Centre for
Economic
and Financial
Research
at
New Economic
School

October 2016

Demand for Alcohol Consumption and Implication for Mortality: Evidence from Russia

Evgeny Yakovlev

Working Paper No 221

CEFIR / NES Working Paper series

Demand for Alcohol Consumption and Implication for Mortality: Evidence from Russia*

Evgeny Yakovlev‡

October 24, 2016

Alcohol abuse is widely blamed for the very high rate of male mortality in Russia. I specify and estimate a simple structural model of the demand for alcohol that incorporates two key features of the Russian context. First, alcohol use – particularly incidents of heavy drinking – often involves friends and (male) family members. Second, there is strong habit persistence in alcohol use: depending on the degree of forward-looking behavior by consumers, responses to a tax policy will depend on beliefs about the future path of prices. I estimate the model using panel data from the Russian Longitudinal Monitoring Survey (RLMS), and two alternative sources of variation in alcohol prices: a 2011 change that shifted the trend in the rate of growth of the excise tax for alcohol, and regional variation in alcohol regulations over the 1995-2014 period. To obtain direct information on peer use of alcohol, I exploit the clustered design of the RLMS, which enables me to find close neighbors for nearly all sample members. The estimation results confirm that both peer influence and habit persistence are critical determinants of the longer-run response of alcohol demand to price changes. One third of the predicted 30% reduction in the rate of heavy drinking caused by a 50% permanent increase in vodka prices, for example, is attributed to the social multiplier effect that emerges as groups of friends jointly reduce their consumption. Finally, I use the RLMS data to relate patterns of heavy drinking to mortality. The estimates imply that permanent increases in alcohol prices would yield significant reductions in male mortality.

---

*Previous version: Peers and Alcohol: Evidence from Russia

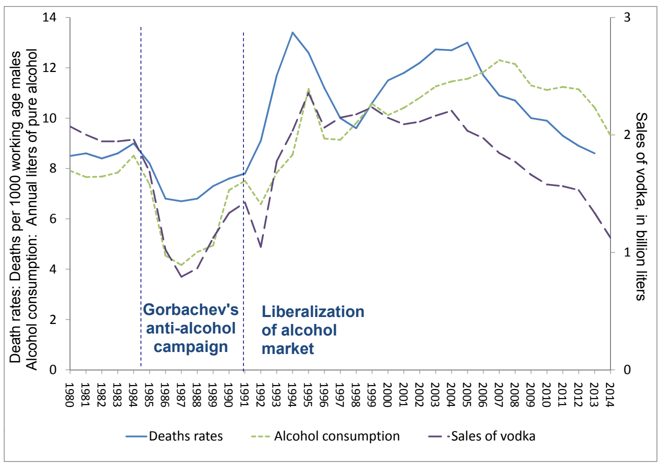
†I am especially grateful to David Card, Denis Nekipelov, Gerard Roland, and Katya Zhuravskaya for discussion of paper and kin support during my PhD study in UC Berkeley. I also thank to Irina Denisova, Ruben Enikolopov, Frederico Finan, Max Kasy, Patrick Kline, Lorenz Kueng, Thomas Lemieux, Alex Mas, Ted Miguel, Enrico Moretti, Anna Obizhaeva, Olexa Shwets, Mu-Jeung Yang, two anonymous referees and seminar participants at Berkeley, CERGE-EI, Harvard School of Public Health, ESEM meeting, NES, HSE, RES meeting and Acumen LLC for helpful discussions and comments. All errors are mine.

‡New Economic School. Address: Novaya str 100A, Skolkovo, Moscow, Russia. E-mail: eyakovlev@nes.ru

1 Introduction

Russian men are notorious for their hard drinking and their high rates of death associated with alcohol abuse.[1] Figure 1 illustrates the strong proximate relationship between male mortality rates and alcohol consumption. During the period of the Gorbachev anti-alcohol campaign in the final years of the Soviet Union (1985-1990), sales of alcohol fell and male mortality was also relatively low. After the collapse of the Soviet Union, the campaign ended and alcohol markets were liberalized, leading to a rise in alcohol consumption and a surge in mortality rates. From 1991 to 1996 alcohol sales doubled and mortality rates increased by 70 percent. Though recent changes in alcohol regulation have partially reversed this trend, both alcohol consumption and male mortality rates remain extremely high.[2]

Figure 1: Alcohol Consumption and Male Mortality Rate



Source: WHO, Treisman (2010), Rosstat. Left axis: Deaths per 1000 working age males (Rosstat); Annual adult per capita consumption, liters of pure alcohol (WHO). Right axis: Sales of vodka in billion of liters (Rosstat).

While the patterns in Figure 1 suggest that policies to restrict alcohol use will reduce consumption and lower male mortality, the magnitudes of the responses and the precise channels linking policy interventions to consumption and mortality are unclear.[3] The goal of this paper is specify and estimate a dynamic model of alcohol demand that can illuminate these issues. The model incorporates two important features of alcohol demand that

[1] Approximately one-third of all deaths in Russia are related to alcohol consumption (see Nemtsov 2002). Most of the burden falls on males of working age: more than half of all deaths of working-age men are accounted for by hazardous drinking (see Leon et al. 2007, Zaridze et al. 2009). Among recent economics studies of the connection between alcohol use and mortality in Russia are: Treisman 2010, Bhattacharya et al. 2013, Brainerd and Cutler 2005, and Kueng and Yakovlev, 2015.

[2] Russian male life expectancy in 2013 is 7 years below the average of the (remaining) BRIC countries and 5 years below the world average. Female life expectancy, by comparison, is 75 years: 5 years higher than the world average, and 2 years above average in the (remaining) BRIC countries. For health statistics, see https://www.cia.gov/library/publications/the-world-factbook/fields/2102.html, http://itbulk.org/population/life-expectancy-by-country/.

[3] Another concern is that some consumers – particularly in the Soviet era - used home made and illegally purchased alcohol.

I show are very important in the Russian context: peer effects in consumption, and habit persistence. Peer effects produce a "social multiplier" effect: decreases in a given consumer's own consumption lead his neighbors to consume less, so that the net effect of an alcohol price increase is amplified. Habit persistence similarly results in an intertemporal multiplier effect: decreases in alcohol consumption today change habits, reducing future preferences toward alcohol, and leading to decrease in alcohol consumption in the future.

I fit the model using micro-level data from the Russian Longitudinal Monitoring Survey (RLMS). RLMS is a nationally-representative panel dataset with a time span of more than 20 years, and it contains information on individual alcohol consumption and local alcohol prices, as well as rich data on individual demographic, health and socioeconomic characteristics.

The price elasticity of drinking is identified using the regression kink design (RK strategy) and instrumental variables regression (IV strategy). To find the price elasticity with the RK strategy, I use a kink in the policy regime of the federal excise tax on vodka. Before 2011 the excise tax had been linked to the CPI growth rate. Since 2011, the growth rate of the excise tax on vodka exceeded the growth rate of the CPI more than twice.

To confirm RK estimates, I use IV approach. In particular, I collect data on the regional regulation of the alcohol market during 1995-2008, when regional authorities had the autonomy to establish their own regulations, and use information on whether regional government imposes additional regulation on producers and on retailers as instruments for the price of alcohol in IV regressions.

To identify neighborhood effects, I exploit the clustered sampling structure of the RLMS survey, that I use for my analysis. RLMS surveys people within narrowly-defined neighborhoods (census blocks). There are sound reasons to believe that neighborhood influence is strong in Russia, given the patterns of dense geographical settlement inherited from the Soviet Union and the low level of mobility. This definition of peers is validated by documenting a strong increase in alcohol consumption around the birthdays of neighbors. The identification of peer effects in my paper relies on the assumption that some peer demographic characteristics affect the utility from alcohol consumption of peers but not the utility of agent (her)himself.

This paper then verifies the predictions of the model with both myopic and forward-looking assumptions about agents' behavior. Although there is no consensus regarding which model is more accurate[4], most literature on policy analysis considers only the myopic assumptions. At the same time, key effects of alcohol consumption – on health, family, and employment status, for example – do not necessarily appear immediately, but rather increasingly manifest themselves 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 con-

[4]In particular, Rust (1994) shows that in a general set-up of dynamic discrete-choice model 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, Fang and Wang 2010, Arcidiacono et al. 2007).

sumption may form a habit and thus affect future behavior. One therefore expects that individuals may behave in a forward-looking manner when determining current alcohol consumption (see rational addiction literature, Becker and Murphy 1988).[5] Possible misspecification from omitting forward-looking agent assumptions might introduce a bias in estimates, and as such might result in incorrect predictions regarding the effects of the proposed changes in the regulation of the alcohol industry.

I find significant price elasticity for heavy drinking and show the importance of peer effects for young age strata (below age 40). 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. The 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 the social multiplier. This cumulative effect can be decomposed in the following way: one's 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. Peer effects increase the estimated price response by 1.5 times for younger generations. The assumption that consumers are forward-looking increases the estimated cumulative effect roughly by an additional 30 percent.

I simulate the consequences of a price-raising alcohol tax on mortality rates and on social welfare. I find significant age heterogeneity in the effect of heavy drinking on the hazard of death with the effect being much stronger for younger generations. Increasing the price of vodka by 50 percent results in a decrease in mortality rates by one-fifth for males ages 18-29, by one-seventh for males ages 30-39, and by one-twentieth for males ages 40-49, with no effect on the mortality of males of older ages. I find also that when agents have bounded rationality (that is, do not take into account the effect of consumption on hazard of death), an increase in vodka price by 50 percent improves welfare. Besides, under certain assumptions about consumer utilities, a tax increases consumers welfare even for fully-rational agents.

Finally, in extension section, I provide a simple model in which I test what assumption, myopic or forward-looking better fits the data. I find that consumers are myopic when dealing with prices of alcohol, abut forward-looking when deal with health issues.

My paper contributes to the existing literature in several ways. First, the paper provides a methodological contribution to the existing studies of drinking and other (unhealthy) behavior by estimating a structural rather model using micro-level data. It allows me to account for and to disentangle different forces that drive decisions on drinking, predict how public policy would affect different subgroups of the population through these forces, and simulate the effect of policy on mortality rates and on the welfare of consumers.

Second, the paper provides several interesting and important examples of statistical

[5]Some studies find empirical evidence to support the rational addiction model (see Becker, Grossman, and Murphy 1991, 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 literature on time-inconsistent preferences, Gruber and Köszegi 2001).

relationships in the data. The kinked stricture of the federal tax policy regime allows me to apply regression kink (RK) estimates (see Card, Lee, Pei and Weber (2015), Lee and Lemieux (2010)). RLMS sampling structure allows me to document a strong increase in alcohol consumption around the birthday of neighbors and show that neighbors are indeed influential for personal decision making.

Finally, this paper contributes to the discussion of the causes and ways of combating the male mortality crisis in Russia. This question is highly policy relevant for Russia as well as for other countries that face similar public health issues. In contrast to the existing studies, the paper provides evidence of a causal relationship between the price of alcohol and alcohol consumption in Russia by addressing endogeneity issues that were present in previous studies.[6]

This paper is organized as follows. In the following section, I present the model. Section 3 describes the data and the variables used in the analysis. Section 4 presents estimation strategy. In Section 5, I discuss results. Section 6 discusses the identification assumptions of the model, provides robustness checks and extensions. Section 7 concludes.

2 Model

I model consumer choice in the tradition of discrete choice models of consumer behavior (see Nevo, 2011 for review).

In the model a consumer chooses among two alternatives, whether to drink heavily or not. I use this discretization because only hard drinking is universally agreed to be harmful for health. The effect of moderate drinking on health is ambiguous: for example, there is evidence that moderate drinking is associated with lower chance of heart diseases, such as coronary heart disease (see for example Cook and Moor, 2000).

The utility of heavy drinking depends on the price of alcohol, the alcohol consumption of one's peers, one's own habits, as well as different demographic and socioeconomic characteristics.

To model the responses to alcohol prices I follow the approach of Berry, Levinsohn, and Pakes (1995) henceforth BLP). I include a set of municipality$\times$year fixed effects in individual-level models and then regress estimated municipality$\times$year effects on the price level.

To model peer interaction I use the methods proposed by Bajary et al (2007) and Bajary et al (2011) for estimating static and dynamic discrete games of incomplete information.[7] It assumes that each member of a peer group makes a decision on whether to drink heavily or not based on the expected average probability of heavy drinking among her (his) peers. In a static (myopic) set up, this assumption leads to a simple two-stage procedure. In the first

[6]Previous studies that demonstrate a negative relationship between price and alcohol consumption using OLS estimates that show correlation rather than causal effect (see Andrienko and Nemtsov (2006) and Treisman (2010)).

[7]For a review of these models see Bajari et al 2011a and Nevo, 2011. For some recent developments see Aguirregabiria and Mira 2007, Berry, Pakes, and Ostrovsky 2007, and Pesendorfer and Schmidt-Dengler 2008.

stage I estimate average expected probability of drinking by each individual's peer group. In the second stage, these forecasts are plugged into each agent's decision model. For a dynamic (forward-looking rational addiction) model, the first stage is the same, but the second stage involves first estimating a polynomial approximation for the value function of not drinking in the current period, and then (building on the Hotz-Miller inversion, see Hotz and Miller, 1993) using this approximation in a third step to approximate to the probability of heavy drinking as a function of state variables and expectations.

Section 2.1 describes the set-up of the model in the event that consumers are myopic. Section 2.2 extends the model for forward-looking consumers.

2.1 Myopic Consumers

The set-up of the model is as follows.

In every period of time t, consumer chooses an binary action, a_{it}, whether to drink heavily $a_{it} = 1$ or not, $a_{it} = 0$.

The consumer utility depends on her (his) choice a_{it}, actions of her (his) peers a_{-it}, the set of observable factors that affect consumer's utility (S_{it}), and private stochastic preference shock, $e_{it}(a_{it})$, unobservable by all, except consumer her(him)self. Set S_{it} includes the socioeconomic and demographic characteristics of agent and the agent's peers, and municipality characteristics, such as the prices of alcohol and local temperature, etc. Following standard practice I call set S_{it} the set of *state* variables.

In a myopic model, consumers, deciding to partake in heavy drinking, only take the current utility of alcohol consumption into account.

The consumer utility consists of a current per-period utility, $\pi_{it}(a_{it}, a_{-it}, S_{it})$ and a private stochastic preference shock, $e_{it}(a_{it})$:

$$U(a_{it}, a_{-it}, S_{it}) = \pi_{it}(a_{it}, a_{-it}, S_{it}) + e_{it}(a_{it}) \tag{1}$$

The per-period utility from drinking has a linear parametrization:

$$\pi_{it}(a_{it} = 1, a_{-it}, S_{it}) = \rho_{mt} + \delta \frac{\sum_{-i} I(a_{jt} = 1)}{N-1} + \gamma habit_{it} + \Gamma' D_{it} + \Upsilon' G_{-it} \tag{2}$$

where municipality×year invariant factors ρ_{mt} capture price variation as well as other factors that affect consumer utility and that vary on the municipality×year level

$$\rho_{mt} = \theta log(Price)_{mt} + \Psi' X_{mt} + u_{mt} \tag{3}$$

Thus, $\pi_{it}(a_{it} = 1, a_{-it}, S_{it})$ depends on the average peer alcohol consumption; *habits* (defined as lagged alcohol consumption), a set of personal demographic and socioeconomic characteristics (D_{it}), (sub) set of peers characteristics G_{-it} and municipality×year invariant factors ρ_{mt}. X_{mt} stands for observable and u_{mt} stands for unobservable by the researcher factors that affect consumer utility and that 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, N stands for number of peers in peer group. For detailed description of all variables see estimation section (section 4).

I model peer interactions using game with incomplete information set up. In every period of time t, consumer and her (his) peers simultaneously choose their actions. The game with incomplete information implies that the consumer does not know the private preference shocks (and so the total payoffs) of his (her) peers. In the context of the model it implies that when someone starts drinking at a party, she (he) does not know exactly how much her (his) peers value drinking today and how much her (his) peers will drink up to the end of the party. Depending on some random factors like current problems with girl-(boy-) friends or parents, stress at work or in school, one can value drinking on a particular day differently and may end up drinking a lot or just one shot. Although consumers do not know exactly, they do guess on how much peers will drink using information that they know about their peers, like personal demographic characteristics, previous level of alcohol consumption etc. These guesses (beliefs) are consistent with the observed equilibrium behavior, and can be estimated using data on (own and peers') set of state variables $S_{it} = U_{j\in\{i,-i\}}\{habit_{jt}, D_{jt}, G_{nt}, \rho_{mt}\}$.

Thus, a consumer's expected (over beliefs) per-period utility from drinking is:

$$E_{e_{-i}}\pi_{it}(a_{it}=1, a_{-it}, S_{it}) = \delta\overline{\sigma_{jt}(a_{jt}=1|S_{it})} + \gamma habit_{it} + \Gamma' D_{it} + \Upsilon' G_{-it} + \rho_{mt} \tag{4}$$

where $\overline{\sigma_{jt}(a_{jt}=1|S_{it})} = \frac{\sum_{-i}\sigma_{jt}(a_{jt}=1|S_{it})}{N-1}$, and $\sigma_{jt}(a_{jt}=1|S_{it})$ stands for the consumer's i belief of what player j will do. I follow this notation throughout this paper.

I assume that private preference shocks of drinking, $e_{it}(a_{it}=1)$, have an i.i.d. logistic distribution.

All components of utility from *not* heavy drinking are normalized to zero: $\pi_{it}(a_{it}=0)=0$ and $e_{it}(a_{it}=0)=0$, and

$$U(a_{it}=0, a_{-it}, S_{it}) = 0 \tag{5}$$

A consumer chooses to drink hard if his or her per-period utility from (heavy) drinking is greater than the utility from not drinking:

$$\delta\overline{\sigma_{jt}(a_{jt}=1|S_{it})} + \gamma habit_{it} + \Gamma' D_{it} + \Upsilon' G_{-it} + \rho_{mt} + e_{it}(a_{it}=1) > 0 \tag{6}$$

Left-hand side of equation (6) is per-period utility from (heavy) drinking, and right-hand side is normalized utility from not (heavy) drinking.

