American Economic Journal: Applied Economics 2018, 10(1): 106–149
https://doi.org/10.1257/app.20130170

Demand for Alcohol Consumption in Russia and Its Implication for Mortality†

By Evgeny Yakovlev*

Alcohol abuse is widely blamed for the very high rate of male mortality in Russia. I estimate a structural model of the demand for alcohol that incorporates two features of alcohol consumption, peer effects and habits. I use a kink in the policy regime of the excise tax on alcohol and regional variation in alcohol regulations to estimate a price elasticity of demand for alcohol. I find that peer influence and habits are critical determinants of the response of alcohol demand to price changes. The estimates imply that increases in alcohol prices would yield significant reductions in mortality. (*JEL* D12, H25, I12, L66, P23, P36)

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

While the patterns in Figure 1 suggest that policies to restrict alcohol use will reduce consumption and lower male mortality, the magnitudes of the responses

*New Economic School, Novaya str 100A, Skolkovo, Moscow, Russia (email: eyakovlev@nes.ru). I am especially grateful to David Card, Denis Nekipelov, Gerard Roland, and Katya Zhuravskaya for discussion of paper and kind support during my PhD study in UC Berkeley. I also thank Irina Denisova, Ruben Enikolopov, Max Kasy, Lorenz Kueng, Thomas Lemieux, Ted Miguel, Enrico Moretti, Anna Obizhaeva, Olexa Shwets, Mu-Jeung Yang, two anonymous referees, and seminar participants at Berkeley, CERGE-EI, Harvard School of Public Health, ESEM meeting, NES, HSE, RES meeting, and Acumen LLC for helpful discussions and comments. All errors are mine. Previous version: Peers and Alcohol: Evidence from Russia.

† Go to https://doi.org/10.1257/app.20130170 to visit the article page for additional materials and author disclosure statement or to comment in the online discussion forum.

[1] Approximately one-third of all deaths in Russia are related to alcohol consumption (see Nemtsov 2002). Most of the burden falls on males of working age: more than half of all deaths of working-age men are accounted for by hazardous drinking (see Leon et al. 2007 and Zaridze et al. 2009). Among recent economics studies of the connection between alcohol use and mortality in Russia are: Treisman (2010), Bhattacharya, Gathmann, and Miller (2013), Brainerd and Cutler (2005), and Kueng and Yakovlev (2014).

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

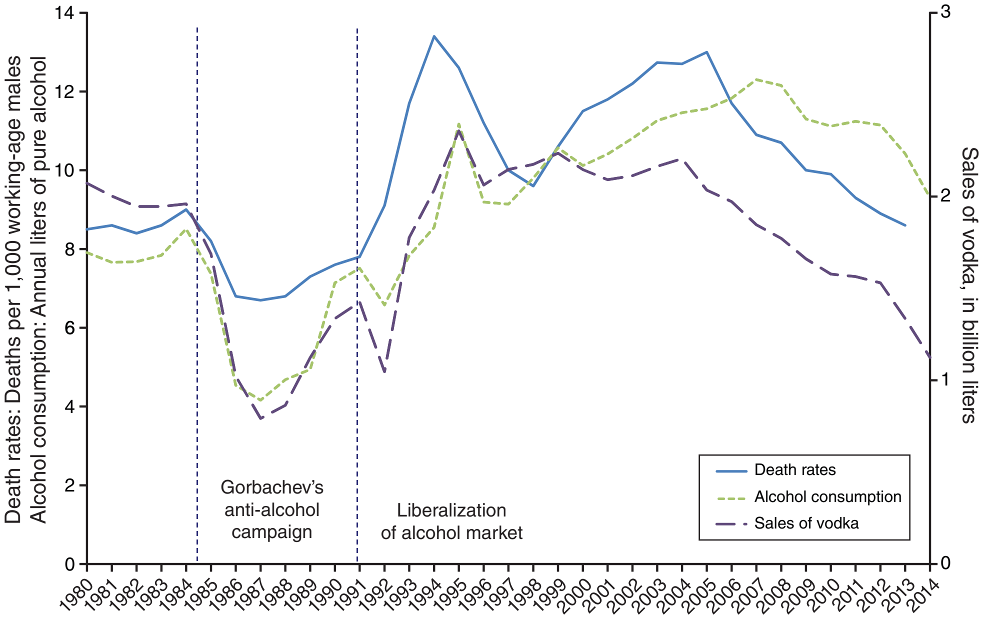

Figure 1. Alcohol Consumption and Male Mortality Rate

Source: WHO (2011), Treisman (2010), Rosstat (www.gks.ru). Left axis: deaths per 1,000 working-age males (Rosstat); annual adult per capita consumption, liters of pure alcohol (WHO 2011). Right axis: sales of vodka in billion of liters (Rosstat).

and the precise channels linking policy interventions to consumption and mortality are unclear.[3] The goal of this paper is to specify and estimate a dynamic model of alcohol demand that can illuminate these issues. The model incorporates two important features of alcohol demand that I show are very important in the Russian context: peer effects in consumption and habit persistence. Peer effects produce a "social multiplier" effect: decreases in a given consumer's own consumption lead his neighbors to consume less, so that the net effect of an alcohol price increase is amplified. Habit persistence similarly results in an intertemporal multiplier effect: decreases in alcohol consumption today change habits, reducing future preferences toward alcohol and leading to decreases in alcohol consumption in the future.

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

The price elasticity of drinking is identified using a regression kink design (RK strategy) and instrumental variables regression (IV strategy). To find the price elasticity with the RK strategy, I use a kink in the policy regime of the federal excise tax

[3] Another concern is that some consumers—particularly in the Soviet era—used homemade and illegally purchased alcohol.

on vodka. Before 2011, the excise tax had been linked to the CPI growth rate. Since 2011, the growth rate of the excise tax on vodka has exceeded the growth rate of the CPI more than twice.

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

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

This paper then verifies the predictions of the model with both myopic and forward-looking assumptions about agents' behavior. Although there is no consensus regarding which model is more accurate, most literature on policy analysis considers only the myopic assumptions.[4] At the same time, key effects of alcohol consumption—on health, family, and employment status, for example—do not necessarily appear immediately, but rather increasingly manifest themselves over the course of the next few years, or even much later in life (see Mullahy and Sindelar 1993 and Cook and Moore 2000). Moreover, alcohol consumption may form a habit and thus affect future behavior. One therefore expects that individuals may behave in a forward-looking manner when determining current alcohol consumption (see rational addiction literature, such as Becker and Murphy 1988).[5] Possible misspecification from omitting forward-looking agent assumptions might introduce a bias in estimates and, as such, might result in incorrect predictions regarding the effects of the proposed changes in the regulation of the alcohol industry.

I find significant price elasticity for heavy drinking and show the importance of peer effects for young age strata (below age 40). To illustrate these findings, I simulate the effect of an increase in vodka price by 50 percent on the probability of being a heavy drinker. The myopic model predicts that five years after introducing a price-raising tax, the proportion of heavy drinkers would decrease by roughly one-third, from 25 to 18 percent. The effect is higher for younger generations because of the nontrivial effect of the social multiplier. This cumulative effect can be decomposed in the following way: one's own one-period price elasticity predicts

[4] In particular, Rust (1987) shows that in a general setup of dynamic discrete-choice model the discounting parameter β is not identified. Although today different identification results are stated, they all are obtained under certain restrictions on parameters (see, for example, Magnac and Thermar 2002; Fang and Wang 2010; and Arcidiacono, Sieg, and Sloan 2007).

[5] Some studies find empirical evidence to support the rational addiction model (see Becker, Grossman, and Murphy 1991; Chaloupka 1991; and Arcidiacono, Sieg, and Sloan 2007). Other studies question this evidence (see Auld and Grootendorst 2004) or provide an alternative to a (fully) rational-model explanation of the evidence (see literature on time-inconsistent preferences, such as Gruber and Kőszegi 2001).

a drop in the share of heavy drinkers by roughly 4.5 percentage points, from 25 to 20.5 percent. Peer effects increase the estimated price response by 1.5 times for younger generations. The assumption that consumers are forward-looking increases the estimated cumulative effect roughly by an additional 30 percent.

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

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

Second, the paper provides several interesting and important examples of statistical relationships in the data. The kinked structure of the federal tax policy regime allows me to use regression kink (RK) estimates (see Card et al. 2015 and Lee and Lemieux 2010). The RLMS sampling structure allows me to document a strong increase in alcohol consumption around the birthdays of neighbors that shows that neighbors are indeed influential in personal decision making.

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

This paper is organized as follows. In the following section, I present the model. Section II describes the data and the variables used in the analysis. Section III presents the estimation strategy. In Section IV, I discuss results. Section V discusses the identification assumptions of the model and provides robustness checks and extensions. Section VI concludes.

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

I. Model

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

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

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

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

To model peer interaction, I use the methods proposed by Bajari et al. (2015) and Bajari et al. (2010) for estimating static and dynamic discrete games of incomplete information.[7] It assumes that each member of a peer group makes a decision on whether to drink heavily or not based on the expected average probability of heavy drinking among his peers. In a static (myopic) setup, this assumption leads to a simple two-stage procedure. In the first stage, I estimate the average expected probability of drinking by each individual's peer group. In the second stage, these forecasts are plugged into each agent's decision model. For a dynamic (forward-looking rational addiction) model, the first stage is the same, but the second stage involves first estimating a polynomial approximation for the value function of not drinking in the current period, and then (building on the Hotz-Miller inversion, see Hotz and Miller 1993) using this approximation in a third step to approximate to the probability of heavy drinking as a function of state variables and expectations.

Section IA describes the setup of the model in the event that consumers are myopic. Section IB extends the model for forward-looking consumers.

A. *Myopic Consumers*

The setup of the model is as follows. In every period of time t, the consumer chooses an binary action, a_{it}, whether to drink heavily ($a_{it} = 1$) or not ($a_{it} = 0$).

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

[7] For a review of these models, see Bajari et al. (2010) and Nevo (2011). For some recent developments, see Aguirregabiria and Mira (2007); Pakes, Ostrovsky, and Berry (2007); and Pesendorfer and Schmidt-Dengler (2008).

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

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

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

The per period utility from drinking has a linear parametrization:

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

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

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

Thus, $\pi_{it}(a_{it} = 1, a_{-it}, S_{it})$ depends on the average peer alcohol consumption, *habit* (defined as lagged alcohol consumption), a set of personal demographic and socioeconomic characteristics (D_{it}), a (sub)set of peers' characteristics G_{-it}, and municipality $\times$ year invariant factors ρ_{mt}. The variable X_{mt} stands for observable factors, and u_{mt} stands for unobservable factors that affect consumer utility and that vary only on the municipality $\times$ year level. Subscripts i, t, and m stand for individual, year, and municipality; subscript $-i$ stands for other individuals within the same peer group; and N stands for the number of peers in the peer group. For a detailed description of all variables, see Section III.

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

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

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

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

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

All components of utility from light/no drinking are normalized to zero: $\pi_{it}(a_{it} = 0) = 0$, $e_{it}(a_{it} = 0) = 0$, and

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

A consumer chooses to drink heavily if his or her per period utility from heavy drinking is greater than the utility from light/no drinking:

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

The left-hand side of equation (6) is per period utility from heavy drinking and the right-hand side is normalized utility from not drinking or light drinking.

B. *Forward-Looking Consumers*

In the myopic model, agents only account for factors that affect the current utility of alcohol consumption. At the same time, alcohol consumption may affect not only the current but also the future flow of utilities. Alcohol consumption is habit forming and thus affects future behavior. Many consequences of alcohol consumption that also affect consumer utility, such as health, family, income, and employment status, do not necessarily appear immediately, but rather they increasingly manifest over a few years (see Mullahy and Sindelar 1993 and Cook and Moore 2000). Indeed, Table 1 shows that heavy drinking affects future health, marital status, and income for Russian males. Conditional upon current health, marital status, and income, heavy drinking results in a decrease in future income (both own and family income), worsening future health, and higher risk of divorce. One would therefore expect that individuals may behave in a forward-looking manner when determining current alcohol consumption.

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

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

I set the annual discounting factor β equal to 0.9.[8] The current per period utility $\pi_{it}(a_{-it}, a_{it}, S_{it})$ and stochastic preference shock $e_{it}(a_{it})$ are similar to the myopic case

[8] Recent studies vary in their estimates of personal discount rate (see Arcidiacono, Sieg, and Sloan 2007; Hausman 1979; Dreyfus and Viscusi 1995; Moore and Viscusi 1990; and Pleeter and Warner 2001). The closest to

TABLE 1—EFFECT OF HEAVY DRINKING ON THE TRANSITION OF INCOME, MARITAL STATUS, AND HEALTH VARIABLES

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

Notes: Table 1 shows the estimates of the effect of heavy drinking on future health, marital status, and income for Russian males. Heavy drinkers are defined as those who belong to the top quarter by total alcohol intake. Columns 1–3 show results of OLS regressions $Y_{it+1} = \alpha + \theta I(heavy\ drinker)_{it} + \beta Y_{it} + u_{it}$. Columns 4–6 show results of probit regressions $\Pr(Y_{it+1} = 1) = \Phi(\alpha + \theta I(heavy\ drinker)_{it} + \beta Y_{it})$. Robust standard errors are in parentheses.

