VIN code availability, origin, and fuel type. In some distributions there are apparent irregularities and spikes that would probably not occur in data with true odometer readings from a random sample of the car population. However, these might be the result of selection in to the used car market or strategic rounding by sellers.

3 Empirical Approach and Identification Issues

The most common strategies for identifying fraud in election or accounting data employ digit tests, which test either (i) the distribution of the first digits or (ii) the distribution of the last digits in each data entry. Both types of tests assume that true data (digits) are produced via a random process, which in turn implies distributional restrictions that can be used in testing the data at hand. Both types also exploit the fact that humans are imperfect randomizers and that if they are asked to generate random numbers, the resulting data will be systematically different from data produced by a true random process (Boland and Hutchinson 2000; Budescu 1987; Diekmann 2007; Mosimann, Wiseman, and Edelman 1995; Rapoport and Budescu 1997; Rath 1966).

First digits tests exploit the fact that first digits of a vector of (count) numbers should follow the Benford distribution (Benford 1938; Fewster 2009; Varian 1972).[10] This has been exploited in a number of studies in order to identify fraudulent data (Cho and Gaines 2007; Debreceny and Gray 2010; Judge and Schechter 2009; Mebane 2010; Mebane and Kalinin 2009; Nigrini 1999; Nye and Moul 2007; Rauch et al. 2011), but also criticized (Deckert, Myagkov, and Ordeshook 2011; Diekmann 2007). More specifically, this method is problematic because deviations from Benford's law may occur in true data if the data is

[10]Let D denote a digit and p a probability, then the Benford probability mass function is defined as $p(D = d) = \log_{10}(d + 1) - \log_{10}d$, where $d = 1, 2, \ldots, 9$.

the result of strategic choices (such as in elections, see Deckert, Myagkov, and Ordeshook 2011 and Mebane 2011) or if it results from deliberately divided population units (such as election districts, see Fewster 2009). It is very plausible that we would find the same to be true in our data because selection into the used car market is strategic in nature, as famously argued by Akerlof (1970). As a result, using a first digits test could suggest the presence of odometer fraud even where none was committed.

A more promising approach is to look at the distribution of the last digits. The idea behind this strategy is that the last digits of count data should follow a uniform distribution.[11] The assumptions that are required for this test are much weaker (see Beber and Scacco (2012) for a formal discussion and simulations). In summary, deviations from uniform distributions can be expected in data with very low standard deviation (e.g. standard deviation below 10) or data with a low mean relative to standard deviation, because such data consist of many small counts (Beber and Scacco 2012). Neither of these caveats is relevant for kilometers driven in our data, which have a mean of 123,000 km and a standard deviation of 79,000 km (see Table 1).

In addition, as noted in the Introduction and corroborated in my discussions with car mechanics, the way odometer fraud is committed today seems well fit for the kind of statistical testing used in election studies. In the old days, cars had mechanical odometers that had to be "rolled back" mechanically using a drill or some other tool. The drill can be viewed as a random number generator, and using this method, statistical testing would probably fail to identify any odometer fraud. Nowadays, however, odometer fraud is committed electronically using a laptop computer connected to the car, with a specialized software to manipulate the odometer readings to an exact sequence chosen by the user.

This setup would be ripe for a straightforward statistical examination if not for two things: (i) the rollback may occur in a different place than the place where the car is sold. If the car is driven between the place of fraud and the place of sale, the odometer reading will once again be generated by a random process; and (ii) people tend to round

[11] Let D be a digit and p a probability, then the uniform distribution is defined as $p(D = d) = 0.1$, where $d = 0, 1, \ldots, 9$.

the odometer readings they provide in their ads up or down, and they do so to different decimal places. Rounding, of course, leads to the last digits deviating from a random distribution. The frequency and pattern of rounding in the data can be seen in Figure 2, which plots histograms of the average values of the last digits computed for each seller. The plots of last, second last, and third last digits demonstrate that sellers commonly round up to whole thousands. However, as the plot of fourth last digits shows, units of thousands are not rounded.