2.2 Forward-looking consumers

In myopic model agents take in account only factors that affect current utility of alcohol consumption. At the same time, alcohol consumption may affect not only the current but also the future flow of utilities. Alcohol consumption is habit-forming and thus affects future behavior. Many consequences of alcohol consumption that also affect consumer utility, such as health, family, income and employment status, do not necessarily appear immediately, but rather they increasingly manifest over the course of the next few years (see Mullahy and Sindelar 1993, Cook and Moore 2000). Indeed, table 3 shows that heavy drinking affects future health, marital status and income for Russian males. Conditional upon current health, marital status and income, heavy drinking results in a decrease in future income (both own and family income), in worsening future health and in higher risk of divorce. One, therefore, would expect that individuals may behave in a forward-looking manner when determining current alcohol consumption.

A forward-looking consumer maximizes not only the current value of the utility but also the discounted expected flow of future utilities. The expected present value of consumer utility consists of the current per period utility, $\pi_{it}(a_{-it}, a_{it}, S_{it})$, discounted expected value function, $\beta E(V_{it+1}(s_{t+1})|a_{-it}, a_{it}, S_{it})$, and a stochastic preference shock, $e_{it}(a_{it})$:

$$U(a_{-it}, a_{it}, S_{it}) = \pi_{it}(a_{-it}, a_{it}, S_{it}) + \beta E(V_{it+1}(s_{t+1})|a_{-it}, a_{it}, S_{it}) + e_{it}(a_{it}) \tag{7}$$

I put annual discounting factor β equal to 0.9.[8] Current per period utility $\pi_{it}(a_{-it}, a_{it}, S_{it})$ and stochastic preference shock $e_{it}(a_{it})$ are similar to myopic case (see equations (2) and (3)). The consumer does not observe the actions of peers and forms expectations over peers actions. The expected per-period utility $E_{e_{-i}}\pi_{it}(a_{-it}, a_{it} = 1, S_{it})$ is the same as in the myopic case (see equation (4)).

A forward-looking consumer chooses to drink hard if his or her expected present value of the utility from (heavy) drinking is greater than the utility from not drinking:

$$\begin{aligned} E_{e_{-i}}\pi_{it}(a_{it} = 1, \quad a_{-it}, \quad S_{it}) + \beta E(V_{it+1}(s_{t+1})|a_{-it}, a_{it} = 1, S_{it}) + e_{it}(a_{it} = 1) \\ > \beta E(V_{it+1}(s_{t+1})|a_{-it}, a_{it} = 0, S_{it}) \end{aligned} \tag{8}$$

For the case of forward-looking consumers, I assume that consumers 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 consumer choice of action. Finally, I restrict equilibrium to be a Markov Perfect Equilibrium, so that a consumer'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). [9] For myopic consumers the model is static, such that none of the assumptions

[8] Recent studies vary in their estimates of personal discount rate (see Arcidiacono et al (2007), Hausman (1979), Dreyfus and Viscusi (1995), Moore and Viscusi (1990), and Warner and Pleeter (2001)). The closest to our study, Arcidiacono et al 2007, found yearly discount factor $\beta = 0.9$. I use this number ($\beta = 0.9$) when estimating a model with "forward-looking" consumers.

[9] These assumptions, together with other assumptions that I made in a paper (such as Markovian state transition process, infinite time horizon, iid logistic error components, etc) are standard assumptions in vast majority of dynamic discrete choice models. See Aguirregabiria and Mira (2010) for review and for discussion of these and other assumptions that are commonly used these models. These assumptions are done in order to simplify and to

described above are needed.

3 Data Description

I use data from the Russian Longitudinal Monitoring survey (RLMS).[10] The RLMS is a nationally-representative annual survey that covers more than 4,000 households (with between 7,413 and 9,444 individual respondents), from 1992 to 2014. The RLMS provides a survey of a very broad set of questions, including a variety of individual demographic and social-economic characteristics, health outcomes (including death events), and consumption data. It also contains data on individual-level alcohol consumption and data on neighborhood characteristics, including – critically – the price of alcoholic beverages in each neighborhood, which allows me to analyze individual price elasticity.

My study utilizes rounds 5 through 23 of RLMS over a time span from 1995 to 2014, except 1997 and 1999.

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. Starting from round 17 dataset that provided by Population Center at The University of North Carolina at Chapel Hill no longer contains identifiers of neighborhoods. I use information from the previous rounds and the HSE version of the RLMS data to collect neighborhood identifiers for rounds 17-23.[11]

The data cover 33 regions – 31 oblasts (krays, republics), Moscow and St. Petersburg. Two of the regions are Muslim. Seventy-five percent of respondents live in an urban area. Forty three percent of respondents are male. The percentage of male respondents decreases with age from 49 percent for ages 13-20 to 36 percent 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. It is the lowest in Moscow and St. Petersbug (60 percent) and the highest in Western Siberia (92 percent). The RLMS team provides a detailed analysis of attrition effects and finds no significant effect of attrition.[12]

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 78237 individuals*year points (2956 to 6616 individuals per year). Summary statistics for the primary demographic and socioeconomic

make possible to implement the nontrivial computational task of estimation of dynamic discrete choice problem.

[10]This survey is conducted by the Carolina Population Center at the University of Carolina at Chapel Hill, and by the Higher 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).

[11]A previous version of the paper uses only data from round 5 to 16 get similar results.

[12]For description of interview completion rates and attrition rates see RLMS web-site, http://www.cpc.unc.edu/projects/rlms-hse/project/samprep.

characteristics are presented in Table 4.

3.1 Alcohol Consumption Variable and Price of Alcohol

Although the negative health and social consequences of hard drinking are widely recognized, there is no documented evidence of negative consequences from moderate drinking. Thus, I focus 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) and zero otherwise. Alcohol consumption is measured as the reported amount of pure alcohol consumed daily during the previous month.[13] The reported threshold level corresponds to the reported amounts greater than 150 grams of pure alcohol per day. This amount corresponds to a daily consumption of 0.35 liters of vodka (or samogon) or 9 bottles (0.33 liters each, 3 liters total) of beer. A summary statistics and age profiles for reported amounts of alcohol consumption are shown in Table 4 and Figure A2 in the appendix.

3.2 "Peers" Definition

RLMS data also allows me to get information on groups of close neighbors and thus to estimate neighborhood (peer) effects.

The Soviet Union left a legacy of communist-style apartment blocks where people live in (uncomfortably) close proximity. I exploit this feature and define peers using geographical locations.

Approximately ten percent of Russian families live in dormitories and communal houses where residents share kitchens and bathrooms. 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. The most common dvors (so called "khrushchevki") are relatively small-size dvors with a population of about 300 people. A photo of a typical dvor is presented in Figure A1 in the appendix. Dvors are the most popular 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.

The important feature of the RLMS survey is that it has a clustered structure.[14] The

[13]The reported amount of pure alcohol is calculated using RLMS data on consumption of all types of alcohol including vodka and other hard drinks, beer, wine, champagne and home-made vodka (samogon) using the following formula: Q(pure alcohol)=0.4Q(hard drinks)+0.12Q(dry vine)+0.12Q(champagne)+0.15Q(fortified vine)+0.05Q(beer)+0.4Q(samogon).

Sometimes a high level of average alcohol consumption is not as harmful for health as one-time drinking binge (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).

[14]See the RLMS web site, http://www.cpc.unc.edu/projects/rlms-hse/project/sampling. A similar data feature- in French LFS data- was explored by Maurin and Moschion (2009) in a analysis of neighborhood effects in mothers labor force participation decisions.

basic sampling unit of the RLMS survey contains one Russian census block. Households within the same set of census blocks are surveyed in every round. The average population of a census block in Russia is 300.[15] A typical census block in Russia contains one dvor; this allows me to use information on neighborhood (and age) to identify peer groups.

I define "peers" as those who live in one neighborhood (census block) and belong to the same 10-years stratum. Age strata are defined by ages 18-29, 30-39, 40-49, and 50-65.

The median number of people in a peer group is 5; the mean is 11; the lower 1 percent is 2, the upper 90 percent is 35, and the largest number is 83. According to the RLMS team, the average population of a census block is 300 people (including females, elderly and children). It implies that the average size of a population of males of age 18-65 is about 90 people and the total size of the peer group is 22 people.[16] I have data on one half of them. On average, I have 794 peer groups per year (each with 2 or more peers). The distribution of the number of peers per peer group is shown in Table 5.

To verify the reliability of my sorting procedure, I implement the following test. I correlate the logarithm of the amount of vodka consumed during the previous month with a dummy variable if a person has a birthday in the previous month and with averages of the birthday dummy variables across peers.

The specification of the regression is as follows:

$$Log(1+vodka)_{it} = \zeta_0 + \zeta_1 I(birthday)_{it} + \zeta_2 \sum_{j\in peers} \frac{I(birthday)_{jt}}{N-1} + \delta_t + \varepsilon_{it} \tag{9}$$

where *vodka* stands for amount of vodka drunk in the last month, $I(birthday)_{kt}$ is an indicator that a person k has a birthday in the previous month ($k \in \{i, j\}$), N is a number of peers in peer group, and δ_t are time fixed effects.[17] Vodka is the most popular alcoholic beverage to serve on birthdays, compared to either beer or wine. Regression suggests that a person's consumption of vodka increases by 18 percent if his birthday is during the previous month, and, by 6 percent if there was a birthday of one of his peers in a group of 5 peers (median peer group size).[18] The results are robust if I eliminate household members from the sample of peers.[19]

---

[15] The RLMS team indicates that population of census blocks in the RLMS survey is in a range between 250 and 400 people. There are 459,000 census blocks in Russia (data on 2010 census). This number implies that average population of the census block 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).

[16] According to national statistics, the share of males of age 18-65 in the total population equals 30 percent.

[17] In the RLMS survey, people report the amount of alcohol they consumed during the last 30 days before survey day. RLMS does not have data on daily consumption, so I cannot estimate correlation using day-level data.

[18] The coefficient ζ_2 in regression (9) equals 0.219. To get meaningful interpretation of the coefficient I look at a The coefficient ζ_2 in regression (9) equals 0.219. To get meaningful interpretation I look at a peer group with size 5 people and calculate the effect of having a birthday of one peer in this peer group. In group of 5, every member has four peers, an so the effect of having one birthday of peers equals to $\zeta_2 \sum_{j\in peers} I(birthday)_{jt}/(N-1) = 0.219 * 1/4 = 0.55$.

Because I do not have data on all peers in a group, OLS estimates shown in Table 1 suffer from attenuation bias.

[19] 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).

Table 1: Birthdays and Alcohol Consumption.

	All peers	Without household members
	log(vodka consumption)	log(vodka consumption)
Birthday of one of the peers	0.055	0.055
	[0.017]***	[0.018]**
Own birthday	0.181	0.182
	[0.040]***	[0.040]***
Year*month FE	Yes	Yes
Observations	64,133	63,886

Notes: Standard errors clustered at neigborhoodXyear level are in brackets.
*significant at 10%; ** significant at 5%; *** significant at 1%

3.3 Mortality

To analyze the effect of alcohol consumption on male mortality I use information on death events that is available in RLMS survey.

Table 5 (panel B) shows the distribution of deaths and death causes for males from different age cohorts. There are 626 death events; out of them 44 (or 7 percent) are deaths of males 18-29 age old; 86 deaths (14 percent) are deaths of males of age 30-39; 149 (24 percent) are deaths of males of age 40-49 and 347 (55 percent) are deaths of males of 50-65 age old.

According to medical studies, the largest contributor to alcohol-related mortality among Russian males are poisoning, accidents, injures, the second largest contributor are cardiovascular diseases (see Nemtsov, 2003, Leon et all 2007, Zaridze et al 2009, Shkolnikov et al 2013). RLMS recorded 6 causes of death, namely heart attack, stroke, external causes of death (accidents, injuries and poisoning), cancer, tuberculosis, and "other" causes. The causes of deaths are reported only in 60 percent of death events. Deaths from poisoning, accidents and violence are prevalent among young age cohorts. Out of deaths with reported causes, the deaths from poisoning, accidents and violence constitute 63, 45, 29 and 9 percent of deaths of males of age 18-29, 20-29, 40-49, and 50-65 correspondingly. Deaths from heart attack and stroke constitute 9, 17, 38 and 51 percent of deaths of males of age 18-29, 20-29, 40-49, and of age 50-65 correspondingly. Deaths from cancer that are mainly not related with alcohol consumption are prevalent among older males. They constitute 6, 7, 9, and 29 percent of deaths of males of age 18-29, 20-29, 40-49, and of age 50-65 correspondingly.

4 Estimation

4.1 Myopic Consumers

Myopic consumers maximize only the current per-period utility, $\pi_{it}(a_{-it}, a_{it}, S_{it})$, and thus discount their future utilities with discount factor $\beta = 0$.[20]

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

At the first stage, I estimate beliefs (predicted probabilities of drinking) $\hat{\sigma}_{jt}(a_{jt} = 1|S_{it})$ as a (arbitrary) function of state variables $S_{i,-i,t}$.[21]

On the second stage, I estimate the remaining parameters of utility function by plugging estimated beliefs into the following logit regression:

$$\begin{aligned} I(heavy\,drinker)_{it} &= \rho_{mt} + \sum_{k} \delta_k I(age\,strata = k)\overline{\hat{\sigma}_{jt}(a_{jt} = 1|S_{it})} \\ &\quad + \gamma habit_{it} + \Gamma' D_{it} + \Upsilon' G_{-it} + e_{it} \end{aligned} \tag{10}$$

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

The set of personal demographic characteristics D_{it} includes weight, education, work status, lagged dummy for smoking, health status, age, age squared, marital status, religion, size of family and log(family income). The (sub) set of peers' characteristics G_{-it} that stands for so-called exogenous effects includes the share of peers with college education, and the share of unemployed peers. Variable *habit* is defined as lagged alcohol consumption.[22] Besides, I allow effect of habits vary by age, i.e. $\gamma habit_{it} = \gamma_0 a_{i,t-1} + \gamma_1 a_{i,t-1} \times \widetilde{age}_{it} + \gamma_2 a_{i,t-1} \times \widetilde{age}^2_{it}$, where $\widetilde{age}$ is demeaned age.

The parameters of the model are identified under the assumption that the utility of one consumer does not depend on the subset of peer demographic characteristics, and that random components of personal utility are independent of peer demographic characteristics

[20] The expected utilities of myopic consumer are as follows: $E_{e_{-i}}U_{it}(0) = 0$, and $E_{e_{-i}}U_{it}(1) = \delta\overline{\sigma_{jt}(a_{jt} = 1|S_{i,-i,t})} + \gamma habit_{it} + \Gamma' D_{it} + \Upsilon' G_{-it} + \rho_{mt} + e_{it}(1)$

[21] The expression for the first stage is as follows: $I(a_{jt} = 1)_{it} = H(s_{it})'\zeta + \varepsilon_{it}$, where $I_i = I(a_{it} = 1)$, $H(s_{it})$ is a set of Hermite polynomials of state variables s_{it} (for a discussion of non-parametric regression with Hermite polynomials see Ai and Chen (2003)). That is, $H(s_{it})$ contains a 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 the dimensionality problem. One important implication 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 $\rho^k_{mt} = \rho_{mt}$. Still, ρ_{mt} will account for any variable (in any power) that varies only on the municipality$\times$year level.

[22] I define state variable $habit_{it}$ as follows. Let state variable $habit_{it} = 0$ if $age_{it} < 18$(years) . The transition process of $habit_{it}$ is 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 the consumer's equilibrium choice of action in previous period, and $\varphi_{i,t}$ is (negligible) smoothing noise. $\varphi_{i,t}$ is added to ensure existence of equilibrium. With this definition of habits, the model satisfies assumptions required for Markov perfect equilibrium (see for example, Assumptions AS, IID and CI-X in Aguirregabiria and Mira, 2007 or Bajari et al 2010) that is requred for dynamic models. A Markov perfect equilibrium (MPE) in this game is a set of strategy functions $a^\star$ such that for any consumer 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 $a^\star_i(S_t, e_{it}) = b(S_t, e_{it}, a^\star_{-i})$.

(see Bajari et al. 2005 for proof).[23] I discuss the robustness of my results in the section 6 and in section 7 in online appendix.

To estimate price elasticity, 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 the cheapest vodka in neighborhood and a set of control variables.

$$\hat{\rho}_{mt} = \theta log(Price)_{mt} + \Psi' X_{mt} + u_{mt} \tag{11}$$

To find the exogenous variation of price, I employ two alternative strategies.

The set of municipality-level factors X_{mt} include average income, level of education and unemployment rate in a region as well as regional and - depending on specification- time fixed effects (IV regression) or smooth function of time trend (RK regression).[24]

In the first specification, I utilize the regression kink design (RK) approach. Similar to the RD (regression discontinuity) method, RK explores the non-smoothness of the policy function to find the exogenous variation of the variable of interest: RK explores the kink structure of policy functions (for example the kink in tax schedule) and uses the variation in the slopes of the policy function around the kink to identify casual relationship (for theoretical treatment, statistical packages and discussion see Card, Lee, Pei and Weber (2012), Calonico, Cattaneo, and Titiunik (2014), Lee and Lemieux (2010)). Under the assumption that all other factors behave smoothly in neighborhood of kink, RKD succeeds in identifying a casual effect by looking at the change in the slope of an outcome variable.

In the RK estimation, I explore a kink in the policy regime of the excise tax on vodka. In 2000 the Russian government introduced a specific excise tax on vodka.[25] Since that and till 2011 the excise tax was updated to catch up with consumer price index (CPI). Since 2011 tax grew twice the rate of CPI growth.