(see equations (2) and (3)). The consumer does not observe the actions of peers and forms expectations over peers' actions. The expected per period utility $E_{e_{-i}}\pi_{it}(a_{-it}, a_{it} = 1, S_{it})$ is the same as in the myopic case (see equation (4)).

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

$$(8) \quad E_{e_{-i}}\pi_{it}(a_{it} = 1, a_{-it}, S_{it}) + \beta E(V_{it+1}(s_{t+1}) \mid a_{-it}, a_{it} = 1, S_{it}) + e_{it}(a_{it} = 1)$$
$$> \beta E(V_{it+1}(s_{t+1}) \mid a_{-it}, a_{it} = 0, S_{it}).$$

In the case of forward-looking consumers, I assume that consumers have an infinite time-planning horizon and that the transition process of state variables is Markovian. This implies that expectations for future periods depend only on a current-period realization of state variables and the consumer's choice of action. Finally, I restrict the equilibrium to be a Markov Perfect Equilibrium, so that a consumer's strategy is restricted to be a function of the current state variables and the realization of a random part of utility (private preference shock).[9] For myopic consumers, the model is static, such that none of the assumptions described are needed.

our study, Arcidiacono et al. (2007), found a yearly discount factor of $\beta = 0.9$. I use this number ($\beta = 0.9$) when estimating a model with forward-looking consumers.

[9] These assumptions, together with other assumptions that I made (such as the Markovian state transition process, infinite time horizon, i.i.d. logistic error components, etc.) are standard assumptions in dynamic discrete-choice models. See Aguirregabiria and Mira (2010) for a review and for discussion of these and other assumptions that are commonly used in these models. These assumptions are done in order to simplify and to make it possible to implement the nontrivial computational task of the estimation of a dynamic discrete-choice problem.

II. Data Description

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

My study utilizes rounds 5 through 23 of RLMS over a time span from 1995 to 2014, except 1997 and 1999. I do not utilize data on earlier rounds because they were conducted by another institution, have different methodology, and are generally agreed to be of worse quality. Starting from round 17, the dataset that was provided by the Population Center at the University of North Carolina at Chapel Hill no longer contains identifiers of neighborhoods. I use information from the previous rounds and the HSE version of the RLMS data to collect neighborhood identifiers for rounds 17–23.[11]

The data cover 33 regions—31 oblasts (krays, republics), Moscow, and St. Petersburg.[12] Two regions have Muslim majority. Seventy-five percent of respondents live in an urban area. Forty-three percent of respondents are male. The percentage of male respondents decreases with age from 49 percent for ages 13–20 to 36 percent for ages above 50. The data only cover individuals older than 13 years.

The RLMS data have a low attrition rate, which can be explained by low levels of labor mobility in Russia (see Andrienko and Guriev 2004). Interview completion exceeds 84 percent. It is the lowest in Moscow and St. Petersburg (60 percent) and the highest in Western Siberia (92 percent). The RLMS team provides a detailed analysis of attrition effects and finds no significant effect of attrition.[13]

My primary population of interest for this research is males between ages 18 and 65. The threshold of 18 years is chosen because it is legal drinking age. The resulting sample consists of 78,237 individuals $\times$ year points (2,956 to 6,616 individuals per year). Summary statistics for the primary demographic and socioeconomic characteristics are presented in Table 2.

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

[11] A previous version of the paper uses only data from rounds 5 through 16 and finds similar results.

[12] Regions are equivalent to states in the United States.

[13] For a description of interview completion rates and attrition rates, see the RLMS website, http://www.cpc.unc.edu/projects/rlms-hse/project/samprep.

Table 2—Summary Statistics of the Key Variables

Variable	Observations	Mean	Standard deviation	Min	Max
Panel data (*males*)					
I(drunk more than 150 grams)	78,235	0.2503	0.433	0	1
log(family income)	78,507	3.988	1.817	0	9.787
Age	78,507	39.02	13.10	18	65
I(diseases)	74,454	0.343	0.474	0	1
Lag *I*(smokes)	61,563	0.633	0.481	0	1
I(employed)	78,453	0.712	0.452	0	1
I(college degree)	78,409	0.237	0.425	0	1
Body weight (kg)	78,071	77.2	13.87	35	250
I(big family)	78,507	0.194	0.395	0	1
I(Muslim)	78,507	0.081	0.272	0	1
Alcohol intake (grams of pure alcohol per day)	78,235	99.78	125.1	0	2,469
I(physical training)	67,483	0.174	0.379	0	1
I(drink tea)	22,415	0.966	0.180	0	1
I(drink coffee)	22,409	0.69	0.458	0	1
Prices and regulation					
log(price of vodka)	714	0.449	0.351	−1.02	1.36
CPI	714	75.43	48.74	4.41	262
Excise tax rate, vodka	600	190.3	116.6	55	500
Sum of regulations	495	0.461	0.738	0	3
Production regulation: Additional document	483	0.102	0.294	0	1
Production regulation: Premises regulation	483	0.129	0.329	0	1
Retail regulation: Additional document	483	0.156	0.357	0	1
Retail regulation: Excise machine	489	0.085	0.274	0	1
Survival regression data					
Death cases, male, >17 years	12,169	0.045	0.207	0	1

Notes: "Panel data (males)" report summary statistics of individual-level data from RLMS. "Prices and regulation" are summarized from data on municipality × cells. "Sum of regulations" is the sum of four indicators, additional document and premises regulation in production regulation, and excise machine and additional document in retail regulation. "Survival regression data" report "between" individual-level data from RLMS that is used in hazard-of-death regressions.

A. *Alcohol Consumption Variable*

Although the negative health and social consequences of hard drinking are widely recognized, there is no documented evidence of negative consequences from moderate drinking. Thus, I focus on an analysis of the personal decision to drink heavily or not. I use a dummy variable that equals one if a person belongs to the top quarter of alcohol consumption (among males of working age) and zero otherwise. Alcohol consumption is measured as the reported amount of pure alcohol consumed daily during the previous month.[14] The reported threshold level corresponds to the reported amounts greater than 150 grams of pure alcohol per day. This

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

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

amount corresponds to a daily consumption of 0.35 liters of vodka (or samogon) or 9 bottles (0.33 liters each, 3 liters total) of beer. Summary statistics and age profiles for reported amounts of alcohol consumption are shown in Table 2 and Figure A1 in the Appendix.

When constructing a measure of heavy drinking, I use data on both official and homemade alcohol (moonshine or samogon) and thus take into account all possible substitution effects. RLMS has data on the consumption of various alcoholic beverages including moonshine. Moonshine consumption is legal in Russia, so there is no reason to expect high (compared to other alcohol) underreporting of moonshine consumption. Indeed, moonshine (an inferior good) became less popular in Russia over the last two decades. According to RLMS data, the average—across all years—share of samogon in total alcohol intake equals 7 percent. Moreover, since 2009, the share of samogon does not exceed 5 percent, and importantly, there is no increase in the share of samogon in 2011–2014, when the price of vodka increased significantly (see Figure A2 in the Appendix).

B. *"Peers" Definition*

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

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

Approximately 10 percent of Russian families live in dormitories and communal houses where residents share kitchens and bathrooms. A majority of the remaining, more fortunate, part of the population lives in complexes of several multistory multi-apartment buildings, called "dvors." These complexes have their own playgrounds, athletic fields, and ice rinks and often serve as the place where people spend leisure time. The most common dvors (so-called "khrushchevki") are relatively small-size dvors with a population of about 300 people. A photo of a typical dvor is presented in Figure A3 in the Appendix. Dvors are the most popular place in Russia to find friends—the very low level of personal mobility in Russia means that most people live in the same place (and therefore the same dvor) for most of their lives.

The important feature of the RLMS survey is that it has a clustered structure.[15] The basic sampling unit of the RLMS survey contains one Russian census block. Households within the same set of census blocks are surveyed in every round. The average population of a census block in Russia is 300.[16] A typical census block in

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

[16] The RLMS team indicates that the population of census blocks in the RLMS survey is between 250 and 400 people. There are 459,000 census blocks in Russia (data from 2010 census). This number implies that the average population of the census block is 310 people (including females, youth, and elderly). This number in turn implies that the average size of a population of males of age 18–65 is about 90 people, and the total size of the peer group is 22 people (adult males in the same age strata).

Russia contains one dvor; this allows me to use information on the neighborhood (and age) to identify peer groups.

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

The median number of people in a peer group is 5, the mean is 11, the lowest 1 percent is 2, the ninetieth percentile is 34, and the largest number is 65.[17] On average, I have 794 peer groups per year (each with 2 or more peers). The distribution of the number of peers per peer group is shown in Table A1 in the Appendix.

To verify that individuals in the same census block and age stratum actually interact, I implement the following test. I correlate the logarithm of the amount of vodka consumed during the previous month with a dummy variable if a member of the peer group has a birthday in the previous month and with averages of the birthday dummy variables across peers. Vodka is the most popular alcoholic beverage to serve on birthdays, compared to either beer or wine.

The specification of the regression is as follows:

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

where *vodka* stands for the amount of vodka drunk in the last month, $I(birthday)_{kt}$ is an indicator that person k has a birthday in the previous month ($k \in \{i, j\}$), N is the number of people in the peer group, and δ_t are time fixed effects.[18] Column 1 of Table 3 reports regression estimates. The regression estimates imply that a person's consumption of vodka increases by 18 percent if his birthday is during the previous month and by 6 percent if there was a birthday of one of his peers in a group of five peers (median peer group size).[19] The results are robust eliminating household members from the sample of peers (see column 2 of Table 3). The results are also 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 4). The evidence therefore suggests that the peer clusters I defined reflect true peer interactions.

C. *Mortality*

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

According to medical studies, alcohol-related mortality represents 45 to 60 percent of deaths of Russian working-age men (see Leon et al. 2007 and Zaridze et al. 2009). The largest contributors to alcohol-related mortality among Russian males

[17] These numbers imply that I have data on about half of the total population of the peer group (see footnote 16).

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

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

Because I do not have data on all peers in a group, OLS estimates shown in Table 3 suffer from attenuation bias.

Table 3—Birthdays and Alcohol Consumption

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

Note: Standard errors clustered at neigborhood × year level are in parentheses.

Table 4—Consumption of Goods and Birthday

	I(drink vodka)	I(smokes)	I(drink tea)	I(drink coffee)
All peers				
$\frac{\sum_{peers} I(birthday)}{N-1}$	0.039	−0.01	−0.009	−0.023
	(0.013)	(0.012)	(0.008)	(0.020)
$I(birthday)$	0.034	0.011	−0.002	0.010
	(0.007)	(0.007)	(0.005)	(0.012)
Year × month FE	Yes	Yes	Yes	Yes
Observations	64,121	64,311	18,837	18,831
Without household members				
$\frac{\sum_{peers} I(birthday)}{N-1}$	0.038	−0.006	−0.007	−0.024
	(0.013)	(0.012)	(0.008)	(0.020)
$I(birthday)$	0.034	0.012	−0.001	0.010
	(0.007)	(0.007)	(0.005)	(0.012)
Year × month FE	Yes	Yes	Yes	Yes
Observations	63,884	64,063	18,788	18,782

Notes: The table checks the effect of peer birthdays on the consumption of different goods. In these regressions, I correlate the consumption variables with a dummy variable if a member of the peer group has a birthday in the previous month and with averages of the birthday dummy variables across peers. Standard errors clustered at neighborhood × year level are in parentheses.

are poisoning, accidents, and injures, followed by cardiovascular diseases (see Nemtsov 2002, Leon et al. 2007, Zaridze et al. 2009, and Shkolnikov et al. 2013). The pattern of alcohol-related mortality in Russia differs significantly from that in Western Europe or in the United States. Whereas the death pool associated with alcoholism or chronic diseases in Russia is relatively low, the mortality rate from hazardous drinking such as accidental poisoning, alcohol-related accidents, and injures is extremely high (see Shkolnikov and Meslé 1996, Nemtsov 2003, Leon et al. 2007, and Zaridze et al. 2009). The largest contributors to alcohol-related mortality are poisoning, accidents, and injures; the next largest contributor is cardiovascular disease, such as sudden heart stop under alcohol intoxication or stroke. The

Table 5—Distribution of Death Events by Age and Causes of Deaths

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

Notes: Table 5 shows the distribution of deaths and death causes for males from different age cohorts. The data on death events and death causes came from the RLMS survey.

main death burden that comes from excessive alcohol consumption lies on males of age 18–50, for whom hazardous drinking is prevalent (see Nemtsov 2003, Leon et al. 2007, Zaridze et al. 2009, and Shkolnikov et al. 2013).

