Essays in Labor Economics and Development Economics

by

Evgeny Yurevich Yakovlev

A dissertation submitted in partial satisfaction
of the requirements for the degree of
Doctor of Philosophy

in

Economics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Denis Nekipelov, co-chair
Professor Gerard Roland, co-chair
Professor David Card
Professor Steven Raphael

Fall, 2012

Essays in Labor Economics and Development Economics

by

Evgeny Yurevich Yakovlev

Abstract

Essays in Labor Economics and Development Economics

by

Evgeny Yurevich Yakovlev

Doctor of Philosophy in Economics

University of California, Berkeley

Professors Denis Nekipelov and Gerard Roland, Chairs

This dissertation contains three essays on labor economics and development economics.

In the first and second chapters, I examine determinants and consequences of alcohol consumption in Russia and quantify the effects of various public policies on mortality rates and on consumer welfare. For the past twenty years, Russia has confronted the Mortality Crisis – the life expectancy of Russian males has fallen by more than five years, and the mortality rate has increased by 50%. Alcohol abuse is widely agreed to be the main cause of this change.

In the first chapter, I employ a rich dataset on individual alcohol consumption to analyze the determinants for heavy drinking in Russia, including the price of alcohol, peer effects, and habits. I exploit unique location identifiers in my data and patterns of geographical settlement in Russia to measure peers within narrowly-defined neighborhoods. This definition of peers is validated by documenting a strong increase in alcohol consumption around the birthday of peers. With natural experiments, I estimate the own-price elasticity of the probability of heavy drinking using variation in alcohol regulations across Russian regions and over time. From these data, I develop a dynamic structural model of heavy drinking to quantify how changes in the price of alcohol would affect the proportion of heavy drinkers among Russian males (and subsequently also affect mortality rates). I find that that higher alcohol prices reduce the probability of being a heavy drinker by a non-trivial amount. An increase in the price of vodka by 50% would save the lives of 40,000 males annually, and would result in an increase in welfare. Peers account for a quarter of this effect.

The second chapter analyzes the consequences of government policy towards light alcohol drinks. Light drinks are commonly viewed as stepping stone to harder drinks, but also as safer substitutes for them. Here, I analyze this trade-off by utilizing micro-level

data on the alcohol consumption of Russian males. I find, first, that beer is a safer drink compared to hard alcohol beverages, in the sense that consumption of hard beverages increases the hazard of death while consumption of beer does not. Second, I find that beer is a substitute for vodka: there is significant positive cross-price elasticity of vodka consumption with respect to beer price. I find also little evidence that beer consumption actually serves as stepping stone for vodka consumption. Initiation of beer consumption instead forms habits for the further consumption of beer. Drinking beer at earlier ages results in higher beer consumption and higher overall alcohol intake in older years, but also results in reduced consumption of hard drinks compared to vodka drinkers and to non-abstainers. Finally, I estimate a multivariate model of consumer choice, and quantify the effect of different government policies on mortality rates, drinking patterns, and consumer welfare. I find that the taxation of beer may decrease consumer welfare and increase mortality rates. In contrast, subsidizing beer consumption will increase consumer welfare and even slightly decrease mortality rates.

The third chapter of my dissertation documents the unequal enforcement of liberalization reform of business regulation across Russian regions with different governance institutions, which leads to unequal effects of liberalization. National liberalization laws were enforced more effectively in sub-national regions with a more transparent government, more-informed population, higher concentration of industry, and stronger fiscal autonomy. As a result, in regions with stronger governance institutions liberalization had a substantial positive effect on the performance of small firms and on the growth of the official small-business sector in general. In contrast, in regions with weaker governance institutions there is no effect from the reform, and in some cases even a negative effect is observed.

To Vanya and Katya

Contents

1 Peers and Alcohol 1

1.1 Introduction 1
1.2 Literature Review 5
1.3 Data Description 7
1.3.1 "Peers" Definition 9
1.3.2 Alcohol consumption variable 10
1.4 Model 11
1.5 Estimation 14
1.5.1 Myopic agents, $\beta = 0$ 14
1.5.1.1 Estimation of utility parameters 14
1.5.1.2 Estimation of the price elasticity 15
1.5.2 Forward-looking agents, $\beta = 0.9$ 16
1.5.2.1 Estimation of utility parameters 17
1.5.2.2 Estimation of price elasticity 18
1.6 Results. 18
1.7 Robustness check 25
1.7.1 Reduced-form elasticity estimates 25
1.7.2 Linear in means peer effect 26
1.7.3 Robustness of dynamic model assumptions 27
1.7.4 Habits versus unobserved heterogeneity 28
1.7.5 Extension 29
1.8 Conclusion 29
1.9 Tables 31
1.10 Appendix 37

2 Beer versus Vodka 52

2.1 Introduction 52
2.2 Data 54
2.3 Alcohol consumption variables and drinking patterns of Russian males 55
2.4 Elasticity of alcohol consumption: hedonic regressions 56
2.5 Hazard of death regression 57
2.6 The stepping-stone effect 59
2.7 The Model 60
2.8 Conclusion 63
2.9 Tables 64

3 The Unequal Enforcement of Liberalization 73

3.1 Introduction 73
3.2 Background and the measures of regulation 77
3.2.1 Russia's liberalization reform of business regulation 77
3.2.2 The MABS survey 79
3.2.2.1 The measures of enforcement of liberalization 81
3.3 Governance institutions and the enforcement of liberalization 83
3.3.1 Hypotheses and measures of institutions 83
3.3.2 Three regulatory areas, taken separately 85
3.3.2.1 Methodology 85
3.3.2.2 Results 92
3.3.2.3 Magnitude: an example of Amur and Samara regions 89
3.3.3 Average effect across the three regulatory areas 90
3.3.4 Testing the assumption about the absence of regional trends .92
3.4 The outcomes of liberalization 92
3.4.1 Firm performance 92
3.4.2 Small business entry and employment 95
3.4.3 Magnitude: the outcomes of liberalization in Amur and Samara regions 97
3.5 Robustness 98
3.6 Conclusions 100
3.7 Tables 101

Chapter 1

Peers and Alcohol

1.1 Introduction

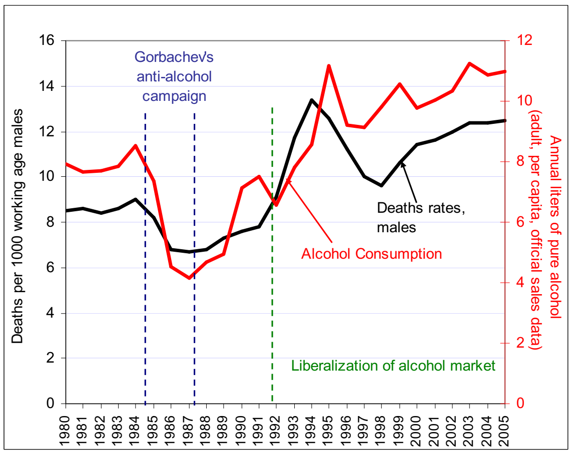
Russian males are notorious for their hard drinking. The Russian (non-abstainer) male consumes an average of 35.4 liters of pure alcohol per year.[1] This amount is equivalent to the daily consumption of 6 bottles of beer or 0.25 liters of vodka. The most notable example of the severe consequences of alcohol consumption is the male mortality crisis – male life expectancy in Russia is only 60 years. This is 8 years below the average in the (remaining) BRIC countries, 5 years below the world average, and below that in Bangladesh, Yemen, and North Korea. High alcohol consumption is frequently cited as the main cause (see for example Bhattacharya et al. 2011, Treisman 2010, Brainerd and Cutler 2005, Leon et al. 2007, Nemtsov 2002).[2] Approximately one-third of all deaths in Russia are related to alcohol consumption (see Nemtsov 2002). Most of the burden lies on males of working age: half of all deaths in working-age men are accounted for by hazardous

[1] See the WHO Global Status Report On Alcohol And Health (2011). More than 90% of Russian males of working age are non-abstainers. Per-adult consumption estimates vary from 11 to 18 liters of pure alcohol per year. Official statistics that take into account only legal sales report 11 liters; however, expert estimates are 15-18 liters (see Nemtsov 2002, WHO 2011, report of Minister of Internal Affairs, http://en.rian.ru/russia/20090924/156238102.html).

[2] In comparison, the situation with female mortality is not so bad. Female life expectancy in Russia is 73 years – 5 years higher than world average, and 2 years above of average in the (remaining) BRIC countries. For health statistics, see https://www.cia.gov/library/publications/the-world-factbook/fields/2102.html

drinking (see Leon et al. 2007, Zaridze et al. 2009, and Figure 1 below).

Figure 1. Alcohol Consumption and Male Mortality Rate.



Source: WHO, Treisman (2010), Rosstat.

Surprisingly, no attempts have been made to quantify the effects of public policy on mortality rates, and there have been few efforts to identify the effects of public policy on alcohol consumption. Moreover, research that identifies the causal effect of price on alcohol consumption and mortality deals with only aggregate (regional-level) data.[3] However, the use of disaggregated data is of particular interest because it allows disentangling the different forces that bear on individual decisions about drinking. Also, it allows an evaluation of the effect of policy on different subgroups.

My paper fills this gap. I utilize micro-level data on the alcohol consumption of Russian males to answer the following two key questions. First, how can we quantify the effects of a price increase for alcohol on the proportion of heavy drinkers and on mortality rates and social welfare? Second, how can we identify the effects of structural forces that influence alcohol consumption, and specifically peer effects and forward-looking assumptions on agent behavior?

Peer effects are agreed to be very important for policy analysis because they produce a (social) multiplier effect. Recent literature emphasizes the importance of peers in making personal decisions, in particular whether to drink or not (see, for example, Gaviria and

[3] Regional-level analysis is done by Treisman (2010) and Bhattacharya et al. (2011).

Raphael 2001, Krauth 2005, Kremer and Levy 2008, Card and Giuliano 2011, Moretti and Mas 2009). There are sound reasons to believe that peer influence is even stronger in Russia because of patterns of the dense geographical settlement inherited from the Soviet Union and the very low level of mobility in Russia. In my paper, I exploit unique location identifiers in the data to measure peers within narrowly-defined neighborhoods. This definition of peers is validated by documenting a strong increase in alcohol consumption around the birthday of peers.

This paper then introduces a model that incorporates these peer effects, and verifies the predictions of the model against both myopic and forward-looking assumptions on agent behavior. Although there is no consensus regarding which model is more true, most literature on policy analysis deals with only myopic assumptions. At the same time, key consequences of alcohol consumption – on health, family, and employment status, for example – do not necessarily appear immediately, but rather increasingly manifest over the course of the next few years, or even much later in life (see Mullahy and Sindelar 1993, Cook and Moore 2000). Moreover, alcohol consumption forms a habit, and thus affects future behavior (see rational addiction literature, Becker and Murphy 1988). Given this, one expects that individuals may behave in a forward-looking manner when determining current alcohol consumption. Possible mis-specification from omitting forward-looking agent assumptions might introduce a significant bias in estimates, and as such might result in incorrect predictions regarding proposed changes in the regulation of the alcohol industry.

In this paper, I employ recent developments in the econometric analysis of static and dynamic models of strategic interactions to model and estimate individual decision problems (for review, see Bajari et al. 2011a). Peer effects are modeled in the context of game with incomplete information. In my model, agents use the demographic characteristics of peers to form beliefs about peers' unobservable decisions regarding drinking. This model is naturally extended to a dynamic framework, where agents have rational expectations about future outcomes (see Bajari et al 2008, Aguirregabiria and Mira 2007, Berry, Pakes, and Ostrovsky 2007, and Pesendorfer and Schmidt-Dengler 2008).

In my estimates, I show the importance of peer effects for young age strata (below age 40). In addition, I find a non-trivial price elasticity for heavy drinking. To estimate the own price elasticity, I explore an exogenous variation in the price of alcohol that comes from changes in alcohol regulations across Russian regions and over time.

To illustrate these findings, I simulate the effect of an increase in vodka price by 50 percent on the probability of being a heavy drinker. A myopic model predicts that five years after introducing a price-raising tax, the proportion of heavy drinkers would decrease by roughly one-third, from 25 to 18 percent. The effect is higher for younger generations because of the non-trivial effect of a social multiplier. This cumulative effect can be decomposed in the following way: own one-period price elasticity predicts a drop in the share of heavy drinkers by roughly 4.5 percentage points, from 25 to 20.5 percent. In addition, peer effects increase the estimated price response by 1.5 times for younger generations. Further, the assumption that agents are forward-looking increases the estimated cumulative effect by roughly an additional 20 percent, although the difference in predicted effects in both models is insignificant.

Then, I simulate the consequences of a price-raising alcohol tax on mortality rates. I find a significant age heterogeneity in the effect of heavy drinking on the hazard of death: this effect is much stronger for younger generations. Increasing the price of vodka by 50 percent results in a decrease in mortality rates by one-fourth for males of ages 18-29, and by one-fifth for males ages 30-39, but with no effect on the mortality of males of older ages.

My results coincide with the regional-level analyses by Treisman (2010) and Bhattacharya et al. (2011), and with the micro-level analyses by Andrienko and Nemtsov (2006) and Denisova (2010). Treisman (2010) utilizes regional-level data for the period 1997-2006, and shows that the increase in heavy drinking resulted largely from an increase in the affordability of vodka. In 1990 – immediately before liberalization of the Russian alcohol market – the price of vodka relative to CPI was four times higher than in 2006. Treisman shows that demand for alcohol is (relatively) elastic, and that variations in vodka price closely match variations in mortality rates. Bhattacharya et al. (2011) use regional-level data from the period of Gorbachev's anti-alcohol campaign, and find that regions experiencing a higher intensity of the campaign also exhibited a higher drop in mortality rates. They argue that the surge in mortality that happened after Gorbachev's campaign can be explained (partly) by a mean reversion effect. Andrienko and Nemtsov (2006) and Denisova (2010) utilize micro-level data on alcohol consumption to reach similar conclusions. Andrienko and Nemtsov (2006) find a negative correlation between the price of alcohol and alcohol consumption. Denisova (2010) studies determinants of mortality in Russia, and finds a correlation between alcohol consumption and hazard of death.

Finally, I analyze the effect of a tax increase on social welfare. I find that when agents have bounded rationality (that is, do not take into account the effect of consumption on hazard of death), a raise in vodka price by 50 percent improves welfare. I find also that under certain assumptions on agent utilities, a tax increases consumer welfare even for fully-rational agents.

This paper is organized as follows. In the following section I review existing empirical literature on peer effects and rational addiction, and on the estimation of dynamic models. In Section 1.3, I describe my data and the variables used in my analysis. Sections 1.4 and 1.5 present the model and estimation strategy. In Section 1.6, I discuss results. Section 1.7 discusses robustness checks. Section 1.8 concludes.

1.2 Literature Review

Recent literature has demonstrated a renewed interest in endogenous preference formation, such as peer influence. Theoretical treatments include those by Akerlof and Kranton (2000), Becker (1996), and others. Empirical research studying social interaction concentrates on resolving the identification problems described in Manski's seminal paper (1993). The naïve approach of analyzing peer effects that was dominant prior to Manski's paper analyzed only the (residual) correlation between individual choice and the average choice of people from a reference group. Manski's primary critique of this approach was that parameters of interest were not identified – the effects would be contaminated by common unobservable factors, non-random reference group selection, the endogeneity of other group members' choices (correlated effects), and the influence of group characteristics (rather than group choice) on individual behavior (contextual effects). In contrast to endogenous peer effects, both contextual effects and correlated effects do not produce a social multiplier.

Different identification approaches have been proposed to solve the problems introduced in Manski's critique. For reviews of these studies, see Blume and Durlauf (2005).

The primary approaches in the empirical labor literature are the random assignment of peers (see Kremer and Levy 2008, Katz et al. 2001) and finding the exogenous variation of peer characteristics (see Gaviria and Raphael 2001). Glaeser, Sacerdote, and Scheinkman (2002) and Graham (2008) use structural models to infer the magnitude of peer effects from aggregate statistics. Krauth (2005) employs a structural approach to directly model endogenous choice and correlated effects.

Empirical industrial organization literature contributes to this by providing an intuitive structural framework for the analysis of peer interaction (see for example Bajari et al. 2008, Aguirregabiria and Mira 2007, Berry, Pakes, and Ostrovsky 2007, and Pesendorfer and Schmidt-Dengler 2008). In this research, the structural framework takes the form of games, with incomplete information. Agents do not observe other people's actions or form beliefs from what people do based on observable state variables. The expected utility of agent therefore does not include the actions of peers, but only the beliefs of the agent. Estimations in this model are very similar to those in the two-stage approach, where in the first stage the researcher estimates the agent's beliefs, and in the second stage the researcher estimates utility parameters, including peer effect. In contrast to other proposed approaches, this approach is structural. Introducing structure to the model allows the researcher to model the effect of policy on different economic factors, such as consumer welfare and the death rate. This approach also allows for analyzing strategic interactions in both static and dynamic contexts.

TThe dynamic nature of the agent problem when the agent consumes addictive goods is emphasized in rational addiction literature, initiated by Becker and Murphy (1988). In their model, individuals choose between immediate gains from the consumption of addictive goods and future costs associated with addiction. This model confronted the prevailing (at that time) view treating agents as myopic, and the empirical studies that follow Becker and Murphy's research offer different results. Some find empirical evidence to support the rational addiction model (see Murphy, Becker, Grossman, and Murphy 1991 and 1994, Chaloupka 2000, Arcidiacono et al. 2007). Other studies question this evidence (see Auld and Grootendorst 2004), or provide an alternative to a (fully) rational-model explanation of the evidence (see Gruber and Köszegi 2001).[4]

[4]Most of the studies that test the validity of the forward-looking hypothesis provide only an indirect test,

Still, there is no consensus regarding which model prevails in explaining and describing addictive behavior. One reason for this is that, in general, the set-up of these models is hardly (or even simply not) distinguishable from the data. Thus, a seminal result from Rust (1994) contrasts with results from dynamic discrete-choice models; he concludes that in a general set-up (with non-parametric utilities) the discounting parameter β is not identified. Although today different identification results are stated, they all are obtained under certain restrictions on parameters (see for example Magnac and Thermar 2002, Hang and Wang 2010, Arcidiacono et al. 2007).

Even though there is no agreement on the β majority of existing empirical literature still uses only the myopic framework to analyze the consumption of addictive goods. In my view, this happens first because myopic models are easier to analyze, and second because until recently dynamic models were very restrictive in requiring discretization of variables, worked with only a small set of variables, and so on. Recent developments in methods of dynamic discrete models have successfully eliminated many of these restrictions. For excellent surveys of the current state of dynamic discrete models, see work by Aguirregabiria and Mira (2010) and Bajari et al. (2011a).

1.3 Data Description

Typical patterns of geographical settlement in Russia – a remainder of the Soviet Union's legacy – allow me to use geographic closeness as a measure of the likelihood of status as a peer. Approximately 10% of Russian families live in dormitories and communal houses, where residents share kitchens and bathrooms.[5] A majority of the remaining, more fortunate, part of the population lives in a complex of several multi-story multi-apartment buildings, called a "dvor." These complexes have their own playgrounds, athletic fields, and ice rinks, and often serve as the place where people spend leisure time.[6] Photos of typical dvors are presented in Figure A2 in the appendix. Dvors are the most popular

looking at the correlation between the current consumption of an addictive product and its future price. These methods are subject to a meaningful drawback, potentially identifying a spurious correlation and so wrongly supporting the rational addiction analysis (Auld and Grootendorst 2004).

[5]See the RLMS web site, http://www.cpc.unc.edu/projects/rlms-hse/project/sampling

[6]The size of dvor vary in range from 200 to more than 2000 inhabitants. The most common dvors are (relatively) small-size dvors with population of roughly 300 people (so called khrushchevki).

place in Russia to find friends – the very low level of personal mobility in Russia means that most people live in the same place (and therefore the same dvor) for most of their lives.

In this study, I utilize data from the Russian Longitudinal Monitoring survey (RLMS)[7], which – fortunately for me – contains data on small neighborhoods where respondents live. The RLMS is a nationally-representative annual survey that covers more than 4,000 households (with between 7413 and 9444 individual respondents), starting from 1992. For every respondent in the survey, the RLMS identifies the census district in which he or she lives. The average population of census district in Russia is 300.[8] Typical census district in Russia contains one dvor or one multi-story building; this allows me to use information on neighborhood (and age) to successfully identify peer groups.[9]

The RLMS also has other advantages over existing data sets. It provides a survey of a very broad set of questions, including a variety of individual demographic characteristics, consumption data, and so on. In particular it includes data on death events, so I can identify the effects of drinking on mortality from micro-level data. Further, it contains rich data on neighborhood characteristics, including – critically – the price of alcoholic beverages in each neighborhood, allowing me to analyze individual price elasticity.

My study utilizes rounds 5 through 16 of RLMS.[10] over a time span from 1994 to 2007, except 1997 and 1999. The data cover 33 regions – 31 oblasts (krays, republics), plus Moscow and St. Petersburg. Two of the regions are Muslim. Seventy-five percent of respondents live in an urban area. Forty three percents of respondents are male. The percentage of male respondents decreases with age, from 49% for ages 13-20, to 36% for ages

[7] This survey is conducted by the Carolina Population Center at the University of Carolina at Chapel Hill, and by the High School of Economics in Moscow. Official Source name: "Russia Longitudinal Monitoring survey, RLMS-HSE," conducted by Higher School of Economics and ZAO "Demoscope" together with Carolina Population Center, University of North Carolina at Chapel Hill and the Institute of Sociology RAS. (RLMS-HSE web sites: http://www.cpc.unc.edu/projects/rlms-hse, http://www.hse.ru/org/hse/rlms).

[8] RLMS team indicates that population of census districts in RLMS survey is in range between 250 and 4000 people. There are 459,000 census districts in Russia (data on 2010 census). This number implies that average population of census district is 310 people (including females, youth and elderly). This number in turn implies, that average population of peer group is 21 (adult males in the same age strata).

[9] Later in the paper I provide a check confirming that this definition of peers has ground.

[10] I do not utilize data on rounds earlier than round 5 because they were conducted by other institution, have different methodology, and are generally agreed to be of worse quality.

above 50. The data cover only individuals older than 13 years.

The RLMS data have a low attrition rate, which can be explained by low levels of labor mobility in Russia (See Andrienko and Guriev 2004). Interview completion exceeds 84 percent, lowest in Moscow and St. Petersbug (60%) and highest in Western Siberia (92%). The RLMS team provides a detailed analysis of attrition effects, and finds no significant effect of attrition.[11]

My primary object of interest for this research is males of ages between 18 and 65. The threshold of 18 years is chosen because it is officially prohibited to drink alcohol before this age. The resulting sample consists of 29554 individuals*year points (2937 to 3742 individuals per year). Summary statistics for primary demographic characteristics are presented in Table 3.

1.3.1 "Peers" Definition

I define "peers" as those who live in one neighborhood (school district) and belong to the same age stratum. Applying this definition, I constructed peer groups. The median number of people in a group is 5; the lower 1% is 2, the upper 90% is 20, and largest number is 66. On average, I have 835 peer groups (each with 2 or more peers) per year. The distribution of the number of peers per peer group is shown in Table 4.

To verify the reliability of my measures, I provide the following test: I correlate log (the amount of vodka consumption) with a dummy variable if a person has a birthday in the previous month, and with averages of the birthday dummy variables across peers.[12]Vodka is the most popular alcoholic beverage to serve on birthdays, compared to beer and for males also to wine. Results for both regressions are positive and statistically significant. Regression suggests that a person's consumption of vodka increases by 16% if his birthday is during the previous month, and by 6% if there was a birthday of one of his peers (in a group of 5 peers). The results are robust if I eliminate household members from the

[11]See http://www.cpc.unc.edu/projects/rlms-hse/project/samprep

[12]The specifications of the regressions are as follows:
$Log(1 + vodka)_{it} = \alpha_1 + \alpha_2 I(birthday)_{it} + \varepsilon_{it}$,
$Log(1 + vodka)_{it} = \zeta_1 + \zeta_2 \sum_{j \in peers} I(birthday)_{jt}/(N-1) + \varepsilon_{it}$,
where $vodka$ stands for amount of vodka have drunk last month (in milliliters).

sample of peers.[13]

Table 1. Birthdays and Alcohol Consumption.

	All peers		Without household members	
	log(vodka)	+1 birthday in group of 5	log(vodka)	+1 birthday in group of 5
$\frac{\sum_{peers} I(birthday)}{(N-1)}$	0.227	0.057	0.212	0.053
	[0.086]***	[0.021]***	[0.086]**	[0.021]***
$I(birthday)$	0.161	0.161	0.161	0.161
	[0.053]***	[0.053]***	[0.053]***	[0.053]***
Year*month FE	Yes	Yes	Yes	Yes
Observations	35995	35995	35995	35995

* significant at 10%; ** significant at 5%; *** significant at 1%

1.3.2 Alcohol consumption variable

Although the negative health and social consequences of hard drinking are widely recognized, there is no evidence for negative consequences from moderate drinking. Thus, I concentrated on an analysis of the personal decision to drink "hard" or not. I use a dummy variable that equals 1 if a person belongs to the top quarter of alcohol consumption (among males of working age). Alcohol consumption is measured as the reported amount of pure alcohol consumed the previous month.[14]

However, alcohol consumption reporting in the RLMS suffers from the common problem of all individual-level consumption surveys: it is significantly under-reported.[15] So, to offer an indication of the actual level of alcohol consumption corresponding to the thresh-

[13] The results are robust using a different measure of vodka consumption. There is no effect (or a small negative effect) of peer birthdays on the consumption of other goods, such as tea, coffee, or cigarettes (see Table A1 in the appendix).

[14] It is worth noting that sometimes a high level of monthly average alcohol consumption is not as harmful for health as one-time binge drinking (with a relatively low average level otherwise). Still, the measure I choose indicates that heavy drinking has huge adverse effect on health (see hazard of death regression).

[15] This is the common problem of all individual-level surveys that study alcohol consumption. Reported threshold level corresponds to reported amount drinking of more 155 grams of pure alcohol per month. A summary statistics and age profiles for reported amounts of alcohol consumption are shown in Table 3 and Figure A1 in the appendix.

old of being a "heavy drinker," I correlate the reports of consumption from the RLMS data with official sales data as a benchmark for average levels of alcohol consumption.

The threshold level for being a "heavy drinker" is 2.6 times the mean alcohol consumption (including women and the elderly) in the RLMS sample. If I take mean alcohol consumption from official sales data (11 liters of pure alcohol per year per person), I can determine that the actual threshold is equivalent to an annual consumption of 29 liters of pure alcohol. This amount corresponds to a daily of consumption of 5 bottles (0.33 liters each, 1.66 liters total) of beer, or 0.2 liters of vodka. If I use (more reliable) expert estimates as a benchmark, then the threshold corresponds to daily consumption of 7 bottles of beer, or 0.29 liters of vodka.

In the Robustness section, I present the results of regressions, where alternative measures of alcohol consumption are used.

1.4 Model

The set-up of the model is as follows.

There are N agents in an (exogenously-given) peer group: $i = \{1, ..., N\}$. In every period of time t agents simultaneously choose an action, a_{it}. The set of actions, a_{it} is binary: whether to drink hard $a_{it} = 1$ or not, $a_{it} = 0$.

The expected present value of agent utility consists of current per period utility, $\pi_{it}(a_{-it}, a_{it}, s_t)$, discounted expected value function, $\beta E(V_{it+1}(s_{t+1})|a_{-it}, a_{it}, s_t)$, and a stochastic preference shock, $e_{it}(a_{it})$:

$$U(a_{-it}, a_{it}, s_t) = \pi_{it}(a_{-it}, a_{it}, s_t) + \beta E(V_{it+1}(s_{t+1})|a_{-it}, a_{it}, s_t) + e_{it}(a_{it})$$

Per-period utility $\pi_{it}(.)$ and private preference shock $e_{it}(.)$ given $a_{it} = 0$ are normalized to zero: $\pi_{it}(a_{it} = 0) = 0$ and $e_{it}(a_{it} = 0) = 0$.

Private preference shocks $e_{it}(1)$ have i.i.d. logistic distribution. Private preference shocks stay personal tastes for heavy drinking, tolerance to alcohol and other factors that observable for the agent, but unobservable for researcher and for other peers in the group.

Further, I will consider two different assumptions on β, that $\beta = 0$ (for myopic agents) and $\beta = 0.9$ (for forward-looking agents).

For the case of forward-looking agents I assume that agents have an infinite time planning horizon, and that the transition process of state variables is Markovian.This implies that expectations for future periods depend on only a current-period realization of state variables and agent choice of action. Finally, I restrict equilibrium to be a Markov Perfect Equilibrium, so that an agent's strategy is restricted to be a function of the current state variables and the realization of a random part of utility (private preference shock). These assumptions ensure identification, and are common in dynamic-choice models. For myopic agents the model is static, such that none of the assumptions described above is needed.

