Peers and Alcohol: Evidence from Russia*

Evgeny Yakovlev‡

November 22, 2011

Abstract

For the last twenty years Russia has confronted the Mortality Crisis– the life expectancy of Russian males has fallen by more than five years, and the mortality rate has increased by 50%. Alcohol abuse is widely agreed to be the main cause of this change. In this paper, I use a rich dataset on individual alcohol consumption to analyze the determinants for heavy drinking in Russia, such as the price of alcohol, peer effects and habits. I exploit unique location identifiers in my data and patterns of geographical settlement in Russia to measure peers within narrowly-defined neighborhoods. The definition of peers is validated by documenting a strong increase of alcohol consumption around the birthday of peers. With natural experiments I estimate the own price elasticity of the probability of heavy drinking. This price elasticity is identified using variation in alcohol regulations across Russian regions and over time. From these data, I develop a dynamic structural model of heavy drinking to quantify how changes in the price of alcohol would affect the proportion of heavy drinkers among Russian males and subsequently also affect mortality rates. I find that that higher alcohol prices reduce the probability of being a heavy drinker by a non-trivial amount. An increase in the price of vodka by 50% would save the lives of 40,000 males annually and would result in an increase of welfare. Peers account for a quarter of this effect.

*Preliminary: Comments welcome. Do not circulate and cite without permission of the author.

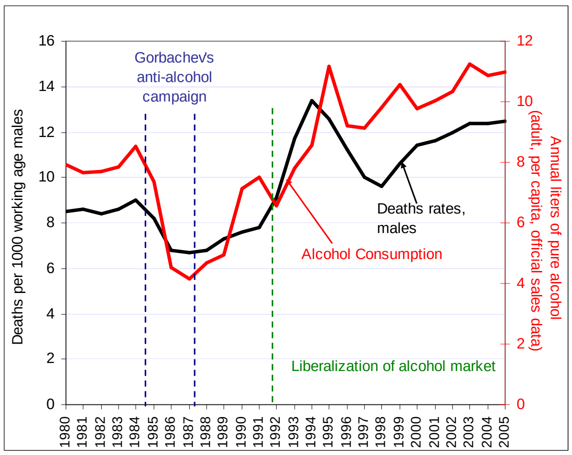
†I am especially grateful to Denis Nekipelov, Gerard Roland, David Card, and Katya Zhuravskaya for discussion of paper and kin support during my PhD study in UC Berkeley. I also thank to Patrick Kline, Enrico Moretti, Lorenz Kueng, Mu-Jeung Yang, Olexa Shwets, Pedro Castro, Max Kazi, and participants of the Labor Lunch and Development Lunch in UC Berkeley for helpful discussions and comments. The financial support of the Economic Department, Graduate Division and International Office of UC Berkeley is gratefully acknowledged.

‡University of California, Berkeley, Berkeley, CA, USA, e-mail: yakovlev@econ.berkeley.edu

Introduction

Russian males are notorious for their hard drinking. The Russian (non-abstainer) male consumes an average of 35.4 liters of pure alcohol per year.[1] This amount is equivalent to the daily consumption of 6 bottles of beer or 0.25 liters of vodka. The most notable example of the severe consequences of alcohol consumption is the male mortality crisis – male life expectancy in Russia is only 60 years. This is 8 years below the average in the (remaining) BRIC countries, 5 years below the world average, and below that in Bangladesh, Yemen, and North Korea. High alcohol consumption is frequently cited as the main cause (see for example Treisman 2010, Leon et al. 2007, Nemtsov 2002, Bhattacharya et al. 2011, Brainerd and Cutler 2005).[2] Approximately one-third of all deaths in Russia are related to alcohol consumption (see Nemtsov 2002). Most of the burden lies on males of working age: more than half of all deaths in working-age men are accounted for by hazardous drinking (see Leon et al. 2007, Zaridze et al. 2009, and Figure 1 below).

Figure 1. Alcohol Consumption and Male Mortality Rate.



Source: WHO, Treisman (2010), Rosstat.

Surprisingly, no attempts have been made to quantify the effects of public policy on mortality rates, and there have been few efforts to identify the effects of public policy on alcohol consumption. Moreover, research that identifies the causal effect of price on alcohol consumption and mortality deals with only

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

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

aggregate (regional-level) data.[3] However, the use of disaggregated data is of particular interest because it allows disentangling the different forces that bear on individual decisions about drinking. Also, it allows an evaluation of the effect of policy on different subgroups.