Table 5 shows the distribution of deaths and death causes for males from different age cohorts. There are 626 death events; 44 (or 7 percent) are deaths of males of age 18–29, 86 (14 percent) are deaths of males of age 30–39, 149 (24 percent) are deaths of males of age 40–49, and 347 (55 percent) are deaths of males of age 50–65. RLMS recorded six causes of death, namely heart attack, stroke, external causes (accidents, injuries, and poisoning), cancer, tuberculosis, and "other" causes. The causes of deaths are reported only in 60 percent of death events. Deaths from poisoning, accidents, and violence are prevalent among young age cohorts. Out of deaths with reported causes, the deaths from poisoning, accidents, and violence constitute 63, 45, 29, and 9 percent of deaths of males of age 18–29, 20–29, 40–49, and 50–65, correspondingly. Deaths from heart attack and stroke constitute 9, 17, 38, and 51 percent of deaths of males of age 18–29, 20–29, 40–49, and of age 50–65, correspondingly. Deaths from cancer that are mainly not related with alcohol consumption are prevalent among older males. They constitute 6, 7, 9, and 29 percent of deaths of males of age 18–29, 20–29, 40–49, and of age 50–65, correspondingly.

III. Estimation

A. *Myopic Consumers*

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

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

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

At the first stage, I estimate beliefs (predicted probabilities of drinking) $\hat{\sigma}_{jt}(a_{jt} = 1 | S_{it})$ as a (arbitrary) function of state variables $S_{i,-i,t}$.[21] For the second stage, I estimate the remaining parameters of utility function by plugging predicted beliefs into the following logit regression:

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

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

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

The parameters of the model are identified under the assumption that the utility of one consumer does not depend on the subset of peer demographic characteristics and that random components of personal utility are independent of peer demographic characteristics (see Bajari et al. 2015 for proof). The exclusion restriction requires that subset G_{-it} does not contain all of the demographic variables. The restriction implies, for example, that the consumer does not have higher utility when he drinks with peers with different body weights or marital or health statuses. In Section V and in the online Appendix, I discuss the identification assumptions as well as different robustness checks of obtained results by allowing different sets of demographic characteristics to be excluded. Results of these regressions are robust to the choice of specifications, and *J*-tests for every specification support the hypothesis that excluded variables are exogenous.

To estimate the price elasticity, I assume that all price variation is captured on a municipality $\times$ year level. I obtain the municipality $\times$ year fixed effects component

[21] The expression for the first stage is as follows: $I(a_{jt} = 1)_{it} = H(s_{it})'\zeta + \varepsilon_{it}$, where $I_i = I(a_{it} = 1)$, $H(s_{it})$ is a set of Hermite polynomials of state variables s_{it} (for a discussion of nonparametric regression with Hermite polynomials, see Ai and Chen 2003). That is, $H(s_{it})$ contains a set of Hermite polynomials up to the third degree of $S_{i,-i,t} = U_{j\in\{i,-i\}}\{habit_{jt}, D_{jt}, G_{nt}, \rho_{mt}\}$. In addition, it includes interactions of state variables $U_{j\in\{i,-i\}}\{habit_{jt}, D_{jt}, G_{nt}\}$. I do not extend the set of polynomials to a larger degree or include a larger set of interactions because of the dimensionality problem. One important implication of this strategy is that ρ_{mt} appears in $H(s_{it})$ only once; this happens because the dummy variable structure of fixed effects implies that $\rho^k_{mt} = \rho_{mt}$. Still, ρ_{mt} will account for any variable (in any power) that varies only on the municipality $\times$ year level.

[22] I define state variable $habit_{it}$ as follows. Let state variable $habit_{it} = 0$ if $age_{it} < 18$ (years). The transition process of $habit_{it}$ is defined in following way: $habit_{it}(S_{t-1}, a_{i,\,t-1}) = a_{i,\,t-1} + \varphi_{i,\,t}$ if $age_{it} \geq 18$, where $a_{i,\,t-1}$ is the consumer's equilibrium choice of action in the previous period, and $\varphi_{i,\,t}$ is (negligible) smoothing noise. $\varphi_{i,\,t}$ is added to ensure the existence of an equilibrium. With this definition of habits, the model satisfies the assumptions required for the existence of a Markov perfect equilibrium (see, for example, Assumptions AS, IID, and CI-X in Aguirregabiria and Mira 2007 or Bajari, Hong, and Ryan 2010) which is required for dynamic models. A Markov perfect equilibrium (MPE) in this game is a set of strategy functions a^* such that for any consumer i and for any $\{S_t, e_{it}\}$, where $S_t = U_{j\in\{i,-i\}}\{habit_{jt}, D_{jt}, G_{nt}, \rho_{mt}\}$, we have $a^*_i(S_t, e_{it}) = b(S_t, e_{it}, a^*_{-i})$.

of utility $\hat{\rho}_{mt}$ and then regress $\hat{\rho}_{mt}$ on a log of the relative price of the cheapest vodka in the neighborhood and a set of control variables:

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

Parameter of interest θ stands for marginal utility of consumer with respect to logarithm of price. The set of municipality-level factors X_{mt} includes average income, level of education, and unemployment rate in a region as well as regional and—depending on specification—time fixed effects (IV regression) or smooth function of time trend (RK regression).[23]

To find the exogenous variation of price, I employ two alternative strategies. In the first specification, I utilize the regression kink design (RK) approach (for theoretical treatment, statistical packages, and discussion, see Card et al. 2015; Calonico, Cattaneo, and Titiunik 2014; and Lee and Lemieux 2010).[24] In the RK estimation, I explore a kink in the policy regime of the excise tax on vodka. In 2000, the Russian government introduced a specific excise tax on vodka.[25] From then until 2011, the excise tax was updated to catch up with the CPI. Since 2011, tax grew twice the rate of the CPI growth, that induced a kink in time profiles of excise tax and price of vodka.

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

Figure 3 shows averages (by year) of vodka prices and alcohol consumption according to RLMS data. Table A2 in the Appendix shows excise tax rates in years 2000–2014. To find the RK estimates, I modify regression (7) to be as follows:

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

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

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

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

[24] RKD explores the kink structure of policy functions (for example, the kink in tax schedule) and uses the variation in the slopes of the policy function around the kink to identify causal relationship. Under the assumption that all other factors behave smoothly in the neighborhood of the kink, RKD succeeds in identifying a causal effect by looking at the change in the slope of an outcome variable.

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

[26] The bandwidth size for the first-stage regression (using Imbens and Kalyanaraman's 2012 procedure) is equal to 2.86. I rounded it to 3 and use it in local polynomial regressions.

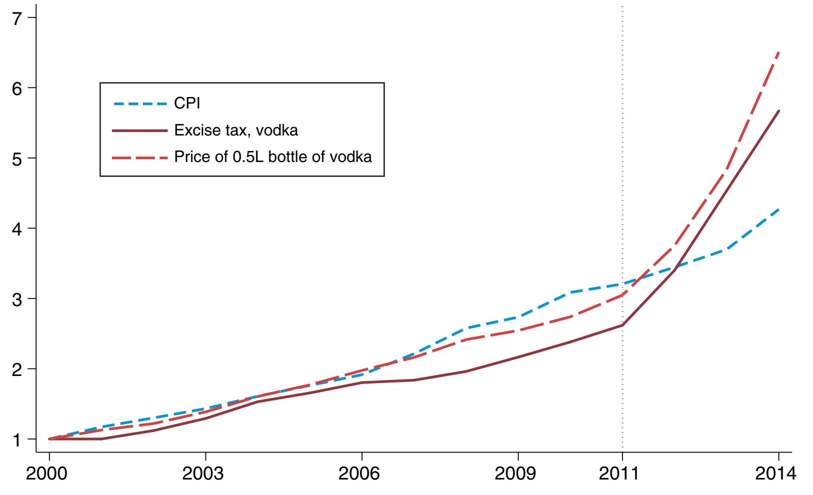

FIGURE 2. EXCISE TAX ON VODKA, THE AVERAGE PRICE OF VODKA, AND CPI

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

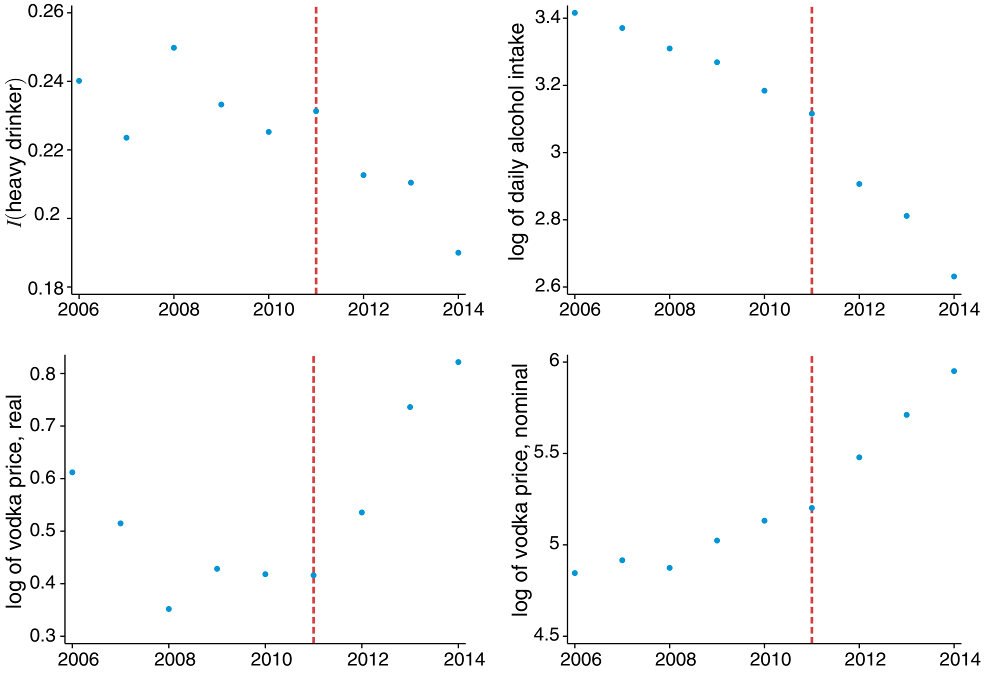

FIGURE 3. AVERAGES (*by year*) OF THE PRICE OF VODKA AND OF ALCOHOL CONSUMPTION

Note: The figures show the average alcohol prices and alcohol consumption around year 2011 (kink date in the policy regime of the excise tax on vodka).

Source: RLMS data, males of age 18–65

order polynomial of the time variable; in the local polynomial approach, $f(t)$ is parametrized as the linear function of time.

For robustness, I rerun the RK regressions discussed with a different instrumental variable. Instead of the kink variable $t \times I(year \geq 2011)$, I use the exact values of the excise tax as an instrument (excise tax profile is shown in Table A2 in the Appendix).

As an alternative approach, I use the data on regional regulation of the alcohol market to instrument the price variable. The time span for IV analysis is 1995–2008. During these years (Yeltsin's presidential terms and the beginning of the Putin administration), Russian regional authorities had substantial freedom to impose regulation procedures on local markets. I collect data on regional regulation of the alcohol market during this time, count additional regulations that the regional government imposes in a particular year, and use the number of additional regulations imposed by the regional government as an instrument in IV regressions.[27]

I obtain IV estimates from the following regression:

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

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

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

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

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

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

[27] As a rule, regional regulations are imposed for two reasons. First, regulations are a popular tool for increasing regional budget revenues: the excise tax and license tax are two of the very few taxes that go directly into the regional budget. Second, the regional regulations are imposed in the result of the lobbying of local firms and/or tollbooth corruption (see Yakovlev 2008; and Slinko, Yakovlev, and Zhuravskaya 2005). This implies that the introduction of new regulation is generally not motivated by public health reasons.

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

B. *Forward-Looking Consumers*

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

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

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

can be rewritten as two moment equations (for derivation see Proof A1 in the Appendix and Bajari et al. 2010, 2015):

The Bellman equation for $V_i(0, s_t)$ is

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

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

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

These two equations together with a moment condition on choice probabilities

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

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

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

In the second step, I estimate $V_{it}(0, S_{it})$ as an arbitrary function of state variables $S_{i,t}$ by solving a sample equivalent of the moment condition (16). To do this, I allow $V_{it}(0, S_{it})$ to be a (hermite) polynomial function of state variables $H(s_{it})'\mu$ and find $\widehat{V_i(0, s_t)} = H(s_{it})'\hat{\mu}$ by finding $\hat{\mu}$ that solves the equation

$$\begin{aligned} &I(a_{it} = 0)[H(s_{it})'\hat{\mu}] \\ &\quad = \beta I(a_{it} = 0)\Big[\log(1 + \exp(\log(\widehat{\sigma_{it+1}(1)}) - \log(\widehat{\sigma_{it+1}(0)})) + H(s_{it+1})'\hat{\mu}\Big]. \end{aligned} \tag{19}$$

[29] For surveys of dynamic discrete models, see research by Aguirregabiria and Mira (2010) and Bajari et al. (2015). Compared to many other studies, the estimation strategy proposed by Bajari et al. (2007) has three advantages. First, this estimation procedure does not require the calculation of a transition matrix on the first stage. Avoiding this calculation decreases errors of estimation. Second, this estimation strategy allows sequential procedure estimation, wherein every step of estimation has closed-form solutions. This means that one can avoid mistakes and problems related to finding a global maximum using a maximization routine. Finally, this estimation procedure does not require discretization of variables. This flexibility of the estimation routine allows me to work with the same extensive set of explanatory variables as in the myopic (static) model and thus makes these two models comparable.