Rounding for advertising purposes is likely to occur independent of any fraud and, even if strategic, is likely to be innocent in terms of the actual damage. This renders the last-digit tests hardly usable for the purpose of identifying fraud in the data as they are. Note, however, that if a non-zero digit follows a zero digit, that tells us we are not looking at a rounded number, and so except for the final digit itself, we can separate rounded from non-rounded zero digits. Take, for instance, the number 160,000. Its fourth-last digit is zero, however it is not clear whether zero was the true reading or whether it is a result of rounding and the following three digits are rounded as well. Now take the number 160,405, whose the fourth-last digit cannot be the result of rounding. Based on this, we can select the ads for which it is reasonable to assume that their odometer readings are not rounded, and obtain a subset of ads susceptible to the last digit test. This is then done using a simple Pearson's chi-squared test.[12]

4 Results

The results are shown in Figure 3, which plots the distributions of the second to fourth last digits where any of the next digits is different from zero. The data are split by the car's origin, fuel type, service history availability, and VIN availability. The p-values reported in each plot are the probabilities that the data follows a uniform distribution. The solid

[12]Let e_d be the frequency with which a digit d is observed and t_d the theoretical frequency given by a uniform distribution, where $d = 0, 1, \ldots, 9$. Then the statistic $\sum_{d=0}^{9} (e_d - t_d)\, t_d^{-1}$ asymptotically approach the χ^2 distribution with $n - 1$ degrees of freedom. For uniform distribution $t_d = \frac{n}{10}$, where n is the number of observations, the statistic simplifies to $\sum_{d=0}^{9} \left(e_d \frac{10}{n} - 1\right)$.

horizontal line represents the expected frequency of each digit in a uniform distribution, which is 0.1. Dashed lines represent bounds for deviations from the expected frequency that are still consistent with a uniform distribution, using a chi-square test and one-percent level of statistical significance. Bars that fall within the dashed lines are filled in white, whereas bars outside that range are filled in a shade that becomes darker with absolute distance from the respective bound.

The first three rows of Figure 3 relate to cars of Czech origin, i.e. whose last owner had the car registered in the Czech Republic. There is no clear difference in the results when comparing cars whose service history is available and those whose service history is unavailable, or when comparing cars with valid VIN codes to those without. For cars with Czech origin, it may appear that the last digits of the odometer readings conform with a uniform distribution better than for cars without both service history and VIN code than for cars with both types of information available. Note, however, that the sample sizes differ by a factor of five, so that the power to reject is less in the sample of ads lacking both service history and VIN code. Overall, the uniform distribution of last digits in the sample of Czech-origin cars is rejected at the one percent level in nine out of 12 subsets, which is more than chance would predict.

The picture is darker in when we look at imported cars, in the bottom half of Figure 3. Visually, the distributions of last digits deviate from uniformity more than the distributions in the top half of Figure 3. As a result, a uniform distribution is rejected in all 12 subsets. Note also that sample sizes are comparable across the corresponding categories of Czech-origin and imported cars, so this result is not an artifact of more statistical power. Overall, the results in Figure 3 suggest that there is greater cause for suspicion of odometer fraud in the imported car data that in the data for cars registered in the Czech Republic. This may not be surprising, as cars are generally imported by businesses, who may have easier access to the relevant technologies and skills. Furthermore, imported cars are disadvantaged on the market because their sale price must include the costs associated with their import on top of the cost at which they were acquired abroad. The free movement of goods within the EU and the fact that cars are tradable goods imply that opportunities for arbitrage are probably scarce and short-lived. This, in turn, implies comparable cars have a uniform value. This setting may give the seller further incentives to roll back the odometer, enabling them to sell an inferior product at the market price of a superior one.

The evidence seen so far, although conclusive, remains rather general. Would it be possible to identify suspicious patterns at the seller level? This is what Figure 4 looks at. It plots the distributions of the second to fourth last digits, where any of the next digits is different from zero, for the ten sellers with the highest numbers of ads. Notably, the odometer readings provided by firms B and C are quite consistent with uniform distributions for all three decimal places. Some other firms' readings, most notably Firm F's, seem to systematically deviate from the expected uniform distribution across the three relevant last digits. This means that there is cause for a suspicion of fraudulent odometer readings in cars offered by the firm F.