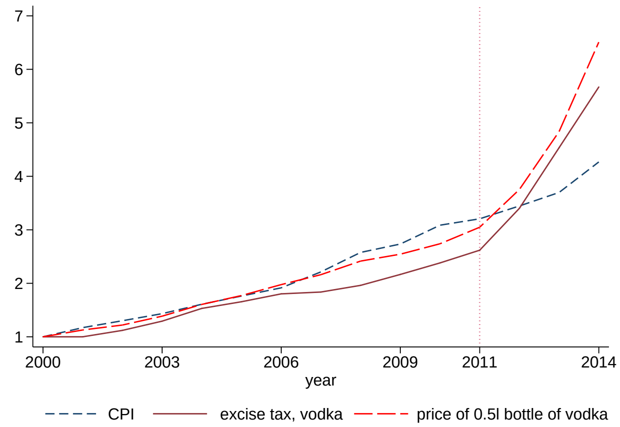
Figure 2 below shows how excise tax, the price of vodka and CPI changed in the last 15 years.

[23] Exclusion restriction requires that subset G_{-it} does not contain all set of demographic variables. From game-theoretic point of view it seems to be reasonable assumption: for example, consumer does not have higher utility when she/he drinks with peers with different weight, different marital or health status. In Sections 6 and in online appendix I will provide discussion of identification assumptions as well as different robustness checks of obtained results, by allowing different sets of demographic characteristics to be excluded as well as allowing. Results of these regressions are robust to choice of specifications, and J-tests for every specification support hypothesis that excluded variables are exogenous.

[24] Regional fixed effects capture factors that affect utility of drinking and that invariant at regional level (such as average temperature or predominant religion) whereas time fixed effects capture time-invariant factors that affect utility of drinking (such as effect of financial crisis of introduction of federal alcohol regulation).

[25] The excise tax on vodka was introduced before 2000. However, before 2000 it was collected as an ad velorem tax that resulted in large scale tax avoidance. Stores under-reported prices on vodka and subsequently under-paid taxes. As a result, starting in 2000, a fixed excise tax per bottle of vodka was introduced.

Figure 2: Excise tax on vodka, the average price of vodka and CPI



Source: Rosstat (www.gks.ru), Rosalcoholregulirovanie (www.fsrar.ru)

Figure 6 show averages (by year) of vodka prices and alcohol consumption according to RLMS data. Table A2 in Appendix shows excise tax rates in years 2000-2014. To find the RK estimates, I modify regression (7) above to be as follows

$$\hat{\rho}_{mt} = \theta log(Price)_{mt} + f(t) + \eta_r + \Psi' D_{mt} + u_{mt} \tag{12}$$

Here, $f(t)$ stands for the smooth function of time variable (defined as $t = year - 2011$), η_r stands for regional fixed effects, and a set of control variables D_{mt} includes log CPI, the regional averages of income, education and employment. $log(Price)_{mt}$ is instrumented by the kink in policy regime of excise tax that was calculated as $t * I(year \geq 2011)$.

I work with two bandwidths. First, I use the whole sample of years for which data on the excise tax is available (global polynomial approach), i.e. years 2000-2014. Second, I use years 2008-2014, i.e bandwidth of size 3 (local polynomial approach).[26] In the global polynomial approach, $f(t)$ is parametrized as second order polynomial of time variable; In the local polynomial approach, $f(t)$ is parametrized as linear function of time.

For robustness, I re-run RK regressions discussed above with different instrumental variable. Instead of the kink variable $t * I(year \geq 2011)$, I use exact values of excise tax as an instrumental variable (excise tax profile is shown in Table A2 in appendix).

As an alternative, I use the data on regional regulation of the alcohol market to instrument the price variable. The time span for IV analysis is years 1995-2008. During these years (Yeltsin's presidential terms and the beginning of the Putin administration) Russian Regional authorities had substantial freedom to impose regulation procedures on local markets. I collect data on regional regulation of the alcohol market during this time,

[26]The bandwidth size is chosen according to Imbens and Kalyanaraman (2012) procedure for first stage regression. The bandwidth size is equal to 2.86. I then choose similar for all specification bandwidth with size of 3.

count down additional regulations that regional government imposes in a particular year, and use number of additional regulations imposed by the regional government as an instrument in IV regressions.[27]

I obtain IV estimates from the following regression

$$\hat{\rho}_{mt} = \theta log(Price)_{mt} + \eta_t + \eta_f + t\eta_f + \Psi' D_{mt} + u_{mt} \tag{13}$$

Here D_{mt} includes log CPI, regional averages of income, education and employment; η_t and η_f stand for time and federal district fixed effects, $t\eta_f$ stand for federal district-specific time trends.[28]

In addition I combine the global polynomial version of regression (10) and regression (11) in one IV regression with two instruments: federal excise tax on vodka and regional regulation combined.

$$\hat{\rho}_{mt} = \theta log(Price)_{mt} + f(t) + \eta_f + t\eta_f + \Psi' D_{mt} + u_{mt} \tag{14}$$

In this case, D_{mt} also includes indicators that the data on the excise tax of vodka is missing (for years 1995-1999) and data on regional regulation is missing (years 2009-2014). Both instruments are set to be 0 in years when data on them is missing.

Finally, because I applied sequential estimation that involves several steps, I calculate standard errors using a bootstrap procedure. The standard errors in regressions $(7) - (11)$ as well as standard errors in the dynamic model are calculated using a bootstrap procedure with re-sampling clustered at the municipality×year level. Reported standard errors are based on 500 replications.

4.2 Forward-looking consumers

Here I present an estimation strategy for forward-looking consumers (with $\beta = 0.9$). My estimation procedure follows Bajari et al. (2007).[29]

---

[27] As a rule, regional regulations are imposed for two reasons. First, regulations are a popular tool for increasing regional budget revenues: the excise tax and license tax are two of the very few taxes that go directly into the regional budget. Second, the regional regulations are imposed in the result of the 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 reasons.

[28] A federal district is a larger territorial unit compared to a Russian region. RLMS surveys people within 8 Federal districts that contain 34 Russian regions. In the robustness section, I estimate IV regressions with regional FE and regional-specific trends. In this case, due to lack of variation that remained after accounting for regional and regional - specific trends, instruments do not have sufficient predictive power: Although instruments are still (statistically) significant, the F-test does not exceed 7. Still, the point estimates of elasticities in this case are very similar to the main IV specifications, although in many cases coefficients lose statistical significance.

[29] For surveys of dynamic discrete models, see research by Aguirregabiria and Mira (2010) and Bajari et al. (2011b). Compared to many other studies, the estimation strategy proposed by Bajari et al. (2007) 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 sequential procedure estimation, wherein every step of estimation has closed-form solutions. This means that one can avoid mistakes and problems related to finding a global maximum using a maximization routine. Finally, this estimation procedure does not require discretization of variables. This flexibility of the 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

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

$$V_i(a_{it}, S_{it}) = E_{e_{-i}}\pi_{it}(a_{-it}, a_{it} = 1, S_{it}) + \beta E(V_i(s_{t+1})|a_{it}, S_{it}) \quad (15)$$

can be rewritten as two moment equations (for derivation see Proof A1 in Appendix and Bajary et al 2007, 2011):

Bellman equation for $V_i(0, s_t)$

$$V_i(0, S_{it}) = \beta E_{t+1}(\gamma - log(\sigma_{it+1}(0)) + V_i(0, S_{it+1})|S_{it}, a_{it} = 0) \quad (16)$$

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

$$\begin{aligned} ln(\sigma_{it}(1)) - log(\sigma_{it}(0)) + V_{it}(0, S_{it})_i &= \pi_{it}(a_{-it}, a_{it} = 1, S_{it}, \theta) \\ &+ \beta E_{t+1}(\gamma + V_{it+1}(0, S_{it+1}) - log(\sigma_{it+1}(0))|a_{it} = 1, S_{it}) \end{aligned} \quad (17)$$

These two equations together with a moment condition on choice probabilities

$$E(I(a_i = k)|S_{it}) = \sigma_{it}(k|S_{it}),\ k \in \{0, 1\} \quad (18)$$

form the system of moments that I estimate in next section.

The first step of the estimation procedure resembles the first step in the estimation of the myopic model: I obtain estimates of beliefs (choice probabilities) $\widehat{\sigma_{it}(0)}$ and $\widehat{\sigma_{it}(1)}$.

On the second step, I estimate $V_{it}(0, S_{it})$ as an (arbitrary) function of state variables $S_{i,t}$ by solving a sample equivalent of the moment condition (16).

To do it I allow $V_{it}(0, S_{it})$ to be a (hermite) polynomial function of state variables $H(s_{it})'\mu$ and find $V_i\widehat{(0, s_t)} = H(s_{it})'\hat{\mu}$ by finding $\hat{\mu}$ that solves equation $I(a_{it} = 0)[H(s_{it})'\hat{\mu}] = \beta I(a_{it} = 0)[(log(1 + exp(log(\widehat{\sigma_{it+1}(1)}) - log(\widehat{\sigma_{it+1}(0)})) + H(s_{it+1})'\hat{\mu}]$.

On the third step, I estimate $\pi(1, S_{it})$ by solving sample the equivalent of moment condition (17). I estimate $\pi(1, s)$ by solving for $\hat{\theta}$ equation $I(a_{it} = 1)[s_t'\hat{\theta} + V_{it}\widehat{(0, s_t)} + log(\widehat{\sigma_{it}(1)}) - log(\widehat{\sigma_{it}(0)})] = \beta I(a_{it} = 1)[\gamma + (log(1 + exp(log(\widehat{\sigma_{it+1}(1)}) - log(\widehat{\sigma_{it+1}(0)})) + V_{it}\widehat{(0, s_{t+1})}]$.[30]

The estimation of price elasticity is similar to that employed in the myopic case.

To simplify the description of the procedure, I start with an estimation of elasticity under the assumption that the government changes the price without changing consumers expectations over future price movement.

To calculate elasticity in this case, I obtain the municipality×year fixed effects components $\hat{\rho}_{mt}(\pi)$, $\hat{\rho}_{mt}(EV1)$, $\hat{\rho}_{mt}(EV0)$ of my estimates of per-period utility of drinking $\pi_{it}(a_{-it}, a_{it} =$

models comparable.

[30]This sequential estimation procedure is not efficient. One can improve efficiency by solving three moment conditions together. 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.

$1, s_t$), and conditional expectation of the future value function, $\beta E(V_i(S_{it+1})|a_{it} = 1, S_{it})$, and $\beta E(V_i(S_{it+1})|a_{it} = 0, S_{it})$. Then, I calculate the aggregate effect of the fixed effect components, $\hat{\rho}_{mt}$:

$$\hat{\rho}_{mt} = \hat{\rho}_{mt}(\pi) + \hat{\rho}_{mt}(EV1) - \hat{\rho}_{mt}(EV0) \tag{19}$$

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

$$\hat{\rho}_{mt} = \theta log(Price)_{mt} + \Psi' X_{mt} + u_{mt} \tag{20}$$

This estimation procedure relies on assumption that consumers, when forming their expectations about future prices, use the rule of price motion guessed from their previous experience. In Russia, the price of alcohol is volatile, and the rule of price motion demonstrates significant mean reversion (see Table A5 in appendix). Therefore, the estimation above implies that consumers believe that the current increase in price comes before it's future decrease. If the government increases price permanently and credibly promise that the price will not decrease in the future then the expectations of consumers should be corrected.

To estimate price elasticity in this case, I make two simplifying assumptions about the price transition process and about the parametrization of the choice-specific value functions.

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: $log(p_{i,t+1}) = \phi_0 + \phi_1 log(p_{it}) + \omega_{it}$, where $E(\omega_{it}|p_{it}) = 0$. Second, I assume the following parametrization of the choice-specific Value functions: $V_i(S_{it}, a_{t-1} = j) = \vartheta_j log(p_t) + V_i(\{S_{it}/p_t\})$, where $j \in \{0, 1\}$, and $\{S_{it}/p_t\}$ is a set of state variables excluding price.

Under these assumptions, price elasticity can be estimated from the regression of modified fixed effect component $\widetilde{\rho}_{mt}$:[31]

$$\widetilde{\rho}_{mt} = \hat{\rho}_{mt}(\pi) + \frac{1}{\widehat{\phi}_1}(\hat{\rho}_{mt}(EV1) - \hat{\rho}_{mt}(EV0)) \tag{21}$$

on the log of the relative price of the cheapest vodka in neighborhood:

$$\widetilde{\rho}_{mt} = \theta log(Price)_{mt} + \Psi' X_{mt} + u_{mt} \tag{22}$$

In the dynamic model, I use the only the regional regulation variable as instrument for price. I do not explore RK estimates because of data restriction. The data on $\widetilde{\rho}_{mt}$ is not calculated for the last year of 2014 because, when calculating $\widetilde{\rho}_{mt}$, I use information on the leads of variables (see step 3 of estimation procedure). Without year 2014, not enough data remains to the right side of the kink (2011).

[31] See note 1 in appendix top for proof.

I estimate the model under two different normalizations of the consumer's utility. In contrast to the myopic case, the dynamic model's estimator of parameters depends on the chosen normalization. In base specification I normalize the utility of not (heavy) drinking to be 0. In the second specification, I normalize the utility of (heavy) drinking to be 0.

4.3 Effect of mortality

To model the effect of a change in vodka price on mortality rates I estimate the effect of heavy drinking on death rates using hazard of death regression

$$\lambda(t,X) = exp(X\beta)\lambda_0(t) \tag{23}$$

where $\lambda_0(t)$ is the baseline hazard, common for all units of population.

I use a semi-parametric Cox specification of baseline hazard. A set of explanatory variables X includes I(heavy drinker), I(smokes), log of family income, Health self-evaluation, body 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 be 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.

5 Results

Estimates of per-period utility parameters are shown in Table 2 below and in Tables 6 through 8. For myopic consumers, the per-period (indirect) marginal utility with respect to log(price) is equal to -0.5 and -0.838 for base RK and for IV regressions respectively.[32]

For a myopic consumer with a mean level of all demographic characteristics, the marginal utility of -0.5 implies that, for example, an increase in the price of vodka by 50 percent will lead to a decrease in the probability of heavy drinking by 4 percentage points (from 0.25 to 0.21).

For forward-looking consumers, the per-period (indirect) marginal utility with respect to log(price) is equal to -0.7. To evaluate the effect of a change in price on forward-looking consumers, one must know not only the consumer's per-period utility but also have an expectation of the consumer's future value function. The marginal value function of consumers with respect to log(price) is equal to -1.025.[33] The marginal value function of -1.025

[32] The RK estimates (with different bandwidth sizes, instruments and kernel specifications) vary in a range from -0.3 to -0.6 (see Table 7). Coefficients are statistically significant for RK estimates with a bandwidth of size 11. In RK regressions with bandwidth size 3 coefficients are same in magnitude, but lose statistical significance due to the decrease in sample size (and therefore loss of power).

[33] Elasticity is calculated under assumption that a price increase is permanent. In the event that the government cannot ensure that the change in price is permanent, the elasticity is -0.765. For description of the calculation procedure see Appendix. It worth noting that estimations of utilities and response functions, although different, do not differ dramatically in the myopic and forward-looking models. A possible explanation of this phenomenon is

implies that an increase in the price of vodka by 50 percent leads to a decrease in the probability of becoming a heavy drinker by 6.5 percentage points.

In both myopic and forward-looking specifications, 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 consumers, $\widehat{\delta}$ equals to 1.4, 0.83, 0.305, and 0.205 for ages 18-29, 30-39, 40-49, and 50-65 respectively. For forward-looking consumers, $\widehat{\delta}$ equals to 1.223, 0.709, 0.354, and 0.367 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 average per 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 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. Again, the forward-looking model does not allow immediate statistical interpretation. 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 2: Consumer's utility parameters. Point estimates.

	Myopic consumers			Forward-looking consumers	
				Per-period utility	Value function
Log(vodka price)	-0.500***	-0.838**	-0.444***	-0.700**	-1.025**
peer effect, $\widehat{\delta}$:					
age 18-29	1.404***	1.404***	1.404***	1.223***	1.275***
age 30-39	0.833***	0.833***	0.833***	0.709**	0.845**
age 40-49	0.305**	0.305**	0.305**	0.354	0.413
age 50-65	0.205	0.205	0.205	0.367	0.465
Instruments for price	(1)	(2)	(1)+(2)	(2)	(2)

Note: Sets of instruments: (1) Kink in excise tax, (2) Regional regulation
* significant at 10%** significant at 5%; *** significant at 1%

The description of utility parameters above does not offer a full picture of what happens with consumer 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 into 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

as follows: During the lengthy period in my analysis, Russia was in a period of transition. During 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 consumers expectations about future value function are noisy, possibly not correlating with current state variables or having a strong effect on consumer decision. In this case, even if in reality consumers are forward-looking, an estimated "myopic" indirect utility may be a good enough approximation of the choice-specific value function. Table A5 in the appendix illustrates this point. My data imply that in this case, consumers should expect a significant mean reversion in price movement. According to column 2 of Table A5, a ten percent change in price today is associated with only a four percent change in the expected price next year.

habits (and other state variables). To do this, I simulate consumer response to a permanent fifty percent increase in price for the 5-year period after the price change.

Figure 3 illustrates the decomposition of the cumulative response to the change in price for males age 18-29 for the myopic model, base RK specification. Dashed lines show the effect of a price increase on myopic consumers 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 consumers. The model predicts a decrease in the proportion of heavy drinkers by 6 percentage points, from 23.3 percent to 17.2 percent over five years. Taking into account only peer effects or only habit formation leads to a prediction of smaller changes (4 percentage points in case with (only) habits and 5 percentage points in case with (only) peer effects). Finally, own price elasticity results in a one-time change of 3 percentage points, which is approximately half of the cumulative effect.

Figure 4 below illustrates the simulated effect of an increase in price for myopic and forward-looking consumers in different age strata. In this example, I work with estimates obtained for the myopic model in base RK specification and for the forward-looking model in base IV specification.[34]

According to the base myopic model in five years after the introduction of a price-raising tax, the proportion of heavy drinkers will decrease by one-fourth. The effect is higher for younger generations because of the non-trivial social multiplier. In the base model with forward-looking assumptions on consumer behavior, the predicted magnitude of change in the proportion of heavy drinkers is 1.5 times larger.

