Direct Estimation of Hidden Earnings: Evidence From Administrative Data

Serguey Braguinsky, Sergey Mityakov, and Andrey Liscovich[•]

Abstract

We estimate hidden earnings by matching car registries to employers' records of paid earnings for a panel of individuals and households in Moscow. The identification strategy is based on the idea that reported earnings may be falsified, but car registries are accurate. Hidden earnings comprise over 75 percent of actual earnings of the large majority of car owners, at least twice as high as estimated in previous studies using less direct methods. There is also a lot of heterogeneity across employers. Foreign-owned firms, large firms and state-owned firms in capital-intensive industries report earnings more transparently than do small firms and firms in labor-intensive industries, where actual earnings may be more than five times higher than reported earnings. Differentials of similar magnitude are found in public services, especially among educators. Our findings shed new light on the perceived links between firm ownership, size and productivity in countries with large hidden economies.

Keywords: hidden economy, foreign-owned firms, size-wage effect, corruption
JEL classification codes: K42, P37.

This draft: December 3, 2010

[•] Braguinsky, Department of Social and Decision Sciences, Carnegie Mellon University, e-mail: sbrag@andrew.cmu.edu; Mityakov, John E. Walker Department of Economics, Clemson University, e-mail: smityak@clemson.edu; Liscovich, Kennedy School of Government, Harvard University, e-mail: liscovic@fas.harvard.edu. We thank William Dougan, Steven Klepper, Tom Mroz, Peter Thompson and seminar participants at Carnegie Mellon University, Clemson University and Florida International University for helpful comments. The usual disclaimer applies.

1. Introduction

Measuring the hidden economy is notoriously difficult. Yet it is essential to be able to assess not only the magnitude but also the distribution of hidden activities across sectors of the economy. Knowing the magnitude allows us to evaluate to what extent recorded differences and growth in GDP per capita reflect real differences and improvement in the standards of living. Knowing the distortions stemming from the heterogeneity of hidden activity across industries and public services is a necessary condition for devising meaningful economic policies aimed at improving efficiency and fighting corruption.

In this paper we employ unique administrative data from Moscow, Russia to obtain a direct estimate of the hidden economy at a disaggregated level. Our approach starts from the observation that it is relatively easy to misreport earnings, but it is very costly to drive an unregistered vehicle.[1] This difference is the key to our identification strategy, as we match administrative data on wages and salaries to car ownership data, to measure hidden earnings. We also use this estimation procedure to gauge unrecorded incomes in government employment and public services.

The magnitude of hidden earnings we find is stunning: for the vast majority of Moscow car owners, estimated true earnings exceed reported earnings by 200 percent or more. We also find a lot of variance systematically related to employer categories. Our methodology detects almost no hidden earnings in foreign-owned firms and large state-owned companies in capital-intensive industries, such as utilities. The same methodology results, however, in an estimate of 85-90 percent of earnings hidden in smaller private companies, especially in labor-intensive industries, such as trade and services, private banking and finance, as well as in public education and health care.

Of course, we are not the first to attempt to measure the hidden economy. Previous empirical approaches have exploited macroeconomic relationships, such as the share of cash in the money supply (Cagan [1958], Tanzi [1983]), electricity consumption (Kaufmann and Kaliberda [1996], Schneider and Enste [2000], Alexeev and Pyle [2003]), or multiple indicators aimed at econometrically estimating the hidden activity as an 'unobserved variable' (Frey and

[1] Moscow police routinely conduct traffic stops to check the paperwork. Unregistered vehicles may be impounded and can be recovered only after paying a fine and producing the registration document.

Weck-Hanneman [1984], Giles [1999]). Other studies used survey data on household incomes and consumption expenditures (Pissarides and Weber [1989], Lyssiotou et al. [2004], Ivanova et al. [2005]], Gorodnichenko and Sabirianova Peter [2007], Gorodnichenko et al. [2009]). Yet other studies employed direct evidence from tax audits (Klepper and Nagin [1989], Feinstein [1999]).

The approach we take in this paper is different in a few key respects. First, we use recorded administrative data. This gives us an exact amount of earnings reported at their source and also a standardized and objective way to estimate with reasonable degree of precision the values of cars owned by individuals and households. We are thus able to avoid the potential problem that survey respondents with large underreported incomes might have incentives not to report their expenditures accurately.

Second, the earnings data come from employers. This allows us to obtain direct evidence on hidden earnings disaggregated by employer type. To the best of our knowledge, our approach produces the first estimates in the literature that measure heterogeneity of hidden earnings separately by ownership, industry, firm size and type of government employment in a large and representative dataset.

Our empirical strategy consists of three steps. First, we select from the universe of all data a representative sample of car owners because we need information about car ownership to identify true (as opposed to generally falsely reported) earnings. Second, we use the data on earnings reported by businesses we believe to be accurate, and we exploit the variation in incomes and car values within this subsample to estimate the income elasticity of demand for the stock of cars. Finally, we use this estimate, coupled with the information about car values in the rest of our sample, to measure the hidden earnings among other car owners.

Comparison of the results from our approach with those obtained in past studies significantly broadens the contentious range of plausible estimates of the hidden economies in developing and post-communist countries (e.g., Fiege and Urban [2008]). In the case of Russia, the most recent estimates in Gorodnichenko et al. [2009] using the data from the Russian Longitudinal Monitoring Survey (RLMS) imply that hidden earnings comprise 33-36 percent of actual earnings[2] versus our findings of 75 percent or more. We conduct tests aimed at

[2] The consumption expenditure and earnings data in RLMS imply a saving rate of minus 30 percent, while the actual saving rate was plus 3-6 percent. The authors carefully note that their estimates should perhaps be interpreted as a

elucidating how much our findings may apply to non-car owners later in the paper and do not find indications of a significant upward bias.

We also show that car owners are relatively more productive workers in the private sector and higher-ranked government employees than non-car owners. Thus, at the very least, our study demonstrates that hiding among the business and government elite has been happening on a much larger scale than previously envisioned. Since successful development and market reform policies are most naturally associated with incentivizing workers and government employees with high-level human capital, the policy implications are immense.

Past studies looking into the balance of benefits and costs of hiding have anticipated the direction but, arguably, not the magnitudes of some of our findings.[3] Particularly striking, in this regard, is the reversal of the conventional size-wage effect that we find after taking into account much larger hiding by smaller firms. Using the data on developed countries, the literature has consistently found that larger firms pay higher wages to similar workers than do smaller firms (Brown and Medoff [1989], Abowd et al. [1999]). More recently, this size-wage effect was found to be even larger in developing countries (e.g., Ströbl and Thornton [2002]). While there are reasons higher efficiency may be correlated with larger firm size in advanced market economies, the opposite might be true in developing and post-communist countries, as smaller firms tend to operate in less rigidly regulated sectors, while larger firms predominantly populate state-controlled or crony-type parts of the economy. In line with this, we find that car values monotonically and strongly *decrease* across the firm-size distribution, even though officially reported earnings strongly *increase* with firm size.

Another important insight from our examination of the heterogeneity of hiding comes from comparing foreign-owned and private Russian firms. Foreign-owned firms in our data pay on average three times higher official wages than do domestic firms. This evidence is in line with the findings in the literature (Bloom et al. [2010], Görg et al. [2007], Sabirianova Peter et al.

---

lower bound (p. 513).

[3] The benefits of hiding have been associated with avoiding bureaucratic corruption (Johnson et al. [2000]), evading taxes (Schneider and Enste [2000]), and retaining financial resources to pay to rackets (Frye and Zhuravskaya [2000]). The costs consist of possible punitive sanctions and an inability to use some government services (Djankov et al. [2003]). Small labor-intensive firms have a balance of benefits and costs favoring more hiding than large capital-intensive factories. See also Kaufmann and Kaliberda [1996].

[2005]). But looking at the market values of cars, we see only marginal evidence of higher actual earnings in foreign-owned firms. Thus our study suggests that differences in transparency need to be taken into consideration when assessing the contribution of foreign-owned firms to productivity.

Our findings also carry important macroeconomic and structural implications. The economic decline after the collapse of communism is often linked to markets not performing their role of reallocating labor from less productive to more productive lines of activity. Our analyses indicate, however, that a lot of the statistical evidence on both the GDP decline itself and the misallocation of labor might be an artifact of underreporting. Thus, Russia may indeed be much closer to a "normal country" (Shleifer and Treisman [2005]) than official statistics would suggest.

In the next section we describe the data and sample construction. Sections 3-4 present the evidence, including a simple empirical model to capture hidden earnings at a disaggregated level. In Section 5 we entertain possible alternative explanations for our findings and conduct robustness checks. Section 6 concludes. More details of the sample construction and estimation procedures are presented in a separate on-line Appendix.

2. Data and sample

In the early-mid 1990s, amid official statistics showing sharply declining GDP and plunging real incomes of the vast majority of the population, one could not help but notice the ever-increasing number of expensive foreign cars in the streets of Moscow, St. Petersburg and other Russian cities. For Russia as a whole, the car ownership rate more than doubled from 58.6 vehicles per 1,000 people in 1990 to 122 vehicles per 1,000 people in 1998 (the last year of the official GDP decline).[4] For Moscow, the corresponding number almost tripled, from 70.6 to 200.4. Part of this explosion in car ownership, especially in early years, followed from the scarcity of cars in the Soviet Union. Still, new car purchases could have hardly happened on such a scale if real earnings were indeed declining sharply. This obvious discrepancy between the trends in officially measured earnings and in car ownership suggested that we might learn more about hidden earnings by matching car ownership to individual earnings than could be

[4] Unless explicitly stated otherwise, all the macroeconomic data come from the official Federal State Statistics Service (http://www.gks.ru).

done using less direct methods.

2.1 Sample construction

Our data come from three sources. One is the residency registry for the year 2002. It contains the names and addresses of all registered residents of Moscow, which can be used to identify members of the same household.

The second source is the 2005 auto registration database (containing retrospective data) that can be matched to the residency database. We use the vehicle's identification number (VIN) to trace its history of owners. We then use the information about the make, model and year to impute the market value of the car in a given year according to a standardized procedure. Appendix 1 describes this procedure in detail.

The third source comprises administrative databases of incomes between 1999 and 2003.[5] These data are official records of all payments and withheld taxes made by all employers (payment sources) registered in Moscow that entered the public domain (see Guriev and Rachinsky [2006] for the discussion of these data). The data contain unique identification numbers, addresses and phone numbers for each employer and employee, which we use to retrieve information from open sources about employer's sector of economic activity and the type of ownership.

Income databases can be matched across years as well as to the residency and vehicle registration databases. As is usual with administrative data, there is a problem of missing information. Information is also recorded in different formats across the different databases so the matching procedure had to be done almost entirely by hand. This made it too costly to work with the universe of all the data and forced us to choose a smaller-sized sample.

We started with 30,000 households (97,141 individuals) randomly selected from the residency database and matched individuals in this sample to the incomes and auto databases using full names, dates of birth and addresses.[6] We then eliminated duplicate income and auto records, commercial vehicles (such as trucks and buses) and motorbikes, and dropped

[5] The 2004 database is also available but we did not use it because the spread of consumer credit that started in Moscow in that year might compromise our identification strategy.

[6] To ensure privacy, all individual-indentifying information used in the process of constructing the sample was subsequently purged from the data.

households consisting solely of retirees. This left 6,101 households (21,617 individuals), for which we could match at least one member to the auto registration database and to officially reported earnings in each of the five income databases. The latter condition ensures that we have five consecutive years of officially reported earnings, thus excluding temporary residents and those whose only source of income was the informal sector in any year between 1999-2003. Incomes from multiple sources are added together, with the sector of employment assigned based on the source of the highest income. Appendix 2 describes the procedure of constructing the sample in detail.

The total number of person-year observations with non-zero reported earnings is 56,893, of which 36,403 (64.0 percent) originate in the Russian private sector, 1,046 (1.8 percent) originated in the foreign-owned sector, and 19,444 (34.2 percent) originated in the government (federal and city/local) and state-owned sector. The total number of person-year observations with non-zero car values is 40,281, but about 30 percent of those have missing VINs. We elected to eliminate these autos from the analysis below because we could not trace car histories and could not determine the period during which an individual owned a given car.[7] We also eliminated a very small number of individuals who owned more than five cars with valid VINs in any given year out of concerns about the quality of the data in those cases.

Since one of the main contributions of our paper is to obtain estimates of the variation in the extent of income underreporting by type of formal employment, in the statistical analysis below we take some extra steps to eliminate potential problems from possible income sources unrelated to official employment. First, in addition to requiring car owners in our sample to have had official employment in all five years 1999-2003, we exclude from all years those who earned less than the official minimum wage in any given year (5-7 percent of observations, depending on the year). We also exclude car owners whose reported earnings exceeded the equivalent of $100,000 in any given year (less than 0.3 percent of observations). Second, we looked at all sources of income of car owners in our sample and identified windfall earnings, such as lottery prizes and income from assets. We excluded from our estimates about 120 individuals (and their households) whose main source of income came from such sources. Our sense is that in the majority of the remaining cases extra informal income sources (if any) tend to be closely related

[7] We also conducted all the estimations presented below assuming that cars with missing VINs were owned by car owners in our sample throughout 1999-2003. The basic results were not affected.

to formal employment (such as a teacher engaged in private tutoring, a construction worker moonlighting in a side construction project, or a traffic policeman taking bribes from motorists).

2.2 Summary statistics and sample characteristics

The average age of individuals in the sample in 2003 was 43.6 years, 51.5 percent are male. For each year, we calculated the percentile of an individual in our sample in the overall earnings distribution of his or her employer and the size of the employer by counting the total number of entries pertaining to its identification number. All ruble values were converted to US dollars using average market exchange rates for each year.

Table 1 presents the basic summary statistics on reported earnings and car values in our sample. Average car values exceed average annual earnings for all years pooled together and in every single year, although there is a pronounced downward time trend. In Appendix 4 we compare these data with the U.S. data from the National Longitudinal Study of the Youth (NLSY) and show that while differentials in car values are roughly in line with relative GDP per capita in the two countries, the differential in earnings is three times larger.

Even though earnings of car owners in our sample are low compared to their car values, they are on average 2.2 times higher than for the whole Moscow workforce and 3.3 times higher than mean earnings estimated from the RLMS data for the overlapping years of 2000-2002 (see Appendix 3). But the growth rates in earnings for the overlapping years are almost exactly the same between our sample and RLMS, and earnings at the 40th percentile of our sample are almost identical to mean earnings estimated from RLMS. Hence, a higher fraction of hidden earnings estimated using our approach as compared to using the RLMS data does not appear to be driven by particularly low reported incomes of car owners in our sample.

The fact that car owners in our sample have significantly higher reported earnings than the whole workforce does carry one significant implication, however. Even though employers may show only a fraction of true earnings they pay to employees in official employment contracts, relative reported earnings closely reflect relative true earnings, so that a worker at a higher percentile in the earnings distribution is also more productive and/or more important in the employer's hierarchy than a worker at a lower percentile.

Table 2 shows that the average percentile of car owners in overall earnings distribution in the whole Moscow earnings database is 0.74, with the corresponding numbers for private and government employment being 0.75 and 0.70, respectively. Thus, an average car owner employed in the private sector in our sample is 50 percent more productive than an average private-sector worker. Similarly, an average government employee in our sample is positioned 40 percent higher in the hierarchy than an average government employee. The median earner in our sample of car owners is located at the 80th percentile of the earnings distribution in the whole database, while those above the 90th percentile in the earnings distribution of car owners belong to the top 1 percent of all income earners in Moscow.

3. Empirical strategy

3.1 A simple estimation model

Our basic assumption is that employers conceal a fraction of true earnings of their employees as part of the employment contract. Such a contract may be explicit (as in cases where hidden earnings are actually paid out in unregistered cash), or implicit (as in cases where incomes from kickbacks or bribes are factored into officially low salaries). Specifically, let employee *i*'s earnings at time *t* be reported in the amount $E_{it}^R = \Gamma_{it} E^*_{it}$, where E^*_{it} are true earnings and Γ_{it} is the fraction reported. This fraction depends on the costs and benefits of hiding that vary with employer characteristics, $\mathbf{S_{it}}$. It may also depend on a range of individual- (or household-) specific characteristics $\mathbf{X}_{\mathbf{i}}^{(1)}$ because different employees may have different employment contracts even with the same employer. Finally, there might be time effects in reported earnings caused, for example, by institutional changes. Hence, we assume that

$$\ln E_{it}^R = \ln E^*_{it} + \mathbf{b}'\mathbf{S}_{it} + \mathbf{g}_1'\mathbf{X}_{it}^{(1)} + \varphi_1(t) + \eta_{it}, \qquad (1)$$

where $\mathbf{X}_{it}^{(1)}$ is the vector of individual characteristics that affect the employment contract and $\mathbf{S_{it}}$ is the vector of employers' characteristics (such as ownership type, size and industry). The primary focus of our empirical analysis is the vector of coefficients $\mathbf{b}'$, which measures average income hiding associated with different characteristics of employers $\mathbf{S}$. The more negative the coefficient β_k, the larger is the fraction of hidden earnings in total earnings among individuals employed in the category of employers possessing characteristic *k* and vice versa.

