WORKING PAPER

STATISTICAL FOOTPRINTS OF CORRUPTION: "VANITY FAIR" OF AUTOMOBILE LICENSE PLATES IN RUSSIA

Tom Eeckhout
Timur Natkhov
Leonid Polishchuk
Koen Schoors
Kevin Hoefman

November 2021
2021/1034

GHENT UNIVERSITY

Department of Economics

Faculty of Economics and Business Administration
Campus Tweekerken, St.-Pietersplein 5, 9000 Ghent - BELGIUM

STATISTICAL FOOTPRINTS OF CORRUPTION: "VANITY FAIR" OF AUTOMOBILE LICENSE PLATES IN RUSSIA*

Tom Eeckhout[1], Timur Natkhov[2], Leonid Polishchuk[2],
Koen Schoors[1], Kevin Hoefman[1]

[1]Department of Economics, Ghent University, Belgium.

[2]Department of Economics, HSE University, Moscow, Russia.

Abstract

We offer a novel big data approach to corruption detection and measurement by using statistical anomalies in publicly observable allocations which corruption affects in a predictable manner. While each individual incidence of corruption remains undetectable under the veil of secrecy, systemic corruption changes distributions of observable outcomes, and thus leaves measurable statistical footprints. We apply this approach to measuring corruption in Russian traffic police, which issues automobile license plates. Some of such plates serve as signs of status and prestige, and they are heavily concentrated among more expensive and especially luxury classes and brands, whereas if the official rules were followed, the distributions should have been close to uniform. Such discrepancies provide evidence-based measures of corruption in traffic police, which exhibit significant correlation with road accidents, injuries and fatalities.

Keywords: Corruption, Police, Law Enforcement, Administrative data, Forensic Economics

JEL-Classification: K13, K42, O17, P37

*We are grateful to Alexei Makarin, Maria Petrova, Susan Rose-Ackerman, Andrei Shleifer, and Daniel Treisman for useful comments and suggestions. We would also like to thank participants at IOEA (Cargèse 2017), LCSR April conference (Moscow 2018), AMEC (St. Petersburg 2018), ASEES (Boston 2018), RSSIA (Moscow 2019) and SIOE (Stockholm 2019) for valuable comments on earlier versions of this project. The usual disclaimer applies.

Introduction

Some human activities are illegal and/or contravene social norms, and as such hidden from public eyes. Social scientists, public officials, think tanks and watchdog organizations deploy various tools to assess the scale of such activities. A case in point is corruption, one of the most pernicious government pathologies, which retards economic growth, breeds poverty and inequality, renders public and private sectors dysfunctional, undermines economic and political systems, threatens environment, erodes social morale, etc. Corruption measurement usually makes use of surveys, field studies and expert opinions (Sampford et al., 2006). Corruption indexes derived from such sources could be susceptible to various biases of hard-to-gauge magnitude and even direction, and sometimes are inconsistent with each other (Treisman, 2007).

In this paper, we measure corruption by using statistical anomalies in the distributions of large arrays of publicly observable information. Such anomalies are corruption's statistical footprints, and their magnitudes could serve as objective, evidence-based, measures of corruption, which we derive for Russian regions. Transparency International ranks Russia among 30% of the most corrupt countries in the world, but there are considerable variations of corruption incidence, scope and burden across the vast country (Libman and Kozlov, 2013); (Baranov et al., 2015).

To estimate these variations by means of "forensic" statistical analysis, we make use of automobile license plates issued to motorists by the Russian traffic police. Some of such plates are more coveted than others among image-concerned Russians, and are sought after as signs of status and prestige. Russia does not have a vanity plate system, common elsewhere in the world, and the official procedure for the issuance of license plates provides for their release in an incremental numerical order on the first come-first served basis. In a major departure from this rule, we observe very high concentration of sought-after license plates among upmarket and especially luxury cars, as illustrated by Figure 1, indicating reallocation of such plates to wealthy car owners. During our observation period, there were no reasons and mechanisms other than corruption, that could have produced such statistical anomalies, and deviations of the observed distributions of preferred license plates in various regions from what would be expected if the official procedures were followed, provide objective measures of corruption in the regional traffic police service.

Such measures differ from one region to another, reflecting uneven corruption in regional police across Russia.

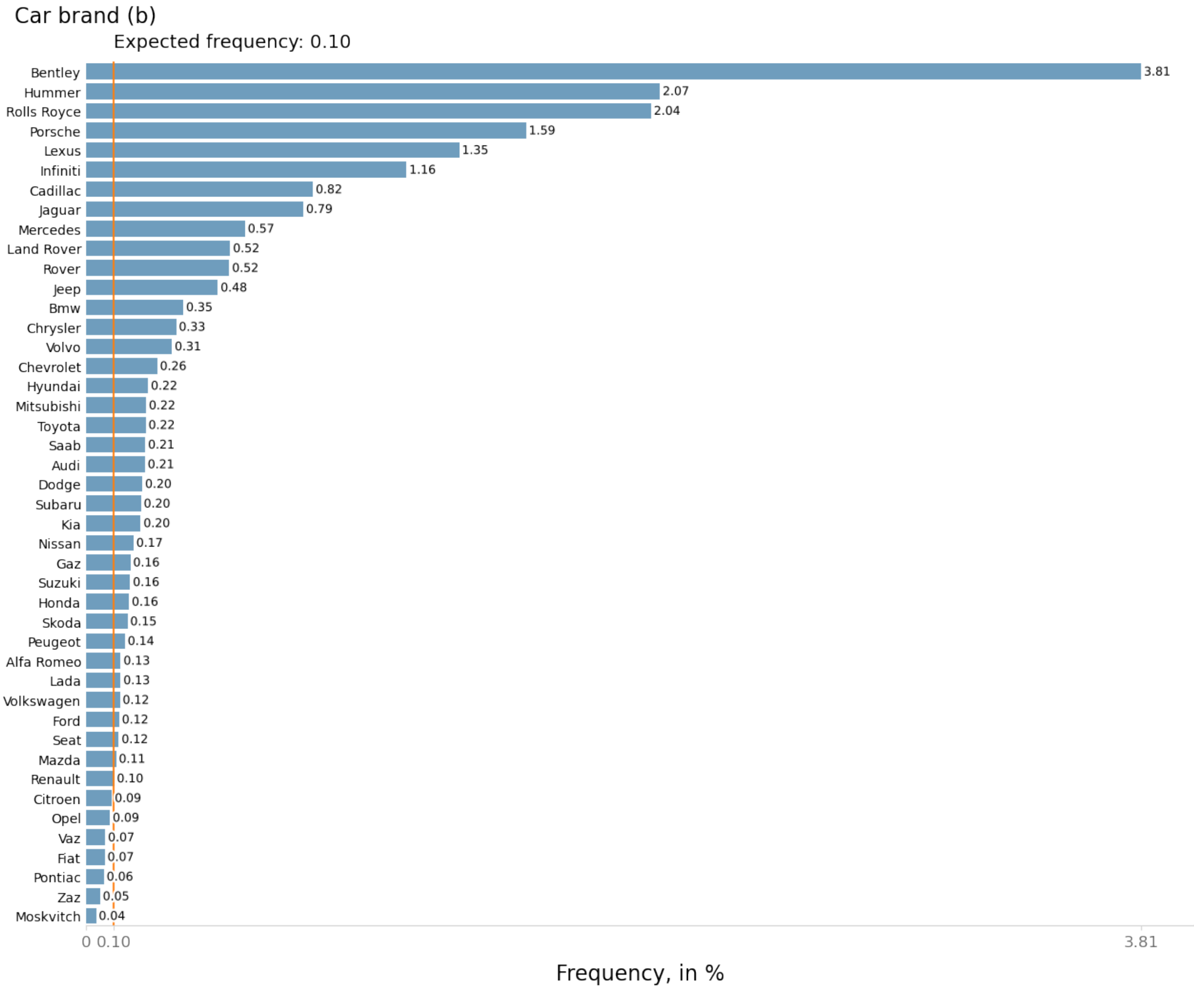

Figure 1: Frequency of license plates with "001" numerical combination across automobile brands in Russia, 1994-2011. Expected frequency without corruption is 1 in 999 or 0.10%. Data source: Russian car registration databases

We validate such objective evidence-based corruption measures by crosschecking them with other corruption indexes for Russian regions. We also observe strong correlation between the license plate-based corruption measures and traffic outcomes, especially those that are harder to conceal, such as road fatalities. Selling illegally prestigious license plates to highest bidders is in and of itself a relatively harmless form of corruption, except that it diverts a significant potential source of public revenue from the Treasury to corrupt officials. However, such "benign bribery" is a publicly observable tip of the iceberg, and signals an across-the-board breakdown in an important public service, which entails much graver losses of life and social welfare. We show that even a narrowly based corruption measure, supported by objective evidence and derived by rigorous analytical tools, could have significance and value for corruption prevention.

The paper demonstrates the potential of measuring massive illicit behavior by its statistical footprints. A single act of such behavior does not leave compelling evidence, but an established *practice* of such behavior biases observable distributions leading to evidence-based measures[1], which are precious few in corruption measurement.

How (not) to measure corruption

Corruption measurement is required to detect and assess corruption, design policy responses and evaluate their impact (Fisman and Golden, 2017). Since the 1990s, such measurement has been a "cottage industry" of development and governance studies. Transparency International was a pioneer, making its annual corruption rankings of countries of the world available since 1995. A growing number of think tanks and consultancies, including the Freedom House, Gallup International, Economist Intelligence Unit, etc., followed suit with their own indexes.

