Drinking is different! Examining the role of locus of control for alcohol consumption

Marco Caliendo\(^1,2,3,4\) · Juliane Hennecke\(^2,5,6\)

Received: 28 May 2021 / Accepted: 19 January 2022 / Published online: 10 March 2022
© The Author(s) 2022

Abstract
Locus of control (LOC) measures how much an individual believes in the causal relationship between her own actions and her life’s outcomes. While earlier literature has shown that an increasing internal LOC is associated with increased health-conscious behavior in domains such as smoking, exercise or diets, we find that drinking seems to be different. Using very informative German panel data, we extend and generalize previous findings and find a significant positive association between having an internal LOC and the probability of occasional and regular drinking for men and women. An increase in an individual’s LOC by one standard deviation increases the probability of occasional or regular drinking on average by 3.4% for men and 6.9% for women. Using a decomposition method, we show that roughly a quarter of this association can be explained by differences in the social activities between internal and external individuals.

Keywords Locus of control · Alcohol consumption · Health behavior · Risk perception · social activity

JEL Codes I12 · D91

The authors are grateful to Deborah Cobb-Clark, Daniel Graeber, Jan Marcus, Ronnie Schöb, Nicolas Ziebarth, participants at the ESPE 2019 Bath and research seminars at Victoria University of Wellington and University of Otago as well as two anonymous referees for helpful comments.

Juliane Hennecke
juliane.hennecke@ovgu.de
Marco Caliendo
caliendo@uni-potsdam.de

\(^1\) University of Potsdam, Potsdam, Germany
\(^2\) IZA, Bonn, Germany
\(^3\) DIW, Berlin, Germany
\(^4\) IAB, Nuremberg, Germany
\(^5\) Otto von Guericke University Magdeburg, Universitätsplatz 2, 39106 Magdeburg, Germany
\(^6\) NZWRI, Auckland, New Zealand
1 Introduction

The personality trait locus of control (LOC) can be characterized as the “generalized attitude, belief, or expectancy regarding the nature of the causal relationship between one’s own behavior and its consequences” (Rotter 1966) and describes whether individuals believe in the effects of their own actions on their life’s future outcomes. While an individual with an internal LOC believes that she is in control of the consequences of her own actions, an external individual attributes her life’s outcomes to luck, chance, fate or other external forces. LOC has already been shown to have an important effect on behavior and decision-making in areas such as human capital investment (Coleman and DeLeire 2003; Caliendo et al. 2020), job search effort (Caliendo et al. 2015; McGee and McGee 2016), labor force participation (Heckman et al. 2006; Hennecke 2020), savings (Cobb-Clark et al. 2016), labor market mobility (Caliendo et al. 2019) as well as investment behavior (Salamanca et al. 2016; Pinger et al. 2018), and it has also been found to have an insuring effect against negative effects of adverse economic and health-related life events on well-being and mental health (Buddelmeyer and Powdthavee 2016; Schurer 2017; Cuesta and Budría 2015).

Additionally, Cobb-Clark et al. (2014) have shown that LOC is positively linked to health-conscious behavior, such as abstaining from smoking, healthy diets and regular exercise, but also excessive alcohol consumption, i.e., binge drinking, raising the question whether “Drinking is Different?”. The existing medical and social science literature already provides considerable evidence for that. One notable distinction between alcohol consumption and other forms of unhealthy behavior is the important differentiation between different levels of consumption. While hazardous and excessive drinking are consistently found to be associated with negative consequences on mental and physical health (see, e.g., Chatterji et al. 2004; Marcus and Siedler 2015; Cotti et al. 2014; Grønbæk 2009; Corrao et al. 2004) as well as social and economic outcomes (see, e.g., Francesconi and James 2019; Jones and Richmond 2006; Macdonald and Shields 2004; Mangiavacchi and Piccoli 2018), studies often link moderate and responsible drinking with beneficial medical outcomes (Sayed and French 2016; Grønbæk 2009; Ronksley et al. 2011) as well as higher earnings and employment probabilities (Ziebarth and Grabka 2009; Peters and Stringham 2006; Ours 2004) and better social networks and social integration (Leifman et al. 1995; Buonanno and Vanin 2013). According to data from the World Health Organization 71.7% of US-Americans and 79.4% of Germans over the age of 15 consumed alcohol in 2016, while 36.4% (Germany) and 43.1% (USA) of these drinkers engaged in a heavy drinking episode in the past 30 days (World Health Organization 2018). Based on these numbers and the existing findings about distinct differences in the consequences of light and heavy drinking, an empirical investigation of intrinsic drivers of different levels of alcohol consumption is of high importance in order to enable policymakers to challenge the unwanted costs resulting from it. The economic burden of excessive alcohol consumption in the USA has been estimated at $249 billion in 2010 (Sacks et al. 2015).
We contribute to the literature by investigating the role of LOC for alcohol consumption. The existing literature on the association between LOC and alcohol consumption has mainly focused on excessive drinking and is rather inconclusive due to heterogeneous and selective samples and different ways of defining their drinking variable. For example, Steptoe and Wardler (2001) found that the perception of high, health-related external control is associated with a higher probability of frequent alcohol consumption in a sample of European university students, while Mendolia and Walker (2014) analyzed the effect of LOC and self-esteem on health-related behavior for a group of adolescents aged 15–16 years and find a weak, positive link of having an external perception of control with the frequency of getting drunk when drinking, but no significant association with regular drinking. Lassi et al. (2019) find an association between external LOC and hazardous drinking for samples of teenagers in the UK, and Chiteji (2010) analyzes the effect of self-efficacy (which is strongly linked to LOC) on drinking and exercising, and finds a negative association with drinking using the 1972-sample of male household heads in the Panel Study of Income Dynamics (PSID). Using data from the Household, Income and Labour Dynamics in Australia (HILDA) survey, Cobb-Clark et al. (2014) rely on a self-efficacy scale as a proxy for LOC and identify significantly positive effects of LOC on binge drinking.