5 Conclusions

Using a unique dataset of 250,000 used car sale ads, this paper has applied statistical tests previously used to detect election and accounting fraud, in order to detect the presence of odometer fraud among used cars marketed in the Czech Republic. The simplest and most assumption-free test consists in looking at the distribution of the last digits of the odometer readings and test its consistency with a uniform distribution. However, selection concerns and the practice of rounding, which may be innocent, make the identification of fraud by this method more challenging in the case of used car ads. As a result, in order to avoid false positives, the tests were performed only for the second-to-last and higher order digits, and only on a subset of ads, in which at least one of the digits following the one tested is not zero. This ensured that the tested digit had not been affected by rounding.

Suspicious patterns were found to be more prevalent in the odometer readings for imported cars, than for cars registered domestically. No other data cut resulted in any systematic differences in patterns. My analysis also tested the data at seller-level, demonstrating that such tests may be of use to authorities and market participants, such as websites for used cars.

The results of the analysis in this paper are rather qualitative and inevitably include some false negatives. Further research using datasets from different countries might yield valuable insights. Ideally, the presence of odometer fraud would be evaluated on a dataset containing each car's full history. Such data are collected by certain private firms that offer car history checks to potential buyers of used cars, but could not be obtained for the purpose of this study. Both market participants and public authorities might well benefit from further research into odometer fraud based on car history data.

References

Akerlof, George A. 1970. "The Market for "Lemons": Quality Uncertainty and the Market Mechanism." *Quarterly Journal of Economics* 84:488–500.

Beber, Bernd and Alexandra Scacco. 2012. "What the Numbers Say: A Digit-Based Test for Election Fraud." *Political Analysis* 20:211–234.

Benford, Frank. 1938. "The law of anomalous numbers." *Proceedings of the American Philosophical Society* :551–572.

Boland, Philip J. and Kevin Hutchinson. 2000. "Student Selection of Random Digits." *Journal of the Royal Statistical Society. Series D (The Statistician)* 49:519–529.

Budescu, David V. 1987. "A Markov model for generation of random binary sequences." *Journal of Experimental Psychology: Human Perception and Performance* 13:25–39.

Cho, Wendy K. Tam and Brian J. Gaines. 2007. "Breaking the (Benford) Law: Statistical Fraud Detection in Campaign Finance." *American Statistician* 61:218–223.

Debreceny, Roger S. and Glen L. Gray. 2010. "Data mining journal entries for fraud detection: An exploratory study." *International Journal of Accounting Information Systems* 11:157–181.

Deckert, J., M. Myagkov, and P. C. Ordeshook. 2011. "Benford's Law and the Detection of Election Fraud." *Political Analysis* 19:245–268.

Diekmann, Andreas. 2007. "Not the First Digit! Using Benford's Law to Detect Fraudulent Scientific Data." *Journal of Applied Statistics* 34:321–329.

Fewster, R. M. 2009. "A Simple Explanation of Benford's Law." *American Statistician* 63:26–32.

Jacob, Brian A. and Steven D. Levitt. 2003. "Rotten Apples: An Investigation of the Prevalence and Predictors of Teacher Cheating." *Quarterly Journal of Economics* 118:843–877.

Judge, George and Laura Schechter. 2009. "Detecting Problems in Survey Data Using Benford's Law." *Journal of Human Resources* 44:1–24.

Klimek, Peter, Yuri Yegorov, Rudolf Hanel, and Stefan Thurner. 2012. "Statistical detection of systematic election irregularities." *Proceedings of the National Academy of Sciences* 109:16469–16473.

Mebane, Jr., Walter R. 2010. "Fraud in the 2009 presidential election in Iran?" *CHANCE* 23:6–15.

———. 2011. "Comment on "Benford's Law and the Detection of Election Fraud"." *Political Analysis* 19:269–272.

Mebane, Jr., Walter R., and Kirill Kalinin. 2009. "Comparative election fraud detection." Toronto, Canada: Annual Meeting of the American Political Science Association.