In the third step, I estimate $\pi(1, S_{it})$ by solving the sample analog of the moment condition (17). I estimate $\pi(1, s)$ by solving for $\hat{\theta}$ the equation[30]

$$(20)\quad I(a_{it} = 1)\left[s_t'\hat{\theta} + \widehat{V_{it}(0, s_t)} + \log(\widehat{\sigma_{it}(1)}) - \log(\widehat{\sigma_{it}(0)})\right]$$
$$= \beta I(a_{it} = 1)\left[\gamma + \log\left(1 + \exp\left(\log(\widehat{\sigma_{it+1}(1)}) - \log(\widehat{\sigma_{it+1}(0)})\right)\right) + \widehat{V_{it}(0, s_{t+1})}\right].$$

The estimation of the price elasticity is similar to that employed in the myopic case. To simplify the description of the procedure, I start with an estimation of elasticity under the assumption that the government changes the price without changing consumers' expectations over future price movement.

To calculate the elasticity in this case, I obtain the municipality $\times$ year fixed effects components $\hat{\rho}_{mt}(\pi)$, $\hat{\rho}_{mt}(EV1)$, $\hat{\rho}_{mt}(EV0)$ of my estimates of the per period utility of drinking, $\pi_{it}(a_{-it}, a_{it} = 1, s_t)$, and conditional expectation of the future value function, $\beta E(V_i(S_{it+1})\,|\,a_{it} = 1, S_{it})$, and $\beta E(V_i(S_{it+1})\,|\,a_{it} = 0, S_{it})$. Then, I calculate the aggregate effect of the fixed effect components, $\hat{\rho}_{mt}$:

$$(21)\qquad \hat{\rho}_{mt} = \hat{\rho}_{mt}(\pi) + \hat{\rho}_{mt}(EV1) - \hat{\rho}_{mt}(EV0)$$

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

$$(22)\qquad \hat{\rho}_{mt} = \theta \log(Price)_{mt} + \Psi' X_{mt} + u_{mt}.$$

This estimation procedure relies on the assumption that consumers, when forming their expectations about future prices, use the rule of price motion guessed from their previous experience. In Russia, the price of alcohol is volatile, and the rule of price motion demonstrates significant mean reversion (see Table A3 in the Appendix). Therefore, the estimation implies that consumers believe that the current increase in price comes before its future decrease. If the government increases the price permanently and credibly promises that the price will not decrease in the future, then the expectations of consumers should be corrected.

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

First, I assume that the price-transition process is independent of all other state variables and personal choice of action and that it follows the AR rule of motion: $\log(p_{i,t+1}) = \phi_0 + \phi_1 \log(p_{it}) + \omega_{it}$, where $E(\omega_{it}\,|\,p_{it}) = 0$. Second, I assume the following parametrization of the choice-specific value functions: $V_i(S_{it}, a_{t-1} = j) = \vartheta_j \log(p_t) + V_i(\{S_{it}/p_t\})$, where $j \in \{0, 1\}$, and $\{S_{it}/p_t\}$ is a set of state variables excluding price.

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

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

$$\tilde{\rho}_{mt} = \hat{\rho}_{mt}(\pi) + \frac{1}{\hat{\phi}_1}\left(\hat{\rho}_{mt}(EV1) - \hat{\rho}_{mt}(EV0)\right) \tag{23}$$

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

$$\tilde{\rho}_{mt} = \theta \log(Price)_{mt} + \Psi' X_{mt} + u_{mt}. \tag{24}$$

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

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

C. *Effect of Mortality*

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

$$\lambda(t, X) = \exp(X\beta)\,\lambda_0(t), \tag{25}$$

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

I use a semi-parametric Cox specification of baseline hazard. A set of explanatory variables X includes I(heavy drinker), I(smokes), log of family income, health self-evaluation, body weight, current work status, and educational level. I allow heavy drinking to have a heterogeneous (by age stratum) effect on the 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.

IV. Results

Estimates of per period utility parameters are shown in Table 6 and in Tables 7 through 9. For myopic consumers, the per period (indirect) marginal utility with respect to log(price) is equal to -0.5 and -0.898 for base RK and for IV regressions,

[31] See Note 1 in the Appendix for proof.

Table 6—Consumer's Utility Parameters: Point Estimates

	Myopic consumers			Forward-looking consumers	
				Per-period utility	Value function
log(vodka price) peer effect, $\hat{\delta}$:	−0.500	−0.898	−0.458	−0.744	−1.021
Age 18–29	1.444	1.444	1.444	1.272	1.366
Age 30–39	0.861	0.861	0.861	0.711	0.879
Age 40–49	0.326	0.326	0.326	0.363	0.412
Age 50–65	0.209	0.209	0.209	0.385	0.464
Instruments for price	(1)	(2)	(1) + (2)	(2)	(2)

Notes: The sets of instruments are (1) kink in excise tax and (2) regional regulation. Price elasticity estimates come from Tables 7 and 8. Peer effects' estimates come from Table 9.

Table 7—Estimates of Price Elasticity: Myopic Consumers; RK Regression

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

(*continued*)

respectively (see Table 6 and the first columns of Tables 7 and 8).[32] For a myopic consumer with a mean level of all demographic characteristics, the marginal utility of −0.5 implies that, for example, an increase in the price of vodka by 50 percent

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

Table 7—Estimates of Price Elasticity: Myopic Consumers; RK Regression (*continued*)

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

Notes: Panel A reports fuzzy RK estimates of price elasticity for a model with myopic agents. Different columns report estimates from different specifications (with varying bandwidth sizes, instruments, and kernel specifications). Panel B reports the corresponding first stages of fuzzy RK estimates. Bootstrapped standard errors are in parentheses.

Table 8—Elasticity Estimates: IV Regressions; Forward-Looking and Myopic Assumptions

		Forward-looking				
	Myopic	Per-period utility	Value function		Value function	
	(1)	(2)	(3)	(4)	(5)	(6)
log(vodka price)	−0.898	−0.744	−1.021	−0.846	−1.349	−1.308
	(0.293)	(0.410)	(0.490)	(0.367)	(0.707)	(0.470)
Share with college degree	−1.604	−0.880	−2.081	−1.322	−2.355	−2.234
	(0.413)	(0.379)	(0.520)	(0.402)	(0.715)	(0.525)
Average income	0.065	0.035	0.188	0.091	0.121	0.094
	(0.074)	(0.064)	(0.087)	(0.068)	(0.116)	(0.089)
Employment	0.950	0.643	0.597	0.626	1.365	1.196
	(0.365)	(0.297)	(0.364)	(0.305)	(0.508)	(0.395)
I(city)	0.169	0.074	0.260	0.142	0.316	0.254
	(0.091)	(0.080)	(0.117)	(0.088)	(0.149)	(0.116)
Share with diseases	0.888	0.456	1.101	0.693	1.126	1.010
	(0.254)	(0.232)	(0.298)	(0.237)	(0.412)	(0.312)
Constant	−3.766	−0.290	−1.293	−0.659	−1.509	−1.419
	(0.473)	(0.451)	(0.527)	(0.452)	(0.742)	(0.537)
Observations	415	414	414	414	414	414
F-test, first stage	39.78	39.87	39.87	39.87	39.87	39.87
Normalization		*U*(not drink) = 0	*U*(not drink) = 0	*U*(not drink) = 0	*U*(drink) = 0	*U*(drink) = 0
Commit to permanent price change			Yes	No	Yes	No

Notes: The table reports IV estimates of price elasticity for models with myopic agents and for models with forward-looking agents. Instrumental variables are regional regulations. Different columns report estimates from different specifications (with different normalization of per-period utility and different condition on price change commitment). *F*-test shows the *F*-statistic from the test for joint significance of the instruments in first-stage regression. Bootstrapped standard errors are in parentheses.

Table 9—Consumer's Utility Parameters

	Agent's (per-period) utility			Agent's (per-period) utility	
	$\beta = 0$	$\beta = 0.9$		$\beta = 0$	$\beta = 0.9$
Peer effect, $\hat{\delta}$:					
Age 18–29	1.444	1.272	log(family income)	−0.033	−0.015
	(0.215)	(0.502)		(0.012)	(0.021)
Age 30–39	0.861	0.711	*Age*	0.127	0.116
	(0.154)	(0.347)		(0.011)	(0.045)
Age 40–49	0.326	0.363	*Age*^{2}	−0.001	−0.001
	(0.154)	(0.414)		(0.0001)	(0.0006)
Age 50–65	0.209	0.385	*Body weight*	0.008	0.006
	(0.215)	(0.597		(0.001)	(0.002)
			I(*diseases*)	−0.017	−0.005
				(0.026)	(0.057)
Habit:					
Lag *I*(heavy drinker)	1.456	1.401	*I*(*big family*)	0.062	0.058
	(0.04)	(0.071)		(0.031)	(0.096)
Lag *I*(heavy drinker) × *Age*	−0.028	−0.016	Lag *I*(*smokes*)	0.505	0.427
	(0.015)	(0.015)		(0.029)	(0.063)
Lag *I*(heavy drinker) × *Age*^{2}	0.0004	0.000	*I*(*work*)	−0.155	−0.175
	(0.00017)	(0.000)		(0.084)	(0.084)
			I(*college degree*)	−0.170	−0.189
				(0.032)	(0.089)
			I(*Muslim*)	−0.272	−0.176
				(0.063)	(0.067)
Municipality × year FE	Yes	Yes			
Peers' mean characteristics	Yes	Yes			
Observations	50,763	50,763			

Notes: Columns "$\beta = 0$" report estimates of per-period utility parameters for a model with myopic agents (see Section IIIA for detailed description). Columns "$\beta = 0.9$" report estimates of per-period utility parameters for a model with forward-looking agents (see Section IIIB for detailed description). Bootstrapped standard errors are in parentheses.

will lead to a decrease in the probability of heavy drinking by 4 percentage points (from 0.25 to 0.21).

For forward-looking consumers, the per period (indirect) marginal utility with respect to log(price) is equal to −0.74. To evaluate the effect of a change in price on forward-looking consumers, one must know not only the consumer's per period utility but also have an expectation of the consumer's future value function. The marginal value function of consumers with respect to log(price) is equal to −1.021 (see Tables 6 and 8).[33] The marginal value function of −1.021 implies that an increase in the price of vodka by 50 percent leads to a decrease in the probability of becoming a heavy drinker by 6.5 percentage points.

In both the myopic and forward-looking specifications, I find that peers have a strong effect on younger generations, with the effect decreasing as age increases.

[33] The elasticity is calculated under the assumption that a price increase is permanent. In the event that the government cannot ensure that the change in price is permanent, the elasticity is −0.765. For the description of the calculation procedure, see the Appendix.

For the two youngest strata, the effect is statistically significant. For myopic consumers, $\hat{\delta}$ equals 1.444, 0.861, 0.326, and 0.209 for ages 18–29, 30–39, 40–49, and 50–65, respectively (see columns 1–3 of Table 6 and column 1 of Table 9). For forward-looking consumers, $\hat{\delta}$ equals 1.272, 0.711, 0.363, and 0.385 for ages 18–29, 30–39, 40–49, and 50–65, respectively (see column 4 of Table 6 and column 2 of Table 9).

The myopic model allows for an immediate statistical interpretation of the peer coefficients: an increase in the average per alcohol consumption of 0.2 (corresponding to a situation in which one out of five peers in a group becomes a heavy drinker) will increase the probability of becoming a heavy drinker for the "mean" person in the age group 18–29 by 5.4 percentage points and for the "mean" person in the age group 30–39 by 2.8 percentage points.[34] Again, the forward-looking model does not allow immediate statistical interpretation. In Table 6, I present estimates of the marginal utility and marginal value function of peers, evaluated at the mean value of other state variables.

It is worth noting that estimates 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: for most of my analysis, Russia was in a period of transition. During this time, people were uncertain about the future, particularly about the realization of state variables such as future alcohol prices, future career, and income. In the context of my model, this may imply that consumers' expectations about the future value function are noisy, possibly not correlated with current state variables, or having a strong effect on consumer decisions. In this case, even if in reality consumers are forward-looking, an estimated "myopic" indirect utility may be a good enough approximation of the choice-specific value function. Table A3 in the Appendix illustrates this point. My data imply that in this case, consumers should expect a significant mean reversion in price movement. According to column 2 of Table A3, a 10 percent change in price today is associated with only a 4 percent change in the expected price next year.