As noted above, employer-level heterogeneity in hiding comes from the differential costs and benefits of hiding. On the one hand, it could be argued that ability to misreport incomes may be linked to ownership of the establishment. Foreign companies from Western countries could be subject to legal actions in their home countries for breaking the laws in other countries or/and lack connections to escape scrutiny of Russian tax authorities. Thus, other things equal, we would expect coefficient β for domestic employers to be more negative than for foreign-owned establishments.

On the other hand, the costs and benefits of underreporting could also differ across sectors of economic activity. One can argue that smaller and more labor-intensive companies would find it easier to escape the oversight of tax authorities than would larger firms and firms located in more capital-intensive sectors (Kaufmann and Kaliberda [1996], Djankov et al. [2003]). We divide all employers in our sample into 19 categories (see Appendix 6 for the full list). We expect to find more negative coefficients on dummies associated with labor-intensive occupations such as retail-wholesale trade, services than for capital-intensive occupations such as utilities and manufacturing.

We also look for income underreporting in the state-owned establishments. Since employer is unlikely to falsify reported earnings in those cases, one might conjecture that these hidden incomes arise from a number of unofficial sources (including moonlighting, unreported tutoring, and bribes).[8] The Global Corruption Barometer identifies law enforcement, education and health care as the most corrupt branches of the state sector in Russia.[9] Thus, we would expect more negative estimates of β_k for those sectors as well.

The identifying assumption used in assessing the degree of underreporting in different sectors is that the fraction of reported income depends on sector of employment, but demand for the stock of cars has the same (controlling for individual specific characteristics) functional form in all sectors. In particular, we consider the following car stock demand equation:

$$\ln C_{it} = \lambda \ln E^*_{it} + \mathbf{g}_2' \mathbf{X}_{it}^{(2)} + \varphi_2(t) + u_{it} \qquad (2)$$

That is, the demand for the stock of cars depends on actual earnings E^*_{it}, individual characteristics $\mathbf{X}_{it}^{(2)}$, time effects $\varphi_2(t)$, and an individual and time specific disturbance term u_{it}, $E\left[u_{it}\right] = 0$.

Substituting for $\ln E^*$ from equation (1) into (2):

---

[8] Note that if a person officially works somewhere else (even part-time), then his/her earnings will be part of the records in our database. Thus, extra incomes can come only from activities not reported officially.

[9] See http://www.transparency.org/policy_research/surveys_indices/gcb/2005)

$$\ln C_{it} = \lambda \ln E_{it}^{R} - \lambda \mathbf{b}'\mathbf{S_{it}} + \mathbf{g}'\mathbf{X_{it}} + \varphi(t) + \varepsilon_{it} \tag{3}$$

Here $\mathbf{g} = \mathbf{g_2} - \lambda \mathbf{g_1}$, and $\varphi(t) = \varphi_2(t) - \lambda\varphi_1(t)$ measure the combined effect of observables and time effects,[10] $\varepsilon_{it} = u_{it} - \lambda\eta_{it}$ is the combined disturbance term, and $\mathbf{X}_{it}$ combines individual characteristics from $\mathbf{X}_{it}^{(1)}$ and $\mathbf{X}_{it}^{(2)}$.

Unfortunately, we cannot estimate equation (3) consistently, because reported income E_{it}^{R} is correlated with the error term ε (which contains disturbance η from income underreporting equation (1)). However, if we were able to consistently estimate the income elasticity of demand for the stock of cars λ, then it would be possible to estimate sector-specific income hiding from the following regression:

$$\ln E_{it}^{R} - \frac{1}{\lambda}\ln C_{it} = \mathbf{b}'\mathbf{S_{it}} + \mathbf{g}'\mathbf{X_{it}} + \varphi(t) + \varepsilon_{it}. \tag{4}$$

In principle, one could estimate λ from other datasets where falsifying reported earnings is not that much of a problem, such as the NLSY data. In Appendix 4 we performed such estimation and obtained $\lambda = 0.34$. However, this approach could raise questions of applicability of U.S-based results to the Russian case.

We thus look for parts of Moscow data that allow us to perform such estimation directly. In particular, we wanted a subsample where earnings are reported truthfully ($E^{R} = E^{*}$). We could then estimate λ by substituting reported incomes in place of actual in regression (2).

Evidence from the raw data presented below and comparisons with the NLSY data in Appendix 4 suggest that employees of foreign-owned firms might be an appropriate subset. Because of a relatively small number of observations on such employees in our main sample, we went back to the original databases and oversampled car owners from all firms already identified in the main sample as being foreign-owned.

Using the benchmark subsample of foreign-owned firms, we estimate the value of the parameter λ from 0.3-0.4 in pooled data and year-by-year, with the exception of 1999, where it is substantially lower.[11] The estimates are not affected by the inclusion of firm size and industry controls, consistent with our main identification assumption (see Appendix 5 for details).

---

[10] This means that we cannot identify separately the effects of individual characteristics and time effects on income hiding, unless we impose some exclusion restrictions. We are identifying sectoral income hiding by assuming that sectoral dummies are not part of the car stock demand equation.

[11] We believe that 1999 may not be a representative year because of the financial meltdown in the autumn of 1998.

Moreover, the value of $\lambda = 0.34$ estimated from the NLSY data (Appendix 4) is also in the same range. We thus chose the value of $\lambda = 0.35$ as our baseline value in regression (4) but we also conducted sensitivity analysis with respect to this value (see below) finding similar results.

3.2 Evidence from raw data

Before we proceed to the regression estimates it is instructive to try to identify the extent of hidden earnings in raw data. Table 3 and Figure 1 present annual average earnings and car values for 2003 for full-time wage earners (defined as receiving positive labor income in at least nine months of the year), averaged by income deciles. The picture is quite telling. Car values are flat or even declining except in the top two deciles of the earnings distribution. Car owners in the first decile of the wage earnings distribution reportedly earned on average just \$529 in 2003, 40 percent less than the official poverty line set by the Russian government at \$896 for the same year. Yet, the average market value of first decile earners' cars exceeds \$7,500, higher than those of wage earners in the fourth to eighth reported earnings deciles. Only wage earners in the tenth decile have significantly higher mean car values than wage earners in the first three.

To understand how much the results might be affected by low-income people owning cars purchased for them by wealthier people (such as young adults driving cars purchased by parents), the last three columns in Table 3 present the data for the same sample but including also earnings and car values of other members of their households. The qualitative picture, as can be seen, remains the same: reported earnings of the vast majority of car-owning households are very likely to substantially understate their true earnings.

The aggregate data hide large differences not just across the earnings distribution but also across different types of employers. Table 4 presents the 2003 data on earnings and market values of cars among full-time wage earners as in the previous table, along with the corresponding ratios and coefficients of variation disaggregated by employer ownership type (private Russian firms, government/state sector and foreign-owned firms), by two representative industries (wholesale and retail trade versus the mass media, communications/software and utilities), and by firm size below and above the median in the private sector. It also presents the data separately for government employees in law enforcement, secondary education and health care and all other government employees.

Recall from our earlier discussion the notable differences in reported income between workers in domestic and foreign-owned enterprises. It is telling that the ratio of car values to reported income for employees of foreign-owned entities is nearly the same as in the U.S. data (see Appendix 4).

The data show that officially reported earnings were 3.6 percent higher in government/state sector than in the private sector (the difference is not statistically significant at conventional levels), but average car values in the state-owned sector were 33 percent below those in the private sector (statistically significant at 1 percent level). There are also considerable differences within government employment itself.

The last two rows in Table 4 show that car owners employed in law enforcement, health care and secondary education have on average more expensive cars than other government/state sector employees (although the difference is not statistically significant), but their officially recorded earnings are more than 46 percent lower (statistically significant at 1 percent level). Note also that the coefficient of variation of earnings in law enforcement, health care and education is the lowest among all other categories in Table 4 but the coefficient of variation of market values of cars is among the highest. The larger coefficient of variation in these occupations suggests sizeable idiosyncratic unreported incomes in those sectors.

There are also some noteworthy differences across industries. The ratio of car values to earnings in labor-intensive industries such as trade is twice as high as in relatively knowledge- or capital-intensive industries such as media, communications/IT and utilities. These patterns are studied more systematically in the following sections.

4. Estimation results

4.1. Relative income hiding in the production sector

In this section, we concentrate on detecting hidden earnings in the production sector of the economy, so we exclude federal and city government employees and employees in public services (law enforcement and state-run education, research and health care systems). The next section deals separately with income underreporting in government and public services.[12] We control for individual observable characteristics such as age, gender, ethnicity, the number of members of the household and the number of earners in the household to account for life cycle

[12] We also conducted the estimations presented below including observations on government employees, and the results were basically the same.

considerations,[13] as well as for differences in family composition and preferences. We also include year dummies and the percentile of earnings of a given individual in the earnings distribution of his/her employer and its square term.

4.1.1. Ownership and size

We first use equation (4) to estimate the impact of firm ownership and size on employment contract transparency by including ownership dummies and (the log of) firm size in the vector of employer characteristics $\mathbf{S_{it}}$. Table 5 presents the results for the baseline value of the income elasticity parameter $\lambda = 0.35$ and for parameter values $\lambda = 0.25$ and $\lambda = 0.45$ to check the sensitivity of our findings.[14]

The omitted ownership dummy corresponds to private domestic firms. Compared to those, foreign ownership has large positive effect on income reporting, robust across all specifications. The effect has high economic importance: in the baseline specification, foreign-owned firms report on average 3.43 times (exp(1.487)-1) higher earnings paid to their employees with the same car values than do private Russian firms.

Firm size is a significant determinant of income underreporting as well. Doubling firm size increases reported incomes for the same car values by 34.9 percent. Firms at the 95th percentile of the firm size distribution (FSD) are estimated to report 2.85 times more earnings for the same car values than do firms at the 5th percentile of FSD. The coefficients on the state ownership dummy, however, are much smaller and statistically barely significant.

Among the demographic variables, only age has a statistically significant effect on the transparency of employment contracts. As expected from life-cycle considerations, older car owners are somewhat more likely to have higher officially reported earnings for the same car values. Gender, ethnicity and family composition variables are not statistically significant and the coefficients are not shown.

The estimated effects of the percentile in employer's earnings distribution (EED) and its square term, on the other hand, are highly statistically significant and economically large. The

---

[13] Hidden earnings are not counted when calculating social security payments, so workers who are closer to retirement age may prefer employment contracts with less hidden earnings than do younger people.

[14] We also did our estimation for car-income ratios (which amounts to setting $\lambda = 1$). The results were even stronger than reported here.

relationship is also non-linear. The coefficients imply that the transparency of employment contracts is increasing until about the 65^{th} percentile in the EED, after which it starts declining. Car owners at the 30^{th} percentile in the overall earnings distribution and in the top 1 percent are both estimated to report about 30 percent less earnings for the same car values than car owners at the 65^{th} percentile.[15]

Coefficients on year dummies (with 1999 as a basis for comparison) are positive, significant and increasing over time until 2002 in the baseline specification. In particular, the magnitude of the coefficient more than doubles from 2000 to 2001. This may reflect a one-time increase in the transparency of reporting following the flat income tax reform (cf. Gorodnichenko et al. [2009]).

4.1.2. Relative hidden earnings by industries in the private sector

We next estimate regression (4) including industry dummies. To focus on relative hiding across industries in the domestic private sector, we exclude state-owned enterprises, along with government employment and foreign-owned firms. We also do not include the firm size variable in this regression. The average size of establishments varies considerably across industries, with industries such as trade being populated by much smaller-sized firms than industries such as utilities. Thus, coefficients on industry dummies in this section estimate the relative composite effect of a particular industry, including differences stemming from establishment size, which we think is by itself an interesting question.[16] The omitted industry is banking and finance.

The coefficients on various industries are shown in Table 6, in the same three specifications as in Table 5. All regressions include demographic and time controls as in Table 5 although we do not show the corresponding estimates (both the magnitude and the statistical significance of the coefficients on age, position in EED, its square term and year dummies

---

[15] We were worried about company cars provided to top managers, which may result in less need for personal cars. If this were the case, however, there would be more, not less earnings reported for the same car values at the top of the earnings distribution. We also re-estimated regression (4) excluding top managers (those above the 90^{th} percentile in EED) and the results were qualitatively the same.

[16] We also estimated regression (4) including both industry dummies and log number of employees as explanatory variables. The qualitative features remain the same, but most industry dummies become statistically insignificant presumably due to multicollinearity with size. The coefficient on log number of employees itself, on the other hand, is estimated to be basically the same as in Table 5.

remain basically the same).

Compared to banking and finance, utilities – which is the most capital-intensive and heavily regulated industry – exhibits much higher reported earnings for the same values of cars. The magnitude of the coefficient on the utilities sector dummy in the baseline specification implies the difference in transparency with the banking sector of 118 percent. Employers in wholesale and retail trade, on the other hand, are estimated to report 43 percent (exp(-0.559)-1) less earnings for the same car values as employers in banking and finance. We also find substantial underreporting (relative to the banking industry) in services, private security, sports and entertainment, and self-employment even though high standard errors make most coefficients statistically insignificant at conventional levels.

One possible concern about the results presented above is that they do not take into account household composition. For example, a high-earning worker may buy a car not just for himself but also for a spouse or a child, and individual-level data would not reflect this properly on either side. To check if this affects our findings, we re-estimated regression (4) with earnings and car values aggregated to the household level.[17] The coefficients on firm ownership and size, as well as statistically significant coefficients on industry dummies in Table 6, are very similar (details are available upon request).

4.2. Estimating absolute income hiding and corruption: Foreign-owned firms as a benchmark

The analysis in the previous section established the heterogeneity of hidden earnings across employer categories but it did not give us a sense of the absolute amounts hidden in each category. We also excluded government employees and employees in public services. In this section, we gauge the absolute amounts of hidden earnings in various industries and the amounts of side incomes in government employment overall and year-by-year, using the estimation equation (4) with the omitted category representing a group assumed to have accurate income reporting: foreign-owned firms with 20 or more car owners in the sample. To increase the statistical power, especially for year-by-year estimations, we report the estimation results based on comparing car owners in our sample with a larger sample of car owners from foreign-owned

[17] In household regressions instead of sectoral dummies we included shares of different sectors earnings in total household income to measure income underreporting in different industries.

firms, including the oversampled part (see Appendix 5 for details of the oversampling procedure) Limiting the benchmark to observations from our original sample leads to basically the same results for all years pooled together, but coefficients and standard errors become unstable in year-by-year regressions because of too few observations remaining in the foreign-owned companies' benchmark.

4.2.1 How much do different industries hide?

Table 7 presents the estimation results for all years pooled together. Year-by-year results are reported in Appendix 7. Private and state ownership dummies are interacted with industry (sector of employment) variables to distinguish between the amount underreported in private and state-owned firms within the same industry. We only show the results using the baseline value of $\lambda = 0.35$ but the results are practically not sensitive to using alternative values of the parameter λ.

The vast majority of coefficients on industry dummies are negative and high in absolute value, indicating large underreported earnings in many industries in both the private and state-owned sectors. For example, the magnitude of the coefficient estimated on employment in a private banking, finance or insurance company implies that 88.3 percent of total earnings were not reported. In this particular case we had the opportunity to compare our estimates for 2003 against some direct evidence on two medium-sized private Russian banks provided to us by a person with insider knowledge of true individual earnings. The average fraction of hidden earnings among top and middle-level managers in those two banks in the year 2003 was 89.9 percent (see Appendix 8), while our estimates put the fraction of hidden earnings for the whole private banking sector in Moscow for the same year at exp(-2.17)-1, or 88.6 percent (Table A7.3)

In most other occupations the underreporting is of similar magnitudes. There are some exceptions, however. State-owned utility companies, in particular, come through as reporting earnings similar to the benchmark, as the coefficient on the dummy capturing employment in this industry is statistically indistinguishable from zero. And the coefficients on the dummies capturing the effects of employment in state-owned banking and finance, state-owned transportation and manufacturing companies as well as private secondary schools, although negative, are statistically not significant at conventional levels.

Further examination of these coefficients year-by-year reveals interesting patterns over time (see Appendix 7 for details). Overall there is a moderate trend towards more transparent income reporting, which is also captured by the coefficients on year dummies reported in Table 7 However, it is confined to the industries that were more transparent to begin with: utilities and manufacturing. There is almost no change for sectors like retail, wholesale trade, or services, which we previously identified to be among the top income hiders.

How widespread is income misreporting? Car owners employed in industries identified in Table 7 as having more or less transparent income reporting (including those employed in foreign-owned companies), comprise less than 10 percent of our sample of car owners. Car owners employed in industries where the estimated coefficient was around -2 or lower (indicating that 85 percent or more earnings are hidden), however, represent two thirds of the sample. Hence aggregate hidden earnings, at least among car owners in Moscow, are likely to be extremely high.

4.2.2 Estimating the amount of corrupt earnings in public services

The coefficients on the dummy variables representing employment in federal and Moscow city governments, law enforcement and state-run public education, research and health care are of particular interest. As argued before, "unreported earnings" in this case likely represent side payments stemming from semi-legal or outright corrupt sources.