Mosimann, James E., Claire V. Wiseman, and Ruth E. Edelman. 1995. "Data fabrication: Can people generate random digits?" *Accountability in Research* 4:31–55.

Nigrini, Mark. 1999. "I've Got Your Number: How a mathematical phenomenon can help CPAs uncover fraud and other irregulaities." *Journal of Accountancy* 187:79–83.

Nye, John and Charles Moul. 2007. "The political economy of numbers: on the application of Benford's law to international macroeconomic statistics." *B.E. Journal of Macroeconomics* 7:1–14.

Rapoport, Amnon and David V. Budescu. 1997. "Randomization in individual choice behavior." *Psychological Review* 104:603–617.

Rath, Gustave J. 1966. "Randomization by Humans." *American Journal of Psychology* 79:97–103.

Rauch, Bernhard, Max Göttsche, Gernot Brähler, and Stefan Engel. 2011. "Fact and Fiction in EU-Governmental Economic Data." *German Economic Review* 12:243–255.

Varian, Hal R. 1972. "Benford's Law." *American Statistician* 26:65.

Table 1: Summary statistics

Statistic	Mean	St. Dev.	Min	Pctl(25)	Median	Pctl(75)	Max
Mileage (Km)	123,362	78,915	0	66,000	127,511	170,950	5,555,616
Year of production	2005.30	6.38	0	2002	2006	2009	2013
Service history available (in %)	68.56	46.43	0	0	100	100	100
VIN check passed (in %)	62.48	48.42	0	0	100	100	100
Make (in %)							
Audi	3.76	19.03	0	0	0	0	100
BMW	3.62	18.68	0	0	0	0	100
Citroen	4.17	19.98	0	0	0	0	100
Fiat	3.22	17.65	0	0	0	0	100
Ford	9.99	29.98	0	0	0	0	100
Mercedes-Benz	3.80	19.12	0	0	0	0	100
Opel	3.99	19.57	0	0	0	0	100
Peugeot	5.75	23.28	0	0	0	0	100
Renault	5.99	23.73	0	0	0	0	100
Skoda	23.73	42.54	0	0	0	0	100
VW	9.95	29.93	0	0	0	0	100
Other	22.04	41.45	0	0	0	0	100
From (in %)							
Czech Republic (=1)	46.99	49.91	0	0	0	100	100
Germany	19.33	39.49	0	0	0	0	100
Elsewhere	15.04	35.75	0	0	0	0	100
N/A	18.64	38.94	0	0	0	0	100
First digits							
Digit no. 1	2.56	2.49	0	1	1	3	9
Digit no. 2	4.17	2.80	0	2	4	6	9
Digit no. 3	4	3.10	0	1	4	7	9
Digit no. 4	2.40	3	0	0	0	5	9
Last digits							
Digit no. 1	1.84	2.84	0	0	0	3	9
Digit no. 2	2.08	2.92	0	0	0	4	9
Digit no. 3	2.71	3.06	0	0	1	5	9
Digit no. 4	4.59	2.92	0	2	5	7	9
Digit no. 5	4.35	2.77	0	2	4	7	9
Number of observations	249,996						