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

Figure 4 illustrates the decomposition of the cumulative response to the change in price for males aged 18–29 for the myopic model base RK specification. Dashed lines show the effect of a price increase on myopic consumers for three situations: in a model where peer effects and habit formation are included, in a model without peer effects, and in a model without habit formation. The difference in effects refers

[34] These estimates are based on logistic regression estimates that are shown in columns "$\beta = 0$" (columns 1 and 3) of Table 9 and are evaluated for a person with mean level of socioeconomic characteristics.

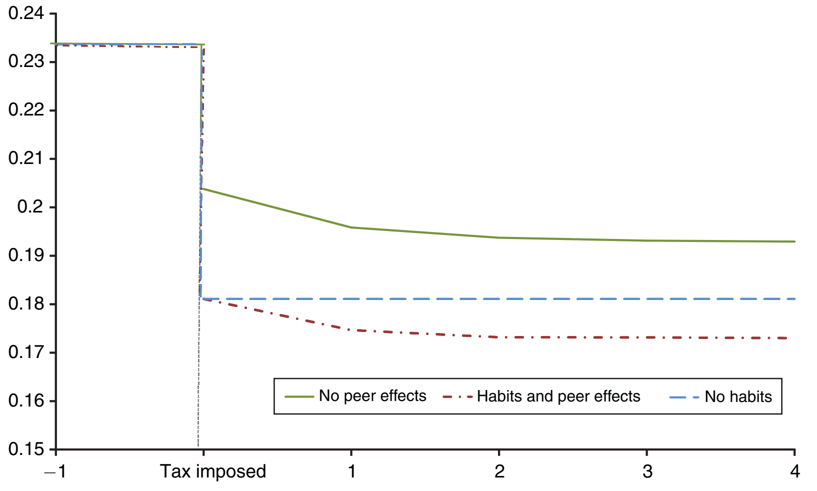

FIGURE 4. EFFECT OF TAX ON Pr(*heavy drinker*), AGE 18–29

Notes: The figure shows the decomposition of the effect of a 50 percent increase in the price of vodka on the share of heavy drinkers among young male adults. The horizontal axis is the years before and after imposing tax. The vertical axis is the share of heavy drinkers.

Source: RLMS, males of age 18–29

to the effect of the social multiplier and of the "habit multiplier." Solid lines show the effect of a price-increasing tax for forward-looking consumers. The model predicts a decrease in the proportion of heavy drinkers by 6 percentage points, from 23.3 percent to 17.2 percent over 5 years. Taking into account only peer effects or only habit formation leads to a prediction of smaller changes (4 percentage points in the case with only habits and 5 percentage points in the case with only peer effects). Finally, one's own price elasticity results in a one-time change of 3 percentage points, which is approximately half of the cumulative effect.

Figure 5 illustrates the simulated effect of an increase in price for myopic and forward-looking consumers in different age strata. In this example, I work with estimates obtained for the myopic model in the base RK specification and for the forward-looking model in the base IV specification.[35]

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

A. *The Effect of a Change in Vodka Price on Mortality Rates*

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

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

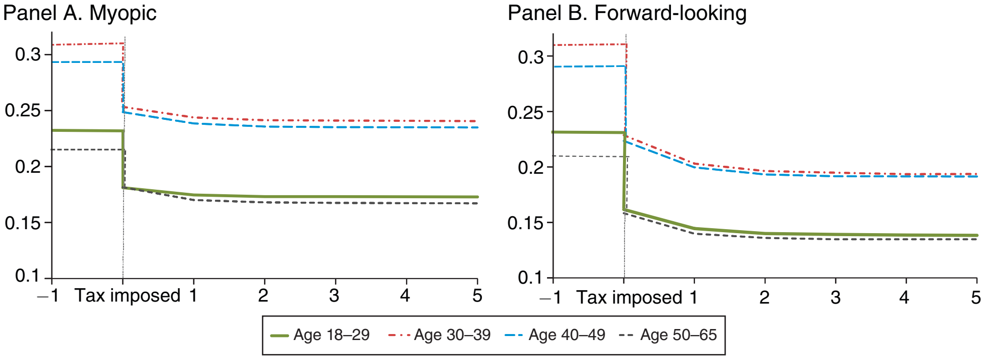

Figure 5. Effect of a 50 Percent Tax on the Share of Heavy Drinkers

Notes: The figure shows the simulated effect of a 50 percent increase in the price of vodka on the share of heavy drinkers in different age cohorts. The horizontal axis is the years before and after imposing tax. The vertical axis is the share of heavy drinkers.

Source: RLMS, males of age 18–65

Table 10 shows the estimates of the effect of heavy drinking on the hazard of death.[36] The effect of heavy drinking is highly heterogeneous by age. The hazard of death for heavy drinkers aged 18–29 is 9.8 times higher than for other males of the same age. The hazard of death for heavy drinkers of age 30–39 is 5.5 times higher. The hazard of death for heavy drinkers of age 40–49 is 1.8 times higher. There is no statistically significant difference between hazard rates for heavy drinkers and non-heavy drinkers aged 50–65. Absence of correlation between hazard of death and alcohol consumption for oldest age cohort might be due to the fact that people with serious illnesses consume little alcohol and at the same time have a higher risk of death. The bias due to this confounding factor is especially high for old people with high rates of chronic diseases, cancer, and other illnesses. Even when controlling for observable health indicators, the unobservable differences in health may drive this result. Also, because regression estimations are done for a relatively short period of 19 years, they do not capture the very long-run (negative) consequences of alcohol consumption.

Using hazard-of-death regression estimates, I simulate the effect of a change in vodka price on mortality rates. Figure 6 shows the simulated effect of increasing the price of alcohol on mortality rates for males of the three youngest age strata. The simulated effect (in the case of myopic consumers) of introducing a 50 percent tax is a decrease in mortality rates by one-fifth (from 0.45 percent to 0.36 percent) for males aged 18–29 years, by one-seventh (from 0.71 percent to 0.62 percent) for males aged 30–39 years, and by one-twentieth (from 1.1 percent to 1.05 percent) for

[36]Table A2 in the online Appendix reports estimates of the hazard of death by different causes of death. Unfortunately, the causes of death are reported in less than 60 percent of death cases. The RLMS recorded six causes of death, namely heart attack, stroke, external causes of death (accidents, injuries, and poisoning), cancer, tuberculosis, and "other" causes. Splitting death events into different groups reduces the power of the tests and increases the standard errors of the coefficients. For young generations, correlation between heavy drinking and the hazards of death is positive for all causes but tuberculosis and statistically significant for death due to accidents, poisoning, and other causes. The only positive correlation between heavy drinking and the risk of death for the older age cohort (ages 50–65) is found for the hazards of death due to poisoning, violence, and accidents. Table A2 in the online Appendix reports the hazard of death with different measures of heavy drinking. Results are the same.

TABLE 10—MORTALITY AND HEAVY DRINKING

	Coefficient	Hazard ratio		Coefficient
I(heavy drinker) age 18–29	2.295 (0.467)	9.826	log(family income)	−0.414 (0.036)
I(heavy drinker) age 30–39	1.704 (0.353)	5.485	I(smokes)	0.595 (0.124)
I(heavy drinker) age 40–49	0.582 (0.315)	1.789	I(college degree)	−0.075 (0.132)
I(heavy drinker) age 50–65	−0.268 (0.246)		Body weight	−0.005 (0.004)
Bad health (self-evaluation)	1.401 (0.164)		I(work)	0.089 (0.146)
Observations	12,125			

Notes: The table reports results of a hazard-of-death regression $\lambda(t, X) = \exp(X\beta)\lambda_0(t)$. A semi-parametric Cox specification of baseline hazard $\lambda_0(t)$ is used. Both coefficients (β) and hazard rates $(\exp(X\beta))$ are reported for cohort-specific effects of I(heavy drinker). Regressions are based on a sample of males of age 18–65. Standard errors are in parentheses.

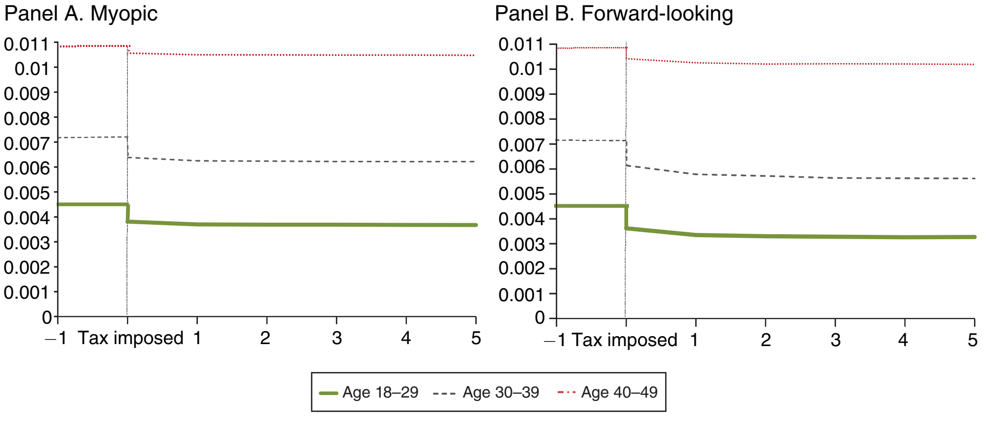

FIGURE 6. EFFECT OF 50 PERCENT TAX ON MORTALITY RATES

Notes: The figure shows the simulated effect of a 50 percent increase in the price of vodka on mortality rates among males of age 18–49. The horizontal axis is the years before and after imposing tax. The vertical axis is the share of heavy drinkers.

Source: RLMS, males of age 18–59

males aged 40–49 years. There is little immediate effect on the mortality of males of older ages. In other words, a 50 percent increase in the price of vodka would save 30,000 (male) lives annually. This is a lower bound (in magnitude) estimate of the effect. Under the forward-looking assumption as well as in the other specification of the myopic model (IV regression), the effect of this policy is more than 50,000 saved lives.

I find also that when agents have bounded rationality (that is, they do not take into account the effect of consumption on the hazard of death), the value of saved lives overweights the losses in consumer and producer surpluses, and in result, an increase in vodka price by 50 percent improves welfare. In both the forward-looking and myopic models presented, consumers have bounded rationality: they do not take into account the effect of heavy drinking on the hazard of death.[37] Within these models, the tax corrects a negative externality that appears from the bounded rationality of consumers. The welfare effect of the 50 percent tax is a 30 percent loss in consumer surplus.[38] At the same time, the tax saves 30,000–50,000 young male lives annually, which is 0.04–0.06 percent of the working-age population. The rough estimation of the value of their lives is the present value of the GDP that they generate. With a time discount of $\beta = 0.9$, the value of saved lives is equal to 0.4–0.6 percent of GDP, which equals the size of the whole alcohol industry in Russia (0.48 percent of GDP). This speculative calculation suggests that a 50 percent tax is actually likely to be smaller than the optimal one.[39] Additionally, under certain assumptions about consumer utilities, a tax increases consumers' welfare even for fully rational agents. (See the online Appendix for an explanation and discussion of this result.)

V. Identification Assumptions, Robustness Checks, and Discussion

A. *Discussion of Identification Assumptions*

In this section, I discuss the identification assumptions of the price elasticity estimation in my model. A discussion of identification assumptions for peer effect and habits can be found in the online Appendix.

Tables 7 and 8 show results of the *F*-test for the relevance of instruments in both fuzzy RK and IV regressions, respectively. Column 7 of Table 7 and Table 8 show *F*-statistics for the base RK and IV regressions, respectively. The *F*-statistics equal 27.15 for the base RK and 39.8 for the base IV regression. Both *F*-statistics exceed ten, so the instruments are strong. In other global polynomial RK regressions, *F*-statistics range from 27 to 51 depending on the specification (see columns 8–10 of Table 7). In the local polynomial RK regression, the *F*-statistic is equal to eight due to small sample size and lack of power to provide the corresponding test.

The identification of RK estimates relies on several additional assumptions. First, the price policy regime should have a kink at year 2011. Second, predetermined covariates that affect outcome Y should be smooth at 2011. I test these assumptions by checking for the presence of a kink and discontinuity in the following regressions:

$$Y_{mt} = \alpha_0 [I(after2011)_{mt} \times t] + \alpha_1 t + \alpha_2 t^2 + \eta_r + u_{mt} \tag{26}$$

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

[38] Consumer welfare is the expected (over realization of private utility shocks) present value of the flow of utilities. Under my model assumptions, $\Delta E(CS) = \frac{1}{\alpha_i}\left[\ln\left(\sum \exp(V_{ij})\right)|tax - \ln\left(\sum \exp(V_{ij})\right)|\, no\ tax\right]$, where V_{ij} is the choice-specific value function (for a consumer i and choice j), and α_n is the marginal utility of income (negative coefficient with price).