We extend and generalize previous analyses in several respects: First, we consider different levels of drinking and thus paint a more comprehensive picture of the relationship between LOC and alcohol consumption. Second, we use a representative adult sample for which we can credibly use LOC as an exogenous covariate. Third, we use an extensive list of control variables such as socio-economic information, health status and other personality traits and, via this, substantially reduce the risk of omitted variable bias. Lastly, we use a general, not domain-specific LOC measurement which is closely linked to the original measure developed by Rotter (1966).

Our estimations are based on extensive information available in the Socio-Economic Panel (SOEP 2017), a large representative household panel from Germany. As opposed to the analysis of Cobb-Clark et al. (2014), our outcome variables are defined in a way which considers more common levels of alcohol consumption and focuses more on frequency rather than amounts of drinking. We thus look into whether individuals drink either occasional or regularly, which does include habitual drinking but is different from excessive drinking such as binge drinking. We find that an internal LOC is associated with a higher probability of reporting occasional or regular drinking even if we control for socio-economic information, health status and other personality and preference measures. Men with a medium or high internal LOC are, on average, about 3.1%–3.7% more likely to be occasional or regular drinkers compared with men with a low internal LOC. Women with a high internal LOC are 6.5% more likely to be moderate drinkers compared with women in the lowest LOC category. As a theoretical explanation for this relationship, we propose that the future risks of alcohol consumption might be underestimated if individuals believe in their own ability to cope with or prevent the negative consequences of unhealthy behavior.

Based on these findings, our paper makes another significant contribution to the literature by discussing social activity as a potential indirect mechanism behind the association. We propose that LOC is especially likely to be highly predictive of individual investment in social networks and thus drinking opportunities, given that attending
social gatherings is often inextricably linked with alcohol consumption. We estimate the extent of this indirect effect by including self-reported social activities as mediators into our model and decomposing the estimated coefficients. We find that roughly a quarter of the positive association between LOC and drinking can be explained by different levels of social activity for men and women. These findings are highly policy relevant as increased drinking due to an increase in social interactions is likely to be linked to more moderate levels of alcohol consumption with distinctly different economic and medical consequences as opposed to increased drinking behavior due to a mis-estimation of risks.

2 Data and empirical approach

Building upon the existing literature, we estimate the relationship between an internal LOC and self-reported alcohol consumption. The estimations are conducted using the extensive information available from the Socio-Economic Panel (SOEP 2017), a large representative longitudinal household panel from Germany (see Goebel et al. 2019, for more information). The SOEP includes detailed socio-economic information and surveys individuals’ LOC as well as their health behavior—including alcohol consumption—on a regular basis. While this is also true for other international surveys such as HILDA or NLSY79, the SOEP is the only data source that also enables us to observe important endogenous variables—such as risk and time preferences—as well as social interactions of individuals on a regular basis. This enables us to paint a more detailed picture about potential channels behind the estimated relationship. We restrict our sample to all observations for individuals between the age of 20 to 70 years for the 2006, 2008 and 2010 waves, within which we observe the self-assessed and reported amount of alcohol consumption. The sample is further reduced by item non-response in the LOC and other explanatory variables. Table A1 gives an overview over the sample restriction steps and observation loss due to item non-response. The final estimation sample comprises 33,765 observations for 14,841 individuals. Of these, 7496 individuals are observed three times, while 3413 and 3932 are observed once and twice, respectively. The later estimations will always be reported separately for men (48% of the sample) and women (52%) to take care of important, gender-specific heterogeneity, as is common in personality and health literature (see, e.g., Cobb-Clark et al. 2014). Table A2 provides an overview of the main summary statistics for the sample.

2.1 Locus of control

For our sample, LOC is measured in 2005 and 2010, in which years SOEP respondents were asked how closely a series of ten statements (items) characterized their
Table 1 Components of Locus of Control in SOEP

| No. | Item                                                                 | All   | Men   | Women  |
|-----|----------------------------------------------------------------------|-------|-------|--------|
|     | The following statements apply to different attitudes toward life and the future. To what degree do you personally agree with the following statements? |       |       |        |
|     | Scale: 1 (disagree completely)–7 (agree completely)                   |       |       |        |
| I1  | How my life goes depends on me                                       | 5.45  | 5.47  | 5.42***|
| I2  | Compared to other people, I have not achieved what I deserve (−)      | 3.26  | 3.34  | 3.19***|
| I3  | What a person achieves is above all a question of fate or luck (−)    | 3.52  | 3.43  | 3.61***|
| I4  | If a person is soc. active, she can have an effect on soc. conditionsa | 3.60  | 3.61  | 3.59   |
| I5  | Other people have a controlling influence over my life (−)            | 3.13  | 3.15  | 3.10***|
| I6  | One has to work hard in order to succeed                             | 6.01  | 6.02  | 6.00   |
| I7  | If I run up against difficulties in life, I doubt my own abilities (−) | 3.29  | 3.02  | 3.54***|
| I8  | Opportunities in life are determined by social conditions (−)         | 4.52  | 4.43  | 4.59***|
| I9  | Inborn abilities are more important than any efforts one can makea    | 4.86  | 4.89  | 4.84***|
| I10 | I have little control over the things that happen in my life (−)      | 2.66  | 2.64  | 2.67   |
|     | Observations                                                         | 33,765| 16,300| 17,465 |

Source SOEP, 2005 and 2010 waves for LOC input into waves 2006, 2008 and 2010 representing the estimation sample, version 33, https://doi.org/10.5684/soep.v33
Significance stars refer to the significance level of a t-test for mean equivalence between men and woman:
* p < 0.1, ** p < 0.05, *** p < 0.01. Items marked with a (−) are reversed prior to factor analysis.
a Items 4 and 9 are not included in the analysis

views about the extent to which they influence what happens in life. Responses were measured on a seven-point Likert scale ranging from 1 (‘disagree completely’) to 7 (‘agree completely’). A list of the set of items used, as well as the means of the observed responses in the full sample and separated by gender, can be found in Table 1.