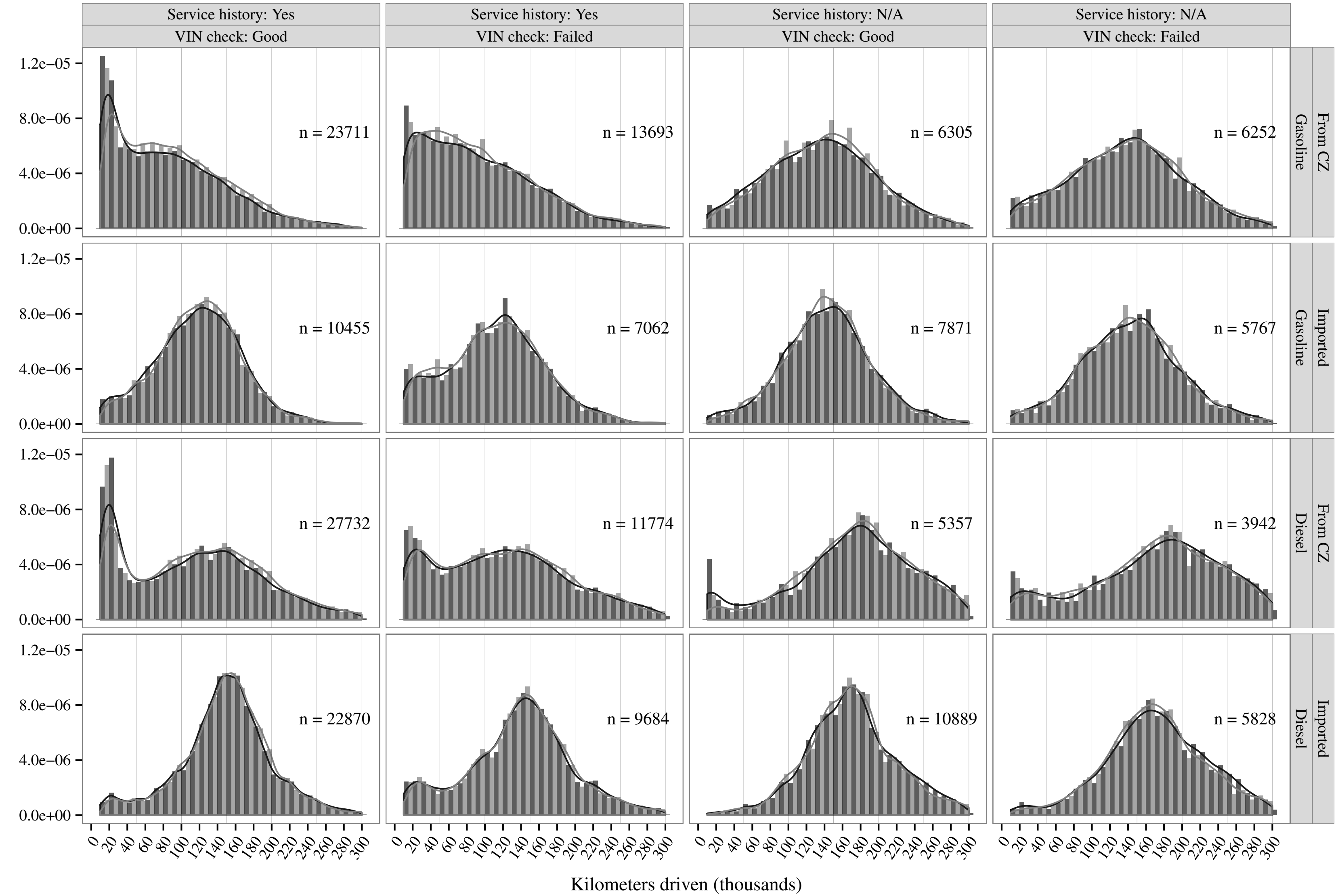

Figure 1: Distributions of odometer readings. Data are split by country of origin, fuel type, service history availability, and VIN availability. The bin width is 5,000 km; darker bars include data within the range 10,000 km to 14,999 km and its multiples, lighter bars include data within the range 15,000 to 19,999 km and its multiples. Data are capped at 300,000 km.