As can be seen from Table 7, the biggest fraction of "hidden earnings" as compared to the benchmark (87 percent) is found in public secondary education. This does not necessarily come as a surprise. Even though the police tops the list of institutions perceived as "extremely corrupt" in Russia (49 percent of respondents subscribed to this view, according to The Global Corruption Barometer),[18] the education system also gets a very high score, with 27 percent of respondents describing it as "extremely corrupt" (*ibid*.).[19] The state-run higher education and research system is the second most corrupt public service sector in our sample. Compared to this we estimate that car owners employed in the police received "only" 79 percent of their incomes from hidden sources, while federal government employees are relatively even less corrupt, with

[18] See http://www.transparency.org/policy_research/surveys_indices/gcb/2005.

[19] As this paper was being written, a scandal erupted in one of the elite Moscow secondary schools, where the director and his colleagues were suspected of extorting over $7 million from parents over the course of three years, collecting illegal fees for everything, from admission to school to extracurricular activities http://lenta.ru/news/2010/09/30/school

"just" 73.5 percent less reported earnings for the same car values as the benchmark.

We can also detect a time trend toward a substantially smaller gap between car values and earnings over time for the federal government and law enforcement agencies (see Appendix 7). The latter two categories of government employees received large pay increases coupled with tougher enforcement of anti-corruption measures in the first few years of the Putin presidency. These measures appear to have produced some effect, although of a partial nature. Even in 2003 55.5 percent of actual earnings remained "unaccounted for" in the official data for federal government employees and 62.7 percent for law enforcement officials. Our estimates detect even less progress in corruption in education and health care, and no progress in the Moscow city government.[20]

5. Alternative explanations and robustness

The gap between car values and reported earnings is so large that it seems impossible to explain it by anything else but rampant income hiding and corruption. Still, we entertain various alternative explanations in this section and conduct several robustness checks. Even though they do not change our conclusions, these checks shed some additional light on the nature of income hiding in businesses and government.

5.1. Evidence from prestigious cars

As mentioned above, the interpretation of our results as measuring the degree of hiding hinges on the assumption that the distribution of the demand for the stock of cars among car owners is independent of employers' characteristics. For example, if employees of private Russian firms somehow had a higher demand for expensive cars than did employees of foreign-owned firms, then employees of foreign-owned firms would have a higher estimated value of the coefficient β_k. Similarly, if employees of small firms or in some specific industries used their passenger cars as a capital asset for work (commercial vehicles are excluded from our sample), this could also explain larger car value-earnings gaps in those firms and industries. We probe this by looking at the distribution of the types of cars across different groups of car owners.

One of the biggest symbols of high-income status in Russia is the ownership of a

[20] As this paper was being completed, President Medvedev fired the long-standing Mayor of Moscow amidst wide-spread allegations of corruption in the city administration.

prestigious car.[21] In contrast, owning a Russian-made car or a Chinese or Korean car (at least for the period covered in our analysis) is considered to be much less prestigious. In Table 8 we show the fraction of prestigious and non-prestigious cars owned in our sample, sliced by various employer characteristics. We also present average reported earnings for owners of prestigious and non-prestigious cars.

First, employment in foreign-owned firms is associated with a relatively high propensity to own prestigious cars and a low propensity to own non-prestigious cars. This is perhaps as clear an indication of higher earnings in foreign-owned firms as we could extract from our data. At the same time, this evidence is not consistent with interpreting the vastly higher coefficients on the foreign ownership dummy in our regressions above as reflecting less intensive preferences for owning expensive cars. If anything, employees of foreign-owned firms do exhibit a preference for fancy cars and nevertheless, as we saw, the gap between car values and reported earnings is several times lower for them than for car owners employed elsewhere.

Second, the relationship between earnings and the fractions of prestigious and non-prestigious cars is reversed for firms of small and large size. Employees of private firms in the smallest size quintiles have a fraction of prestigious cars that is 2.5 times higher than that of employees of private firms in the largest quintile. The fraction of non-prestigious cars is also monotonically increasing from smaller to larger firms. This is not consistent with employees of smaller firms using their cars as workhorses for business. Similarly, the fractions of prestigious cars are higher while the fractions of non-prestigious cars are lower than the average in the private sector among car owners employed in trade, services, sports, entertainment, and private security. Once again, it seems rather implausible that luxury cars would be used to run business errands, so car owners in these industries likely own expensive cars because they can afford them even though their officially reported earnings seem to tell a very different story.

Finally, the patterns found in public services provide further evidence that a lot of hidden earnings do represent incomes from idiosyncratic side earnings. In particular, public secondary education and health care sectors have average reported earnings well below average for the state sector as a whole. Nevertheless, the fractions of prestigious cars owned by workers

---

[21] We assigned a car to this category based on its characteristics. Included are higher-end models of Audi, BMW, Mercedes, Volvo, Saab, Lexus, Infinity, some high-end models of Alpha-Romeo, Rover, Renault, Toyota, Ford, GM and Chrysler, as well as Cadillacs, Porsches, Jaguars, Bentleys, etc. The full list is available upon request.

in these sectors are higher than for other government/state sector employees. At the same time, the fractions of non-prestigious cars in these sectors are also quite high, suggesting a sharp differentiation among corrupt and non-corrupt car owners. The same pattern can be detected in law enforcement, although to a less extent. In contrast, workers in utilities have the second-highest fraction of non-prestigious cars and the lowest fraction of prestigious cars among all categories, despite the fact that reported earnings in this sector are considerably above average.

The differences in average earnings between owners of prestigious and non-prestigious cars are also telling. Even though the *levels* of reported earnings are very different for car owners employed in private and foreign-owned firms, the *ratios* of reported earnings between owners of prestigious and non-prestigious cars are quite similar. This renders support to the notion that employment contracts in the private sector hide a fraction of true earnings but preserve ordering. The same can be seen to be generally true across the firm size distribution and for various private industries in Table 8 – in all these cases, with the notable exception of sports and entertainment sector, reported earnings of owners of prestigious cars are significantly higher than those of owners of non-prestigious cars.

The reversal of relative reported earnings between owners of prestigious and non-prestigious cars in the sports and entertainment sector is, without doubt, an indicator that hidden earnings in this sector are idiosyncratic. The same pattern, not surprisingly, is observable in public services identified previously as those with the largest corrupt incomes, namely, secondary education, health care and Moscow city government.

5.2. Evidence from high-paying jobs

It can be argued that selecting the sample based on car ownership may bias our estimates of the car value-earnings gap, especially for low-income individuals. More specifically, suppose that low-paid workers for the most part can not and do not buy cars, but some of them may still end up owning a car because they could afford it due to, for example, a big positive income shock that happened before 1999 or because they had saved money for many years.

To probe this, we re-estimated equation (4) with car owners employed by foreign-owned firms as a benchmark, but excluding all workers below the median in the employer's earnings distribution (EED) and then also those below the 75^{th} percentile. Since the relative positions in EED may mean little in small firms, we also limited the sample to car owners employed in establishments with 50 or more employees. Apart from providing a robustness check for our previous results, these estimates are interesting on their own. Especially in the sample limited to car owners above the 75th percentile we are looking at the cream of the crop of the Moscow workforce, government employment and public services.

Table 9 presents the estimation results. The magnitudes of some coefficients do decline somewhat compared to Table 7, but the qualitative picture remains remarkably similar. In particular, elite workers in trade, services, private security, and not-for-profits are still estimated to hide 80 percent or more of their actual earnings, while incomes from unaccounted sources comprise 60 percent or more of actual earnings of top-ranking public servants in federal government, education, law enforcement, and health care.

We also compare the occupational distribution of car owners and their earnings in our sample with the statistical data for the whole Moscow workforce (Figures A3.1 and A3.2 in Appendix 3). Under the alternative explanation above, our sample should have very few car owners in low-paying occupations and those car owners, furthermore, should have officially reported earnings not significantly different from non-car owners.

The evidence presented in Appendix 3 is not consistent with this. For example, our sample does contain relatively more individuals employed in the higher-paying banking and finance sector and relatively fewer individuals employed in the lower-paying utilities sector than the Moscow workforce as a whole. But given car values, car owners employed in utilities have actually higher, not lower, reported earnings compared to car owners employed in banking and finance, contrary to the notion that the former can afford to have a car only because of an unobserved income shock. Moreover, average reported earnings of car owners in our sample are significantly higher than the corresponding earnings reported by official statistics for the whole of Moscow in all employment sectors, high-paying and low-paying alike.

5.3. Fixed Effects Estimation

Our empirical specification in (4) and the estimations presented in the previous sections did not include fixed effects. One reason why using a fixed-effects estimator in the context of our data might be problematic is the time lag in adjusting the stock of cars, which is likely to be sizeable relative to our five-year panel. Another reason is that fixed-effects may produce biased estimates of the coefficients on our variables of interest. Other things equal, individuals who choose to work in sectors with more transparent employment contracts (such as foreign-owned firms) have "revealed preference" for avoiding large fractions of "black cash" in their earnings. Hence, even when they move to sectors where hidden earnings are on average higher, we would expect them to sort to employers with relatively less hidden earnings, so that $E\left[\eta_{it+1}\right]-E\left[\eta_{it}\right]>0$ in equation (1) above, attenuating the estimate of the corresponding coefficient β. More generally, if preferences for transparency of reporting are "sticky," we would expect sorting in accordance with those among sector movers to result in $E\left[\eta_{it+1}\right]-E\left[\eta_{it}\right]$ having the opposite sign to that of $\beta_{k't+1}-\beta_{kt}$. Thus, fixed effects estimates are likely to provide lower bounds on income underreporting in different sectors.

Nevertheless, we did check how our results are affected by controlling for individual fixed effects and confirmed that the qualitative features of the analyses below remain intact, although the magnitudes of most coefficients of interest become smaller. The results of these estimates, with foreign-owned companies as a benchmark, are presented in Appendix 9. For example, employment in wholesale and retail trade, which is estimated to be associated with hiding 94 percent of earnings in our preferred pooled OLS estimation, is estimated to be associated with hiding 78 percent of earnings in the fixed-effect regression. Similarly, the estimated fraction of hidden earnings in public secondary education is 78 percent using fixed effects, as opposed to 87 percent in pooled OLS, the estimated fraction of hidden earnings in law enforcement is 60 percent in fixed-effects estimates as opposed to 79 percent in pooled OLS and so on. We also conducted fixed-effect estimates for high-paying jobs only as in the previous subsection and the results, once again, were similar.

5.4. Income changes and car purchases of job changers

We also examine the subsample of job changers more directly to see whether changes in their income reporting and car ownership agree with our results above. The basic idea is very simple. Suppose, for example, that the large differences in earnings between foreign-owned and private Russian firms were real. Then a person from a domestic company who switches to a lucrative foreign-company job is probably more likely to purchase a better car then a person who remains employed in the domestic or foreign sector the whole time. Similar logic applies to moving from smaller to larger firms and between different industries within the private sector.

Thus, we can look at probabilities of purchasing new cars and changes in values of cars to see if these are consistent with the changes in recorded earnings associated with a move.

5.4.1. Movers across ownership types

Table 10 presents the results of three regressions estimating the effects of moving from an employer with one type of ownership in year *t*–1 to an employer with a different type of ownership in year *t* for the same individuals on three different variables, (A) the probability of purchasing a new car in year *t*; (B) the difference in logs of car market values between year *t* and year *t*-1; and (C) the difference in logs of (reported) earnings between year *t* and year *t*-1. The omitted category is workers who stayed employed with a foreign-owned employer. All three regressions control for demographic characteristics and year dummies, although the coefficients are not reported.

As can be seen from the last column of Table 10, switching to a foreign-owned employer results in a big upward jump in reported earnings (by exp(0.885)-1=142 percent when moving from the private sector and by exp(0.963)-1=198 percent when moving from the state/government sector). However, the hypothesis that these individuals would be more likely to purchase new cars following such a lucrative move compared to those who are continuously being employed in the foreign-owned sector is strongly rejected. A similar effect is observed among car owners who stay in the Russian private and state/government sectors; their year-to-year reported earnings increase relative to reported earnings of car owners who are continuously employed by foreign-owned firms (by 14 and 17 percent, respectively) but their probabilities of purchasing a new car are statistically and economically the same as those of car owners continuously employed by foreign-owned firms, while relative car values even decline by 5.5 and 8.5 percent, respectively. The increase in relative earnings thus seems to reflect increased transparency over time, rather than any real relative income changes between these sectors and the foreign-owned sector.

Among negative changes in reported earnings, the effect of moving from a foreign-owned firm to a Russian-owned private firm is the most strongly pronounced one (resulting in a 37 percent drop in earnings compared to staying with a foreign-owned employer). Once again, this large negative income shock is not associated with any significantly lower probability of purchasing a new car, or decrease in car stock. It is also noteworthy that the drop in reported earnings when moving from a foreign-owned firm to a private firm is much less than the increase in reported earnings associated with the opposite move, while the move from a foreign-owned firm to a state-owned firm can not be meaningfully associated with any such drop (the coefficient on the corresponding dummy is positive and large in absolute value but not significant statistically as the number of observations is rather small). This suggests that car owners who had been employed in the foreign-owned firms tend to carry the "culture of transparency" with them even when they move outside of the foreign-owned sector.

5.4.2. Changes across firm size and industries

Table 11 reports the results of an exercise similar to the above, where we estimate the impact for the same individuals of moving to a new employer of a larger or smaller size and into two big industries previously identified as among the most polar by the transparency of reported earnings: utilities and wholesale and retail trade. The dummy for moving to a new employer of a larger (smaller) size is set equal to one if the individual changed employers between year t-1 and t and the new employer belonged to a higher (lower) quintile of firm size distribution than the previous employer, zero otherwise. The dummy capturing moving into utilities (wholesale or retail trade) is set equal to one if the individual derived his or her primary earnings from some other industry in the private sector in year t-1 and his or her primary earnings from utilities (wholesale or retail trade) in year t, zero otherwise. The dependent variables are the same as in Table 10: the probability of purchasing a car, the differences in log car values and the differences in log of (reported) earnings.

Once again, the results are consistent with our previous findings. For example, a worker moving up one quintile in establishment size experiences on average an increase in reported earnings of 32.5 percent as compared to the omitted category (workers staying with the same employer or moving to a new employer in the same size quintile), while a worker making the opposite move experiences a decrease in reported earnings of 26 percent. Nevertheless, we do not see any significant differences in the probability of purchasing a car, and the increases in the values of owned cars are also indistinguishable from one another.

Our hypothesis is that the decision how much to hide is basically a firm-level decision. A very strong test of this hypothesis would be to compare what happens to reported earnings and car values of those individuals who do not change jobs, but whose employer moves up or down the firm size distribution. The last two rows in Table 11 report the result of this test where the sample is limited to car owners whose main source of earnings was from the same employer in the Russian private sector between two adjacent years. There is no effect of the firm becoming larger (in 15 percent of observations) or smaller (in 9 percent of observations) on the value and purchases of cars, but reported earnings in firms that shrink in size between the two years go down by about 10 percent while reported earnings in firms that increase their size go up by about 14 percent. It thus appears that firms that are downsizing simply start hiding more of their earnings and vice versa (so that part of the decision to become smaller may be related to the decision to increase the fraction of hidden activity).

6. Conclusions and discussion

We have developed a novel approach to measuring the hidden economy by juxtaposing reported incomes with cars owned by the same people. Using administrative data on earnings of Moscow car owners together with data on cars they own we obtained a striking result. Over the period of five years from 1999-2003 hidden earnings may have on average constituted more than 75 percent of total earnings. These findings have profound implications for the ways economists interpret Russian economic performance since the collapse of communism.

Before its collapse in 1991 the Soviet Union was viewed as the world's second superpower, although in hindsight it is clear that the strength of the Soviet economy was exaggerated. The analyses of the Russian transition to a market economy, however, may have gone too far in the opposite direction because of underestimating the size of the hidden economy. Official sources have put the fraction of the shadow economy in Russia at about 20-25 percent. While some economists have come up with larger numbers (e.g., 45.6 percent in Alexeev and Pyle [2003], using electricity consumption data and extrapolating from their estimates of the fraction of hidden economy in the former Soviet Union), our estimates indicate that it might be even much higher than that. Of course, our sample of car owners in Moscow may be not representative of the whole economy. But even if that were the case, our findings still carry important policy implications. In particular, they imply that the capital city might account for a much larger fraction of the actual economic activity in the country than the officially measured 22-24 percent. Given that car owners also represent the elite part of the workforce and of the federal government, this virtually guarantees that no economic reform policy can be meaningful without addressing the issue of the hidden economy and corruption first.

In more recent years, the Russian economy appears to have been resurgent. The real GDP index had bounced back to 70.7 percent of its 1990 level by 2001, to almost 80 percent by 2003, and to 96.7 percent by 2006. The nominal GDP converted to US dollars at the market exchange rate increased by 152 percent between 1999 and 2003 for the country as a whole and by 188 percent for the city of Moscow.[22] Officially reported earnings in our data recorded very similar increases over the same period (161 percent in US dollar terms on average for private firms and 184 percent on average for the government/state sector). The implied growth in car values (with income elasticity equal to 0.35 as in our baseline specification) is 56.5 and 64.5 percent respectively. Instead, car values only grew by 37 percent in the private sector and by 24.3 percent in the state-owned sector. These numbers suggest that increased transparency might indeed be responsible for a large share of growth in statistically measured GDP (cf. Gorodnichenko et al. [2009]). The increase in transparency may have also been more pronounced in the state-owned than in private sector.