As a first step in constructing our LOC variable, we conduct an exploratory factor analysis in which we investigate the way in which these items load onto latent factors. The factor analysis reveals that items 1 and 6 have a negative loading and items 2, 3, 5, 7, 8 and 10 have a positive loading onto a first factor. The factor’s eigenvalue is 1.84. A second factor has an eigenvalue of only 0.54 and can be neglected. Item 4 does not clearly load onto the first factor and item 9 has an unintuitive attribution, such that we exclude both in line with the earlier literature.

Subsequently, we use a two-step process to create a continuous, unidimensional LOC factor variable, consistent with previous literature (see, e.g., Piatek and Pinger 2016). Based on the exploratory factor analysis, we first reverse the scores for the external items (items 2, 3, 5, 7, 8 and 10) such that all eight items are increasing in internality. Secondly, we use confirmatory factor analysis to extract a single factor for each year separately. 4 This has the advantage of avoiding equal weighting of all items and instead relies on the data to determine how each item is weighted in the overall

---

4 For the estimation of the factor loadings within the confirmatory factor analysis, the information from all available LOC observation years is used simultaneously to minimize the risk of temporary measurement error issues affecting the factor loadings. Thus, the factor loadings are constant over time but the item values and the LOC factor are still time variant.
As per Piatek and Pinger (2016), simply averaging the items risks measurement error and attenuation bias. The resulting factor is therefore increasing in internal LOC, and its distribution is shown in Fig. 1.

Additionally, Fig. 1 reports the kernel densities of the LOC factor separately for men and women. It can be seen from both the distribution in Fig. 1 as well as most of the items in Table 1 that men are more internal than women. We account for these gender differences in our empirical analysis by using fully separated estimation models and standardizing the continuous LOC factor as well as generating dichotomous indicators separately within both sub-samples.

Due to the lack in overlap of observations waves of LOC and alcohol consumption, the information on LOC is imputed forward into the years in which we observe alcohol consumption, i.e., the LOC from 2005 is used as the explanatory variable for alcohol consumption in 2006 and 2008 and LOC from 2010 is used for consumption in 2010.

Exogeneity of LOC The stability of LOC during adulthood, and thus the exogeneity of the trait in models of decision making, have been heavily discussed in the psychological and economic literature. Although psychological literature exists, which finds variation in LOC during adulthood (Nowicki et al. 2018; Specht et al. 2013), the economic literature widely concluded that, after controlling for age, observed changes are not large enough to be economically relevant (Cobb-Clark and Schurer 2013) or largely unsystematic, driven by situation-specific and temporary measurement inaccuracy in reporting (Preuss and Hennecke 2018). Preuss and Hennecke (2018) find that the reported LOC does change after an exogenous labor market shock but that this observed change is solely driven by the labor force status during the second interview and no permanent changes are observable if individuals are reemployed. Endogeneity issues caused by omitted variable bias or non-random variation in the reporting of LOC are thus less likely if a large set of characteristics of, e.g., labor force or family...
status are included into the models. Given that we observe LOC only once for 34% of the sample and keeping in mind that variation for the remaining part is likely to arise from temporary variation in reporting, we will not be able to use a fixed-effects framework later on. Nevertheless, besides including a very rich set of control variables into the model, we additionally make sure to reduce the risk of endogeneity to a minimum in two steps: first, we tackle potential reverse causality. Potential endogeneity concerns caused by reverse causality are likely to apply to excessive consumption only and such consumption occurs seldomly in our sample. Nevertheless, in order to minimize the risks in this respect, we ensure that the LOC factor is never measured after the period in which we measure alcohol consumption by the method of forward imputation described above. Second, this would not solve the issue of non-random short-term differences in reporting as described by Preuss and Hennecke (2018), if it has behavioral implications. Thus, we additionally test the robustness of our results to an alternative specification of the LOC indicator in Sect. 3.3. Instead of a forward imputation, we average LOC over all available observation periods and thus generate a LOC measure which is less likely to be affected by temporary volatility in reporting in certain periods. The average LOC is assumed to draw a more consistent picture of the long-term latent trait as it wipes out all within-variation in LOC.

### 2.2 Alcohol consumption

In 2006, 2008 and 2010, individuals were asked to rate their consumption of four different types of alcoholic beverages (beer, wine, spirits, and mixed drinks) on a scale from 1 (regularly) to 4 (never). Based on a combination of all those answers and guided by the work of Ziebarth and Grabka (2009), we generate an ordinal measure of alcohol consumption. The variable categorizes individuals into the following four groups:

1. **Abstainers** No consumption of all four types,
2. **Rare Drinkers** Seldom drinking of at least one type, no occasional drinking,
3. **Occasional Drinkers** Occasional drinking of at least one type, no regular drinking,
4. **Regular Drinkers** Regular drinking of at least one type.