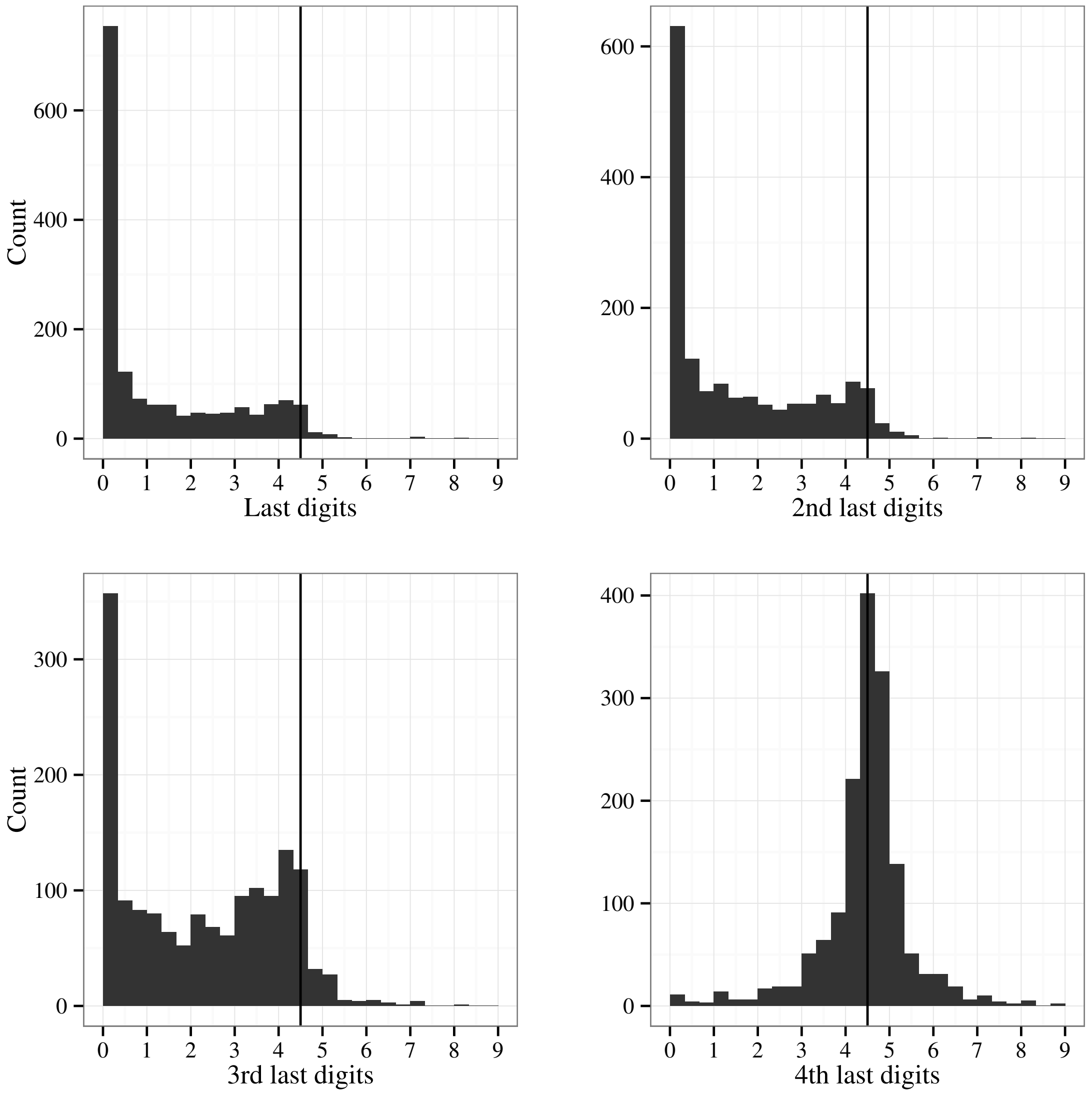

Figure 2: Rounding patterns by sellers. The unit of observation is a seller and the value is the mean of the respective last digits of the odometer reading computed across the seller's ads.