I also assume that given choice $a_{it} = 1$ the per-period utility of the agent has the linear parameterization:

$$\pi_{it}(a_{-it}, a_{it} = 1, s_t) = \delta \frac{\sum_{-i} I(a_{jt} = 1)}{N-1} + \gamma habit_{it} + \Gamma' D_{it} + \Upsilon' G_{-it} + \rho_{mt}$$

Thus, $\pi_{it}(a_{-it}, a_{it} = 1, s_t)$ depends on average peer alcohol consumption, habits $(a_{i,t-1})$[16], a set of personal demographic characteristics (D_{it}), (sub) set of peers characteristics G_{-it} and municipality*year invariant factors ρ_{mt}.

The set of personal demographic characteristics D_{it} includes weight, education, work status, lagged I(smokes), I(Muslim), health status, age, age squared, marital status, size of family and log(family income). The (sub) set of peers characteristics G_{-it} that stands for so-called exogenous effects includes share of Muslims, share of peers with college education, share of unemployed.[17] I include municipality*year invariant factors ρ_{mt} to account

[16] I define state variable $habit_{it}$ as follows. Let state variable $habit_{it} = 0$ if $age_{it} < 18$(years) and let transition process of $habit_{it}$ be defined in following way: $habit_{it}(S_{t-1,}a_{i,t-1}) = a_{i,t-1} + \varphi_{i,t}$ if $age_{it} \geq 18$, where $a_{i,t-1}$ is agent equilibrium choice of action in previous period, and $\varphi_{i,t}$ is (negligible) smoothing noise. $\varphi_{i,t}$ is added to ensure existense of equilibrium. With this definition of habits, the model satisfies assumptions requred for MPE (see for example, Assumptions AS, IID and CI-X in Aguirregabiria and Mira, 2007 or Bajari et al 2010). A Markov perfect equilibrium (MPE) in this game is a set of strategy functions $a^\star$ such that for any agent i and for any $\{S_t,\ e_{it}\}$, where $S_t = U_{j\in\{i,-i\}}\{habit_{jt},\ D_{jt},\ G_{nt},\ \rho_{mt}\}$ we have that $a_i^\star(S_t, e_{it}) = b(S_t, e_{it}, a_{-i}^\star)$.

[17] Exclusion restriction requires that subset G_{-it} does not contain all set of demographic variables. It seems to be reasonable assumption: for example, agent does not have higher utility when drink with peers with different weight, different marital or health status. Actually my estimates show that agent does not

for price, weather and other factors that affect an agent's utility, and that (I assume) vary only on the municipality*year level.

Subscripts $i,\ t,\ m$ stand for individual, year, and municipality; subscript $-i$ stands for other individuals within the same peer group.

I assume a game with an incomplete information set up.[18] Agents do not observe peer choices and do not observe realization of peer private shocks, $e_{it}(a_{it})$. They form expectations of other peer actions. The expectations are based on agent (consistent) beliefs of what peers do. These beliefs depend on a set of state variables, observed by agents. In my case, beliefs are based on (own and peers') set of variables $S_{i,-i,t} = U_{j\in\{i,-i\}}\{habit_{jt}, D_{jt}, G_{nt}, \rho_{mt}\}$.

Thus, an agent's expected (over beliefs) per-period utility in case of $a_i = 1$ is:

$$E_{e_{-i}}\pi_{it}(a_{-it}, a_{it} = 1, s_t) = \delta\overline{\sigma_{jt}(a_{jt} = 1|S_{i,-i,t})} + \gamma habit_{it} + \Gamma' D_{it} + \Upsilon' G_{-it} + \rho_{mt}$$

The term $\overline{\sigma_{jt}(a_{jt} = 1|S_{i,-i,t})} = \dfrac{\sum_{-i}\sigma_{jt}(a_{jt} = 1|S_{i,-i,t})}{N-1}$, where $\sigma_{jt}(a_{jt} = 1|S_{i,-i,t})$ stands for the agent's i belief of what player j will do. I follow this notation throughout this paper.

Finally, an agent chooses to drink hard if his or her expected present value of the utility of (heavy) drinking is greater than the utility of not drinking:

$$\begin{aligned} E_{e_{-i}}\pi_{it}(a_{-it}, a_{it} &= 1, s_t) + \beta E(V_{it+1}(s_{t+1})|a_{-it}, a_{it} = 1, s_t) + e_{it}(a_{it} = 1) \\ &> \beta E(V_{it+1}(s_{t+1})|a_{-it}, a_{it} = 0, s_t) \end{aligned}$$

In the following section, I discuss the estimation procedure for two parametrizations of the discount factor, $\beta = 0$ and $\beta = 0.9$. Case $\beta = 0$ refers to "myopic" agents, while $\beta = 0.9$ refers to "forward-looking" agents.[19]

have any preferences about G_{-it}: all coefficients in Υ' are insignificant.

[18] In both games with complete and incomplete information agents do not observe actions of others if they make their decisions simultaneously. Within game with an incomplete (rather than complete) information set-up agents do not know payoffs of other players because these payoffs include private preference shocks $e_{it}(1)$. When starting drinking, people do not know how much their peers will drink: they may end up to drink a lot or just one shot. Game of incomplete information gives me the game-theoretic motivation to use demographic characteristics of peers as instruments for their drinking behavior.

[19] I discuss both of the models because there is no consensus in the literature regarding which assumption

To simplify the exposition of the model and estimation, I start with the less-technical case, the myopic agent model.

1.5 Estimation

1.5.1 Myopic agents, $\beta = 0$

Under the assumption that agents are myopic, the expected utility of agent is simplified to the following expression:

$E_{e_{-i}}U_{it}(1) = \delta\overline{\sigma_{jt}(a_{jt} = 1|S_{i,-i,t})} + \gamma habit_{it} + \Gamma' D_{it} + \Upsilon' G_{-it} + \rho_{mt} + e_{it}(1)$, and
$E_{e_{-i}}U_{it}(0) = 0$

An agent chooses to drink hard if his or her expected utility of heavy drinking is greater than zero: $EU_{it}(1) > 0$.

1.5.1.1 Estimation of utility parameters

Estimation of the model proceeds in two steps. These steps are similar to the standard 2SLS regression procedure.

On the first stage, I (non-parametrically) estimate beliefs $\hat{\sigma}_{jt}(a_{jt} = 1|S_{i,-i,t})$:

$$I(a_{jt} = 1)_{it} = H(s_{it})'\zeta + \varepsilon_{it}$$

where I_i =$I(a_{it} = 1)$, $H(s_{it})$ is a set of Hermite polynomials of state variables s_{it}.[20] That is, $H(s_{it})$ contains set of Hermite polynomials up to the third degree of $S_{i,-i,t} = U_{j\in\{i,-i\}}\{habit_{jt},\ D_{jt},\ G_{nt},\ \rho_{mt}\}$. In addition it includes interactions of state variables $U_{j\in\{i,-i\}}\{habit_{jt},\ D_{jt},\ G_{nt}\}$. I do not extend the set of polynomials to a larger degree or include a larger set of interactions because of dimensionality problem. One important implication (for me) of this strategy is that ρ_{mt} appears in $H(s_{it})$ only once: this happens because the dummy variable structure of fixed effects implies that $\rho^k_{mt} = \rho_{mt}$.[21]

is more relevant for the analysis of drinking behavior. In general set-up, a discount factor is not identified (see Rust 1994).

[20]For a discussion of non-parametric regression with Hermite polynomials see Ai and Chen (2003).

[21]Still, ρ_{mt} will account for any variable (in any power) that varies only on municipality*year level.

On the second stage, I estimate the remaining parameters of utility function using logit regression:

$$\begin{aligned} E_{e_{-i}} u_{it}(1) &= \sum_k \delta_k I(age\, strata = k)\overline{\hat{\sigma}_{jt}(a_{jt} = 1|S_{i,-i,t})} \\ &\quad + \gamma habit_{it} + \Gamma' D_{it} + \Upsilon' G_{-it} + \rho_{mt} + e_{it}(1) \end{aligned}$$

where

$$\sigma_{it}(a_{it} \widehat{= 1}|S_{i,-i,t}) = H(s_{it})'\hat{\zeta}$$

are agent beliefs, estimated in the first stage.

I assume age heterogeneity in peer effects, so I estimate δ separately for every age stratum.

Parameters of the model are identified under the assumption that the utility of one agent does not depend on subset of peer demographic characteristics, and that random components of personal utility are independent of peer demographic characteristics (see Bajari et al. 2005 for proof). I discuss the robustness of my results in the Robustness section.

1.5.1.2 Estimation of the price elasticity

To estimate elasticity, I employ following strategy.

I assume that all price variation is captured on a municipality*year level. I obtain the municipality*year fixed effects component of utility $\hat{\rho}_{mt}$, and then regress $\hat{\rho}_{mt}$ on a log of the relative price of cheapest vodka in neighborhood.

$$\hat{\rho}_{mt} = \theta ln(Price)_{mt} + \delta_t + u_{mt}$$

I use data on regional regulation of the alcohol market to instrument the price variable. I use following variables as instruments: I(regional government imposes tax on producers), I(regional government imposes tax on retailers), I(regional government imposes additional measure to controls for alcohol excise payments).[22] The latter measure is a

[22]As a rule, regional regulations are imposed both to increase regional budget revenues (excise tax and license tax are two of the very few taxes that go directly into the regional budget) and as a result of the

popular tool in Russia because it controls the tax evasion of sellers of alcoholic beverages.

1.5.2 Forward-looking agents, $\beta = 0.9$

Here I present an estimation strategy for forward-looking agents (with $\beta = 0.9$).

Literature on the estimation of dynamic discrete models originated in 1987, after the seminal work of Rust (1987). During the last 20 years, tremendous progress has been made in this field. Further work significantly simplified the estimation procedure (Holtz and Miller 1993), discussed identification restrictions (Rust 1994), and extended dynamic discrete choice to the estimation of dynamic discrete games (Bajari et al. 2011, Aguirregabiria and Mira 2002, Berry, Pakes, and Ostrovsky 2007, and Pesendorfer and Schmidt-Dengler 2008). For excellent surveys of dynamic discrete models, see research by Aguirregabiria and Mira (2010) and Bajari et al. (2011b).

My estimation procedure follows Bajari et al. (2007). Compared to many other studies, the estimation strategy proposed by Bajari et al. has three advantages. First, this estimation procedure does not require the calculation of a transition matrix on the first stage. Avoiding this calculation decreases errors of estimation. Second, this estimation strategy allows using sequential procedure estimation, wherein every step of estimation has closed-form solutions. This means that one can avoid mistakes and problems related with finding a global maximum using a maximization routine. Finally, this estimation procedure does not require discretization of variables. This flexibility of estimation routine allows me to work with the same extensive set of explanatory variables as in the myopic (static) model, and thus makes these two models comparable.

The idea of this estimation is as follows. After applying two well-known relationships – Hotz-Miller inversion and expression for Emax (ex ante Value function) function – the choice-specific Bellman equation

$$V_{it}(a_{it}, s_t) = E_{e_{-i}} \pi_{it}(a_{-it}, a_{it} = 1, s_t) + \beta E(V_{it+1}(s_{t+1})|a_{it}, s_t)$$

can be rewritten as two moment equations (for derivation see Proof A1 in the ap-

lobbying of local firms and/or tollbooth corruption (see Yakovlev 2008, Slinko et al. 2005). This implies that the introduction of new regulation is generally not motivated by public health.

pendix):

Bellman equation for $V_i(0, s_t)$

$$V_{it}(0, s_t) = \beta E_{t+1}(log(1 + exp(log(\sigma_{it+1}(1)) - log(\sigma_{it+1}(0))|s_t, a_{it} = 0) \\ +\beta E_{t+1}(V_{it+1}(0, s_{t+1})|s_t, a_{it} = 0) \tag{1.1}$$

Bellman equation for $V_i(1, s_{it})$

$$log(\sigma_{it}(1)) - log(\sigma_{it}(0)) + V_{it}(0, s_t)_i = \pi_{it}(a_{-it}, a_{it} = 1, s_t, \theta) \\ +\beta E_{t+1}(V_{it+1}(0, s_{t+1}) - log(\sigma_{it+1}(0))|a_{it} = 1, s_t) \tag{1.2}$$

These two equations together with a moment condition on choice probabilities

$$E(I(a_i = k)|s_t) = \sigma_{it}(k|s_t), \ k \in \{0, 1\} \tag{1.3}$$

form the system of moments I estimate in next section.

1.5.2.1 Estimation of utility parameters

A shortcut of the estimation procedure is as follows[23]

The first step resembles the first step in in the estimation of the myopic model: I obtain estimates of choice probabilities $\widehat{\sigma_{it}(1)}$, $\widehat{\sigma_{it}(0)}$ from a sieve regression of $I(a_{it} = k)$ on Hermite polynomials of state variables:

$\widehat{\sigma_{it}(1)} = H(s_{it})'\hat{\zeta}$, $\widehat{\sigma_{it}(0)} = 1 - \widehat{\sigma_{it}(1)}$.

On the second step, I obtain nonparametric estimates of $V_{it}(0, s)$ by solving a sample equivalent of moment condition (1):

$$V_{it}\widehat{(0, s_{it})} = H(s_{it})'\hat{\mu}$$

I find $V_i\widehat{(0, s_t)}$ by finding $\hat{\mu}$ that solves following sample equivalent of moment condi-

[23]My sequential estimation procedure is not efficient. One can improve efficiency by solving three moment conditions altogether. In this case, however, there is no closed-form solution, and so one will face computational difficulties related to the problem of finding the (correct) global maximum of the GMM objective function with many variables.

tion (3):

$$I(a_{it}=0)[H(s_{it})'\hat{\mu}] = \beta I(a_{it}=0)[(log(1+exp(log(\widehat{\sigma_{it+1}(1)})-log(\widehat{\sigma_{it+1}(0)}))+H(s_{i+1})'\hat{\mu}]$$

On final step, I estimate $\pi(1,s)$ by solving for $\hat{\theta}$ sample equivalent of moment condition (2):

$$\begin{aligned} I(a_{it} &= 1)[s_t'\hat{\theta} + V_{it}\widehat{(0,s_t)} + log(\widehat{\sigma_{it}(1)}) - log(\widehat{\sigma_{it}(0)})] \\ &= \beta I(a_{it}=1)[(log(1+exp(log(\widehat{\sigma_{it+1}(1)})-log(\widehat{\sigma_{it+1}(0)}))+V_{it}\widehat{(0,s_{t+1})}]] \end{aligned}$$

1.5.2.2 Estimation of price elasticity

Here, I follow a procedure similar to that employed in the myopic case. From the estimation above, I obtain municipality*year fixed effects components $\hat{\rho}_{mt}(\pi)$, $\hat{\rho}_{mt}(EV1)$, $\hat{\rho}_{mt}(EV0)$ of my estimates of per-period utility $\pi_{it}(a_{-it}, a_{it}=1, s_t)$, and conditional expectation of future Value function, $\beta E(V_{it+1}(s_{t+1})|a_{it}=1, s_t)$, and $\beta E(V_{it+1}(s_{t+1})|a_{it}=0, s_t)$. Then I calculate the aggregate effect of fixed effect components, $\hat{\rho}_{mt}$:

$$\hat{\rho}_{mt} = \hat{\rho}_{mt}(\pi) + \hat{\rho}_{mt}(EV1) - \hat{\rho}_{mt}(EV0)$$

and then regress $\hat{\rho}_{mt}$ on log of the relative price of the cheapest vodka in neighborhood (with the same set of instruments as in myopic case):

$$\hat{\rho}_{mt} = \theta ln(Price)_{mt} + \delta_t + u_{mt}$$

1.6 Results

Estimates of per-period utility parameters are shown in Table 2 below, and in Tables 5 through 7 at the end of paper.

In both specifications (myopic and forward-looking agents), I find that peers have a strong effect on younger generations, with the effect decreasing with increasing age. For the two youngest strata, the effect is statistically significant. For myopic agents, $\widehat{\delta}$ equals to 1.355, 0.688, 0.039, and 0.09 for ages 18-29, 30-39, 40-49, and 50-65 respectively. For forward-looking agents, $\widehat{\delta}$ equals to 0.932, 0.456, 0.128, and 0.214 for ages 18-29, 30-39, 40-49, and 50-65 respectively.

The myopic model allows for an immediate statistical interpretation of the coefficients: an increase in peer average alcohol consumption of 0.2 (corresponding to a situation in which one out of five peers in a group becomes a heavy drinker) will increase the probability of becoming a heavy drinker for the "mean" person in age group 18-29 by 5.4 percentage points, and for "mean" person in age group 30-39 by 2.8 percentage points. The forward-looking model does not allow for immediate statistical interpretation; to evaluate how an increase in peer alcohol consumption affects agent decision, one must know not only the agent's per-period utility, but also have an expectation of the agent's future value function. In Table 6, I present point estimates of the marginal utility and marginal value function of peers, evaluated at the mean value of other state variables. Table 6 shows that in the forward-looking model, marginal value function (of peers) does not differ much from marginal per-period utility. The predicted marginal value function for the youngest age stratum is smaller than the marginal utility of myopic agents.

The per-period (indirect) marginal utility of myopic agents with respect to log(price) is equal to -0.82 and -0.68 for myopic and forward-looking agents respectively. For a myopic agent with mean level of all demographic characteristics, this coefficient implies that, for example, an increase in the price of vodka by 10% will lead to a decrease in the probability of heavy drinking by 6.5 percentage points (from 0.25 to 0.185). To evaluate the effect of a change in price on forward-looking agents, one must know not only the agent's per-period utility, but also have an expectation of the agent's future value function. The per-period marginal value function of agents with respect to log(price) is equal to -0.968. This number implies a (slightly) higher elasticity for forward-looking agents - an increase in the price of vodka by 50% leads to a decrease in the probability of becoming a heavy drinker by 7.8 percentage points.

Table 2. Agent's utility parameters. Point estimates.

	Myopic	Forward-looking	
	Per-period utility	Per-period utility	Value function
Log(vodka price)	-0.82**	-0.68*	-0.968**
Peers effect, $\widehat{\delta}$:			
age 18-29	1.355***	0.932***	0.961***
age 30-39	0.688***	0.456 ***	0.609***
age 40-49	0.039	0.128	0.073
age 50-59	0.09	0.214	0.18
Habit: lag I(heavy drinker)	1.27***	1.234***	

Note: * significant at 10%** significant at 5%;*** significant at 1%

In elasticity estimates standard errors are clustered on municipality*year level

However, the description of utility parameters above does not offer a full picture of what happens with agent decisions regarding heavy drinking when the price of alcohol changes. One needs to calculate new equilibrium consumption levels after the price has changed, as well as to take in account that the change in price will have an effect on future consumption through a change in habits. To evaluate the response of a consumer to a price change, I evaluate the cumulative effect of own elasticity, the peer effect, and the effect of a change in habits (and other state variables). To do this, I simulate agent response to a 50% increase in price for the 5-year period after the price change.

Figure 2 illustrates the decomposition of the cumulative response to change in price for males age 18-29. Dashed lines show the effect of a price increase on myopic agents for three situations: in a model where peer effects and habit formation are included, in a model without peer effects, and in a model without habit formation. The difference in effects refers to the effect of the social multiplier and of the "habit multiplier." Solid lines show the effect of a price-increasing tax for forward-looking agents. The forward-looking model predicts a decrease in the proportion of heavy drinkers by 8 percentage points, from 22.5% to 14.5% over five years. The myopic model predicts a (slightly) smaller decrease of 7.5 percentage points, from 22.5% to 15%. Taking into account only peer effects or only habit formation leads to a prediction of smaller changes: 5.3 percentage points versus 5.6 percentage points. Finally, own price elasticity results in a one-time change of 4.3 percentage points, which is approximately half of the cumulative effect.

Figure 2. Effect of tax on Pr(heavy drinker), age 18-29.

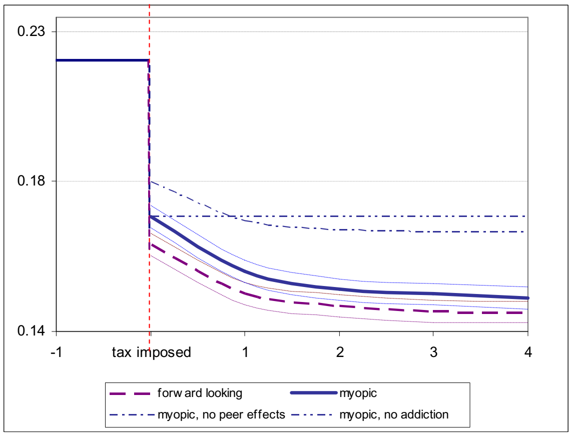
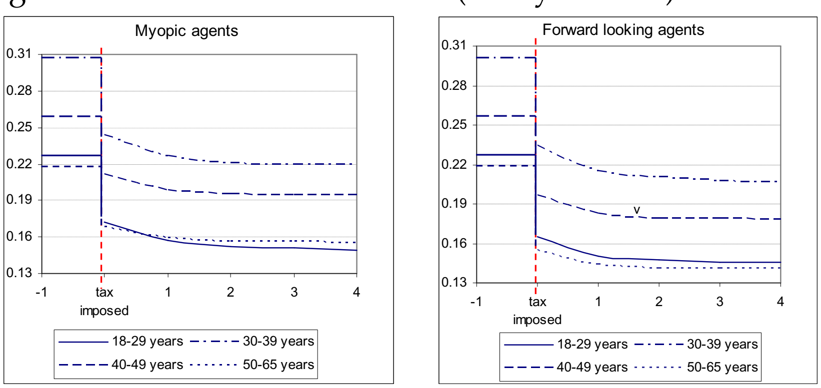


Figure 3 below illustrates the simulated effect of an increase in price for myopic and forward-looking agents in different age strata. Overall, five years after the introduction of a price-raising tax, the proportion of heavy drinkers will decrease by one-third. The effect is higher for younger generations because of the non-trivial social multiplier.

In the model with forward-looking assumptions on agent behavior, the predicted magnitude of change in the proportion of heavy drinkers is 1.2 times higher (although the difference in response between myopic and forward-looking models is not significant). The difference in the effect of a price-raising tax on different age strata is not large, because of smaller differences in estimated peer effects.

Figure 3. Effect of a 50% tax on Pr(heavy drinker) in different age cohorts.



In my second experiment, I model the effect of a change in vodka price on mortality rates.

To do this I estimate the effect of heavy drinking on death rates using the hazard specification

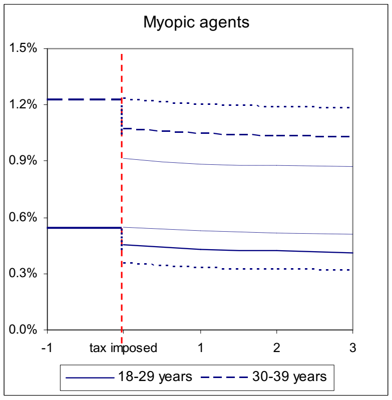
$$\lambda(t, x) = exp(x\beta)\lambda_0(t)$$

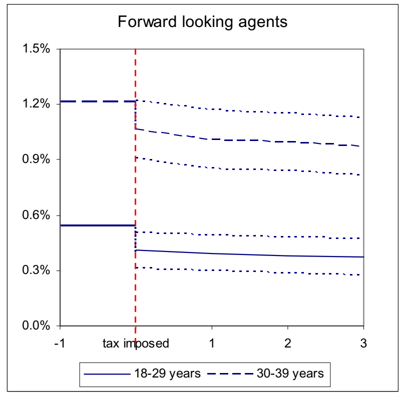
where $\lambda_0(t)$ is the baseline hazard, common for all units of population. I use a semi-parametric Cox specification of baseline hazard. Explanatory variables includes I(heavy drinker), I(smokes), log of family income, I(deceases), weight, current work status, and educational level. I allow heavy drinking to have a heterogeneous (by age stratum) effect on hazard of death. Younger males are more likely to engage in hazardous drinking, which increases hazard rates. For younger people, other factors that affect hazard of death – such as chronic diseases – play a smaller role, and so the relative importance of heavy drinking as a factor of mortality is high.

Results of the estimation are presented in Table 8. The effect of heavy drinking is highly heterogeneous by age. The hazard of death for heavy drinkers age 18-29 is 7.4 times higher than for other males of the same age. The hazard of death for heavy drinkers in age 30-39 is 4.5 times higher. There is no difference between hazard rates for heavy drinkers and non-heavy drinkers age 40-65. It is worth noting that these estimations are done for a relatively-short period of 12 years, and so do not capture in account very long run consequences of alcohol consumption.

Figure 4 shows the simulated effect of increasing the price of alcohol on mortality rates for males of the youngest age strata. The simulated effect of introducing a 50 percent tax is a decrease in mortality rates by one-fourth (from 0.55% to 0.4%) for males age 18-29 years, and by one-fifth (from 1.23% to 1.02%) for males age 30-39 years. There is no effect on the mortality of males of older ages. In other words, a 50 percent increase in the price of vodka would save 40,000 (male) lives annually.

Figure 4. Effect of 50% tax on mortality rates.





In my final experiment, I model the effect of tax policy on consumer welfare.

In both the forward-looking and myopic models presented above, agents have bounded rationality: they do not take into account the effect of heavy drinking on hazard of death.[24] Within these models, tax corrects a negative externality that appears from the bounded rationality of agents. The welfare effect of the 50 % tax is as follows. The tax results in a 30% loss in consumer surplus. At the same time, the tax saves 40,000 young male lives annually, which is 0.055% of the working-age population. The rough estimation of the value of their lives is the present value of the GDP that they generate. With time discount $\beta = 0.9$ value of saved lives equals to 0.55% of GDP, which is more than the size of the whole alcohol industry in Russia (0.48% of GDP). This speculative calculation suggests that a 50% tax is actually likely to be smaller than optimal one.[25]

Besides, , my model, under certain assumptions of utilities, implies that the effect of a vodka tax on consumer surplus would be positive even for fully-rational agents, forward-looking agents who take into account the hazard of death associated with heavy drinking. The model I describe in the main body of my paper implies that peer effects and the effect of habits are positive: all other things being constant, an agent has higher utility if he or she drank within the previous period and if he or she has peers that are heavy drinkers. These forces, however, can equally run an agent's utility to the negative. First, quitting

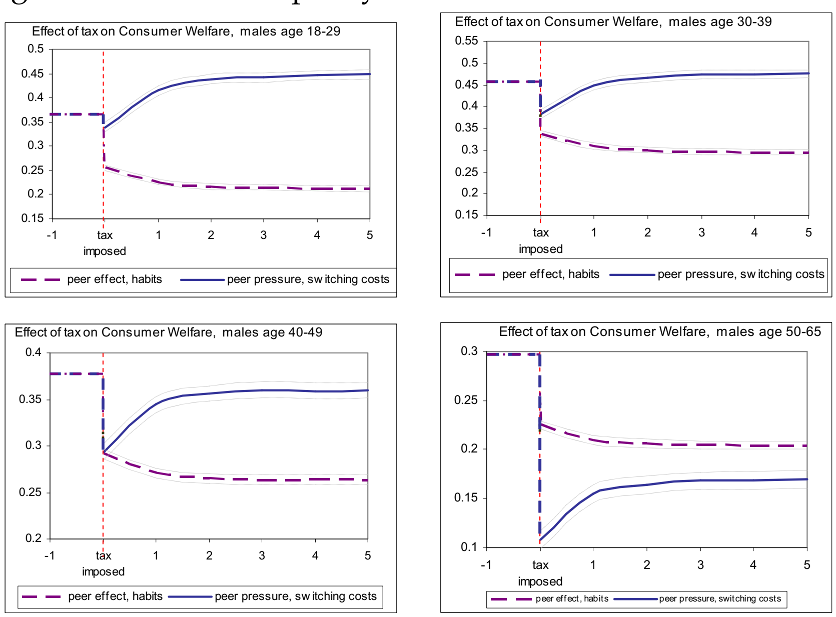
[24]I analyze the model where agents do take in account the effect of drinking on hazard of death in the appendix (table A2, column 2). Results are similar to those of forward looking model in main body of text (with slightly lower magnitude).

[25]My model does not take into account that the tax almost certainly saves other lives (children, females, the elderly), decreases crimes committed under alcohol intoxication, decreases car accidents, and so on.

heavy drinking is costly. Second, an agent who decides not to drink may suffer from the fact that peers are drinking – the agent may experience peer pressure, or agent may suffer if no peer wishes to participate in alternative (to drinking) activities, such as playing soccer or doing other sports.[26]Thus, in the Robustness section I find that peer decisions matter for an agent if he or she decide to do physical training. These alternative assumptions on utilities, although barely distinguishable from the data, have different implications for the analysis of consumer welfare.[27] In this case, case, a 50% tax on vodka results in an increase in the consumer welfare of young males below age 40.[28]

Figure 5 below illustrates this point.

Figure 5. Effect of tax policy on Consumer Welfare.



The final point I want to discuss is my finding that estimations of utilities and response

[26]In this case, an agent's per-period choice specific utilities are as follows:
$\pi_{it}(0) = -\delta \overline{I(a_j = 1|S_{i,-i,t})} - \gamma a_{i,t-1}$, $\pi_{it}(1) = \Gamma' D_{it} + \Upsilon' G_{-it} + \rho_{mt}$

[27]In "myopic" case peer effect and peer pressure jointly are not identified. One can identify only difference between them. In "forward-looking" case they are identified under additional assumptions. See proof of identification results in the appendix (Proof A3). In appendix I provide results of estimation for the following model: $\pi_{it}(0) = \delta \overline{\sigma(a_j = 1|S_{i,-i,t})} + \gamma a_{i,t-1}$, $\pi_{it}(1) = \alpha \overline{\sigma(a_j = 1|S_{i,-i,t})}$. Point estimates of δ, γ and α are -1.373, -1.141, 0.114 correspondingly (see Table A9b).