5.1 The effect of a change in vodka price on mortality rates

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

Alcohol-related mortality stands for 45 to 60 percent of deaths of Russian working-age men (see Leon et all 2007, Zaridze et al 2009). The pattern of alcohol-related mortality in Russia differs significantly from that in Western Europe or in US. Whereas the death pool associated with alcoholism or chronic diseases in Russia is relatively low, the mortality from hazardous drinking such as accidental poisoning, alcohol-related accidents and injures is extremely high (see Shkolnikov and Messe, 1996, Nemtsov, 2003, Leon et all 2007, Zaridze et al 2009). The largest contributors to alcohol-related mortality are poisoning, accidents and injures, the second largest contributor is cardiovascular diseases, such as sudden heart stop under alcohol intoxication or stroke. The main death burden that comes from excessive alcohol consumption lies on males of age 18-50 for which hazardous drinking is prevalent (see Nemtsov, 2003, Leon et all 2007, Zaridze et al 2009, Shkolnikov et al 2013).

[34]The base IV estimates for myopic model predict a price response that lies in the middle between predictions of these two models.

Table 9 shows the estimates of the effect of heavy drinking on hazard of death.[35] The effect of heavy drinking is highly heterogeneous by age. The hazard of death for heavy drinkers age 18-29 is 9.9 times higher than for other males of the same age. The hazard of death for heavy drinkers of age 30-39 is 5.5 times higher. The hazard of death for heavy drinkers of age 40-49 is 1.8 times higher. There is no statistically significant difference between hazard rates for heavy drinkers and non-heavy drinkers of age 50-65. Absence of correlation between hazard of death and alcohol consumption for oldest age cohort might be due to the fact that people with serious illness consume little alcohol and at the same time have higher risk of death. The bias due to this confounding factor is especially high for old people with high rates of chronic diseases, cancer and other illnesses. Even when controlling for observable health indicators, the unobservable differences in health may drive this result. Also, because regression estimations are done for a relatively-short period of 19 years, they do not capture very long run (negative) consequences of alcohol consumption.

Using hazard-of-death regression estimates, I simulate the effect of a change in vodka price on mortality rates.

Figure 5 shows the simulated effect of increasing the price of alcohol on mortality rates for males of the three youngest age strata. The simulated effect (in case of myopic consumers) of introducing a 50 percent tax is a decrease in mortality rates by one-fifth (from 0.45 percent to 0.36 percent) for males age 18-29 years, by one-seventh (from 0.71 percent to 0.62 percent) for males age 30-39 years, and by one-twentieth for males age 40-49 years. There is little immediate effect on the mortality of males of older ages. In other words, a 50 percent increase in the price of vodka would save 30,000 (male) lives annually. This is a lower bound (in magnitude) estimate of the effect. Under the "forward looking" assumption as well as in the other specification of the myopic model (IV regression) the effect of this policy is more than 50,000 saved lives.

I find also that when agents have bounded rationality (that is, do not take into account the effect of consumption on hazard of death), the value of saved lives overweight the losses in consumer and producer surpluses, and in result an increase in vodka price by 50 percent improves welfare. Besides, under certain assumptions about consumer utilities, a tax increases consumers welfare even for fully-rational agents. (See Online Appendix for elaboration and discussion of this result).

[35] Table OA4 in online appendix reports estimates of hazard of death by different causes of death. Unfortunately, the causes of death are reported in less than 60 percent of death cases. RLMS recorded 6 causes of death, namely heart attack, stroke, external causes of death (accidents, injuries and poisoning), cancer, tuberculosis, and "other" causes. Splitting death events in different groups reduces of power of the tests, and increases standard errors of coefficients. For young generations correlation between heavy drinking and hazards of death is positive for all causes but tuberculosis, and statistically significant for death due to accidents and poisoning, due to other reasons. The only positive correlation between heavy drinking and risk death for older age cohort (of age 50-65) is found for hazard of death due to poisoning, violence and accidents. Table OA5 in online appendix reports of hazard of death with different measures of heavy drinking. Results are the same.

6 Identification Assumptions, Robustness checks and Discussion

6.1 Discussion of Identification Assumptions

In this section, I discuss the identification assumptions of price elasticity estimation in my model. Discussion of identification assumptions of peer effect and habits can be found in the online appendix (see Sections 8.1 and 8.2).

Tables 7 and 8 show results of the F-test for the relevance of instruments in both fuzzy RK and IV regressions correspondingly. Column 7 of table 7 and table 8 show F-statistics for base RK and IV regressions correspondingly. F-statistics equal 27.15 for the base RK and 43.3 for the base IV regression. Both F-statistics exceed 10, so instruments are strong. In other global polynomial RK regressions, F-statistics are in range from 27 to 51 depending on specification (see columns 8-10 of table 7). In the local polynomial RK regression F-statistics is equal to 8, due to small sample size and lack of power to provide corresponding test.

The identification of RK estimates relies on several additional assumptions.

First, the price policy regime should have a kink at year 2011. Second, predetermined covariates that affect outcome Y should have a smooth - at year 2011- profile.

I test these assumptions by checking for presence of kink and discontinuity from the following regression

$$Y_{mt} = \alpha_0[I(after2011)_{mt} \times t] + \alpha_1 t + \alpha_2 t^2 + \eta_r + u_{mt} \tag{24}$$

and

$$Y_{mt} = \alpha_0[I(after2011)_{mt}] + \alpha_1 t + \alpha_2 t^2 + \eta_r + u_{mt} \tag{25}$$

where Y_t stands for prices and alcohol consumption variables as well as pre-determined characteristics (average educational level, income, employment). Here t and t^2 stand for time and time squared and η_r stands for regional fixed effects. Coefficient α_0 in regression (18) shows the size of the kink; coefficient α_0 in regression (23) shows the size of the jump in year 2011. The regressions are estimated for the sample from 2000 to 2011.

Table 10 shows the estimation results. It shows a statistically significant kink (α_0) for regressions with price and alcohol consumption, but no evidence of kink in regressions where Y stands for demographic and socioeconomic characteristics. It also shows no evidence of a jump, except for one regression.[36]

In addition, I perform a simulation experiment where I move a placebo date of kink from year 2006 till year 2013 and estimate regression (23) above for the sub-sample of years within an interval of three years from the placebo kink date.[37] Figure 7 shows the levels

[36] α_0 is statistically significant in one regression where educational level is the dependent variable.

[37] In these regressions, I add linear function of time instead of quadratic polynomial function of time.

and 95 percent confidence interval of α_0 for the regressions with different placebo dates of kink. With the presence of kink one should expect that graphs should have a U (or inverse U - depending on the sign of the kink) form with the top (bottom) around 2011. Indeed, Figure 7 shows exactly this pattern for regressions with prices, but not for demographic or socioeconomic characteristics.

6.2 Alternative Elasticity Estimates

Table A3a in the appendix presents point estimates of elasticity for alternative specifications. Table A3a shows RK and IV estimates for different sets of instrumental variables with different set of regional fixed effects, which are included in the regression, as well as under a different assumption about price movement in the forward-looking model. All estimates lie in a range from -0.066 to -1.787, with a mean of -0.823 and median of -0.802.

Table A3b shows RK estimates for alternative definitions of heavy drinkers. In first model heavy drinkers are defined as those who belong to top 25% by alcohol intake within every 10 years age cohort. In second model, heavy drinkers are defined as those who belong to 50% by alcohol intake. In third model heavy drinkers are defined as those who belong top 25% by frequency of alcohol consumption (days per week). According to table A3b, the price elasticities of heavy drinking for these three models are in range from -0.36 to -0.61.

Table A3c shows RK estimates for regional sales of alcohol. According to table A3c, the price elasticities of alcohol consumption are in range of -0.56 to -0.81.

Table A3d in the appendix presents reduced-form elasticity estimates from a linear global RK regression

$$Share\ of\ heavy\ drinkers_{mt} = \theta log(Price)_{mt} + \Psi' X_{mt} + u_{mt} \tag{26}$$

The variable $Share\ of\ heavy\ drinkers_{mt}$ stands for the share of heavy drinkers in the neighborhood. The RK specification is similar to the global polynomial specification discussed above.The set of control variables X_{mt} includes a second order polynomial of time (running) variable and averages of the following variables: education, work status, lagged I(smokes), lagged I(heavy drinker) interacted with quadratic function of age , health status, weight, age, age squared, marital status, religion, size of family and log(family income).

In addition, I obtain reduced estimates of elasticity for different age cohort groups.[38]

Table A3d in the appendix reports price elasticity estimates. Columns 1-4 show the estimates for elasticity for two kernel types, triangular and rectangular, and for two instruments, $Excise\ tax\ on\ vodka$ and $run * I(after2011)$. The elasticity estimates lie in a range from -0,053 to -0,113, and all are statistically significant.[39] In the main specification, regression

[38]The regression specification in this case as follows: $Share\ of\ heavy\ drinkers_{mt} = \sum \theta_c log(Price)_{mt} \delta_c + \Psi' X_{mt} + u_{mt}$, where δ_c are cohort fixed effects.

[39]Remind that Table OA1 presents results of linear regression whereas main specification regressions (Tables 5-7) present results of logit regression, that make direct comparison of coefficients senseless.

with triangular kernel the elasticity of heavy drinking with respect to price of vodka is equal to -0.092. This means that increase in price of vodka by 50 percent will result in an (immediate) decrease in the share of heavy drinkers by 4.6pp: from an average of 25pp to 20.4pp, i.e. an effect similar in magnitude to those discussed in Figure 4. Columns 5-8 of Table A3d show estimates for cohort specific elasticities. In all specifications, the elasticities for the two youngest cohorts of age 18-29 and age 31-40) are higher than that for older cohorts, which coincides with the observation of higher social multiplier effects for younger generations.

6.3 Comparison with elasticity estimates from other studies

The simulation example discussed in the Results Section (see Figure 4) implies that the short-run elasticity of heavy drinking equals -0.44 and long run elasticity equals -0.52.[40] [41]

These estimates are comparable with elasticity estimates that come from meta-analysis studies (See Leung et al 1993, Wagenaar et al. (2009). The latest meta-analysis study, Wagenaar et al, 2009 reports an average elasticity of -0.44 for total alcohol intake, and of -0.28 for heavy drinking.[42] In our data, the fact that the average price elasticity of overall alcohol intake is -0.44 implies that the price elasticity of my measure of heavy drinking is equal to -0.53. The available estimates of elasticity of alcohol in Russia (see Andrienko and Nemtsov, 2006, Treisman, 2010) report elasticities of total alcohol intake in range from -0.145 to -0.67. Again, our estimates are in a range between these two numbers.

Finally, when constructing a measure of heavy drinking I use data on both official and home-made alcohol (moonshine or samogon) and thus take in account all possible substitution effects. RLMS has data on consumption of various alcoholic beverages including moonshine. Moonshine consumption is legal in Russia so there is no reason to expect high (compared to other alcohol) under-reporting of moonshine consumption. Indeed, moonshine (an inferior good) became less popular in Russia. According to RLMS data, the average - across all years - share of samogon in total alcohol intake equals 7 percent . Moreover since year 2009 the share of samogon does not exceed 5 percent, and, importantly, there is no increase in share of samogon in 2011-2014, when of price of vodka increased significantly (see Figure A3 in appendix).

6.4 Robustness of Dynamic Model Assumptions

First, I did not model the idea that consumers probably correctly estimate their hazard of death, and so I now take this into account. I verify the robustness of the results after ac-

[40] In Figure 4 elasticities are as follows: $SR\ Elasticity = \frac{\%Share\ of\ heavy\ drinkers}{\%price} = \frac{(0.233-0.181)/0.233}{0.5} = 0.44$; $LR\ Elasticity = \frac{\%Share\ of\ heavy\ drinkers}{\%price} = \frac{(0.233-0.172)/0.233}{0.5} = 0.52$

[41] As a reminder, because I use logit regression, the coefficients in the regression estimates themselves are not informative.

[42] Own price elasticities for different kinds of alcoholic beverages suggest higher elasticities than that for total alcohol intake because of the substitution effect.

counting for this factor. In this robustness experiment, a consumer has discounting factor $\beta\lambda(t,s)$, where hazard rates depends on state variables, and also on a consumer's decision about heavy drinking. The results of this estimation are presented in Table A4 in the appendix. Again, utility parameters do not differ from those shown above because actual hazard of death is very small, especially for the younger generation.

6.5 Extension

In this section, I provide an informal toy test of which model, myopic or forward-looking, does a better job of explaining the 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 parametric restrictions in order to obtain identification from the model. In addition, in order to estimate of the discounting factor, I have to impose additional strong simplifications. Therefore, this informal test should be treated at most as only suggestive. More detailed elaboration and estimation of the discounting factor is out of the scope of this paper.

In the main text of this paper, I use sequential procedure of estimation of the parameters, which provides little guidance regarding which β is better in describing the 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.

I test separately whether agents behave as forward-looking or as myopic when dealing with three different factors - price of alcohol, self-evaluated health status and hazard of death. All of these factors may serve as instruments for policymakers who can change taxation and provision of information about future consequences of unhealthy habits.

In all three models, I assume that consumer utility depends on two variables. The first variable, habits, is similar in all three models. The second variable is model-specific: the price of vodka in the first model, self-evaluated health status in the second model and hazard of death in the third model. In the model with hazard of death I estimate a model in which consumer utility depends on habits and hazard of death. The hazard of death is evaluated from the Cox hazard model (see equation 23) as a function of alcohol consumption, self-evaluated health status, indicators that the person had surgery last year, marital status and employment status.

To implement nested fixed points algorithm, I have to discretize variables that I use in these models. Habits already were discrete (with two values), price of vodka, hazard of death and health status are discretized and have five values.

Table A6a in the appendix shows point estimates of β that brings the maximum of likelihood functions.

The log likelihoods for both myopic and forward-looking models are almost the same, with a slightly-higher likelihood in the myopic model with the price of alcohol and slightly-higher likelihood in the forward-looking model in the model with health status. In the model with price, β that maximizes the likelihood function equals to 0. In the model with health status, β equals 0.99. In the model with hazard of death, the β equals 0.47.

The result that agents discount so heavily the future flow of utilities may be explained by the fact that heavy drinkers do not behave fully rationally. An other explanation is an unpredictability of the factors that enter in utilities in my models.

Since the price of alcohol is volatile in Russia, agents may not be able correctly predict prices, and, as a result may ignore future prices in their current decision on heavy drinking.

A similar issue is relevant for the hazard of death. In contrast to the model with self-evaluated health status, in which agents directly observe their own health and use previous experience, the effect of drinking on the hazard of death is more difficult to predict.[43] The largest contributors to alcohol-related mortality of Russian working-age males are poisoning, accidents and violence, but not alcohol-related diseases, such as liver cirrhosis (see Zaridze et al 2009).[44] Poisoning, accidents and violence may happen at random with healthy people too, and therefore it may be hard to predict the hazard of death based on, say, personal health evaluation. Indeed, Table A6c shows that, although heavy drinking, as well as bad health, significantly increase the hazard of death, these effects do not interact. In regression with binge drinking and (bad) health variables, interaction terms between bad health and binge drinking are statistically insignificant. Besides, excluding heavy drinking from regressions does not change coefficients on (bad) health indicators. In contrast, some factors that can potentially help heavy drinkers not to get in accidents or other risky events, such as marriage and religion, do decrease the harmful effect of heavy drinking. The interaction terms between heavy drinking and these factors are negative.

To further check results, I provide a simple reduced-form test for myopic behavior similar to that in Chaloupka (1991). I regress the indicator of heavy drinking on current prices of vodka as well as on lead and lag prices of vodka (instrumenting them by current, lead, and lag of excise tax of vodka). I find that heavy drinking depends on current prices, but not on lead prices (see Table A6b). This result confirms the observation that consumers do not take future prices into consideration when deciding on heavy drinking. In contrast, when I regress indicator of heavy drinking on current, lead and lag health status, I find that not only current, but also lead health affect current alcohol consumption: Both coefficients on current and lead health status are statistically significant (see Table A6b).

[43] Arcidiacono et al (2007) studied decisions on smoking and alcohol consumption for old adults in US. In their model the utility of agents depends on health status, alcohol consumption and smoking. They find that likelihood function for $\beta = 0.9$ is similar to, still slightly exceeds that for $\beta = 0$, and β that brings maximum to likelihood equals to 0.91. In contrast to our study old adults in US obtain strong signals of increasing risk of death due to heavy drinking through changes in health, and thus react correspondingly.

[44] According to official statistics, *one third* of deaths of working-age males in Russia are due to external causes such as alcohol poisoning, accident, and violence that mainly happen during alcohol intoxication (see Rosstat data, www.gks.ru).

7 Conclusion

In this paper, I estimate a dynamic model of drinking behavior that incorporates several important determinants of drinking such as price of alcohol, neighborhood (peer) effects and drinking habits.

I fit the model to Russian micro level data. The nature of Russian data allows me to identify several key parameters of the model.

To estimate the price elasticity of heavy drinking, I explore a kink in the policy regime of excise tax on vodka. In 2011 the Russian government introduced a new tax policy regime. Before 2011 the excise tax on vodka was growing proportionally to CPI index. Since 2011 the tax growth rate twice exceeded the CPI growth. I use the kink in tax policy regime to apply regression kink estimation to establish the causal relationship between price and drinking. I show that RK estimates are similar to the results of instrumental variables regressions where variation in regulations of regional alcohol markets was used as instrument for price of alcohol.

The clustered structure of the dataset I use allows me to find the effect of close neighbors (peers) on individual drinking behavior. In particular, I show that neighbors indeed affect individual decision-making regarding drinking behavior by documenting a strong increase in alcohol consumption around the birthday of neighbors.

These results are especially important from the policy perspectives, since alcohol consumption is a big problem in Russia itself. Over the past twenty years, Russia has experienced one of the largest historical surge in mortality during peace time, and it is widely attributed to heavy alcohol consumption.

I find that the probability of being a heavy drinker is (relatively) elastic with respect to the price of alcohol. I also find that peers play a significant role in the decision-making regarding drinking of Russian males below age 40. The presence of a social multiplier results in significantly higher elasticity of alcohol consumption for younger cohorts. Finally, I find that the assumption that consumers are forward-looking gives higher estimates of price elasticity compared to "myopic" case.