More often than not, such measures reflect *corruption perception* in expert and business communities and in the general population (Sampford et al., 2006);(Treisman, 2007). Corruption measures so derived are usually strongly correlated with each other, which is taken by their proponents as evidence of an underlying latent variable, i.e. corruption, which individual measures reflect, perhaps with some idiosyncratic noises. A less sanguine view has it that corruption perception measures represent opinions and expectations, rather than direct objective knowledge (Treisman, 2007), and as such could be influenced by various pre-conceptions. As a result, assessments of corruption are based on its assumed causes and consequences, such as a lack of democracy and political competition, economic backwardness, resource rent, and other "usual suspects", whereas more successful nations or subnational jurisdictions receive higher subjective scores due to a "halo effect" (Bardhan, 2004). This "reverse causality" devalues corruption perception measures as analytical tools. Cross-jurisdictional comparisons of such measures could be

[1]Statistical anomalies in Russian administrative data were used earlier to detect electoral fraud ("election forensics", see e.g. Klimek et al. (2012); Enikolopov et al. (2013)), corruption in public procurement (Mironov and Zhuravskaya, 2016), tax evasion (Braguinsky et al., 2014), etc. See also Fisman and Miguel (2007) for another example of using automobile and traffic data in corruption studies.

problematic due to differences in the local perception and framing. It is also possible that respondents are influenced by stereotypes prevailing in their communities, and could exaggerate corruption in a relatively clean public service (Oldenburg, 1987).

The second most popular approach to corruption measurement reflects actual *corruption experience* of survey respondents. As such, it is not susceptible to the above biases, and is recommended as a more reliable alternative to corruption perception measures. However, this approach has flaws of its own, such as reticence of respondents to give candid answers to sensitive questions about partaking in illegal activities. According to Kraay and Murrell (2016), this could cause a multifold downward bias of corruption experience measures, rendering this approach unusable. Furthermore, grassroots corruption experience does not reveal large-scale corruption (Olken, 2009).

Aggregation of corruption measures obtained from several independent sources could reduce biases of individual indicators. This approach has been employed e.g. in the Worldwide Governance Indicators project (Kaufmann et al., 2010), which since 1996 calculates annual Corruption Prevention Indexes for various countries. Knack (2006) cautions that such aggregation could lead to the loss of conceptual precision, making a single source data a better alternative. Also gains in statistical precision could be modest, if individual indexes are strongly correlated with each other, possibly due to common origins and biases.

Finally, corruption could be ascertained and estimated based on its various footprints, i.e. discrepancies between what is observed and what should be expected when government is clean ("the null hypothesis"[2]). Examples of such footprints are anomalies in the cost-benefit ratios in the public sector (Golden and Picci, 2005), gaps between official income and consumption of public servants (Gorodnichenko and Peter, 2007), reaction of suspected officials and agencies to anti-corruption campaigns (Di Tella and Schargrodsky, 2003), sluggish response to tariff reform (Sequeira, 2016), etc. When corruption is a massive grassroots phenomenon, it could produce statistical anomalies in publicly observable allocations, and the scale of such anomalies becomes a corruption measure. We derive such measure using the allocation of automobile license plates in Russia.

[2]This reflects a general approach in "forensic economics", whereby violation of a "null hypothesis" points to hidden and usually illicit behavior (Zitzewitz, 2012).

Russian license plates

An automobile license plate in Russia is a combination of alphanumeric characters in the following format: a letter, followed by three digits from 0 to 9, followed by another two letters. A license plate also has a region identifier, which is another two (sometimes three) digits. Figure 2 illustrates a Russian license plate (issued in the capital city of Moscow, as indicated by Moscow's regional identifier "197"). For a given region, the standard license plate format is "LDDDLL", where "L" is a letter, and "D" – a digit. The letter part of a plate in combination with the regional identifier indicates the plate's series. Each series potentially includes 999 license plates with various three-digit combinations ("000" combination is not used).

Figure 2: A Russian license plate. Source: gosnomerus.com

Automobile registration in Russia is a federal service, administered by regional branches of the national traffic police, which is subordinate to the federal Interior Ministry. According to a 2008 federal regulation, license plates are issued within a given series in their consecutive numerical order until the series is exhausted; the regulation disallows advance reservation of particular license plates for individual applicants. Until 2014, license plates were not transferable between automobiles and/or their registered owners, but after 2014, such transfers are possible through a sequence of transactions, some of them fictitious, which opened up a legitimate market for private license plate trade.

Russia does not have an official vanity plate system, and all license plates are issued for a flat nominal processing fee. Yet traditionally some combinations of letters and/or digits, popularly known as *krasivye nomera* – "nice numbers", thereafter *NNs*, are valued and sought after as visible signs of exception, status and prestige. Certain series could signal affiliation with powerful government agencies, while numerical combinations are valued if they stand out among ordinary draws, such as e.g. "00D" or "DDD". In some other countries or subnational units with similar sentiments and preferences, motor vehicle administrations auction off such numbers or make them available for a surcharge. The Russian system of license plate issuance has no such provisions, despite the willingness to pay for *NNs* by motorists who are sufficiently vain to want an *NN*, and sufficiently wealthy to afford one.[3] In what follows, we restrict our analysis to three-digit numerical parts of "nice numbers", which allows for cross-regional comparisons.

NNs in Russia could be obtained at source by luck within the established procedure, or by expending additional resources and efforts, e.g. paying the going market price at the secondary market, which has opened up in 2014. Market data point to a stratified (vertically differentiated) license plates market, where owners of expensive cars seek *NNs* in the premium category and owners of mid-market models make do with more generic *NNs*. Market stratification is sustained not only by the ability to pay for a high-end *NN* a price exceeding the cost of a cheaper car, but also by a complementarity between a car and a license plate in projecting status through conspicuous consumption (Truyts, 2010).[4]

The official procedure of car registration in Russia, where license plates are issued in a pre-set order on a first come-first served basis, should produce a close to uniform allocation of *NNs* across car classes and brands, but presently the secondary market reallocates *NNs* to more expensive cars. Recall however that until 2014 a secondary market for license plates in Russia did not exist, and at that time the only alternative to essentially winning a lottery in obtaining an *NN* was to resort to corrupt means – pay a bribe and/or use connections to circumvent the established order.[5] If higher concentrations of *NNs* among more expensive models were also observed before 2014, such lopsided distributions must had originated at source. Deviations that were far in excess of what could be expected if the rules were followed, provide *statistical* evidence of corruption at the issuing agencies. Below we elaborate this approach to corruption identification and measurement.

[3]According to the specialized web site `gosnomerus.com/plate/238495-A500AA197.html`, the asking price for the license plate shown on Figure 2 and considered in the premium category, at the time of writing was Ruble 3,000,000, or USD 42,000. Market prices for less prestigious *NNs* are lower, and "economy class" combinations are available for the equivalent of USD 1,000 and less.

[4]If a premium *NN* adorns a cheaper model, such mismatch would fail to impress the public, which would discount the plate as obtained by luck, and not view it as a signal of wealth or status of the owner. In a separating signaling equilibrium, such public expectations are rational, leading to a stratified market.

[5]Hence, another common colloquial name for such license plates – blatnye nomera, from "blat" – Russian slang for connections, influence, exchange of favors, etc. (Ledeneva, 1998).

Revealed preferences for license plates

One possibility to observe statistical anomalies pointing to corruption in the Russian automobile registration service is to look at the concentration of premium *NNs* in the premium segment of cars, where according to Figure 1 and in agreement with the presented above logic, such concentration should be particularly stark. However, corruption measures so derived would be strongly affected by the composition of car population, especially by the presence of luxury and other upmarket models, high in some Russian regions and nearly absent in others. This would make measures based solely on premium cars less suitable for *interregional* corruption comparisons. To avoid such biases, we develop a more encompassing approach, tracing statistical anomalies in the allocations of a broader set of *NNs*, where corruption could also be expected in multiple segments of a stratified market.

Numerical parts of *NNs* in Russia are valued for their rarity and/or aesthetic properties (such as same digit combinations ("111"), symmetric combinations ("101"), ascending or descending combinations ("123"), etc.). We use two independent sources of administrative and market information to identify a set of *NNs*, same for all Russian regions, that would be exclusive ("nice") enough to ensure willingness to pay by car owners, and at the same time broad enough to ensure that such transactions are sufficiently numerous for statistical inference even in less prosperous regions. In both cases, we employ the revealed preferences logic, standard in economics, whereby underlying preferences for particular license plates are deduced from observable transactions and market outcomes. We distinguish between periods before and after 2014. For the latter period, we use market information (Appendix A) on *NN* availability and trade to identify three-digit combinations which are (i) more common in such transactions, (ii) adorning more expensive cars, and (iii) being reallocated from cheaper to more expensive car models and classes.

For the period before 2014, when private trade in license plates was not possible and the market was based solely on corruption, we look for instances when car registration dates for particular license plates were breaking out of the established monotonic order prescribed by the regulation on the book. Such license plates were issued either earlier than those with numerically lower combinations in the same series (e.g. "555" issued before "527"), or later than those with numerically higher combinations (e.g. "666" issued after "689"). The first option reveals a "mining" pattern, whereby a preferred license plate is pulled ahead of time from those left in the series, whereas the second – a "holding" pattern, whereby a preferred plate, when it is reached in the numerical order, is stashed to be issued later to the right applicant. Our source of data in such analysis is information about over 70 million cars driven in Russia in the 1994-2011 period, available from various Internet sources (for more details see Appendix A)[6]. An observation in this dataset is a car with the following attributes: license plate number, brand, class (small, medium, large, executive, luxury, SUV, sports etc.) and the first registration date. Fig. B1, in Appendix B, presents results of such screening for randomly picked five license plate series allocated in a certain Russian region (not identified to avoid de-anonymization). Some numerical combinations, including "001", "002", "111", "222", etc., all the way to "999", clearly and consistently break out from the otherwise monotonic increase of the numerical combinations issued one after another.

From each of the above sources we derive alternative measures of preferences shown by Russian car owners for particular numerical combinations on license plates, assuming that corrupt officials cater to such preferences (for detailed descriptions of such measures see Appendix B). These preferences remained stable over time – popularity measures obtained for the periods – before and after 2014 – are strongly correlated with each other.[7] We aggregate several such measures to obtain a single license plate attractiveness index, and retain 56 numerical combinations with attractiveness above a certain threshold (the most sought-after numerical combination is "001", followed by "777" and "007"; full list is presented in Appendix B). This set of revealed *NNs* (thereafter *RNN*), provides a basis for corruption measurement.

[6]These and similar sources of Russian administrative data were used by a number of authors, including Mironov (2013), Braguinsky et al. (2014), Mironov and Zhuravskaya (2016).