[28]Determining this optimal tax rate is a question for my future research.

functions, although different, do not differ dramatically in the myopic and forward-looking models. A possible explanation of this phenomenon is as follows. During the lengthy period in my analysis, Russia was in period of transition. This time people were uncertain about the future, and in particular about the realization of state variables such as future alcohol prices, future career, and income. In the context of my model, this may imply that agent expectations about future Value function are noisy, possibly not correlating with current state variables or having a strong effect on agent decision. In this case, even if in reality agents are forward-looking, an estimated "myopic" indirect utility may be a good enough approximation of the choice-specific Value function. Table A2 in Appendix illustrates this point. My data implies that in this case agents should expect a significant mean reversion in price movement. According to column 2 of Table A2, a 10% change in price today leads to only a 4% change in the expected price next year.

In my last experiment, I calculate the response to price change in the case where the government can credibly commit that the new (increased) price will not decrease in the future, and then I correspondingly change the agent expectations regarding price movement (see calculation in Appendix). My calculations imply that in this case price elasticity is 1.73 times higher than in the myopic case.

1.7 Robustness check

In this section I provide several robustness checks for my results.

1.7.1 Reduced-form elasticity estimates

Table A3 in the appendix presents reduced-form elasticity estimates from linear 2SLS regression.

$I(heavy\ drinker)_{it} = \alpha + \theta log(vodka\ price)_{mt} + \Gamma' D_{it} + \rho_t + e_{it}$

The price of vodka is instrumented by the same set of regulatory variables described above. Results are consistent with my estimates: reduced-form elasticity is 1.5 times higher than the own-price elasticity from my model, and represents the cumulative effect of own-price elasticity and the social multiplier.

1.7.2 Linear in means peer effect

In this section I provide a robustness check for my estimates of peer effects on the two younger age groups.

The results of my estimations can be contaminated if (i) peers have the same with agent unobservable shocks that affect their choice, and (ii) these unobservable shocks are independent of the set of peers demographic characteristics (see Manski, 1993).

I check the validity of my results using a non-structural, linear in means assumption for peer effects. The main regression specification is the following:

$$I_{it}(heavy\ drinker) = \sum_k \delta_k I(age\ strata = k)\overline{I(heavy\ drinker)} + \gamma I_{it-1}(heavy\ drinker) + \Gamma' D_{it} + \Upsilon' G_{-it} + \rho_{mt} + e_{it}$$

where $\overline{I(heavy\ drinker)}$ is instrumented by average (across peers) demographic characteristics.[29]

Table A4 the appendix presents IV regression results, as well as the results of different robustness checks. After correcting for the difference in the magnitude of coefficients of the logit and linear probability models, the results have the same magnitude as the myopic model.[30]

First, I present estimates of peer effects using average peer demographic characteristics as instruments. I estimate the model using the entire sample and also separately for different age strata, and for sub-samples without the two regions with a Muslim majority (the Tatarstan and Karachaevo-Cherkessk republics). I verify the robustness of my results

[29] One can show that, under the assumption that beliefs are linear, the structural model I describe in the main body of this paper can be rewritten as a 2SLS regression with average peer demographics used as instruments. To simplify exposition of material, I do not follow structural specification. Within this structural framework, every particular set of instruments potentially changes the model itself. For example, I should add additional game with fathers to the model if I wanted use paternal demographics as instrumental variables.

[30] To compare coefficients in the logit model (Table 5) with those in the linear probability model (Table A4) one need to multiply coefficients in Table A4 on 5.3. To compare marginal effects of LPM and logit regression, one need to divide coefficients in LPM on $p(1-p)$, where p is the probability of being a heavy drinker. In our case $(p(1-p))^{-1} = 5.3$.

by including different sets of fixed effects. Results are similar to those elsewhere in this paper.

I then check the robustness of my results by using the demographic characteristics of the fathers of peers, rather than of the peers themselves, as instruments in my regression. The fathers of peers likely do not face shocks in common with the agent. Finally, I verify the robustness of my results by estimating IV regression on only a sub-sample of respondents who just returned from military service. These people are likely not to face shocks common to their peers. All estimates have the same magnitude, and most of them are statistically significant.

I also employ alternative measures of alcohol-consumption frequency as a measure of alcohol consumption. I use a dummy (who drinks two-or-more times per week, so is in the top 21% of drinkers) as an indicator for a heavy drinker, from which I get similar results with a slightly lower magnitude (see Table A4 in the appendix). In addition, I check the model by applying a similar strategy to tea, coffee, and cigarette consumption, and to hours of physical training. I find no evidence that peers affect either tea, or coffee consumption. At the same time, I find a positive and statistically-significant (for younger groups) peer effect on the personal decision to undertake physical training (see Table A5 in the appendix). The effect of peers on smoking is marginally significant for two age strata.

1.7.3 Robustness of dynamic model assumptions

First, I verify the robustness of the results of the dynamic model under different normalizations of utility: in contrast to the myopic case, the dynamic model's estimator of parameters depends on the chosen normalization. I normalize the utility of heavy drinking to be 0. Results qualitatively are the same, with slightly higher own price elasticity , and a slightly lower magnitude of peer effects (see table A6 in the appendix). In addition, I check the results of the model by allowing all parameters of utilities to vary by age cohort. Utility estimates are similar to those described above (see Table A6 in the appendix).

Second, I did not model that agents probably correctly estimate their hazard of death, and so I now take this into account. I verify the robustness of results after accounting for this factor. In this robustness experiment, an agent has discounting factor $\beta\lambda(t, s)$, where hazard rates depends on state variables, and also on an agent's decision about heavy drinking. Results of this estimation are presented in Table A6 in in the Appendix. Again, utility parameters do not differ from those shown above, because actual hazard of death is very small, especially for young generation.

Finally, I re-estimate the model under the assumption that unobserved utility $e_{it}(1)$ has a uniform (rather than logistic) distribution. The evaluation of moment equations that I use to estimate utility parameters relies largely on the functional form of logistic distribution. To check the robustness of my results against different distributional assumptions, I re-estimate the model with the assumption that $e_{it}(1)$ has U[-1,0] distribution, so that the moment condition can be rewritten in the following way (for the derivation of moment conditions, see Proof A2 in the appendix):

$E[V_{it}(0,s_t) - \beta V_{it+1}(0,s_{t+1}) + \sigma_{it}(1) + \beta\sigma^2_{it+1}(1) + \pi_{it}(a_{-it},1,s_t,\theta)|a_{it}=1,s_t)] = 0$

$E[V_{it}(0,s_t) - \beta V_{it+1}(0,s_{t+1}) + \beta\sigma^2_{it+1}(1)|a_{it}=0,s_t] = 0$

$E(I(a_{it}=k)|s_t) = \sigma_{it}(k|s_t)$, $k \in \{0,1\}$

Table A6 in the appendix presents the results of estimations for both myopic and forward-looking agents. Again, results qualitatively are similar, although in this specification, the price elasticity of forward-looking agents is twice as high as that for myopic agents.

Finally, I estimate the primary specification of the dynamic model separately for every stratum. Results are presented in Table A7 in the appendix. The magnitude of peer effects is slightly lower in this case.

1.7.4 Habits versus unobserved heterogeneity

To provide evidence that the observable correlation between current and lagged level of consumption is driven not by only individual heterogeneity, but also by habit formation, I estimate an instrumental variable regression:

$$I_{it}(heavy\quad drinker) = \alpha + \gamma I_{it-1}(heavy\quad drinker) + \Gamma' D_{it} + \rho_i + \delta_t + e_{it}$$

I use personal demographic characteristics (including current health status) to control for observed individual heterogeneity, and individual fixed effects to control for unobserved heterogeneity. I use lagged health status as an instrument for lagged $I(heavy\ drinker)$. Results of regression are presented in Table A8 in Appendix. Table A8 shows results of regressions with lagged $I(heavy\ drinker)$ as well as results of regressions with average across two and three lags of $I(heavy\ drinker)$. Regression results suggest that habits are important, with the same magnitude as elsewhere in my paper.

1.7.5 Extension

In this section, I provide an informal toy test of which model, myopic or forward-looking, does the better job of explaining my data.

To start, it is worth noting that the seminal result of Rust (1994) states that in general, set-up cannot identify the discounting parameter. One must impose a strong parametric restrictions in order to obtain identification from the model. Therefore, this informal test should be treated at most as only suggestive. In main text of this paper, I use a sequential procedure of estimation for my parameters, which provides little guidance regarding β is better in describing my data. To provide an informal test I first simplify my model, and then use maximum likelihood with the nested fixed-point estimation algorithm described by Rust (1987) instead of the sequential algorithm described above.

In my toy model I assume that agent utility depends on a simplified model with only two variables - habits (lag of I(heavy drinker)) and beliefs about peer actions, $\overline{\hat{\sigma}(a_j = 1|S_{i,-i,t})}$. Table A9 in the appendix shows the level of log likelihood functions, as well as estimated peer effects and the effect of habit for different age strata. Log likelihood for both models is almost the same, with a slightly-higher likelihood in the myopic model for young generations, and a slightly-higher likelihood in the forward-looking model for the oldest generation.

1.8 Conclusion

Over the past twenty years, the life expectancy of male Russian citizens has fallen by more than five years, and the mortality rate has increased by fifty percent. Now, male life expectancy in Russia is only 60 years, below that in Bangladesh, Yemen, and North Korea. Heavy alcohol consumption is widely agreed to be the main cause of this change.

In this paper, I present a structural model of heavy drinking behavior that accounts for the presence of peer effects and habit formation, and with forward-looking assumptions on agent behavior, in order to quantify the effect of public policy (specifically, higher taxation) on the number of heavy drinkers and on mortality rates

First, I find that peers play a significant role in the decision-making of Russian males below age 40. Second, I find that the probability of being a heavy drinker is (relatively) elastic with respect to the price of alcohol. Finally, I find that the assumption that agents are forward-looking gives me higher estimates of price elasticity (although the difference is insignificant).

To illustrate this finding, I simulate the effect on heavy drinkers of increasing the price of vodka by 50%. The myopic model predicts that five years after introducing a price-raising tax, the proportion of heavy drinkers will decrease by roughly one-third – from 25 to 18 percentage points. The effect is higher for young generations because of the non-trivial effect of the social multiplier. This cumulative effect can be decomposed in following way: own one-period price elasticity predicts a drop in the proportion of heavy drinkers by roughly 4.5 percentage points, from 25 to 20.5 percent. In addition, peer effects and habit formation, and a forward-looking assumption, increase the estimated price elasticity by 1.9 times for younger generations, and by about 1.4 times for the older generation. In a model with forward-looking agents, the effect of a change in price is higher by roughly 20 percent. With this established, I simulate the effect on mortality rates of this increase in the price of alcohol. I find significant age heterogeneity in the effect of heavy drinking on the hazard of death: the hazard is much stronger for younger generations. The simulated effect of introducing a 50% tax leads to a decrease in mortality rates by one-fourth for males age 18-29 years, and by one-fifth for males age 30-39 years (with little effect on the mortality of males of older ages). In terms of actual numbers, a 50% tax on the price of vodka will save 40,000 (male) lives annually, or 1% of young male adult lives in six years

1.9 Tables

Table 3. Summary statistics.

Variable	Obs	Mean	Std. Dev.	Min	Max
Panel data (males)					
I(Drunk more than 150 gr last month)	41261	0.285	0.451	0	1
Log(family income)	41395	2.681	3.848	-10.37	8.79
Age	41395	38.77	13.04	18	65
Age squared	41395	1674	1064	324	4225
I(deceases)	41379	0.137	0.343	0	1
I(big family)	41395	0.485	0.500	0	1
Lag I(heavy drinker)	32515	0.284	0.451	0	1
Lag I(Smokes)	32530	0.651	0.477	0	1
I(works)	40734	0.713	0.452	0	1
I(college degree)	41391	0.429	0.495	0	1
I(Muslim)	41395	0.088	0.283	0	1
Weight	37956	75.87	13.25	35	250
I(big family)	41395	.455	.498	0	1
Liters of pure alcohol drunk last month	41261	0.114	0.143	0	2.69
I(physical training)	41395	0.137	0.344	0	1
I(drink tea)	22104	0.966	0.181	0	1
I(drink coffee)	22098	0.698	0.459	0	1
Survival regression data					
Death cases, total population	25697	0.058	0.226	0	1
Death cases, male, >17 years	10894	0.078	0.259	0	1
Drunk more than 150 gr last month	10895	0.250	0.433	0	1
Smokes	10900	0.701	0.458	0	1
Health evaluation (5 = good, 1 = bad)	10881	2.690	0.648	1	5
Married	10307	0.645	0.479	0	1
University education	10900	0.588	0.492	0	1
Weight	10627	74. 78	12.65	36	215

Table 4. Distribution of # of peers in peer groups.

# of peers	(Peer group)-level data			Individual - level data		
in peer group	Freq.	Percent	Cum. %	Freq.	Percent	Cum. %
2	3,373	37.98	37.98	6,746	18	17.71
3	2,383	26.83	64.81	7,149	19	36.48
4	1,253	14.11	78.92	5,012	13	49.64
5	653	7.35	86.27	3,265	8.57	58.21
6	326	3.67	89.94	1,956	5.14	63.35
7	174	1.96	91.9	1,218	3.2	66.55
8	129	1.45	93.36	1,032	2.71	69.26
9	66	0.74	94.1	594	1.56	70.82
10	46	0.52	94.62	460	1.21	72.02
11	57	0.64	95.26	627	1.65	73.67
12	37	0.42	95.68	444	1.17	74.84
13	28	0.32	95.99	364	0.96	75.79
14	28	0.32	96.31	392	1.03	76.82
15	22	0.25	96.55	330	0.87	77.69
16	31	0.35	96.9	496	1.3	78.99
17	19	0.21	97.12	323	0.85	79.84
18	17	0.19	97.31	306	0.8	80.64
19	17	0.19	97.5	323	0.85	81.49
20 and more	222	2.5	100	7,050	18.51	100
Total	8,881	100		38,087	100	

Note: 3642 peers groups that contain 1 peer are excluded

Table 5. Agent's utility parameters.

	$\beta = 0$	$\beta = 0.9$
Peers effect, $\widehat{\delta}$: age 18-29	1.355	0.932
	*[0.273]****	*[0.239]****
$\widehat{\delta}$, age 30-39	0.688	0.456
	*[0.211]****	*[0.183]****
$\widehat{\delta}$, age 40-49	0.039	0.128
	[0.255]	*[0.201]*
$\widehat{\delta}$ age 50-59	0.090	0.214
	[0.244]	*[0.234]*
Habit: Lag I(heavy drinker)	1.270	1.234
	*[0.038]****	*[0.032]****
Log (family income)	0.004	0.003
	[0.012]	*[0.009]*
Age	0.120	0.079
	*[0.026]****	*[0.021]****
Age squired	-0.001	-0.001
	*[0.0004]***	*[0.0003]****
Weight	0.007	0.005
	*[0.001]****	*[0.001]****
I(deceases)	-0.096	-0.093
	*[0.062]**	*[0.042]***
I(big family)	-0.002	-0.010
	[0.038]	*[0.024]*
Lag I(smokes)	0.505	0.429
	*[0.046]****	*[0.029]****
I(work)	-0.241	-0.222
	*[0.051]****	*[0.040]****
I(college degree)	-0.147	-0.127
	*[0.062]***	*[0.042]****
I(Muslim)	-0.263	-0.186
	*[0.102]****	*[0.070]****
*municipality*year FE*	*Yes*	*Yes*
Peers mean characteristics	*Yes*	*Yes*
Observations	25042	25042

Note: Bootstrapped standard errors in brackets

* significant at 10%;** significant at 5%;*** significant at 1%

Table 6. Marginal utility of peers

	Myopic agents	Forward looking agents	
	MU ($du/d\overline{\sigma(a_j = 1)}$)	MV ($dV/d\overline{\sigma(a_j = 1)}$)	MU ($du/d\overline{\sigma(a_j = 1)}$)
age 18-29	1.355	0.961	0.932
age 30-39	0.688	0.609	0.456
age 40-49	0.039	0.073	0.128
age 50-59	0.09	0.18	0.214

Table 7. Estimates of price elasticity

	Myopic agents	Forward looking agents		First stage
	MU (du/dlogP)	MV (dV/dlogP)	MU (du/dlogP)	log(vodka price)
log(vodka price)	-0.82	-0.968	-0.68	
	[0.336]**	[0.453]**	[0.356]*	
I(excise)				0.137
				[0.050]***
I(tax-producers)				0.135
				[0.039]***
I(tax-retail)				0.117
				[0.037]***
Constant	-0.245	1.324	1.196	0.4
	[0.174]	[0.224]***	[0.175]***	[0.028]***
Year FE	yes	yes	yes	yes
Observations	25042	25042	25042	25042
F-stat (clustered errors)	9.75	9.75	9.75	9.75
F-stat	724	724	724	724
J-test, p-value	0.97	0.61	0.45	

Note: Robust standard errors, clustered at regionXyear level in brackets
* significant at 10%; ** significant at 5%; *** significant at 1%

Table 8. Mortality and heavy drinking.

	all males		all males	
	coefficient	hazard ratio	coefficient	hazard ratio
I(heavy drinker) age 18-29	1.993	7.337		
	[0.519]***			
I(heavy drinker) age 30-39	1.541	4.669		
	[0.357]***			
I(heavy drinker) age 40-49	-0.031	0.969		
	[0.324]			
I(heavy drinker) age 50-64	0.108	1.114		
	[0.243]			
I(heavy drinker), age18-64			0.39	1.477
			[0.147]***	
Log (family income)	-0.322	0.725	-0.321	0.725
	[0.016]***		[0.016]***	
I(deceases)	0.34	1.405	0.365	1.441
	[0.128]***		[0.128]***	
Lag I(smokes)	0.561	1.527	0.563	1.756
	[0.099]***		[0.099]***	
I(college degree)	-1.504	0.222	-1.53	0.217
	[0.228]***		[0.228]***	
Weight	-0.002	0.998	-0.001	0.999
	[0.003]		[0.003]	
I(work)	-0.299	0.742	-0.29	0.748
	[0.134]**		[0.133]**	
Observations	7735		7735	

Standard errors in brackets; * significant at 10%** significant at 5%;*** significant at 1%

1.10 Appendix

Figure A1. Alcohol consumption: age profile

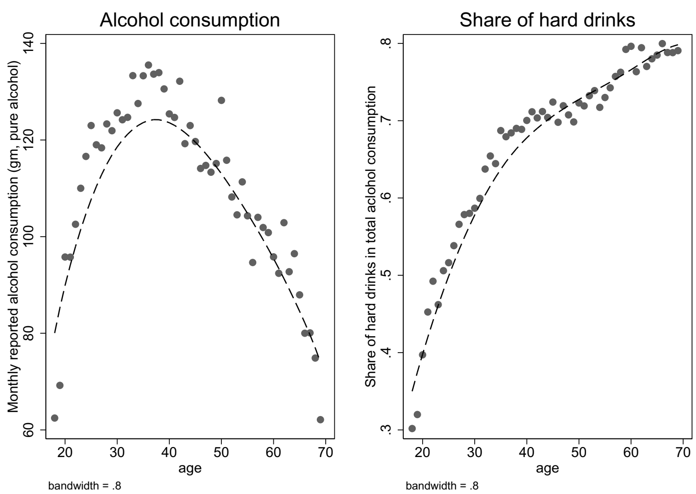
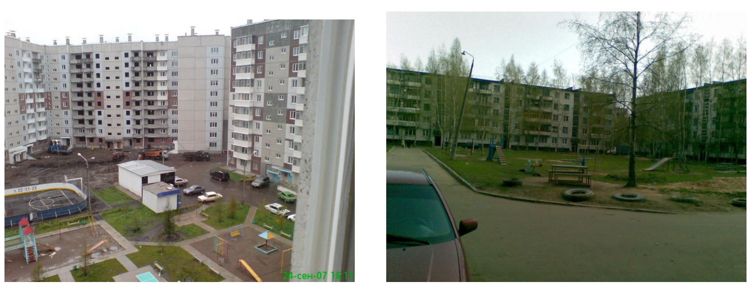


Figure A2. Dvors in Russia.



Source: www.miel.ru, www.su155.ru, www.yandex.ru

Table A1. Consumption of goods and birthday.

	I(drink vodka)	I(smokes)	I(drink tea)	I(drink coffee)
All peers				
$\frac{\sum_{peers} I(birthday)}{(N-1)}$	0.042	-0.029	-0.01	-0.013
	[0.015]***	[0.015]*	[0.007]	[0.019]
$I(birthday)$	0.028	0.025	-0.002	0.008
	[0.009]***	[0.009]***	[0.005]	[0.012]
Year*month FE	Yes	Yes	Yes	Yes
Observations	39534	39515	20450	20444
Without household members				
$\frac{\sum_{peers} I(birthday)}{(N-1)}$	0.039	-0.028	-0.008	-0.015
	[0.015]**	[0.015]*	[0.007]	[0.019]
$I(birthday)$	0.028	0.026	-0.002	0.007
	[0.009]***	[0.009]***	[0.005]	[0.012]
Year*month FE	Yes	Yes	Yes	Yes
Observations	35995	35977	18253	18247

* significant at 10%; ** significant at 5%; *** significant at 1%

Table A2. Lag (Log vodka price) is not good predictor for current Log(Vodka Price)

	$log(vodka\ price)_t$		$log(vodka\ price)_t - log(vodka\ price)_{t-1}$
$log(vodka\ price)_{t-1}$	0.007	0.392	
	[0.005]	[0.039]***	
$log(vodka\ price)_{t-1} - log(vodka\ price)_{t-2}$			-0.419
			[0.052]***
Year FE	YES	NO	NO
Region FE	YES	NO	NO
Observations	36307	36307	28403
R-squared		0.18	0.19

Robust standard errors clustered at municipalityXyear level are in brackets
* significant at 10%; ** significant at 5%; ***significant at 1%

Table A3a. Reduced form elasticity estimates. Individual-level 2SLS regression.

	1st stage : log(vodka price)	2nd stage : I(heavy drinker)
log(vodka price)		-0.338
		[0.133]**
I(excise)	0.051	
	[0.018]***	
I(tax, producers)	0.084	
	[0.016]***	
I(tax, retail)	0.034	
	[0.016]**	
Log (family income)	0.022	0.007
	[0.002]***	[0.003]**
Age	0	0.013
	[0.001]	[0.001]***
Weight	-0.001	0.001
	[0.000]***	[0.000]***
I(deceases)	0.009	-0.013
	[0.007]	[0.009]
I(big family)	-0.033	-0.029
	[0.010]***	[0.010]***
Lag I(smokes)	0.026	0.127
	[0.007]***	[0.009]***
I(work)	0.018	-0.017
	[0.011]*	[0.009]*
I(college degree)	0.028	-0.021
	[0.010]***	[0.011]*
I(Muslim)	-0.31	-0.215
	[0.078]***	[0.054]***
Year FE	YES	YES
Constant	0.521	0.032
	[0.034]***	[0.067]
Observations	33193	33103
R-squared	0.31	
F-test		154.62
F-test (robust st.errors)		9.58
J-test, p-val		0.12

Standard errors clustered at neighborhood level in brackets

* significant at 10%; ** significant at 5%; *** significant at 1%

Table A3b. Elasticity of Vodka consumption.

	Neigborhood-level, OLS
	Log(vodka consumption)
log(vodka price)	-0.242
	[0.024]***
Constant	5.053
	[0.014]***
Observations	4850
R-squared	0.02

Standard errors clustered at neighborhood level in brackets
** significant at 5%; *** significant at 1%

Table A4. Linear in means peer effects. Robustness checks under different specification.

I(heavy drinker)						
			age 18-65			
	IV-1	IV-2	IV-3	IV-4	OLS-1	OLS-2
Peers effect, $\hat{\delta}$:						
age 18-29	0.264	0.297	0.242	0.255	0.193	0.119
	[0.04]***	[0.05]***	[0.04]***	[0.09]***	[0.03]***	[0.02]***
age 30-39	0.194	0.218	0.181	0.16	0.17	0.111
	[0.03]***	[0.04]***	[0.03]***	[0.065]**	[0.02]***	[0.01]***
age 40-49	0.063	0.089	0.053	0.063	0.121	0.057
	[0.030]**	[0.037]**	[0.031]*	[0.059]	[0.02]***	[0.01]***
age 50-65	-0.005	0.015	-0.022	0.009	0.088	0.03
	[0.033]	[0.041]	[0.033]	[0.056]	[0.02]***	[0.016]*
Demographics	Yes	Yes	Yes	Yes	Yes	Yes
Munic*year FE	Yes	Yes	Yes			Yes
Individual FE				Yes		
Year FE				Yes		
Muslim region excluded?			Yes			
Instruments	Peers 1	Peers 2	Peers 1	Peers 1		
Observations	29554	29554	27400	29554	29923	29923
F-test	79.9	36.29	72.02	17.02		
J-test, p-value	0.22	0.13	0.26	0.02		
			age18-29			
	IV-5	IV-6	IV-7	IV-8		
Peers effect, $\hat{\delta}$:	0.211	0.197	0.225	0.359		
	[0.09]**	[0.136]	[0.14]*	[0.180]**		
Demographics	Yes	Yes	Yes	Yes		
Munic*year FE	Yes	Yes	Yes	Yes		
Just came from military service?				Yes		
Instruments	Peers 1	Fathers 1	Fathers 2	Peers 1		
Observations	7750	8152	8152	149		
F-test	34.24	16.52	28.97	6.85		
J-test, p-value	0.06	0.4	0.86	0.17		

* significant at 10%; ** significant at 5%, *** significant at 1%. St. errors clustered clustered at municipality*year in brackets. Instrument set: Peers: (1) average demographics (2) average demographics without lag I(heavy drinker)
Peer fathers: (1) average demographics (2) average demographics-subset

Table A5. Linear in means peer effects. Peer effects for different products/activities.

	Peer effect			
year	age 18-29	age 30-39	age 40-49	age 50-64
I(drink tea)	-0.016	-0.016	-0.003	-0.006
I(drink coffee)	0.02	0.055	0.055	0.057*
I(smoking)	0.016	0.021*	0.014	0.018*
I(physical training)	0.14***	0.127***	0.141***	0.073
I(Drink 2 days/week)	0.195***	0.118***	-0.014	0.009

* significant at 10%; ** significant at 5%, *** significant at 1%

Table A6. Forward looking agents. Point estimates of utility parameters. Different robustness checks.

	Utility parameters			
Utility parameters:				
Peers effect, $\widehat{\delta}$:				
age 18-29	0.644	0.948	0.198	0.358
age 30-39	0.201	0.49	0.132	0.321
age 40-49	-0.031	0.152	0.014	0.052
age 50-59	0.051	0.253	-0.008	0.019
Habit: lag I(heavy drinker)	1.34	1.23	0.262	0.261
Elasticity:				
log(vodka price)	-1.069	-0.858	-0.157	-0.344
Normalization	U(drink)=0	U(not drink)=0	U(not drink)=0	U(not drink)=0
Forward looking?	Yes	Yes	Myopic	Yes
Distribution of private shocks	Logistic	Logistic	Uniform[-1.0]	Uniform[-1.0]
Discounted by hazard of death	No	Yes	No	No
Demographics	Yes	Yes	Yes	Yes
Municipality*year FE	Yes	Yes	Yes	Yes
Peers mean characteristics	Yes	Yes	Yes	Yes

Note: In first column I revert signs of coefficients on opposite.

Table A7. Point estimates of utilities for forward looking agents. Separate regression for every age strata.

	age: 18-29	age: 30-39	age: 40-49	age: 50-65
Peer effects, $\widehat{\delta}$	0.793	0.558	0.001	0.143
Havit: lag I(heavy drinker)	1.074	1.338	1.38	1.441

Table A8. Habits versus unobserved heterogeneity.

	Y					
	log(1+alcohol consumption)			I(heavy drinker)		
Mean(Lag Y, LagLag Y, LagLagLag Y)	0.423			0.666		
	[0.207]**			[0.323]**		
Mean(Lag Y, LagLag Y)		0.472			0.901	
		[0.233]**			[0.462]*	
Lag Y			0.313			0.604
			[0.235]			[0.497]
I(health problems)	-0.006	-0.005	-0.007	-0.01	-0.001	-0.009
	[0.002]**	[0.003]*	[0.003]***	[0.010]	[0.015]	[0.013]
Individual FE	Yes	Yes	Yes	Yes	Yes	Yes
Demographics	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	33812	33810	33735	33814	33814	33814
Number of individuals	5814	5814	5814	5814	5814	5814
F-test for instruments (with robust se)	19	14.9	14.78	9.77	6.02	4.82

Note: Instruments are Mean(Lag X, LagLag X, LagLagLag X), Mean(Lag X, LagLag X), and Lag X correspondingly, where X stands for I(health problems).