To illustrate this finding, I estimate the impact of public policy (specifically, higher taxation) on the demand for heavy drinking and consequently on mortality rates. I simulate the effect of imposing the tax that increases the price of vodka by 50 percent. The myopic model predicts that five years after the introduction of the price-raising tax, the proportion of heavy drinkers will decrease by roughly one-forth – 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 the 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 assumptions 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 consumers, the effect of a change in price is higher by roughly 30 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. A fifty percent tax on the price of vodka will save 30,000-50,000 (male) lives annually, or 1 percent of young male adult lives in six years.

References

Aguirregabiria, Victor, and Pedro Mira. 2007. "Sequential estimation of dynamic discrete games." *Econometrica,* 75(1), 1.

Aguirregabiria Victor, and Pedro Mira. 2010. "Dynamic discrete choice structural models: A survey." *Journal of Econometrics*, 156, (2010) 38-67.

Ai, Chunrong, and Xiaohong Chen. 2003. "Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions." *Econometrica*, 71(6), 1795–1843.

Andrienko Yury, and Alexandr Nemtsov. 2006. "Estimation of Individual Demand for Alcohol," CEFIR/NES Working paper #89, Centre for Economic and Financial Research at New Economic School, Moscow, 2006.

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

Arcidiacono, Peter, Holger Sieg, and Frank Sloan. 2007. "Living rationally under the volcano? An empirical analysis of heavy drinking and smoking." *International Economic Review* 48, 3765.

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

Bajari, Patrick, Victor Chernozhukov, Han Hong, and Denis Nekipelov, 2015, Identification and Efficient Semiparametric Estimation of a Dynamic Discrete Game, NBER Working Paper, w21125.

Bajari, Patrick, Han Hong, Krainer J., and Denis Nekipelov. 2011. "Estimating Static Models of Strategic Interactions.", *Journal of Business and Economic Statistics*.

Bajari, Patrick, Han Hong, and Denis Nekipelov. 2013. "Econometrics for Game Theory", in Advances in Economics and Econo metrics: Theory and Applications, 10th World Congress, 2013.

Bajari, Patrick, Han Hong and Stephen P. Ryan. 2009. "Identification and Estimation of a Discrete Game of Complete Information." Econometrica 78(5): 1529-68

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

Bhattacharya, Jay, Christina Gathmann, and Grant Miller, 2013, "The Gorbachev Anti-Alcohol Campaign and Russia's Mortality Crisis." *American Economic Journal: Applied Economics* Vol. 5, Issue 2, April 2013

Brainerd, Elizabeth and David 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

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

Becker, Gary, Michael Grossman, and Kevin Murphy. 1991. "Rational Addiction and the Effect of Price on Consumption." *American Economic Review*, 81 (2), 237.

Berry, Steven, James Levinsohn, and Ariel Pakes, 1995 "Automobile Prices in Market Equilibrium", *Econometrica*, Vol. 63, No. 4. (Jul., 1995), pp. 841-890

Card David, David Lee, Pei Z, Weber A, 2015, "Inference on Causal Effects in a Generalized Regression Kink Design." Econometrica, 83(6): 2463-2483, Nov 2015.

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

Chenet, Laurent, Martin McKee, David Leon, Vladimir Shkolnikov, and Sergei 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. and 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.

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

Dreyfus, M. and Viscusi, K. (1995). Preferences and Consumer Valuations of Automobile Safety and Fuel Efficiency. Journal of Law and Economics, 38(1), 79–105.

Fang, Hanming and Yang Wang. 2010. "Estimating Dynamic Discrete Choice Models with Hyperbolic Discounting, with an Application to Mammography Decisions", NBER Working Paper No. 16438.

Gotz, G. A. and J. J. McCall, 1984. "A dynamic retention model for air force o¢ cers: Theory and estimates," RAND, R-3028-AF.

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

Gruber, Jonathan and Botond Kőszegi. 2001. "Is Addiction 'Rational?' Theory and Evidence." *Quarterly Journal of Economics* (2001), 116(4), pp. 1261-1305.

Graham, Brian. 2008. "Identifying social interactions through conditional variance restrictions." *Econometrica* 76 (3): 643 - 660, 2008.

Hausman, J. (1979). Individual Discount Rates and the Purchase and Utilization of Energy-Using Durables. Bell Journal of Economics, 10(1), 33–54.

Hotz, Joseph and Robert A. Miller. 1993. "Conditional Choice Probabilities and Estimation of Dynamic Models", *Review of Economic Studies*, 60, 497.

Imbens, Guido and Kalyanaraman K. Optimal Bandwidth Choice for the Regression

Discontinuity Estimator. Review of Economic Studies, 2012;79(3):933-959

Kueng, Lorenz and Evgeny Yakovlev, 2015, "How Persistent Are Consumption Habits? Micro-Evidence From Russia", 2015, NBER working paper 20298

Lee David, Lemieux Thomas, 2010, "Regression Discontinuity Designs in Economics." Journal of Economic Literature, 48(2), 281-355.

Leon, David, Lyudmila Saburova, Susannah Tomkins, Evgueny Andreev, Nikolay Kiryanov, Martin McKee, and Vladimir M Shkolnikov. 2007. "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.

Magnac, Thierry and David Thesmar. 2002. "Identifying dynamic discrete decision processes." *Econometrica* 70, 801816.

Maurin, Eric and Julie Moschion. 2009. "The Social Multiplier and Labor Market Participation of Mothers." American Economic Journal: Applied Economics, 1(1): 251-72.

Miller, Robert A. 1984. "Job matching and occupational choice". *Journal of Political Economy* 92: 1086-1120.

Moore, M. and Viscusi, K. (1990). Models for Estimating Discount Rates in Long-Term Health Risks Using Labor Market Data. Journal of Risk and Uncertainty, 31(4), 381–401.

Mullahy, John and Jody L. Sindelar. 1993. "Alcoholism, Work, and Income." *Journal of Labor Economics*, Vol. 11, No. 3. (Jul., 1993), pp. 494-520.

Nemtsov, Alexandr. 2002. "Alcohol-Related Human Losses in Russia in the 1980s and 1990s." *Addiction*, 97, pp. 1413-25.

Nevo, Aviv, 2011, "Empirical Models of Consumer Behavior", Annual Review of Economics , Volume 3 (2011), 51- 75

Pakes, Ariel. 1986. " Patents as options: Some estimates of the value of holding European patent stocks". *Econometrica* 54: 755-84.

Pesendorfer, Martin and Philipp Schmidt-Dengler, 2008, "Asymptotic Least Squares Estimators for Dynamic Games." *Review of Economic Studies*, 901-928.

Rust, John, 1987, "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurchner." *Econometrica*, 55 (5), 999.

Shkolnikov V.M., Meslé F., 1996, The Russian epidemiological crisis as mirrored by mortality trends. In: Julie DaVanzo ed., Russia's Demographic "Crisis", RAND, Santa Monica, pp. 113-161.

Shkolnikov, Vladimir M., Evgeny M. Andreev, Martin McKee, David A. Leon, 2013, Components and possible determinants of the decrease in Russian mortality in 2004-2010, Demographic Research, Vol. 28, pp 917-950, April 2013

Slinko Irina, Evgeny Yakovlev and Ekaterina Zhuravskaya. 2005."Laws for Sale: Evidence from Russia." *American Law and Economics Review*, 2005.

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

Wagenaar, A. C., Salois, M. J., & Komro, K. A., 2009,. Effects of beverage alcohol price and tax levels on drinking: A meta-analysis of 1003 estimates from 112 studies. Addiction, 104, 179-190.

Warner, J. and Pleeter, S. (2001). The Personal Discount Rate: Evidence from Military Downsizing Programs. American Economic Review, 91(1), 33–53.

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

Wolpin, K. I. 1984, "A dynamic stochastic model of fertility and child mortality". *Journal of Political Economy* 92: 852-74.

Yakovlev, Evgeny. 2008. "Political Economy of Regulation: Evidence from Russian Alcohol Market", mimeo, 2008.

Zaridze, David, Paul Brennan, Jillian Boreham, Alex Boroda, Rostislav Karpov, Alexander Lazarev, Irina Konobeevskaya, Vladimir Igitov, Tatiana Terechova, Paolo Boffetta, and Richard Peto. 2009. "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.

Table 3: Effect of heavy drinking on transition of income, marital status and health variables

	(1)	(2)	(3)	(4)	(5)	
			Dependent variables: Y_{it+1}			
	Log income	Log family income	Health evaluation	I(surgery last year)	I(married)	I(employed)
I(heavy drinker)	-0.121***	-0.153***	-0.017***	0.074***	-0.062***	-0.013
	[0.015]	[0.015]	[0.005]	[0.023]	[0.020]	[0.015]
Y_{it}	0.542***	0.513***	0.548***	0.828***	3.090***	1.962***
	[0.005]	[0.005]	[0.005]	[0.040]	[0.022]	[0.018]
Constant	1.868***	2.065***	1.511***	-1.968***	-1.287***	-0.658***
	[0.022]	[0.024]	[0.019]	[0.013]	[0.015]	[0.014]
Observations	57,276	61,402	60,835	61,330	58,430	61,396
R-squared	0.315	0.277	0.298			

Notes: Robust standard errors in brackets; ***p<0.01, **p<0.05, * p<0.1.
Heavy drinkers are defined as those who belong to top quarter by total alcohol intake.
Columns (1) - (3) show results of OLS regressions $Y_{it+1} = \alpha + \theta I(heavy\, drinker)_{it} + \beta Y_{it} + u_{it}$.
Columns (4)-(6) show results of probit regressions $Pr(Y_{it+1} = 1) = \Phi(\alpha + \theta I(heavy\, drinker)_{it} + \beta Y_{it})$.

Table 4: Summary statistics

Variable	Obs	Mean	St. D.	Min	Max
Panel data (males)					
I(drunk more than 150 gr)	78235	.2503	.433	0	1
Log(family income)	78507	3.988	1.817	0	9.787
Age	78507	39.02	13.10	18	65
Age squared	78507	1695	1073	324	4225
I(diseases)	74454	.343	.474	0	1
I(big family)	78507	.194	.395	0	1
Lag I(heavy drinker)	61403	.255	.436	0	1
Lag I(Smokes)	61563	.633	.481	0	1
I(employed)	78453	.712	.452	0	1
I(college degree)	78409	.237	.425	0	1
Body weight	78071	77.2	13.87	35	250
I(big family)	78507	.194	.395	0	1
I(Muslim)	78507	.081	.272	0	1
Alcohol intake	78235	99.78	125.1	0	2469
I(physical training)	67483	0.174	0.379	0	1
I(drink tea)	22415	.966	.180	0	1
I(drink coffee)	22409	.69	.458	0	1
I(surgery last year)	78451	.0323	.177	0	1
Health self-evaluation	78084	3.378	0.68	1	5
Prices and Regulation					
log(price of vodka)	714	.449	.351	-1.02	1.36
CPI	714	75.43	48.74	4.41	262
Excise tax rate, vodka	600	190.3	116.6	55	500
Sum of regulations	495	.461	.738	0	3
Production regulation					
Additional document	483	.102	.294	0	1
Premises regulation	483	.129	.329	0	1
Retail regulation					
Additional document	483	.156	.357	0	1
Excise machine	489	.085	.274	0	1
Survival regression data					
Death cases, male, >17 years	12169	.045	.207	0	1
Drunk more than 150 gr	12167	.239	.29	0	1
Smokes	12169	.612	.439	0	1
Bad Health (self-evaluation)	12167	0.074	0.194	0	1
Employment	12163	.698	.408	0	1
Log(family income)	12169	4.035	1.38	0	7.39
College degree	12159	.248	.408	0	1
Body weight	12163	77.23	13.16	40	215

Notes: Prices and regulation are summarized from data on municipality×year cells. Survival regression data is " between" individual-level data.

Table 5: Distribution of variables

Panel A. Distribution of # of peers in peer groups

# of peers in peer group	(Peer group)-level data Freq.	percent	Cum. %	Individual - level data Freq.	percent	Cum. %
2	5,601	36.44	36.44	11,202	15.73	15.73
3	3,951	25.70	62.14	11,853	16.65	32.38
4	2,205	14.34	76.48	8,820	12.39	44.76
5	1,161	7.55	84.04	5,805	8.15	52.91
6	641	4.17	88.21	3,846	5.40	58.31
7	351	2.28	90.49	2,457	3.45	61.77
8	232	1.51	92.00	1,856	2.61	64.37
9	142	0.92	92.92	1,278	1.79	66.17
10	115	0.75	93.67	1,150	1.61	67.78
11	97	0.63	94.30	1,067	1.50	69.28
12	72	0.47	94.77	864	1.21	70.49
13	64	0.42	95.19	832	1.17	71.66
14	60	0.39	95.58	840	1.18	72.84
15	49	0.32	95.90	735	1.03	73.87
16	47	0.31	96.20	752	1.06	74.93
17	36	0.23	96.44	612	0.86	75.79
18	41	0.27	96.70	738	1.04	76.82
19	33	0.21	96.92	627	0.88	77.71
20 and more	474	3.08	100.00	15,876	22.29	100.00
Total	15,372	100.00		71,210	100.00	

Notes: 7027 peer groups that contain 1 peer are excluded

Panel B. Distribution of deaths events by age and causes of deaths

	Number of deaths				Share in deaths with reported cause (in percent)			
Age cohort:	18-29	30-39	40-49	50-65	18-29	30-39	40-49	50-65
Cause of death								
Heart attack	2	7	24	63	6.25	12.07	23.30	25.1
Stroke	1	3	15	66	3.13	5.17	14.56	26.29
Cancer	2	4	9	70	6.25	6.90	8.74	27.89
Poisoning, injuries, accidents	20	26	30	22	62.5	44.83	29.13	8.76
Tuberculosis	1	3	2	4	3.13	5.17	1.94	1.59
Other	6	15	23	26	18.75	25.86	22.33	10.36
Not reported	12	28	46	96				
Total	44	86	149	347				

Table 6: Consumer utility parameters

	Agent's (per-period) Utility			Agent's (per-period) Utility	
	$\beta = 0$	$\beta = 0.9$		$\beta = 0$	$\beta = 0.9$
Peer effect, $\widehat{\delta}$:			Log (family income)	-0.033***	-0.018***
age 18-29	1.444***	1.269***		[0.012]	[0.021]
	[0.215]	[0.502]	*Age*	0.127***	0.115***
age 30-39	0.833***	0.709**		[0.011]	[0.045]
	[0.154]	[0.347]	Age^2	-0.001**	-0.001***
age 40-49	0.326**	0.361		[0.0001]	[0.0006]
	[0.154]	[0.414]	*Body Weight*	0.008***	0.006***
age 50-65	0.209	0.374		[0.001]	[0.002]
	[0.215]	[0.597]	*I(diseases)*	-0.017	-0.006
Habit:				[0.026]	[0.057]
Lag I(heavy drinker)	1.456***	1.401***	*I(big family)*	0.062**	0.059***
	[0.04]	[0.071]		[0.031]	[0.096]
Lag I(heavy drinker)$\times$*age*	-0.028**	-0.016***	*Lag I(smokes)*	0.505***	0.427***
	[0.015]	[0.015]		[0.029]	[0.063]
Lag I(heavy drinker)$\times age^2$	0.0004***	0.000***	*I(work)*	-0.155***	-0.176***
	[0.00017]	[0.000]		[0.084]	[0.084]
			I(college degree)	-0.170***	-0.188***
				[0.032]	[0.089]
			I(Muslim)	-0.272***	-0.171***
				[0.063]	[0.067]
municipality$\times$year FE	*Yes*	*Yes*			
Peers mean characteristics	*Yes*	*Yes*			
Observations	50,763	50,763			

Notes: Bootstrapped standard errors in brackets.

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

Table 7: Estimates of price elasticity. Myopic consumers. RK Regression.

	(1)	(2)	(3)	(4)	(5)	(6)
	MU (du/dlogP)					
log(vodka price)	-0.500***	-0.595***	-0.403**	-0.451***	-0.306	-0.465
	[0.224]	[0.199]	[0.169]	[0.157]	[0.379]	[0.370]
Time	0.062	0.074*	0.055*	0.061*	0.026	0.048
	[0.015]	[0.014]	[0.010]	[0.010]	[0.036]	[0.036]
$Time^2$	0.005**	0.006***	0.003**	0.003***		
	[0.002]	[0.002]	[0.001]	[0.001]		
CPI	-0.446	-0.520*	-0.459**	-0.499**	-0.263	-0.357
	[0.297]	[0.286]	[0.225]	[0.220]	[0.325]	[0.333]
I(city)	0.097	0.103	0.152*	0.154*	0.104	0.119
	[0.092]	[0.093]	[0.083]	[0.083]	[0.125]	[0.128]
Employment	0.516	0.559	0.222	0.233	1.226*	1.312*
	[0.384]	[0.384]	[0.340]	[0.339]	[0.591]	[0.587]
Share with	-1.186***	-1.214***	-1.208***	-1.216***	-1.203**	-1.276**
college degree	[0.432]	[0.432]	[0.386]	[0.384]	[0.577]	[0.586]
Average income	-0.000	-0.000	-0.001	-0.001	-0.000	-0.001
	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]
Constant	2.467	2.877*	2.674**	2.900**	1.084	1.612
	[0.345]	[0.332]	[0.319]	[0.319]	[0.396]	[0.403]
Observations	523	523	561	561	257	257
R-squared	0.343	0.314	0.375	0.367	0.467	0.430
kernel	triangle	triangle	uniform	uniform	uniform	uniform
IV	afterXrun	excise tax	afterXrun	excise tax	afterXrun	excise tax
BW size	11	11	11	11	3	3

	(7)	(8)	(9)	(10)	(11)	(12)
	First Stage: log(vodka price)					
I(after 2011)×Time	0.165***		0.191***		0.104***	
	[0.039]		[0.034]		[0.038]	
Excise tax		0.002***		0.003***		0.001***
		[0.000]		[0.000]		[0.000]
Time	0.054**	0.017	0.037**	-0.002	0.085***	0.064**
	[0.017]	[0.022]	[0.013]	[0.018]	[0.019]	[0.026]
$Time^2$	-0.001	-0.001	-0.003**	-0.003***		
	[0.002]	[0.002]	[0.001]	[0.001]		
Socioeconomic vars	Yes	Yes	Yes	Yes	Yes	Yes
kernel	triangle	triangle	uniform	uniform	uniform	uniform
IV	afterXrun	excise tax	afterXrun	excise tax	afterXrun	excise tax
BW size	11	11	11	11	3	3
F-test	27.15	29.71	47	50.98	8.16	8.38

Notes: Bootstrapped standard errors are in brackets; * significant at 10%; ** significant at 5%; *** significant at 1%. Set of socioeconomic variables in 1st stage is the same as in 2nd stage.