Table 2 provides an overview of the shares of alcohol consumption in the sample. In the full sample, 12% can be characterized as abstainers (no alcohol consumption at all) and 60% of the individuals are counted as being occasional (42%) or regular (18%) drinkers. In line with expectations, the share of drinkers is distinctly lower for women (40% occasional drinkers and 9% regular drinkers) than for men (44% occasional drinkers and 27% regular drinkers). 15% of all women are abstainers compared with 9% of men.

---

5 The main drawback of this measurement is the rather vague and subjective character, as no concrete information about the exact quantity of alcohol consumption is collected. We conduct a sensitivity check to test our measures and results in this respect and show the robustness of our results against the use of a more objective measure of alcohol consumption (see Sect. 3.3 for more detail).

6 Of the regular drinkers, observed in the data, only 5% report that they drink either spirits or mixed drinks regularly, which corresponds to less than 1% of all individuals in the data and does not allow us to draw any empirical conclusions on excessive drinking from the data at hand. The presented results do not change if these “excessive drinkers” are dropped from the sample.

---


d Springer
Table 2  Summary statistics and descriptive analysis—alcohol consumption

|                | All | Men | Women |
|----------------|-----|-----|-------|
|                | All | External | Internal | All | External | Internal |
| Alcohol consumption |     |         |         |     |         |         |
| Abstainers      | 0.12 | 0.08 | 0.11 | 0.06*** | 0.15 | 0.18 | 0.12*** |
| Rare drinkers   | 0.29 | 0.21 | 0.22 | 0.19*** | 0.36 | 0.37 | 0.34*** |
| Occasional drinkers | 0.42 | 0.44 | 0.41 | 0.46*** | 0.40 | 0.37 | 0.43*** |
| Regular drinkers | 0.18 | 0.27 | 0.26 | 0.29*** | 0.09 | 0.08 | 0.11*** |
| Observations    | 33,765 | 16,300 | 17,465 |     |         |         |

Source: SOEP, 2006, 2008, 2010 waves, version 33, https://doi.org/10.5684/soep.v33, own calculations
Individuals are grouped into internals and externals based on whether their LOC is lower/equal (external) or higher (internal) than the median. Significance stars refer to the significance level of a t-test for mean equivalence between externals and internals: *p < 0.1, **p < 0.05, ***p < 0.01

Additionally, Table 2 summarizes the results of a first descriptive analysis of the relationship between LOC and alcohol consumption. The results of the t-tests for mean equality indicate that for both men and women, the share of individuals who indicate that they are occasional or regular drinkers is significantly higher in the group of internal individuals (individuals with a LOC larger than the sample median). The share of occasional drinkers in the internal men category is 46%, while the share in external men is 41%. Internal men are also more likely to be regular drinkers (29% as opposed to 26%) and less likely to be abstainers (6% as opposed to 11%). All differences hold similarly for women.

2.3 Estimation strategy

Based on the available data, the obvious modeling choice would be to estimate an ordered response model. However, this model is based on the proportional odds assumption, which can be easily tested with a Brant (1990) test. The statistics of the Brant test for parallel regressions indicate a strong violation of the proportional odds assumption in the full model for men and women in our case, such that we refrain from using an ordered response model. Instead, we estimate four separate binary choice models based on the four drinking indicators $D_j$ with $j = \{1, 2, 3, 4\}$ summarized in Table 3.

As our main indicator—$D_{i1}$—we estimate the average marginal effects of an individual’s LOC on her probability of being an occasional or regular drinker as opposed to be an abstainer or rare drinker in Sect. 3.1. The choice of this indicator as our main explanatory variable is based on the assumption that rare drinkers are very similar to abstainers in their decision making while the same holds true for occasional and regular drinkers. This assumption is discussed and empirically tested in Sect. 3.2. In

---

7 We conduct the Brant test as an omnibus test for the entire model, and separately for each of the independent variables. The test statistics indicate a strong violation of the proportional odds assumption in the full model for men and women, as well as for the LOC factor for men. The results are available in Table S.1 in the supplementary online material.
Drinking is different! Examining the role of ...

Table 3  Outcome Variables in the Estimation Models

| Drinking indicator | Never | Rare | Occasional | Regular |
|--------------------|-------|------|------------|---------|
| **Main indicator (Sect. 4.1)** |       |      |            |         |
| $D_1$ Occasional + Regular | $D_1 = 0$ | $D_1 = 0$ | $D_1 = 1$ | $D_1 = 1$ |
| **Supplementary indicators (Sect. 4.2)** |       |      |            |         |
| $D_2$ Extensive: Any | $D_2 = 0$ | $D_2 = 1$ | $D_2 = 1$ | $D_2 = 1$ |
| $D_3$ Intensive I: Occasional + Regular | Missing | $D_3 = 0$ | $D_3 = 1$ | $D_3 = 1$ |
| $D_4$ Intensive II: Regular | Missing | Missing | $D_4 = 0$ | $D_4 = 1$ |

a further step, we take a closer look at potential differences along the extensive and intensive margin by also estimating the relationship between LOC and the probability of drinking any alcohol ($D_{2it}$), i.e., the extensive margins, as well as the probability of being an occasional or regular drinker as opposed to being a rare drinker ($D_{3it}$) and a regular drinker as opposed to being an occasional drinker ($D_{4it}$), which serve as two different cutoffs at the intensive margin. Nevertheless, it should be noted that the estimated relationships using the latter two outcome variables are at risk of sample selection bias. Therefore, the estimated results should be interpreted with care and only serve as ancillary evidence, additional to the main indicator.