* significant at 10%; ** significant at 5%; *** significant at 1%

Robust standard errors, clustered on individual level, are in brackets

Table A9a. Log likelihoods for different betas. Rust approach.

	age 18-29	age 30-39	age 40-49	age 50-65
β=0				
Lag I(heavy drinker)	1.407	1.42	1.425	1.466
Peer effect	1.399	0.98	0.866	0.757
Log Likelihood	-3555.43	-3723.54	-3877.12	-3591.9
β=0.9				
Lag I(heavy drinker)	1.432	1.42	1.425	1.468
Peer effect	1.257	0.767	0.673	0.596
Log Likelihood	-3556.5	-3723.52	-3877.1	-3591.34

Table A9b. Peer effects vs Peer pressure. Rust approach.

	age 18-29
β=0.9	
Lag I(heavy drinker), γ	-1.373
Peer effect, α	0.114
Peer pressure, δ	-1.141
Log Likelihood	-3554.9

Note: In this case, an agent's per-period choice specific expected utilities are as follows:
$\pi_{it}(0) = \delta\overline{\sigma(a_j = 1|S_{i,-i,t})} + \gamma a_{i,t-1}$, $\pi_{it}(1) = \alpha\overline{\sigma(a_j = 1|S_{i,-i,t})}$.
$\overline{\hat{\sigma}_{jt}(a_{jt} = 1|S_{i,-i,t})}$ is discretized to set {0.2, 0.4, 0.6. 0.8, 1}.

Price elasticity under commitment on price stability: Calculation

Remember that price enters in into the Value function twice: first in per-period utility, and second in an agent's expectation of future Value functions. Change in Ep_{t+1} then will affect only the second component.

To simplify this analysis, I make two (strong) assumptions on the price transition process and on parametrization of the Value function.

First, I assume that the price-transition process is independent of all other state variables and personal choice of action, and that it follows the AR rule of motion:

$log(p_{i,t+1}) = \rho_0 + \rho_1 log(p_{it}) + \omega_{it}$, where $E(\omega_{it}|p_{it}) = 0$,

Second, I assume the following parameterization of the Value function:

$$V_{it}(S_t, a_{t-1} = j) = \vartheta_j log(p_t) + V_{it}(\{S_t/p_t\}),$$

where $j \in \{0, 1\}$, and $\{S_t/p_t\}$ is set of state variables excluding price.

This parameterization implies that for a given level of habits, the Value function is separable in price and linear in (log) price component.

Under these assumptions,

$$\frac{\partial}{\partial log(p_t)}[E(V_{it+1}(S_{t+1})|1, S_t) - E(V_{it+1}(S_{t+1})|0, S_t)] = (\vartheta_1 - \vartheta_0)\frac{\partial E_p(log(p_{t+1}))}{\partial log(p_t)}$$

Without a commitment on price stability, $\frac{\partial E_p(log(p_{t+1}))}{\partial log(p_t)} = 0.392$ (see Table A2). Once the government can commit that the price will not revert, then $\frac{\partial E_p(log(p_{t+1}))}{\partial log(p_t)} = 1$, and the price elasticity of expectations over future Value functions goes up by 2.5 times (=1/0.392). The result is that the price elasticity of the Value function will not be -0.968, but -1.418:

$$\begin{aligned}\frac{\partial Value\ function}{\partial log(p_t)} = \quad & \frac{\partial}{\partial log(p_t)}[E_{e_{-i}}\pi_{it}(a_{-it}, a_{it} = 1, s_t)] \\ & + \frac{\partial}{\partial log(p_t)}[E(V_{it+1}(S_{t+1})|1, S_t) - E(V_{it+1}(S_{t+1})|0, S_t)] \\ & = -(0.68 + 0.288 * 2.5) = -1.416\end{aligned}$$

This means that in this case price elasticity is 1.73 times higher than in the myopic case.

Proof A1

Derivation of moment conditions, model with forward looking assumption (withβ=0.9). Agent's choice specific value function is

$$V(a_{it}, s_t) = E_{e_{-i}}\pi_{it}(a_{-it}, a_{it}, s_t) + \beta E(V_{it+1}(s_{t+1})|a_{it}, s_t)$$

where $E(V_{it+1}(s_{t+1})|a_{it}, s_{it})$ is ex ante value function (or so called Emax function):

$$V_{it+1}(s_{t+1}) = E_{e_{it+1}}(max_{a_{it+1}}[V(a_{it+1}, s_{t+1})_{it+1} + e_{it+1}(a_{it+1})])$$

To derive moment conditions for my further estimation I will use two well-known relationships. Both of these relationship based on properties of logistic distribution of private utility shock (random utility component).

First relationship, is called Hotz-Miller inversion (see Hotz and Miller, 1993):

$$V(1, s_t)_i - V(0, s_t)_i = log(\sigma_{it}(1)) - log(\sigma_{it}(0))$$

Second equation states relationship between Emax function and choice specific value functions:

$$V(s_t) = log(exp(V(0, s_t)) + exp(V(1, s_t)))$$

Applying these relationships to equation for value function:

$$\begin{aligned} V(a_{it}, s_t) &= \pi_{it}(a_{-it}, a_{it}, s_t, \theta) + \beta E(log(exp(V(0, s_{t+1})) + exp(V(1, s_{t+1}))|a_{it}, s_t) \\ &= \pi_{it}(a_{-it}, a_{it}, s_t, \theta) + \beta E(log(exp(V(0, s_{t+1})) \\ &\quad + exp(V(0, s_{t+1}))\sigma_{it+1}(1)/\sigma_{it+1}(0))|a_{it}, s_t) \\ &= \pi_{it}(a_{-it}, a_{it}, s_t, \theta) + \beta E(V(0, s_{t+1}) - log(\sigma_{it+1}(0))|a_{it}, s_t) \end{aligned}$$

When put $a_{it} = 0$, and $a_{it} = 1$ in equation above I have:

Moment condition on $V_i(0, s_{it})$:

$$V_i(0, s_{it}) = \beta E_{t+1}[log(1 + exp(log(\sigma_{it+1}(1)) - log(\sigma_{it+1}(0)) + V_i(0, s_{it+1})|s_t, a_{it} = 0]$$

Moment condition on $V_i(1, s_{it})$:

$$\begin{aligned} V(1, s)_{it} &= log(\sigma_{it}(1)) - log(\sigma_{it}(0)) + V(0, s)_{it} \\ &= \pi_{it}(a_{-it}, a_{it} = 1, s_t, \theta) + \beta E_{t+1}(V(0, s_{t+1}) - log(\sigma_{it+1}(0))|a_{it} = 1, s_t) \end{aligned}$$

These two equations, together with moment equation on choice probabilities

$$E(I(a_i = k)|s_t) = \sigma_i(k|s_t),\ k \in \{0, 1\}$$

form system of moments I estimated:

$$\begin{gathered} E[\pi_{it}(a_{-it}, a_{it} = 1, s_t, \theta) + V_i(0, s)_{it} - \beta V(0, s_{t+1}) + log(\sigma_{it}(1)) \\ -log(\sigma_{it}(0)) + \beta log(\sigma_{it+1}(0))|a_{it} = 1, s_t)] = 0 \end{gathered}$$

$$\begin{gathered} E[V_i(0, s_t) - \beta V(0, s_{t+1}) + \beta log(\sigma_{it+1}(0))|a_{it} = 0, s_t] = 0 \\ E(I(a_i = k)|s_t) = \sigma_i(k|s_t),\ k \in \{0, 1\} \end{gathered}$$

Proof A2

Derivation of moment conditions with assumption of uniform distribution of unobserved component of utility: $e_{it}(1)$ is distributed uniformly on [-1,0], $e_{it}(0)$ is normalized to 0.

I use the same notation I used in Proof A1. To derive moment conditions for my estimation I will use "uniform" analogs of relationships I discussed in Proof A1:

First lemma establishes relationship between choice probability and choice specific value functions:

Lemma 1

$V(1,s)_{it} - V(0,s)_{it} = \sigma_{it}(1)$

Proof:

$$\begin{aligned} Pr(1) &= Pr(V(1,s)_{it} + e_{it}(1) > V(0,s)_{it} + e_{it}(0)) \\ &= Pr(e_{it}(0) - e_{it}(1) < V(1,s)_{it} - V(0,s)_{it} = V(1,s)_{it} - V(0,s)_{it} \end{aligned}$$

Second lemma states relationship between Emax function and choice specific value functions:

Lemma 2

$V(s) = V(0,s)_{it} + (V(1,s)_{it} - V(0,s)_{it})^2$

Proof:

$$\begin{aligned} V(s) &= E_{e1}(max(V(1,s)_{it} + e_{it}(1), V(0,s)_{it})) \\ &= Pr(V(1,s)_{it} + e_{it}(1) > V(0,s)_{it})[V(1,s)_{it} + E(e_{it}(1)|e_{it}(1) > V(0,s)_{it} - V(1,s)_{it})] \\ &+ Pr(V(1,s)_{it} + e_{it}(1) < V(0,s)_{it})V(0,s)_{it} \\ &= (V(1,s)_{it} - V(0,s)_{it})[V(1,s)_{it} + (V(0,s)_{it} - V(1,s)_{it})/2] \\ &+ (1 - V(1,s)_{it} + V(0,s)_{it})V(0,s)_{it} \\ &= V(0,s)_{it} + (V(1,s)_{it} - V(0,s)_{it})^2/2 \end{aligned}$$

Applying these relationships to equation for value function:

$$\begin{aligned} V(a_{it}, s_t) &= \pi_{it}(a_{-it}, a_{it}, s_t, \theta) + \beta E(Emax|a_{it}, s_t) \\ &= \pi_{it}(a_{-it}, a_{it}, s_t, \theta) + \beta E(V(0, s_{t+1}) + (\sigma_{it+1}(1))^2|a_{it}, s_t)/2 \end{aligned}$$

When put $a_{it} = 0$, and $a_{it} = 1$ in equation above I have:

Moment condition on $V_i(0, s_{it})$:

$$V_i(0, s_{it}) = \beta E_{t+1}((\sigma_{it+1}(1))^2/2 + V_i(0, s_{it+1})|s_t, a_{it} = 0)$$

Moment condition on $V_i(1, s_{it})$:

$$\begin{aligned} V(1, s)_{it} &= \sigma_{it}(1) + V(0, s)_{it} \\ &= \pi_{it}(a_{-it}, a_{it} = 1, s_t, \theta) + \beta E_{t+1}(V_i(0, s_{t+1}) + (\sigma_{it+1}(1))^2/2|a_{it} = 1, s_t) \end{aligned}$$

These two equations, together with moment equation on choice probabilities

$$E(I(a_i = k)|s_t) = \sigma_i(k|s_t),\ k \in \{0, 1\}$$

form system of moments:

$$E[\pi_{it}(a_{-it}, a_{it} = 1, s_t, \theta) + V_i(0, s)_{it} - \beta V_i(0, s_{t+1}) + \sigma_{it}(1) + \beta(\sigma_{it+1}(1))^2/2|a_{it} = 1, s_t)] = 0$$

$$E[V_i(0, s_t) - \beta V_i(0, s_{t+1}) + \beta(\sigma_{it+1}(1))^2/2|a_{it} = 0, s_t] = 0$$

$$E(I(a_i = k)|s_t) = \sigma_i(k|s_t),\ k \in \{0, 1\}$$

Proof A3

Lemma

Let z_{it} be a state variable that enters both in $\pi_{it}(1)$ and in $\pi_{it}(0)$:

$\pi_{it}(0) = \rho_0 z_{it}$

$\pi_{it}(1) = \rho_1 z_{it} + \Gamma' S_{it} + e_{it}(1)$

then

i) In myopic model ρ_0 and ρ_1 are not identifiable

ii) In forward looking model, ρ_0 and ρ_1 are identifiable iff there is no $f(s_t, z_{it})$ such that $f(s_t, z_{it}) - \beta * E[f(s_{t+1}, z_{it+1})|a_{it} = j, s_t, a_{-it}] = \phi_j * z_{it}$ for $j \in \{0, 1\}$

Proof

i) In myopic model agent decides to drink if

$\pi_{it}(1) - \pi_{it}(0) = (\rho_1 - \rho_0)z_{it} + \Gamma' S_{it} + e_{it}(1) > 0$

Then for any number b , pairs(ρ_1, ρ_0) and $(\rho_1 + b, \rho_0 + b)$ are observationally equivalent.

ii)$\Rightarrow$ From the data we know population parameters $\sigma(0)$ and $\sigma(1)$ and operators $E_{t+1}(.|1)$, $E_{t+1}(.|0)$.

In case of forward looking agent's value function is fully characterized by two equations:

$$V(0_{it}, s_t) = \rho_0 z_{it} + \beta E_{t+1}(exp(V(0, s) - log(\sigma(0))|0_{it}, s_t) \quad (1.4)$$

$$V(0_{it}, s_t) + log(\sigma(1)/(\sigma(0)) = \rho_1 z_{it} + \pi_{it}(a_{-it}, a_{it}, s_t, \theta) + \beta E_{t+1}(V(0, s) - log(\sigma(0)))|1, s_t) \quad (1.5)$$

Suppose that exists another pair$V(0_{it}, s_t)', \rho'_j$ for which these two equations hold

Define $\Delta_j = \rho'_j - \rho_j$, $f(s_t, z_{it}) = V(0_{it}, s_t) - V(0_{it}, s_t)'$

Equations above imply

$f(s_t, z_{it}) - \beta * E[f(s_{t+1}, z_{it+1})|a_{it} = j, s_t, z_{it}] = \Delta_j * z_{it}$, so contradiction.

$\Leftarrow$

Assume that $\exists f(s_t, z_{it}) : f(s_t, z_{it}) - \beta * E[f(s_{t+1}, z_{it+1})|a_{it} = j, s_t, a_{it}] = \phi_j * z_{it}$

and let $V(0_{it}, s_t), \rho_j$ is solution of equations above. Then $V(0_{it}, s_t)', \rho'_j$, such as $V(0_{it}, s_t)' = f(s_t, z_{it}) + V(0_{it}, s_t)$, and $\rho'_j = \rho_j + \phi_j$ will be solution of equations (4) and (5).

•

Note: Example where we can not identify ρ_1 and ρ_0.

If there are ϕ_j, such that $E(z_{it+1}|a_{it} = j, s_t) = \zeta + \phi_j * z_{it}$, then we can not identifyρ_0 and ρ_1 simultaneously.

Proof:

Let $V(0_{it}, s_t)' = V(0_{it}, s_t) + z_{it} + \zeta/(1 - \beta)$ and $\rho_j' = \rho_j + 1 - \beta\phi_j$, and we have that equations (4) and (5) above hold for new$V(0_{it}, s_t)', \rho_j'$

Chapter 2

Beer versus Vodka:
Stepping-Stone or Safer Drink, What is More Important?

2.1 Introduction

Russian males are notorious for their hard drinking. The most notable example of the severe consequences of alcohol consumption is the male mortality crisis – male life expectancy in Russia is only 60 years. This is 8 years below the average in the (remaining) BRIC countries, 5 years below the world average, and below even the life expectancy in Bangladesh, Yemen, and North Korea. High alcohol consumption is frequently considered to be the main cause of this (see for example Treisman 2010, Leon et al. 2007, Nemtsov 2002, Bhattacharya et al. 2011, Brainerd and Cutler 2005). Approximately one-third of all deaths in Russia are related to alcohol consumption (see Nemtsov 2002). Most of the burden lies on males of working age: more than half of all deaths in working-age men are accounted for by hazardous drinking (see Leon et al. 2007, Zaridze et al. 2009, and Figure 1 below). Russian authorities at both the federal and regional level have responded to this problem by regulating the alcohol industry; during the past twenty years various federal and regional laws have been imposed – some to good effect, and some not (see Yakovlev, 2006).

In particular, vibrant policy debates have arisen about the 2011 federal law that imposed additional time and money costs for beer production.[1] These debates in Russia parallel current policy discussions on the legalization of marijuana in California. On the one hand, the Russian debates emphasize the potential consequences of consuming light alcohol beverages or light drugs at early ages – that such use may serve as a "stepping-stone" to harder substances, and so can have negative long-run consequences on public health (see "gateway effect" literature, Mills and Noyes 1984, Van Ours 2003, Deza, 2012). On the other hand, light alcohol beverages or light drugs may serve as safer substitutes for harder drinks or drugs, and thus may instead have a direct positive consequences on current public health. Moreover, consumption of light alcohol beverages at early ages may form habits for these goods, and thus prevent a person from consuming harder substances (see "habit formation" literature, Becker and Murphy 1988, Cook and Moor 1995, Heien and Durham 1991, Beenstock and Rahav 2002). All of these effects are well known, but there is a surprising lack of research to analyze the effect of these three forces together. There are strong reasons to analyze them in a bundle, because these forces affect individual behavior simultaneously. My paper fills this gap.

In this paper, I analyze the trade-off between these consumption effects by utilizing micro-level data on the alcohol consumption of Russian males. I find, first, that beer is a more healthy drink compared to hard alcohol beverages: only the consumption of hard beverages affects hazard of death, while beer does not. Second, I find that beer is a substitute for vodka consumption: there is significant positive cross-price elasticity between the prices of vodka and beer. Finally, I find that the stepping-stone effect of beer is small: beer generally creates habits only for future beer consumption. Drinking beer at early ages results in higher beer consumption and higher alcohol intake at older ages, but also results in lower consumption of hard drinks (vodka) compared to those who drink vodka habitually, and even compared to general non-abstainers. Finally, I estimate a multivariate model of consumer choice that allows unobserved heterogeneity in consumer preferences regarding alcohol consumption, and quantify the effect of different government policies on mortality rates, drinking patterns, and consumer welfare. The statistical challenge in estimating models with stepping-stone effects and habits is to separate state dependence from the unobserved heterogeneity of individual tastes (see Heckman 1981, Keane 1997).

[1] Law proposed increase on excise tax and time restriction on retail sales of alcohol.

A higher propensity to consume alcohol results in higher alcohol consumption at both younger and older ages, which creates in an inter-temporal correlation of alcohol consumption that can be incorrectly interpreted as evidence of state dependence – that is, as a habit or stepping-stone effect. In contrast to state dependence, unobserved heterogeneity in tastes does not have policy implication: change in consumption patterns today will not result in changing consumption patterns tomorrow. My model deals with this issue by allowing for individual-level unobserved heterogeneity in consumer preferences regarding alcohol consumption. My simulation shows that a decrease in the price of beer relative to vodka results in a decrease in mortality rates in both short and long-run time horizons. I find that subsidizing beer consumption will increase consumer welfare and (slightly) decrease mortality rates. In contrast, taxation of beer will make the situation worse: mortality rates will increase, and consumer welfare will decrease. This paper is organized as follows. In Sections 2.2 and 2.3, I describe my data and the variables employed in my analysis. Section 2.4 discusses estimations of the elasticity of alcohol consumption using hedonic regressions, Section 2.5 discusses the effect of alcohol consumption on hazard of death, and Section 2.6 discusses the stepping-stone effect. In Section 2.7, I estimate a model of consumer choice among different kinds of alcoholic beverages, and simulate the effect of government policies on the share of heavy drinkers, on consumer welfare, and on mortality rates. Section 2.8 concludes.

2.2 Data

In this study, I utilize data from the Russian Longitudinal Monitoring survey (RLMS)[2]. The RLMS is a nationally-representative annual survey that covers more than 4,000 households (with between 7413 and 9444 individual respondents), starting from 1992. My study

[2]This survey is conducted by the Carolina Population Center at the University of Carolina at Chapel Hill, and by the High School of Economics in Moscow. Official Source name: "Russia Longitudinal Monitoring survey, RLMS-HSE," conducted by Higher School of Economics and ZAO "Demoscope" together with Carolina Population Center, University of North Carolina at Chapel Hill and the Institute of Sociology RAS. (RLMS-HSE web sites: http://www.cpc.unc.edu/projects/rlms-hse, http://www.hse.ru/org/hse/rlms).

utilizes rounds 5 through 16 of RLMS.[3] Time span of my study is from 1994 to 2007, except 1997 and 1999. The data cover 33 regions – 31 oblasts (krays, republics), plus Moscow and St. Petersburg. Two of the regions are Muslim. Seventy-five percent of respondents live in an urban area. Forty three percents of respondents are male. The percentage of male respondents decreases with age, from 49% for ages 13-20, to 36% for ages above 50. The data cover only individuals older than 13 years.

The RLMS data have a low attrition rate, which can be explained by low levels of labor mobility in Russia (See Andrienko and Guriev 2004). Interview completion exceeds 84 percent, lowest in Moscow and St. Petersbug (60%) and highest in Western Siberia (92%). The RLMS team provides a detailed analysis of attrition effects, and finds no significant effect of attrition.[4]

My primary object of interest for this research is males of ages between 13 and 65. The threshold of 13 years is a minimum age of respondents in the survey. Summary statistics for primary demographic characteristics are presented in Table 1.

2.3 Alcohol consumption variables and drinking patterns of Russian males

My primary measures of alcohol consumption are indicators of whether a person drank vodka or beer (or both) during the previous month. In addition to these main measures, I also use a log of reported monthly alcohol consumption, as well as indicators of whether a person is a heavy drinker of beer and/or vodka. For each beverage, I employ a dummy variable that equals 1 if a person belongs to the top quarter by consumption of this beverage (among males of working age). Table 1 summarizes the statistics of different measures of alcohol consumption.

Seventy-three percents of males reported that they drank alcohol during the previous

[3]I do not utilize data on rounds earlier than round 5 because they were conducted by other institutions, have different methodology, and are generally agreed to be of worse quality.

[4]See http://www.cpc.unc.edu/projects/rlms-hse/project/samprep

month. Vodka and beer are the most popular alcoholic drinks among Russian males: the share of vodka in total alcohol intake (calculated in ml of pure alcohol intake) is 54%, and the share of beer is 27%. Fifty-three percent of males drank vodka and 44% of males drank beer during the previous month. Figure 1a shows that drinking patterns are different for males from different age strata: older males prefer vodka, whereas beer is the most popular drink among males below age 24. Share of beer consumption drops from 56% at age 18 to only 11% at age 65, whereas share of vodka increases from 28% at age 18 to 61% at age 65. However, these patterns are driven by individual-level differences in preferences for people from different age cohorts, but not for a given person over time. Figure 1b shows the evidence for how drinking patterns change with age for a particular person. After subtracting personal averages (among periods of observation), opposite trends are observed: a person starts to consume more beer and less vodka as he grows older. [5]

2.4 Elasticity of alcohol consumption: hedonic regressions

This section presents the results of hedonic regressions for the price of alcohol.

Table 1 presents OLS and tobit estimates for own and cross-price elasticities for different measures of vodka and beer consumption, as well as for total alcohol intake measures, for the following hedonic regressions:

(1) $Y_{it} = \alpha + \gamma_b Log(P_{beer})_{it} + \gamma_v Log(P_{vodka})_{it} + \Gamma' D_{it} + \delta_r + e_{it}$

(2) $Y_{it} = \alpha + \gamma_{bv} P_{beer}/P_{vodka} + \Gamma' D_{it} + \delta_r + e_{it}$

Y_{it} stands for alcohol consumption, D_{it} is a set of demographic characteristics, and δ_r is the regional fixed effects. The set of demographic characteristics includes log(family income), health status, age, I(Muslim), I(college degree), and personal body weight.

Table 2 below and Table 2a at the end of the section illustrate significant negative own and positive cross-price elasticities. According to tobit estimates, a decrease in the price of beer by 10% results in an increase in consumption of beer by 6% and a decrease in the

[5]When these statistics are demeaned I employ only a subset of data with more than one observation per person.

consumption of vodka by 7%. Similarly, a decrease in the price of vodka by 10% will result in an increase in consumption of vodka by 9% and a decrease in the consumption of beer by 10%.

Estimated own-price elasticities (-0.6 for beer and -0.9 spirits) are within the range of those obtained in other studies. Leung and Phelps (1993) and Fogarty (2010) survey estimates of price sensitivity of demand for alcoholic beverages. Leung and Phelps (1993) find that average estimates for the elasticity of beer and spirits are -0.3 and -1.5 correspondingly. Fogarty (2010) finds that average own elasticities are -0.44 for beer and −0.73 for spirits.

Table 2. Demand elasticities: price hedonic regression.

	Log(beer consumption)		Log(vodka consumption)	
	OLS	Tobit	OLS	Tobit
$Log(P_{beer})$	-0.33	-0.582	0.358	0.673
	[0.085]	[0.195]	[0.110]	[0.208]
$Log(P_{vodka})$	0.5	1.029	-0.48	-0.894
	[0.068]	[0.155]	[0.089]	[0.168]

Standard errors in brackets

2.5 Hazard of death regression

Table 3 presents estimates of the effect of alcohol consumption on hazard of death for the following hazard specification:

$$\lambda(t,x) = exp(x\beta)\lambda_0(t)$$

where $\lambda_0(t)$ is the baseline hazard, common for all units of population. I use a semi-parametric Cox specification of baseline hazard. The set of explanatory variables includes alcohol consumption variables, log of family income, health status, weight, age, employment status, and educational level.

Table 3 shows that the probability of death is strongly positively-related with the consumption of vodka. As such, drinking vodka increases the hazard of death twice (=exp(0.68)). However, the hazard of death is high even for males who reported only moderate average monthly vodka consumption. This is because even with moderate average consumption a person can die as a result of one-time hazardous binge drinking. Indeed, most alcohol-related deaths in Russia are not due to diseases that result from long-time alcohol consumption (such as cirrhosis), but rather to (probably occasional) one-time hazardous drinking.

First, 6% of all deaths of Russian males are caused by alcohol poisoning. The main cause of poisoning is not poor quality of the alcohol, but rather imbibing so much alcohol that the amount in the blood causes the heart to stop (see Djoussé and Gaziano 2008). Thus, it takes binging vodka only once to force the heart to stop. In contrast to vodka, beer consumption is safer: one must consume eight times more beer to get the same amount of alcohol in the blood.

Second, 35% of deaths are due to external causes – vehicular and other accidents, or homicides, for example – that occurred largely under the effects of alcohol intoxication. Again even with moderate average vodka consumption, it is enough to binge only once and get into an accident. However, beer consumption does not result in an increase of death hazard, and people who drink beer have a smaller chance of death compared to those who drink vodka, as well to those who do not drink or drink beverages other than beer or vodka. The number of non-drinkers in Russia is very low (less than 10% of males reported that they did not drink in the previous month over three consecutive years), and there is possible negative selection to non-drinkers: non-drinkers have smaller incomes and lower levels of education, do not perform more physical training, and do not have lower rates of disease.

2.6 The stepping-stone effect

It has been proposed that beer may serve as a "stepping-stone" for harder substances such as vodka, and so may have negative long-run consequences on public health. Several recent studies have tested hypotheses regarding a stepping-stone effect against alternative explanations, with unobserved individual heterogeneity in preferences, and have reached the opposite results. Deza (2012) and Mills and Noyes (1984) have found evidence of a modest stepping-stone effect of consuming marijuana and alcohol for later consumption of harder drugs. Beenstock, and Rahav (2002) find a stepping-stone effect in consuming cigarettes for the later consumption of marijuana. However, Van Ours (2003) finds that unobserved individual heterogeneity and a stepping-stone effect explains patterns of drug consumption.

I check for the presence of a stepping-stone effect for 3 subgroups of population: males of age 13-30 for whom I have data on the consumption of alcohol during their high-school years (age 13-18); a sample of all males of age 13-65, and a sub-sample of males who have been abstainers for 2 consecutive years over a period of observation.

Table 4 shows the probabilities of falling into different groups by alcohol consumption, conditional on initial patterns of alcohol consumption.

Panels A and B show conditional probabilities for males of age 20-30 and 25-30, correspondingly. Y_{t-1} for these panels corresponds to alcohol consumption during high school. According to the RLMS survey, 38% of males tried alcohol in high school: 24% tried only beer, 9.3% tried vodka, and 2.6% tried both beer and vodka. Panels A and B show that trying beer in high school does not increase the probability of being a (heavy) drinker of vodka; in fact, these probabilities are even smaller compared to abstainers or to those who consumed vodka or other alcoholic beverages in high school. Panel A and Panel B show that both beer and vodka form product-specific habits: people who start with the consumption of certain kinds of alcoholic beverages in high school keep consuming that same beverage at later ages. Panel C shows conditional probabilities for males of age 18-65 who have abstained for at least 3 consecutive years. Y_{t-1} for this panel corresponds to alcohol consumption after 3 years of being an abstainer. Similar to Panels A and B, Panel C shows little evidence of a stepping-stone effect. Panel D shows conditional probabilities for all males of age 13-65. Y_{t-1} in this panel corresponds to alcohol consumption after 3 years of being an abstainer. Panel D does show some evidence of a stepping-stone effect: drinking beer in a previous period increased the chance of drinking vodka compared to those who were abstainers during the previous period.

2.7 The Model

In this section, I simulate the effects of taxation on alcohol consumption, consumer welfare, and mortality rates.

To do this, I estimate dynamic consumer choice among different kinds of alcoholic beverages.