Our disaggregated estimates of the hidden economy corroborate the long-held beliefs that hiding is more prevalent in some sectors of the economy than in others. In line with these beliefs, we found that smaller firms hide much more than large firms, firms in labor-intensive industries hide much more than firms in capital-intensive industries and foreign-owned firms hide dramatically less than indigenous firms. Even though the general direction may not be surprising, it was important to put some numbers on these qualitative perceptions.

For example, one of our most robust findings is that when hidden earnings are taken into account, the firm-size effect on wages is reversed. Larger firms only seemingly pay more to their workers but in fact, they are paying less. In the industrial organization literature (that mostly uses theoretical models developed for institutionally sound market economies and the corresponding empirical evidence), larger firm size has been firmly associated with higher productivity. Our analysis indicates that although the official data from countries with less than stellar institutional environments may suggest the same, that is not necessarily the case in reality. At the same time, our study also issues a strong note of caution with regard to policy measures promoting small businesses for development. Such measures, if not accompanied by provisions ensuring increased transparency and openness, can end up feeding the hidden economy at the expense of open-market economic activity.

---

[22] Our calculations based on official statistical data available at http://www.gks.ru/dbscripts/cbsd/DBInet.cgi

Our findings also have implications for the way in which the role of foreign direct investment is perceived for countries like Russia. According to the administrative data (which form the basis of official statistics), the performance of foreign-owned firms in Moscow was nothing short of miraculous as they managed to generate wages for their employees 200 percent or higher than domestic firms. In reality we cannot even be sure whether the earnings in foreign-owned firms were indeed higher than in private Russian firms, so the appearance of a "miracle" is due to a stark difference in the transparency of reporting. It thus appears that foreign direct investment cannot be counted on to quickly elevate the true productivity of the economy.

At the same time, our findings indicate that foreign-owned firms do introduce the culture of more transparent accounting into the economic environment characterized by a lot of hidden activity. Just as we did in this study, Russian regulatory authorities can use those firms as a benchmark to put pressure on other firms by targeted audits of those of them that fall behind most egregiously in reported earnings. This seems to be a realistic way of gradually squeezing out the hidden economy, at least in areas where the regulatory authority has the means of monitoring the situation closely. Promoting foreign direct investment and foreign ownership, while not a panacea, can thus be seen to be important in bringing economic activity more into the open – something that may be a pre-requisite for productivity growth. We intend to develop this insight further in our future work.

We also found disturbingly large hidden earnings, most likely from corrupt sources, in all areas of government employment. A corrupt government and law enforcement not only lack credibility and moral authority to fight against tax evasion in the private economy; they also have incentives *not* to fight it seriously so as to increase sources of their corrupt incomes. This problem is going to complicate Russia's push to a more transparent economy for years to come. But a potentially even more disturbing finding is that corruption seems to be at its worst in the state-run education system. In particular, a lot of concerned voices have pointed out that the quality of education is deteriorating sharply. The huge gap between official and estimated earnings of educators that we found corroborates this view, albeit in an indirect way. The hidden economy is eating up not just Russia's present generation but also corrupts its future generations, and this may be the worst news to come out of our study.

We conclude by noting some general lessons. Lacking firm statistical evidence, economists are understandably reluctant to embrace anecdotes suggesting that the hidden economy has a much larger scale than suggested by survey data or macroeconomic indicators. The increased availability of matched employer-employee datasets, and the relative ease with which data on personal assets such as cars can be available and matched to those datasets mean that our methodology can perhaps find more widespread use. This can be expected to contribute to better understanding of the magnitude and, most importantly, the structural aspects of the hidden economy and corruption, something badly needed for policy evaluations and recommendations.

References

Abowd, John M., Francis Kramarz, and David N. Margolis, 1999. "High Wage Workers and High Wage Firms," *Econometrica*, 67 (2), 251-333.

Alexeev, Michael, and William Pyle, 2003. "A Note on Measuring the Unofficial Economy in the Former Soviet Republics," *Economics of Transition*, 11 (1), 153-175.

Bloom, Nicholas, Aprajit Mahajan, David McKenzie, and John Roberts, 2010. "Why Do Firms in Developing Countries Have Low Productivity?", *American Economic Review: Papers and Proceedings*, 100 (2), 619-623.

Brown, Charles, and James Medoff, 1989. "The Employer Size-Wage Effect," *Journal of Political Economy*, 97 (5), 1027-1059.

Cagan, Phillip, 1958. "The Demand for Currency Relative to the Total Money Supply," *Journal of Political Economy*, 66 (4), 303-328.

Djankov, Simeon, Ira Lieberman, Joyita Mukherjee, and Tatiana Nenova, 2003. "Going Informal: Benefits and Costs," Chapter 3 in Belev, Boyan, *The Informal Economy in the EU Accession Countries: Size, Scope, Trends, and Challenges to the Process of EU Enlargement,* Sofia, Center for the Study of Democracy.

Feinstein, Jonathan S., 1999. "Approaches for estimating noncompliance: examples from federal taxation in the United States," *Economic Journal*, 109 (June), F360-F369.

Fiege, Edgar L., and Ivica Urban, 2008. "Measuring Underground (Unobserved, Non-Observed, Unrecorded) Economies in Transition Countries: Can We Trust GDP?" *Journal of Comparative Economics*, 36 (2), 287-306.

Frey, Bruno and Hannelore Weck-Hanneman, 1984. "The Hidden Economy as an 'Unobserved' Variable," *European Economic Review*, 26 (1-2), 33-53.

Frye, Timothy, and Zhuravskaya, Ekaterina, 2000. "Rackets, Regulation, and the Rule of Law," *Journal of Law, Economics and Organization*, 16 (2), 478-502.

Giles, David E. A., 1999. “Measuring the Hidden Economy: Implications for Econometric Modelling’, *Economic Journal*, 109 (June), F370–F389

Gorodnichenko, Yuri, Jorge Martinez-Vazquez, and Klara Sabirianova, 2009. “Myth and Reality of Flat Tax Reform: Micro Estimates of Tax Evasion Response and Welfare Effects in Russia,” *Journal of Political Economy*, 117 (3), 504-554.

Guriev, Sergei and Andrey Rachinsky, 2006. “The Evolution of Personal Wealth in the Former Soviet Union and Central and Eastern Europe.” Working Papers RP2006/120, World Institute for Development Economic Research (UNU-WIDER). Published in Davies, James, ed. *Personal Wealth from a Global Perspective*, Oxford University Press: 2008.

Holger Görg, Eric Strobl, and Frank Walsh, 2007. “Why Do Foreign-Owned Firms Pay More?” *Review of World Economics*, 143 (3), 464-482.

Ivanova, Anna, Michael Keen, and Alexander Klemm, 2005. “The Russian ‘flat tax’ reform.” *Economic Policy*, Vol. 20 (43), pages 397-444.

Johnson, Simon, Daniel Kaufmann, John McMillan, and Christopher Woodruff, 2000. “Why Do Firms Hide? Bribes and Unofficial Activity After Communism,” *Journal of Public Economics*, 76, 495-520.

Kaufmann, Daniel, and Aleksander Kaliberda, 1996. “Integrating the Unofficial Economy into the Dynamics of Post-Socialist Economies: A Framework of Analysis and Evidence,” World Bank Policy Research Working Paper, 1691.

Klepper, Steven and Daniel Nagin, 1989. “The Anatomy of Tax Evasion,” *Journal of Law, Economics and Organization*, 5 (1), 1-24.

Lyssiotou, Panayiota, Panos Pashardes, and Thanasis Stengos, 2004. “Estimates of the Black Economy Based on Consumer Demand Approaches,” *Economic Journal*, 114 (July), 622-640.

Pissarides, Christopher A., and Guglielmo Weber, 1989. “An expenditure-based estimate of Britain’s black economy”, *Journal of Public Economics*, 39 (1), 17–32.

Sabirianova, Klara, Jan Svejnar, and Katherine Terrell, 2005. “Foreign Investment, Corporate Ownership, and Development: Are Firms in Emerging Markets Catching Up to the World Standard?” IZA Discussion Paper, No. 1457.

Schneider, Freidrich, and Dominik Enste, 2000. "Shadow Economies: Size, Causes, and Consequences," *Journal of Economic Literature*, 38 (1), 77-114.

Shleifer, Andrei, and Daniel Treisman, 2005. "A Normal Country: Russia After Communism," *Journal of Economic Perspectives*, 19 (1), 151-174.

Strobl, Eric, and Robert Thornton, 2002. "Do Large Employers Pay More in Developing Countries? The Case of Five African Countries," IZA Discussion Papers 660, Institute for the Study of Labor (IZA).

Tanzi, Vittorio, 1983. "The underground economy in the United States: annual estimates 1930–1980," *IMF Staff Papers*, 30, 283-305.

Table 1. Average earnings and car values (1999-2003)

Years	# of obs.	Earnings (E)			Car values (C)			Ratio C/E
		Mean	Std. Dev.	Mean/ St.Dev.	Mean	Std. Dev.	Mean/ St.Dev.	Means
Average	22,862	6,049	10,509	1.74	7,446	11,061	1.49	1.23
1999	3,392	3,760	8,328	2.21	6,347	10,485	1.65	1.69
2000	3,970	4,341	8,335	1.92	6,423	9,733	1.52	1.48
2001	4,697	6,017	10,636	1.77	7,424	10,962	1.48	1.23
2002	5,163	6,644	10,789	1.62	7,670	10,695	1.39	1.15
2003	5,640	8,108	12,109	1.49	8,641	12,469	1.44	1.07

Note: the sample consists of car owners with non-zero officially reported earnings in all five years, excluding those with 5 or more cars, 17 and younger in 1999 and 61 and older in 2003, those earning less than the annual equivalent of the legal minimum wage or more than the equivalent of US $100,000 in a given year and those whose primary earnings in any year 1999-2003 came from lottery winnings, veteran charitable foundation or interest and dividend incomes from the state savings bank or two largest state-owned corporations. COV is the coefficient of variation (standard deviation over the mean).

Table 2. Car owners in the sample in the overall earnings distribution

	All car owners	Private employment	Government employment
Mean (st.dev.)	0.74 (0.24)	0.75 (0.24)	0.70 (0.24)
Percentiles			
5	0.25	0.27	0.23
10	0.37	0.38	0.32
25	0.58	0.60	0.54
50	0.80	0.81	0.76
75	0.94	0.94	0.91
90	0.99	1.00	0.97
95	1.00	1.00	0.99

Note: The numbers represent the percentile of car owners in the sample in the overall earnings distribution in all databases at the mean and at the corresponding percentiles. For example, a car owner at the fifth percentile in the sample is at the 25th percentile in the earnings distribution overall and so on. Private employment includes car owners employed in domestic private employers. Government employment includes car owners employed in federal and Moscow city governments, law enforcement agencies, as well as in the state-run education, research and health care systems and the mass media.

Table 3. Earnings and car values by earnings deciles in 2003, US dollars

	Individuals				Households			
Earnings deciles	# of obs.	Earnings (E)	Car values (C)	C/E	# of obs.	Earnings (E)	Car values (C)	C/E
1	466	529	7,558	14.29	424	844	10,787	12.78
2	427	1,038	8,184	7.88	425	1,865	11,152	5.98
3	425	1,594	8,806	5.53	424	2,933	8,750	2.98
4	406	2,318	7,088	3.06	423	4,062	10,603	2.61
5	414	3,154	7,310	2.32	424	5,373	9,481	1.76
6	419	4,150	6,194	1.49	426	6,856	10,751	1.57
7	427	5,358	6,429	1.20	426	8,798	11,139	1.27
8	449	7,022	6,898	0.98	424	11,559	11,641	1.01
9	461	10,687	8,259	0.77	426	16,982	15,307	0.90
10	487	40,752	14,299	0.35	424	54,159	21,262	0.39

Note: the sample is limited to 2003 wage earners (households with primary earner a wage earner) with non-zero wage earnings in at least 9 months, with no more than 5 cars per individual (10 per household).

Figure 1. Individual earnings and car values from Table 3, ratio scale

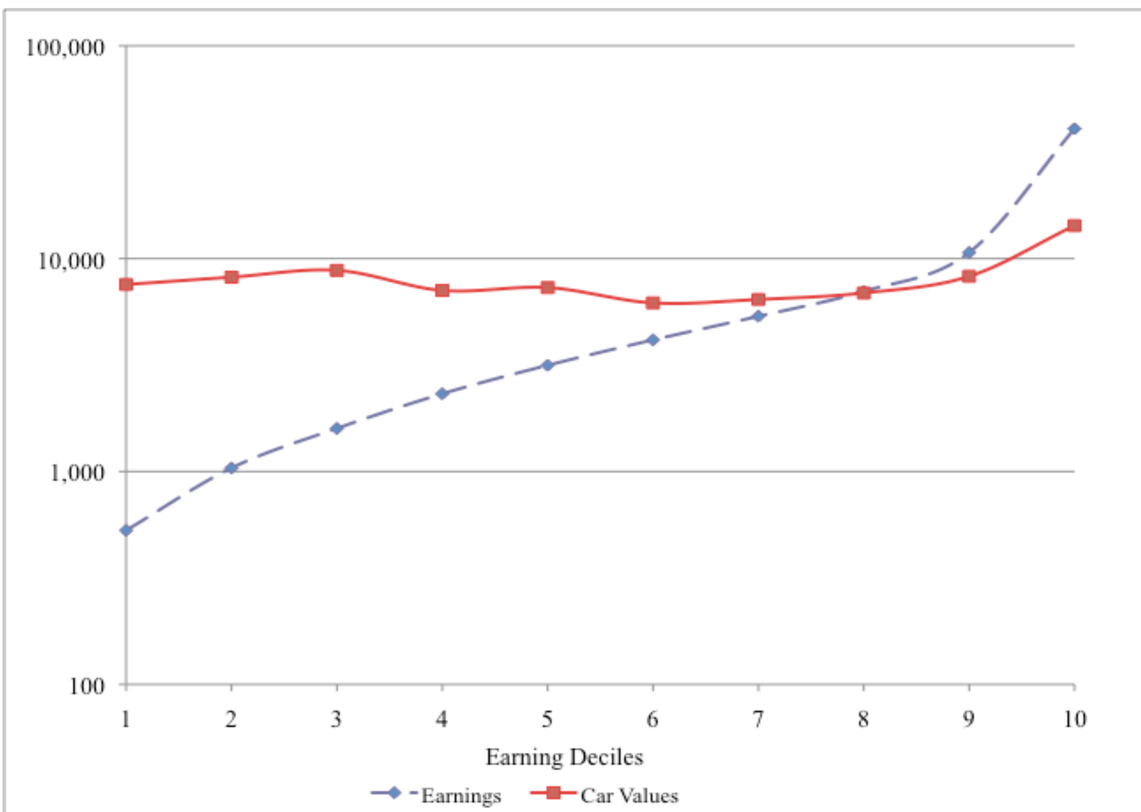


Table 4. Car values and earnings in 2003 by some employer types

	Means			Mean/St.Dev.	
	Earnings (A)	Car Values (B)	Ratio (B)/(A)	Earnings (A)	Car Values (B)
Private Russian firms	6,291	9,471	1.51	1.61	1.47
Government/state sector	6,514	6,331	0.97	1.15	1.37
Foreign-owned firms	19,473	9,296	0.48	0.93	0.97
Among Private Russian firms					
Wholesale and retail trade	4,626	9,683	2.09	1.81	1.37
Mass media, IT, utilities	9,683	7,987	0.82	1.14	1.25
Firm size below median	3,932	10,959	2.79	1.90	1.42
Firm size above median	8,665	7,973	0.92	1.36	1.49
Among government employers					
Law enforcement, secondary education, health care	3,862	6,471	1.68	0.69	1.47
Other government employment	7,197	6,294	0.87	1.13	1.34

Note: the sample is limited to 2003 wage earners with non-zero wage earnings in at least 9 months, with no more than 5 cars per individual.

Table 5. Estimates of regression (4): ownership and size

Specifications	$\lambda = 0.35$	$\lambda = 0.25$	$\lambda = 0.45$
Employer Characteristics (Ownership, Size), Private Domestic Dummy is omitted			
State ownership	0.275*	0.392**	0.210*
	(0.144)	(0.199)	(0.114)
Foreign ownership	1.487***	1.352***	1.563***
	(0.235)	(0.325)	(0.186)
Log number of employees	0.349***	0.406***	0.317***
	(0.020)	(0.027)	(0.016)
Individual (household primary earner) characteristics			
Percentile in EED	7.379***	9.443***	6.232***
	(0.778)	(1.075)	(0.617)
Percentile in EED squared	-5.741***	-7.995***	-4.489***
	(0.626)	(0.866)	(0.497)
Age in 2003	0.054***	0.074***	0.044***
	(0.005)	(0.007)	(0.004)
Year dummies			
2000	0.119**	0.077	0.142***
	(0.049)	(0.066)	(0.040)
2001	0.243***	0.081	0.333***
	(0.060)	(0.081)	(0.049)
2002	0.311***	0.083	0.438***
	(0.067)	(0.092)	(0.054)
2003	0.279***	-0.032	0.452***
	(0.072)	(0.099)	(0.058)
Constant	-22.701***	-33.420***	-16.745***
	(0.356)	(0.491)	(0.282)
Adjusted R^2	0.103	0.083	0.127

Notes: The dependent variable is the difference between Log of reported earnings and income elasticity-adjusted Log of car values: $\log E^R - 1/\lambda \log C$. The number of observations (individuals) is 15,972 (4,782) in all regressions. Pooled OLS with robust clustered standard errors in parentheses. ***, **, and * indicate significance at 1 percent, 5 percent, and 10 percent levels, respectively. All three regressions control also for gender, ethnicity, the number of members in the household and the number of earners in the household (coefficients are statistically insignificant and are not shown). Federal and city government employees and employees in public services (law enforcement, education and health care) are excluded. The sample is restricted to car owners with 5 or less cars, 18 and older in 1999 and 60 and younger in 2003, excluding those earning less than the annual equivalent of the legal minimum wage or more than the equivalent of US $100,000 in a given year and those whose primary earnings in any year 1999-2003 came from lottery winnings, veteran charitable foundation or interest and dividend incomes from the state savings bank or two largest state-owned corporations.