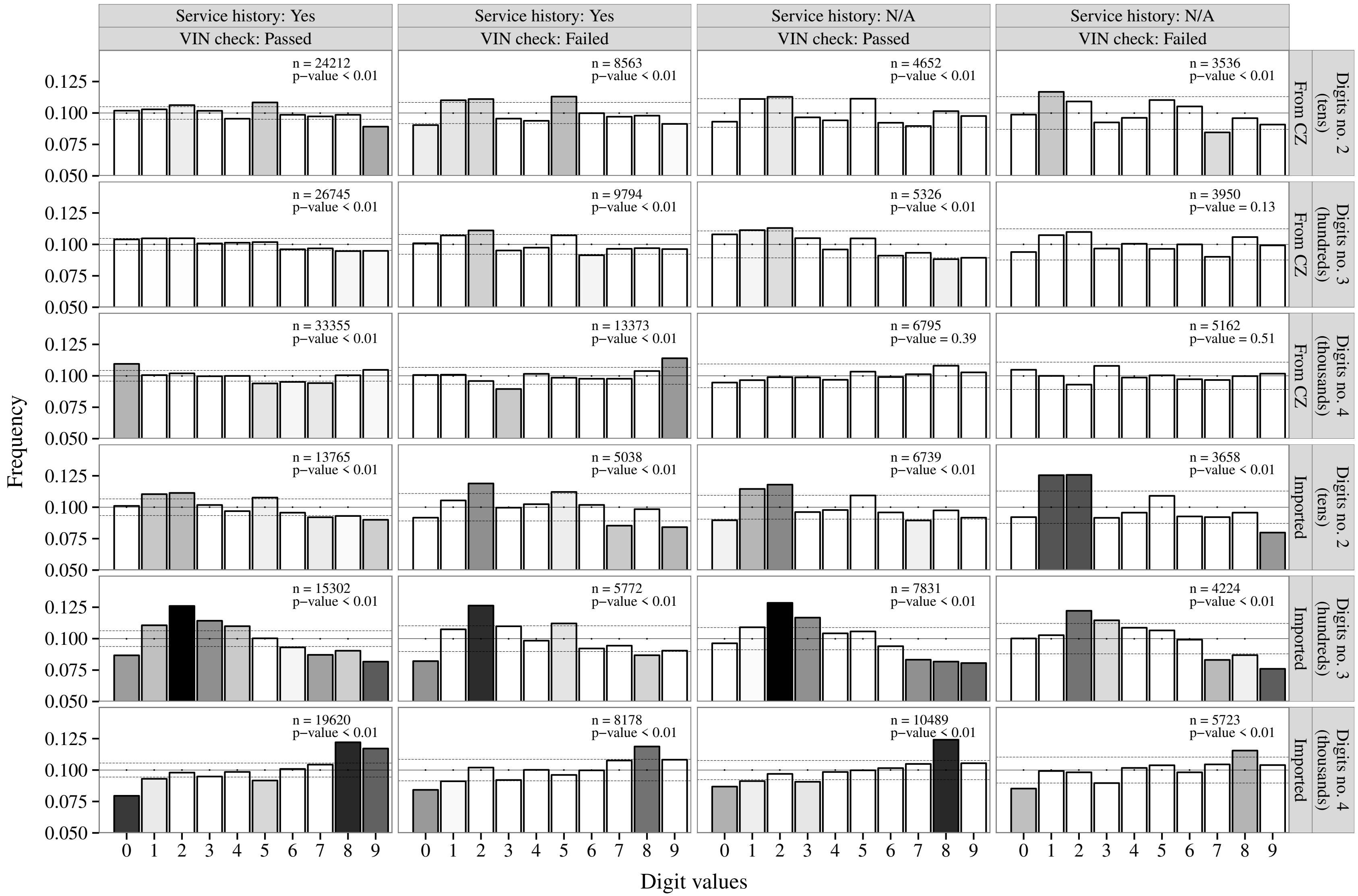

Figure 3: Distributions of last digits where any of the next digits is different from zero. The data are split by the car's origin, fuel type, service history availability and VIN availability. The p-values reported in each plot are the probabilities that the data follow a uniform distribution. Solid horizontal line represents the expected frequency of each digit in a uniform distribution, which is 0.1. Dashed lines represent bounds of deviations from the expected frequency that are still consistent with a uniform distribution using a chi-square test and one-percent level of statistical significance. Bars that fall within the dashed lines are filled in white. Bars outside that range are filled with a gray color, whose shade a is a function of the absolute distance from the respective bound.