In my models, agents are assumed to be myopic. Consumers have four choices: drink both vodka and beer $(1,1)$, drink only vodka $(1,0)$, drink only beer $(0,1)$, or drink neither beer or vodka $(0,0)$. Indirect utilities of consumers are assumed to have linear parameterization:

$$\begin{aligned} U(k,j) = \quad & \alpha_{kj} + \gamma_{kj} P_{beer,it}/P_{vodka,it} + \beta_{vkj} I(drink\ vodka)_{it-1} \\ & + \beta_{bkj} I(drink\ beer)_{it-1} + \Gamma' D_{itjk} + \delta_{rkj} + \nu_i + e_{itkj} \end{aligned}$$

Indexes $k, j \in \{0,1\}$ stand for personal choices. The indirect utility of non-drinking is normalized to zero: $U(0,0) = 0$; and γ is normalized to 0 the $(0,1)$ choice, "drink only beer":$\gamma_{01} = 0$. In my model, I normalize price of vodka to 1. With this normalization, a change in beer price results in a change in P_{beer}/P_{vodka}[6]. β_{b10} stays for stepping stone effect for choice "drink only vodka". β_{b11} captures both the stepping-stone effect of beer and beer habit formation for choice "drink both vodka and beer." β_{b01} captures habit formation for the choice "drink only beer." Vodka habit formation effects are captured by β_{v10} and β_{v11}. D_{it} is a set of demographic characteristics that affects utility. This set of

[6] In my simulation experiments I do not analyze the effect of a change in the price of vodka.

demographic characteristics includes log(family income), health status, age, I(Muslim), I(college degree), and personal body weight.δ_{rkj} stands for (unobservable for researcher and observable for individual) regional-specific factors that affect utility, such as official religion, temperature and so on. e_{itkj} is a choice-specific utility component that is unobserved for a researcher but observed for a consumer.

Finally, ν_i stands for an individual-specific taste for alcohol, unobservable to the researcher but observable for the individual. This term captures unobserved personal heterogeneity in tastes for alcohol consumption that do not vary across time and kinds of alcohol. Further, I provide estimation of utilities under different assumptions on unobserved tastes: with and without allowing for unobserved heterogeneity in tastes (with and without individual fixed effects (ν_i)) as well as allowing $e_{itkj}$to be correlated with lagged consumptions (using control function approach with lagged prices used as instruments)[7].

Estimates of utility parameters are shown in Table 5 below, and in Tables 5a and 5b at the end of the chapter[8] Tables 5, 5a, and 5b show strong habit-formation effects, but rather small (if at all) stepping-stone effects. Indeed, Tables 5a and 5b show positive switching costs for changing patterns of drinking (from drinking only beer to drinking vodka) and a strong effect of habits over the same pattern of consumption. Drinking only beer in a previous period positively affects the utility of drinking only beer now, and negatively affects the utility of drinking vodka (with or without beer). The results are similar under any specification with much higher magnitude in CF regression (possibly due to weak instrument bias). Table 5 also shows that the relative price of beer has a negative effect on consumer utility specific to the choice of drinking beer.

Table 5. Estimates of utility parameters. Multivariate logit.

[7]See Train (2003) for description of control function approach.

[8]Table 5a shows estimation results for models with and without unobseved hetergeneity in tastes. Table 5b shows estimation results for control function (CF) regressions as well as estimates for following parametrization of indirect utility:

$$U(k,j) = \alpha_{kj} + \gamma_{kj} P_{beer,it}/P_{vodka,it} + \beta_{vkj} I(drink\ only\ vodka)_{it-1} + \beta_{bkj} I(drink\ only\ beer)_{it-1} + \beta I(drink\ vodka\&beer)_{it-1} + \Gamma' D_{itjk} + \delta_{rkj} + \nu_i + e_{itkj}$$

	U(0,1)	U(1,0)	U(1,1)	U(0,1)	U(1,0)	U(1,1)
Lag I(drink vodka)	-0.439	0.980	0.858	0.139	1.624	1.456
	[0.048]	[0.043]	[0.042]	[0.046]	[0.041]	[0.039]
Lag I(drink beer)	1.011	-0.338	0.961	1.604	0.177	1.500
	[0.049]	[0.047]	[0.043]	[0.047]	[0.045]	[0.040]
$P_{beer,it}/P_{vodka,it}$	-1.557		-0.710	-1.575		-0.719
	[0.242]		[0.175]	[0.240]		[0.173]
Individual FE	Yes	Yes	Yes	No	No	No

Note: Standard errors in brackets.

My next exercise simulates effects of two government policies on consumer drinking patterns, consumer welfare, and mortality rates. I consider two policies: taxation and subsidization of beer consumption. Figure 2 and Table 6 demonstrate the effect of these policies over a 5-year period, specifically the effect of a two-times decrease or two-times increase in the price of beer.

My simulation results show that an increase in share of those who drink beer does not result in a later increase in the share of those who drink vodka. Subsidizing beer results in an increase in the share of those who drink beer, and decreases the share of vodka drinkers and those who do not drink beer or vodka. This policy results in a 17% increase in consumer welfare, a decrease in the share of vodka drinkers from 53 to 49%, and an increase in the share of beer drinkers from 46 to 51%. .[9] Switching consumer patterns toward safer beer consumption results in decrease in mortality rates from 1.4% to 1.1%.

Further, the taxation of beer results in a decrease in the share of beer drinkers from 46 to 37%, an increase in the share of vodka drinkers from 53 to 55%, and an increase in the share of those who drink neither beer nor vodka from 31 to 41%. This policy also results in a 21% decrease in consumer welfare, and an increase in mortality rates from 1.4% to 1.68%.

[9] See Train (2003) for a description of the estimation of model parameters, choice probabilities, and consumer welfare.

2.8 Conclusion

Light alcohol drinks are commonly viewed to be a stepping stone for harder drinks, but also as a safer substitute for them. In this paper, I analyze this trade-off by utilizing micro-level data on the alcohol consumption of Russian males. I find, first, that beer is a safer drink compared to hard alcohol beverages, in the sense that consumption of hard beverages increases the hazard of death while consumption of beer does not. Second, I find that beer is a substitute for vodka: there is significant positive cross-price elasticity in vodka consumption with respect to the price of beer. I find also little evidence that beer consumption actually serves as stepping stone for vodka consumption. Rather. initiation of beer consumption forms habits for further consumption of beer. Drinking beer at earlier ages results in higher beer consumption and higher general alcohol intake at older years, but also results in lower consumption of hard drinks (vodka) compared to vodka drinkers and to non-abstainers in general. Finally, I estimate a multivariate model of consumer choice and quantify the effect of different government policies on mortality rates, drinking patterns, and consumer welfare. I find that taxation of beer may decrease consumer welfare and increase mortality rates. In contrast, subsidizing beer consumption will improve consumer welfare, and even slightly decrease mortality rates.

2.9 Tables

Table 1. Summary statistics.

Variable	Obs	Mean	Std. Dev.	Min	Max
I(drink alcohol)	41721	0.73	0.44	0	1
I(drink vodka)	41721	0.53	0.50	0	1
I(drink beer)	41721	0.44	0.50	0	1
I(drink beer and/or vodka)	41721	0.68	0.47	0	1
log(alcohol intake)	41721	3.46	2.22	0	7.9
Share of beverages in total alcohol intake					
vodka	30532	0.54	0.39	0	1
beer	30532	0.27	0.34	0	1
homemade liquors	30532	0.09	0.25	0	1
vine	30532	0.07	0.2	0	1

Figure 1a. Share of beer and vodka consumption by age.

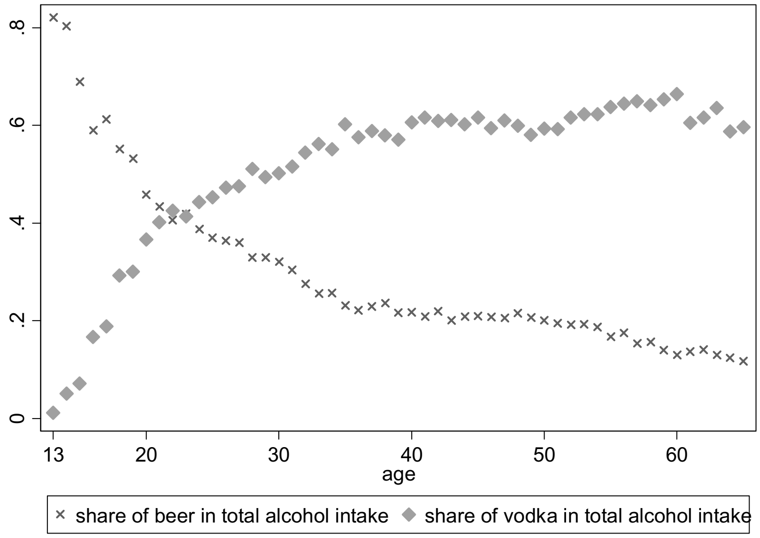


Figure 1b. (De-meaned) share of beer and vodka consumption by time.

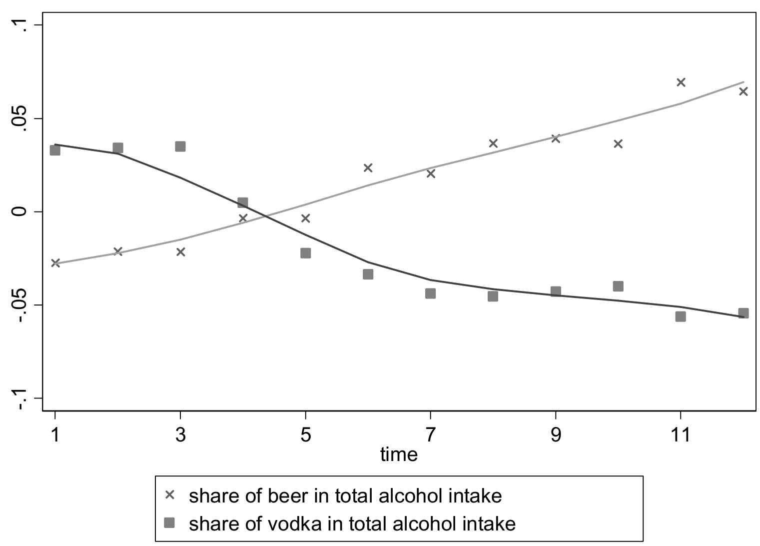


Table 2b. Demand elasticities: price hedonic regression.

Beer consumption	(1)	(2)	(3)	(4)
$Log(P_{beer})$	-0.33	-0.093	-0.072	-0.582
	[0.085]***	[0.019]***	[0.023]***	[0.195]***
$Log(P_{vodka})$	0.5	0.093	0.125	1.029
	[0.068]***	[0.016]***	[0.018]***	[0.155]***
P_{beer}/P_{vodka}	-0.549	-0.104	-0.141	-1.329
	[0.102]***	[0.024]***	[0.027]***	[0.231]***
Vodka consumption	(5)	(6)	(7)	(8)
$Log(P_{beer})$	0.358	0.014	0.015	0.673
	[0.110]***	[0.016]	[0.008]*	[0.208]***
$Log(P_{vodka})$	-0.48	-0.038	-0.07	-0.894
	[0.089]***	[0.013]***	[0.010]***	[0.168]***
P_{beer}/P_{vodka}	0.891	0.056	0.157	1.383
	[0.158]***	[0.020]***	[0.028]***	[0.247]***
Alcohol consumption	(9)	(10)	(11)	
$Log(P_{beer})$	0.156	0.02	0.26	
	[0.104]	[0.007]***	[0.143]*	
$Log(P_{vodka})$	-0.085	-0.002	-0.145	
	[0.084]	[0.009]	[0.115]	
P_{beer}/P_{vodka}	-0.332	0.01	0.235	
	[0.059]***	[0.026]	[0.170]	

Standard errors in brackets;

* significant at 10%; ** significant at 5%; *** significant at 1%;

Dependent variables:

(1) Log(beer consumption) (2) I(top 25% by beer consumption)

(3) I(drink beer) (4) Log(beer consumption), Tobit model

(5) Log(vodka consumption) (6) I(top 25% by vodka consumption)

(7) I(drink vodka) (8) Log(vodka consumption), Tobit model

(9) Log(alcohol intake) (10) I(abstainer) (11) Log(alcohol intake), Tobit

Table 3. Hazard of death estimates

	Hazard of death					
I(drink vodka)	0.68	0.82				
	[0.156]***	[0.165]***				
I(drink beer)		-0.516				
		[0.212]**				
I(heavy drinker: beer)			-0.503	-0.687		
			[0.280]*	[0.279]**		
I(heavy drinker: vodka)			0.555	0.935		
			[0.206]***	[0.224]***		
I(moderate drinker: beer)				-0.542		
				[0.290]*		
I(moderate drinker: vodka)				0.907		
				[0.183]***		
Log(beer consumption)					-0.14	
					[0.056]**	
Log(vodka consumption)					0.171	
					[0.034]***	
Share of vodka in alcohol consumption						0.301
						[0.080]***
Log(alcohol consumption)						-0.151
						[0.080]*
log(family income)	-0.323	-0.311	-0.314	-0.31	-0.31	-0.295
	[0.021]***	[0.022]***	[0.021]***	[0.022]***	[0.022]***	[0.023]***
Age	-0.278	-0.279	-0.272	-0.286	-0.279	-0.263
	[0.017]***	[0.017]***	[0.017]***	[0.018]***	[0.017]***	[0.017]***
Health evaluation	-0.576	-0.573	-0.562	-0.574	-0.579	-0.524
	[0.132]***	[0.132]***	[0.132]***	[0.132]***	[0.132]***	[0.140]***
I(smokes)	0.473	0.464	0.5	0.456	0.445	0.493
	[0.132]***	[0.132]***	[0.134]***	[0.133]***	[0.133]***	[0.146]***
I(college)	-1.836	-1.767	-1.841	-1.754	-1.753	-1.825
	[0.298]***	[0.300]***	[0.297]***	[0.301]***	[0.300]***	[0.308]***
Body weight	-0.002	-0.002	-0.003	-0.002	-0.002	-0.004
	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]	[0.005]
I(employed)	-0.727	-0.696	-0.574	-0.683	-0.674	-0.559
	[0.145]***	[0.144]***	[0.144]***	[0.146]***	[0.144]***	[0.151]***
Observations	7069	7069	7069	7069	7069	6516

Standard errors in brackets; * significant at 10%; ** significant at 5%; *** significant at 1%

Table 4. Estimators of $Pr(Y_t|Y_{t-1})$

	Y_t					
	drink alcohol	drink vodka	drink beer	heavy drinker	heavy drinker, beer	heavy drinker, vodka
Panel A. Age 20-30						
Pr(. | abstainer)	0.339	***0.160***	0.287	0.079	0.130	***0.025***
Pr(. | try only beer)	0.660	***0.135***	0.631	0.101	0.239	***0.017***
Pr(. | try beer&vodka)	0.826	0.499	0.748	0.266	0.324	0.082
Pr(. | try only vodka)	0.785	0.548	0.508	0.282	0.288	0.147
Pr(. | try other)	0.699	0.293	0.365	0.148	0.171	0.043
Panel B. Age 25-30						
Pr(. | abstainer)	0.758	***0.458***	0.641	0.249	0.318	***0.079***
Pr(. | try only beer)	0.773	***0.409***	0.727	0.318	0.500	***0.045***
Pr(. | try beer&vodka)	1.000	0.733	0.867	0.267	0.467	0.067
Pr(. | try only vodka)	0.902	0.537	0.878	0.390	0.512	0.244
Pr(. | try other)	0.848	0.587	0.739	0.217	0.500	0.109
Panel C. All ages, previously abstainers						
Pr(. | abstainer)	0.205	***0.124***	0.143	0.064	0.068	***0.031***
Pr(. | try only beer)	0.522	***0.152***	0.477	0.089	0.192	***0.022***
Pr(. | try beer&vodka)	0.530	0.384	0.430	0.184	0.191	0.087
Pr(. | try only vodka)	0.490	0.381	0.209	0.165	0.100	0.102
Pr(. | try other)	0.492	0.140	0.191	0.117	0.094	0.033
Panel D. All ages						
Pr(. | abstainer)	0.375	***0.213***	0.207	0.100	0.084	***0.054***
Pr(. | try only beer)	0.801	***0.358***	0.657	0.227	0.320	***0.066***
Pr(. | try beer&vodka)	0.898	0.726	0.694	0.403	0.370	0.180
Pr(. | try only vodka)	0.796	0.666	0.286	0.289	0.120	0.172
Pr(. | try other)	0.755	0.295	0.257	0.270	0.094	0.067

Table 5a. Estimates of utility parameters. Multivariate logit.

	U(0,1)	U(1,0)	U(1,1)	U(0,1)	U(1,0)	U(1,1)
$P_{beer,it}/P_{vodka,it}$	-1.557		-0.710	-1.575		-0.719
	0.242		*0.175*	*0.240*		*0.173*
Lag I(drink vodka)	-0.439	0.980	0.858	0.139	1.624	1.456
	0.048	*0.043*	*0.042*	*0.046*	*0.041*	*0.039*
Lag I(drink beer)	*1.011*	*-0.338*	*0.961*	1.604	0.177	1.500
	0.049	*0.047*	*0.043*	*0.047*	*0.045*	*0.040*
Age	*-0.030*	*0.019*	*-0.016*	-0.033	0.016	-0.019
	0.002	*0.002*	*0.002*	*0.002*	*0.002*	*0.002*
Health evaluation	0.218	0.162	0.186	0.252	0.208	0.223
	0.054	*0.046*	*0.046*	*0.052*	*0.044*	*0.044*
Income	0.001	-0.0009	0.001	0.002	0	0.002
	0.0003	*0.0004*	*0.0003*	*0.0003*	*0.0004*	*0.0003*
I(college)	*0.296*	*-0.237*	*0.138*	0.281	-0.236	0.134
	0.045	*0.041*	*0.039*	*0.044*	*0.039*	*0.038*
Body weight	-0.007	-0.003	0.003	-0.005	0.000	0.005
	0.002	*0.001*	*0.001*	*0.002*	*0.001*	*0.001*
I(Muslim)	*-0.295*	*-0.285*	*-0.434*	-0.117	-0.097	-0.253
	0.112	*0.097*	*0.097*	*0.108*	*0.092*	*0.092*
Regional fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Heterogeneity in tastes,ν_i	Yes	Yes	Yes	No	No	No

Note: standard errors in italic.

Table 5b. Estimates of utility parameters. Multivariate logit. Alternative Specifications.

	U(0,1)	U(1,0)	U(1,1)	U(0,1)	U(1,0)	U(1,1)	U(0,1)	U(1,0)	U(1,1)
$P_{beer,it}/P_{vodka,it}$	-1.428		-0.607	-1.394		-0.575	-0.699		0.032
	0.197		*0.149*	*0.195*		*0.147*	*0.247*		*0.197*
Lag I(drink only beer)	0.523	**-0.834**	**-0.964**	0.261	**-1.101**	-1.226			
	0.072	***0.085***	***0.068***	*0.067*	***0.081***	*0.063*			
Lag I(drink only vodka)	-0.889	0.373	-1.056	-1.167	0.104	-1.329			
	0.076	*0.060*	*0.058*	*0.073*	*0.056*	*0.054*			
Lag I(drink vodka&beer)	0.594	0.640	1.765	1.646	1.676	2.827			
	0.063	*0.058*	*0.054*	*0.060*	*0.056*	*0.052*			
Lag I(drink vodka)							-2.223	3.049	2.204
							0.092	*0.073*	*0.088*
Lag I(drink beer)							3.775	-1.610	3.203
							0.088	*0.090*	*0.090*
CFs for lag consumptions	No	No	No	No	No	No	Yes	Yes	Yes
Demographics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Regional fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Heterogeneity in tastes	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes

Figure 2. Effect of government policy on consumer welfare and on drinking patterns.

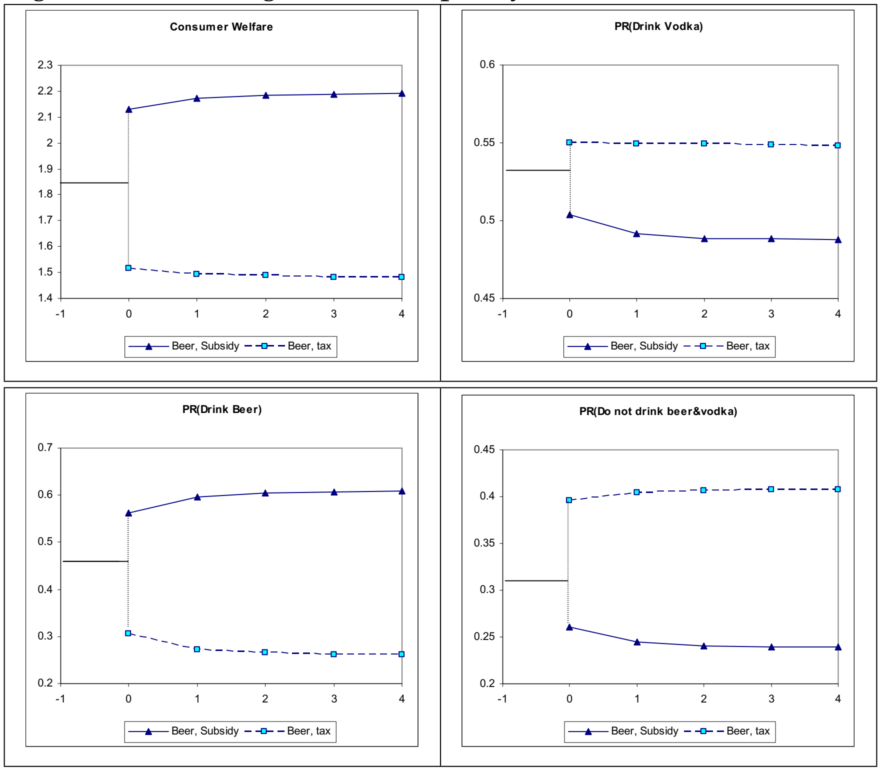


Table 6. Share of drinkers and mortality rates under different policies

	No tax/subsidy	Subsidy on beer	Tax on beer
Drinking patterns:			
Do not drink beer/vodka	0.31	0.24	0.41
Drink only beer	0.15	0.27	0.04
Drink only vodka	0.22	0.15	0.33
Drink beer&vodka	0.31	0.34	0.22
Hazard of death (1)	0.0137	0.0113	0.0168
Hazard of death (2)	0.0152	0.0149	0.0154
CS	1.87	2.19	1.48

Note: Hazard of death estimates in (1) based on Table 5, column 2 estimates, hazard of death estimates (2) based on Table 5, column 1 estimates.

Chapter 3

The Unequal Enforcement of Liberalization

by Evgeny Yakovlev and Ekaterina Zhuravskaya

3.1 Introduction

In recent years, liberalization of business regulations has become very popular among policymakers all over the world. For example, in 2005-2007 sixty-two countries undertook reforms to cut the administrative costs of starting a business and getting a license (World Bank, 2006, p. 4; 2007, p. 4). The effects of such policy experiments have been widely studied (see Djankov, 2009, for a survey). So far, much of this work has focused on estimating the average effect of reforms and has neglected the fact that such effects may depend heavily on the local institutional environment. Aghion et al. (2008) were the first to show that liberalization reform had different effects depending on local labor market institutions in the context of Indian de-licensing reform. In the context of Russian reform of inspections, licenses and registration, this paper also documents unequal effects of liberalization and provides evidence on one of the channels through which local institutions affect the results of liberalization, namely, the level of enforcement. We show that local governance institutions improved the level of enforcement of national liberalization laws in Russia and by means of influencing enforcement had a positive effect on such liberalization outcomes such as firm performance and small business employment. Previous work focused on the effects of changes in *de jure* regulations on outcomes without taking enforcement in consideration. We show that differences in the level of enforcement of liberalization laws, i.e., the wedge between *de facto* and *de jure* regulations, give rise to the variation in reform outcomes across different institutional environments.[1]

Between 2001 and 2004, Russia undertook a drastic liberalization reform of business regulation. Three consecutive national laws focused on liberalization of entry and operation of existing businesses in the areas of inspections, licenses, and registration. Regarding inspections, the relevant agencies (e.g., fire, sanitary, labor, or certification) were limited to no more than one inspection of any particular firm every two years. A substantial delicensing took place during this time, with over one hundred business activities that previously had required licenses becoming exempt. Registration of new firms was transformed from authorization-based to notification-based (by abolishing the need for startups to obtain several admissive documents from various government agencies before starting their operations). Prior to this reform, many scholars pointed to the excessive regulatory burden on Russian firms and argued that over-regulation was among the most important reasons for Russia's poor economic performance during the first eight years of transition.[2] The proclaimed goal of the reform was to increase entry and the growth of small business.

In this paper we study how local governance institutions affected whether this reform succeed in bringing down the administrative costs of doing business and whether it reached the ultimate goal of boosting small business development. We use a unique panel survey data of small firms with questions about their *actual* regulatory burden, which allow us to measure the enforcement of liberalization reform. For each of the three regulatory areas liberalized by the reform (inspections, licenses, and registration), we construct a firm-level measure of reform enforcement by comparing the reform target to the actual

[1]The lack of enforcement has been recognized as an important reason for ineectiveness of regulation at least since Stigler and Friedland (1962). Empirical research, however, has had little to say about the obstacles and driving forces behind the enforcement of liberalization reforms.

[2]See, for instance, Frye and Shleifer (1997); Shleifer (1997); Johnson, Kaufmann and Shleifer (1998); Shleifer and Vishny (1998); Frye and Zhuravskaya (2000).

regulatory burden faced by the firm. The data spans a selection of sub-national regions, and therefore we are able to observe varying degrees of success of reforms in different regions. As Russia's regions are relatively homogenous in culture, but differ greatly in governance institutions, we can study the effect of regional institutions on reform and its outcomes.

As the first of the two steps in our analysis, we study the determinants of the reform enforcement. We consider several aspects of regional institutional environment that potentially can affect the quality of government at the regional level and, therefore, may influence the extent to which the local bureaucrats who administer regulations comply with national liberalization laws. As liberalization takes away rents from these bureaucrats, they may be reluctant to decrease the regulatory burden on firms (Shleifer and Vishny, 1993, 1994). Indeed, we find that the reform was far from perfectly enforced. Despite some improvement in the regulatory burden after the reform, inspections continued at rates well beyond the target set by the liberalization law on inspections; firms still had to apply for licenses for activities that were no longer required according to the de-licensing law, and new firms still had to obtain authorization from various local agencies to start operations, despite the new notification-based registration. Local public officials who administer regulations are expected to have particularly strong incentives to sabotage liberalization when they are not well monitored by the public and businesses, and when they have no fiscal gain from supporting business growth. Indeed, we find that the enforcement of liberalization reform in all three areas of regulation was better 1) in regions with higher transparency of authorities and higher Internet penetration and, therefore, more effective monitoring on the part of the general public; 2) in regions with higher industrial concentration and, therefore, more effective monitoring on the part of large businesses, and 3) in regions with higher fiscal incentives and, therefore, less likely benefit from liberalization. In addition, we find that institutional characteristics affect the enforcement of liberalization of entry and of the operations of established firms in the same way. Our empirical methodology is difference in differences. We estimate the differential effect of introduction of liberalization laws on the wedge between *de jure* liberation targets (i.e., the maximum level of regulation permitted after liberalization) and *de facto* regulations (i.e., the actual level of regulation faced by firms) depending on pre-reform institutional environment.

Second, we use the interaction between the timing of liberalization and the institutional determinants of enforcement of liberalization as an exogenous source of variation in the level of actual regulations to estimate a causal effect of reform on performance and entry of small firms. In particular, we instrument *de facto* change in the level of regulation with *de jure* change interacted with regional institutional variables, controlling for both time and region fixed effects. Instrumenting regulation is important because of reverse causality going from outcomes to regulation as bureaucrats who administer regulations have higher incentives to over-regulate best-performing firms because the higher potential bribe tax that can be collected from these firms. Since the reform is aimed at boosting small business growth, we consider the following reform outcomes: sales growth at the firm-level, small businesses entry to the official sector, and official small business employment at the regional level. Using 2SLS, we find a significant positive effect of de-licensing and of liberalization in the area of inspections on sales growth of firms located in the regions with better governance institutions, and no effect or even small negative effect in regions with poor governance institutions. In addition, liberalization of registration had a significant positive impact on the number of small businesses and the share of employment in small businesses, but also only in regions with strong governance institutions.[3]

The fact that regions with better-monitored and incentivized authorities achieve better reform progress in liberalization of business regulations is consistent with the public choice theory of regulations (e.g., Tullock, 1967).

Our main contribution is to the burgeoning literature on the effects of regulation surveyed in Djankov (2009). A distinguishing feature of our work compared to previous studies is that we measure the enforcement of liberalization reforms by comparing changes in legislation to changes in the actual regulatory burden and demonstrate that, in regions with poor governance institutions, liberalization reforms are poorly enforced and liberalization laws understate the actual regulatory burden. A large body of literature, starting

[3]As we only have data from the official sector, the increases following liberalization in the number of small businesses and the share of small business employment in regions with strong governance institutions reflect the actual business formation as well as the shift of business activity between the official and unofficial sectors. Both have important first-order effects on the economy (Johnson, Kaufmann and Shleifer, 1998; Johnson et al., 2000).

with the pioneering work of Djankov et al. (2002), estimates the effects of various regulations across and within countries.[4] Our results confirm that there is significant variation in the regulatory burden within a country and that looking only at the largest city (as in Djankov et al., 2002, and related work) may give a misleading picture about the state of regulation in the country as a whole. In addition, panel data allow us to control for unobserved regional and firm-level variation as well as time trends and, therefore, to improve on the cross-sectional analysis of many previous studies. Our paper is most closely related to Aghion et al. (2008); the two papers study complementary channels through which local institutions affect the outcomes of a nationwide liberalization.