All estimations are conducted using the following estimation equation:

$$P(D_{jit} = 1) = P(\beta_1 + \beta_2 loc_{it} + \beta_3 I_{it} + \beta_4 D_{it} + \beta_5 P_{i} + \beta_6 E_{it} + \beta_7 LM_{it} + \beta_8 H_{it} + \epsilon_{it} > 0),$$

where $D_{jit}$ is one of the four indicators for alcohol consumption $j = \{1, 2, 3, 4\}$ of individual $i$ at time $t$ and $loc_{it}$ is the (imputed) LOC of individual $i$ in $t$. Each model pools observations from the 2006, 2008, and 2010 waves and contains a list of interview characteristics $I_{it}$, such as the interview mode, interview year, interview month as well as the day of the week. Additionally, we include an extensive list of exogenous control variables, such as demographic characteristics $D_{it}$ (age, nationality, region of residence, number of children in the household, an indicator for expecting parents as well as young children (aged 0–1 and 1–7) in the household, family status, and religious affiliation) as well as standardized personality and preferences measures averaged over all available time periods $P_{i}$ (i.e., the Big Five personality traits, general and health-related risk aversion as well as patience and impulsiveness as a proxy for individual time preferences). In order to further investigate the extent of the indirect link between LOC and alcohol consumption, we additionally include a list of potentially endogenous variables into the model. The inclusion of these variables likely moderates the effect of locus of control. These variables include educational controls $E_{it}$ (school degree, vocational degree and university degree), current individual labor market controls $LM_{it}$ (net household income, gross labor income, labor force status and occupational autonomy) and the individual health status and behavior $H_{it}$ (indicator for officially assessed, severe disability or working incapability, subjective health, body mass index...
and mental and physical health scores as well as the frequency of smoking, healthy diets and exercise). See Table A2 for the full list of control variables.

Equation (1) is estimated using binary logit models, and average marginal effects are reported in the following. LOC is always standardized and categorized within the sub-samples such that, e.g., having a high LOC corresponds to a high LOC compared to all other individuals within the selected sub-sample (e.g., only drinkers in the analysis at the intensive margin) and within the same gender. Standard errors are clustered at the personal level to account for serial correlation in the error terms, which occurs due to the panel nature of the data.

3 Results

3.1 Main indicator

Table 4 summarizes the average marginal effects based on the logit model using the binary indicator of occasional or regular drinking \( D_1 \) as the dependent variable. Column (1) shows the descriptive raw difference. Consistent with the descriptive differences in Table 2, the estimates show a positive raw gap in the probability of occasional and regular drinking between internals and externals. The more internal that an individual is, the higher the probability of reporting at least occasional drinking. We can see that this descriptive raw difference between internals and externals becomes smaller but remains significantly positive when we include additional sets of exogenous control variables in column (2). An increase in an individual’s LOC by one standard deviation increases the probability of occasional or regular drinking on average by 2.4 percentage points for men and by 3.4 percentage points for women, holding all other variables constant. This corresponds to a relative effect of 3.4% for men and 6.9% for women based on the sample means of 71% and 49%, respectively (see Table 2). In order to get a more conservative estimate of the association between LOC and alcohol consumption which considers the role of possible mediators, we include a long list of potentially endogenous control variables in columns (3)–(5). The inclusion of educational controls as well as health behavior has a particularly strong effect on the estimated relationship between LOC and drinking for women, explaining close to two-thirds of the estimated association in column (2). For men all three groups of variables (education, labor market and health) have a similarly large effect on the estimated relationship and explain roughly half of it. Column (5) contains the results for the full model, which we will use as our main specification and to which we will refer in the following. An increase in an individual’s LOC by one standard deviation on average increases the probability of occasional or regular drinking by 1.1 percentage points for men and by 0.9 percentage points for women, holding all other variables constant. This corresponds to a relative effect of 1.5% for men and 1.8% for women. While we observe a substantial decrease in the estimated effect on \( D_1 \) from column (2) to column (5), it is important to note that our data contain a very rich set of control

---

8 For more details on the used health scores, see Andersen et al. (2007).

9 The full estimation results for the specification in column (5) can be found in Table S.2.
Table 4  Main results (logit, marginal effects)—main drinking variable

| Outcome: occasional or regular drinking $D_{t}$ | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
|-----------------------------------------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Men LOC Factor (cont.)                        | 0.036*** | 0.024*** | 0.019*** | 0.016*** | 0.011** | (0.004) | (0.005) | (0.005) | (0.005) | (0.005) |
| Locus of control terciles (Ref.: [$LOC_{min}, LOC_{P33}$]) | | | | | | | | | | |
| ($LOC_{P33}, LOC_{P66}$) | 0.070*** | 0.050*** | 0.041*** | 0.036*** | 0.026** | (0.011) | (0.011) | (0.010) | (0.010) | (0.010) |
| ($LOC_{P66}, LOC_{max}$) | 0.078*** | 0.051*** | 0.039*** | 0.033*** | 0.022*** | (0.011) | (0.012) | (0.012) | (0.012) | (0.012) |
| Pseudo $R^2$ | 0.005 | 0.033 | 0.040 | 0.045 | 0.059 | 0.005 | 0.033 | 0.040 | 0.045 | 0.059 |
| Observations | 16,300 | 16,300 | 16,300 | 16,300 | 16,300 | 16,300 | 16,300 | 16,300 | 16,300 | 16,300 |
| Women LOC Factor (cont.)                       | 0.051*** | 0.034*** | 0.024*** | 0.019*** | 0.009* | (0.005) | (0.005) | (0.005) | (0.005) | (0.005) |
| Locus of control terciles (Ref.: [$LOC_{min}, LOC_{P33}$]) | | | | | | | | | | |
| ($LOC_{P33}, LOC_{P66}$) | 0.070*** | 0.044*** | 0.033*** | 0.026** | 0.012 | (0.011) | (0.011) | (0.011) | (0.011) | (0.011) |
| ($LOC_{P66}, LOC_{max}$) | 0.121*** | 0.085*** | 0.064*** | 0.052*** | 0.032*** | (0.012) | (0.012) | (0.012) | (0.012) | (0.012) |
| Pseudo $R^2$ | 0.007 | 0.048 | 0.059 | 0.066 | 0.085 | 0.007 | 0.048 | 0.059 | 0.066 | 0.086 |
| Observations | 17,465 | 17,465 | 17,465 | 17,465 | 17,465 | 17,465 | 17,465 | 17,465 | 17,465 | 17,465 |
| Interview | ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ *
| Demographics | ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ *
| Personality | ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ *
| Education | ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ *
| Labor Market | ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ *
| Health | ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ *