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

Peer effects are agreed to be very important for policy analysis because they produce a (social) multiplier effect. Recent literature emphasizes the importance of peers in making personal decisions, in particular whether to drink or not (see, for example, Gaviria and Raphael 2001, Krauth 2005, Kremer and Levy 2008, Card and Giuliano 2011, Moretti and Mas 2009). There are sound reasons to believe that peer influence is even stronger in Russia because of patterns of the dense geographical settlement inherited from the Soviet Union and the very low level of mobility in Russia. In my paper, I exploit unique location identifiers in the data to measure peers within narrowly-defined neighborhoods. This definition of peers is validated by documenting a strong increase in alcohol consumption around the birthday of peers.

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

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

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

outcomes (see Bajari et al 2008, Aguirregabiria and Mira 2007, Berry, Pakes, and Ostrovsky 2007, and Pesendorfer and Schmidt-Dengler 2008).

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

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

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

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

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

This paper is organized as follows. In the following section I review existing empirical literature on peer effects and rational addiction, and on the estimation of dynamic models. In Section 3, I describe my data and the variables used in my analysis. Section 4 presents the model and estimation strategy. In Section 5, I discuss results. Section 6 discusses robustness checks. Section 7 concludes.

Literature Review

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

Different identification approaches have been proposed to solve the problems introduced in Manski's critique. For reviews of these studies, see Blume and Durlauf (2005). The primary approaches in the empirical labor literature are the random assignment of peers (see Kremer and Levy 2008, Katz et al. 2001) and finding the exogenous variation of peer characteristics (see Gaviria and Raphael 2001). Glaeser, Sacerdote, and Scheinkman (2002) and Graham (2008) use structural models to infer the magnitude of peer effects from aggregate statistics. Krauth (2005) employs a structural approach to directly model endogenous choice and correlated effects.

Empirical industrial organization literature contributes to this by providing an intuitive structural framework for the analysis of peer interaction (see for example Bajari et al. 2008, Aguirregabiria and Mira 2007, Berry, Pakes, and Ostrovsky 2007, and Pesendorfer and Schmidt-Dengler 2008). In this research, the structural framework takes the form of games, with incomplete information. Agents do not observe other people's actions or form beliefs from what people do based on observable state variables. The expected utility of agent therefore does not include the actions of peers, but only the beliefs of the agent. Estimations in this model are very similar to those in the two-stage approach, where in the first stage the researcher estimates the agent's beliefs, and in the second stage the researcher estimates utility parameters, including peer effect.

In contrast to other proposed approaches, this approach is structural. Introducing structure to the model allows the researcher to model the effect of policy on different economic factors, such as consumer welfare and the death rate. This approach also allows for analyzing strategic interactions in both static and dynamic contexts.

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

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

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

Data Description

Typical patterns of geographical settlement in Russia – a remainder of the Soviet Union's legacy – allow me to use geographic closeness as a measure of the

[4]Most of the studies that test the validity of the forward-looking hypothesis provide only an indirect test, looking at the correlation between the current consumption of an addictive product and its future price. These methods are subject to a meaningful drawback, potentially identifying a spurious correlation and so wrongly supporting the rational addiction analysis (Auld and Grootendorst 2004).

likelihood of status as a peer. Approximately 10% of Russian families live in dormitories and communal houses, where residents share kitchens and bathrooms.[5] A majority of the remaining, more fortunate, part of the population lives in a complex of several multi-story multi-apartment buildings, called a "dvor." These complexes have their own playgrounds, athletic fields, and ice rinks, and often serve as the place where people spend leisure time.[6] Photos of typical dvors are presented in Figure A2 in the appendix. Dvors are the most popular place in Russia to find friends – the very low level of personal mobility in Russia means that most people live in the same place (and therefore the same dvor) for most of their lives.

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

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

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

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

[6]The size of dvor vary in range from 300 to 3000 inhabitants.

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

[8]"Neighborhood" is defined as a school district. This is not a precise definition. The RLMS has data on belonging to a census district, which in most cases is equivalent to a school district.