[7]Opening up a market for private trade in license plates in 2014 did not eliminate corruption in the automobile registration agencies (see Footnote 9 below). Operators of trading platforms could be intermediaries between corrupt officials and their clients (reliance on intermediaries is a well-known means to reduce risks and other transaction costs of corruption; see e.g. Polishchuk et al. (2008)). However, such corruption does not leave direct attributable statistical footprints similar to those left before 2014.

Anomalies in license plate allocation and regional corruption measures

If corruption is suspected in the allocation of license plates in the *RNN* set, we should observe higher concentration of such license plates, issued before 2014, among premium cars. As explained above, owners of mid-market models also showed preferences for some *RNN* license plates. To capture all segments of the corruption-driven market, we consider various car categories, usually defined by a combination of car brand and class, e.g. Toyota SUV; some less numerous brands are considered as categories in their own without further subdivision to classes.

We use the above-mentioned dataset of 70 million plus cars driven in Russia over the 1994-2011 period, to calculate the frequency of *RNN* license plates issued in each category. If the official car registration rules were followed, and each car enters such process at a random time, the expected value of such frequency should be equal to the share of *RNN* combinations among all three-digit numbers, i.e. $\mathbb{E}_{RNN} = 56/999 = 5.6\%$. If in a given category, the observed frequency exceeds $\mathbb{E}_{RNN}$, it could indicate preferential treatment of car owners in this category, making them (relative) recipients of *NNs*. Vice versa, frequencies below $\mathbb{E}_{RNN}$ show that owners of cars in such categories were underserved with *RNNs*, being de facto *NNs*' donors. Recipients should be expected in the upper, and donors – in the lower segments of the car population, and this is what our data clearly show nationwide across car brands (Figure 3). A similar pattern is observed across car classes (Appendix C). For luxury cars, the observed frequency is more than three times the expected no-corruption average, and for top luxury brands, it is more than five times of what should be expected without corruption.[8]

A Chi-squared test decisively rejects the null hypotheses that such distributions could have occurred without interferences in a regular process with random entry of license plate applicants (p-value of the test is below a computational threshold of 2.22e-308). Similar tests for separate regions reject the non-interference hypothesis with comparable levels of confidence. As explained earlier, the only conceivable cause of the observed

[8]Gains of *RNNs* by donors and losses by recipients should not necessarily match each other due to the "mining" pattern whereby *RNNs* issued to bribing clients are pulled out from upper parts of a series, which could remain otherwise undistributed.

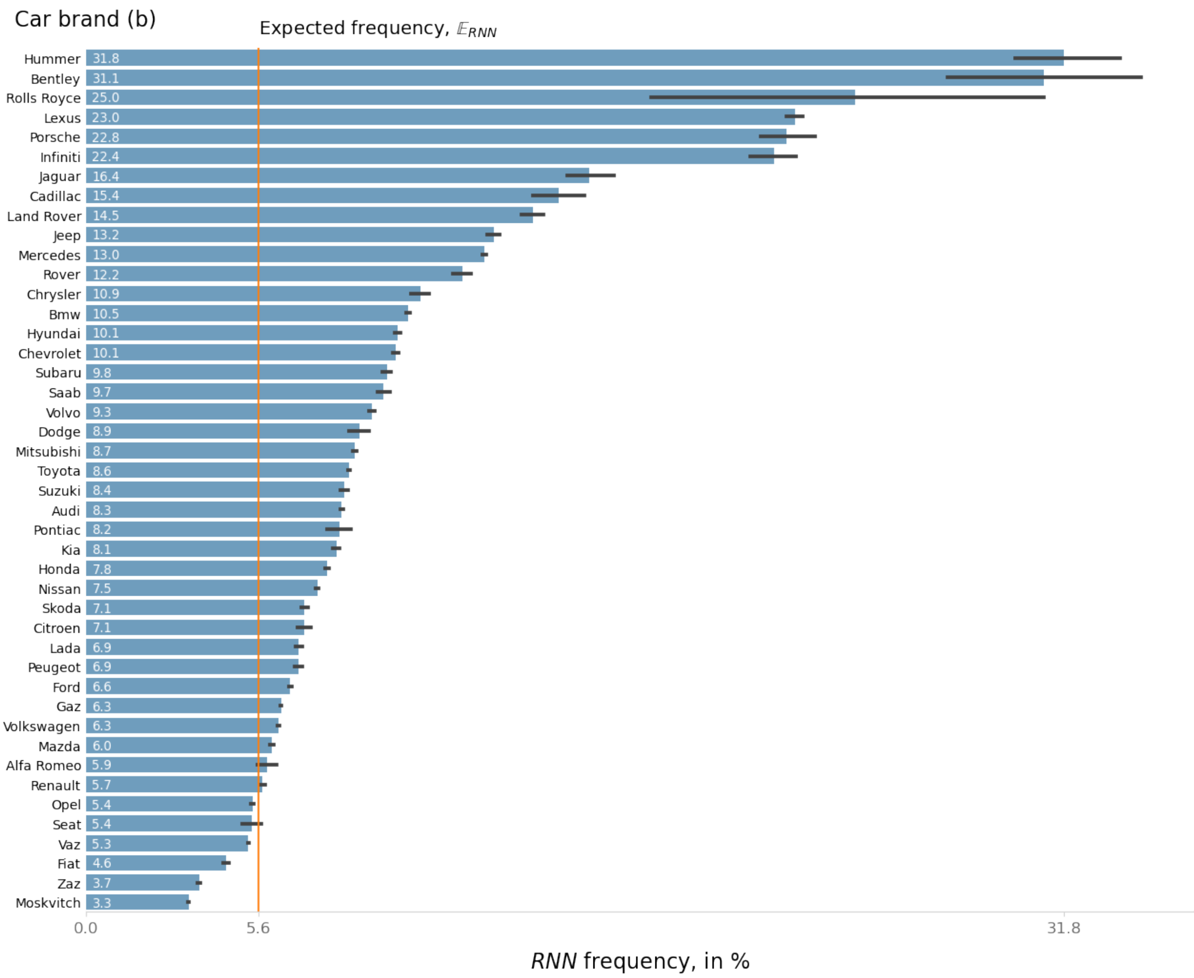

Figure 3: Frequency of *RNN* license plates across a selection of car brands. The horizontal line indicates $\mathbb{E}_{RNN}$. Narrow black vertical bars indicate 95% confidence intervals. Data source: Russian car registration databases.

distortions is corruption, which has thus left clear statistical footprints.

Obtaining statistical evidence of corruption in a Russian public agency is hardly surprising, given Russia's overall poor standing in corruption prevention, and the main analytical and policy value of our method is in seeing how these corruption-caused distortions, large as they may be, varied across the country, and whether they can serve as regional corruption measures. To this end, we calculate absolute values of the differences between the actual frequencies and $\mathbb{E}_{RNN}$ for all car categories in a given region, and take weighted averages across categories of such deviations, using shares of different categories in the regional car population as weights. After normalizing such weighted averages by dividing them over $2 * \mathbb{E}_{RNN}$, we obtain deviation indexes R_j (j stands for a region), which reflect *regional* anomalies in the allocations of RNN license plates. These are our raw measures of corruption in regional car registration agencies.

These measures need to be calibrated by deducting no-corruption benchmarks, which vary across regions, reflecting uneven concentrations of regional car populations across categories. Without corruption, in each category the actual share of RNN license plates would be the sample average for a binomial distribution, which according to the Law of Large Numbers approximates $\mathbb{E}_{RNN}$ as the sample size grows. Such averages would be closer to $\mathbb{E}_{RNN}$ for categories with many cars, and therefore the no-corruption benchmark would be lower when the car population in a region is concentrated in a few big categories with many cars in each, whereas if the car population is spread among a larger number of small and mid-size categories, the no-corruption benchmark would be higher. To account for such differences, we calculate no-corruption benchmarks R_j^0, following the procedure used to calculate R_j, except that the actual license plate number is replaced by a random draw from the uniform distribution over $[001 - 999]$. In doing so we use the actual sizes of categories in the regional car population. We average the results across 100 simulation runs for every region, and thus obtain the benchmarks R_j^0. The calibrated corruption measure for a region now obtains as $C_j = R_j - R_j^0$. Such corruption measures derived from the allocation of license plates (hereafter C_j) exhibit significant variation across Russia (figure 4), allowing for meaningful interregional corruption comparisons.

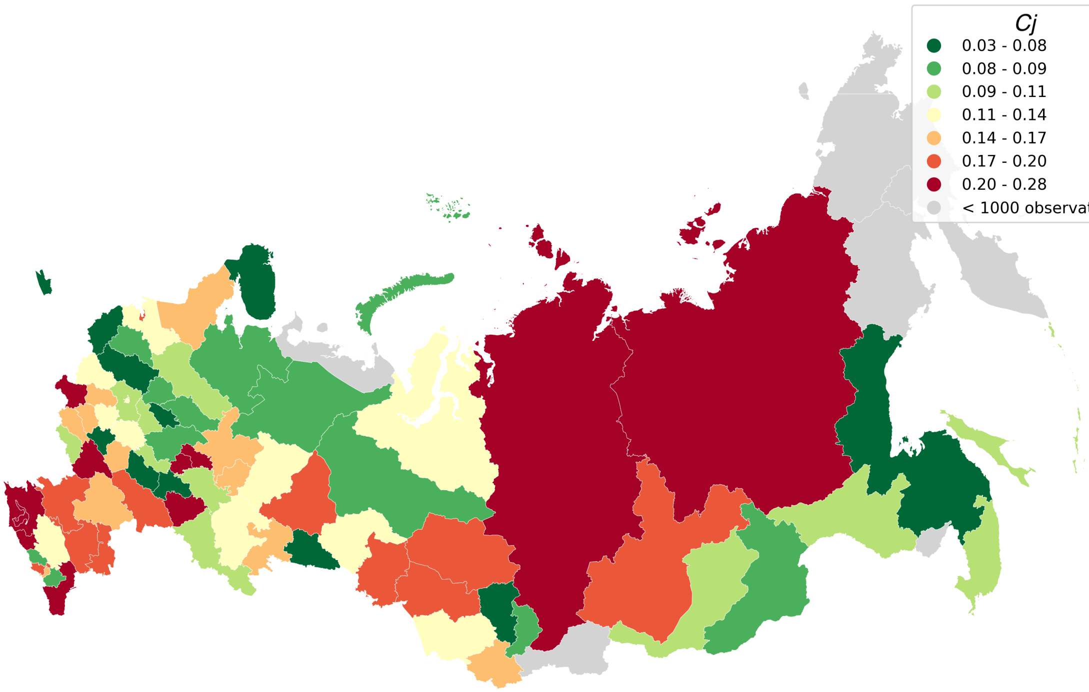

Figure 4: C_j corruption measures for Russian regions (the estimated share of all vanity plates that are redistributed in a region across car brands and classes). Data source: authors' calculations