Table 6. Estimates of regression (4): industries in the private sector

Specifications	$\lambda = 0.35$	$\lambda = 0.25$	$\lambda = 0.45$
Industries/sectors (omitted sector: banking, finance and insurance)			
Mass media	-0.137	-0.140	-0.136
	(0.364)	(0.502)	(0.290)
Construction	0.203	0.462*	0.059
	(0.203)	(0.278)	(0.163)
Utilities	1.158***	1.456***	0.993***
	(0.309)	(0.428)	(0.244)
Transportation	0.464	0.837*	0.257
	(0.337)	(0.467)	(0.267)
Wholesale and retail trade	-0.559***	-0.435*	-0.628***
	(0.192)	(0.263)	(0.154)
Manufacturing	0.010	0.194	-0.092
	(0.193)	(0.266)	(0.154)
Sports, entertainment	-0.471	-0.254	-0.592
	(0.511)	(0.701)	(0.408)
Services	-0.324	-0.164	-0.413**
	(0.201)	(0.275)	(0.162)
Communications and IT	-0.005	-0.036	0.012
	(0.258)	(0.347)	(0.210)
Private security	-0.304	0.053	-0.503**
	(0.305)	(0.421)	(0.243)
Self employed	-0.713	-0.411	-0.881*
	(0.654)	(0.898)	(0.528)
Non-education not-for-profit	0.029	0.257	-0.098
	(0.327)	(0.440)	(0.267)
Constant	-21.375***	-32.080***	-15.427***
	(0.406)	(0.556)	(0.325)
Adjusted R^2	0.047	0.042	0.056

Notes: The dependent variable is the difference between Log of reported earnings and income elasticity-adjusted Log of car values: $\log E^R - 1/\lambda \log C$. The number of observations (individuals) is 13,682 (4,286) in all regressions. Pooled OLS with robust clustered standard errors in parentheses. ***, **, and * indicate significance at 1 percent, 5 percent, and 10 percent levels, respectively. All three regressions include also age, gender, ethnicity, the number of members in the household, the number of earners in the household and year dummies. The sample is restricted to car owners employed in Russian capital-owned private firms with 5 or less cars, 18 and older in 1999 and 60 and younger in 2003, excluding those earning less than the annual equivalent of the legal minimum wage or more than the equivalent of US $100,000 in a given year and those whose primary earnings in any year 1999-2003 came from lottery winnings, veteran charitable foundation or interest and dividend incomes from the state savings bank or two largest state-owned corporations.

Table 7. Hidden earnings and corruption relative to the benchmark

Ownership	Private	State	Ownership	Private	State
Banking, finance	-2.147***	-0.281	Manufacturing	-2.200***	-0.326
	(0.206)	(0.357)		(0.180)	(0.431)
Federal Government	NA	-1.327***	Sports, entertainment	-2.663***	-2.166***
	NA	(0.355)		(0.505)	(0.626)
City, local government	NA	-1.454***	Services	-2.502***	-1.736***
	NA	(0.372)		(0.189)	(0.601)
Law enforcement	NA	-1.557***	Communications and IT	-2.237***	-1.881***
	NA	(0.257)		(0.248)	(0.359)
Higher education	-2.675***	-1.761***	Private security	-2.478***	NA
	(0.388)	(0.235)		(0.298)	NA
Secondary education	-0.398	-2.073***	Self employed	-3.015***	NA
	(1.422)	(0.386)		(0.651)	NA
Health care	-2.231***	-1.659***	Non-education not-for-profit	-2.222***	-3.781***
	(0.562)	(0.310)		(0.313)	(0.932)
Mass media	-2.384***	-0.942*	Year dummies		
	(0.354)	(0.496)	2000	0.149***	
Construction	-1.996***	-1.541***		(0.041)	
	(0.190)	(0.266)	2001	0.230***	
Utilities	-1.033***	-0.059		(0.051)	
	(0.297)	(0.548)	2002	0.350***	
Transportation	-1.698***	-0.276		(0.057)	
	(0.329)	(0.277)	2003	0.413***	
Wholesale and retail trade	-2.787***	-2.367***		(0.062)	
	(0.177)	(0.562)	# of observ.	21,397	
			Adjusted R^2	0.088	

Notes: The dependent variable is the difference between log of reported earnings and income elasticity adjusted log of car values: $\log E^R - 1/\lambda \log C$. Income elasticity is $\lambda = 0.35$. Robust clustered standard errors in parentheses. The omitted category is the benchmark sample of foreign-owned employers in Appendix 5. Regressions also include age, gender, percentile in EED and its square term, and the constant term. The sample is restricted to car owners with 5 or less cars, 18 and older in 1999 and 60 and younger in 2003, excluding those earning less than the annual equivalent of the legal minimum wage or more than the equivalent of US $100,000 in a given year and those whose primary earnings in any year 1999-2003 came from lottery winnings, veteran charitable foundation or interest and dividend incomes from the state savings bank or two largest state-owned corporations.. ***, **, and * indicates significance at 1 percent, 5 percent, and 10 percent levels respectively.

Table 8. Prestigious and non-prestigious cars and earnings by employer types

	Fraction of:		Earnings (US$):	
	non-prestigious cars	prestigious cars	non-prestigious car owners	prestigious car owners
Ownership				
Private	64.76	13.34	3,392	5,195
Government/State	72.75	9.66	3,649	4,585
Foreign	50.00	17.57	11,133	15,764
Firm size quintiles in the private sector				
First	55.98	17.70	1,703	2,499
Second	66.44	13.46	2,785	5,999
Third	69.18	10.51	4,514	7,977
Fourth	72.71	9.45	4,911	9,550
Fifth	75.32	7.02	5,765	11,145
Industries (sectors) of employment				
Banking and finance (private)	59.47	15.34	6,797	9,795
Federal government	66.57	10.22	3,728	4,532
City and local government	68.68	10.57	4,204	3,103
Law enforcement	80.21	9.23	2,539	2,880
Higher education/research (public)	75.16	8.22	2,922	4,849
Secondary education (public)	71.81	11.28	2,082	1,036
Health care (public)	72.49	10.29	2,500	2,053
Mass media	58.87	17.10	5,381	11,207
Construction	76.58	9.23	3,033	4,799
Utilities	80.14	5.50	4,990	5,114
Transportation	74.59	8.12	3,452	4,412
Wholesale and retail trade	58.26	17.04	2,761	3,778
Manufacturing	68.70	11.80	3,452	4,913
Sports and entertainment	62.84	18.01	2,251	1,794
Services	61.05	13.92	3,214	5,494
Communications and IT	61.73	12.40	4,659	7,741
Private security	71.11	14.55	1,401	2,024
Not-for-profit	65.00	11.82	3,413	5,219

Note: Non-prestigious cars are Russian, Korean and Chinese-made cars. Prestigious cars include luxury and exotic brands (Porsche, Bentley, Jaguar, etc.) as well as higher-end models of Audi, BMW, Mercedes-Benz, Saab, Volvo, Lexus, Infinity, Cadillac, Toyota Crown, Rover, and so on. The full list is available upon request.

Table 9. Hidden earnings and corruption in higher-paying jobs

Employers with 50 or more workers	Sample limited to: Percentile in EED above median		Sample limited to: Percentile in EED above 75th	
Ownership	Private	State	Private	State
Banking, finance and insurance	-1.919***	-0.165	-1.761***	-0.027
	(0.222)	(0.375)	(0.265)	(0.409)
Federal Government	NA	-1.083***	NA	-1.193**
	NA	(0.418)	NA	(0.541)
City and local government	NA	-1.185***	NA	-0.308
	NA	(0.416)	NA	(0.514)
Law enforcement	NA	-1.500***	NA	-1.376***
	NA	(0.275)	NA	(0.342)
Higher education, research	-2.351***	-1.287***	-1.837***	-0.989***
	(0.484)	(0.262)	(0.564)	(0.299)
Secondary education	0.378	-1.676***	0.684	-1.321**
	(1.837)	(0.433)	(1.997)	(0.535)
Health care	-2.230***	-1.548***	-1.627***	-1.242***
	(0.494)	(0.339)	(0.554)	(0.424)
Mass media	-1.831***	-0.834	-1.826***	-0.520
	(0.379)	(0.537)	(0.470)	(0.588)
Construction	-1.290***	-1.335***	-1.278***	-1.113***
	(0.208)	(0.297)	(0.250)	(0.315)
Utilities	-0.956***	-0.309	-0.783**	-0.248
	(0.310)	(0.447)	(0.386)	(0.569)
Transportation	-1.001**	-0.278	-0.954*	-0.132
	(0.433)	(0.288)	(0.514)	(0.364)
Wholesale and retail trade	-2.043***	-2.296***	-2.040***	-2.478***
	(0.223)	(0.634)	(0.263)	(0.847)
Manufacturing	-1.624***	-0.414	-1.590***	-0.771*
	(0.200)	(0.460)	(0.227)	(0.418)
Sports, entertainment	-1.950***	-1.532***	-1.490	-1.424**
	(0.623)	(0.544)	(0.931)	(0.593)
Services	-1.763***	-1.666***	-1.964***	-1.910***
	(0.242)	(0.632)	(0.274)	(0.704)
Communications and IT	-1.670***	-1.827***	-1.648***	-1.438***
	(0.304)	(0.383)	(0.348)	(0.418)
Private security	-2.568***	NA	-2.482***	NA
	(0.317)	NA	(0.365)	NA
Self employed	-1.646*	NA	-1.901	NA
	(0.995)	NA	(1.561)	NA
Non-education not-for-profit	-2.119***	-3.956**	-2.093***	-3.739***
	(0.379)	(1.586)	(0.414)	(1.399)
# of observations, adj. R^2	13,375	0.058	9,812	0.063

Notes: The dependent variable is the difference between log of reported earnings and income elasticity adjusted log of car values: $log\, E^R - 1/\lambda\, log\, C$, with $\lambda = 0.35$. Robust clustered standard errors in parentheses. The omitted category is the benchmark sample of foreign-owned employers with 20 or more observations on car owners per year as in Appendix 5. Regressions also include age, gender, percentile in EED, year dummies, and the constant term. The sample is restricted to car owners with 5 or less cars, 18 and older in 1999 and 60 and younger in 2003, excluding those earning less than the annual equivalent of the legal minimum wage or more than the equivalent of US $100,000 in a given year. ***, **, and * indicates significance at 1 percent, 5 percent, and 10 percent levels, respectively.

Table 10. Mobility across ownership types and changes in car ownership and earnings

Dependent variable		(1) Probability of purchasing a car in *t*	(2) Log car value *t* - Log car value *t*-1	(3) Log earnings *t* - Log earnings *t*-1
Ownership change dummies; equal to 1 if:				
Private in both *t*-1 and *t*	Coefficient	0.031	-0.057*	0.129***
	St. Error	(0.025)	(0.031)	(0.029)
Gov./state in both *t*-1 and *t*	Coefficient	0.010	-0.089***	0.159***
	St. Error	(0.026)	(0.032)	(0.029)
Private in *t*-1, gov/state in *t*	Coefficient	0.021	-0.069	0.467***
	St. Error	(0.037)	(0.043)	(0.072)
Private in *t* -1, foreign in *t*	Coefficient	0.046	-0.090	0.885***
	St. Error	(0.061)	(0.108)	(0.169)
Gov./state in *t* -1, private in *t*	Coefficient	0.029	-0.065	-0.151**
	St. Error	(0.035)	(0.042)	(0.061)
Gov./state in *t* -1, foreign in *t*	Coefficient	0.022	-0.229***	1.092**
	St. Error	(0.160)	(0.038)	(0.487)
Foreign in *t* -1, private in *t*	Coefficient	0.040	-0.111	-0.463***
	St. Error	(0.055)	(0.079)	(0.142)
Foreign in *t* -1, gov./state in *t*	Coefficient	0.198	0.365	0.846
	St. Error	(0.253)	(0.395)	(0.651)

Notes: The omitted category is those employed by foreign-owned employers in both *t*-1 and *t*. Demographic and time effects controls are included but coefficients not reported. The first column shows marginal effects at the mean from a probit regression. Robust clustered standard errors in parentheses. The sample in models (1)-(3) is restricted to car owners with 5 or less cars registered in their names, ages 18 and older in 1999 and 60 and younger in 2003, excluding those who earned less than the annual equivalent of the legal minimum wage or more than the equivalent of US $100,000 in a given year and those whose primary earnings in any year 1999-2003 came from lottery winnings, veteran charitable foundation or interest and dividend incomes from the state savings bank or two largest state-owned corporations. In addition, the sample in model (2) excludes those who did not own cars in the previous year. The total number of observations is 15,871 in models (1) and (3) and 13,768 in model (2). ***, **, and * indicates significance at 1 percent, 5 percent, and 10 percent levels respectively.

Table 11. Mobility across employers' size and industries and changes in car ownership and earnings in the private sector

pendent variable	(1) Probability of purchasing a car in t	(2) Log car value t - Log car value t-1	(3) Log earnings t - Log earnings t-1
changers; size and industry change dummies: equal to 1 if			
naller employer in t- larger in t	-0.005	0.023	0.305***
	(0.020)	(0.028)	(0.052)
rger employer in t -1, naller in t	-0.015	-0.010	-0.306***
	(0.019)	(0.024)	(0.046)
ilities: not in t -1, yes t	0.018	0.026	0.630***
	(0.056)	(0.074)	(0.151)
holesale / retail trade: t in t -1, yes in t	-0.024	-0.016	-0.199***
	(0.028)	(0.031)	(0.067)
stayers; size change dummies: equal to 1 if			
nployer size smaller t than in t-1	-0.013	-0.021	-0.109***
	(0.024)	(0.031)	(0.038)
nployer size larger in nan in t -1	-0.011	0.004	0.131***
	(0.019)	(0.027)	(0.028)

Notes: Year dummies and demographic controls are included but not reported. The first column shows marginal effects at the mean from a probit regression. Robust clustered standard errors in parentheses. The sample is limited to car owners employed in the private Russian sector in both year t-1 and year t. It is restricted to car owners with 5 or less cars registered in their names, ages 18 and older in 1999 and 60 and younger in 2003, excluding those who earned less than the annual equivalent of the legal minimum wage or more than the equivalent of US $100,000 in a given year and those whose primary earnings in any year 1999-2003 came from lottery winnings, veteran charitable foundation or interest and dividend incomes from the state savings bank or two largest state-owned corporations. In addition, the sample in model (2) excludes those who did not own cars in the previous year. The total number of observations for job changers is 10,346 in models (1) and (3) and 9,011 in model (2). The total number of observations for job stayers is 7,993 in models (1) and (3) and 6,951 in model (2). ***, **, and * indicates significance at 1 percent, 5 percent, and 10 percent levels respectively.

Appendix (for referees' attention only)

1. Imputation of car values

We develop a procedure to assign prices to the vehicles owned by individuals in our sample. For each car, our data contain the car's make, model and the year it was produced. For example: "Make: Hyundai; Model: Avant; Production year: 1999," or "Make: Jaguar; Model: XJ6VP; Production year: 1993." There are 625 unique make-model combinations in our data. No information on the presence of optional features or the vehicle's condition is available. Hence, we could only assign prices to vehicles based on the median market value of cars of the same make, model and year of production. Since our identifying assumption is that the demand for cars is independent of the sector of employment, the imputation of prices does not pose a problem for our analysis. The details of the procedure used to impute prices are described below

Determining used/new status

New cars sell at a substantial premium over used cars, so accurately assigning a price to a vehicle requires determining whether it was purchased new or used. To do so, we used Vehicle Identification Numbers (VINs) to search the Vehicle Registration Database and determine the car's date of first registry. We dropped vehicles lacking a valid VIN, but this affected only a relatively small number of older, low-value vehicles.

We designated a car "new" if it was first registered in the year it was produced and in the name of the current owner. We considered a car "used" if the database showed prior registrations by different owners. We also considered a car "used" if either: (i) it was produced two or more years prior to the date of the first recorded purchase, or (ii) the first recorded purchase occurred after June 30th of the year following the production year. This (somewhat arbitrary) rule applied to less than 5 percent of cars in our sample (these cars also all proved to be relatively dated and therefore heavily depreciated by the time of our analysis). The results are also not sensitive to dropping these cars (and their owners) from the sample.