Table 8: Elasticity estimates. IV regressions. Forward-looking and myopic assumptions

	(1)	(2)	(3)	(4)	(5)	(6)
	Myopic	Forward-looking				
		Per-period utility	Value function		Value function	
log(vodka price)	-0.839***	-0.730***	-1.008**	-0.832***	-1.275***	-1.235***
	[0.293]	[0.410]	[0.490]	[0.367]	[0.707]	[0.470]
Share with college degree	-1.570***	-0.849**	-2.065***	-1.296***	-2.286***	-2.186***
	[0.413]	[0.379]	[0.520]	[0.402]	[0.715]	[0.525]
Average income	0.060	0.032	0.184**	0.088	0.115	0.088
	[0.074]	[0.064]	[0.087]	[0.068]	[0.116]	[0.089]
Employment	0.963***	0.654**	0.623*	0.642**	1.363***	1.193***
	[0.365]	[0.297]	[0.364]	[0.305]	[0.508]	[0.395]
I(city)	0.166*	0.069	0.255**	0.137	0.313**	0.248**
	[0.091]	[0.080]	[0.117]	[0.088]	[0.149]	[0.116]
Share with diseases	0.988***	0.561**	1.412***	0.874***	1.335***	1.214***
	[0.254]	[0.232]	[0.298]	[0.237]	[0.412]	[0.312]
Constant	-0.596	-0.410	-0.298	-0.369	-0.502	-0.780
	[0.473]	[0.451]	[0.527]	[0.452]	[0.742]	[0.537]
Observations	415	414	414	414	414	414
F-test	43.27	43.35	43.35	43.35	43.35	43.35
Normalization		u(not drink)=0	u(not drink)=0	u(not drink)=0	u(drink)=0	u(drink)=0
Commit that price change is permanent			YES	NO	YES	NO

Notes: Bootstrapped standard errors in brackets; * significant at 10%; ** significant at 5%; *** significant at 1%

Table 9: Mortality and heavy drinking

	Coefficient	Hazard ratio		Coefficient
I(heavy drinker) age 18-29	2.296***	9.934	Log (family income)	-0.413***
	[0.467]			[0.036]
I(heavy drinker) age 30-39	1.704***	5.496	I(smokes)	0.591***
	[0.353]			[0.124]
I(heavy drinker) age 40-49	0.580*	1.786	I(college degree)	-0.076
	[0.315]			[0.132]
I(heavy drinker) age 50-65	-0.268		Body Weight	-0.005
	[0.246]			[0.004]
Bad health (self evaluation)	1.399***		I(work)	0.090
	[0.164]			[0.146]
Observations	12,125			

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

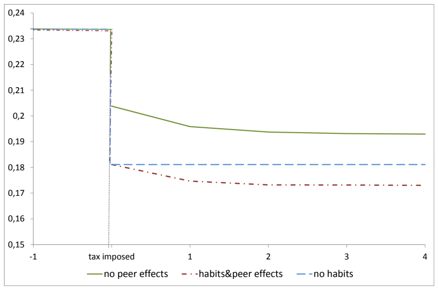
Table 10: Test for Smoothness: Price, Alcohol Consumption and Social-Economic characteristics.

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: test for kink	$\widehat{\rho_{mt}}$	Share of heavy drinkers	log(vodka price)	Employment	Share with college degree	Average income
$I(after2011) \times run$	-0.073*	-0.012*	0.226***	0.001	0.001	-1.529
	[0.042]	[0.007]	[0.033]	[0.006]	[0.006]	[2.645]
run	0.029	0.003	0.009	-0.001	0.001	14.96***
	[0.026]	[0.004]	[0.018]	[0.003]	[0.003]	[1.603]
run^2	0.004**	0.001***	-0.002	-0.000*	0.000	-0.096
	[0.002]	[0.000]	[0.001]	[0.000]	[0.000]	[0.115]
Observations	561	561	561	561	561	561

Panel B: test for jump	$\widehat{\rho_{mt}}$	Share of heavy drinkers	log(vodka price)	Employment	Share with college degree	Average income
$I(after2011)$	0.064	-0.010	0.063	-0.007	0.023**	-2.526
	[0.085]	[0.014]	[0.069]	[0.011]	[0.010]	[5.362]
run	-0.012	0.000	0.089***	0.001	-0.003	14.807***
	[0.025]	[0.004]	[0.018]	[0.003]	[0.003]	[1.486]
run^2	0.001	0.001***	0.005***	-0.000	0.000	-0.114
	[0.002]	[0.000]	[0.001]	[0.000]	[0.000]	[0.098]
Observations	561	561	561	561	561	561

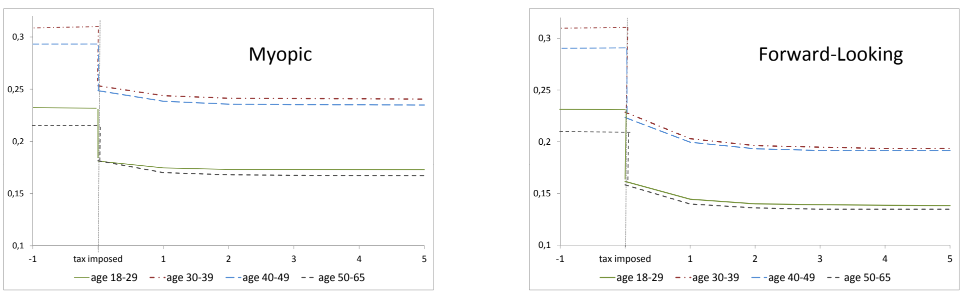
Note: Robust standard errors are in brackets. Set of controls includes regional fixed effects, CPI, I(city)

Figure 3. Effect of tax on Pr(heavy drinker), age 18-29



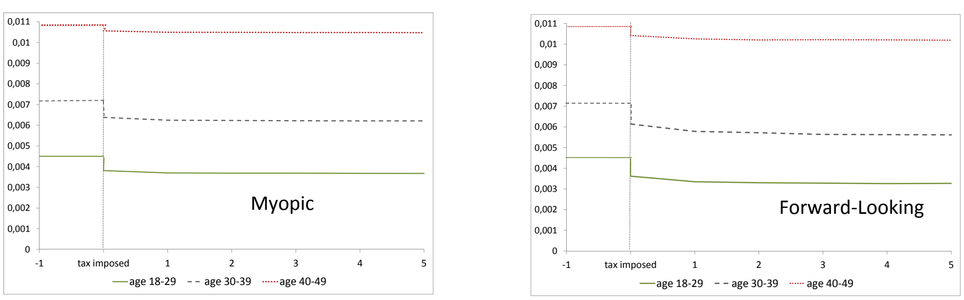
Notes: The figure shows the decomposition of the effect of 50 percent increase percent price of vodka on share of heavy drinkers among young male adults. Source: RLMS, males of age 18-29. Horizontal axis: years before and after imposing tax. Vertical axis: Share of heavy drinkers.

Figure 4. Effect of a 50 percent tax on the share of heavy drinkers.



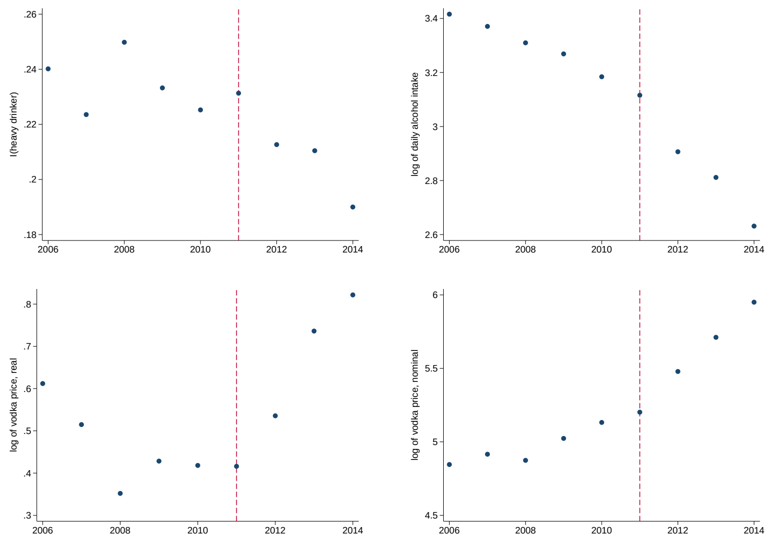
Notes: The figure shows the simulated effect of 50 percent increase in the price of vodka on share of heavy drinkers in different age cohorts. Source: RLMS, males of age 18-65. Horizontal axis: years before and after imposing tax. Vertical axis: Share of heavy drinkers.

Figure 5. Effect of 50% tax on mortality rates.



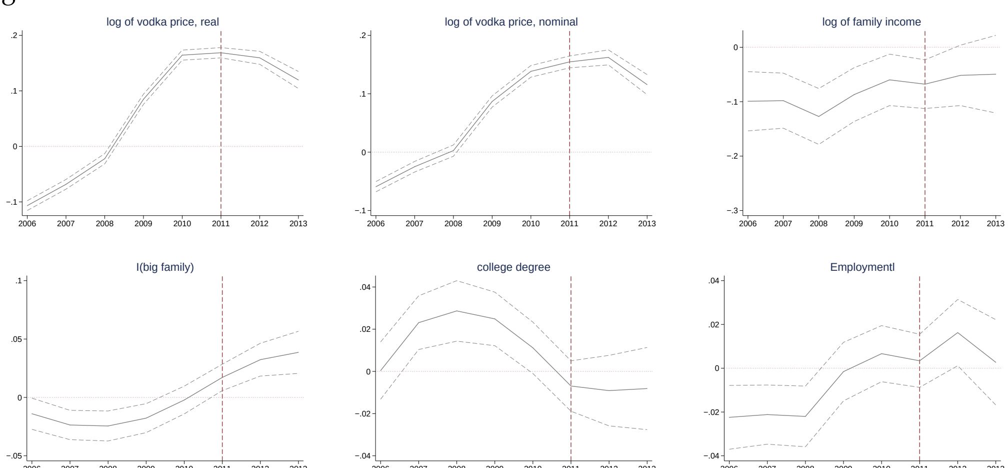
Notes: The figure shows the simulated effect of 50 percent increase in the price of vodka on mortality rates among males of age 18-49. Source: RLMS, males of age 18-59. Horizontal axis: years before and after imposing tax. Vertical axis: Share of heavy drinkers.

Figure 6. Averages (by year) of the price of vodka and of alcohol consumption.



Notes: The figures show average alcohol prices and alcohol consumption around year 2011 (kink date in the policy regime of the excise tax on vodka). Source: RLMS data, males of age 18-65.

Figure 7. Placebo for kink:



Notes: Figure 7 shows the levels and 95% confidence intervals of size of the kink (α_0) for the regressions with different placebo dates of kink. α_0 is estimated from regressions

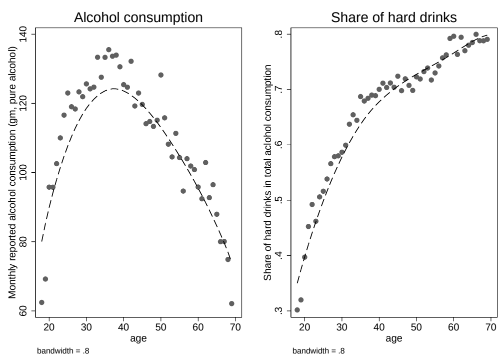
$Y_{mt} = \alpha_0[I(after\,placebo\,date) \times t] + \alpha_1 t + \alpha_2 t^2 + \eta_r + u_{mt}$. Placebo dates of kink on horizontal axis; α_0 on vertical axis.

Appendix

Figure A1. Typical dvor ("khrushevka") in Russia.

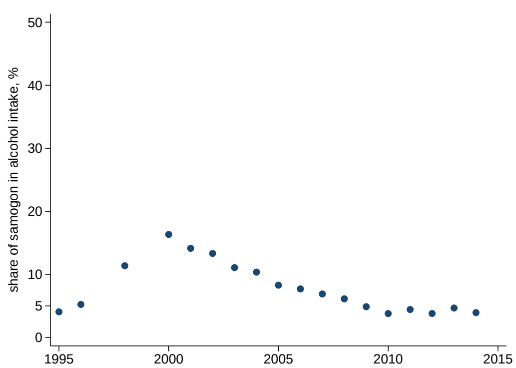
Source: www.photographer.ru (Petr Antonov)

Figure A2. Alcohol consumption: age profile



Source: RLMS, subsample of males of age 18-65.

Figure A3. Share of samogon (moonshine) in total alcohol intake.



Source: RLMS, subsample of males of age 18-65.

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

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

Table A2.Excise tax on vodka

year	Excise tax on vodka
2000	88,2
2001	88,2
2002	98,78
2003	114
2004	135
2005	146
2006	159
2007	162
2008	173
2009	191
2010	210
2011	231
2012	300
2013	400
2014	500
2015	600

Notes: Excise tax on vodka is in rubles per 1 liter of pure alcohol

Table A3a. Point estimates of price elasticity for different RK and IV specifications

				Static, RKD				
log(price of vodka)	-0.595	-0.500	-0.451	-0.403	-0.465	-0.306		
kernel	triangle	triangle	uniform	uniform	uniform	uniform		
IV	excise tax	afterXrun	excise tax	afterXrun	excise tax	afterXrun		
BW size	11	11	11	11	3	3		
				Static, IV				
log(price of vodka)	-0.839	-1.029	-1.075	-0.450	-0.202	-0.444		
Specification	IV 1	IV 1	IV 1	IV 2	IV 2	IV 3		
Regional FE		YES	YES		YES			
Regional Trends			YES		YES			
Fedokrug FE, Trends	YES			YES		YES		
				Dynamic, IV, Normalization: U(no drink=0)				
log(price of vodka)	-1.008	-0.832	-1.002	-0.824	-0.614	-0.746	-0.854	-0.702
Permanent price change	YES	NO	YES	NO	YES	NO	YES	NO
Specification	IV 1	IV 1	IV 1	IV 1	IV 1	IV 1	IV 2	IV 2
Fedokrug FE, Trends	YES	YES					YES	YES
Regional FE			YES	YES	YES	YES		
Regional Trends					YES	YES		
				Dynamic, IV, Normalization: U(drink=0)				
log(price of vodka)	-1.771	-1.275	-1.829	-1.351	-1.774	-1.296	-1.048	-1.038
Permanent price change	YES	NO	YES	NO	NO	YES	YES	NO
Specification	IV 1	IV 1	IV 1	IV 1	IV 1	IV 1	IV 2	IV 2
Fedokrug FE, Trends	YES	YES					YES	YES
Regional FE			YES	YES	YES	YES		
Regional Trends					YES	YES		

Note: IV regressions: IV1: instrument is sum of regional regulations; IV2: instruments is set of four regional regulation variables; IV3: instruments are sum of regional regulations and federal excise tax

Table A3b. Model parameters estimates under different definitions of heavy drinkers.

	(1)	(2)	(4)	(5)	(7)	(8)
	Model 1		Model 2		Model 3	
log(price of vodka)	-0.609**	-0.521*	-0.509**	-0.477*	-0.366*	-0.356
	[0.270]	[0.293]	[0.243]	[0.255]	[0.210]	[0.228]
Peer effect, $\widehat{\delta}$:						
age 18-29	1.856***		1.274***		1.108***	
age 30-39	0.420***		0.845***		0.611***	
age 40-49	-0.012		0.297*		0.197**	
age 50-65	0.507***		0.536***		0.126	
Habits	1.432***		1.658***		1.602***	
IV	excise tax	afterXrun	excise tax	afterXrun	excise tax	afterXrun
F-test	29.65	27.06	29.65	27.06	29.65	27.06

Notes: ***p<0.01, **p<0.05, * p<0.1; Robust Standard errors are in Brackets.
Heavy drinking definitions: Model (1): Top 25% by alcohol intake within 10 years age cohorts;
Model (2): Top 50% by alcohol intake; Model (3): Top 25% by days of alcohol consumption (per month)
In all models price elasticity estimates come from global RK estimates with triangle kernel.

Table A3c. Regional-level regression. RK estimates of elasticity.

	(1)	(2)	(3)	(4)	(5)	(6)
	log(sales of alcohol)					
log(price of vodka)	-0.798***	-0.815***	-0.697***	-0.722***	-0.565***	-0.564***
	[0.256]	[0.283]	[0.196]	[0.202]	[0.175]	[0.178]
run	0.033	0.036	0.021	0.024	-0.015	-0.015
	[0.040]	[0.044]	[0.028]	[0.029]	[0.032]	[0.032]
run^2	0.005	0.005	0.004**	0.004**		
	[0.004]	[0.004]	[0.002]	[0.002]		
log(CPI)	0.810	0.806	0.377	0.359	0.389	0.388
log(GDP per capita)	0.400***	0.403***	0.385***	0.389***	0.377***	0.377***
Unemployment	-0.072***	-0.072***	-0.055***	-0.055***	-0.075***	-0.075***
Population	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***
Observations	847	847	925	925	534	534
IV	excise tax	afterXrun	excise tax	afterXrun	excise tax	afterXrun
kernel	triangle	triangle	uniform	uniform	uniform	uniform
Sample	2003-2014	2003-2014	2003-2014	2003-2014	2008-2014	2008-2014
F-test	121.8	109.1	213.5	197.4	282.6	266.2

Notes: The data source: Rosstat data for 78 Russian regions. www.gks.ru.
Robust standard errors clustered at regional level are in brackets. ***p<0.01, **p<0.05, * p<0.1.