The paper is organized as follows. In Section 3.2, we describe the reform and the regulations data. Section 3.3 focuses on the estimation of the institutional determinants of the enforcement of the liberalization reform. Section 3.4 reports the estimates of the effect of the reform on outcomes. Section 3.5 discusses robustness. Section 3.6 concludes.

3.2 Background and the measures of regulation

3.2.1 Russia's liberalization reform of business regulation

The level of regulatory burden prior to the Russian liberalization reform of business regulations was extremely high. The goal of the reform was to cut costs to firms associated with inspections, licensing, and registration. The reform consisted of a package of three laws passed at different points in time during 2001-2004: the law on inspections – on August 8, 2001; the law on de-licensing – on February 11, 2002; the law on registration – on January 1, 2004.

The liberalization reform introduced clear measurable limits for the regulatory burden in some areas of regulation and abolished other regulations completely (e.g., Shehovtzov et al., 2005). In particular, the law on inspections stipulated that each inspecting agency is

[4]See, for instance, Djankov et al. (2003); Botero et al. (2004); Mulligan and Shleifer (2005a,b); Klapper, Laeven and Rajan (2006); Djankov, McLiesh and Ramalho (2006); Aghion et al. (2005, 2008); Monteiro and Assuncao (2006); Bruhn (2007); Kaplan, Piedra and Seira (2007); Chari (2007).

allowed to conduct a maximum of one regular (or so-called "planned") inspection of each firm in a two-year period. If no violation is found during the inspection, the next visit can take place no earlier than in two years. If violations are found, they need to be officially recorded by the inspectors, an official fine should be levied on the firm, and inspectors may return to confirm correction of the violation. The previous legislation did not put a limit to the number of "planned" visits by inspectors. Before the new law took force, inspectors came to visit firms very often and they rarely officially recorded violations, instead extracting unofficial payments from businessmen and not requiring them to correct violations.

The de-licensing law reduced the number of business activities that required licenses from 250 to 103. For example, the following business activities became exempt from licensing in 2002: real estate agents, pawn shops, publishing houses, audio studios, private certification firms, antique shops, construction firms, bread bakers, drilling and drill manufacturing, and service work in sea ports.[5]

The registration law introduced a so-called "one-stop shop" rule for registration, and formalized the list of required documents. Previously, any start-up had to obtain authorizations with several different government agencies, such as the pension fund, the social security department, the statistical and fire departments, and the local administration. Additionally, the rules for registration differed across localities. According to the new law, a start-up needs to submit all necessary documents to the local branch of the tax ministry and no permission is necessary to start operations.

We study the effects of these three changes in the legislation.[6]

[5]This law also increased the minimum length of license validity from three to five years.

[6]Another important change to Russian legislation, passed on January 1, 2003, simplified the tax system for small businesses. This law increased the scope of application of the existing system of simplified tax administration, which allows small firms to pay a single "unified" tax with a flat rate on either profit or revenue instead of many taxes such as VAT, profit, sales, and property taxes, and reduced the tax rate for the "unified" tax. The timing of this law is such that it is not a confounding factor to the liberalization reform that we consider. In addition, on July 1, 2002 and on July 1, 2003, two laws streamlined the procedures for product certification and registration. But, unlike the liberalization laws that we consider, they did not liberalize any regulatory areas and were not aimed at reducing regulatory burden.

3.2.2 The MABS survey

The Center for Economic and Financial Research in Moscow conducted a long-term project of Monitoring of Administrative Barriers to Small business (MABS). The project collected data on the regulatory burden on Russian firms by means of regularly repeated surveys of top managers in 2,000 small firms in a selection of 20 regions of Russia. During face-to-face interviews, top managers were asked about firms' actual quantifiable costs associated with inspections, licensing and registration.[7] Two survey instruments are used: one inquires about the regulatory burden on firms established more than a year ago and the other is designed for newly registered start-ups in order to monitor the administrative costs of entry. Panel data are collected to monitor the administrative burden on existing firms that comes from inspections and continuation licenses and a repeated cross-section is collected to monitor the costs of registration and acquisition of start-up licenses. New start-ups constitute about 20% of the total sample in each survey round. The samples were constructed separately in each region. The sample of established firms was drawn at random from the census of regional small and medium-size businesses as of 2000 and the sample of start-ups was drawn at random from the official list of firms registered in the region during the last half year.

The MABS data set includes the results of all six rounds of the survey conducted in the spring and the fall of 2002, the spring of 2003, 2004 and 2005, and the fall of 2006.[8] Each round collected information about all aspects of the regulatory burden on firms for the immediately preceding six months (e.g., the fifth round took place in the spring of 2005 and collected all variables for the second half of 2004. In addition, the first round of the survey (which took place in the spring of 2002) collected information about inspections in the first half of 2001.

Figure 1 presents the timing of the stages of liberalization reform and the periods covered by the data. The first round of the survey collected baseline information from the time

[7]The survey also collected objective information on certification and tax administration and asked managers about their subjective perceptions of the business climate. In this paper, we focus exclusively on the objective data on the regulatory burden in the areas affected by liberalization.

[8]See reports on survey results at www.cefir.org/index.php?l=eng&id=25 and an interactive database at www.cefir.ru/monitoring.

before any of the liberalization laws came into force. The data from the second round onwards allow evaluation of reform progress after the enactment of the law on inspections. The data from the third round onwards enable an assessment of the effect of de-licensing law. The last two rounds allow evaluation of the impact of the registration law.[9]

The sampling procedure was as follows. In each round and each region, 20 newly-registered firms were chosen at random from the list of the population of all firms which registered in this region in the half-year preceding the survey round. In the first round of the survey, in each region, 80 established firms were chosen at random from the registry of existing small businesses with the following quotas that ensured over-representation of construction and manufacturing firms: 8 construction firms and 25 manufacturing firms.[10] From the second round onward, the aim was to keep as many established firms in the sample as possible in order to ensure the panel structure of the data. In every round starting with the second one, 88% of established firms come from the previous round sample. Out of them, 85% come from the established-firms sample of the previous round and 15% come from the new-firms sample of the previous round. The attrition from the sample established firms, therefore, was 25% if compared to the previous round sample. It is, however, over-stated, as 9% of firms that do not appear in current round reappear in next two rounds. So, the attrition rate in the panel of established firms over 4 rounds is 22%.[11] The replacements for firms that dropped out of the panel were chosen at random, first from the pool of firms that appeared in the sample of newly-registered firms in previous

[9] All of these data are in half-year increments. The enactment of the law on registration fell exactly between rounds 4 and 5 of the survey. This is not the case for the laws on licensing and inspections. In our empirical exercise, we assume that the law on inspections took force between rounds 1 and 2, even though in reality the law took force in the *middle* of round 1. Similarly, we assume that the law on licenses took force between rounds 2 and 3 (rather than in the middle of round 2). This is done for two reasons: First, one should expect at least a few months lag between the enactment of the law and its implementation; and second, during the half-year period when each of these laws were enacted, inspectors and license authorities may have deliberately shifted their activities earlier in the respective half-year periods in order to avoid the need to comply with the new laws. The results are robust to making an alternative assumption about the timing; this, however, requires the use of retrospective data for inspections in the first half of 2001, which are subject to a recall bias.

[10] Selection was based on the industry code originally reported by the firms at the time of registration, which often was different from the actual industry reported during the interview.

[11] There is no data on the reasons for attrition, which could include exit from the market, relocation, or refusal to participate in the survey.

rounds, and then from the registry of existing small businesses.[12]

3.2.2.1 The measures of enforcement of liberalization

We measure at the firm level whether the regulatory burden meets the targets set by the liberalization reform. At each round of the survey for every firm in the sample, we construct dummies for whether the actual inspections and licenses of firms comply with the liberalization laws on inspections and licenses. And for every newly registered firm in the sample, we construct a dummy for whether registration procedure complies with the liberalized registration law.

For inspections, our measure of meeting the liberalization target is a dummy indicating whether there was no more than one sanitary inspection per six-month period.[13] We focus on sanitary inspection because it is one of the most frequent in our sample.[14]

To describe the measure of meeting the liberalization target in de-licensing, let us first define the terms. We call a license *"legitimate"* if it is issued for a business activity that is supposed to be licensed according to the 2002 de-licensing law. In turn, we call a license *"illegitimate"* if it is granted for an activity that is not supposed to be licensed according to this law.[15] We consider a dummy for having no illegitimate licenses in a firm as an

[12] Note that because of the differences in sampling rates across regions and across industries, regional-level regressions presented below somewhat over-represent firms in construction and manufacturing as well as regions with smaller number of small firms. Sampling rates differ across regions, as the sample has the same number of firms per region, but the number of small firms in different regions differs.

[13] The dummy equals zero only when extreme violations of the liberalization target occurs, because the law limits the number of inspections to one in *two years*, whereas we look at the situations with two or more inspections in a firm during *six months* in order to avoid autocorrelation in our panel. These extreme violations are not rare: in 2001, 12% of all firms had more than one sanitary inspection in six months; the situation improved by 2006 (five years after the law took force), but the rate of violations of this deregulation target remained non-trivial: 6.4% of firms.

[14] According to our data, 36% of firms dealt with sanitary inspections. There is some industry-level variation in frequency of sanitary inspections. In the food industry 85% of firms had sanitary inspections. In the high-tech and construction industries one quarter of firms had sanitary inspections. In other industries this number varies from 35 to 46%. We control for industry dummies in all specifications.

[15] For example, if a realty firm applied for and was granted a license to operate after 2002, we record a violation of the law and call this license illegitimate. The data show that many firms applied for and were granted licenses for activities that did not require licenses according to the new de-licensing law after it took

indication that the de-licensing target is met.

We measure compliance with the liberalization target in the area of registration by a dummy indicating whether registration of a new firm did not require admissive documents. More precisely, it takes the value of one if the firm had to visit only the local branch of tax ministry for registration and takes the value of zero if the firm had to visit and obtain permission to enter the market from any government agencies apart from the local branch of tax ministry.

Before the liberalization laws took force, the three measures indicate whether liberalization reform was binding in each of the respective areas of regulation. After the liberalization laws took force, the three dummies indicate the level of enforcement of respective liberalization laws.[16]

Summary statistics for the measures of meeting liberalization targets are reported in Table 1 for before and after the reform. The table shows that in all three dimensions of reform, the level of attainment of liberalization targets had increased after the reform compared to before the reform. Yet, the change in the compliance with liberalization targets is not very high on average, particularly for inspections and licensing. 88% of firms had fewer than two inspections per half-year period before liberalization of inspections compared to 93% after the liberalization. 77% of firms had no illegitimate licenses before de-licensing and 79% after it. 25% of new firms registered without having to visit more than one government agency for registration before liberalization of entry, and 43% after it. Table A1 in the Appendix, summarizes the levels of *de facto* regulations that were used to calculate the compliance dummies. On average, established firms had 0.7 sanitary inspections per half-year period and 1.2 illegitimate licenses before liberalization, and 0.4 sanitary inspections and 0.9 illegitimate licenses after liberalization. Startups had to visit 4 government agencies for registration on average before liberalization and 2.7 after liber-

force. In focus group interviews, firm managers said that it was cheaper for them to pay for the illegitimate licenses than to defend their right to operate without a license in court. Most illegitimate licenses had been granted by regional authorities.

[16]It is important to note that since our data are comprised of firms that actually exist (i.e., entered the market and survived to the time of the survey), there is an inherent problem of sample selection. Ideally, one would have liked to know the level of regulatory burden for firms that were not able to enter the market or exited because the regulatory burden they faced was too high. This sample selection problem, however, is shared by all studies in this literature.

alization.[17] Importantly, these average changes in compliance with liberalization targets and regulatory levels may not be driven by the liberalization reform, as the level of regulations can change over time with macro-economic trends and other time-varying factors. Figure 2 plots the means of the measures of enforcement of liberalization by the rounds of the MABS survey and Figure A1 in the Appendix presents the dynamics of the level of respective regulations. The figures illustrate that there is no *obvious* discontinuous jump in the compliance with liberalization targets or *obvious* discontinuous drop in the levels of regulation at the time of liberalization. Instead, we observe time trends and some fluctuations around them. This suggests that the enforcement of the liberalization laws on average was rather poor and that it is essential to control for the overall trends in order to estimate the impact of liberalization on the actual regulatory burden across different institutional environments.

3.3 Governance institutions and the enforcement of liberalization

3.3.1 Hypotheses and measures of institutions

The incentives of bureaucrats who administer regulations at the local level are important for the actual implementation of reforms of business regulations and, in particular, liberalization reforms. In this paper, we consider institutional characteristics that potentially can affect incentives of bureaucrats at the local level to meet the targets of liberalization laws.

Since the initial level of regulatory burden on firms was excessive–as reflected in the general consensus among academics, politicians, and businessmen–it is reasonable to as-

[17] Note that data are missing for newly-registered firms in round 4 for 11 out of 20 regions. The reason was the resignation of Russia's cabinet of ministers leading to a situation in which nobody in the government knew where the data on the registration of firms were located; these data were needed for sampling of new firms in round 4 of the survey. Data are also missing for Altaisky Krai in the 3rd round due to a reorganization of the regional survey agency that was supposed to conduct the survey.

sume that the general public as well as managers of small businesses were in favor of liberalization. In contrast, we expect local bureaucrats to be interested in maintaining high levels of regulation and opposing liberalization because they benefited from excessive regulations, and liberalization would take their rents away (according to the public choice theory of regulation, e.g., Tullock, 1967; Shleifer and Vishny, 1993; Djankov et al., 2002). Therefore, in regions where the general public can better monitor bureaucrats, one should expect better enforcement of liberalization. We consider two aspects of the ease of monitoring of regional governments by the public: government transparency and public access to independent media, and we expect better enforcement of liberalization laws in regions with higher government transparency and greater access of the public to independent media.

Regional governments in Russia are often influenced by large regional businesses (e.g., Slinko, Yakovlev and Zhuravskaya, 2005). Holding the initial level of regulation constant, we expect the extent to which local bureaucrats are influenced by large business to facilitate enforcement of liberalization reform, at least with regard to regulations of established firms. This is because all established firms benefit from de-licensing and reducing the number of inspections. The monitoring and control of regional bureaucrats by large industry incumbents, however, may have an ambiguous effect on the enforcement of liberalization of entry. On the one hand, industry incumbents may be in favor of higher regulation of entry in order to protect themselves from potential competition (e.g., Stigler, 1971). On the other hand, they may be in favor of boosting small business entry as it is politically less costly to shed excess labor–a characteristic of large industrial firms in Russia–when laid off workers can find jobs in small businesses. Following Grossman and Helpman (1994), we use concentration among industrial firms to proxy for the extent of monitoring of regional bureaucrats by large businesses.

In addition, the strength of fiscal incentives of regional governments, i.e., the correlation between business growth and the actual size of the disposable regional budgets, is also expected to increase enforcement of liberalization (holding the initial level of regulation constant). Local government has stronger incentives to enforce liberalization reform in order to maximize the tax base when the local budget primarily relies on own revenues (e.g., Zhuravskaya, 2000; Jin, Qian and Weingast, 2005).

We take the following four variables as baseline measures of potential institutional determinants of enforcement of liberalization: government transparency, Internet penetration, industrial concentration, and fiscal incentives. The exact definitions of all institutional measures are presented in the Data Appendix and summarized in Panel A of Table A2. We verified that our results are robust to using various alternative measures of governance institutions (as described in the Data Appendix). Note that institutional variables do not vary over time and were measured in 2000, before liberalization had started.

3.3.2 Three regulatory areas, taken separately

3.3.2.1 Methodology

In this section we explore the differential impact of liberalization laws on the attainment of liberalization targets depending on the initial regional institutional environment. We use the difference-in-differences (DD) methodology to study the effect of the pre-determined (i.e., pre-reform) institutional characteristics on the local enforcement of national liberalization reform, exogenously-mandated from the point of view of the regions.

First, we consider each area of liberalization, i.e., inspection, licensing, and registration, separately. We regress each of the three measures of meeting liberalization targets on the interaction between the onset of liberalization dummy and a potential institutional determinant of enforcement of liberalization. We control for time fixed effects and region or firm fixed effects depending on whether we are looking at new startups for which we have repeated cross-sections, or established firms for which we have panel data. The firm-level panel dataset on established firms contains information on licensing and inspections; repeated cross-sections of new firms contain information on licensing and registration.

Thus, for licensing and inspections in established firms, we estimate the following equation with firm and time fixed effects (ϕ_f and ρ_t):

$$L_{ft} = \alpha I_r A_t + \beta \bar{L}_{rt_0} A_t + \delta' \mathbf{X}_{ft} + \mu' \mathbf{Z}_{rt} + \phi_f + \rho_t + \varepsilon_{ft}; \tag{3.1}$$

whereas for licensing and registration of new firms, the estimated equation has region and time fixed effects (ϕ_r and ρ_t):

$$L_{ft} = \alpha I_r A_t + \beta \bar{L}_{rt_0} A_t + \delta' \mathbf{X}_{ft} + \mu' \mathbf{Z}_{rt} + \phi_r + \rho_t + \varepsilon_{ft}. \tag{3.2}$$

Subscript f indexes firms; subscript t indexes time periods (i.e., rounds of MABS survey); and r refers to the region, where firm f is located. Dependent variable L_{ft} stands for one of the three measures of the attaintment of liberalization targets in firm f at time t (described in Section 2 and summarized in Table 1). I_r stands for institutional characteristics described in the previous section. A_t is the "after liberalization" dummy (or "AFTER" for short) which takes the value of one when the respective liberalization law takes force. Firm and region fixed effects (ϕ) control for all time-invariant characteristics of firms and regions. Time effects (ρ) control for over-time variation in the level of regulation.

The main coefficient of interest in this specification, α, is a DD estimate of the impact of institutional characteristics on the enforcement of liberalization. To be precise, it estimates the differential effect of the liberalization reform, i.e., the enactment of liberalization laws, on the level of compliance with liberalization targets in an average firm depending on the level of regional institutional characteristics. The main assumption necessary for the validity of our estimation strategy is that in the absence of institutional variation, the average change in the attainment of liberalization targets as a result of liberalization would have been the same across regions conditional on the set of covariates, described below. (We discuss the validity of this assumption after the presentation of the baseline results.)

It is important to allow for differential effect of reform depending on the initial level of regulation because the institutional environment is often correlated with the initial level of regulation (i.e., initial attainment of liberalization targets). Therefore, we control for $\bar{L}_{rt_0} A_t$, the interaction of the initial level of regulation ($\bar{L}_{rt_0}$) and the "after liberalization" dummy (A_t). $\bar{L}_{rt_0}$ is calculated as the average of L_{ft_0} across all firms for each region r at t_0. The initial time period (t_0) refers to the first round of the survey which measures the benchmark level of regulations before any of the reform laws took effect, i.e., the second half of 2001.[18] Without this covariate, one could have found a spurious correlation

[18] It is with noting that L_{ft_0} varies only across regions and not over time, and therefore, our panel is not a dynamic panel.

between the progress of reforms and institutions due to the omitted effect of the initial level of regulation. This is because institutional characteristics correlate with the initial level of regulatory burden and the initial level of regulatory burden correlates with the costs of reform implementation. Indeed, as we report below, reform progress was higher in regions with a particularly high regulatory burden to start with. In addition, industrial concentration is negatively significantly correlated with the initial level of attainment of liberalization targets (L_{ft_0}), whereas government transparency has a positive and significant correlation with measures of L_{ft_0}. Even though it is necessary to control for $\bar{L}_{rt_0}A_t$ to avoid omitted variables bias in the estimates, as a robustness exercise we verify that all our results qualitatively are robust to exclusion of this control. (The results of this robustness check are available from the authors upon request.)

We include the following variables in the list of covariates: $\mathbf{X}_{ft}$ is a vector of controls for basic firm characteristics, i.e., age, size allowing for a quadratic term, legal firm, state vs. private ownership, and industry. $\mathbf{Z}_{rt}$ is a vector of additional regional covariates. It includes the logarithm of regional population to control for the regional size and the mean individual income to control for prosperity of the region. We correct standard errors to allow for clustering of error terms (ε_{ft}) within region before and after the reform to account for residual correlation among firms and overtime within region. All control variables are summarized in Table A2 in the Appendix.

In addition, in order to estimate the full average effect of liberalization reform, we replace time dummies with the linear time trend and include A_t as a covariate. So that equation 3.1 transforms into: $L_{ft} = \gamma A_t + \alpha(I_r - \bar{I})A_t + \beta(\bar{L}_{rt_0} - \bar{L}_{t_0})A_t + \delta'\mathbf{X}_{ft} + \mu'\mathbf{Z}_{rt} + \phi_f + \sigma t + \varepsilon_{ft}$; and we do a similar transformation for equation 3.2 as well.[19]

3.3.2.2 Results

The results are presented in Table 2. The upper panel presents the results for the sample of established firms; the lower panel presents the results for newly registered firms. The

[19]In order to interpret the coefficient γ as the full effect of reform at the mean level of institutional and regulatory environment, we subtract the sample means ($\bar{I}$ and $\bar{L}_{t_0}$) from I_r and $\bar{L}_{rt_0}$, respectively, before taking their cross-terms with A_t.

first four columns for each outcome report regression results for interactions of $AFTER$ with institutions included one-by-one and the last column includes all of these interactions together. In addition, in the last column, we replace time dummies by $AFTER$ and the linear trend. As can be seen from the estimated coefficients on the cross-terms, all considered institutional measures, i.e., government transparency, Internet penetration, industrial concentration, and fiscal incentives, significantly improve the local enforcement of liberalization in the areas of inspections and registration; and all institutional measures, with the exception of fiscal incentives, significantly improve enforcement of the liberalization of licensing regulations. There is no difference in the direction of the effect of institutional measures in general, and industrial concentration in particular, for the enforcement of liberalization of entry regulations and those of established businesses. Thus, we can conclude that large incumbent firms in Russia lobby for liberalization of entry as well as of day-to-day operations of existing firms.[20]

Regional institutional characteristics are positively and significantly correlated (we report correlation matrix in Table A3 of the online Appendix). Thus, once we include all of the interactions as covariates at the same time, coefficients at some of them become statistically insignificant and occasionally change the sign. Yet, the cross-terms of $AFTER$ and institutional measures are jointly statistically significant as reflected in the results of the F-test presented in the last two rows of the table.[21]

[20]This result may seem contradictory to regulatory capture theory (Stigler, 1971), which postulates that regulations are created to protect incumbents from the competition of potential entrants. Yet, owners of large industrial firms–who are the main regional lobbyists in Russia (Slinko, Yakovlev and Zhuravskaya, 2005) often have political benefits from the entry of small businesses and, therefore, from lower entry regulations. As we mentioned above, they are often interested in shedding excess labor (which is a legacy of the Soviet economy). Emerging small businesses absorb laid-off workers and make layoffs less politically costly for large businesses.

[21]Overall, the results are consistent for the regressions with region and firm fixed effects, with time dummies or linear trends and for the samples of old firms and startups. To ensure robustness of our results, we also use several additional institutional measures described in the online Appendix: the industrial concentration of output, non-zero subscription to the only two independent (at that time) national business newspapers *Vedomosti* and *Kommersant*, and the presence of a signal of the largest independent radio station *Echo Moscow* in the area. Also, we re-estimated all regressions using the *levels of actual regulation* as dependent variables rather than the dummies indicating the attainment of liberalization targets. The results using alternative institutional and regulatory measures are very similar to the baseline. In addition, our results do not depend on the inclusion of the regional control variables, i.e., population and income, which may be endogenous. We also verified that the results are robust to controlling for regional-level labor productivity.

3.3.2.3 Magnitude: an example of Amur and Samara regions

As institutional measures are positively correlated, the magnitude of the results is better understood by a comparison of a typical "good" region and a typical "bad" region in terms of the whole cluster of institutions. As an example of a "good" region, we take the Samara region (*Samarskaya Oblast*). It has one of the highest levels of government transparency, own revenues, and industrial concentration: out of 20 regions, it is the 4th from the top in terms of government transparency, the 2nd in terms of fiscal incentives (i.e., the size of own revenue share), and the 1st in industrial concentration. It is also among the top third of regions in terms of Internet penetration. As an example of a "bad" region, we take the Amur region (*Amurskaya Oblast*). It is 17th out of 20 regions in terms of government transparency, has the second lowest levels of fiscal incentives and Internet penetration, while it is the 6th from the top in industrial concentration.[22] We plug in the values of institutional measures for the two regions to the estimated results presented in the last column in Table 2 for each regulatory area. This exercise yields that a typical "good" and a typical "bad" region (which differ in terms of institutional environment but are similar with respect to control variables) attain very different levels of regulatory burden following liberalization. A region with institutions at a level similar to Samara is expected to have 8 and 11 percentage point increases in the probabilities of attainment of liberalization targets for inspections and registration, respectively, compared to just one percentage point increase in the respective probabilities for both areas of regulation in a region with institutional environment similar to Amur. As far as de-licensing reform is concerned, in a good region like Samara, de-licensing is expected to decrease the probability of having illegitimate licenses among old firms by 4 percentage points and among new firms by 52 percentage points, whereas in a bad region like Amur, the share of firms with illegitimate licenses is expected to *increase* by 4 percentage points among established firms and 18 percentage points among newly registered firms despite the liberalization

[22]The values of institutional measures for Amur and Samara regions are as follows: government transparency 3.3 and 10.9, industrial concentration 0.22 and 0.44, internet penetration 3.36 and 4.36, and fiscal incentives 0.63 and 0.96.

reform.[23]

To summarize, our main finding in this section is that government transparency, industrial concentration, and Internet penetration consistently, significantly and robustly affected the enforcement of liberalization reform.

3.3.3 Average effect across the three regulatory areas

As the effect of institutions on enforcement of liberalization is consistent across all regulatory areas, we can estimate the average impact of adoption of a liberalization law on enforcement by pooling data for all areas of regulations. Let L_{ift} denote the measure of attainment of the liberalization target in the regulatory area i, where $i \in$ (inspections, licenses, registration). We estimate the following equation:

$$L_{ift} = \gamma A_{it} + \alpha(I_r - \bar{I})A_{it} + \beta(\bar{L}_{irt_0} - \bar{L}_{t_0})A_{it} + \delta' \mathbf{X}_{ft} + \mu' \mathbf{Z}_{rt} + \phi_{ir} + \rho_t + \sum_i^3 \eta_i t d_i + \varepsilon_{ift}. \quad (3.3)$$

It is a modification of equation 3.2 allowing for estimation of an average effect across all i. In this specification, the "after liberalization" dummy varies across regulations as well as over-time reflecting the fact that liberalization took place at different points in time in the three regulatory areas. To control for trends in the level of regulations, we also include linear time trends specific to each regulation (td_i, where t is linear trend and d_i is a dummy for regulation i). The results are robust to allowing regulation-specific linear trends to change slope after the reform. The rest of the notation is as above. Again, standard errors are corrected to allow for clusters of error terms (ε_{ift}) within each regulation, each region and periods, before and after the reform.

In equation 3.3, the γ coefficient is a DD estimate of the average enforcement of the liberalization reform across all regulatory areas; and α, which is our main coefficient of

[23]In reality, the share of firms without illegitimate licenses increased in the Samara region by 11 percentage points and in Amur region it actually decreased by 3 percentage points following the de-licensing law; while the shares of firms that had no more than one sanitary inspection increased by 16 and 12 percentage points in the Samara and Amur regions, respectively, following the liberalization law on inspections.

interest, is an estimate of the differential enforcement of reform across different institutional environments averaged out across regulations.[24] Table 3 presents the results of the estimation of this equation. First, it confirms our previous findings by showing that governance institutions significantly facilitate enforcement of liberalization. Coefficients on cross-terms with all institutional measures, when included individually, are positive and statistically significant, and when included together they all are positive and three out of the four are statistically significant. In addition, we find that on average across all regions and all regulatory areas, a liberalization law increases the probability of compliance with liberalization targets by 6.8 to 9.1 percentage points (depending on specification) as can be seen from the estimates of the coefficient on $AFTER$. The magnitude of the differential effects is as follows: following liberalization, one expects the compliance with liberalization targets to increase by 16 percentage points in a region similar to Samara in terms of institutional environment and by 7 percentage points in a region similar to Amur.