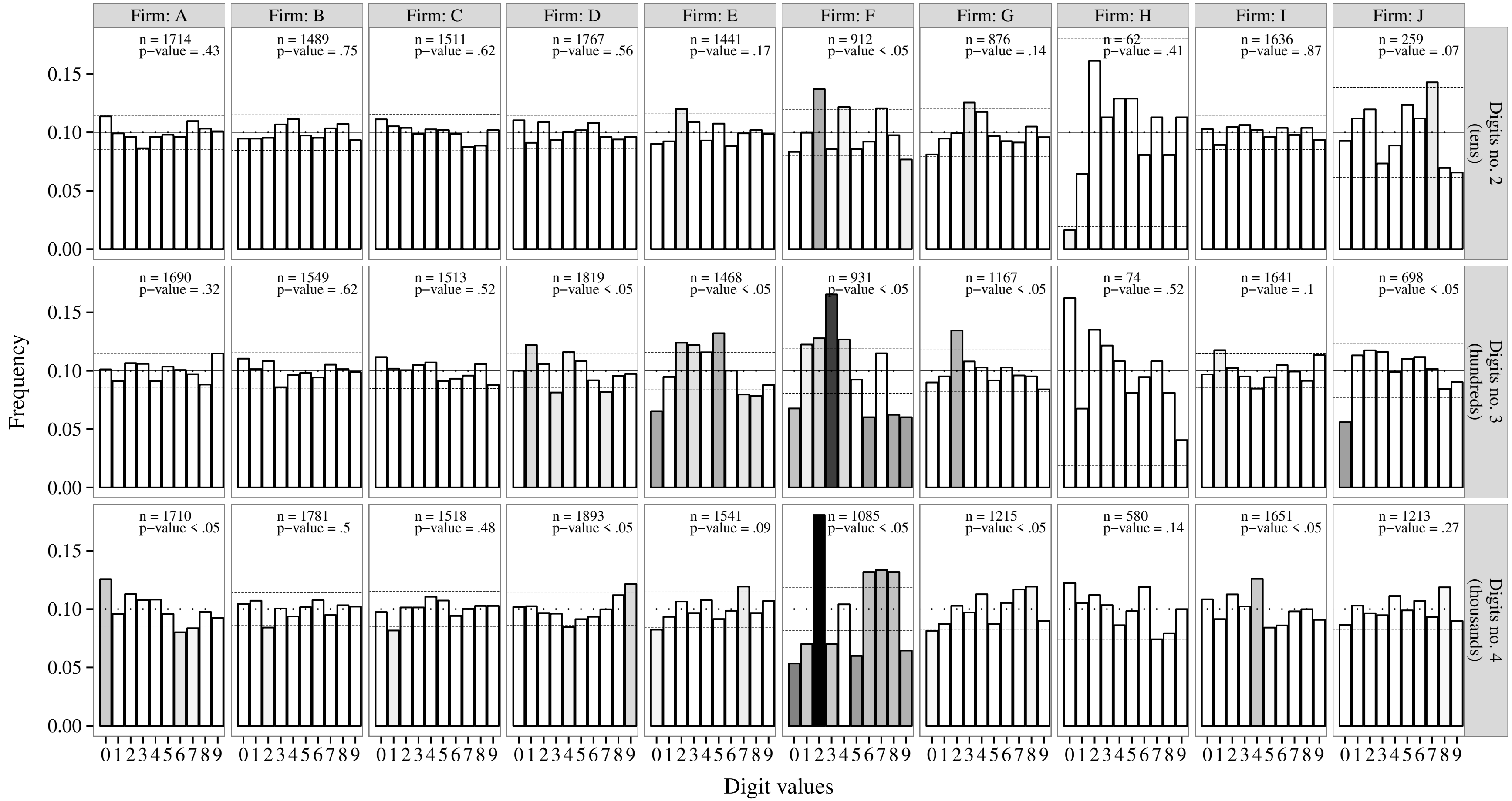

Figure 4: Distributions of last digits where any of the next digits is different from zero for the ten sellers with the highest numbers of ads. The p-values reported in each plot are the probabilities that the data follow a uniform distribution. The solid horizontal line represents the expected frequency of each digit in a uniform distribution, which is 0.1. Dashed lines represent bounds on deviations from the expected frequency that are still consistent with a uniform distribution using a chi-square test and five-percent level of statistical significance. Bars that fall within the dashed lines are filled in white. Bars outside that range are filled with a gray color, whose shade a is a function of the absolute distance from the respective bound.