Obtaining prices

Russia lacks an authoritative source of car price information analogous to the "Blue Book" in the United States. Instead, we relied upon prices listed on the two large auto-trading websites that were operating in Moscow during 2005 and 2006.

The first website---www.autonet.ru---contained online sales advertisements from various private owners and used-car dealers in Moscow and provided information on a large variety of makes, models and years of production. For the majority of cars in our sample we were able to find multiple matching offers (often more than 10), and we took the median asking price as the market value of the vehicle as of 2005. We also referred to the second website---www.automosk.ru (which is no longer operating)---to collect pricing data on the new vehicles in our sample. Whenever we could not find a price for a given combination of make, model, and production year, we used the most similar model available. For example, for 2003 Mercedes models 200 and 200E, we used the price of the 2003 Mercedes model 200D.

We use these data to estimate an exponential depreciation rate, as well as category-specific new-car premiums for seven classes of vehicles: 1) Luxury models, 2) German and Swedish cars, 3) Japanese cars, 4) American cars 5) other European (non-German or Swedish) cars, 6) Russian cars, and 7) Korean and Chinese cars (the full inventory of models and category assignments is available upon request).

To estimate category-specific new car premiums and the annual depreciation rate we employed the universe of about 1,043 car make/model/year prices we gathered from the on-line sites above. Formally, let X_{in}^{st} denote the price of a car of make-model i, of new/old status n, produced in year s, observed in year t. We assume that this price is given by

$$X_{in}^{st} = \overline{X}_i \exp\left\{-\delta(t-s) + \gamma_{k_i} D(n=1) + \varepsilon_{in}^{st}\right\},$$

where $\overline{X}_i$ is the price of a given make-model i in the most recent year for which we have car price data from the on-line sites above (for most cars it is 2005 but for some cars the most recent available price was for the 2004 model), γ_{k_i} is the new car premium for the category into which make-model i falls, δ denotes the depreciation rate, and ε_{in}^{st} is the error term. Taking logs, we obtain the regression model:

$$\ln X_{in}^{st} = \ln \overline{X}_i - \delta(t-s) + \gamma_{ki} D(n=1) + \varepsilon_{in}^{st},$$

which we estimate by ordinary least squares. The depreciation rate and category-specific new car premiums estimated by this method are presented in Table A1.1. We also experimented with category-specific depreciation rates, but the results were very similar.

Table A1.1. Estimated new car premium and depreciation coefficients for different categories of cars

Variable	Coefficient	Std. Err.	t-value	P>t
New car premium				
Luxury	0.353	0.019	18.69	0.000
Russian	0.097	0.022	4.36	0.000
German	0.182	0.031	5.87	0.000
Japanese	0.111	0.024	4.59	0.000
American	0.076	0.046	1.66	0.098
Korean/Chinese	0.026	0.037	0.69	0.489
European	0.180	0.039	4.64	0.000
Depreciation				
Each additional year of age	0.123	0.002	72.94	0.000
R-squared: 0.929, Number of observations: 1,043				

We then use the depreciation rate and new car premiums estimated above to compute the estimated price $\hat{X}_{in}^{st}$ of all make-model-year combinations in our data, taking into account also if the car was purchased new or used.

The estimated values of cars in the sample vary a lot. They range from less than \$100 in 2003 for Soviet-made cars of the 1970s-early 1980s to \$244,442 in 2003 for a new Porsche Cayenne. The mean estimated car value is \$6,352 with the median of \$3,445 and the standard deviation of \$10,113. Details of the estimated prices for each car in our sample are available upon request.

2. Details of the sample construction procedure

Data sources

We employed three separate sets of data.

Residency Registry Database, 2002: This database contains information on all officially registered residents of Moscow as of August 2002, including their full names, dates of birth and registered addresses. It also provides us with an opportunity to group individual residents by households, with a household being defined as a group of people registered at the same address (and often, though not always, having the same last name). The total number of raw entries in this database is 10,155,157.

Vehicle Registration Database, 2005: Contains the list of all recorded instances of vehicle registrations in Moscow as of April 2005, along with the corresponding list of owners. The former provides the detailed description of vehicle characteristics including model, make, year of manufacturing, license plate number, unique Vehicle Identifying Number (VIN), and the date of registration. The latter contains data on each owner's full name, date of birth, residential address and passport number. Each entry in the vehicle list represents an instance of registration. Repeated registrations of the same vehicle are recorded as separate entries. We therefore define continuous periods of ownership for each car as intervals between its consecutive registrations by distinct owners. And to find all relevant entries that correspond to a given car we use its VIN number. The total number of raw entries is 8,308,881 vehicle registrations and 8,141,122 owners

Administrative Databases of Income, 1999–2003: This is a collection of five separate databases filed by all registered employers (sources of income) in Moscow for their employees (recipients of income). Each database covers one year from 1999–2003. Individual records in all of the five files provide full names, dates of birth, personal tax IDs, passport numbers, residential addresses, annual gross and taxable incomes, employer's names and employer's IDs. Additionally, for years 1999, 2000 and 2003 we have information on monthly breakdown of incomes; for year 2003 on income types (such as labor income, income from selling stock, income from intellectual property, etc.) and social security numbers; and for year 2000 on primary/non-primary status of income. The total number of raw entries in each database is as follows: 8,711,103 (1999); 10,361,320 (2000); 10,019,144 (2001); 7,029,376 (2002); 9,355,493 (2003).

Data issues

All three datasets above appear to have originated from manually digitized paper-based records and that leads to the following common problems:

Errors and missing data: A substantial number of entries contain artifacts of manual input: violations of the format, misspellings, typos, idiosyncratic abbreviations, missing data in certain fields, etc. This poses a challenge for matching entries across databases, as it reduces the amount and reliability of identifying information. As a result, we were not able to positively identify all legitimate matches, however, due to the random nature of imperfections in the data, we do not expect these missing matches to cause any systematic bias in our estimates.

Duplicate entries: We found that approximately 10 percent of all entries in our datasets are in fact virtual duplicates of some other entries contained in the same files. Some of them are fully identical to (and are thus indistinguishable from) the originals; the rest have slight modifications compared to the originals, caused usually by typos or partially missing data. In the former case we automatically merged the identical entries together. In the latter case, since the imperfect duplicates are sometimes linked to distinct records in other files, we manually tracked those connections and merged variables (incomes or cars) associated with both original and duplicate entries in the initial database.

Sample

Due to the issues outlined above, we performed the matching of the data across different databases almost entirely by hand. The significant amount of manual work involved in this process forced us to restrict our attention to a smaller-sized representative subsample of all the available data.

Selection procedure

In a nutshell, we took a randomly selected group of households from the pre-cleaned residency registry database; then, using combinations of full name, date of birth and address matched these people to their income and vehicle ownership records. We then used additional identifying information (personal tax ID, passport number, etc.) available in those matched records to find other matches that were not identified before. Finally, in our working sample we kept only the households that had at least one member of pre-retirement age, having both at least one car and recorded incomes in all five years.

In more detail, the selection procedure was as follows:

Step 1: We started by eliminating poor quality data from the residency registry database to increase the efficiency of subsequent matching. Specifically, we left out all households whose members either had inconsistencies in their full names (abbreviations, obvious typos or non-alphabetic characters), or lacked information on the date of birth, or address. This left us with 1,474,610 out of an initial 3,327,648 households, containing 4,757,680 out of an initial 10,155,157 individual members.

Step 2: From this set, we then drew a random sample of 30,000 households with 97,141 individual members.

Step 3: We matched this sample with vehicle registration database and all five administrative databases of incomes using either a combination of full name and date of birth, or a combination of full name and address. Apart from the data on income and cars, these databases provide complementary personal identifying information that we further employed to leverage the matching procedure.

Step 4: Using the passport numbers, personal tax IDs, social security numbers, combinations of full name and employer's ID available in the already established matches, we reiterated the matching procedure. This allowed us to find income and vehicle ownership records that had not yet been discovered due to missing (or incorrect) data in the fields previously used as identifiers.

Step 5: We repeated step 4 until no new matches were found.

Step 6: After that we further narrowed down the sample to contain only those households with at least one household member represented in the vehicle registration database and in all of the five income tax returns databases. This step left 6,362 out of the initial 30,000 households.

Step 7: Finally, we eliminated all duplicate income and vehicle records, and also dropped all households comprised solely of retirees. In the end, our sample contained 6,101 households with 21,617 members.

Representativeness of the sample

The design of the selection procedure outlined above has the following implications for representativeness of the final sample.

Step 1 should not have affected the statistical properties of the sample relative to the general population, as the eliminated poor quality data resulted from typing errors and other random omissions. Step 2 constituted a random draw and thus added no systematic bias. Steps 3–5 did not modify the composition of the sample itself, as it only involved matching of the sample entries to other databases. The most restrictive step in the procedure was step 6, which excluded 23,638 households, 13,446 of which had no car owners and the rest had no recorded income earners in at least one year from 1999–2003. Excluding non-car-owners was inevitable because car ownership is the only source of identification of true (as opposed to reported) earnings in our data (see the discussion in the main text). As for households that had no recorded income earners in at least one year from 1999–2003, recall that there are various reasons why this may happen, some of which have nothing to do with the representativeness of our sample. For example, we suspect that the biggest reason may be that official residency registration in Moscow is hard to obtain, so individuals and whole households that had actually moved out of the capital city (temporarily or sometimes even permanently) tend not to annul the registration and thus remain in the residency (but obviously not in the income) database while they actually live elsewhere. Also, recall that the residency database is a one-time snapshot as of 2002, so it will include households that did not reside in Moscow in some year either before or after 2002. In both these cases our procedure simply excludes households that did not reside in Moscow throughout the period of our analysis.

There is also a non-negligible fraction of individuals employed entirely in the unofficial economy in any given year, and our procedure excludes these from the sample. While this (together with step 7 where we excluded retired households) makes the sample not fully representative of the general population (and even of all car owners), it is a deliberate choice dictated by our desire to measure the degree of income underreporting in official employment. Thus, the results of the paper should be properly interpreted as applicable to a representative household with at least one car owner and at least one consistent income earner in the official (pre-retirement) employment. In most estimations, we restrict this sample further to income earners between age 17 in 1999 and 60 in 2003, earning more than the minimum wage in any given year and not deriving their main income solely from interest, dividends or windfall sources (see the main text).

Structure of the sample

For each individual in our sample we were able to directly obtain the following information from the databases: full name; age (from the information about the date of birth), personal tax ID; gross and taxable incomes for years 1999-2003; monthly breakdown of incomes for years 1999, 2000, 2003; income type for year 2003; primary income indicator for year 2000; number of owned cars; name of primary employer's ID; number of members in the household. We define the primary employer to be the source of the individual's highest recorded income.

Additionally, we supplemented our sample with the following data:

Gender: Imputed from gender-specific endings of middle names, which are characteristic of the Russian language.

Ethnicity: Inferred manually from full names. We conservatively assigned specific ethnicity only to people with distinctively ethnic full names. We added all undetermined cases to the majority Slavic group. Table A2.1 presents the summary of the resulting assignment.

Table A2.1. Ethnic composition of the sample

	Ethnicity	Individuals	Share
1	Slavic and uncertain	20266	55.745%
	of which positively Russian	14738	40.539%
2	Jewish	377	1.037%
3	Georgian	68	0.187%
4	Armenian	179	0.492%
5	"Muslim" (Tatar, Bashkir, Chechen, etc.)	493	1.356%
6	Other (Korean, Spanish, German, etc.)	234	0.644%

Sector classification and type of ownership: We classified 13,774 distinct primary employers in our sample into 19 sectors and also assigned each of them to one of the four types of ownership (see Appendix A6 for more details). Namely, we checked the presence of sector-specific keywords (such as bank, insurance, factory, police, etc.) in employers' names to do the initial automatic sector assignment and then manually assigned sectors to the employers that were left out by the script. Similarly, we used another list of keywords to infer the type of ownership (e. g. JSC, Ltd, State, etc.). We then manually resolved all the remaining undetermined cases. Sector classification of each individual's primary employer is represented in the working sample by a 19-component binary vector. Additionally, for every household we calculated an average of the respective "sector"-vectors of its members' primary employers, weighted by the incomes received from those employers.

Size of primary employer: Total number of individuals who received payments from a given employer (income source) in a given year.

Total wage bill of primary employer: Sum of all payments made by a given employer (income source) to individuals (income recipients) in a given year.

Income percentile: Percentile of individual's primary income in the overall wage distribution of his/her primary employer in a given year.

Even though all the data used by us came from the public domain, to ensure privacy we have purged all the individual-identifying information (names, addresses, id numbers) after we finished the construction of the sample. All the data used in the paper (without individual-identifying information) and our estimation codes will be available for the purposes of replicating our results. We can also provide the scripts used to clean the data and to conduct the selection/matching process described in steps 1-7, which can be employed to replicate our sample construction procedure using the original databases.

3. Comparisons between our sample and the general population in Moscow and Russia

This section compares the distributions of earnings and employers in our sample of car owners to the data on the whole Moscow workforce and the data from past studies.

Table A3.1. Officially reported earnings in our sample and in RLMS data (in Russian rubles).

Years	RLMS	Wage earners in our data	
	Mean earnings	Mean earnings	40th percentile
1998	13,104	N/A	N/A
1999	N/A	58,910	16,802
2000	26,088	86,578	26,296
2001	39,720	130,881	38,645
2002	51,984	169,153	51,235
2003	N/A	225,338	61,575

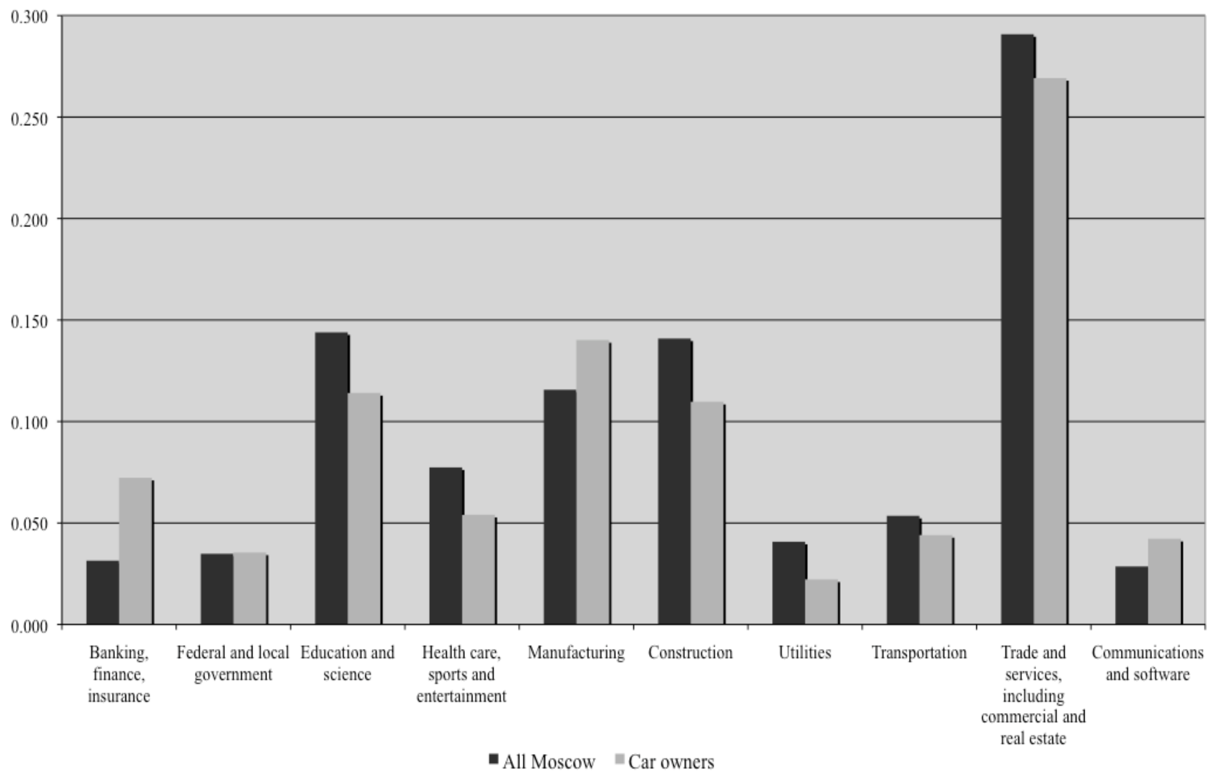
RLMS data from Table 4 in Ivanova et al. [2005] (multiplied by 12 to adjust to annual data). Our data on wage earners among car owners (as defined in the main texts), excluding employees of foreign-owned companies.

In Table A3.1 we compare reported earnings in our sample to earnings estimated by Ivanova et al. [2005] using the RLMS (Russian Longitudinal Monitoring Survey) questionnaire. For overlapping years, the dynamics of wages in the 40th percentile of our distribution almost exactly parallel the dynamics of mean earnings estimated from RLMS (for the whole country). The growth rates exhibited by mean earnings from 2000-2001 and from 2001-2002 are also remarkably similar (52 percent in the RLMS data versus 51 percent in our data for 2000-2001, and 31 percent and 29 percent, respectively, for 2001-2002).