Validation by other corruption measures

For external validation of the C_j corruption measure, we compare it to other independently derived corruption indexes for Russian regions. A variety of such indexes have been proposed (reviewed in Libman and Kozlov (2013)), most of them survey-based and reflecting corruption perception and/or experience. These regional measures are susceptible to the same biases as those similarly produced for countries of the world, and are subject to the same caveats. Correlation of such measures with the evidence-based C_j measure would corroborate the former, and at the same time show that the results of objective corruption measurement are in broad agreement with what other approaches suggest.[9]

One of the most popular and perhaps more reliable of such measures is the regional corruption index jointly produced in 2002 by the Russian chapter of Transparency International and INDEM think tank (Libman and Kozlov, 2013). While this index characterizes re-

[9]For a similar validation of a non-conventional corruption measure (using Internet document frequencies) see (Saiz and Simonsohn, 2013).

gional corruption at large, rather than in a particular government service, it should still be positively correlated with the C_j corruption measure, since corruption tends to spread across government agencies (Shleifer and Vishny, 1993), and various facets and manifestations of corruption usually reveal a general lack of accountability and transparency in the government. Indeed, the correlations between C_j and TI/INDEM corruption measures is 0.21 and significant at the 0.1 level. Still, these two measures are far from being identical and reveal significant discrepancies.

Our next validation test uses a corruption perception index based on a 2009 survey of the Georating project – a public opinion and attitudes polling program implemented in Russia since 2003 (Libman and Kozlov, 2013). Georating's surveys used large samples of Russian households, representative in every covered region of the country. Some surveys involved specific questions about corruption, and the one held in 2009 included a question about respondents' general concern about corruption in their lives. We take regional averages of the expressed corruption concern, and the correlation of such index with the C_j measure is 0.29, significant at the 0.05 level. Here we observe stronger correlation, both numerically and statistically, than with the TI/INDEM measure, because the 2009 Georating index reflected grassroots perception of corruption, which should be closer to the experience of dealing with corrupt traffic police, than the overall assessments of regional corruption by TI/INDEM. Russian capital Moscow is an outlier in this comparison, and without it, the correlation rises to 0.33 and becomes significant at the 0.01 level.

Finally, we use the same 2009 Georating survey, in which respondents were also asked a specific question about corruption in police and judiciary. Correlation of regional averages of the answers to this question with the C_j measure equals 0.36 (0.4 without Moscow) and is significant at the 0.01 level. Predictably, this correlation is even stronger than with two previous measures of corruption, because now both indexes reflect corruption in the same government service.

Therefore, as we zero in earlier developed corruption measures on the public sector's segment where the C_j measure is derived, we observe increasing proximity between the former and the latter (Figure 5). This could be seen as a bilateral validation of our approach and some of the alternative corruption measures.

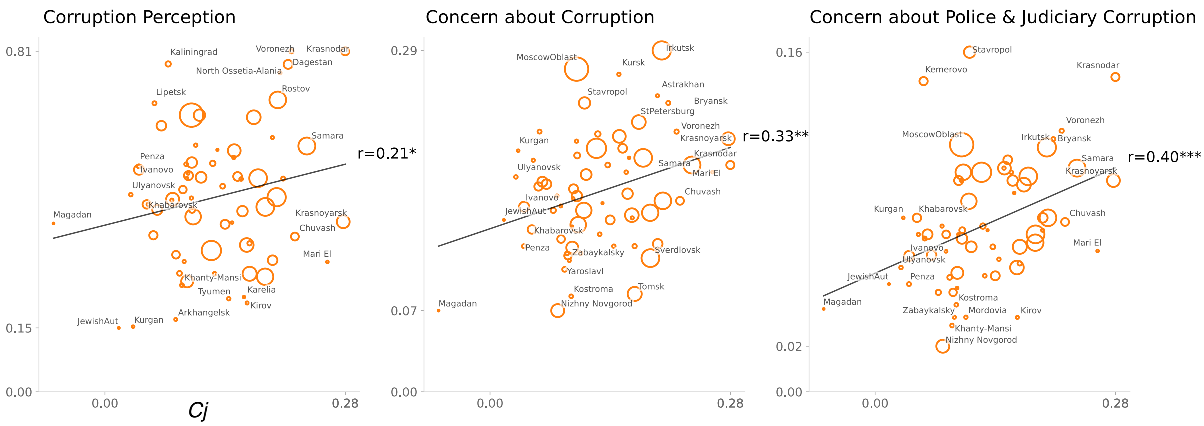

Figure 5: C_j vs earlier corruption measures (without Moscow). Scatterplots and Pearson correlations between C_j and Corruption Perception Index (a), Corruption Concerns (b) and Police & Judiciary Corruption Concerns (c). *, ** and *** correspond with 0.1, 0.05 and 0.01 significance of the correlation. Circle sizes are proportional to the logarithm of the number of observations available for a region. Sources: Transparency International and INDEM; Georating Project; authors' calculations.

Social cost of corruption

Bribery in the distribution of vanity plates is a relatively harmless form of corruption. According to the classification offered by Shleifer and Vishny (1993), this is an example of corruption without theft, as all official fees are fully collected and remitted. Such corruption still could be detrimental to social welfare, if it misallocates scarce resources controlled by the government, such as permits, licenses etc. Obviously, this was not the case when bribes were paid for *NN* license plates – on the contrary, until 2014 corruption substituted for a legitimate market for prestigious plates, making those available in accordance to the willingness to pay. Therefore one could argue that corruption, serving as a market-clearing mechanism (Rose-Ackerman and Palifka, 2016), produced a more efficient allocation of such license plates, especially if emotional components of utility reflecting vanity, prestige etc. are included in the social welfare calculus.[10]

A de facto market for *NN* license plates caused no social harm on either demand or supply sides (unlike other officially disallowed markets for endangered species, narcotics, human trafficking, body parts, votes etc.), and therefore no public case could be made against such market. The most likely reason it has not been yet officially endorsed in Russia, unlike most other countries of the world, is not a public interest, but rather a public choice argument (Djankov et al., 2002). The beneficiaries of corruption around *NN* license plates have a strong stake in preserving the status quo, and apparently managed to protect their source of illicit gains, despite decade-long attempts to introduce an official surcharge for such license plates, or auction them off to highest bidders.[11]

The stakes are high indeed. Our back-of-the-envelope estimation of the bribes collected over the 2000-2010 period, based on the data presented earlier in the paper and using published market prices for *NN* plates in 2017 from the specialized trading platform

[10]See Andreoni (2006) for a discussion whether utility derived from emotions should count in public decision-making.

[11]kommersant.ru/doc/1532276;ria.ru/20140930/1026282566.html; themoscowtimes.com/2013/06/04/personalized-license-plates-to-boostbudget-a24644; gazeta.ru/auto/2017/07/14_a_10784564.html; www.rbc.ru/society/20/01/2020/5e25593f9a7947630ee13410. In contrast, in the Russian mobile phone market, which is served by several private providers, customers can select "nice numbers" from official menus for an additional fee depending on the number's rarity, simplicity, aesthetic features etc.

nomera.net, produces the ballpark total of Ruble 17.2 billion, or 2020 USD 242 million. This is in all likelihood a conservative estimate, since it does not include license plates valued for their non-numerical parts.

The diversion of such considerable source of public revenues to corrupt officials is perhaps the only tangible downside of corruption as a means of quasi-market allocation of a scarce resource. However, this amount is not very large in the grand scheme of things, given the overall scale of corruption in Russia, estimated by various sources in the range of hundreds of billions of dollars.[12] Furthermore, at least some of that amount is a flow-through for the state budget, since corruption makes civil servants content with below-the-market wages (Rose-Ackerman and Palifka, 2016).

However, C_j corruption measures have far greater public significance than the volume of bribes paid for *NN* plates could suggest. License plates are selected as a measuring tool not because they are central to corruption in the Russian traffic police, but because this is where corruption leaves clear publicly observable and measurable footprints. Assessments based on such footprints, where license plates provide prima facie evidence, could be indications of greater harm caused by other kinds of corruption in the same service, which are however more difficult to ascertain directly. Examples of more deleterious corruption in the Russian traffic police include bribes paid to avoid official penalties for traffic rule violations or for sidestepping road use regulations (so that penalties fail to serve as a deterrence), extortion of tollbooth-like bribes (Shleifer and Vishny, 1993), etc.[13]

Russian traffic police is notorious for alleged corruption in its ranks[14], which corrodes the agency and adversely affects its performance. To the extent that C_j measures reflect overall corruption in the service, we should expect a significant correlation of this measure

[12]See e.g. an interview with Russia's Deputy Chief Prosecutor Alexander Buksman rg.ru/2006/11/07/buksman.html

[13]This can be illustrated by recent media reports (see e.g. bbc.com/russian/news-57903424) about corrupt traffic police officers in a Russian region, who enriched themselves by selling *NN* license plates and illegal passes for trucks exceeding weight limits and/or violating other road maintenance and traffic safety requirements. The second type of corruption, tied to the first one, is evidently far more damaging to social welfare.