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

Table 11—Test for Smoothness: Price, Alcohol Consumption, and Social-Economic Characteristics

	$\widehat{\rho_{mt}}$ (1)	Share of heavy drinkers (2)	log(vodka price) (3)	Employment (4)	Share with college degree (5)	Average income (6)
Panel A. Test for kink						
$I(after2011) \times run$	−0.073	−0.012	0.226	0.001	0.001	−1.529
	(0.042)	(0.007)	(0.033)	(0.006)	(0.006)	(2.645)
run	0.029	0.003	0.009	−0.001	0.001	14.96
	(0.026)	(0.004)	(0.018)	(0.003)	(0.003)	(1.603)
run^2	0.004	0.001	−0.002	−0.000	0.000	−0.096
	(0.002)	(0.000)	(0.001)	(0.000)	(0.000)	(0.115)
Observations	561	561	561	561	561	561
Panel B. Test for jump						
$I(after2011)$	0.064	−0.010	0.063	−0.007	0.023	−2.526
	(0.085)	(0.014)	(0.069)	(0.011)	(0.010)	(5.362)
run	−0.012	0.000	0.089	0.001	−0.003	14.807
	(0.025)	(0.004)	(0.018)	(0.003)	(0.003)	(1.486)
run^2	0.001	0.001	0.005	−0.000	0.000	−0.114
	(0.002)	(0.000)	(0.001)	(0.000)	(0.000)	(0.098)
Observations	561	561	561	561	561	561

Notes: The table reports results of regressions that test for smoothness in profiles of predetermined covariates. I test for smoothness by checking for the presence of a kink (panel A) and discontinuity (panel B) in the following regressions $Y_{mt} = \alpha_0[I(after2011)_{mt} \times t] + \alpha_1 t + \alpha_2 t^2 + \eta_r + u_{mt}$ (panel A) and $Y_{mt} = \alpha_0[I(after2011)_{mt}] + \alpha_1 t + \alpha_2 t^2 + \eta_r + u_{mt}$ (panel B). In these regressions, Y_{mt} stands for the prices and alcohol consumption variables as well as the predetermined characteristics. The set of controls includes regional fixed effects, CPI, and I(city). Robust standard errors are in parentheses.

and

$$Y_{mt} = \alpha_0[I(after2011)_{mt}] + \alpha_1 t + \alpha_2 t^2 + \eta_r + u_{mt}, \tag{27}$$

where Y_t stands for prices and alcohol consumption variables as well as predetermined characteristics (average educational level, income, and employment). Here, t and t^2 stand for time and time squared, and η_r stands for regional fixed effects. Coefficient α_0 in regression (26) shows the size of the kink; coefficient α_0 in regression (27) shows the size of the jump in 2011. The regressions are estimated for the sample from 2000 to 2011.

Table 11 shows the estimation results. There is a statistically significant kink (α_0) for regressions with price and alcohol consumption but no evidence of a kink in regressions where the dependent variable is demographic and socioeconomic characteristics. It also shows no evidence of a jump except in one regression.[40]

In addition, I perform a simulation experiment where I move a placebo kink from 2006 until 2013 and estimate regression (26) for the sub-sample of years within an interval of three years from the placebo kink date.[41] Figure 7 shows the levels and 95 percent confidence intervals of α_0 for the regressions with different placebo dates of kink. With the presence of a kink, one should expect that graphs would have a U

[40] α_0 is statistically significant in one regression where educational level is the dependent variable.
[41] In these regressions, I add a linear function of time instead of a quadratic polynomial function of time.

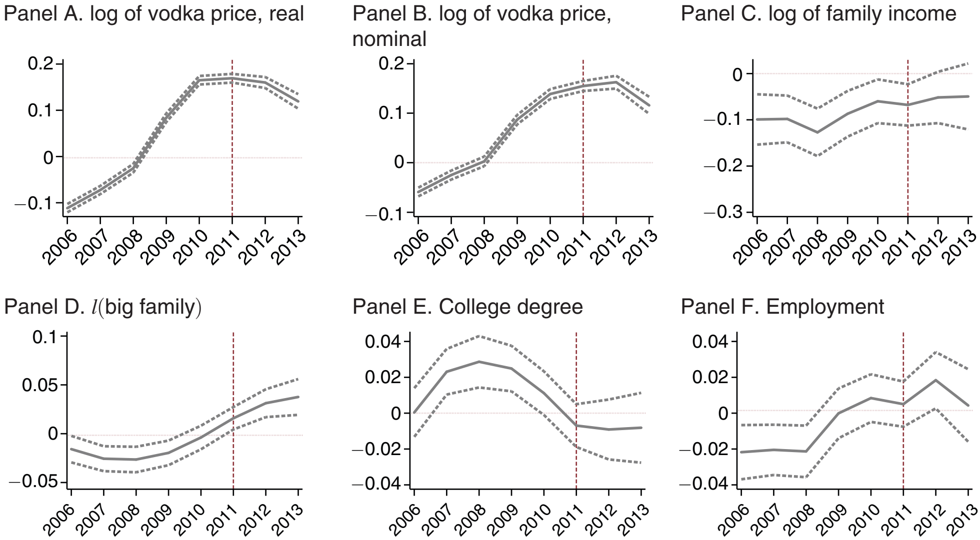

Figure 7. Placebo for Kink

Notes: Figure 7 shows the levels and 95 percent confidence intervals of the size of the kink (α_0) for the regressions with different placebo dates of kink. α_0 is estimated from regressions $Y_{mt} = \alpha_0 [I(\textit{after placebo date}) \times t] + \alpha_1 t + \alpha_2 t^2 + \eta_r + u_{mt}$. Placebo dates of kink are on the horizontal axis; α_0 is on the vertical axis.

(or an inverse U—depending on the sign of the kink) form with the top (bottom) around 2011. Indeed, Figure 7 shows exactly this pattern for regressions with prices, but not for demographic or socioeconomic characteristics.

B. *Alternative Elasticity Estimates*

Table A5 in the Appendix presents point estimates of elasticity for alternative specifications. Table A5 shows RK and IV estimates for different sets of instrumental variables with different sets of regional fixed effects as well as different assumptions about price movement in the forward-looking model. All estimates lie in a range from −0.218 to −2.194 with a mean of −0.878 and median of −0.841.

Table A6 shows RK estimates for alternative definitions of heavy drinkers. In the first model, heavy drinkers are defined as those who belong to the top 25 percent by alcohol intake within each ten-year age cohort. In the second model, heavy drinkers are defined as those who belong to the top 50 percent by alcohol intake. In the third model, heavy drinkers are defined as those who belong to the top 25 percent by frequency of alcohol consumption (days per week). According to Table A6, the price elasticities of heavy drinking for these three models are in range from −0.36 to −0.61.

Table A7 shows RK estimates for regional sales of alcohol. According to Table A7, the price elasticities of alcohol consumption range from −0.56 to −0.81. Table A8 in the Appendix presents reduced-form elasticity estimates from a linear global RK regression

$$\textit{Share of heavy drinkers}_{mt} = \theta \log(\textit{Price})_{mt} + \Psi' X_{mt} + u_{mt}. \quad (28)$$

The variable *Share of heavy drinkers*$_{mt}$ stands for the share of heavy drinkers in the neighborhood. The RK specification is similar to the global polynomial specification discussed. The set of control variables X_{mt} includes a second order polynomial of the time (running) variable and averages of the following variables: education, work status, I(smokes), health status, weight, religion, size of family, and log(family income).

In addition, I obtain reduced estimates of elasticity for different age cohort groups.[42] Table A8 in the Appendix reports price elasticity estimates. The elasticity estimates lie in a range from -0.068 to -0.093, and all are statistically significant.[43] Estimates of cohort-specific elasticities show higher elasticities for the two youngest cohorts of age 18–29 and 30–39, which coincides with the observation of higher social multiplier effects for younger generations.

Finally, I check the robustness of the dynamic model assumptions. Earlier I did not model the idea that consumers are likely to correctly estimate their hazard of death, and so I now take this into account. I verify the robustness of the results after accounting for this factor. In this robustness experiment, a consumer has the discounting factor $\beta\lambda(t, s)$ where hazard rates depend on state variables and on a consumers decision about heavy drinking. The results of this estimation are presented in Table A4 in the Appendix. Again, utility parameters do not differ from those shown because the actual hazard of death is very small, especially for the younger generation.

C. *Comparison with Elasticity Estimates from Other Studies*

The simulation example discussed in the results section (see Figure 4) implies that the short-run elasticity of heavy drinking equals -0.44 and that the long-run elasticity equals -0.52.[44,45] These estimates are comparable with elasticity estimates that come from meta-analysis studies (see Leung and Phelps 1993; and Wagenaar, Salois, and Komro 2009). The latest meta-analysis study, Wagenaar, Salois, and Komro (2009), reports an average elasticity of -0.44 for total alcohol intake and -0.28 for heavy drinking.[46] In our data, the fact that the average price elasticity of overall alcohol intake is -0.44 implies that the price elasticity of my measure of heavy drinking is equal to -0.53. The available estimates of elasticity of alcohol in Russia (see Andrienko and Nemtsov 2006 and Treisman 2010) report elasticities of total alcohol intake ranging from -0.145 to -0.67. Again, our estimates are in a range between these two numbers.

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

[43] Recall that Table A8 presents results of linear regression whereas main specification regressions (Tables 7–9) present results of logistic regressions that make direct comparison of coefficients senseless.

[44] In Figure 4, the elasticities are as follows:

$$SR\ Elasticity = \frac{\%Share\ of\ heavy\ drinkers}{\%price} = \frac{(0.233 - 0.181)/0.233}{0.5} = 0.44;$$

$$LR\ Elasticity = \frac{\%Share\ of\ heavy\ drinkers}{\%price} = \frac{(0.233 - 0.172)/0.233}{0.5} = 0.52.$$

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

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

VI. Conclusion

In this paper, I estimate a dynamic model of drinking behavior that incorporates several important determinants of drinking such as the price of alcohol, neighborhood (peer) effects, and drinking habits. I fit the model to Russian micro-level data. The nature of Russian data allows me to identify several key parameters of the model. To estimate the price elasticity of heavy drinking, I explore a kink in the policy regime of the excise tax on vodka. In 2011, the Russian government introduced a new tax policy regime. Before 2011, the excise tax on vodka was growing proportionally to the CPI. Since 2011, the tax growth rate twice exceeded the CPI growth. I use the kink in tax policy regime to apply regression kink estimation to establish the causal relationship between price and drinking. I show that RK estimates are similar to the results of instrumental variables regressions where variation in regulations of regional alcohol markets was used as instrument for the price of alcohol.

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

These results are especially important from a policy perspective, since alcohol consumption is a big problem in Russia itself. Over the past 20 years, Russia has experienced one of the largest historical surges in mortality during peace time, and it is widely attributed to heavy alcohol consumption.

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

To illustrate this finding, I estimate the impact of public policy (specifically, higher taxation) on the demand for heavy drinking and consequently on mortality rates. I simulate the effect of imposing the tax that increases the price of vodka by 50 percent. The myopic model predicts that five years after the introduction of the price-raising tax, the proportion of heavy drinkers will decrease by roughly one-fourth—from 25 to 18 percentage points. The effect is higher for younger generations because of the nontrivial effect of the social multiplier. This cumulative effect can be decomposed in the following way: one's own one-period price elasticity predicts a drop in the proportion of heavy drinkers by roughly 4.5 percentage points, from 25 to 20.5 percent. In addition, peer effects and habit formation assumptions increase the estimated price elasticity by 1.9 times for younger generations, and by about 1.4 times for the older generation. In a model with forward-looking consumers, the effect of a change in price is higher by roughly 30 percent.

With this established, I simulate the effect on mortality rates from this increase in the price of alcohol. I find significant age heterogeneity in the effect of heavy drinking on the hazard of death: the hazard is much stronger for younger generations. A 50 percent tax on the price of vodka will save 30,000–50,000 male lives annually, or 1 percent of young male adult lives in 6 years.