Table A3d. Reduced form elasticity estimates. RK regression.

	(1)	(2)	(3)	(4)
Panel A	I(Heavy drinker)			
Price elasticity				
log (price of vodka)	-0.092**	-0.089**	-0.060**	-0.053*
	[0.036]	[0.040]	[0.026]	[0.028]

	(5)	(6)	(7)	(8)
Panel B	I(Heavy drinker)			
Cohort-specific Price Elasticity				
log (price of vodka)				
Cohort: age 18-29	-0.117***	-0.087**	-0.161***	-0.115***
	[0.045]	[0.041]	[0.047]	[0.040]
Cohort: age 30-39	-0.074	-0.056	-0.078	-0.067
	[0.049]	[0.045]	[0.050]	[0.043]
Cohort: age 40-49	0.012	0.008	0.031	0.012
	[0.042]	[0.041]	[0.045]	[0.038]
Cohort: age 50-65	-0.018	-0.017	-0.014	-0.024
	[0.041]	[0.040]	[0.041]	[0.036]

Notes: Robust standard errors in brackets; ***p<0.01, **p<0.05, * p<0.1

Table A4. Forward-looking model with hazard of death discounting

	Per-period utility		Value function	Value function	Per-period utility
Peer effect, $\widehat{\delta}$:					
age 18-29	1.155***	log(vodka price)	-1.088***	-0.832***	-0.683**
	[0.051]		[0.422]	[0.318]	[0.280]
age 30-39	0.762***	Commit that price			
	[0.036]	change is permanent	Yes	No	
age 40-49	0.368***				
	[0.038]				
age 50-65	0.279***				
	[0.051]				
Habits	1.396***				
	[0.006]				

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

Table A5. Lag (Log vodka price) is not a good predictor for current Log(Vodka Price)

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

Notes: Robust standard errors clustered at municipality×year level are in brackets
* significant at 10%; ** significant at 5%; *** significant at 1%

Table A6a. Annual discounts that brings best fit. Rust's (NFP) approach.

Model with Price		Model with Hazard of death		Model with self-evaluated health	
Best fit		Best fit		Best fit	
β	0	β	0.47	β	0.99
Price	-0,108	Hazard of death	-4,01	Bad Health	-0,177
Habits	1,615	Habits	1,614	Habits	0,614
Constant	-1,548	Constant	-5,06	Constant	-2,59
Log Likelihood	-31315,3	Log Likelihood	-26297,3	Log Likelihood	-31315,6

Utility parameters are estimated from the model with following utility specifications
$U(not\,drink)_{it} = 0$; $U(drink)_{it} = \alpha + \beta Habit_{it} + \gamma Var_{it} + u_{it}$, where
Var_{it} is Price in model (1), hazard of death in model (2), and self -evaluated health in model (3)

Table A6b.Heavy drinking and current, lead and lagged prices and health.

	(1)	(2)	(3)
		Dependent Variable, Y: $I(heavy\,drinker)$.	
Independent Variable, X:	Panel A: Log Real price	Panel B: Log Nominal prices	Panel C: I(bad health)
X	-0.097**	-0.101**	-0.030***
	[0.049]	[0.039]	[0.010]
Lead X	0.044	0.033	-0.023**
	[0.047]	[0.028]	[0.009]
Lagged X	-0.005	-0.010	-0.016*
	[0.046]	[0.019]	[0.010]
Individual FE			YES
Observations	42,677	42,677	48,514
F-test, 1st stage	217.1	426.6	

Note: In all regression dependent variable is a $I(heavy\,drinker)$. Robust standard errors in brackets.
***p<0.01, **p<0.05, * p<0.1. Columns (1) and (2) are based on IV regression
$I(heavy\ drinker)_{it} = \alpha + \gamma_0 Log(Price)_{it} + \gamma_1 Log(Price)_{it+1} + \gamma_3 Log(Price)_{it-1} + t + t^2 + e_{it}$
where prices are instrumented by current, lead and lag excise tax of vodka. Column (3) is based on regression
$I(heavy\ drinker)_{it} = \alpha + \gamma_0 Health_{it} + \gamma_1 Health_{it+1} + \gamma_2 Health_{it-1} + \rho_i + e_{it}$, where ρ_i stands for individual fixed effects.

Table A6c. Hazard of death as a function of bad health and heavy drinking.

	Panel 0	Panel 1 Health var: I(has surgery last year)			Panel 2 Health var: I(bad health (self-evaluation))		
I(heavy drinker)	0.826***	0.796***	0.801***		0.987***	1.073***	
	[0.150]	[0.150]	[0.159]		[0.151]	[0.175]	
(Bad) health		1.712***	1.740***	1.822***	1.801***	1.898***	1.694***
		[0.362]	[0.474]	[0.366]	[0.145]	[0.177]	[0.143]
(Bad) health x I(heavy drinker)			-0.074			-0.419	
			[0.799]			[0.450]	
Observations	12,167	12,167	12,167	12,169	12,165	12,164	12,167

	Panel 3 Demographics var: I(married)			Panel 4 Demographics var: I(Muslim)		
I(heavy drinker)	0.810***	1.258***		0.828***	0.900***	
	[0.151]	[0.288]		[0.151]	[0.153]	
Demographics	-0.194	0.012	-0.235*	0.029	0.492**	-0.044
	[0.124]	[0.172]	[0.123]	[0.169]	[0.228]	[0.168]
Demographics x I(heavy drinker)		-0.609*			-2.555**	
		[0.336]			[1.046]	
Observations	12,164	12,164	12,166	12,167	12,167	12,169

Standard errors in brackets. Regressions are based on sample of males of age 18-65. The hazard regression specification is as follows: $\lambda(t,X) = exp(X\beta)\lambda_0(t)$. I use a semi-parametric Cox specification of baseline hazard, $\lambda_0(t)$.

Table A6d. 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, a consumer 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}.

Note 1. Calculation of marginal (with respect to price) value function.

Remind that, 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:

$log(p_{i,t+1}) = \phi_0 + \phi_1 log(p_{it}) + \omega_{it}$, where $E(\omega_{it}|p_{it}) = 0$, i.e. $\frac{\partial E_p(log(p_{t+1}))}{\partial log(p_t)} = \phi_1$

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

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

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

Under these assumptions,

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

Without a commitment on price stability, $\frac{\partial E_p(log(p_{t+1}))}{\partial log(p_t)} = \phi_1$. 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 therefore

$$\begin{aligned}\frac{\partial Value\,function}{\partial log(p_t)} = \quad & \frac{\partial}{\partial log(p_t)}[E_{e_{-i}}\pi_{it}(a_{-it}, a_{it} = 1, s_t)] \\ & + \frac{\partial}{\partial log(p_t)}[E(V_i(S_{t+1})|1, S_t) - E(V_i(S_{t+1})|0, S_t)] \\ & = \frac{\partial \rho_{mt}(\pi)}{\partial log(p_t)} + \frac{1}{\phi_1}(\frac{\partial \rho_{mt}(EV1)}{\partial log(p_t)} - \frac{\partial \rho_{mt}(EV0)}{\partial log(p_t)})\end{aligned}$$

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 are 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:

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

where $\gamma = 0.577$ is Euler constant.

Applying these relationships to equation for value function:

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

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

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

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

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

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

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:

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

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

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

Proof A2

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 consumer value function is fully characterized by two equations:

$$V(0_{it}, s_t) = \rho_0 z_{it} + \beta E_{t+1}(exp(V(0, s) - log(\sigma(0))|0_{it}, s_t) \tag{27}$$

$$V(0_{it}, s_t) + log(\sigma(1)/(\sigma(0)) = \rho_1 z_{it} + \pi_{it}(a_{-it}, a_{it}, s_t, \theta) + \beta E_{t+1}(V(0, s) - log(\sigma(0)))|1, s_t) \tag{28}$$

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 (27) and (28).

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 (27) and (28) above hold for new $V(0_{it}, s_t)', \rho_j'$

8 Online Appendix

8.1 Effect of tax policy on consumer welfare.

In this part I model the effect of tax policy on consumer welfare.

In both the forward-looking and myopic models presented above, consumers have bounded rationality: they do not take into account the effect of heavy drinking on hazard of death.[45] Within these models, the tax corrects a negative externality that appears from the bounded rationality of consumers. The welfare effect of the 50 percent tax is as follows: The tax results in a 30 percent loss in consumer surplus.[46] At the same time, the tax saves 30,000-50,000 young male lives annually, which is 0.04-0.06 percent 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 a time discount $\beta = 0.9$, the value of saved lives is equal to 0.4-0.6 percent of GDP, which equals the size of the whole alcohol industry in Russia (0.48 percent of GDP). This speculative calculation suggests that a 50 percent tax is actually likely to be smaller than the optimal one.[47]

Besides, under certain assumptions about utilities, my model implies that the effect of a vodka tax on consumer surplus would be positive even for fully-rational consumers, forward-looking consumers who take into account the mortality risk associated with heavy drinking. The model implies that peer effects and the effect of habits are positive: all other things being constant, a consumer has higher utility if he or she drank within the previous period and if he or she has peers who are heavy drinkers. These forces, however, can equally run a consumer utility into the negative. First, quitting heavy drinking is costly. Second, a consumer who decides not to drink may suffer from the fact that peers are drinking – the consumer may experience peer pressure, or the consumer may suffer if no peer wishes to participate in alternative (to drinking) activities, such as playing soccer or doing other sports.[48] Thus, in the section 8 in online appendix, I find that peer decisions matter for a consumer if he or she decides to do physical training. These alternative assumptions about utilities, although barely distinguishable from the data, have different implications for the analysis of consumer welfare.[49] In this case, a 50 percent tax on vodka results in an

[45] I analyze the model where consumers do take into account the effect of drinking on hazard of death in the appendix (see table A4 in appendix). Results are similar to those of the forward-looking model in the main body of text (with slightly higher magnitude).

[46] Consumer welfare is expected (over realization of private utility shocks) present value of the flow of utilities. Under my model assumptions, $\triangle E(CS) = \frac{1}{\alpha_i}[ln(\sum(exp(V_{ij}))|tax - ln(\sum(exp(V_{ij}))|notax]$, where V_{ij} is choice-specific Value function (for and consumer i and choice j), α_n is marginal utility of income (negative coefficient with price).

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

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

[49] In the "myopic" case, peer effect and peer pressure are not identified jointly. One can identify only the difference between them. In the "forward-looking" case, they are identified under additional assumptions. See proof of identification results in the appendix (Proof A2). In the 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 A6d).

increase in the consumer welfare of young males below age 40.[50] Figure OA1 illustrates this point.

8.2 Hazard of death regressions: robustness checks

Table OA4 in online appendix reports estimates of hazard of death by different causes of death. Unfortunately, the causes of death are reported in less than 60 percent of death cases. RLMS recorded 6 causes of death,namely heart attack, stroke, accidents&poisoning, cancer, tuberculosis, and "other" causes. Heavy drinking results in higher rates of death due to accidents and poisoning, due to other reasons, and in higher rate of death for which the cause was not reported.

Table OA5 in online appendix reports of hazard of death with different measures of heavy drinking. Heavy drinking definitions: I check regression results where heavy drinker defined as those who belong to a) Top 25% by alcohol intake; b) to top 25% by alcohol intake within 10-years age cohorts;c) to top 50% by alcohol intake; and d) to top 25% by days of alcohol consumption (per month). In all specifications, heavy drinking results in higher mortality rates, and the effect is highest for youngest age cohort.

8.3 Discussion of identification assumptions in peer effect estimation

This model relies on two assumptions: exclusion restriction, and uniqueness of equilibria.

8.3.1 Exclusion restriction

Exclusion restriction requires that (i) the subset of demographic characteristics G_{-it} does not contain all of the set of demographic variables D_{it}, and that (ii) excluded demographic characteristics are independent of private utility shocks.

Although my estimates show that consumers do not have any preferences regarding G_{-it}, all coefficients at G_{-it} are insignificant, still the assumption (i) on which exclusion restriction is based is strong. One can argue that any demographic characteristics of peers may affect the utility of the consumer, and so should be included in G_{-it}; the utility of drinking may be greater for a consumer when she or he drinks with peers of same marital status, peers with better health, etc. In addition, some of the excluded demographic variables may respond to alcohol consumption. Further, excluded demographic variables (and alcohol consumption) may be affected by unobservable shocks or the selection on unobservable characteristics (see Manski's "reflection problem" 1993, Moffit 2001). In result, excluded instruments may be econometrically endogenous. To verify the reliability of my model, I provide different robustness checks for the obtained results.

Recent literature emphasizes the importance of peers in making personal decisions, in particular whether to drink or not (see, for example, Akerlof and Kranton 2000, Card and

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

Giuliano 2011, Cooley 2016, Gaviria and Raphael 2001, Krauth 2005, Kremer and Levy 2008, Moretti and Mas 2009). The estimation of peer effects is a difficult task because it can be contaminated by common unobservable factors, non-random reference group selection, and the endogeneity of other group members' choices (Manski 1993, Moffit 2001). Recent literature responds to this problem by using random assignments of peers in peer groups, or by using quasi-experiments (see for example Kremer and Levy 2008, Katz et al. 2001, Oreopoulos 2003). However, as Card and Giuliano (2011) and Carrell, Sacerdote, and West (2011) argue, peer relationships that occur within randomly assigned groups may significantly differ from those occurring in natural environments where people grow up together and friendships naturally occur.

Further, studies that employ random assignment usually cover only relatively narrow groups within a population and a relatively short-run time horizon. In my paper, however, the task of quantifying the effect of government alcohol policy on the alcohol consumption, mortality, and welfare of all Russian males makes it particularly important to infer the consequences of alcohol consumption and peer interactions for a broad heterogeneous group of people and over a relatively long-run time horizon.

First, I employ linear-in-means specification with the same set of instruments to test endogeneity and relevance of instruments. The main regression specification is as follows:

$$I(heavy\,drinker)_{it} = \sum_k \delta_k I(age\,strata = k)\overline{I(heavy\,drinker)}+ \\ \gamma I(heavy\,drinker)_{it-1} + \Gamma' D_{it} + \Upsilon' G_{-it} + \rho_{mt} + e_{it} \quad (29)$$

where $\overline{I(heavy\,drinker)}$ is instrumented by average (across peers) demographic characteristics.[51]

Table OA1 in 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.[52]

Column IV-1 of Table OA1 shows the results of regressions where the set of explanatory variables (D_{it}, G_{-it}, ρ_{mt}, $I_{it-1}(heavy\,drinker)$) is the same as in the main model discussed in the text.

The P-value of the J-test for the exogeneity of instruments is 0.22, so based on this re-

[51]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 the 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 an additional game with fathers to the model if I wanted use paternal demographics as instrumental variables.

[52]To compare coefficients in the logit model (Table 6) with those in the linear probability model (Table OA1), one needs to multiply the coefficients in Table OA1 on 5.3. To compare marginal effects of LPM and logit regression, one needs to divide the coefficients in LPM on $p(1-p)$, where p is the probability of being a heavy drinker. In our case $(p(1-p))^{-1} = 5.3$.

gression one cannot reject a hypothesis of exogeneity for instruments. The F-statistic of the test for relevance of excluded instruments is 72 (with errors clustered on the municipality×year level), which shows that instruments are relevant. The J-statistics confirms a model assumption that demographic characteristics respond to alcohol consumption in a non-elastic way, and that after accounting for municipality×year fixed effects (that capture all shocks and all possible selection on municipality×year level) and individual demographics, average peer demographics are exogenous instruments for peer alcohol consumption. In addition, Column IV-2 of table OA1 shows similar results in regression when I use a subset of instruments. Column IV-3 of Table OA1 shows that the results are also similar when I change municipality×year fixed effects on individual fixed effects, which captures possible effects of omitted factors that are constant at the individual level.

8.3.2 Peer effects for Alternative Measures of Alcohol consumption and for Other Goods

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, and thus 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 OA2 in the appendix).

Besides I provide the additional test to check that observed correlation between own consumption and peer consumption is driven by peer effects, but not by common unobservable shocks. I check the model by applying a similar strategy to tea, coffee, and cigarette consumption, and to hours of physical training (see Table OA2 in the appendix). In case if the correlation is driven by common shocks (say local prices) I should find evidence of peer effects for every good. Indeed, 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 sports, that are social activity (we play soccer or basketball in groups). The effect of peers on smoking is marginally significant for the two age strata.

8.3.3 Alternative Instruments: Military service and Father Characteristics

According to my paper, the strongest peer effects are observed for the younger generation of males (males of age 18-29). The assumption of exogeneity of subset demographic characteristics seems to be more reliable for this group because alcohol consumption does not have an immediate effect on demographic characteristics but rather manifests over the long-run.

Further, for this sub-population of males, I have the opportunity to provide additional tests of my results based on other peer characteristics used as instruments in the regression above.

First, I use the share of peers who returned from military service. Members of this group have a higher probability of being a heavy drinker. To control for selection bias, I include a share of peers who ever served, or who will ever serve in the Russian army (according to my data) as a control for the regression. In addition, I include individual-level variables – I(served in army) and I(ever served or will ever serve in army) – as control variables in the regression. The regression specification is as follows:

$$I(heavy\,drinker)_{it} = \delta\overline{I(heavy\,drinker)_{-it}} + \gamma I(heavy\,drinker)_{it-1} + \Gamma' D_{it} + \Upsilon' G_{-it} + \rho_{mt} + e_{it} \quad (30)$$

where D_{it} contains, in addition to a standard set of demographic characteristics, $I(already\ served\ in\ army)$, $I(served\ or\ will\ ever\ serve\ in\ army)$, and G_{-it} contains $\overline{I(served\ or\ will\ ever\ serve\ in\ army)}$. Average peers alcohol consumption $\overline{I(heavy\ drinker)}$ is instrumented by share of peers who came from army, $\overline{I(served\ in\ army)}$.