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

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

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

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

"Peers" Definition

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

To verify the reliability of my measures, I provide the following test: I correlate log (the amount of vodka consumption) with a dummy variable if a person has a birthday in the previous month, and with averages of the birthday dummy variables across peers.[12]Vodka is the most popular alcoholic beverage to serve on birthdays, compared to beer and for males also to wine. Results for both regressions are positive and statistically significant. Regression suggests that a person's consumption of vodka increases by 16% if his birthday is during the previous month, and by 6% if there was a birthday of one of his peers (in a group of 5 peers). The results are robust if I eliminate household members from the sample of peers.[13]

---

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

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

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

Table 1. Birthdays and Alcohol Consumption.

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

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

Alcohol consumption variable

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

However, alcohol consumption reporting in the RLMS suffers from the common problem of all individual-level consumption surveys: it is significantly under-reported.[15] So, to offer an indication of the actual level of alcohol consumption corresponding to the threshold of being a "heavy drinker," I correlate the reports of consumption from the RLMS data with official sales data as a benchmark for average levels of alcohol consumption.

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

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

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

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

Model

The set-up of the model is as follows.

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

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

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

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

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

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

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

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

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

Thus, $\pi_{it}(a_{-it}, a_{it} = 1, s_t)$ depends on average peer alcohol consumption, habits ($a_{i,t-1}$)[16], a set of personal demographic characteristics (D_{it}), (sub) set of

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

peers characteristics G_{-it} and municipality*year invariant factors ρ_{mt}.

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

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

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

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

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

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

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

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

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

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

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

[19]I discuss both of the models because there is no consensus in the literature regarding which assumption is more relevant for the analysis of drinking behavior. In general set-up, a discount factor is not identified (see Rust 1994).

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

Estimation

Myopic agents, $\beta = 0$

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

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

$E_{e_{-i}}U_{it}(0) = 0$

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

Estimation of utility parameters

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

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

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

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

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

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

where

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

are agent beliefs, estimated in the first stage.

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

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

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

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

Estimation of the price elasticity

To estimate elasticity, I employ following strategy.

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

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

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

Forward-looking agents, $\beta = 0.9$

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

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

My estimation procedure follows Bajari et al. (2007). Compared to many other studies, the estimation strategy proposed by Bajari et al. has three advantages. First, this estimation procedure does not require the calculation of a transition matrix on the first stage. Avoiding this calculation decreases errors

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

of estimation. Second, this estimation strategy allows using sequential procedure estimation, wherein every step of estimation has closed-form solutions. This means that one can avoid mistakes and problems related with finding a global maximum using a maximization routine. Finally, this estimation procedure does not require discretization of variables. This flexibility of estimation routine allows me to work with the same extensive set of explanatory variables as in the myopic (static) model, and thus makes these two models comparable.

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

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

can be rewritten as two moment equations (for derivation see Proof A1 in the appendix):

Bellman equation for $V_i(0, s_t)$

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

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

$$\begin{aligned} log(\sigma_{it}(1)) - log(\sigma_{it}(0)) + V_{it}(0, s_t)_i = \pi_{it}(a_{-it}, a_{it} = 1, s_t, \theta) \\ +\beta E_{t+1}(V_{it+1}(0, s_{t+1}) - log(\sigma_{it+1}(0))|a_{it} = 1, s_t) \end{aligned} \quad (2)$$

These two equations together with a moment condition on choice probabilities

$$E(I(a_i = k)|s_t) = \sigma_{it}(k|s_t),\ k \in \{0, 1\} \quad (3)$$

form the system of moments I estimate in next section.

Estimation of utility parameters

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

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

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

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

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

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

I find $\widehat{V_i(0, s_t)}$ by finding $\hat{\mu}$ that solves following sample equivalent of moment condition (3):

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

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

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

Estimation of price elasticity

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

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

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

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

Results

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

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

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

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

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

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

Note: * significant at 10%** significant at 5%;*** significant at 1%
In elasticity estimates standard errors are clustered on municipality*year level