Appendix

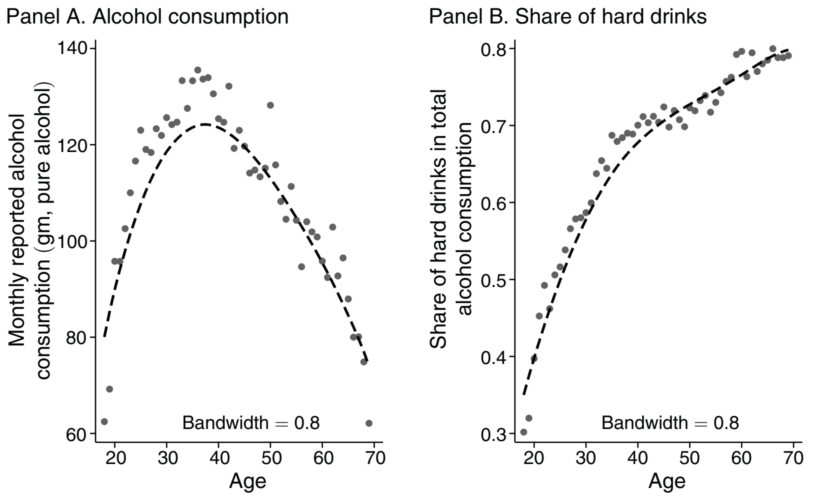

Figure A1. Alcohol Consumption: Age Profile

Source: RLMS, subsample of males of age 18–65

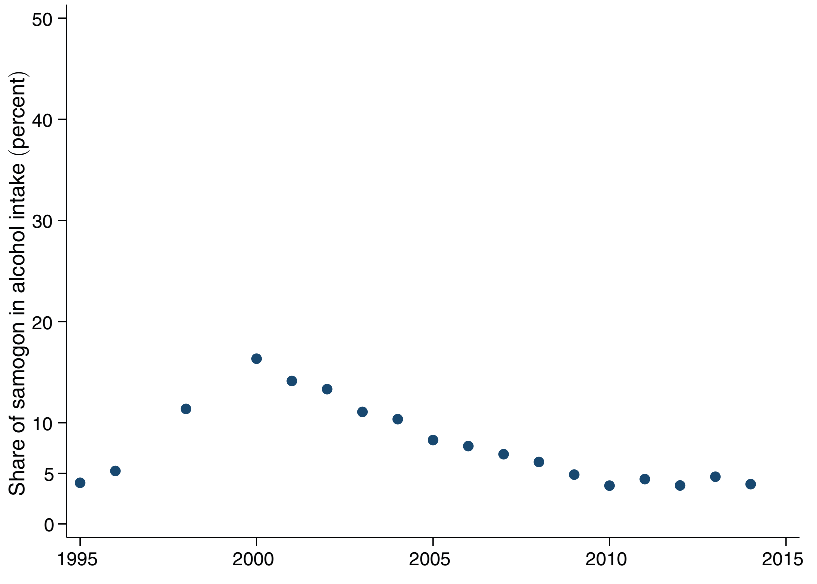

Figure A2. Share of Samogon (*moonshine*) in Total Alcohol Intake

Source: RLMS, subsample of males of age 18–65

Figure A3. Typical Dvor (*"khrushevka"*) in Russia

Source: E. Yakovlev

Table A1—Distribution of the Number of Peers in Peer Groups

	(Peer group)-level data			Individual-level data		
Number of peers in peer group	Freq.	Percent	Cum. %	Freq.	Percent	Cum. %
2	4,954	37.78	37.78	9,908	17.31	17.31
3	3,487	26.59	64.37	10,461	18.28	35.59
4	1,896	14.46	78.82	7,584	13.25	48.83
5	983	7.50	86.32	4,915	8.59	57.42
6	489	3.73	90.05	2,934	5.13	62.55
7	240	1.83	91.88	1,680	2.94	65.48
8	171	1.30	93.18	1,368	2.39	67.87
9	74	0.56	93.75	666	1.16	69.04
10	77	0.59	94.33	770	1.35	70.38
11	73	0.56	94.89	803	1.40	71.78
12	68	0.52	95.41	816	1.43	73.21
13	54	0.41	95.82	702	1.23	74.44
14	44	0.34	96.16	616	1.08	75.51
15	28	0.21	96.37	420	0.73	76.25
16	39	0.30	96.67	624	1.09	77.34
17	27	0.21	96.87	459	0.80	78.14
18	29	0.22	97.09	522	0.91	79.05
19	23	0.18	97.27	437	0.76	79.81
20 and more	358	2.73	100.00	11,555	20.19	100.00
Total	13,114	100.00		57,240	100.00	

Notes: Table A1 reports the distribution of peers in peer group. Peers are defined as those males who live in one neighborhood (census block) and belong to the same ten-year stratum. Excluded are 5,780 peer groups that contain 1 peer.

Table A2—Excise Tax on Vodka

Year	Excise tax on vodka	Year	Excise tax on vodka
2000	88.2	2008	173
2001	88.2	2009	191
2002	98.78	2010	210
2003	114	2011	231
2004	135	2012	300
2005	146	2013	400
2006	159	2014	500
2007	162	2015	600

Note: The table shows the value of excise tax on vodka (in rubles per one liter of pure alcohol).

Table A3—Lag log(Vodka Price) Is Not a Good Predictor for Current log(Vodka Price)

	$\log(vodka\ price)_t$	$\log(vodka\ price)_t - \log(vodka\ price)_{t-1}$
$\log(vodka\ price)_{t-1}$	0.336	
	(0.037)	
$\log(vodka\ price)_{t-1} - \log(vodka\ price)_{t-2}$		−0.436
		(0.044)
Observations	38,297	30,463
R^2	0.128	0.194

Notes: The table reports estimates of alcohol price motion using AR(1) specification for log prices and for (log) price changes. The time span in regressions is the same as in the dynamic model. Robust standard errors clustered at municipality × year level are in parentheses.

Table A4—Forward-Looking Model with Hazard-of-Death Discounting

	Per-period utility		Value function	Value function	Per-period utility
Peer effect, $\hat{\delta}$:					
Age 18–29	1.155	log(vodka price)	−1.152	−0.878	−0.719
	(0.053)		(0.441)	(0.325)	(0.283)
Age 30–39	0.774	Commit to permanent price change	Yes	No	
	(0.038)				
Age 40–49	0.372				
	(0.038)				
Age 50–65	0.275				
	(0.053)				
Habits	1.396				
	(0.007)				

Notes: The table reports estimates of utility parameters for the model with forward-looking consumers, where consumers take into account the effect of drinking on the hazard of death. In the model consumers discount, future utilities flow with an additional discount factor, hazard of death. Bootstrapped standard errors are in parentheses.

Table A5—Point Estimates of Price Elasticity for Different RK and IV Specifications

Panel A. Static, RKD								
log(price of vodka)	−0.595	−0.500	−0.451	−0.403	−0.465	−0.306		
Kernel	triangle	triangle	uniform	uniform	uniform	uniform		
IV	excise tax	after × run	excise tax	after × run	excise tax	after × run		
BW size	11	11	11	11	3	3		
Panel B. Static, IV								
log(price of vodka)	−0.898	−1.127	−0.841	−0.736	−0.218	−0.458		
Specification	IV1	IV1	IV1	IV2	IV2	IV3		
Regional FE		Yes	Yes		Yes			
Regional trends			Yes		Yes			
Fedokrug FE, trends	Yes			Yes		Yes		
Panel C. Dynamic, IV, normalization: U(no drink=0)								
log(price of vodka)	−1.021	−0.846	−0.822	−0.724	−0.291	−0.467	−0.890	−0.731
Permanent price change	Yes	No	Yes	No	Yes	No	Yes	No
Specification	IV1	IV1	IV1	IV1	IV1	IV1	IV2	IV2
Fedokrug FE, trends	Yes	Yes					Yes	Yes
Regional FE			Yes	Yes	Yes	Yes		
Regional trends					Yes	Yes		
Panel D. Dynamic, IV, normalization: U(drink=0)								
log(price of vodka)	−1.349	−1.308	−2.194	−1.714	−1.177	−0.921	−1.081	−1.059
Permanent price change	Yes	No	Yes	No	No	Yes	Yes	No
Specification	IV1	IV1	IV1	IV1	IV1	IV1	IV2	IV2
Fedokrug FE, trends	Yes	Yes					Yes	Yes
Regional FE			Yes	Yes	Yes	Yes		
Regional trends					Yes	Yes		

Notes: The table reports point estimates of price elasticity for various regression specifications. Panels A and B report estimates for models with myopic consumers. Panels C and D report estimates for models with forward-looking consumers. IV regression specifications are as follows. IV1: the instrument is the sum of regional regulations; IV2: the instruments are the set of four regional regulation variables; and IV3: the instruments are the sum of regional regulations and federal excise tax.

Table A6—Model Parameters' Estimates under Different Definitions of Heavy Drinkers

	Model 1		Model 2		Model 3	
	(1)	(2)	(3)	(4)	(5)	(6)
log(price of vodka)	−0.609	−0.521	−0.509	−0.477	−0.366	−0.356
	(0.270)	(0.293)	(0.243)	(0.255)	(0.210)	(0.228)
Peer effect, $\hat{\delta}$:						
Age 18–29	1.856		1.274		1.108	
Age 30–39	0.420		0.845		0.611	
Age 40–49	−0.012		0.297		0.197	
Age 50–65	0.507		0.536		0.126	
Habits	1.432		1.658		1.602	
IV	excise tax	after × run	excise tax	after × run	excise tax	after × run
F-test	29.65	27.06	29.65	27.06	29.65	27.06

Notes: The table reports estimates of price elasticity, peer effects, and habits in models with myopic consumers under different definitions of heavy drinking. The heavy drinking definitions are: model 1: top 25 percent by alcohol intake within ten years of age cohorts; model 2: top 50 percent by alcohol intake; and model 3: top 25 percent by days of alcohol consumption (per month). In all models, price elasticity estimates come from global RK estimates with triangle kernel. Robust standard errors are in parentheses.

Table A7—Regional-Level Regression: RK Estimates of Elasticity

	log(sales of alcohol)					
	(1)	(2)	(3)	(4)	(5)	(6)
log(price of vodka)	−0.798	−0.815	−0.697	−0.722	−0.565	−0.564
	(0.141)	(0.162)	(0.109)	(0.119)	(0.175)	(0.178)
run	0.033	0.036	0.021	0.024	−0.015	−0.015
	(0.026)	(0.028)	(0.021)	(0.023)	(0.032)	(0.032)
run^2	0.005	0.005	0.004	0.004		
	(0.002)	(0.002)	(0.002)	(0.002)		
log(CPI)	0.810	0.806	0.377	0.359	0.389	0.388
log(GDP per capita)	0.400	0.403	0.385	0.389	0.377	0.377
Unemployment	−0.072	−0.072	−0.055	−0.055	−0.075	−0.075
Population	−0.000	−0.000	−0.000	−0.000	−0.000	−0.000
Observations	847	847	925	925	534	534
IV	excise tax	after × run	excise tax	after × run	excise tax	after × run
Kernel	triangle	triangle	uniform	uniform	uniform	uniform
Sample	2003–2014	2003–2014	2003–2014	2003–2014	2008–2014	2008–2014
F-test, first stage	121.8	109.1	213.5	197.4	282.6	266.2

Notes: The table reports results for fuzzy RK estimates of price elasticity of alcohol consumption using regional-level data. The dependent variable is the logarithm of regional sales of alcohol (measured in pure alcohol). The data source is Rosstat data for 78 Russian regions (see www.gks.ru). Robust standard errors clustered at the regional level are in parentheses.

Table A8—Reduced-Form Elasticity Estimates: RK Regression

	(1)	(2)	(3)	(4)
Panel A. I(*Heavy drinker*)				
Price elasticity				
log(price of vodka)	−0.093	−0.087	−0.078	−0.068
	(0.042)	(0.045)	(0.035)	(0.035)
	(5)	(6)	(7)	(8)
Panel B. I(*Heavy drinker*)				
Cohort-specific price elasticity				
log(price of vodka)				
Cohort: age 18–29	−0.136	−0.112	−0.184	−0.129
	(0.051)	(0.048)	(0.053)	(0.045)
Cohort: Age 30–39	−0.151	−0.133	−0.156	−0.127
	(0.055)	(0.052)	(0.056)	(0.048)
Cohort: Age 40–49	−0.011	−0.027	0.036	0.009
	(0.048)	(0.048)	(0.051)	(0.045)
Cohort: Age 50–65	−0.070	−0.073	−0.013	−0.039
	(0.048)	(0.046)	(0.048)	(0.045)

Notes: The table reports reduced-form price elasticity estimates from linear global RK regression. Columns 1–4 show the estimates of elasticity for two kernel types, triangular and rectangular, and for two instruments, *Excise tax on vodka* and *run* × *I*(*after*2011). Columns 5–8 show estimates of cohort-specific elasticities. Robust standard errors are in parentheses.

Note 1. *Calculation of Marginal (with respect to price) Value Function*

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

$$\log(p_{i,t+1}) = \phi_0 + \phi_1 \log(p_{it}) + \omega_{it},$$

where

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

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

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

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

Under these assumptions,

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

Without a commitment on price stability, $\frac{\partial E_p(\log(p_{t+1}))}{\partial \log(p_t)} = \phi_1$. Once the government can commit to the price not reverting, then $\frac{\partial E_p(\log(p_{t+1}))}{\partial \log(p_t)} = 1$, and therefore

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

PROOF A1:

Derivation of the moment conditions, model with forward-looking assumption (with $\beta = 0.9$).