All regressions presented in Tables 2 and 3 show that the initial level of the regulatory burden itself is also a very important determinant of reform progress. The coefficients on the interaction of the initial level of regulatory burden and the "after liberalization" dummy are negative, statistically significant and large in magnitude. The reform partially equalized the level of regulatory burden across firms. According to Table 3: a 10 percentage point higher compliance with liberalization targets prior to reform leads to a 4 percentage point less progress (i.e., lower increase in the compliance with liberalization targets) as a result of reform.[25]

[24]Similarly to the analysis above, the empirical strategy imbedded in the estimation of equation 3.3 is valid only if the following two assumptions hold (subject to holding all covariates constant): 1) in the absence of liberalization, different regulatory measures i would have followed their own overtime trends (td_i) and would not have had a discrete shift at the time when the reform actually took place; and 2) in the absence of institutional variation across regions, the impact of liberalization on *de facto* level of regulations would have been uniform.

[25]Potentially, a negative correlation between the enforcement of liberalization and the initial level of compliance with liberalization targets could be generated by a mean reversion process due to a measurement error in the level of regulation. First, for the law liberalizing registration, we checked directly that there was no mean reversion before the reform; we could not do it for licensing and inspections laws, as there are not enough data prior to reform. Second, we verified that the exclusion of the cross-term of the post-reform dummy, and the initial level of regulation from the list of covariates do not substantially affect the results regarding the effect of institutions, i.e., estimates of α.

3.3.4 Testing the assumption about the absence of regional trends

The main identifying assumption in the regressions for the determinants of enforcement of liberalization is the absence of a correlation between institutional environment and pre-reform trends in the regional regulatory burden. We perform two tests of this assumption. First, we regress the degree of attainment of liberalization targets at the firm level for the three dimensions of regulation on the time trend interacted with institutional variables before reform, controlling for region and time fixed effects. Second, we regress first differences in the attainment of liberalization targets at the regional level on the institutional measures, also prior to reform, controlling for time dummies. These exercises yield 24 regressions (i.e., 2 specifications x 3 areas of regulation x 4 institutions). In only 4 of these regressions do we find a statistically significant (at 10% level) negative relationship, and in approximately the same number of cases we find a positive relationship between the dynamics of regulation and institutions. In addition, the number of positive and negative coefficients is approximately the same. Thus, we conclude that this assumption is reasonable, subject to an important caveat regarding data limitations. In particular, for licensing and registration laws, there are only two data points pre-reform, and for the inspections law there are only two data points including the retrospective data.[26]

3.4 The outcomes of liberalization

3.4.1 Firm performance

Russia's liberalization was aimed at boosting small business growth. What were the effects of the reform in the light of proclaimed goals? This section addresses this question. A common problem with figuring out the effects of a liberalization reform on business

[26] In the robustness section 2, we discuss the results of a placebo experiment, in which we vary the timing of laws. Had there been region-specific trends in regulations, the results of this placebo experiment would have been different. In addition, section 2 reports how our results change with the inclusion of the interaction of the linear trend with institutional variables as an additional control variable.

growth is endogeneity of regulations. According to the public choice theory of regulations (Tullock, 1967), predatory regulators are attracted disproportionately to well-performing firms and regions with thriving business. This is because they can generate more rents by preying on successful and profitable firms. Therefore, there may be reverse causality from business growth to higher regulatory burden. Without finding an exogenous source of variation in the level of regulatory burden, causal claims based on correlation between the level of regulation and economic outcomes are problematic.

We use the determinants of the variation in enforcement of Russia's liberalization laws as instruments in order to solve this endogeneity problem. Our goal is to estimate the relationship between firm performance and the level of regulation:

$$Y_{ft} = \xi L_{ft} + \eta' \mathbf{X}_{ft} + \zeta' \mathbf{Z}_{rt} + \phi_f + \rho_t + \varepsilon_{ft}, \tag{3.4}$$

where Y_{ft} is sales growth in firm f and time t (defined as percentage change in sales over the 6-month period). This variable is available for each firm in the MABS survey, as summarized in Table A2). We can estimate the equation 3.4 for licenses and inspections as we have firm-level panel data with information on L_{ft} in these regulatory areas. For registration, the data are a repeated cross-section. So, as above, we replace firm fixed effects with region fixed effects, and error terms are clustered by region before and after liberalization.

Since our measure of the level of regulations, i.e., the attainment of liberalization targets L_{ft}, is endogenous, we need to find an exogenous source of variation in regulations. The previous section documented that the liberalization reform was enforced differently across regions. We argue that the magnitude of the decrease in *de facto* regulations, i.e., the increase in compliance with liberalization targets, at the onset of reform is exogenous to firm performance once we control for the overall trends in outcomes and cross-region differences in institutional environments. Liberalization should be the only reason for a discrete shift in *de facto* regulations at the particular time of enactment of each of the deregulation laws, while the institutional environment determines the magnitude of this shift. Thus, we use as instruments the interaction of the "after liberalization" dummy for a particular deregulation law (A_t) with an institutional measure (I_r).[27] Therefore, equation 3.1 is the first stage to predict an arguably exogenous component in variation of regulatory levels in inspections and licensing; and equation 3.2 – in registration. The inclusion of time and region fixed effects into the list of covariates is crucial to the validity of these instruments. First, regional institutions (I_r) may have a direct effect on firm performance (Y_{ft}). Second, the dependent variable has an over-time trend. As, by construction, both the time trend and cross-sectional institutional differences are correlated with the instruments ($I_r A_t$), controlling for the direct effect of the time trend with round dummies and of institutions with firm/region fixed effects. With these controls included, the instrument picks out the exogenous impact of imperfectly enforced liberalization on the regional regulatory environment.

The results are presented in Table 4. For each regulatory area, we present three regressions: first stage, second stage, and OLS (for comparison). As reflected in the magnitude of F-statistics reported at the bottom of the table, the instruments are sufficiently strong not to worry about a weak instrument problem.[28] The second stage estimations on the sample of established firms reveal that liberalization of licensing and inspections leads to a significantly higher sales growth in established firms (see columns 3 and 6 of the table). As for liberalization of registration, we run regressions on the sample of newly registered firms and find no statistically significant relationship between their sales growth and liberalization of registration. Since, theoretically, there is no unambiguous relationship between sales growth of firms once they enter, and the liberalization of entry, it is more reasonable to expect liberalization of entry to affect new firm entry rather than firm performance. In the next subsection, we turn to the estimation of the effect of liberalization of registration on entry and the share of small business employment.

The direction and size of the bias in OLS estimates can be seen from the comparison of 2SLS and OLS estimates. In all regressions, OLS estimates are smaller than 2SLS estimates. This points to a negative and rather large bias in the OLS estimates, which is consistent

[27] As the model is over-identified (we have several potential instruments and one endogenous regressor), we test for the validity of over-identification restrictions and find that Hansen's J-test does not reject the null hypothesis of the validity of restrictions.

[28] In every case, we choose a particular institutional measure to be used for the instrument in order to maximize the F-statistic for the excluded instrument in the first stage.

with the view that predatory regulators are attracted to more vibrant and growing business.[29]

3.4.2 Small business entry and employment

To test the relationship between liberalization reforms and entry, we run the following regression on a panel of regions:

$$S_{rt} = \xi \bar{L}_{rt} + \zeta' \mathbf{Z}_{rt} + \phi_r + \rho_t + \varepsilon_{rt}. \tag{3.5}$$

The dependent variable (S_{rt}) stands for one of the following regional outcomes: net entry into the official sector (measured by the log number of small businesses) and official small business employment share (measured by the number of employees in the small business sector over the labor force). These variables are summarized in Panel B of Table A2. They come from the official Russian statistical agency *Rosstat*, and are available for all regions annually up to and including 2004, i.e., for the period from the first to the fifth round of the survey. There are no reliable data on the size of the unofficial sector.

$\bar{L}_{rt}$ stands for a *regional-level* measure of attainment of liberalization targets. We construct regional-level measures by aggregating firm-level regulation measures across firms in the same region and round. The aggregation takes two steps. First, we partial out the effect of basic firm characteristics ($\mathbf{X}_{ft}$) from L_{ft} by taking residuals of the OLS regression: $L_{ft} = \lambda' \mathbf{X}_{ft} + \varepsilon_{ft}$. Second, we take simple averages of these residuals by region in each round of the survey: $\bar{L}_{rt} = \frac{1}{N} \sum_{f=1}^{N} \hat{L}_{ft}$, where N is the number of firms in each region and round.[30] The rest of the notation is as above.

[29]Such endogeneity of regulation can explain why Klapper, Laeven and Rajan (2006) find that more benign entry regulations are not associated with higher entry in corrupt countries, whereas there is a strong relationship in uncorrupt countries.

[30]The use of firm employment to construct regional regulation measures could introduce a simultaneity problem if regional and firm employment co-vary. However, the point estimates of ξ in the second stage remain unchanged if we construct regional regulation measures as simple averages without controlling for firm characteristics. As a baseline, we control for firm characteristics because this increases the power of the instruments in the first stage. In addition, the results are robust to using region*round fixed effects rather than averages of residuals to aggregate regulation measures.

Since our outcome variables are measured at the regional rather than firm level, and, therefore, the number of observations declines dramatically, we cannot include many other control variables that potentially could have an effect on the outcomes. Particularly, in the case of the analysis of firm entry, region-specific trends are important to control for trends in entry that might be correlated with institutional environment. However, we do not have enough statistical power to include these controls here. Thus, one should treat the results of this aggregated analysis with caution, and mostly focus on firm-level evidence. We verified, however, that the results are robust to controlling for regional averages of the firm-level controls used in the firm-level analysis and for using plain regional averages of firm-level measures of attainment of liberalization targets as regional level measures. In addition, as in Aghion et al. (2008), we control for the average technological level of firms with the average regional labor productivity.

The first stage is an aggregation of equation 3.2 to the regional level:

$$\bar{L}_{rt} = \alpha I_r A_t + \mu' \mathbf{Z}_{rt} + \phi_r + \rho_t + \varepsilon_{rt}. \tag{3.6}$$

Table A4 in the Appendix reports the first stage along with the results of the F-test. For each regression, we use instruments that maximize F-statistics for the excluded instruments in the first stage. For registration and licensing, the instruments are sufficiently strong, whereas for inspections the instruments are weak and, therefore, the second stage results for inspections may be biased due to the weak instrument problem (we use criteria for weak instruments from Stock, Wright and Yogo, 2002).[31] In order to deal with the weak instruments problem, we report two sets of standard errors for our estimates in the second stage: the conventional robust standard errors and standard errors calculated using the conditional likelihood ratio approach developed by Moreira (2003) and Andrews, Moreira and Stock (2007), which gives reliable confidence intervals in the case of weak instruments.

The results are presented in Table 5. Our main focus here is on the effects of liberalizing registration. We find no statistically significant effect of liberalizing registration on the number of small businesses, and a large, statistically significant, and robust effect on small

[31] Again, the Hansen's J-test does not reject the null hypothesis of the validity of identifying restrictions.

business employment as a share of labor force. As far as the effects of liberalization of licenses and inspections is concerned, second stage estimates yield significant positive effects of liberalization of inspections and de-licensing on the number of small businesses, and also of liberalization of inspections on the small business employment. Yet, only the effect of liberalization of inspections on the number of small businesses is robust to using the conservative standard errors calculated using the conditional likelihood ratio approach.

As with the firm-level regressions presented in Table 5, in all regional-level regressions the bias in OLS estimates is negative. This provides additional evidence in favor of the public choice theory of regulations, as predatory regulators are drawn disproportionately to regions with higher small business entry into the official sector.

3.4.3 Magnitude: the outcomes of liberalization in Amur and Samara regions

Let us illustrate the magnitude of the results on the effect of liberalization reform on outcomes using the example of the Amur and Samara regions. First, our results suggest that because of the differences in the enforcement of reform under different institutional environments, the performance of small firms is expected to be affected in a drastically different way by liberalization of inspections and licenses in regions with "good" versus "bad" governance institutions. In a region with an institutional environment similar to the Samara region, where the reform is relatively well enforced, liberalization of inspections leads to a 12 percentage point increase in sales growth for average small firms (an amount equal to one fifth of the standard deviation of sales growth across firms). In contrast, in a region with institutions at a level similar to the Amur region, growth of sales is expected to rise only by 1.6 percentage points as a result of liberalization of inspections as it does not translate into a significant change in the actual regulatory level. Even more striking is the difference in the effects of de-licensing reform in the two types of regions: growth in sales following de-licensing is expected to increase by 4.5 percentage points in a region with an institutional environment comparable to that of Samara, and decrease by

4.5 percentage points in a region with an institutional environment comparable to that of Amur. Again, these differences are driven by the differences in enforcement of the reform across regions.

In addition, our regional-level results show that in a region similar to Samara in terms of institutional environment, the liberalization of registration leads to an increase in the share of small business employment by 1 percentage point (which is approximately one quarter of its standard deviation). In contrast, in a region similar to Amur, the share of employment of small businesses is expected to fall by 0.3 percentage points as a result of reform, because the liberalization of registration is expected to be poorly enforced.[32]

3.5 Robustness

In this section, we describe various robustness checks for our baseline results.

Placebo experiment. One could argue that, independent of the reform, different regional institutional environments may be associated with different trends in regulation levels. In addition, standard errors in the dif-in-dif regressions may be biased downwards due to residual autocorrelation (Bertrand, Du o and Mullainathan, 2004). To address these concerns, we conduct a placebo experiment. We consider all possible combinations of dates for liberalization in the three areas of regulation, such that these dates are different from the dates of the real liberalization reforms. If these dates happen to be after the dates of the real reforms they are at least two rounds away from the enactment of the actual liberalization laws. We take these combinations as the dates of the enactment of placebo laws in placebo reform packages. We exclude dates for placebo laws that take place two

[32]Since one possible goal of licensing and inspection authorities may be to curb pollution and monitor product quality, we also tried to check if liberalization had an effect on pollution (measured by log emissions of contaminants into the atmosphere in a region in a year) and public health (measured by morbidity from injuries and poisoning per 1,000 people in a region in a year). We used the same methodology as for regional-level regressions on entry of small businesses. We find no robust effects (not reported). The public health and pollution variables, however, may be poorly measured and the regulations we consider may aim at resolving other market failures. Thus, one should treat the evidence of no relationship between pollution and morbidity, on the one hand, and regulations on the other hand, merely as suggestive.

rounds after the actual law because of a possible delay in the implementation of the real laws. Altogether there are 140 such combinations. For these placebo liberalizations, we re-ran all regressions in Tables 2-3. All combinations of placebo timing for each liberalization law and institutional measures yield 140 regressions. We find a significant positive effect of institutional measures in facilitating enforcement of placebo liberalizations at the 10% level in 10% of all placebo regressions (14 cases), at the 5% level — in 5.7% of placebo regressions (8 cases), and at the 1% level in 2.1% of regressions (3 cases). In addition, there is one placebo regression (0.7% of all placebo regressions), which α is negative and significant at the 5% level. Thus, although it could be the case that the standard errors are slightly biased downwards as the share of significant coefficients is slightly higher than one should expect in the case of ideal identification, this cannot explain the strong and robust effects that we find for the real laws. Figure A2 in the Appendix provides a graphical illustration of our placebo experiment by plotting the coefficients (along with their confidence intervals) on the interaction between institutional measures and lags and leads of $AFTER$ in a specification similar to Equation 3.3, which is a subset of our placebo regressions. On the horizontal axis, we plot the placebo timing such that zero coincides with the timing of the actual liberalization; -1 is as if liberalization occurred one round before the actual liberalization; +1 is as if liberalization occurred one round after the actually liberalization, etc. The figure shows that for the industrial concentration and fiscal incentives, only the coefficients on the interaction with the actual timing of the laws are statistically significant. For the transparency of government, in addition to the interaction with the actual timing of liberalization, the interaction with the lead of $AFTER$ is also statistically significant, suggesting a somewhat sluggish implementation of liberalization. The plot of coefficients on interaction with Internet penetration is more problematic, however, as we find a significant effect of the interactions with two leads of $AFTER$, in contrast to all other institutional measures. It is worth noting that we do not use Internet penetration to construct instruments in our analysis of the effect of liberalization on outcomes.

Region-specific linear trends. In order to make sure that our results are not driven by region-specific trends, we also re-ran regression equations 3.1, 3.2, and 3.3 with region-specific linear trends as additional regressors. The direction of the estimated effects of institutions on the reform progress remains the same (α coefficients remain positive), the magnitude of the effects decreases slightly in some cases, and the magnitude of the standard errors increases quite substantially, but in the majority of regressions the coefficients of interest remain statistically significant. To be more precise, in Tables 2a and 2b, we report 20 regressions estimating γ coefficients (5 regressions for each institutional measure), in which we find significant effects of the considered institutions on the reform progress in 18 (90%) of these regressions. Once we include region-specific linear trends, significance is preserved in 14 (70%) of all regressions. In the vast majority of the cases, the statistical significance is lost because of an increase in the standard errors rather than a decrease in the magnitude of point estimates. This suggests that our baseline specification does not suffer from the omitted variable bias. The most vulnerable result to the inclusion of region-specific trends turns out to be the effect of Internet penetration, which remains significant only in 2 out of 5 regressions. One should note, however, that many of the alternative measures of access to independent media remain significant after controlling for region-specific trends. Thus, the results are qualitatively the same, but become somewhat weaker statistically with the inclusion of region-specific trends. This, however, is to be expected considering that we have only 6 time periods. In addition, we verified that our results are not driven by any particular region-outlier.

Overall, our results prove to be robust.

3.6 Conclusions

We study the outcomes of a drastic national liberalization reform of inspections, licenses, and registration in Russia. We find that liberalization had positive effects on firm performance and small business employment in regions with good governance institutions, and no or even negative effects in regions with bad governance institutions. We also find that the channel for the unequal effects of liberalization in regions with different institutional environments is the differential enforcement of liberalization. In regions with higher transparency of government, better access of the public to independent sources of information, more concentrated large businesses, and better fiscal incentives, the liberalization reform was better enforced and led to a significantly higher drop in the actual regulatory burden, and as a result, better outcomes.

Our evidence that regions with transparent governments and more informed populations are the ones that achieve better progress in liberalizing regulation supports the public choice theory of regulations.

3.7 Tables

Figure 1. The Timing and Content of Liberalization Reform and Rounds of MABS Survey

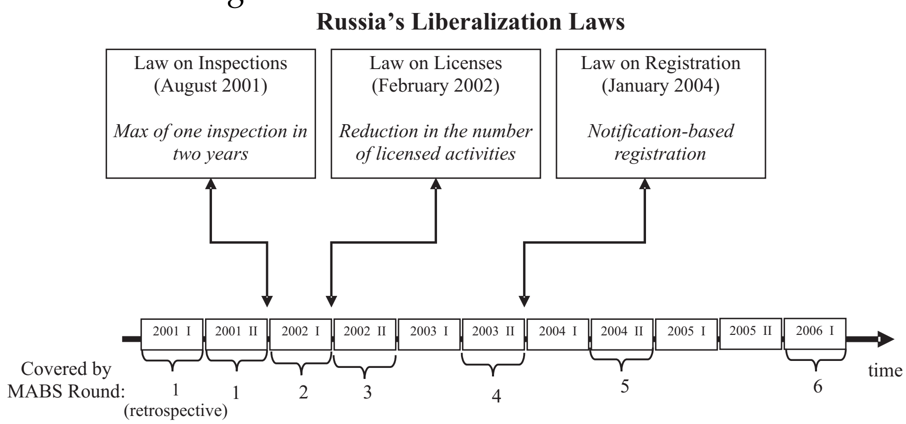
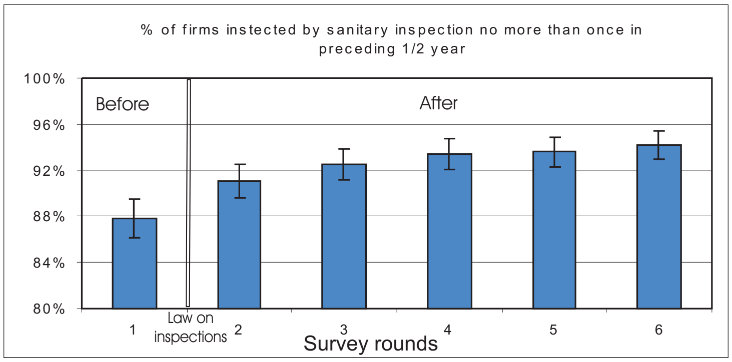
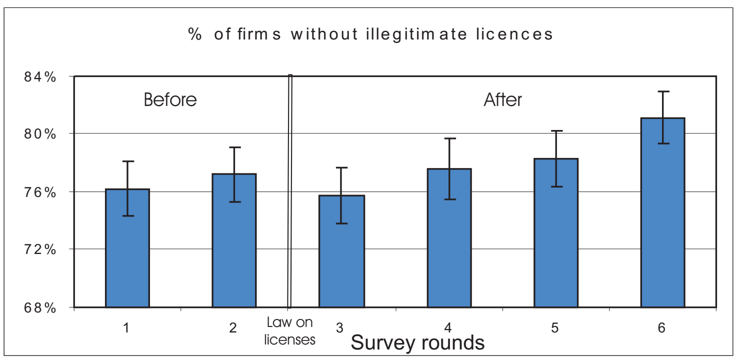


Figure 2. Attainment of reform targets before and after liberalization

Inspections



Lincenses



Registration

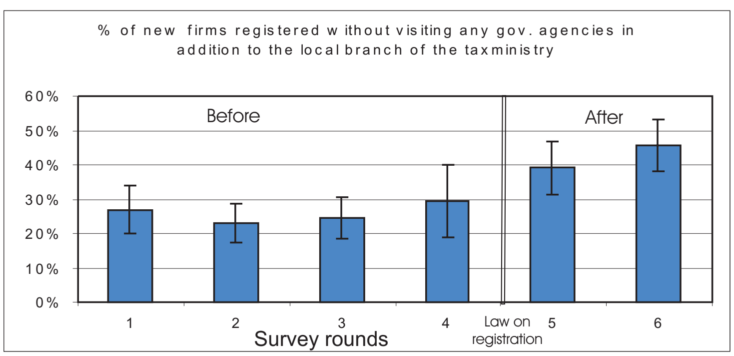


Table 1. Summary statistics for measures of compliance with liberalization targets

	Before reform					After reform				
Dummies for meeting liberalization targets:										
	Obs.	Mean	SD	SE	# of rounds	Obs.	Mean	SD	SE	# of rounds
inspections	1534	0.878	0.327	0.008	1	7512	0.929	0.257	0.003	5
licensing	3942	0.767	0.423	0.007	2	7648	0.792	0.406	0.005	4
registration	688	0.251	0.434	0.017	4	343	0.426	0.495	0.027	2

Table 2a. Enforcement of liberalization and institutions

	Meeting the liberalization targets for new firms, region fixed effects									
	Registration					Licensing				
Transparency	0.024				0.008	0.009				0.004
x AFTER	[0.010]**				[0.009]	[0.004]**				[0.003]
Ind. concentr.		0.935			0.323		0.486			0.543
x AFTER		[0.30]***			[0.240]		[0.09]***			[0.135]***
Internet			0.039		0.024			0.015		0.012
x AFTER			[0.019]**		[0.014]*			[0.005]***		[0.004]***
Fiscal incentives				0.844	0.664				-0.099	-0.285
x AFTER				[0.23]***	[0.22]***				[0.095]	[0.071]***
AFTER					0.203					0.059
					[0.081]**					[0.027]**
Initial level	-0.835	-0.614	-0.651	-0.655	-0.655	-0.873	-0.785	-0.887	-0.855	-0.767
x AFTER	[0.18]***	[0.18]***	[0.17]***	[0.16]***	[0.18]***	[0.09]***	[0.07]***	[0.12]***	[0.12]***	[0.08]***
Round FE	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	No
Linear trend	No	No	No	No	Yes	No	No	No	No	Yes
Observations	809	809	809	809	809	2015	2015	2015	2015	2015
R-squared	0.2	0.2	0.2	0.2	0.2	0.08	0.09	0.08	0.08	0.09
F-stat	6.13	9.46	4.24	14.07	4.17	6.1	27.89	8.6	1.09	9.94
p-value	0.02	0.00	0.05	0.00	0.01	0.02	0.00	0.01	0.30	0.00

Note: Robust standard errors adjusted for clusters within region before and after liberalization are in brackets.

*significant at 10%; **significant at 5%; ***significant at 1%. All regressions include firm-level and regional-level controls.

Table 2b. Enforcement of liberalization and institutions

	Meeting the liberalization targets for established firms, firm fixed effects									
	Inspections					Licensing				
Transparency	0.003				0.001	0.007				0.003
x AFTER	[0.002]*				[0.002]	[0.002]***				[0.004]
Ind. concentr.		0.162			0.127	0.007				0.237
x AFTER		[0.064]**			[0.064]*	[0.002]***				[0.129]*
Internet			0.004		0.002		0.291			0.006
x AFTER			[0.002]*		[0.002]		[0.114]**			[0.005]
Fiscal incentives				0.126	0.087			0.009		0.023
x AFTER				[0.05]***	[0.040]**			[0.004]**		[0.083]
AFTER					0.03				0.098	-0.029
					[0.015]**				[0.081]	[0.023]
Initial level	-0.778	-0.778	-0.779	-0.78	-0.78	-0.636	-0.635	-0.635	-0.639	-0.632
x AFTER	[0.03]***	[0.03]***	[0.03]***	[0.03]***	[0.03]***	[0.03]***	[0.03]***	[0.03]***	[0.03]***	[0.03]***
Round FE	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	No
Linear trend	No	No	No	No	Yes	No	No	No	No	Yes
Observations	5305	5305	5305	5305	5305	6594	6594	6594	6594	6594
R-squared	0.65	0.65	0.65	0.65	0.65	0.57	0.57	0.57	0.57	0.57
F-stat	3.26	6.44	3.85	7.92	3.54	8.08	6.57	5.9	1.46	4.73
p-value	0.08	0.02	0.06	0.01	0.01	0.01	0.01	0.02	0.23	0.00

Note: Robust standard errors adjusted for clusters within region before and after liberalization are in brackets.

*significant at 10%; **significant at 5%; ***significant at 1%. All regressions include firm-level and regional-level controls.

Table 3 Average enforcement across all regulations and institutions

	Meeting the liberalization targets, average across regulatory areas				
Transparency x AFTER	0.008				0.004
	[0.001]***				[0.002]**
Ind. concentr x AFTER		0.294			0.22
		[0.050]***			[0.054]***
Internet x AFTER			0.009		0.005
			[0.002]***		[0.002]*
Fiscal incentives x AFTER				0.111	0.02
				[0.043]**	[0.033]
AFTER	0.089	0.083	0.087	0.091	0.068
	[0.019]***	[0.021]***	[0.022]***	[0.023]***	[0.020]***
Initial regulation x AFTER	-0.407	-0.366	-0.393	-0.418	-0.363
	[0.078]***	[0.082]***	[0.083]***	[0.087]***	[0.073]***
Firm and region controls	Yes	Yes	Yes	Yes	Yes
Region*Regulation FE	Yes	Yes	Yes	Yes	Yes
Round FE	Yes	Yes	Yes	Yes	Yes
Regulation specific trends	Yes	Yes	Yes	Yes	Yes
Observations	21219	21219	21219	21219	21219
R-squared	0.19	0.19	0.19	0.19	0.19
Number of clusters	120	120	120	120	120
F-test	37.37	34.31	23.14	6.69	29.88
p-val	0.00	0.00	0.00	0.01	0.00

Note: Robust standard errors adjusted for clusters within region before and after the reform separately for each regulation in brackets; *significant at 10%; **significant at 5%;***significant at 1%.

Table 4.Firm performance and liberalization

	First	Change in sales		First	Change in sales		First	Change in sales	
	stage	2SLS	OLS	stage	2SLS	OLS	stage	2SLS	OLS
Meeting liberalization targets:									
Licensing		1.062	-0.0001						
		[0.51]**	[0.03]						
Inspections					1.554	0.065			
					[0.91]*	[0.04]*			
Registration								0.181	0.106
								[0.23]	[0.06]*
Institution	0.010			0.297			0.066		
x AFTER	[0.003]***			[0.10]***			[0.02]***		
Initial level	-0.475	0.493	-0.015	-0.764	1.205	0.067	-0.801	0.522	0.497
x AFTER	[0.03]***	[0.25]**	[0.05]	[0.02]***	[0.70]*	[0.08]	[0.48]*	[0.33]	[0.32]
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No
Round FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5860	5318	5860	4826	4474	4826	340	340	340
R-squared	0.08		0.54	0.26		0.48	0.16		0.09
F-stat, instrument	11.85	11.85		8.8	8.8		10.24	10.24	
Number of firms	2170	1628	2170	1676	1324	1676	340	340	340
Instrument	Transparency x AFTER			Ind. concentr x AFTER			Transparency x AFTER		

Note: Robust standard errors adjusted for clusters within region before and after liberalization are in brackets. *significant at 10%; **significant at 5%; ***significant at 1%. All regressions include firm-level and regional-level controls. Regressions for licensing and inspections are estimated on the sample of established firms, regressions for registration are estimated on the sample of newly registered firms.