[14]According to a 2018 nation-wide public opinion poll, road police is the second (after health care) most corrupt service in the country; wciom.ru/analytical-reviews/analiticheskii-obzor/korrupcziya-v-rossii-monitoring

	Accidents		Injuries		Fatalities		Traffic Violations	
C_j	2.396+	2.300	3.313+	2.201	0.669**	1.133*	-0.898	-2.435+
	(1.91)	(0.50)	(1.90)	(0.35)	(3.25)	(2.43)	(-1.45)	(-2.03)
Constant	13.92***	13.69***	16.86***	16.56***	2.271***	2.205***	0.507***	0.857**
	(60.65)	(23.96)	(55.36)	(21.75)	(67.47)	(35.50)	(5.92)	(3.56)
Region FE	✓	✓	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓	✓
Region-Year $\geq$ 1000 Obs	✗	✓	✗	✓	✗	✓	✗	✓
Regions	77	54	77	54	77	54	13	8
Region-Years	239	151	239	151	239	151	34	23

t statistics in parentheses

+ $p<.1$, * $p<0.05$, ** $p<0.01$

Table 1: Estimated effects of C_j on regional traffic accidents, injuries and mortality. All traffic variables are expressed per capita (per 10000).

with various traffic outcomes, which could be adversely affected by corruption in the traffic police. We test this hypothesis for four traffic outcomes – number of traffic accidents, injuries, fatalities and traffic rule violations, all normalized per 10,000 population. Russian state statistical agency tabulates such data annually for every region. To this end, we re-calculate C_j measures C_{jt} annually for years t from 2004 through 2011, and estimate the following panel regression:

$$TO_{jt} = \alpha * C_{jt} + \beta * X_{jt} + \epsilon_j + \delta_t$$

,where TO_{jt} is a traffic outcome index in region j and year t, with region and time fixed effects and controls X_{jt}. Estimation results for the above listed traffic outcomes are presented in Table 1 (the results are shown without controls X_{jt}; to check robustness, we add various controls reflecting climate, road density, demographics, etc., which leaves estimation results qualitatively intact).

We find statistically significant increases in the numbers of accidents, injuries, and especially fatalities when C_j corruption measure rises, which confirm the above hypothesis. Notice that this association is particularly sharp (0.05 significance, rising to 0.01 when we allow region-years with fewer than 1000 observations) in the case of road fatalities, which are most difficult to conceal or misreport. It is also noteworthy that the number of traffic violation is negatively correlated with C_j measure, which is in all likelihood an indication of more frequent settlement of traffic violations on the spot by paying a bribe, in which case the incident is not entered into the official records. Therefore, we observe a power of the C_j measure to predict corruption at large in the traffic police, and the grave damage that it causes.

Discussion

Corruption, as any other illegal activity, is shrouded in secrecy and usually requires in-depth investigation to be exposed and proven. Public law-enforcement agencies responsible for criminal investigations cannot always be trusted with fighting corruption, in which they themselves could be dealing. Political opposition, independent media and investigative journalism are often suppressed by corrupt regimes, making corruption's shielding from public eyes nearly watertight. This greatly complicates corruption measurement, as acknowledged in the vast literature on the subject. Various palliative solutions are not evidence-based and/or suffer from multiple biases, leaving development, governance and political economy studies, as well as governments and societies, without reliable information on a matter of central importance.

We offer a solution to this problem by making use of the simple fact that corruption carries favors to willing-to-pay customers, leading to inequality in the allocation of scarce resources. If such resources are traded on legitimate markets, unequal allocation is the standard market outcome reflecting general wealth inequality. However, if resources are controlled by government and are supposed to be distributed in an equitable manner, but in reality, the distribution is skewed towards greater wealth, such inequity becomes a corruption footprint. To make it a reliable basis for corruption measurement, several conditions have to be met. First, allocation decisions (or their essential parts) should be in the public domain and visible to third parties. Second, some indicators of individual wealth should also be publicly observable, to be matched against allocation decisions. And third, the population of resource recipients should be large enough to enable statistical tests that would reject with high confidence the "null hypothesis" of a clean government service, and provide statistical benchmarks to be used to measure the magnitude of the observed biases caused by corruption.

All of these conditions are present in the case of Russian automobile license plates. Preferred license plates are not only observable to the public, but in fact are intended and expected, in the spirit of conspicuous consumption, to broadcast status and prestige. Model and/or class of a car that carries such license plate is also publicly observable and is a reliable wealth proxy. Official procedures for license plate distribution do not provide for preferential treatment of any kind of motorists, and in the absence of private markets, the observed concentration of the preferred license plates among expensive cars indeed reveals corruption. Finally, there were tens of millions of license plates issued in Russia during the observation period, allowing statistical corruption measurement with high degree of precision.

Our analysis reveals massive corruption in the Russian traffic police, manifested by staggering concentration of preferred license plates among upmarket cars. The likelihood of spotting a coveted license plate on a luxury car could be 20-30 times higher than for popular models. This highly uneven allocation mirrors overall wealth inequality in Russia, by some estimates one of the highest in the world (Novokmet et al., 2018). Fueled by corruption, a routine administrative procedure became a multi-million dollar "vanity fair" for the Russian wealthy, with some spoils also left for the middle class. Volumes of transactions on such "fairs", properly normalized and calibrated, exhibit significant variations between Russian regions and serve as regional corruption measures.

Arguably, satisfying one's craving for vanity is not a fundamental human need, and inequality in access to preferred license plates in and of itself should not be a matter of serious concern. A flip side of this observation is that bribery in the distribution of automobile license plates is not a particularly pernicious form of corruption. However, corruption is rarely confined to isolated activities of government agencies, and although regional corruption measures offered in the paper are rather narrowly based, they signal systemic breakdowns in regional traffic police, which impose far greater costs on the society.

References

Andreoni, James, "Philanthropy," *Handbook of the economics of giving, altruism and reciprocity*, 2006, *2*, 1201–1269.

Baranov, Alexey, Egor Malkov, Leonid Polishchuk, Michael Rochlitz, and Georgiy Syunyaev, "How (not) to measure Russian regional institutions," *Russian Journal of Economics*, 2015, *1* (2), 154–181.

Bardhan, Pranab, *Scarcity, conflicts, and cooperation: essays in the political and institutional economics of development*, MIT Press, 2004.

Braguinsky, Serguey, Sergey Mityakov, and Andrey Liscovich, "Direct Estimation of Hidden Earnings: Evidence from Russian Administrative Data," *The Journal of Law and Economics*, 2014, *57* (2), 281–319.

Djankov, Simeon, Rafael La Porta, Florencio Lopez de Silanes, and Andrei Shleifer, "The regulation of entry," *The quarterly Journal of economics*, 2002, *117* (1), 1–37.

Enikolopov, Ruben, Vasily Korovkin, Maria Petrova, Konstantin Sonin, and Alexei Zakharov, "Field experiment estimate of electoral fraud in Russian parliamentary elections," *Proceedings of the National Academy of Sciences*, 2013, *110* (2), 448–452.

Fisman, Raymond and Edward Miguel, "Corruption, norms, and legal enforcement: Evidence from diplomatic parking tickets," *Journal of Political economy*, 2007, *115* (6), 1020–1048.

_ **and Miriam Golden**, "How to fight corruption," *Science*, 2017, *356* (6340), 803–804.

Golden, Miriam A and Lucio Picci, "Proposal for a new measure of corruption, illustrated with Italian data," *Economics & Politics*, 2005, *17* (1), 37–75.

Gorodnichenko, Yuriy and Klara Sabirianova Peter, "Public sector pay and corruption: Measuring bribery from micro data," *Journal of Public economics*, 2007, *91* (5-6), 963–991.

Kaufmann, Daniel, Aart Kraay, and Massimo Mastruzzi, "The worldwide governance indicators: Methodology and analytical issues," *World Bank policy research working paper*, 2010, (5430).

Klimek, Peter, Yuri Yegorov, Rudolf Hanel, and Stefan Thurner, "Statistical detection of systematic election irregularities," *Proceedings of the National Academy of Sciences*, 2012, *109* (41), 16469–16473.

Knack, Stephen F, *Measuring corruption in Eastern Europe and Central Asia: A critique of the cross-country indicators*, Vol. 3968, World Bank Publications, 2006.

Kraay, Aart and Peter Murrell, "Misunderestimating corruption," *Review of Economics and Statistics*, 2016, *98* (3), 455–466.

Ledeneva, Alena, *Russia's economy of favours: Blat, networking and informal exchange*, Vol. 102, Cambridge University Press, 1998.

Libman, Alexander and Vladimir Kozlov, "Sub-national variation of corruption in Russia: what do we know about it?," *REGION: Regional Studies of Russia, Eastern Europe, and Central Asia*, 2013, *2* (2), 153–180.

Mironov, Maxim, "Taxes, theft, and firm performance," *The Journal of Finance*, 2013, *68* (4), 1441–1472.

_ **and Ekaterina Zhuravskaya**, "Corruption in procurement and the political cycle in tunneling: Evidence from financial transactions data," *American Economic Journal: Economic Policy*, 2016, *8* (2), 287–321.

Novokmet, Filip, Thomas Piketty, and Gabriel Zucman, "From Soviets to oligarchs: inequality and property in Russia 1905-2016," *The Journal of Economic Inequality*, 2018, *16* (2), 189–223.

Oldenburg, Philip, "Middlemen in third-world corruption: Implications of an Indian case," *World politics*, 1987, *39* (4), 508–535.

Olken, Benjamin A, "Corruption perceptions vs. corruption reality," *Journal of Public economics*, 2009, *93* (7-8), 950–964.

Polishchuk, Leonid, Oleg Shchetinin, O Shestoperov et al., "Intermediaries between Private Sector and the Government: Helping Business or Participating in Corruption?," *VOPROSY ECONOMIKI*, 2008, *3.*

Rose-Ackerman, Susan and Bonnie J Palifka, *Corruption and government: Causes, consequences, and reform*, Cambridge university press, 2016.

Saiz, Albert and Uri Simonsohn, "Proxying for unobservable variables with internet document-frequency," *Journal of the European Economic Association*, 2013, *11* (1), 137–165.

Sampford, Charles, Arthur Shacklock, Carmel Connors, and Fredrik Galtung, *Measuring Corruption*, Ashgate, 2006.