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

The first relationship is called Hotz-Miller inversion (see Hotz and Miller 1993):

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

The second equation states the relationship between the Emax function and the choice-specific value functions:

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

where $\gamma = 0.577$ is Euler's constant.

Applying these relationships to the equation for the value function:

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

When I put $a_{it} = 0$, and $a_{it} = 1$ in the equation, I have the
moment condition on $V_i(0, s_{it})$:

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

The moment condition on $V_i(1, s_{it})$ is

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

These two equations, together with moment equation on choice probabilities

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

form the system of moments I estimated:

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

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

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

PROOF A2:

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

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

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

Then:

(i) *in the myopic model, ρ_0 and ρ_1 are not identifiable;*

(ii) *in the forward-looking model, ρ_0 and ρ_1 are identifiable if and only if there is no $f(s_t, z_{it})$, such that*

$$f(s_t, z_{it}) - \beta \times E[f(s_{t+1}, z_{it+1}) \,|\, a_{it} = j, s_t, a_{-it}] = \phi_j \times z_{it} \quad \text{for } j \in \{0, 1\}.$$

PROOF:

(i) In the myopic model, the 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 $E_{t+1}(\cdot \,|\, 1)$, $E_{t+1}(\cdot \,|\, 0)$.

In the case of a forward-looking consumer, the value function is fully characterized by two equations:

$$(29) \qquad V(0_{it}, s_t) = \rho_0 z_{it} + \beta E_{t+1}(\exp(V(0, s)) - \log(\sigma(0)) \,|\, 0_{it}, s_t);$$

$$(30) \quad V(0_{it}, s_t) + \log(\sigma(1)/(\sigma(0)) = \rho_1 z_{it} + \pi_{it}(a_{-it}, a_{it}, s_t, \theta)$$
$$+ \beta E_{t+1}(V(0, s) - \log(\sigma(0)) \,|\, 1, s_t).$$

Suppose that there exists another pair $V(0_{it}, s_t)'$, ρ_j' for which these two equations hold.

Define

$$\Delta_j = \rho_j' - \rho_j, \quad f(s_t, z_{it}) = V(0_{it}, s_t) - V(0_{it}, s_t)'.$$

The equations imply

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

so we have established a contradiction.

$\Leftarrow$

Assume that

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

and let $V(0_{it}, s_t)$, ρ_j be a solution to the equations. Then $V(0_{it}, s_t)'$, ρ_j', are such that $V(0_{it}, s_t)' = f(s_t, z_{it}) + V(0_{it}, s_t)$, and $\rho_j' = \rho_j + \phi_j$ will be the solution of equations (29) and (30). ∎

Note: An example of where we can not identify ρ_1 and ρ_0:

if there are ϕ_j, such that $E(z_{it+1} \,|\, a_{it} = j, s_t) = \zeta + \phi_j \times z_{it}$, then we cannot identify ρ_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 equations (29) and (30) hold for new $V(0_{it}, s_t)'$, ρ_j'. ∎

REFERENCES

Aguirregabiria, Victor, and Pedro Mira. 2007. "Sequential Estimation of Dynamic Discrete Games." *Econometrica* 75 (1): 1–53.

Aguirregabiria, Victor, and Pedro Mira. 2010. "Dynamic discrete choice structural models: A survey." *Journal of Econometrics* 156 (1): 38–67.

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

Andrienko, Yury, and Sergei Guriev. 2004. "Determinants of interregional mobility in Russia." *Economics of Transition* 12 (1): 1–27.

Andrienko, Y., and A. Nemtsov. 2006. "Estimation of Individual Demand for Alcohol," Centre for Economic and Financial Research at New Economic School (CEFIR/NES) Working Paper 89.

Arcidiacono, Peter, Holger Sieg, and Frank Sloan. 2007. "Living Rationally under the Volcano? An Empirical Analysis of Heavy Drinking and Smoking." *International Economic Review* 48 (1): 37–65.

Auld, Christopher M., and Paul Grootendorst. 2004. "An empirical analysis of milk addiction." *Journal of Health Economics* 23 (6): 1117–33.

Bajari, Patrick, Victor Chernozhukov, Han Hong, and Denis Nekipelov. 2015. "Identification and Efficient Semiparametric Estimation of a Dynamic Discrete Game." National Bureau of Economic Research (NBER) Working Paper 21125.

Bajari, Patrick, Han Hong, John Krainer, and Denis Nekipelov. 2010. "Estimating Static Models of Strategic Interactions." *Journal of Business and Economic Statistics* 28 (4): 469–82.

Bajari, Patrick, Han Hong, and Denis Nekipelov. 2013. "Game Theory and Econometrics: A Survey of Some Recent Research." In *Advances in Economics and Econometrics: 10th World Congress, Econometrics*, Vol. 3, edited by Daron Acemoglu, Manuel Arellano, and Eddie Dekel, 3–52. Cambridge: Cambridge University Press.

Bajari, Patrick, Han Hong, and Stephen P. Ryan. 2010. "Identification and Estimation of a Discrete Game of Complete Information." *Econometrica* 78 (5): 1529–68.

Becker, Gary S., Michael Grossman, and Kevin M. Murphy. 1991. "Rational Addiction and the Effect of Price on Consumption." *American Economic Review* 81 (2): 237–41.

Becker, Gary S., and Kevin M. Murphy. 1988. "A Theory of Rational Addiction." *Journal of Political Economy* 96 (4): 675–700.

Berry, Steven, James Levinsohn, and Ariel Pakes. 1995. "Automobile Prices in Market Equilibrium." *Econometrica* 63 (4): 841–90.

Bhattacharya, Jay, Christina Gathmann, and Grant Miller. 2013. "The Gorbachev Anti-Alcohol Campaign and Russia's Mortality Crisis." *American Economic Journal: Applied Economics* 5 (2): 232–60.

Brainerd, Elizabeth, and David M. Cutler. 2005. "Autopsy on an Empire: Understanding Mortality in Russia and the Former Soviet Union." *Journal of Economic Perspectives* 19 (1): 107–30.

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

Card, David, David S. Lee, Zhuan Pei, and Andrea Weber. 2015. "Inference on Causal Effects in a Generalized Regression Kink Design." *Econometrica* 83 (6): 2453–83.

Chaloupka, Frank. 1991. "Rational Addictive Behavior and Cigarette Smoking." *Journal of Political Economy* 99 (4): 722–42.

Chenet, Laurent, Martin McKee, David Leon, Vladimir Shkolnikov, and Sergei Vassin. 1998. "Alcohol and cardiovascular mortality in Moscow; new evidence of a causal association." *Journal of Epidemiology and Community Health* 52 (12): 772–74.

Cook, Philip J., and Michael J. Moore. 2000. "Alcohol." In *Handbook of Health Economics*, Vol. 1B, edited by Anthony J. Culyer and Joseph P. Newhouse, 1629–73. Amsterdam: North-Holland.

Denisova, Irina. 2010. "Adult mortality in Russia." *Economics of Transition* 18 (2): 333–63.

Dreyfus, Mark K., and W. Kip Viscusi. 1995. "Rates of Time Preference and Consumer Valuations of Automobile Safety and Fuel Efficiency." *Journal of Law and Economics* 38 (1): 79–105.

Fang, Hanming, and Yang Wang. 2010. "Estimating Dynamic Discrete Choice Models with Hyperbolic Discounting, with an Application to Mammography Decisions." National Bureau of Economic Research (NBER) Working Paper 16438.

Glaeser, Edward L., Bruce I. Sacerdote, and Jose A. Scheinkman. 2003. "The Social Multiplier." *Journal of the European Economic Association* 1 (2–3): 345–53.

Gotz, Glenn A., and John J. McCall, 1984. *A Dynamic Retention Model for Air Force Officers: Theory and Estimates*. RAND. Santa Monica, CA, December.

Graham, Brian S. 2008. "Identifying Social Interactions through Conditional Variance Restrictions." *Econometrica* 76 (3): 643–60.

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

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

Hotz, V. Joseph, and Robert A. Miller. 1993. "Conditional Choice Probabilities and Estimation of Dynamic Models." *Review of Economic Studies* 60 (3): 497–529.

Imbens, Guido, and Karthik Kalyanaraman. 2012. "Optimal Bandwidth Choice for the Regression Discontinuity Estimator." *Review of Economic Studies* 79 (3): 933–59.

Kueng, Lorenz, and Evgeny Yakovlev. 2014. "How Persistent Are Consumption Habits? Micro-Evidence From Russia." National Bureau of Economic Research (NBER) Working Paper 20298.

Lee, David S., and Thomas Lemieux. 2010. "Regression Discontinuity Designs in Economics." *Journal of Economic Literature* 48 (2): 281–355.

Leon, David A., Lyudmila Saburova, Susannah Tomkins, Evgueny Andreev, Nikolay Kiryanov, Martin McKee, and Vladimir M. Shkolnikov. 2007. "Hazardous alcohol drinking and premature mortality in Russia: A population based case-control study." *Lancet* 369 (9578): 2001–09.

Leung, Siu Fai, and Charles E. Phelps. 1993. "'My Kingdom for a Drink...?': A Review of Estimates of the Price Sensitivity of Demand for Alcoholic Beverages." In *Economics and the Prevention of Alcohol-Related Problems*, edited by Michael E. Hilton and Gregory Bloss, 1–32. Rockville, MD: National Institute on Alcohol Abuse and Alcoholism.

Magnac, Thierry, and David Thesmar. 2002. "Identifying Dynamic Discrete Decision Processes." *Econometrica* 70 (2): 801–16.

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

Miller, Robert A. 1984. "Job Matching and Occupational Choice." *Journal of Political Economy* 92 (6): 1086–1120.

Moore, Michael J., and W. Kip Viscusi. 1990. "Models for estimating discount rates for long-term health risks using labor market data." *Journal of Risk and Uncertainty* 3 (4): 381–401.

Mullahy, John, and Jody L. Sindelar. 1993. "Alcoholism, Work, and Income." *Journal of Labor Economics* 11 (3): 494–520.

Nemtsov, A. V. 2002. "Alcohol-related human losses in Russia in the 1980s and 1990s." *Addiction* 97 (11): 1413–25.

Nevo, Aviv. 2011. "Empirical Models of Consumer Behavior." *Annual Review of Economics* 3: 51–75.

Pakes, Ariel. 1986. "Patents as Options: Some Estimates of the Value of Holding European Patent Stocks." *Econometrica* 54 (4): 755–84.

Pakes, Ariel, Michael Ostrovsky, and Steven Berry. 2007. "Simple estimators for the parameters of discrete dynamic games (with entry/exit examples)." *RAND Journal of Economics* 38 (2): 373–99.

Pesendorfer, Martin, and Philipp Schmidt-Dengler. 2008. "Asymptotic Least Squares Estimators for Dynamic Games." *Review of Economic Studies* 75 (3): 901–28.

Pleeter, Saul, and John T. Warner. 2001. "The Personal Discount Rate: Evidence from Military Downsizing Programs." *American Economic Review* 91 (1): 33–53.

Rust, John. 1987. "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher." *Econometrica* 55 (5): 999–1033.

Shkolnikov, Vladimir M., Evgeny M. Andreev, Martin McKee, and David A. Leon. 2013. "Components and possible determinants of the decrease in Russian mortality in 2004–2010." *Demographic Research* 28: 917–50.

Shkolnikov, Vladimir M., and France Meslé. 1996. "The Russian Epidemiological Crisis as Mirrored by Mortality Trends." In *Russia's Demographic* "*Crisis*," edited by Julie Davannzo and Gwendolyn Farnsworth, 113–61. Santa Monica, CA: RAND.

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

Treisman, Daniel. 2010. "Death and prices: The political economy of Russia's alcohol crisis." *Economics of Transition* 18 (2): 281–331.

Wagenaar, Alexander C., Matthew J. Salois, and Kelli A. Komro. 2009. "Effects of beverage alcohol price and tax levels on drinking: A meta-analysis of 1,003 estimates from 112 studies." *Addiction* 104 (2): 179–90.

Wolpin, Kenneth I. 1984. "An Estimable Dynamic Stochastic Model of Fertility and Child Mortality." *Journal of Political Economy* 92 (5): 852–74.

World Health Organization (WHO). 2011. *Global status report on alcohol and health*. Geneva: World Health Organization.

Yakovlev, Evgeny. 2008. "The Political Economy of Regulation: Evidence from the Russian Alcohol Industry." Unpublished.

Yakovlev, Evgeny. 2018. "Demand for Alcohol Consumption in Russia and Its Implication for Mortality: Dataset." *American Economic Journal: Applied Economics*. https://doi.org/10.1257/app.20130170.

Zaridze, David, Paul Brennan, Jillian Boreham, Alex Boroda, Rostislav Karpov, Alexander Lazarev, Irina Konobeevskaya, et al. 2009. "Alcohol and cause-specific mortality in Russia: A retrospective case-control study of 48, 557 adult deaths." *Lancet* 373 (9682): 2201–14.