We next compare our sample of car owners with official statistical data about all the workforce in Moscow to see if there are significant differences in the fractions employed in different sector and in relative earnings by employment type as predicted in Section 2 of the main text.

Figure A3.1 compares fractions of car owners in our sample by different employment sectors with the distribution of the whole Moscow workforce from official statistical sources. For example, we find that ratios of car values to incomes are among the highest in the education and science sector, but the fraction of those with main source of income from this sector in our sample of car owners in 2003 (11.4 percent) is only slightly lower than in the official Moscow statistics for the same year (14.4 percent). Overall, individuals working in banking, finance and insurance as well as those in communications and computer software and in manufacturing are somewhat overrepresented in our sample as compared to all of Moscow, while members of the education and healthcare professions, as well as workers in construction, utilities and transportation sectors are somewhat underrepresented, but these differences are not systematically related to differences in estimated hidden earnings. Overall, our sample is fairly representative of various sectors of employment in Moscow as we do not see any sector that has only a miniscule fraction of its workers represented in our sample when compared to the whole workforce, as we would expect if our sample was totally unrepresentative of some occupations.

Figure A.3.1
Fraction of employment by sector in 2003: All Moscow versus sample of car owners



Source: All Moscow from Russian State Statistical Committee Report on Labor and Employment (2005), available electronically at http://www.gks.ru/wps/PA_1_0_S5/Documents/jsp/Detail_default.jsp?category=1112178611292&elementId=1139916801766; our estimates.

Figure A3.2 compares the mean earnings of car owners in our sample by different employment sectors with mean earnings of the whole Moscow workforce from official statistical sources and, once again, does not show systematic differences that we would expect if workers in sectors estimated as hiding particularly large fractions of the earnings were actually picked up because of unobserved random income shocks. For example, the average reported earnings of car owners in all of our sample in 2003 were 117 percent higher than average statistical earnings of all workers in Moscow (224 thousand rubles versus 103 thousand rubles). Average reported earnings of car owners in the trade and service sector (where we estimate a particularly high fraction of hidden earnings based on cars being more expensive, relative to reported earnings, than in other sectors), however, were 215 percent higher than average statistical earnings of workers employed in trade and services in the whole of Moscow (203 thousand rubles versus 65 thousand rubles). Hence, official earnings of an average worker in this sector of the whole population are actually much lower relative to the average member of the Moscow workforce than the corresponding number for an average car owner employed in this sector in our sample. This is exactly the opposite of what we would expect if car owners in trade and services were able to afford a car for most part because of some unobserved income shocks unrelated to their occupation.

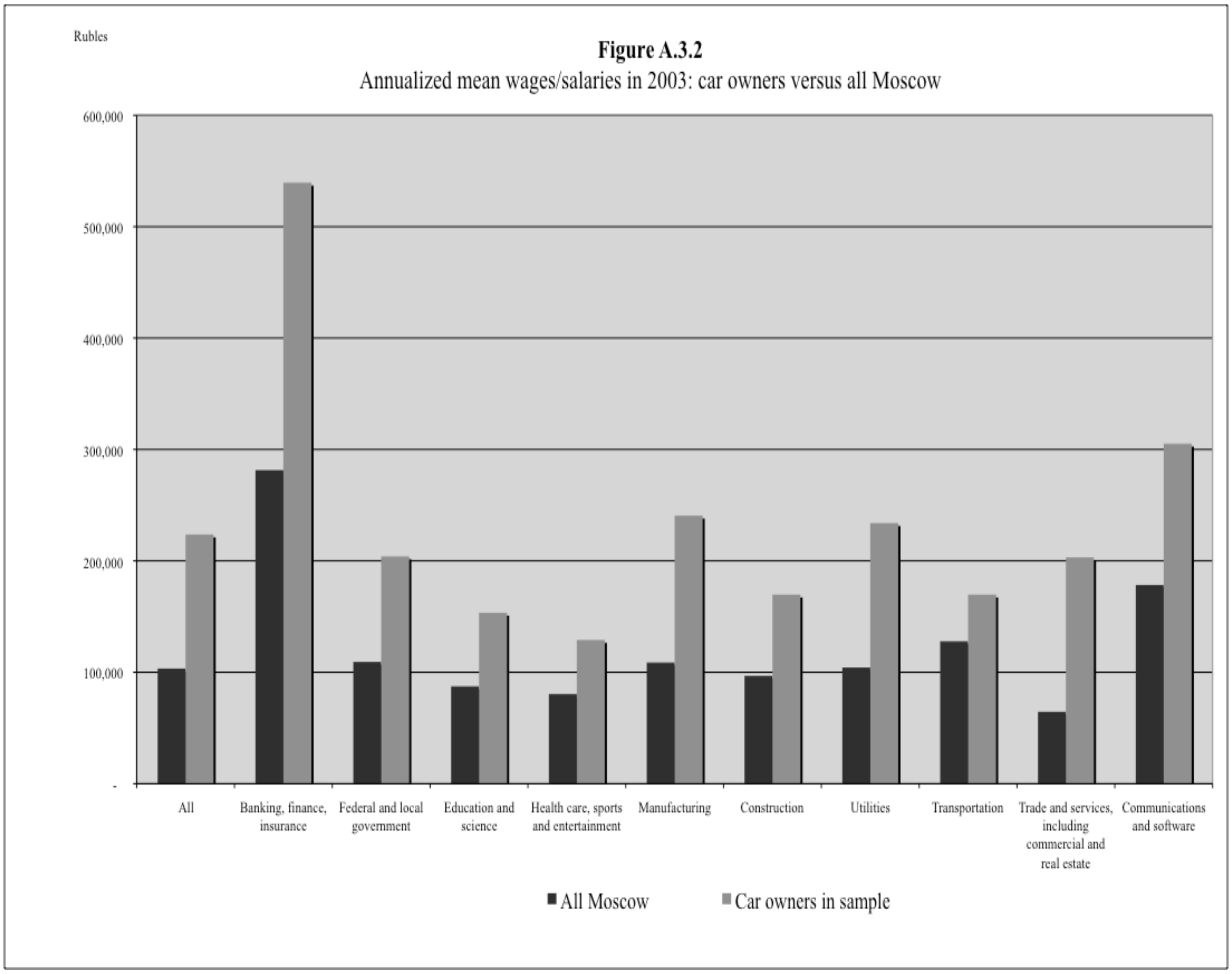


Figure A.3.2
Annualized mean wages/salaries in 2003: car owners versus all Moscow

Source: All Moscow from Russian State Statistical Committee Report on Labor and Employment (2005), available electronically at http://www.gks.ru/wps/PA_1_0_S5/Documents/jsp/Detail_default.jsp?category=1112178611292&elementId=1139916801766; our estimates.

4. Comparisons with the NLSY data

We can compare our data on Moscow car owners with the U.S. data. Needless to say, there is a lot of differences between Moscow and U.S. car owners even apart from the role played by hidden earnings. Nevertheless, the U.S. data may give us one possible benchmark, especially when looking for relatively more "reasonable" patterns in some parts of the disaggregated Moscow data. We utilize the data on respondents who owned at least one but not more than five cars for personal use and earned non-zero incomes in the form of wages and salaries in the 2002 survey of the NLSY 1979 cohort. We first compare some summary statistics of the NLSY sample with the subsample of Moscow car owners whose primary source of income was wage and salary earnings in 2003, who owned no more than 5 cars, and who belonged to the same age bracket in 2002 as the NLSY 1979 cohort (between 35 and 45 years of age). To increase the number of observations, especially for car owners in foreign-owned firms in Moscow, we pool together the data for 1999-2003.

Table A4.1 presents the results. Comparing first the NLSY data to car owners employed by foreign-owned firms in Moscow, car values to incomes ratios are rather close, and the coefficients of variation of both earnings and car values are also very similar. The NLSY car owners are on average 2.25 times richer than Moscow car owners measured by earnings and 1.91 times richer measured by car values.

Table A4.1. Car values and earnings summary statistics, NLSY and Moscow data.

		# of obs.	Mean	COV
Russian-owned private firms	Earnings (E)	4,940	4,944	3.47
	Car values (C)	4,940	8,752	1.41
	Ratio (C)/(E)		1.77	
Government and state-owned sector	Earnings (E)	1,635	3,757	1.24
	Car values (C)	1,635	5,837	1.44
	Ratio (C)/(E)		1.55	
Foreign-owned firms	Earnings (E)	189	18,248	1.19
	Car values (C)	189	9,517	0.93
	Ratio (C)/(E)		0.52	
NLSY 1979 cohort (2002)	Earnings (E)	4,997	41,055	0.92
	Car values (C)	4,997	18,163	0.93
	Ratio (C)/(E)		0.44	

Note: Moscow sample consists of car owners with the primary source of income wages and salaries in 2003, aged 35-45 in 2002, owning no more than 5 cars per person. Earnings and car values for Moscow data are pooled for the five years, 1999-2003. Earnings for the NLSY 1979 cohort are wages and salaries for 2002.

Moscow car owners employed by private Russian firms and by the government/state-owned sector present a strikingly different picture. The NLSY car owners are still about two times richer than car owners employed by the private Russian sector in terms of the market values of their respective cars. But the ratio of average reported earnings is four times higher at 8.3. Compared to Moscow car owners employed in the government and state-owned sector, the gap in car values with the NLSY wage earners is 3.1, but the gap in reported earnings is 10.9. As a result, the car values to earnings ratio is 4 times higher in the Russian private sector and 3.5 times higher in the government and state-owned sector than it is in the NLSY data. Extrapolating from the car values to earnings ratio in the NLSY data suggests that 75 percent of earnings may be hidden in the Russian private sector, and about 72 percent in the government and state-owned sector. Crude as they are, these estimates are similar in magnitude to those obtained in the main text.

We also employed the NLSY data to estimate the income elasticity of the demand for the stock of cars and compared it with the corresponding elasticity in the Moscow data. Specifically, we estimated the following regression

$$\log C = a + b\log E + d'Z + \upsilon,$$

where C is the value of cars owned by an individual, E is his or her earnings, while Z is a vector of controls that can affect the relationship between value of cars and earnings, such as age, gender, ethnicity, number of members of the household, etc. Table A4.2 presents the results.

Table A4.2. Coefficient on Log labor earnings in Log car value regressions

	NLSY		Private Russian firms		Government/ State sector	
Coefficient	0.339	***	0.087	***	0.061	*
St. error	(0.017)		(0.019)		(0.037)	
Constant term	5.435	***	8.775	***	9.269	***
St. error	(0.324)		(0.452)		(0.750)	
Adj. R-squared	0.111		0.035		0.014	
# of observations	4,996		4,940		1,635	

The dependent variable is Log car value. Moscow sample consists of car owners with the primary source of income wages and salaries in 2003, aged 35-45 in 2002, owning no more than 5 cars per person. Earnings and car values for Moscow data are pooled for the five years, 1999-2003. Earnings for the NLSY 1979 cohort are wages and salaries for 2002. Other controls include age, gender and ethnic dummies, the total number of household members, and year dummies for Moscow. Robust clustered standard errors are reported. *** indicates that the coefficient is significant at the 1 percent level ** at 5 percent level.

In the NLSY sample, the coefficient on log labor earnings is large and highly significant implying that a 100 per cent increase in labor earnings translates into 34 per cent of increase in the market value of cars. But Moscow data present a very different picture. Among car owners employed in private Russian firm, doubling of reported earnings is associated with just 8.7 percent increase in the market value of the stock of cars, while among car owners employed in government and state-owned sector the corresponding coefficient is even smaller and statistically barely significant.

5. Estimating income elasticity of the demand for the stock of cars.

In order to estimate the appropriate value for the income elasticity of the demand for the stock of cars using Moscow data, we need a subsample of car owners whose earnings are likely to be reported truthfully. While car owners employed in foreign-owned firms seem to be the natural candidates, our main sample did not contain enough observations, especially year-by-year, on these car owners to consistently estimate this parameter.

To create a sample that could produce a consistent and reliable estimate, we decided to go back to the whole database and to oversample car owners employed in foreign-owned firms, using the following procedure.

Step 1: We began by selecting all entries from the administrative databases of income whose source of income was previously classified by us as foreign-owned in the main sample.

Step 2: We then used the identifying data contained in these entries (full names, dates of birth, passport numbers, social security numbers and personal tax IDs) to match the records of distinct individuals across different years and to identify all of their other sources of income. This allowed us to create a panel of individuals who were: 1) present in all five databases of income, and 2) employed by a foreign-owned firm in at least one of the five years.

Step 3: Once matched, these cumulative records often contained more complete identifying information that could be used to find some of the previously missed income entries. And we repeated the matching procedure iteratively until no new matches were found.

Step 4: Finally, we matched the most complete records assembled at the previous step to the vehicle registration database and eliminated all non-car-owners from the sample. To improve the quality of the data used as a benchmark of transparent reporting, we also left in the sample only for-profit foreign-owned employers, for which we had 20 or more observations on car owners per year.[23]

We then estimated the following regression equation, similar to (2) in the main text:

$$\ln C_{it} = \lambda \ln E_{it}^{R} + \gamma_2' X_{it} + \phi_2 t + \varepsilon_{it} \qquad (2')$$

We first included only individual characteristics such as age, age squared, and gender dummy as well as year dummies as controls when estimating (2'). Table A5.1 presents the results. In late 1998 Russia was hit by the default of its banking system, which largely eliminated savings of the middle class and may have forced many to delay a planned car purchase. We conjecture that this factor probably explains the anomalously low value of the parameter λ estimated for 1999. The estimates for other years indicate that the income elasticity of the demand for the stock of cars was in the range of 0.3-0.4. Note that the estimate obtained from the NLSY data (equal to 0.34 in Table A4.2 above) falls right in the middle of this range. Thus, the two benchmarks produce very similar estimates of the income elasticity of the demand for the stock of cars.

Table A5.1. Estimating demand elasticity for the stock of cars in the subsample of employees of foreign-owned firms

	All Years	1999	2000	2001	2002	2003
Log Reported Income	0.319***	0.166***	0.295***	0.306***	0.390***	0.389***
	(0.053)	(0.059)	(0.073)	(0.066)	(0.071)	(0.071)
Age in 2003	-0.075*	-0.014	-0.068	-0.054	-0.060	-0.125***
	(0.040)	(0.064)	(0.053)	(0.048)	(0.045)	(0.047)
Age in 2003 squared	0.001*	0.000	0.001	0.001	0.001	0.001**
	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Male	-0.157	-0.212	-0.250*	-0.203	-0.085	-0.088
	(0.108)	(0.146)	(0.143)	(0.130)	(0.123)	(0.118)
Constant	6.934***	7.103***	6.828***	6.753***	5.963***	7.327***
	(0.882)	(1.477)	(1.278)	(1.053)	(1.022)	(1.093)
Observations	1,923	245	309	413	451	505
Adj. R-squared	0.098	0.038	0.075	0.078	0.105	0.111

Notes: The dependent variable is log of car value. The sample consists of car owners aged between 18 in 1999 and 60 in 2003, excluding those with primary earnings less than the legal minimum wage or in excess of $100,000, with the main source of earnings coming from (for-profit) foreign-owned firms with 20 or more observations on car owners in each year. The regression for all years pooled together includes year dummies. *** indicates that the coefficient

[23] It is easy to see that this step eliminates smaller-sized firms because a firm has to have at least 20 employees to have 20 or more car owners in the sample. Preliminary estimates using our main sample and the oversampled part indicated that serious income hiding may still be present in smaller foreign-owned firms, so that excluding those is necessary to construct a benchmark sample. We experimented with other ways of paring down the sample of foreign-owned firms for this purpose, such as excluding firms below a certain size in the whole database. Overall results using those alternative benchmarks were similar to those presented here, but somewhat less stable in year-by-year estimates (details are available upon request).

is significant at the 1 percent level, ** at 5 percent level, and * at 10 percent level.

We then re-estimated equation (2') including industry dummies and (log of) firm size. If, as our empirical analysis in the main text assumes, there is no difference in the functional form of the demand for the stock of cars depending on employer size or industry, we would expect the estimated value of the parameter λ to remain basically the same in this benchmark sample where the reporting of earnings is more or less transparent, while coefficients on employer size and industry dummies should be economically and statistically insignificant.