Columns 8 and 9 in Table OA1 show the second and first stages of this regression. Column 9 indicates that those who were previously in the army drink more: the F-statistic for the first stage is 68. The second stage shows results similar to the IV regression discussed above.

Additionally, I verify the robustness of my results by estimating the IV regression on a sub-sample of respondents who had just returned from military service. These people are likely not to face shocks common to their peers. All estimates for this sub sample have the same magnitude and statistically significant.

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 common consumer shocks. Moreover, both the model estimates and the 2SLS estimates discussed above show no correlation between one's own and peer alcohol consumption for the old-age strata to which the fathers belong. This should not happen if common unobservable factors affect the behavior fathers or shape their demographics. Regression IV-8 shows the results of an IV regression that uses the demographics of all fathers as instruments. It shows the same magnitude, but the result is statistically-insignificant. Regression IV-9 shows the results of an IV regression where instead a subset of father characteristics with better predictive power is used as instruments, with results that are statistically-significant and similar in magnitude to peer effects.

Finally, in regression IV-10 I include both army IV and father IV in one regression. Again, the regressions show a similar magnitude and statistically-significant peer effects. The P-value of the J-test equals 0.57, indicating that the instruments are exogenous.

8.3.4 Important limitations

The important limitation on my study is that my data does not provide me exact information on members of peer drinking groups. This may cause a bias in estimation of peer effects. First, within same-age and neighborhood groups of people, some may indeed be peers and drink together while the others may not. Second, I do not have data on all peers in peer groups. As a reminder, RLMS contains data on roughly the half of population in peer groups. In this case regression estimates of peer effects in our model suffer from attenuation bias even if I apply IV regression. Thus, I am likely under-estimate the magnitude of peer effects. Indeed, because I use data on half of the peers, the true effect is likely to be twice as high compared to my estimates. The short elaboration of this result is presented in Appendix, Note 2.

It is also possible that hard drinkers are less likely to respond to the RLMS survey. If hard drinkers react differently on changes in price of alcohol and/or if restriction alcohol consumption result in higher drop in mortality rates among hard drinkers, and/or magnitude of peer affects are different for them, then our estimates would be biased.

8.3.5 Uniqueness of Equilibria

The second identification assumption is that although multiple equilibria are possible, only one equilibrium is played out in the data. In the context of my model, this identification assumption states that equation [53]

$$\sigma_{it} = \frac{exp(\delta \sum_{-i} \sigma_{jt}/(N-1) + \gamma habit_{it} + \Gamma' D_{it} + \Upsilon' G_{-it} + \rho_{mt})}{\Sigma_j exp(\delta \sum_{-j} \sigma_{kt}/(N-1) + \gamma habit_{jt} + \Gamma' D_{jt} + \Upsilon' G_{-jt} + \rho_{mt})} = G(\sigma_{-it}, S_{it}, \Theta) \qquad (31)$$

has a unique equilibrium in the data. This assumption is commonly made in empirical studies of games with incomplete information. As was shown in Bajari, Hong, and Ryan (2009), in case of monotone payoff, the expected number of equilibria decreases as the number of states $|S_{it}|$ goes to infinity. In my model setup $S_{it} = U_{j\in\{i,-i\}}\{habit_{jt}, D_{jt}, G_{nt}, \rho_{mt}\}$, and thus $|S|$ is big enough to claim that the probability of observing multiple equilibria is small.

In addition, I ran the following experiment, which confirms the uniqueness of equilibria in my data (albeit does not prove it). Using an iteration procedure, I find the fixed points $\sigma*$ of equation $\sigma_{it} = G(\sigma_{-it}, S_{it}, \Theta))$, starting from different initial values of σ_{it}. First, I find the fixed point using an iteration procedure that starts from a zero level of alcohol consumption (the so-called "low-level equilibrium"). Then, I find the fixed point starting from the highest level of consumption (the high-level equilibrium). Further, I perform 1000 simulations for which starting points are chosen randomly from interval (0,1) for every agent in my data. I find that the estimated σ and utilities are essentially the same (with only a small difference due to computational errors) – estimated utility parameters differ only at the third decimal point. Finally, in the robustness section, I re-estimate my model using linear probability model assumption, for which multiplicity of equilibria is not an issue. Results are robust to choice of specification.

[53] Here S is a set of state variables; Θ is a set of parameters. This equation comes from the expression of probability of choosing heavy drinking option that comes from equation (6).

8.4 Habits versus Unobserved Heterogeneity

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

$$I(heavy \quad drinker)_{it} \;=\; \alpha \;+\; \gamma I(heavy \quad drinker)_{it-1} \;+\; \Gamma' D_{it} \;+\; \rho_i \;+\; \delta_t \;+\; e_{it} \quad (32)$$

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 $I(heavy\, drinker)$. The results of regression are presented in Table OA3 in Appendix.

Table OA3 shows the results of regressions with lagged $I(heavy\, drinker)$ as well as the results of regressions with an average across two and three lags of $I(heavy\, drinker)$. The regression results suggest that habits are important, with the same magnitude as elsewhere in my paper. Point estimates of the coefficient on lagged $I(heavy\, drinker)$ vary from 0.28 to 0.54. In two out of three specifications, the coefficient is statistically significant at the 5% significance level.

References for Online Appendix

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

Cattaneo, Matias, Sebastian Calonico and Rocio Titiunik, 2014, Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs, Econometrica 82(6): 2295–2326, November 2014

Card, David, and Laura Giuliano, 2013 “Peer Effects and Multiple Equilibria in the Risky Behavior of Friends“, Review of Economics and Statistics, 2013.

Carrell, Scott E., Bruce I. Sacerdote and James E. West. 2011. “From Natural Variation to Optimal Policy? The Lucas Critique Meets Peer Effects.” NBER Working Paper No. 16865, National Bureau of Economic Research, Inc.

Cooley, Jane. 2016, Can Achievement Peer Effect Estimates Inform Policy? A View from Inside the Black Box,” forthcoming, Review of Economics and Statistics.

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

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

Krauth, Brian. 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.

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

Moffitt, R. (2001): “Policy Interventions, Low-Level Equilibria, and Social Interactions”, in Social Dynamics, edited by Steven Durlauf and Peyton Young, MIT press

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

Oreopoulos, Philip. 2003. “The Long Run Consequences of Growing Up in a Poor Neighbourhood.” Quarterly Journal of Economics 118(4): 1533-1575.

Table OA1. Linear in means peer effects. Robustness checks.

	Dependent variable: I(heavy drinker)						
			Sample: males of age 18-65				
	IV-1	IV-2	IV-3	IV-4	OLS-1	OLS-2	
Peer effect, $\widehat{\delta}$:							
age 18-29	0.264	0.297	0.255	0.242	0.193	0.119	
	[0.04]***	[0.05]***	[0.09]***	[0.04]***	[0.03]***	[0.02]***	
age 30-39	0.194	0.218	0.16	0.181	0.17	0.111	
	[0.03]***	[0.04]***	[0.065]**	[0.03]***	[0.02]***	[0.01]***	
age 40-49	0.063	0.089	0.063	0.053	0.121	0.057	
	[0.030]**	[0.037]**	[0.059]	[0.031]*	[0.02]***	[0.01]***	
age 50-65	-0.005	0.015	0.009	-0.022	0.088	0.03	
	[0.033]	[0.041]	[0.056]	[0.033]	[0.02]***	[0.016]*	
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	29554	27400	29923	29923	
F-test	79.9	36.29	17.02	72.02			
J-test, p-value	0.22	0.13	0.02	0.26			
			Sample: males of age 18-29				
	IV-5	IV-6	IV-6, 1st st.	IV-7	IV-8	IV-9	IV-10
Peer effect, $\widehat{\delta}$:	0.211	0.25		0.359	0.197	0.225	0.298
	[0.09]**	[0.079]***		[0.180]**	[0.136]	[0.14]*	[0.125]**
Munic×year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
$\overline{I(served\ in\ army)}$			0.211				
			[0.040]***				
$\overline{I(served\ or\ will\ ever\ serve\ in\ army)}$			-0.128				
			[0.039]***				
Just came from military service?				Yes			
Instruments	Peers 1	Army		Peers 1	Fathers 1	Fathers 2	Fathers 2 Army
Observations	7750	5629[+]	5629[+]	149	8152	8152	5629[+]
F-test	34.24	61.05	61.05	6.85	16.52	28.97	12.9
J-test, p-value	0.06			0.17	0.4	0.86	0.58

Notes: Standard errors clustered at municipality×year in brackets. * significant at 10%; ** significant at 5%, *** significant at 1%. Instrument set: Peers: (1) average demographics (2) average demographics without lag I(heavy drinker). Instrument set: Peer fathers: (1) average demographics (2) average demographics-subset. Demographics controls are included in every regression.[+] The # of obs. in IV6 and IV10 is smaller because data on military service available only for subset of rounds

Table OA2. Linear in means peer effects. Peer effects for different products/activities.

	Peer effect			
year	age 18-29	age 30-39	age 40-49	age 50-65
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

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

Table OA3. Habits versus unobserved heterogeneity.

	Y		
	I(heavy drinker)		
Mean(Lag Y, LagLag Y, LagLagLag Y)	0.535**		
	[0.215]		
Mean(Lag Y, LagLag Y)		0.529**	
		[0.221]	
Lag Y			0.283
			[0.236]
I(health problems)	-0.016	-0.015	-0.018*
	[0.010]	[0.011]	[0.011]
Individual FE	Yes	Yes	Yes
Demographics	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Observations	54,429	54,422	54,111
Number of individuals	9,434	9,433	9,402
F-test for instruments (with robust se)	21.48	21.51	17.40

Notes: 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).

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

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

Table OA5. Hazard of Death regressions with different definitions of heavy drinker

	(1)	(2)	(3)	(4)
I(heavy drinker), age 18-29	2.304***	2.247***	1.678***	3.576***
	[0.467]	[0.456]	[0.650]	[0.440]
I(heavy drinker), age 30-39	1.704***	1.912***	1.359***	1.905***
	[0.353]	[0.357]	[0.473]	[0.334]
I(heavy drinker), age 40-49	0.588*	0.470	0.819**	1.133***
	[0.315]	[0.336]	[0.403]	[0.248]
I(heavy drinker), age 50-65	-0.275	-0.317	0.001	-0.420**
	[0.247]	[0.239]	[0.245]	[0.187]
I(Bad Health)	1.397***	1.391***	1.464***	1.418***
	[0.164]	[0.164]	[0.286]	[0.165]
Log (family income)	-0.415***	-0.415***	-0.105	-0.413***
	[0.036]	[0.036]	[0.066]	[0.036]
I(smokes)	0.586***	0.594***	0.917***	0.554***
	[0.124]	[0.124]	[0.210]	[0.126]
I(college degree)	-0.073	-0.074	-0.248	-0.075
	[0.132]	[0.132]	[0.212]	[0.133]
Weight	-0.002	-0.002	0.002	-0.002
	[0.004]	[0.004]	[0.006]	[0.004]
I(work)	0.083	0.087	-0.086	0.003
	[0.146]	[0.146]	[0.242]	[0.147]
Observations	12,109	12,109	9,227	12,109
Heavy drinking definition	0	1	2	3

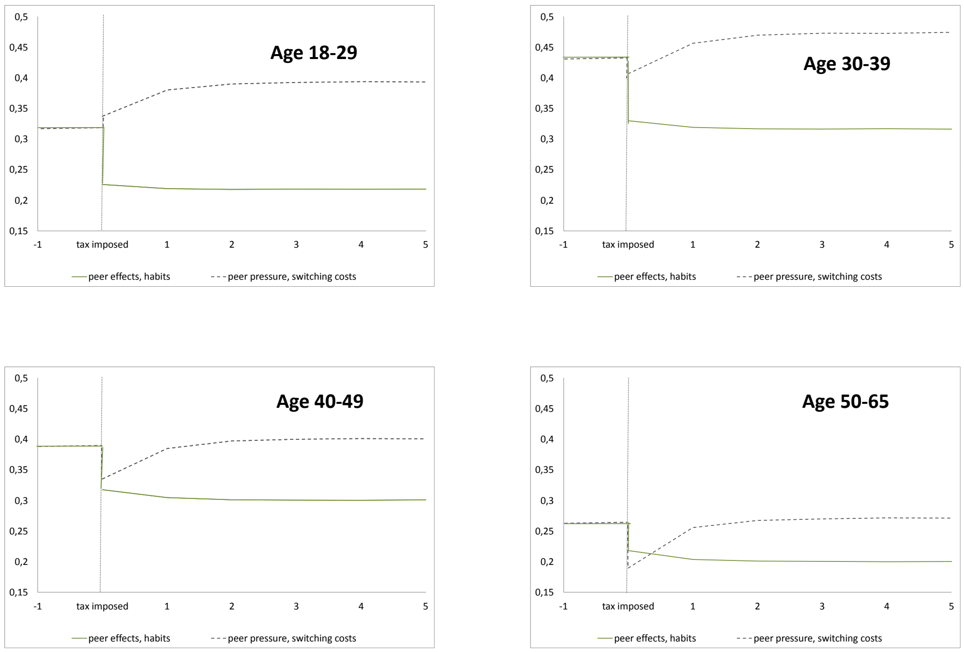
Notes: Standard errors are in brackets. * significant at 10%; ** significant at 5%, *** significant at 1%. Heavy drinking definitions: Model (0): Top 25% by alcohol intake; Model (1): Top 25% by alcohol intake within 10 years age cohorts; Model (2): Top 50% by alcohol intake; Model (3): Top 25% by days of alcohol consumption (per month). In model 3 data is available only for rounds 15-23 of RLMS survey.

Table OA4. Hazard of Death by different causes of death

	(1) Heart attack	(2) Stroke	(3) Cancer	(4) Poisoning, accidents, injuries	(5) Tuberculosis	(6) Other	(7) Not reported
I(heavy drinker), age 18-29	0.366	1.725	1.806	2.263***	-1.499	4.111***	1.405
	[2.483]	[2.900]	[2.263]	[0.729]	[7.706]	[1.066]	[0.959]
I(heavy drinker), age 30-39	0.589	3.170**	1.877	2.502***	-3.837	1.805**	0.732
	[1.281]	[1.339]	[1.449]	[0.593]	[5.034]	[0.860]	[0.730]
I(heavy drinker), age 40-49	1.136	0.541	-0.051	0.165	5.219**	1.454**	-0.513
	[0.713]	[0.947]	[1.248]	[0.778]	[2.134]	[0.721]	[0.661]
I(heavy drinker), age 50-65	-0.092	-0.320	-1.239*	0.489	-1.874	-0.243	0.299
	[0.585]	[0.565]	[0.645]	[0.814]	[2.578]	[0.856]	[0.442]
Log (family income)	-0.113	-0.155	-0.116	-0.397***	-0.514	-0.431***	-0.771***
	[0.100]	[0.105]	[0.110]	[0.086]	[0.318]	[0.101]	[0.065]
Bad health	0.968**	0.951**	1.665***	-0.422	1.997*	1.661***	2.055***
	[0.460]	[0.418]	[0.407]	[0.649]	[1.211]	[0.473]	[0.292]
I(smokes)	0.773***	0.881***	1.018***	0.892**	-0.759	1.031**	-0.001
	[0.299]	[0.318]	[0.326]	[0.390]	[0.971]	[0.469]	[0.218]
I(college degree)	-0.160	0.226	0.029	-0.409	-587.986	-0.284	-0.098
	[0.310]	[0.309]	[0.325]	[0.390]	[0.000]	[0.453]	[0.262]
Body Weight	0.020***	0.015*	-0.002	-0.027***	-0.081**	-0.026**	-0.017**
	[0.008]	[0.008]	[0.009]	[0.010]	[0.041]	[0.012]	[0.007]
I(work)	0.266	-0.644*	0.260	-0.221	-1.521	-0.209	0.666**
	[0.384]	[0.377]	[0.397]	[0.360]	[1.375]	[0.416]	[0.275]
Observations	12,125	12,125	12,125	12,125	12,125	12,125	12,125

Notes: Standard errors in brackets; ***p<0.01, **p<0.05, * p<0.1

Figure OA1. Effect of tax policy on consumer welfare.



Notes: The figures show the simulated effect of 50 percent increase in the price of vodka on consumer welfare. Figures plot response under different assumptions on consumer utility function. Horizontal axis: years before and after imposing tax. Vertical axis: Consumer surplus.

Note 2. Peer effects estimation when data only on the part of peers is available

For simplicity, lets work with the example where every person has exactly two peers in her(his) peer group, and data contains information on only one peer.

The true model we want to estimate is

$Y_i = \phi_0 + \phi_1 \frac{X1+X2}{2}_i + e_i$,

but due to data availability we are restricted to estimate another model

$Y_i = \phi_0 + \phi_1 X1_i + e_i$.

The probability limit of OLS estimate of ϕ_1 in this case is

$plim\widehat{\phi}_{1OLS} = \frac{cov(\phi_1 \frac{X1+X2}{2}_i, x1_i)}{Var(x1_i)} = \phi_1(0.5 + 0.5cov(X2_i, X1_i))$

In case when$cov(X2_i, X1_i) < 1$, $\widehat{\phi_{1OLS}}$ has attenuation bias

Applying IV regression with instrument $Z1$ that correlated with $X1$ and not correlated with $X2$ does not help to eliminate attenuation bias in this model:

The probability limit of IV estimate of ϕ_1 in this case is

$plim\widehat{\phi}_{1IV} = \frac{cov(\phi_1 \frac{X1+X2}{2}_i, Z1_i)}{civ(Z1_i, X1_i)} = \phi_1(0.5 + 0.5\frac{cov(X2_i, Z1_i)}{cov(X1_i, Z1_i)})$

If $cov(X2_i, Z1_i) = 0$, $\widehat{\phi_{1IV}}$ has attenuation bias again.

Indeed $plim\widehat{\phi}_{1IV} = 0.5\phi_1$ in this case.