Table 5a. Liberalization and Entry of Small Businesses

	Log of total number of small businesses					
	2SLS	OLS	2SLS	OLS	2SLS	OLS
Meeting liberalization target,	0.280	0.009				
registration	[0.214]	[0.069]				
	(0.195)					
Meeting liberalization target,			0.768	0.016		
licensing			[0.359]**	[0.203]		
			(0.530)			
Meeting liberalization target,					2.226	0.391
inspections					[0.775]***	[0.405]
					*(1.108)***	
Log (population)	0.116	0.523	0.471	0.327	0.419	0.341
	[1.057]	[1.210]	[0.395]	[0.403]	[0.397]	[0.492]
Log (labor productivity)	-64.502	-61.417	-57.362	-56.117	-39.922	-53.253
	[10.50]***	[9.37]***	[8.83]***	[9.79]***	[12.93]***	[12.31]***
Round and Region FE	Yes	Yes	Yes	Yes	Yes	Yes
R-sqrd	0.99	0.99	0.99	0.99	0.99	0.99
Observations	84	84	99	99	99	99
Number of clusters	40	40	40	40	40	40
F-stat for the instrument (	10.38		11.93		6.62	

Table 5b. Liberalization and Entry of Small Businesses

	Share of small business employment					
	2SLS	OLS	2SLS	OLS	2SLS	OLS
Meeting liberalization target,	0.016	0.007				
registration	[0.008]**	[0.003]**				
	*(0.008)**					
Meeting liberalization target,			0.001	-0.019		
licensing			[0.027]	[0.009]**		
			(0.027)			
Meeting liberalization target,					0.083	0.035
inspections					[0.044]*	[0.026]
					(0.056)	
Log (population)	-0.057	-0.044	-0.092	-0.096	-0.089	-0.091
	[0.039]	[0.040]	[0.027]***	[0.031]***	[0.041]**	[0.051]*
Log (labor productivity)	-0.304	-0.203	-0.469	-0.436	0.133	-0.215
	[0.343]	[0.379]	[0.403]	[0.438]	[0.503]	[0.535]
Round and Region FE	Yes	Yes	Yes	Yes	Yes	Yes
R-sqrd	0.99	0.98	0.98	0.98	0.98	0.98
Observations	84	84	99	99	99	99
Number of clusters	40	40	40	40	40	40
F-stat for the instrument	10.38		11.93		6.62	

Note: Robust standard errors in brackets and normal font; Standard errors adjusted for weak instrument bias in parentheses and italics; *significant at 10%; **significant at 5%; ***significant at 1%.

Appendix: Data on institutional determinants

Summary statistics for all institutional measures are presented in Panel A of Table A1a

Government transparency

As a measure of government transparency, we use the overall index of transparency of regional authorities constructed by an independent informational agency "Strana.ru" and an independent association of journalists "Media Soyuz." This is a composite of indices of transparency of different branches of regional government. The results using these branch-specific indices are very similar. The indices were constructed on the basis of a survey of more than a thousand prominent regional journalists who were asked to evaluate performance of the regions along the following dimensions: accessibility and accuracy of information about decisions of a particular regional authority, impartiality and easiness of journalist accreditation rules, quickness of response on journalist inquiries, presence and quality of internet site, etc.

The transparency ratings are available at www.strana.ru/print/128316.html.

Independent media

In the baseline analysis, we use internet penetration in the region measured by the number of personal computers connected to internet per 100 employees as a proxy for the access of the public to independent media. This variable comes from the official Russia's statistical agency (Rosstat). In addition, we verify that the results are robust to using two alternative measures: a dummy that indicates regions with non-zero subscription to the two main independent (in 2000) daily newspapers, i.e., *Kommersant* and *Vedomosti* and a dummy for availability of the signal in the region of the largest independent radio station, i.e., *Echo Moscow*. The sources of these data are the websites of the respective media outlets: www.kommersant.ru, www.vedomosti.ru, and www.echo.msk.ru.

Monitoring by large businesses

We use the concentration (Herfindahl-Hirschman) index of employment among industrial firms in each region as a proxy for monitoring of regional public officials by large businesses. The source of these data is the Russia's Industrial Registry. We verify that our results are robust to using two alternative measures, which also reflect political power of large regional firms. The first alternative is the concentration (Herfindahl-Hirschman) index of sales among industrial firms in each region (this measure is also from the Russia's Industrial Registry). The second alternative is a measure of regional regulatory capture constructed by and described in (Slinko et all, 2005). This is the concentration of preferential treatments (i.e., subsidies, tax breaks, etc.) given to large firms in each region by the regional laws and regulations. It reflects the extent to which political power is concentrated in the hands of a few large firms.

Fiscal incentives

The share of own budgetary revenues in the total regional budget is used as a simple (and rather crude) proxy for the regional fiscal incentives. The data come from the Treasury of the Russian Federation (www.roskazna.ru/reports/mb.html).

Table A1a. Summary statistics for institutional measures, outcomes, and controls

Panel A: Institutional determinants	Obs	Mean	Std. Dev.				
Transparency of authorities	20	7.478	4.014				
Industrial concentration	20	0.178	0.077				
Internet penetration	20	4.808	3.181				
Fiscal incentives	20	0.829	0.117				
Panel B: Outcomes	Obs	Mean	Std. Dev.	2001 (before)		2004 (after)	
				Mean	Std. Dev.	Mean	Std. Dev.
Change in sales	6990	0.131	0.600	0.169	0.722	0.124	0.565
Small business employment share	99	0.053	0.038	0.049	0.037	0.056	0.035
Log number of small businesses	99	2.559	1.133	2.541	1.116	2.608	1.164
Panel C: Controls	Obs	Mean	Std. Dev.	2001 (before)		2004 (after)	
				Mean	Std. Dev.	Mean	Std. Dev.
New firm dummy	11245	0.184	0.388	0.194	0.396	0.197	0.398
Firms age	11218	1.384	0.940	1.367	0.900	1.420	0.959
State firm dummy	11590	0.018	0.132	0.016	0.127	0.015	0.122
Log(1+Firms size)	11163	2.414	1.017	2.363	0.975	2.402	1.043
Log(1+Firms size) squared	11163	6.860	5.558	6.535	5.195	6.858	5.707
Firms size	11163	19.088	31.135	17.258	27.64	19.516	33.38
Log (population)	99	7.740	0.712	7.755	0.702	7.725	0.755
Log (mean pc income)	99	7.751	0.457	7.593	0.461	7.888	0.471

Table A1b. Summary statistics for institutional measures, outcomes, and controls

Panel C- continued: Controls	Obs	Mean	Std. Dev.	2001 (before)		2004 (after)	
Industry dummies				Mean	Std. Dev.	Mean	Std. Dev.
1. Manufacturing	11222	0.103	0.304	0.139	0.346	0.083	0.276
2. Services	11222	0.306	0.461	0.212	0.409	0.343	0.475
3. Commerce	11222	0.297	0.457	0.289	0.453	0.304	0.460
4. Agriculture, hunting, fishing	11222	0.011	0.103	0.017	0.129	0.008	0.087
5. Construction	11222	0.106	0.308	0.112	0.315	0.085	0.280
6. Food industry	11222	0.030	0.171	0.041	0.199	0.023	0.149
7. Science intensive technologies	11222	0.060	0.238	0.077	0.267	0.061	0.239
8. Other	11222	0.087	0.282	0.113	0.317	0.094	0.291
Legal form dummies							
1. Person-entrepreneur	11243	0.012	0.109	0.015	0.121	0.014	0.116
2. Private entreprise	11243	0.026	0.160	0.025	0.155	0.021	0.142
3. Federal state enterprise	11243	0.004	0.060	0.002	0.045	0.003	0.057
4. Regional state enterprise	11243	0.002	0.043	0.002	0.039	0.002	0.047
5. Municipal state enterprise	11243	0.004	0.067	0.003	0.051	0.004	0.066
6. Partnership	11243	0.004	0.062	0.004	0.064	0.006	0.077
7. Partnership limited	11243	0.804	0.397	0.786	0.410	0.822	0.383
8. Cooperative	11243	0.009	0.095	0.010	0.098	0.007	0.081
9. Closed cooperative	11243	0.092	0.289	0.113	0.317	0.080	0.271
10. Open cooperative	11243	0.027	0.163	0.030	0.171	0.026	0.158
11. Joint venture	11243	0.000	0.019	0.001	0.023	0.000	0.000
12. Subsidiary	11243	0.000	0.009	0.000	0.000	0.000	0.000
13. Other	11243	0.015	0.121	0.011	0.103	0.016	0.127

Table A2. Correlation between institutional measures

	Transparency	Ind. concentr	Internet
Ind. concentr	0.288	1	
	(0.000)		
Internet	0.573	0.019	1
	(0.000)	*(0.039)*	
Fiscal incentives	0.345	0.199	0.198
	(0.000)	*(0.000)*	*(0.000)*

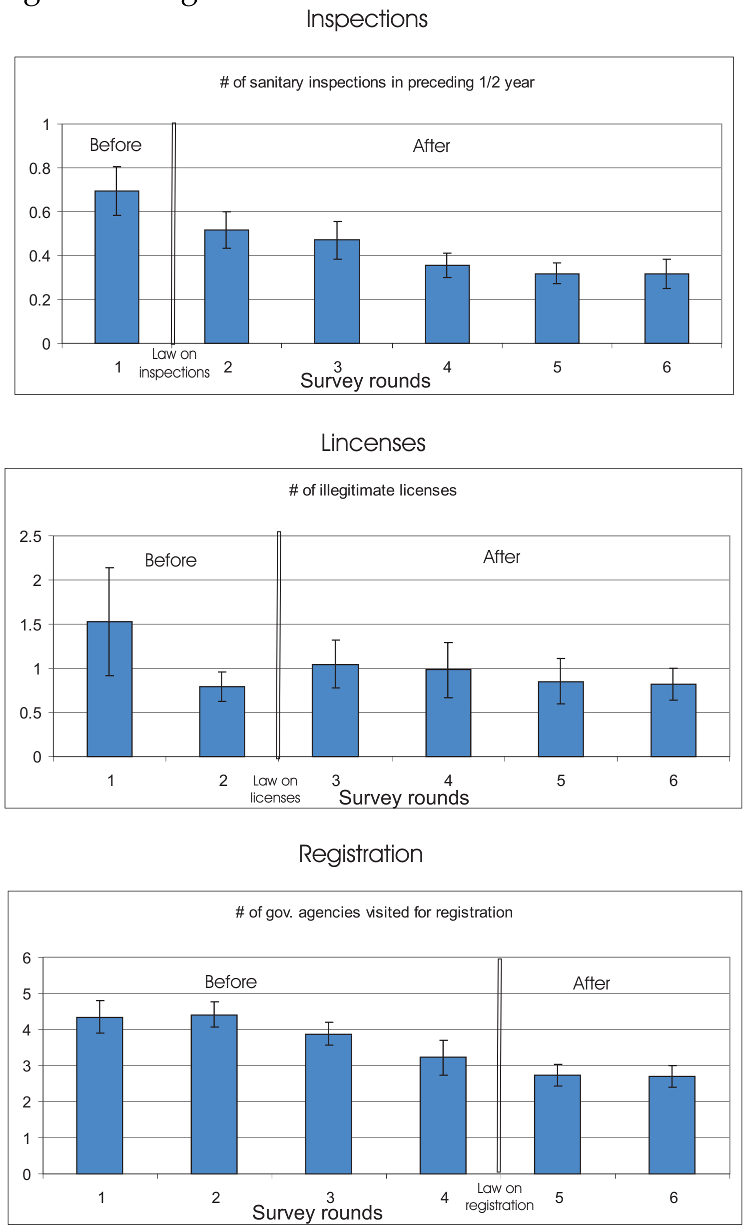
Note: p-values for pair-wise correlations in parentheses.

Table A3. The first stage

	Meeting liberalization targets		
	Registration	Licensing	Inspections
Fiscal incentives x AFTER	1.266		
	[0.403]***		
Transparency x AFTER		0.012	0.003
		[0.003]***	[0.002]*
Initial level x AFTER			-0.531
			[0.162]***
Log (population)	0.469	-0.467	-0.115
	[1.349]	[0.224]**	[0.100]
Log (labor productivity)	-0.013	0.007	-0.073
	[0.255]	[0.081]	[0.037]*
Round and Region FE	Yes	Yes	Yes
Observations	84	99	99
F-stat for instruments (1st stage)	9.86	11.97	6.64

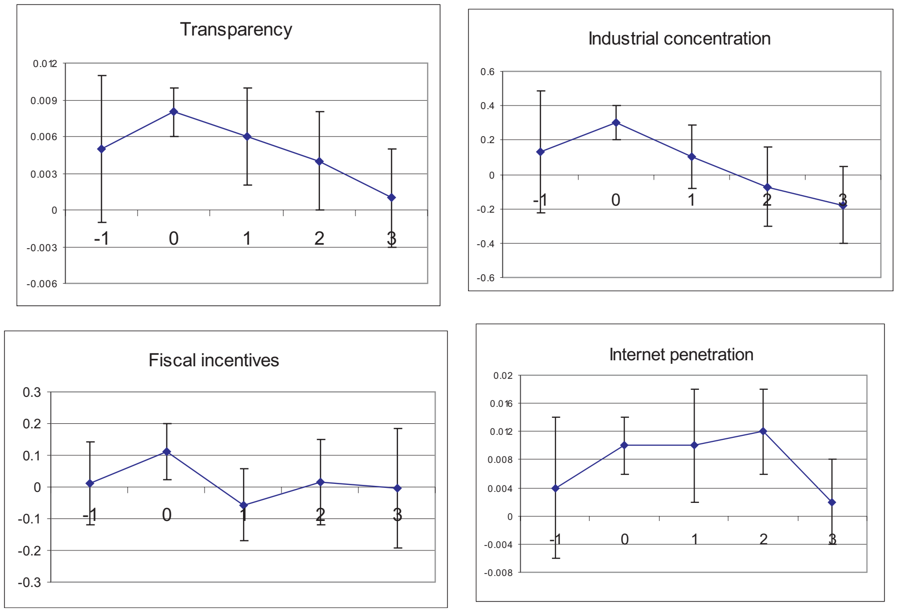
Note: Robust standard errors in brackets; *significant at 10%; **significant at 5%; ***significant at 1%. The choice of a particular set of instruments is guided by maximization of the F-statistic for the excluded instruments.

Figure A1.Regulation level before and after liberalization



bb

Figure A2.Interaction between lags and leads of $AFTER$ with institutional measures



Bibliography

Aghion, Philippe, Robin Burgess, Stephen Redding and Fabrizio Zilibotti. 2005. “Entry Liberalization and Inequality in Industrial Performance." Journal of the European Economic Association 3(2-3):291-302.

Aghion, Philippe, Robin Burgess, Stephen Redding and Fabrizio Zilibotti. 2008. "The Unequal Effects of Liberalization: Evidence from Dismantling the License Raj in India." American Economic Review 98(4):1397-1412.

Aguirregabiria, V., and P. Mira, 2007, “Sequential estimation of dynamic discrete games," Econometrica, 75(1), 1.

Aguirregabiria V., and P. Mira, 2010, “Dynamic discrete choice structural models: A survey”, Journal of Econometrics, 156, (2010) 38-67

Ai, C., and X. Chen, 2003, “Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions,” Econometrica, 71(6), 1795–1843.

Akerlof, G. A. and Kranton, R. E., 2000. "Economics And Identity," The Quarterly Journal of Economics, MIT Press, vol. 115(3), pages 715-753, August

Andrews, Donald W.K., Marcelo J. Moreira and James Stock. 2007. "Performance of Conditional Wald Tests in IV Regression with Weak Instruments." Journal of Econometrics (39):116-132.

Andrienko Y., and A. Nemtsov, 2006, “Estimation of Individual Demand for Alcohol”, CEFIR working paper.

Andrienko Y., and S. Guriev, 2004, "Determinants of inter-regional mobility in Russia," The Economics of Transition, vol. 12(1), pages 1-27, 03.

Arcidiacono, P., Sieg, H., Sloan, F., 2007. Living rationally under the volcano? An empirical analysis of heavy drinking and smoking. International Economic Review 48, 3765.

Auld, M.C. and P. Grootendorst, 2004 "An empirical analysis of milk addiction". Journal of Health Economics, 23, 1117-33.

Bajari, P., V. Chernozhukov, H. Hong, and D. Nekipelov, 2007 "Semiparametric estimation of a Dynamic Game of Incomplete Information," working paper, University of Michigan and Stanford University

Bajari, P., Hong, H., Krainer J. and Nekipelov D., 2011, "Estimating Static Models of Strategic Interactions", forthcoming, Journal of Business and Economic Statistics

Bajari, P., Hong, H., and Nekipelov D., 2011, "Game Theory and Econometrics: A Survey of Some Recent Research", mimeo

Becker, G., Accounting for tastes. Harvard University Press, Cambridge, Massachusetts.

Becker, G. and Murphy, K., 1988. "A Theory of Rational Addiction", Journal of Political Economy, 96 (4), 675

Becker, G., Grossman, M., and Murphy, K., 1991. Rational Addiction and the Effect of Price on Consumption. American Economic Review, 81 (2), 237

Beenstock, Michael, and Giora Rahav, "Testing Gateway Theory: do cigarette prices affect illicit drug use?", Journal of Health Economics 21 (2002) 679–698

Berry, S., Ostrovsky, T. and Pakes, A., 2007, "Simple Estimators for the Parameters of Discrete Dynamic Games (with Entry/Exit Examples)", RAND Journal of Economics, Summer.

Bertrand, Marianne, Esther Dufflo and Sendhil Mullainathan. 2004. "How Much Should We Trust Differences-in-Differences Estimates?" The Quarterly Journal of Economics 119(1):249-275.

Bhattacharya, J., C. Gathmann, C., and G. Miller, 2011, "The Gorbachev Anti-Alcohol Campaign and Russia's Mortality Crisis"

Blume, L. and S. Durlauf. 2005. "Identifying Social Interactions: A Review. " In Methods in Social Epidemiology, J. M. Oakes and J. Kaufman, eds. San Francisco: Jossey-Bass

Botero, Juan, Simeon Djankov, Rafael La Porta, Florencio Lopez-de Silanes and Andrei Shleifer. 2004. "The Regulation of Labor." The Quarterly Journal of Economics 119(4):1339- 1382.

Brainerd,E and D. Cutler, 2005. "Autopsy on an Empire: Understanding Mortality in Russia and the Former Soviet Union," Journal of Economic Perspectives, American Economic Association, vol. 19(1), pages 107-130, Winter

Bruhn, Miriam. 2007. "License to Sell: The Effect of Business Registration Reform on Entrepreneurial Activity in Mexico." Department of Economics, MIT.

Card D., and L. Giuliano "Peer Effects and Multiple Equilibria in the Risky Behavior of Friends", NBER Working Paper No. 17088 May 2011

Chaloupka, F. J., "Rational Addictive Behavior and Cigarette Smoking," Journal of Political Economy, XCIX (1991), 722–742

Chari, Amalavoyal. 2007. "License Reform in India: Theory and Evidence." Department of Economics, Yale University, New Haven, Connecticut. 36

Chenet, L., M. McKee, D. Leon, V. Shkolnikov and S. Vassin. 1998. "Alcohol and cardiovascular mortality in Moscow; new evidence of a causal association," Journal of Epidemiology and Community Health, Vol 52, 772-774

Cook, Philip J. & Moore, Michael J., 2000. "Alcohol," Handbook of Health Economics, in: A. J. Culyer & J. P. Newhouse (ed.), Handbook of Health Economics, edition 1, volume 1, chapter 3 Community Health, Vol 52, 772-774

Cook and Moor, 1995, "Habits and heterogeneity in the youthful demand for alcohol", NBER working paper #5152, 1995

Denisova. I., 2010, "Adult mortality in Russia: a microanalysis", Economics of Transition, Vol. 18(2), 2010, 333-363

Deza, Monica "Is There a Stepping-Stone Sect in Drug Use? Separating State Dependence from Unobserved Heterogeneity Within and Across Illicit Drugs", Job Market Paper, UC Berkeley, 2012

Djankov, Simeon. 2009. "The Regulation of Entry: A Survey." World Bank Research Observer 24(2):183-203.

Djankov, Simeon, Caralee McLiesh and Rita Ramalho. 2006. "Regulation and Growth." Economics Letters 92(3):395-401.

Djankov, Simeon, Rafael La Porta, Florencio Lopez-de Silanes and Andrei Shleifer. 2002. "The Regulation of Entry." The Quarterly Journal of Economics 117(1):1-37.

Djankov, Simeon, Rafael La Porta, Florencio Lopez-de Silanes and Andrei Shleifer. 2003. "Courts." The Quarterly Journal of Economics 118(2):453-517.

Djoussé L, Gaziano JM, 2008. "Alcohol consumption and heart failure: a systematic review". Curr Atheroscler Rep 10 (2): 117–20

Dynan, Karen E. “Habit Formation in Consumer Preferences: Evidence from Panel”, The American Economic Review, Vol. 90, No. 3 (Jun., 2000), pp. 391-406

Fogarty, James “The Demand for Beer, Wine and Spirits: A Survey of the literature”, Journal of Economic Surveys, Volume 24, Issue 3, pages 428–478, July 2010

Frye, Timothy and Andrei Shleifer. 1997. "The Invisible Hand and the Grabbing Hand." American Economic Review 87(2):354-358.

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

Gaviria A., and Raphael, S. 2001. "School-Based Peer Effects And Juvenile Behavior," The Review of Economics and Statistics, MIT Press, vol. 83(2), pages 257-268, May.

Glaeser E., Sacerdote B. I. and Scheinkman, J. 2003. "The Social Multiplier," Journal of the European Economic Association, MIT Press, vol. 1(2-3), pages 345-353, 04/05.

Grossman, Gene M. and Elhanan Helpman. 1994. "Protection for Sale." American Economic Review 84(4):833-850.

Gruber, J. and K˝oszegi B., 2001 “Is Addiction ‘Rational?’ Theory and Evidence”, Quarterly Journal of Economics (2001), 116(4), pp. 1261-1305.

Graham, B., “Identifying social interactions through conditional variance restrictions”, Econometrica 76 (3): 643 - 660, 2008

Heckman, J.J “Heterogeneity and state dependence”. Studies in Labor Markets, 1991, 31:91

Hotz, J. and Miller, R., 1993 “Conditional Choice Probabilities and Estimation of Dynamic Models”, Review of Economic Studies, 60, 497

Jin, Hehui, Yingyi Qian and Barry R. Weingast. 2005. "Regional decentralization and fiscal incentives: Federalism, Chinese style." Journal of Public Economics 89(9-10):1719-1742.

Johnson, Simon, Daniel Kaufmann and Andrei Shleifer. 1998. "The Unofficial Economy in Transition." Brookings Papers on Economic Activity (2):159-239.

Johnson, Simon, Daniel Kaufmann, John McMillan and Christopher Woodru. 2000. "Why 37 Do Firms Hide? Bribes and Unofficial Activity Under Communism." Journal of Public Economics 76(3):495-520.

Kaplan, David S., Eduardo Piedra and Enrique Seira. 2007. "Entry Regulation and Business Start-ups: Evidence from Mexico." Policy Research working paper 4322, World Bank, Washington DC.

Keane, M.P. "Modeling heterogeneity and state dependence in consumer choice behavior", Journal of Business & Economic Statistics, 15(3):310-327, 1997.

Klapper, Leora, Luc Laeven and Raghuram G Rajan. 2006. "Entry Regulation as a Barrier to Entrepreneurship." Journal of Financial Economics 82(3):591-629.

Krauth, B. V., 2005, "Peer effects and selection effects on smoking among Canadian youth", Canadian Journal of Economics/Revue canadienne d'économique, Volume 38, Issue 3, pages 735–757, August 2005

Kremer, Mi., and Levy D., 2008. "Peer Effects and Alcohol Use among College Students." Journal of Economic Perspectives, 22(3): 189–206.

Leon, D. A., Saburova L., Tomkins S., Andreev E., Kiryanov N., McKee M., Shkolnikov V., "Hazardous alcohol drinking and premature mortality in Russia: a population based case-control study", Lancet 369, 2007, pp. 2001–2009.

Leung S. F., and Phelps, C. E. "My kingdom for a drink. . . ?" A review of estimates of the price sensitivity of demand for alcoholic beverages. In: Hilton, M. E. and Bloss, G., eds. Economics and the Prevention of Alcohol-Related Problems. NIAAA Research Monograph No. 25, NIH Pub. No. 93–3513. Bethesda, MD: National Institute on Alcohol Abuse and Alcoholism, 1993. pp. 1–32.

Manski, Ch., "Identification of Endogenous Social Effects: The Reflection Problem," Review of Economic Studies, Vol. 60, No. 3, 1993, pp. 531-542.

Mills C.J. and H.L. Noyes. "Patterns and correlates of initial and subsequent drug use among adolescents". Journal of Consulting and Clinical Psychology, 52(2):231, 1984

Monteiro, Joana and Juliano Assuncao. 2006. "Outgoing the Shadows: Estimating the Impact of Bureaucratic Simplification and Tax Cuts on Informality and Investment." Pontica Universidade Catolica, Department of Economics, Rio de Janeiro, Brazil.

Moretti, E. and Mas A. "Peers at Work" , American Economic Review, 99(1), 2009

Moreira, Marcelo J. 2003. "A Conditional Likelihood Ratio Test for Structural Models." Econometrica (71):1027-1048.

Mulligan, Casey B. and Andrei Shleifer. 2005a. "Conscription as Regulation." American Law and Economics Review 7(1):85-111.

Mullahy, J and Sindelar J. L. “Alcoholism, Work, and Income”, Journal of Labor Economics, Vol. 11, No. 3. (Jul., 1993), pp. 494-520.

Mulligan, Casey B. and Andrei Shleifer. 2005b. "The Extent of the Market and the Supply of Regulation." The Quarterly Journal of Economics 120(4):1445-1473.

Nemtsov, A. V. 2002. “Alcohol-Related Human Losses in Russia in the 1980s and 1990s,” Addiction, 97, pp. 1413-25.

Pesendorfer, M. and P. Schmidt-Dengler, 2008, “Asymptotic Least Squares Estimators for Dynamic Games”, Review of Economic Studies, 901-928.

Rust, J. 1987, “Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurchner”, Econometrica, 55 (5), 999

Shehovtzov, Alexey O., Sergey V. Migin, Alexander V. Demin and Alexey S. Belov. 2005. Evolution of State Policy in the Area of Deregulation and Debureaucratization, and of Removal of Administrative Barriers to Entrepreneurship: EU-Russia Cooperation Program "Deregulation of Economy and Removal of Administrative Barriers". Technical report. In Russian. EUROPEAID/114008/C/SV/RU. 38

Shleifer, Andrei. 1997. "Government in Transition." European Economic Review 41(3-5): 385410.

Shleifer, Andrei and Robert W. Vishny. 1993. "Corruption." Quarterly Journal of Economics 108(3):599-617.

Shleifer, Andrei and Robert W. Vishny. 1994. "Politicians and Firms." Quarterly Journal of Economics 109(4):995-1025.

Shleifer, Andrei and Robert W. Vishny. 1998. The Grabbing Hand: Government athologies and Their Cures. Cambridge, MA: Harvard University Press.

Slinko, Irina, Evgeny Yakovlev and Ekaterina Zhuravskaya. 2005. "Laws for Sale: Evidence from Russian Regions." American Law and Economics Review 7(1):284-318.

Stigler, George J. 1971. "The Theory of Economic Regulation." Bell Journal of Economics 2(1):3-21.

Stigler, George J. and Claire Friedland. 1962. "What Can Regulators Regulate? The Case of Electricity." Journal of Law and Economics (4):1-16.

Stock, James H, Jonathan H Wright and Motohiro Yogo. 2002. "A Survey of Weak Instruments and Weak Identication in Generalized Method of Moments." Journal of Business & Economic Statistics 20(4):518-29.

Train, K., Discrete choice methods with simulation. Cambridge Univ Pr, 2003.

Treisman, Daniel, "Death and Prices: The Political Economy of Russia's Alcohol Crisis. Economics of Transition", Vol. 18, Issue 2, pp. 281-331, April 2010.

Tullock, Gordon. 1967. "The Welfare Cost of Tariffs, Monopoly, and Theft." Western Economic Journal 5:224-232. 39

Van Ours, J.C. "Is cannabis a stepping-stone for cocaine?" Journal of Health Economics, 22(4):539-554, 2003.

WHO report, "Global status report on alcohol and health", 2011, World Health Organization, 2011

Williams, 2005 "Habit formation and college students' demand for alcohol", Health Economics, 14: 119–134, 2005

World Bank, The. 2006. Doing Business 2007: How to Reform. Washington, D.C.: World Bank and International Finance Corporation.

World Bank, The. 2007. Doing Business 2008. Washington, D.C.: World Bank and Inter- national Finance Corporation.

Yakovlev, E., Political Economy of Regulation: Evidence from Russian Alcohol Market, CEFIR, mimeo, 2008

Zaridze, D., P. Brennan, J. Boreham, A. Boroda, R. Karpov, A. Lazarev, I. Konobeevskaya, V. Igitov, T. Terechova, P. Boffetta, R. Peto, "Alcohol and cause-specific mortality in Russia: a retrospective case–control study of 48 557 adult deaths", The Lancet, Volume 373, Issue 9682, 27 June 2009-3 July 2009, Pages 2201-2214.

Zhuravskaya, Ekaterina V. 2000. "Incentives to Provide Local Public Goods: Fiscal Federalism, Russian Style." Journal of Public Economics 76(3):337-368