Sequeira, Sandra, "Corruption, trade costs, and gains from tariff liberalization: Evidence from Southern Africa," *American Economic Review*, 2016, *106* (10), 3029–63.

Shleifer, Andrei and Robert W Vishny, "Corruption," *The quarterly journal of economics*, 1993, *108* (3), 599–617.

Tella, Rafael Di and Ernesto Schargrodsky, "The role of wages and auditing during a crackdown on corruption in the city of Buenos Aires," *The Journal of Law and Economics*, 2003, *46* (1), 269–292.

Treisman, Daniel, "What have we learned about the causes of corruption from ten years of cross-national empirical research?," *Annu. Rev. Polit. Sci.*, 2007, *10*, 211–244.

Truyts, Tom, "Social status in economic theory," *Journal of Economic Surveys*, 2010, *24* (1), 137–169.

Zitzewitz, Eric, "Forensic economics," *Journal of Economic Literature*, 2012, *50* (3), 731–69.

A Data sources

Our main dataset is obtained by merging 104 traffic registration databases covering 50 Russian regions and available from the Internet. Each regional data set covers cars that were registered, unregistered, involved in an accident and/or traffic violation in the region over the 1994-2011 period. Initial datasets were inconsistent with each other and used various formats and software. Once cleaned (e.g. by removing duplicate entries), reconciled and organized in a single format, the dataset includes 72,682,678 entries, each representing a car with the following attributes: license plate number (with a regional identifier), brand, and class (small, medium, large, executive, luxury, SUV, sports etc.). For some regions, dataset entries also include car registration date, when a license plate was issued. Regional coverage of the dataset varies over the observation period, and is wider than the 51 regions where the data originated. The regional coverage was expanded beyond the original 51 regions, because regional databases included cars registered in other regions, if such cars were involved in accident and/or traffic violation in the given region. In such case, we used regional identifiers, which are parts of Russian license plates, to assign a car to another region. This approach yielded data for 86 Russian regions (from the total of 89 Russian subnational units at the beginning of the observation periods). For 76 regions, our dataset has more than 1,000 observation per region. For the period after 2014, when de facto private trade by license plates was made possible, we use data from the specialized online license plate trading platform nomera.net. We also use a car trading portal car.ru as a source of car market prices.

B Identification of vanity plates (RNN numerical combinations)

Vanity plates are commonly identified by their three-digit numerical combinations (see Figure 2 in the paper) (the rest of an alphanumerical license plate, comprising the letter part and the regional identifier, is the plate's series)[15]. Instead of imputing motorists' preferences for some numerical combinations, we reveal such preferences from the available data. We do so by using seven heuristic approaches, elaborated below in this Appendix, to calculate indexes $\nu_i(p), i = 1, \ldots, 7$ of preferences demonstrated by Russian motorists for numerical combinations p from 001 through 999. These independently obtained measures are strongly correlated with each other, and we aggregate them in a single license plate attractiveness index. License plates for which such index is above a certain threshold are *RNN* vanity plates used in the calculation of C_j corruption measures. The first four measures make use of the main dataset, covering a period before 2014, when private trade in license plates was not possible and hence no conventional market data was available to assess desirability of particular license plates. Instead, we look for incidences of violations of the prescribed incremental order of license plate issuance. As argued in the paper, such violations reflect either holding of license plate for a "right" applicant, in which case the plate is issued past its prescribed place in the incremental order, or "mining" for a desired license plate among those left in the series, in which case a plate is issued ahead of its place in the incremental order. While such deviations in all likelihood reveal corruption, we use them at this stage not yet to detect and measure corruption (which is done later on in our analysis), but to gauge preferences for particular license plates shown by applicants, under the assumption that corrupt officials cater to such preferences. The following figure presents results of issuance dates screening for five randomly picked license plate series allocated in a certain Russian region (not identified to avoid de-anonymization). Horizontal axis shows the date of registration, whereas the vertical one – numerical combinations of license plates issued on a particular date. Some numerical combinations clearly and consistently break out from the otherwise monotonic increase of the numerical combinations issued one after another.

[15]Motorists could also value particular series – in their own and/or in conjunction with numerical parts, but our analysis is confined to numerical parts only.

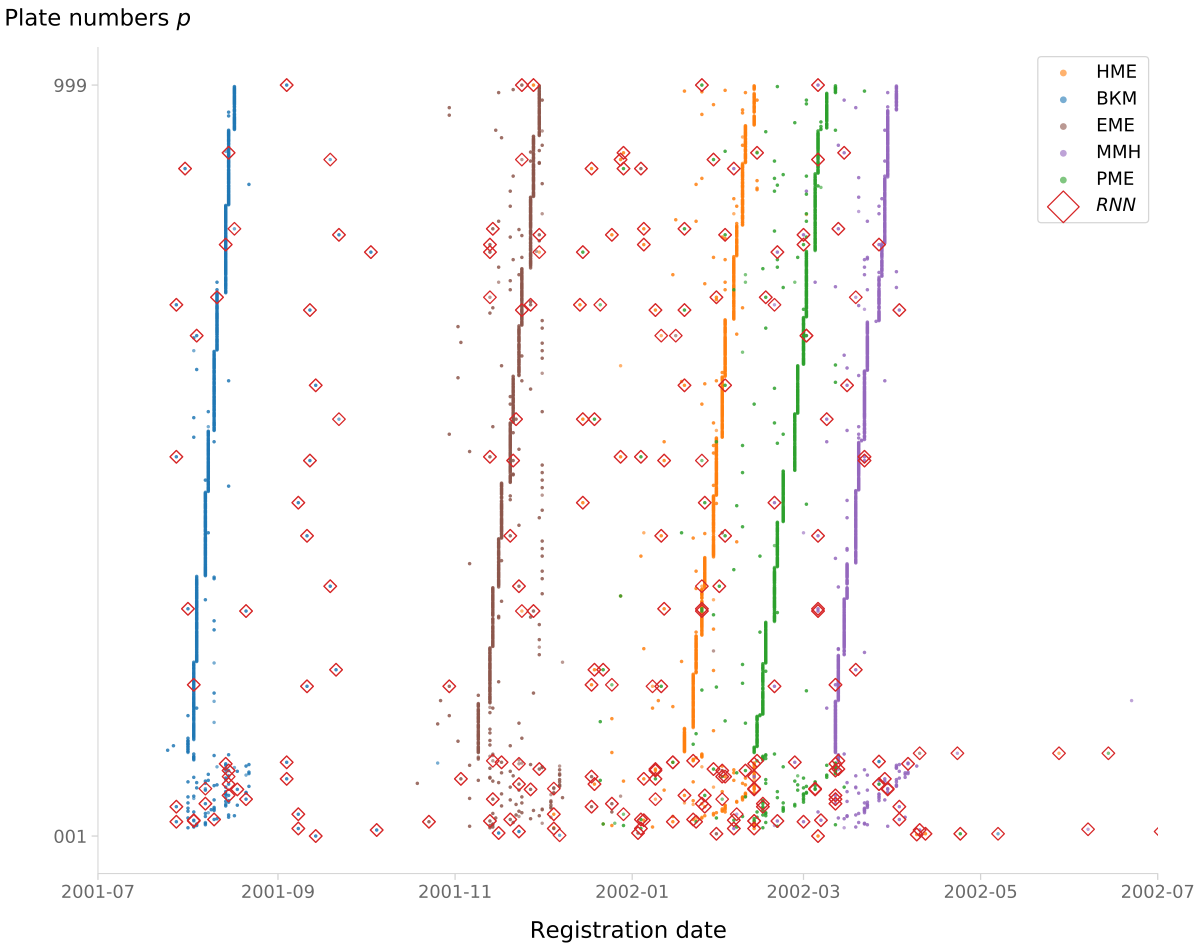

Figure B.1: First registrations in a Russian region. The horizontal axis is the date of first registration, and the vertical one – the numerical combination of a license plate. According to the law, plate numbers should increase monotonically over time, implying upwards sloping (broken) lines. Some plates deviate from these issuance lines, many of them (indicated with red diamonds) from the *RNN* group. Data source: Russian car registration databases

For a license plate (p,s), where p is the three-digit numerical part, and s – series, denote r(p,s) the registration date when the license plate was issued. Denote r(s) the average registration date for all license plates issued in the given series, and $\sigma(s)$ the standard deviation of registration dates in the given series. Since we want to capture both the holding and mining patterns, when a plate is issued either ahead of or past its turn in the incremental order, we calculate the (standardized) deviation of the issuance date from the average for a given series. Next, we take the average of such deviations across all series $s \in S$ in our dataset, for which car registration dates are available (the total number of such series is N):

$$d_p \equiv \frac{\sum_{s \in S} \frac{|r(p,s) - \overline{r}(s)|}{\sigma(s)}}{N}$$

Measure d(p) follows a V-shaped pattern, deviations from which point to preferred license plates. To identify such deviations, we deduct from d(p) the trend component $\alpha|p-500|$, where α is the slope of the V-shaped curve, assuming the same for the descending and ascending parts[16], and the absolute value of the difference

$$\nu_1(p) \equiv d(p) - \alpha|p - 500|$$

is our first measure of the numerical combination's attractiveness. This measure alone immediately captures some NN combinations, which clearly break out from the issuance trend.