Table A5.2. Estimating demand elasticity for the stock of cars in the subsample of employees of foreign-owned firms (including employer characteristics)

	All Years	1999	2000	2001	2002	2003
Log of reported income	0.330***	0.224***	0.394***	0.337***	0.367***	0.372***
	(0.060)	(0.068)	(0.096)	(0.083)	(0.086)	(0.081)
Log number of employees	-0.007	-0.021	0.092	0.036	-0.042	-0.076
	(0.041)	(0.058)	(0.060)	(0.053)	(0.052)	(0.050)
Wholesale and retail trade	-0.109	0.010	0.029	-0.205	-0.096	-0.109
	(0.134)	(0.198)	(0.198)	(0.164)	(0.170)	(0.160)
Manufacturing	-0.237**	-0.341**	-0.027	-0.256**	-0.201	-0.246*
	(0.110)	(0.147)	(0.151)	(0.129)	(0.131)	(0.136)
Services	0.030	0.264	0.482**	-0.025	-0.010	-0.198
	(0.151)	(0.211)	(0.207)	(0.181)	(0.185)	(0.185)
Communications/software	0.062	0.227	0.137	-0.040	0.090	-0.001
	(0.142)	(0.186)	(0.179)	(0.163)	(0.166)	(0.176)
Age in 2003	-0.079*	-0.030	-0.074	-0.055	-0.068	-0.136***
	(0.041)	(0.063)	(0.054)	(0.049)	(0.046)	(0.048)
Age in 2003 squared	0.001*	0.000	0.001	0.001	0.001	0.001**
	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Male	-0.162	-0.230	-0.263*	-0.197	-0.078	-0.077
	(0.107)	(0.144)	(0.146)	(0.130)	(0.123)	(0.118)
Constant	7.080***	7.115***	5.397***	6.432***	6.699***	8.386***
	(1.036)	(1.670)	(1.688)	(1.261)	(1.312)	(1.341)
Observations	1,923	245	309	413	451	505
Adj. R-squared	0.106	0.065	0.089	0.079	0.105	0.116

Notes: The dependent variable is log of car value. The sample consists of car owners aged between 18 in 1999 and 60 in 2003, excluding those with primary earnings less than the legal minimum wage or in excess of $100,000, with the main source of earnings coming from (for-profit) foreign-owned firms with 20 or more observations on car owners in each year. The omitted industry is banking and finance. The regression for all years pooled together includes year dummies. *** indicates that the coefficient is significant at the 1 percent level, ** at 5 percent level, and * at 10 percent level.

Table A5.2 presents the results. The tests of cross-equation constraints fail to reject the hypothesis that coefficients on Log of reported income are equal between Tables A5.1 and A5.2 in the whole sample at 5 percent significance level, and in year-by-year estimations at 5-10 percent significance level in all years with the exception of 2000. The omitted industry is banking and finance, so that the coefficients on industry dummies should be interpreted relative to it. The coefficients on the (log of the) employer size and industry dummies are jointly not statistically significant at conventional levels in the regression for all years pooled together and in year-by-year regressions for 2001-2003. Overall, the results suggest that the demand for the stock of cars estimated for the benchmark sample indeed does not depend on firm size or industry. The estimated values of λ remain between 0.3 and 0.4 for all years except 1999. The estimate for all years pooled together is almost identical to the estimate obtained previously using the NLSY data. We thus use the value of parameter $\lambda = 0.35$ in the main text regressions.

6. List of sectors of economic activity

We classified the employers in our sample by the type of ownership and the sector of economic activity. Overall, we have 10,456 private companies (including joint ventures with federal and local government); 2,836 government/public sector employers; and 367 foreign-owned companies. The breakdown by sectors of economic activity is as follows.

Table A6. List of sectors of economic activity

Sector	Employers	Share
Banking, finance, insurance	689	4.74%
Federal government	168	1.16%
City and local government	262	1.80%
Law enforcement	173	1.19%
Higher education and research	966	6.64%
Secondary education	344	2.37%
Health care and medical services	540	3.71%
Mass media	345	2.37%
Construction	1,610	11.07%
Utilities	41	0.28%
Transportation	467	3.21%
Wholesale and retail trade	2,517	17.31%
Manufacturing	2,117	14.56%
Sports and entertainment	292	2.01%
Services	1,780	12.24%
Communications, IT	410	2.82%
Private security	391	2.69%
Self employed	111	0.76%
Non-education not-for-profits	449	3.09%
Unknown (unable to classify)	868	5.97%

Note: Employers assigned to more than one sector (such as those operating in both trade and services) are counted in each of those sectors.

7. Year-by-year estimates of hidden earnings and corruption relative to the benchmark

The following tables present year-by-year estimates of the regression used to estimate hidden earnings and corruption relative to the benchmark of foreign-owned firms for all years pooled together in Table 7 in the main text.

Table A7.1. Hidden earnings and corruption relative to the benchmark (1999-2000)

	1999		2000	
Ownership	Private	State	Private	State
Banking, finance and insurance	-2.724***	-1.573***	-2.835***	-0.743*
	(0.309)	(0.429)	(0.276)	(0.437)
Federal Government	NA	-1.953***	NA	-1.534***
	NA	(0.500)	NA	(0.469)
City and local government	NA	-1.584***	NA	-1.814***
	NA	(0.451)	NA	(0.470)
Law enforcement	NA	-2.276***	NA	-2.310***
	NA	(0.370)	NA	(0.378)
Higher education, research	-3.189***	-1.961***	-3.425***	-2.080***
	(0.543)	(0.336)	(0.549)	(0.296)
Secondary education	0.362	-3.043***	0.370	-2.944***
	(1.639)	(0.605)	(1.983)	(0.490)
Health care	-2.080**	-2.315***	-2.817***	-2.377***
	(0.923)	(0.416)	(0.784)	(0.386)
Mass media	-3.253***	-1.669**	-2.926***	-1.257*
	(0.612)	(0.676)	(0.509)	(0.712)
Construction	-2.100***	-2.074***	-2.183***	-2.037***
	(0.275)	(0.423)	(0.258)	(0.387)
Utilities	-1.245***	0.610	-1.918***	0.078
	(0.466)	(0.781)	(0.422)	(0.818)
Transportation	-1.890***	-0.822*	-1.943***	-1.160***
	(0.456)	(0.464)	(0.461)	(0.369)
Wholesale and retail trade	-3.055***	-3.468***	-3.093***	-2.611***
	(0.268)	(0.726)	(0.249)	(0.747)
Manufacturing	-2.805***	-0.686	-2.738***	-0.829
	(0.262)	(0.624)	(0.239)	(0.616)
Sports, entertainment	-3.458***	-2.534***	-3.139***	-2.346***
	(0.690)	(0.760)	(0.592)	(0.876)
Services	-2.592***	-2.348***	-2.632***	-2.397***
	(0.276)	(0.747)	(0.261)	(0.686)
Communications and IT	-2.893***	-2.036***	-2.820***	-2.880***
	(0.417)	(0.516)	(0.380)	(0.443)
Private security	-2.639***	NA	-3.000***	NA
	(0.436)	NA	(0.451)	NA
Self employed	-4.434***	NA	-2.651***	NA
	(1.506)	NA	(0.814)	NA
Non-education not-for-profit	-2.401***	-3.031**	-2.416***	-4.169***
	(0.622)	(1.382)	(0.451)	(1.228)
# of observations, adj. R^2	3,166	0.084	3,702	0.091

Notes: The dependent variable is the difference between log of reported earnings and income elasticity adjusted log of car values: $\log E^R - 1/\lambda \log C$. Income elasticity is $\lambda = 0.35$. Robust clustered standard errors in parentheses. The omitted category is the benchmark sample of foreign-owned employers in Appendix 5. Regressions also include age, gender, percentile in EED and its square term, and the constant term. The sample is restricted to car owners with 5 or less cars, 18 and older in 1999 and 60 and younger in 2003, excluding those earning less than the annual equivalent of the legal minimum wage or more than the equivalent of US $100,000 in a given year and those whose primary earnings in any year 1999-2003 came from lottery winnings, veteran charitable foundation or interest and dividend incomes from the state savings bank or two largest state-owned corporations.. ***, **, and * indicates significance at 1 percent, 5 percent, and 10 percent levels, respectively.

Table A7.2. Hidden earnings and corruption relative to the benchmark (2001)

Pooled OLS	2001	
Ownership	Private	State
Banking, finance and insurance	-1.554***	0.238
	(0.253)	(0.468)
Federal Government	NA	-1.393***
	NA	(0.401)
City and local government	NA	-1.105***
	NA	(0.427)
Law enforcement	NA	-1.584***
	NA	(0.321)
Higher education, research	-2.354***	-1.529***
	(0.518)	(0.277)
Secondary education	-0.037	-2.101***
	(1.601)	(0.429)
Health care	-2.315***	-1.723***
	(0.663)	(0.363)
Mass media	-2.472***	-0.982
	(0.409)	(0.732)
Construction	-1.759***	-1.226***
	(0.245)	(0.353)
Utilities	-0.631	0.047
	(0.387)	(0.712)
Transportation	-1.334***	-0.052
	(0.436)	(0.370)
Wholesale and retail trade	-2.656***	-1.977***
	(0.221)	(0.743)
Manufacturing	-1.893***	-0.273
	(0.222)	(0.545)
Sports, entertainment	-3.301***	-2.072**
	(0.741)	(0.830)
Services	-2.350***	-1.404**
	(0.234)	(0.670)
Communications and IT	-1.686***	-1.937***
	(0.301)	(0.460)
Private security	-2.146***	NA
	(0.418)	NA
Self employed	-3.702**	NA
	(1.587)	NA
Non-education not-for-profit	-2.397***	-2.525*
	(0.425)	(1.347)
# of observations, adj. R^2	4,398	0.080

Notes: The dependent variable is the difference between log of reported earnings and income elasticity adjusted log of car values: $log\, E^R - 1/\lambda\, log\, C$. Income elasticity is $\lambda = 0.35$. Robust clustered standard errors in parentheses. The omitted category is the benchmark sample of foreign-owned employers in Appendix 5. Regressions also include age, gender, percentile in EED and its square term, and the constant term. The sample is restricted to car owners with 5 or less cars, 18 and older in 1999 and 60 and younger in 2003, excluding those earning less than the annual equivalent of the legal minimum wage or more than the equivalent of US $100,000 in a given year and those whose primary earnings in any year 1999-2003 came from lottery winnings, veteran charitable foundation or interest and dividend incomes from the state savings bank or two largest state-owned corporations.. ***, **, and * indicates significance at 1 percent, 5 percent, and 10 percent levels, respectively.

Table A7.3. Hidden earnings and corruption relative to the benchmark (2002-2003)

Pooled OLS	2002		2003	
Ownership	Private	State	Private	State
Banking, finance and insurance	-1.848***	-0.031	-2.169***	0.433
	(0.249)	(0.448)	(0.257)	(0.435)
Federal Government	NA	-1.275***	NA	-0.811*
	NA	(0.403)	NA	(0.424)
City and local government	NA	-1.443***	NA	-1.511***
	NA	(0.503)	NA	(0.470)
Law enforcement	NA	-1.323***	NA	-0.986***
	NA	(0.358)	NA	(0.359)
Higher education, research	-2.415***	-1.606***	-2.491***	-1.800***
	(0.482)	(0.282)	(0.410)	(0.273)
Secondary education	-0.294	-1.873***	-1.658	-1.177***
	(1.415)	(0.391)	(1.462)	(0.441)
Health care	-2.420***	-1.083***	-1.652**	-1.342***
	(0.644)	(0.373)	(0.666)	(0.354)
Mass media	-2.122***	-0.626	-1.823***	-0.663
	(0.407)	(0.642)	(0.407)	(0.591)
Construction	-2.007***	-1.515***	-2.074***	-1.209***
	(0.228)	(0.372)	(0.224)	(0.352)
Utilities	-0.817**	-0.234	-0.825**	-0.493
	(0.387)	(0.581)	(0.346)	(0.583)
Transportation	-1.846***	0.011	-1.693***	0.194
	(0.419)	(0.335)	(0.364)	(0.333)
Wholesale and retail trade	-2.769***	-2.189***	-2.604***	-1.826***
	(0.210)	(0.661)	(0.211)	(0.698)
Manufacturing	-1.978***	-0.295	-1.982***	0.123
	(0.219)	(0.453)	(0.212)	(0.490)
Sports, entertainment	-1.862***	-2.055***	-2.269***	-2.054***
	(0.655)	(0.681)	(0.640)	(0.609)
Services	-2.384***	-0.999	-2.643***	-1.974***
	(0.229)	(0.664)	(0.215)	(0.749)
Communications and IT	-2.005***	-1.543***	-2.206***	-1.357***
	(0.304)	(0.403)	(0.293)	(0.461)
Private security	-2.319***	NA	-2.504***	NA
	(0.333)	NA	(0.321)	NA
Self employed	-5.455***	NA	-2.440***	NA
	(0.185)	NA	(0.695)	NA
Non-education not-for-profit	-1.965***	-4.125***	-2.036***	-5.357***
	(0.420)	(1.014)	(0.507)	(1.506)
# of observations, adj. R^2	4,831	0.082	5,300	0.090

Notes: The dependent variable is the difference between log of reported earnings and income elasticity adjusted log of car values: $\log E^R - 1/\lambda \log C$. Income elasticity is $\lambda = 0.35$. Robust clustered standard errors in parentheses. The omitted category is the benchmark sample of foreign-owned employers in Appendix 5. Regressions also include age, gender, percentile in EED and its square term and the constant term. The sample is restricted to car owners with 5 or less cars, 18 and older in 1999 and 60 and younger in 2003, excluding those earning less than the annual equivalent of the legal minimum wage or more than the equivalent of US $100,000 in a given year and those whose primary earnings in any year 1999-2003 came from lottery winnings, veteran charitable foundation or interest and dividend incomes from the state savings bank or two largest state-owned corporations.. ***, **, and * indicates significance at 1 percent, 5 percent, and 10 percent levels, respectively.

8. Direct evidence from two private banks

The difference between the fraction of hidden earnings estimated by us and in other studies is so striking that we sought for an opportunity to check it on some actual employers. Such an opportunity presented itself in the form of an acquaintance with insider knowledge of the actual employment contracts of a group of middle- and top-level managers in two medium-sized private Russian banks. This individual (hereinafter referred to as "the source") agreed, on the condition of anonymity, to match the 2003 actual earnings of those managers to the administrative data at our disposal.[24]

Table A8 presents average actual and recorded earnings in those two banks (separately and together) as well as the implied fraction of hidden earnings. Among 45 top and middle managers of the relatively larger bank A, reported earnings represent on average just about 11 percent of true earnings (the fraction hidden is 89 percent), while among 34 top and middle managers of the relatively smaller bank B reported earnings represent on average less than 9 percent of true earnings (the fraction hidden is 91 percent). The weighted average fraction of hidden earnings across all 79 individuals is almost 90 percent. These numbers are in the range estimated by us in our sample, using foreign-owned firms as a benchmark (see Table A7.3).

Table A8. Direct evidence from two private banks (2003 data)

	Number of observations	Averages of:		
		True earnings	Recorded earnings	Fraction hidden
Bank A	45	46,613	3,737	0.887
Bank B	34	25,032	471	0.914
Average (weighted)	79	37,325	2,332	0.899

Note: recorded earnings come from the 2003 administrative database. True earnings provided by the source.

[24] The source also described a popular form of paying out hidden earnings, which consisted of the bank purchasing foreign currency from an employee and selling it back, often on the same day but at a different exchange rate, thereby generating profit for the individual involved (and loss for the bank). The monthly amount of such "foreign currency transaction profits" was part of the verbal employment contract. According to the source, a loophole in the tax code technically allowed these payments to go unreported as earnings, without violating the letter (though certainly not the spirit) of the tax code.

9. Fixed-effects estimation results

Table A9. Hidden earnings and corruption relative to the benchmark

Ownership	Private	State	Ownership	Private	State
Banking, finance	-1.117***	-0.655	Manufacturing	-1.427***	-0.670
	(0.333)	(0.504)		(0.323)	(0.445)
Federal Government	NA	-0.957**	Sports, entertainment	-1.626***	-2.163***
	NA	(0.408)		(0.461)	(0.571)
City, local government	NA	-0.672	Services	-1.526***	-0.719
	NA	(0.446)		(0.321)	(0.584)
Law enforcement	NA	-0.909**	Communications and IT	-1.175***	-1.393**
	NA	(0.438)		(0.370)	(0.706)
Higher education	-1.351***	-1.342***	Private security	-1.697***	NA
	(0.374)	(0.344)		(0.393)	NA
Secondary education	-1.784***	-1.521***	Self employed	-1.128*	NA
	(0.643)	(0.471)		(0.597)	NA
Health care	-1.039**	-0.744*	Non-education not-for-profit	-1.726***	-2.442***
	(0.523)	(0.401)		(0.388)	(0.675)
Mass media	-1.209***	-0.613	Year dummies		
	(0.404)	(0.629)	2000	0.269***	
Construction	-1.384***	-1.164***		(0.031)	
	(0.334)	(0.385)	2001	0.473***	
Utilities	0.200	-0.051		(0.042)	
	(0.528)	(0.472)	2002	0.543***	
Transportation	-1.210***	-0.752*		(0.049)	
	(0.372)	(0.422)	2003	0.567***	
Wholesale and retail trade	-1.515***	-0.699		(0.055)	
	(0.320)	(0.688)	# of observ.	21,397	
Within R^2	0.051		Between R^2	0.045	

Notes: The dependent variable is the difference between log of reported earnings and income elasticity adjusted log of car values: $log\,E^R - 1/\lambda\, log\,C$. Income elasticity is $\lambda = 0.35$. Robust standard errors in parentheses. The omitted category is the benchmark sample of foreign-owned employers in Appendix 5. Regressions also include percentile in EED and its square term and the constant term. The sample is restricted to car owners with 5 or less cars, aged between 18 1999 and 60 in 2003, excluding those earning less than the annual equivalent of the legal minimum wage or more than the equivalent of US $100,000 in a given year and those whose primary earnings in any year 1999-2003 came from lottery winnings, veteran charitable foundation or interest and dividend incomes from the state savings bank or two largest state-owned corporations. ***, **, and * indicates significance at 1 percent, 5 percent, and 10 percent levels, respectively.