Source: SOEP, 2006, 2008, 2010 waves, version 33, https://doi.org/10.5684/soep.v33, own calculations

This table displays the marginal effects estimated based on a logit estimation in line with the model in estimation Eq. (1) with the dependent variable being the binary indicator for occasional or regular drinking. Each column represents a different specification where control variables are included stepwise. Columns (1)–(5) use the continuous LOC factor as the explanatory variable and in columns (6)–(10), LOC is included as binary indicators for the terciles of LOC (with $LOC < LOC_{P33}$ being the reference group). Standard errors (in parentheses) are clustered on the individual level. $^* p < 0.1$, $^{**} p < 0.05$, $^{***} p < 0.01$

variables. The evolution of the estimated effect can be interpreted as evidence that the relationship between LOC and occasional/regular drinking is robust and likely not purely driven by observable mediators. We will further analyze the sensitivity of our results with respect to omitted variables following Oster (2019) in Sect. 3.3.

Although these effects appear to be rather small, concurrent with the low overall explained variability of alcohol consumption in the model (see pseudo $R^2$ in Table 4), they hold considerable economic relevance. The magnitude of the effect is of a similar size to the marginal effects of knowingly important preference measures such as the willingness to take risks and patience (as a proxy for time preferences), which can be found in Table S.2.

To identify potential non-linearities, we consider indicators for being in a different tercile of the LOC distribution as explanatory variables in columns (6)–(10). The
results show a similar picture to that of the continuous LOC measure. For men, having a medium LOC \( (LOC_{P33}, LOC_{P66}) \) on average increases the probability of occasional or regular consumption by 2.6 percentage points (3.7%) compared to having a low LOC \( (LOC_{min}, LOC_{P33}) \). Interestingly, having a high LOC \( (LOC_{P66}, LOC_{max}) \) increases men’s probability of occasional or regular consumption in a similar magnitude by 2.2 percentage points (3.1%). Thus, the association appears to be nonlinear for men, with the highest probability of drinking being found for men with a medium LOC. The overall picture is slightly different for women: Only having a high LOC increases a woman’s probability of occasional or regular consumption by 3.2 percentage points (6.5%), while the effect of a medium LOC is not significant when compared to having a low LOC.

3.2 Supplementary indicators: extensive and intensive margin

As the results from the Brant test already indicated, the effect of LOC is likely to differ between different intensities of alcohol consumption. To further investigate this, we devote some further attention to how the effect might differ at the extensive and intensive margin.

Table 5 summarizes the estimated average marginal effects of LOC using the three supplementary binary indicators \( D_j = \{2, 3, 4\} \) as dependent variables.\(^{10}\) In the first step, we re-estimate the model using an indicator for any drinking (extensive margin) in order to identify whether the estimation results for the main indicator have been driven by differences between abstainers and rare drinkers. In line with the main results, the effects in columns (1) and (2) are significantly positive, indicating a reduced probability of being an abstainer for internal men and women. The effect magnitudes are smaller than for the main outcome variable, indicating some important associations at the intensive margin, but are still of considerable magnitude.\(^{11}\) In the second step, we then analyze potential associations at the intensive margin in columns (3)–(6) of Table 5. If we abstract from potential sample selection bias caused by the restriction of the sample to non-abstainers, the results reveal that for men(women) having a medium(high) LOC increases the probability of occasional and regular drinking in the sub-sample of individuals who drink at least sometimes. The patterns are comparable to the ones estimated for the main drinking indicator in Sect. 3.1, but effect magnitudes are smaller and the effects of the continuous measure lose significance. In the last step, in columns (5) and (6), we do not find a significant association between LOC and differences between regular and occasional drinking for men as well as women. Based on these findings, we can assume that much of the association between LOC and drinking is driven by differences at the extensive margin as well as in parts at the cutoff between rare drinkers and occasional drinkers.

\(^{10}\) The full estimation results for the specifications in columns (1), (3) and (5) can be found in Table S.2.

\(^{11}\) An additional analysis revealed that no significant relationship between LOC and being a rare drinkers (as opposed to being an abstainer) can be found. The results are available in Table S.3.
Table 5  Main results (logit, marginal effects)—intensive vs. extensive margin