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

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

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

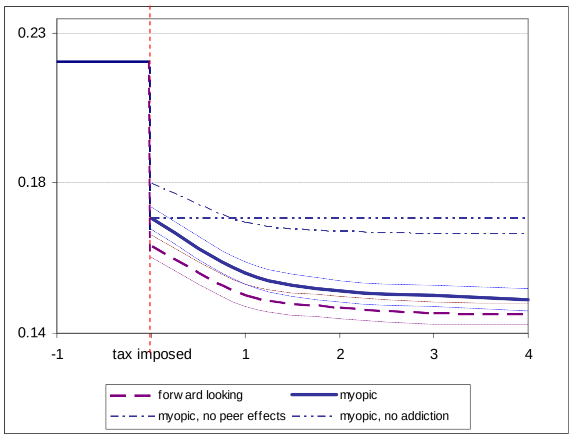
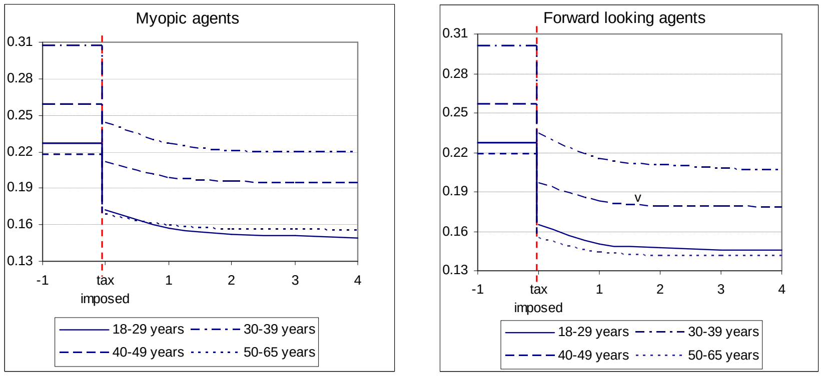


Figure 3 below illustrates the simulated effect of an increase in price for myopic and forward-looking agents in different age strata. Overall, five years after the introduction of a price-raising tax, the proportion of heavy drinkers will decrease by one-third. The effect is higher for younger generations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

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

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



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

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

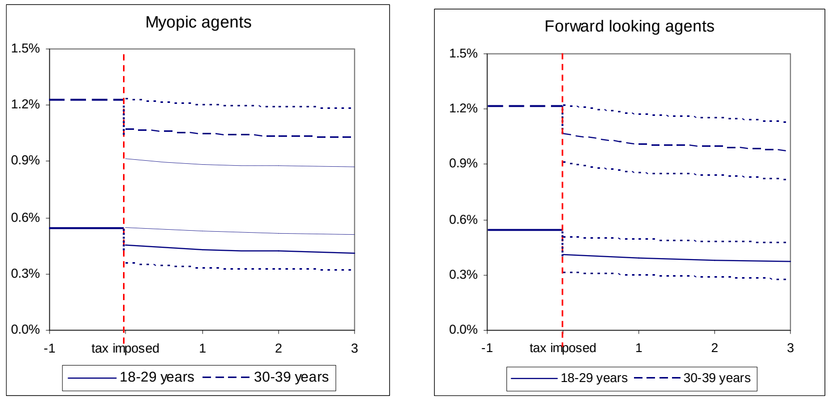
$$\lambda(t, x) = exp(x\beta)\lambda_0(t)$$

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

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

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

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



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

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

Besides, , my model, under certain assumptions of utilities, implies that the effect of a vodka tax on consumer surplus would be positive even for fully-rational agents, forward-looking agents who take into account the hazard of death associated with heavy drinking. The model I describe in the main body of my paper implies that peer effects and the effect of habits are positive: all other things being constant, an agent has higher utility if he or she drank within the previous period and if he or she has peers that are heavy drinkers. These forces, however, can equally run an agent's utility to the negative. First, quitting heavy drinking is costly. Second, an agent who decides not to drink may suffer from the fact that peers are drinking – the agent may experience peer pressure, or agent may suffer if no peer wishes to participate in alternative (to drinking) activities, such as playing soccer or doing other sports.[26]Thus, in the

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

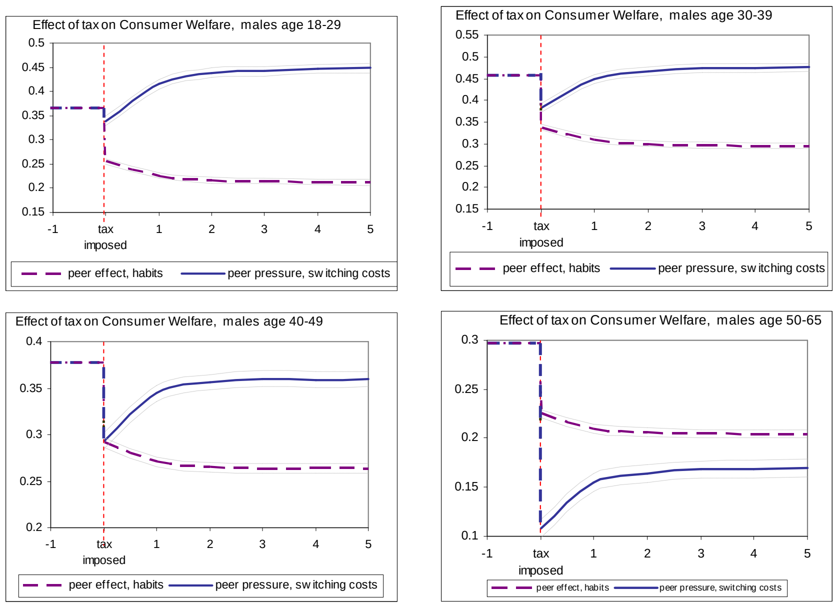
[25]My model does not take into account that the tax almost certainly saves other lives (children, females, the elderly), decreases crimes committed under alcohol intoxication, decreases car accidents, and so on.