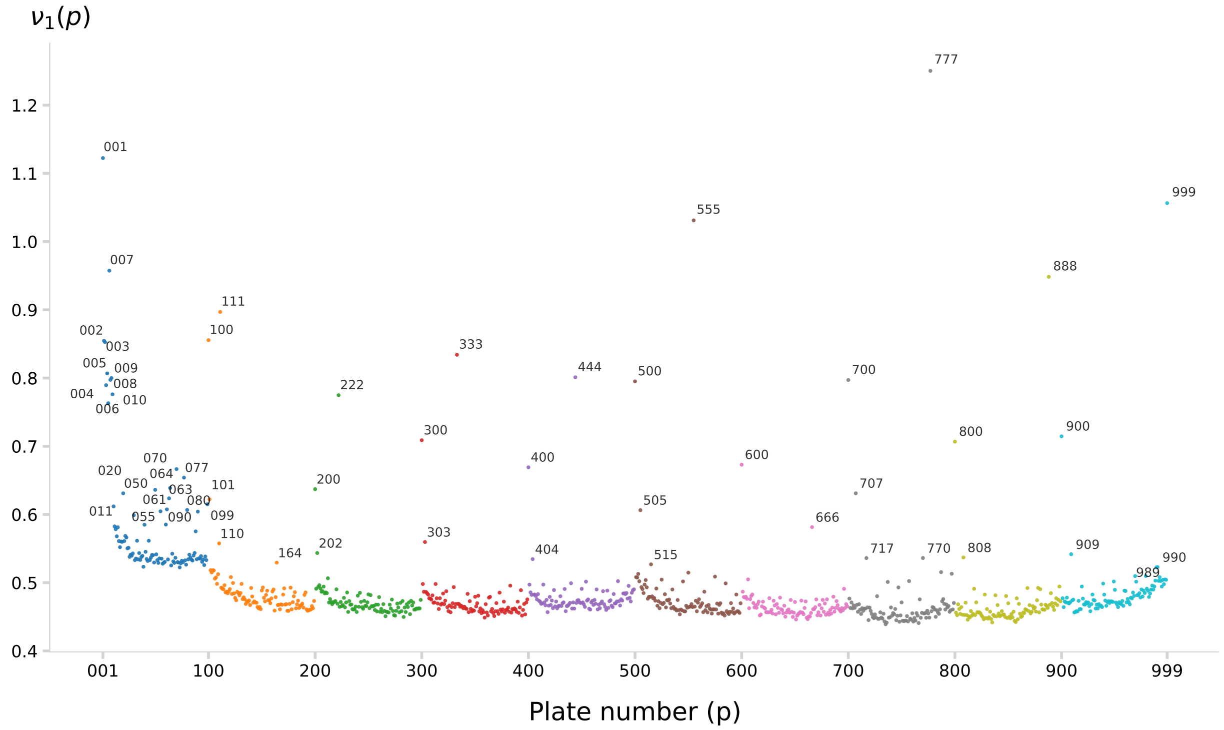

Figure B.2: Measure $\nu_1(p)$ chart. Plate numbers that on average deviate more from their prescribed issuance date are those that are more frequently involved in plate redistribution via corruption. The plate numbers with the top $\nu_1(p)$ values are most popular in corruption transactions and hence are plausible vanity combinations, revealed by anomalies in administrative data. Such numbers include triples, whole hundreds and numbers with some repetition or symmetry. (Different colors are used for consecutive hundred portions of numerical combinations).

The first measure $\nu_1(p)$ is calculated by initial averaging across series, and obtaining a baseline pattern for such averages; deviations from such baseline reveal RNN combinations. In the second measure $\nu_2(p)$, the order of averaging and obtaining baselines is

[16]The slope is calculated as follows: $\alpha = [\frac{d(001)+d(999)}{2} - d(500)]/499$

reversed. Namely, for each series $s \in S$ we estimate an issuance line relating a combination p to the time r(p,s) when this combination was issued. Once such estimation is obtained, the actual issuance time r(p,s) is compared to its fitted value, predicted by the issuance line for the series, and the distances between the two are averaged across series. More specifically, we derive an ad-hoc robust linear fit to each issuance line of individual series $s \in S$. First, all numerical combinations are divided in two equal halves. Next, reference points are determined for each half; a straight line connecting these two points provides fitted values $\tilde{r}(p, s)$ of issuance time for the series.[17] The second measure obtains as

$$\nu_2(p) \equiv \frac{\sum_{s \in S} |r(p, s) - \tilde{r}(p, s)|}{N}$$

Calculation of the third measure reflects concentration of $\nu_1(p)$ values within consecutive hundreds of numerical combinations, as shown e.g. by Figure B.2. To account for this, we modify the original measure as follows:

$$\nu_3(p) \equiv \nu_1(p) - \nu_1^k(p)$$

where ν_1^k is the average value of $\nu_1(p')$ for the numerical combinations p' from $k00$ to $k99, k = 0, 1, \ldots, 9$, and $k(p)$ is the highest integer multiple of 100 preceding p. To derive yet another measure, we combine elements of the above approaches, i.e. estimate slopes of issuance lines for particular series $s \in S$, but do so for hundred-long segments of numerical combinations $p' \in \{k00, \ldots, k99\}, k = 0, 1, \ldots, 9$, as in $\nu_3(p)$. For every such segment of a given series, we use the Ordinary Least Squares (OLS) method to fit the slope of the issuance line $r(p', s)$. For a given numerical combination p, belonging to the segment from $k(p)00$ through $k(p)99$, we calculate the fitted value $\tilde{r}(p, s)$, obtained from the OLS estimation for this segment in series s. The fourth measure now obtains, similarly to the second one, as

$$\nu_4(p) \equiv \frac{\sum_{s \in S} |r(p, s) - \tilde{r}(p, s)|}{N},$$

except that presently fitted values are OLS-estimated for hundred-long segments.

The fifth measure is based on actual market data available for later periods, when private trade in license plates opened up after the 2014 modification of the car registration

[17] Reference points are medians for the 0.30 quantile and 0.40 quantile of the first and the second halves (combinations from 001 through 500 and 501 through 999, resp.). These selections were made by trial and error, using visual fit and standard errors of the fitted lines. Shift of the reference points to the combinations which are lower in the pecking order reflect the prevalence of the holding pattern over the mining one (see p.9 of the paper).

system. As explained in the paper, this modification in all likelihood did not fully eliminate corruption, although irregularities in license plates allocation could not any longer be ascribed to corruption alone. However, such data could still be used to observe preferences for particular numerical combinations, this time revealed in the conventional market transaction - based manner. One can expect stability of preferences for *NN* license plates, in which case market-based preference measures could be added to those obtained from earlier observed administrative transactions. Our source of market data is a license plate online trading platform nomera.net (data collected on 10/24/2017). Using this data, we obtain the index $\nu_5(p)$ as the quantity of license plates with the numerical combination p offered for sale on the above platform.

The two remaining approaches make use of the redistribution of *NN* plates towards more expensive cars. We calculate $\nu_6(p)$ as the average market value of a car with a license plate including the numerical combination p, expecting that for *NN* plates such value would be higher. Car value data are imputed in the administrative data using listed prices on the online trading platform car.ru.[18] Finally, measure $\nu_7(p)$ reflects redistribution of license plates with the numerical combination p between car categories. This measure is calculated similarly to the regional corruption indexes C_j described in the main part of the paper, with two modifications – first, the calculation is performed for a particular numerical combination p, rather than the set of *RNN* combinations, and second, it covers all Russian regions represented in the main dataset. All seven measures are strongly correlated with each other, both in magnitude and significance (Table B.1). Therefore, all these measures, using different methodologies, types and sources of data (administrative and market), and reflecting different observation periods, consistently point out to the same stable patterns. To reduce noise that could be present in particular measures, we standardize them by deducting averages and dividing by standard deviations, thus obtaining $\nu_i^s(p), i = 1, \ldots, 7$, and take the average vanity index

$$\nu(p) \equiv \frac{\sum_{i=1}^{7} \nu_i^s(p)}{7}$$

Inspection of the distribution of the vanity index identifies a cutoff level $\underline{\nu}$[19], and treat com-

[18] Data collected as of 7/15/2017, which is long after the end of the observation period for the main dataset. This time discrepancy is a source of noise in such index. However, relative car values did not exhibit drastic changes, and hence the noise should not be significant. Indeed, as shown below, $\nu_6(p)$ is strongly correlated with other indexes.

[19] $\underline{\nu} := |min(\nu(p))|$

binations p such that $\nu(p) \geq \underline{\nu}$ as "revealed nice numbers" (RNN). Figure B.3 illustrates this procedure, shows RNN numbers and provides additional evidence that individual measures used in the aggregation are in broad agreement with each other.