| Outcome | Extensive $D_2$ | Intensive I $D_3$ | Intensive II $D_4$ |
|---------|-----------------|-------------------|-------------------|
|         | Any (vs. Non)   | Occ./Reg. (vs. Rare) | Regular (vs. Occasional) |
|         | (1)             | (2)               | (3)               | (4)               | (5)               | (6)               |
| Men     | LOC Factor (cont.) | 0.007**           | 0.006            | −0.006            |
|         |                  | (0.003)           |                   |                   |
| Locus of control terciles (Ref.: $[LOC_{min}, LOC_{P33}]$) | | 0.017***          | 0.021**          | 0.016             |
|         |                  | (0.006)           | (0.010)          | (0.013)           |
|         | $[LOC_{P33}, LOC_{P66}]$ | 0.019***          | 0.014            | −0.009            |
|         |                  | (0.007)           | (0.011)          | (0.014)           |
| Locus of control terciles (Ref.: $[LOC_{min}, LOC_{P33}]$) | | 0.009             | 0.008            | 0.003             |
|         |                  | (0.007)           | (0.011)          | (0.012)           |
|         | $[LOC_{P66}, LOC_{max}]$ | 0.018**          | 0.026**          | −0.002            |
|         |                  | (0.008)           | (0.013)          | (0.013)           |
| Observations | 16,300   | 16,300           | 14,920           | 14,920           | 11,566        | 11,566        |
| Women   | LOC Factor (cont.) | 0.008**           | 0.005            | −0.003            |
|         |                  | (0.003)           |                   |                   |
| Locus of control terciles (Ref.: $[LOC_{min}, LOC_{P33}]$) | | 0.009             | 0.008            | 0.003             |
|         |                  | (0.007)           | (0.011)          | (0.012)           |
|         | $[LOC_{P66}, LOC_{max}]$ | 0.018**          | 0.026**          | −0.002            |
|         |                  | (0.008)           | (0.013)          | (0.013)           |
| Observations | 17,465   | 17,465           | 14,905           | 14,905           | 8621          | 8621          |
| Time-fixed effects | ✓          | ✓                | ✓                | ✓                | ✓              | ✓              |
| Demographics     | ✓          | ✓                | ✓                | ✓                | ✓              | ✓              |
| Education        | ✓          | ✓                | ✓                | ✓                | ✓              | ✓              |
| Labor market     | ✓          | ✓                | ✓                | ✓                | ✓              | ✓              |
| Personality      | ✓          | ✓                | ✓                | ✓                | ✓              | ✓              |
| Health           | ✓          | ✓                | ✓                | ✓                | ✓              | ✓              |

Source SOEP, 2006, 2008, 2010 waves, version 33, https://doi.org/10.5684/soep.v33, own calculations

This table displays the marginal effects estimated based on a logit estimation in line with the model in estimation Eq. (1) with the dependent variable being either one of the three alternative drinking indicators ($D_2$, $D_3$ or $D_4$). Columns (1), (3) and (5) use the continuous LOC factor as the explanatory variable and in columns (2), (4) and (6), LOC is included as binary indicators for the terciles of LOC (with LOC $< LOC_{P33}$ being the reference group). All models include the full set of control variables in line with columns (5) and (10) of Table 4. Standard errors (in parentheses) are clustered on the individual level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

3.3 Robustness checks

We test the robustness of our results with respect to both the definition of our main explanatory variable and our outcome variable as well as with respect to potentially omitted variables.

Explanatory variable In a first step, we check the robustness of our estimated effects with respect to the construction and imputation of the LOC measure. Thus, we
construct two alternative LOC measures and re-estimate our main model for these alternative explanatory variables. The results can be found in Table A3. First, we find that our estimated effects are relatively robust against the use of a simple average over all eight LOC items (panel A1), which assumes equal weights for each item on the latent factor. Only the coefficient of the medium LOC for men seems to be sensitive to this change in the definition of the LOC and thus might be more strongly associated with items with higher loadings (e.g., items 5 and 10). Secondly, we check whether our estimated effects are sensitive to the use of an averaged LOC imputed over all available observations, which wipes out all within-variation in LOC for those individuals whom we observe more than once. This adjustment is expected to reduce measurement inaccuracies in the situational measurement of LOC. The estimated effects presented in panel A2 are also robust and effect magnitudes are stronger than for the imputed version of LOC indicating potential attenuation bias due to noise in the measurement of LOC.

Outcome variable—objective amounts

The estimated results might also be biased by the subjective nature of the alcohol consumption variable used. As our main dependent variable is based on the self-assessed amount of consumption, it not only depends on the actual consumption level but also the individual’s perception of the terms ‘regular’ and ‘occasional’. If individuals perceive amounts differently based on their LOC, this would bias our results. We can test the reliability of our measure and the sensitivity of our results with respect to the subjectivity of the reported amounts using measures of concrete frequencies and amounts available in the SOEP 2016 wave. In 2016, individuals do not self-assess their consumption but report more distinct amounts and frequencies. An overview of the descriptive statistics for these variables can be found in Table A4.

The new dependent variables are generated based on the reported frequency of consumption and the reported consumption amount per consumption day. LOC is imputed from the 2015 wave.12 The results of this sensitivity check are reported in panel (B) of Table A3. First, the binary indicator for drinking is one if the individual reports drinking at two or more days per month (“moderate or high frequency”). This behavior is assumed to correspond most closely to “occasional or regular consumption” as per the baseline. The sensitivity check (panel B1) indicates that the results from the baseline are relatively robust with respect to the type of reporting. Although effects lose significance due to the extreme reduction in sample size, the effect sizes remain stable except for the effect of a high LOC for women.13 When we look at high consumption amount—defined as three or more drinks per day—(panel B2) for women we can see that LOC has no effect on the amount of drinks consumed per episode. However, for men, a medium as well as high LOC has a significant positive effect on consumption amounts, too.

Omitted variable bias

Despite our extraordinary rich set of controls—which include detailed socio-economic characteristics, health status and health behavior in other domains as well as a list of other personality traits and preference measures—we

---

12 The measurement of LOC in 2015 as well as the construction of the factor is equivalent to those in 2005 and 2010, which is described in detail in Sect. 2.

13 If we concentrate on the full sample of all individuals and thus a higher sample size, effects are stable in significance level and size for both medium and high LOC. The results are available in Table S.4.
cannot rule out that omitted variables bias our results. In order to address this issue, we (i) add a number of stressful events as additional control variables and (ii) use a bounding analysis.