[26]In this case, an agent's per-period choice specific utilities are as follows:

Robustness section I find that peer decisions matter for an agent if he or she decide to do physical training. These alternative assumptions on utilities, although barely distinguishable (or not distinguishable at all) from the data, have different implications for the analysis of consumer welfare.[27] In this case, case, a 50% tax on vodka results in an increase in the consumer welfare of young males below age 40.[28]

Figure 5 below illustrates this point.

Figure 5. Effect of tax policy on Consumer Welfare.



The final point I want to discuss is my finding that estimations of utilities and response functions, although different, do not differ dramatically in the myopic and forward-looking models. A possible explanation of this phenomenon is as follows. During the lengthy period in my analysis, Russia was in period of transition. This time people were uncertain about the future, and in particular about the realization of state variables such as future alcohol prices, future career, and income. In the context of my model, this may imply that agent expectations about future Value function are noisy, possibly not correlating with current state variables or having a strong effect on agent decision. In this case, even if in reality agents are forward-looking, an estimated "myopic" indirect utility may be a good enough approximation of the choice-specific Value function. Table A2 illustrates this point. Within one region and controlling for time trend, current alcohol prices do not provide a lot of information about future prices. This suggests that agents may not expect current shocks in

$\pi_{it}(0) = -\delta\overline{I(a_j = 1|S_{i,-i,t})} - \gamma a_{i,t-1}$, $\pi_{it}(1) = \Gamma' D_{it} + \Upsilon' G_{-it} + \rho_{mt}$

[27] See proof of identification results in the appendix (Proof A3).

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

price to result in future price changes.

Robustness

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

Reduced-form elasticity estimates

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

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

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

Linear in means peer effect

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

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

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

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

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

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

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

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

First, I present estimates of peer effects using average peer demographic characteristics as instruments. I estimate the model using the entire sample and also separately for different age strata, and for sub-samples without the two regions with a Muslim majority (the Tatarstan and Karachaevo-Cherkessk republics). I verify the robustness of my results by including different sets of fixed effects. Results are similar to those elsewhere in this paper.

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

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

Robustness of dynamic model assumptions

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

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

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

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

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

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

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

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

Habits versus unobserved heterogeneity

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

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

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

Extension

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

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

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

Conclusion

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

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

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

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

References

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Slinko I., E.Yakovlev and E. Zhuravskaya, “Laws for Sale: Evidence from Russia”, American Law and Economics Review, 2005

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

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

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

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

***************TO ADD:** ***************

Graham, K. (1985). Identifying and Measuring Alcohol Abuse among the Elderly: Serious Problems with Existing Instrumentation. Journal of the Studies of Alcohol, 47, 322

*****Elasticity of Alcohol: elasticity of beer/ vine/spirits are 0.3 1 and 1.5

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.

Tables

Table 3. Summary statistics.

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

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

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

Note: 3642 peers groups that contain 1 peer are excluded

Table 5. Agent's utility parameters.

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

Note: Bootstrapped standard errors in brackets

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

Table 6. Marginal utility of peers

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

Table 7. Estimates of price elasticity

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

Note: Robust standard errors, clustered at regionXyear level in brackets

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

Table 8. Mortality and heavy drinking.