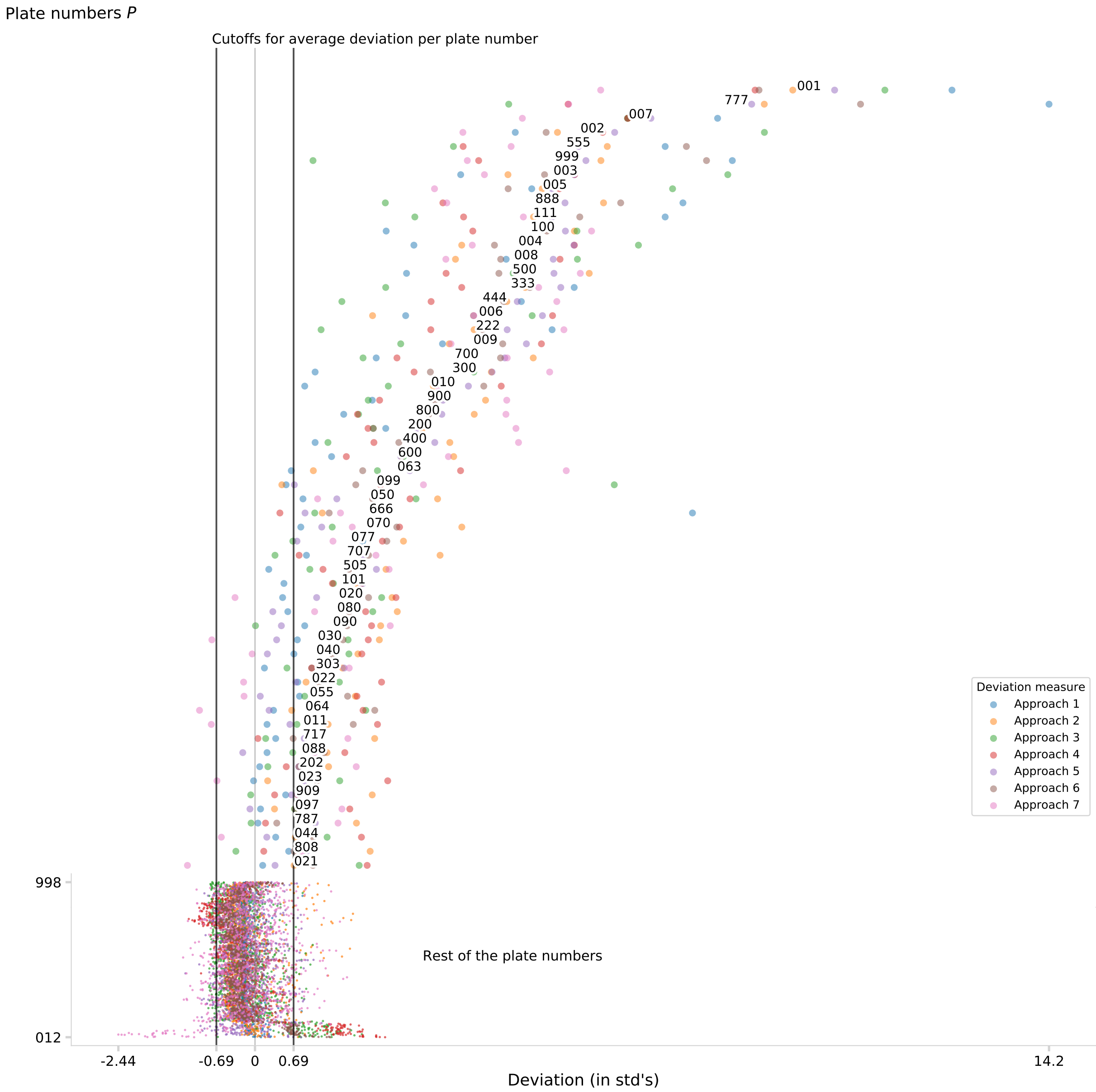

Figure B.3: Identification of *RNN* combinations. Numerical combinations on the graph show the value of the aggregate index $\nu(p)$ for the corresponding combination; colored dot show the values of partial standardized indexes for the same combination.

	Approach 1	Approach 2	Approach 3	Approach 4	Approach 5	Approach 6	Approach 7
Approach 1	1.00	0.89	0.89	0.69	0.88	0.79	0.92
Approach 2	0.89	1.00	0.80	0.58	0.69	0.83	0.75
Approach 3	0.89	0.80	1.00	0.83	0.84	0.72	0.88
Approach 4	0.69	0.58	0.83	1.00	0.65	0.53	0.77
Approach 5	0.88	0.69	0.84	0.65	1.00	0.66	0.84
Approach 6	0.79	0.83	0.72	0.53	0.66	1.00	0.73
Approach 7	0.92	0.75	0.88	0.77	0.84	0.73	1.00

Table B.1: Correlations of Approach 1 - 7. All 7 approaches are highly correlated with each other, even though they use very different data sources.

C Redistribution of *RNN* license plates across automobile classes

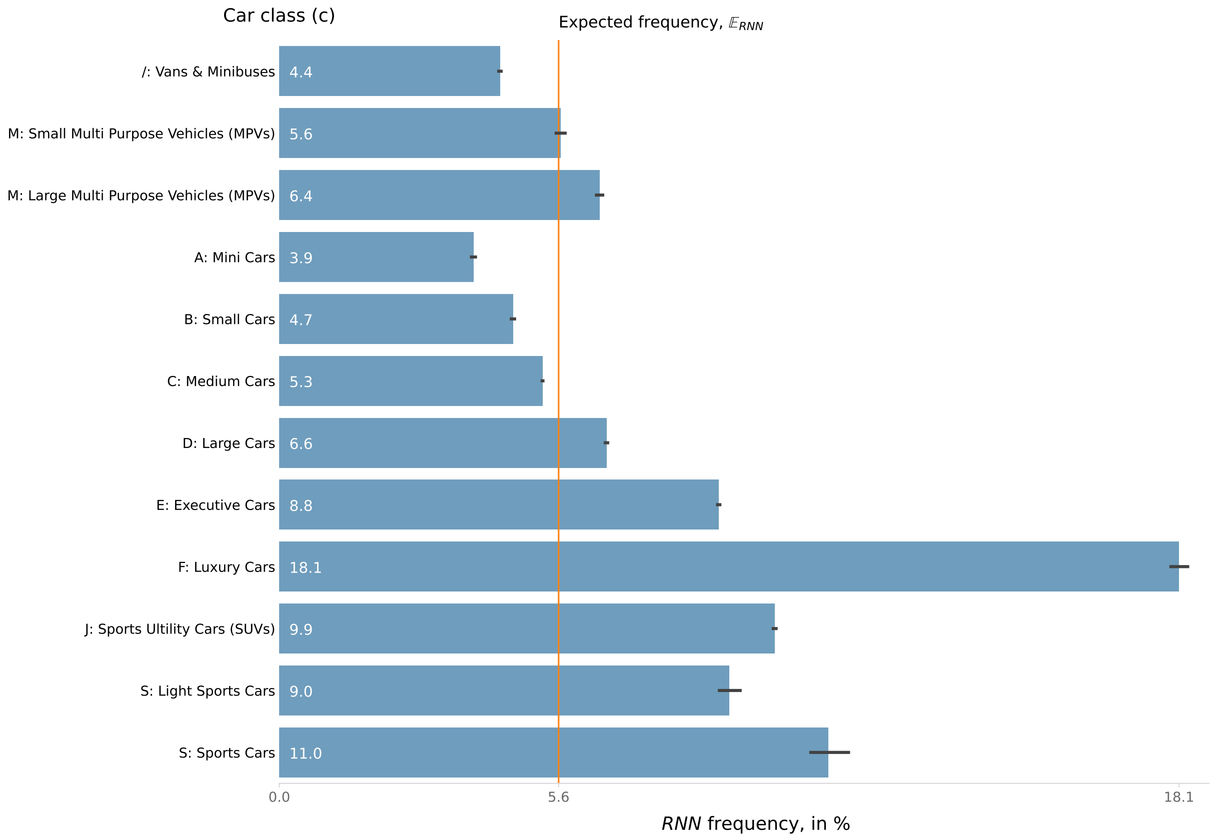

Figure C.1: Frequency of RNN license plates across car classes. Expected frequency without corruption is 56 in 999 or 5.6%. Data source: Russian car registration databases

D C_j by region

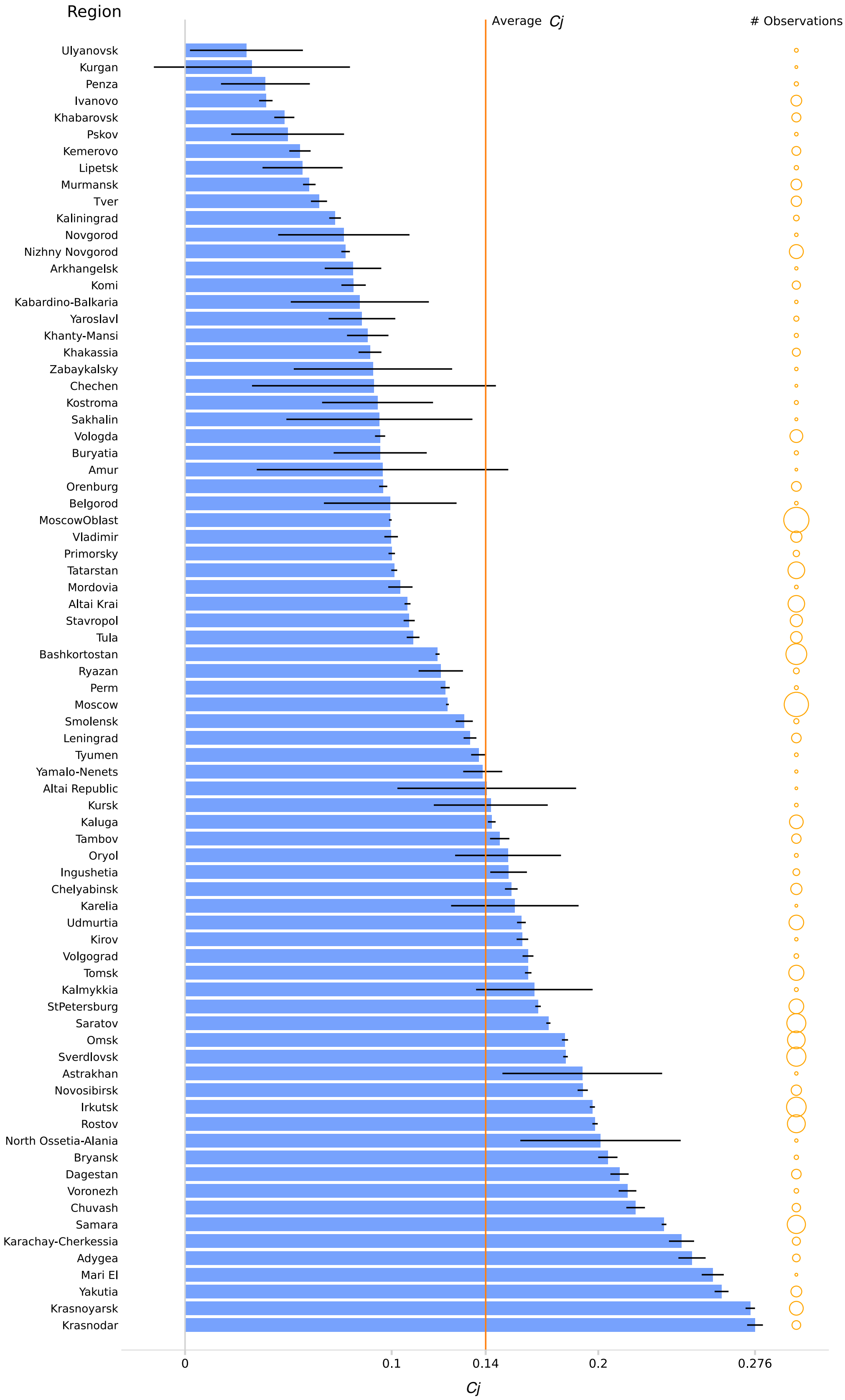

Figure D.1: C_j by Russian region. Average C_j across all regions is indicated by the vertical orange line. Black horizontal bars indicate %90 confidence intervals