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

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

Appendix

Figure A1. Alcohol consumption: age profile

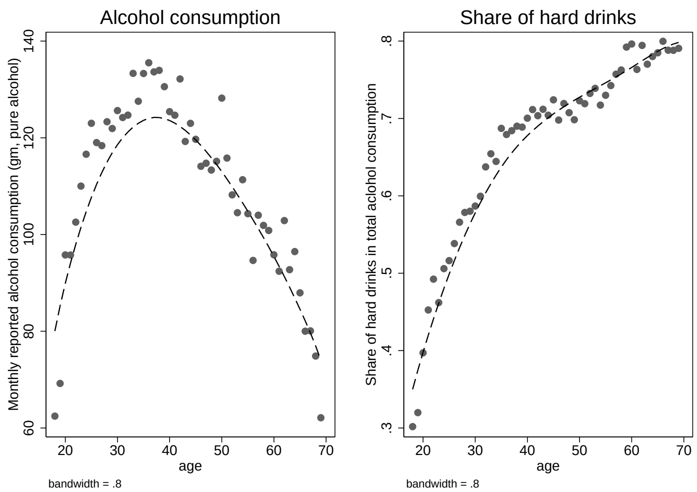


Figure A2. Dvors in Russia.

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

Table A1. Consumption of goods and birthday.

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

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

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

	Log(Vodka Price)
Lag (Log vodka price)	0.007
	[0.005]
Year FE	yes
Region FE	yes
Observations	36307

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

Table A3. Reduced form elasticity estimates. 2SLS regression.

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

Standard errors clustered at neighborhood level in brackets

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

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

	I(heavy drinker)							
	age 18-55				age18-29			
Peers effect, $\widehat{\delta}$:								
age 18-29	0.264	0.297	0.242	0.255	0.211	0.197	0.225	0.359
	[0.04]***	[0.05]***	[0.04]***	[0.09]***	[0.09]**	[0.136]	[0.14]*	[0.180]**
age 30-39	0.194	0.218	0.181	0.16				
	[0.03]***	[0.04]***	[0.03]***	[0.065]**				
age 40-49	0.063	0.089	0.053	0.063				
	[0.030]**	[0.037]**	[0.031]*	[0.059]				
age 50-59	-0.005	0.015	-0.022	0.009				
	[0.033]	[0.041]	[0.033]	[0.056]				
Demographics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Munic*year FE	Yes	Yes	Yes		Yes	Yes	Yes	Yes
Individual FE				Yes				
Year FE				Yes				
Muslim region excluded?			Yes					
Just came from military service?								Yes
Instruments	Peers 1	Peers 2	Peers 1	Peers 1	Peers 1	Fathers 1	Fathers 2	Peers 1
Observations	29554	29554	27400	29554	7750	8152	8152	149
F-test	79.9	36.29	72.02	17.02	34.24	16.52	28.97	6.85
J-test, p-value	0.22	0.13	0.26	0.02	0.06	0.4	0.86	0.17

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

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

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

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

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

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

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

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

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

Table A8. Habits versus unobserved heterogeneity.

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

Note: Instruments are Mean(Lag X, LagLag X, LagLagLag X), Mean(Lag X, LagLag X), and Lag X correspondingly, where X stands for I(health problems). Robust standard errors, clustered on individual level, are in brackets
* significant at 10%; ** significant at 5%; *** significant at 1%

Table A9. Log likelihoods for different betas. Rust approach.

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

Proof A1

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

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

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

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

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

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

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

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

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

Applying these relationships to equation for value function:

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

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

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

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

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

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

These two equations, together with moment equation on choice probabilities

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

form system of moments I estimated:

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

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

Proof A2

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

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

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

Lemma 1

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

Proof:

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

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

Lemma 2

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

Proof:

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

Applying these relationships to equation for value function:

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

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

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

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

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

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

These two equations, together with moment equation on choice probabilities

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

form system of moments:

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

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

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

Proof A3

Lemma

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

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

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

then

i) In myopic model ρ_0 and ρ_1 are not identifiable

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

Proof

i) In myopic model agent decides to drink if

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

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

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

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

$$V(0_{it}, s_t) = \rho_0 z_{it} + \beta E_{t+1}(exp(V(0, s) - log(\sigma(0))|0_{it}, s_t) \tag{4}$$

$$V(0_{it}, s_t) + log(\sigma(1)/(\sigma(0)) = \rho_1 z_{it} + \pi_{it}(a_{-it}, a_{it}, s_t, \theta) + \beta E_{t+1}(V(0, s) - log(\sigma(0)))|1, s_t) \tag{5}$$

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

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

Equations above imply

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

$\Leftarrow$

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

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

•

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

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

Proof:

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