First, in addition to our extensive list of control variables, we conduct a robustness check in Table A3 (panel C) in which we additionally add a list of potentially stressful events (job loss, marriage, residential moves, separation, death of a spouse and birth of a child), which might be at risk of affecting both alcohol consumption and LOC, as control variables. Reassuringly, controlling for these events does not affect the estimated effects of LOC.

Secondly, Oster (2019) provides a method of calculating consistent estimates of bias-adjusted treatment effects given assumptions about (i) the relative degree of selection on observed and unobserved variables \( \delta \), and (ii) the R-squared from a hypothetical regression of the outcome on the treatment and both observed and unobserved controls \( R_{max} \). \( \delta = 1 \) implies that observed and unobserved factors are equally important in explaining the outcome, while \( \delta > 1 \) (\( \delta < 1 \)) implies a larger (smaller) impact of unobserved than observed factors. Given the assumed bounds for \( \delta \) and \( R_{max} \), researchers can then calculate an identified set for the treatment effect of interest. If this set excludes zero, the results from the controlled regressions can be considered robust to omitted variable bias.

Consequently, we focus on our main result—the estimated effect of LOC on our main indicator \( D_1 \) (occasional/regular drinking vs none/rare)—and we re-estimate the results reported in Table 4 using OLS. Table A5 presents the results for the LOC terciles. Comparing Columns (1) and (2) in Table A5 reveals that for men the estimated effect of a medium (high) LOC on \( D_1 \) decreases from 0.067 (0.081) in a model with only interview controls to 0.028 (0.026) in our full specification which includes all sets of control variables. For women, the estimated effect decreases from 0.067 (0.115) to 0.013 (0.030). Guided by the rule of thumb provided in Oster (2019), the maximum \( R^2 \) is set to 1.3 times the \( R^2 \) in the fully-controlled model. Column (3) contains the identified set of coefficients at \( \delta = 1 \), i.e., a situation in which there are unobserved variables that have similarly explanatory power as our large set of explanatory variables. Subsequently, the identified set is \([0.014; 0.028]\) \((0.001; 0.026)\) for men and would still be positive even if we consider the full set of control variables including the potentially endogenous mediators. In fact, the identified set of coefficients only includes zero if \( \hat{\delta} \) exceeds 1.88 (1.05). The identified set for women is \([-0.007; 0.013]\) \((-0.007; 0.030)\) if the reduced baseline effect is compared to the controlled effects which include the potentially endogenous health-related variables and thus includes 0. In this case \( \hat{\delta} \) is 0.79 (0.82). This is driven by the strong effect of the health-related control variables for women. As has already been discussed above, these sets of variables are at risk of introducing endogeneity to the model. We thus re-estimate the selection test for a case in which we exclude them from the fully controlled model. If the health controls are excluded from the fully controlled model in the second panel, the identified set is \([0.013; 0.028]\) \((0.022; 0.051)\) and would only include zero if \( \hat{\delta} \)

---

14 Results for the continuous LOC measure can be obtained from Table S.5.
exceeds 1.81 (1.64). Overall, the robustness analysis is re-assuring and shows that the results are quite robust to potentially omitted variables.\textsuperscript{15}

**Sample restriction** In a set of robustness checks, we analyze the sensitivity of our estimation results with respect to the sample restriction steps as described in Sect. 2. First, in panel (D) of Table A3, we further restrict the age-range of our sample to working-age individuals (i.e., 25–64 years) as LOC is assumed to be more stable in this age period. Estimation results are robust against this sample restriction. Secondly, the robustness checks presented in panel (E) of Table A3 analyze the role of non-random item non-response in the very early sample restriction steps. We re-estimate the raw effect of LOC (without controls) for the unrestricted sample, which also includes individuals who drop out of our main estimation sample to missing information on any of the control variables. Estimation results are also robust with respect to this sample restriction step.

### 3.4 Discussion of results

The results from our empirical analysis stand in contrast to the existing findings on the effect of LOC on health-related behavior in other domains such as smoking, exercise and healthy diet in the previous literature. Health investment models such as in Grossman (1972, 2000) might, thus, not be applicable to the relationship between LOC and alcohol consumption. This doubt is prompted by the missing subjective link between current alcohol consumption and future health consequences. Bennett et al. (1998) state that alcohol consumption might be associated with higher levels of uncertainty about future outcomes as individuals do not see alcohol consumption in reasonable amounts as affecting their health too strongly. Although individual considerations about health investments are likely still at play, they might be on average dominated by other mechanisms in the analyzed population. Due to this uncertainty, we can assume that individual perceptions are highly important in those situations. Individuals must build their own expectations about the probabilities with which their behavior is associated with certain outcomes. In the present case, individuals estimate the likelihood with which their alcohol consumption entails negative future consequences for their health. For example, Sloan et al. (2013) find that heavy drinkers in the USA on average tend to overestimate their ability to handle alcohol while Lundborg and Lindgren (2002) find that young people in Sweden on average tend to overestimate the risks associated with drinking.\textsuperscript{16} In line with the definition of LOC, it is obvious to expect that an internal LOC entails lower levels of these risk perceptions. Multiple studies have already found that LOC has an important effect on individual perceptions about personal risk, e.g., with respect to, e.g., myocardial infarction and cancer (see, e.g., Stürmer et al. 2006; Källmén 2000; Sjoberg 2000). In line with this literature, Cobb-Clark et al. (2014) argue that an increased perception of control might

\textsuperscript{15} In line with the nonlinearity of effects discussed above, the estimates for the continuous measure (see Table S.5) are found to be more sensitive to the selection test. In this case, \( \delta \) is only > 1 if health controls are excluded from the fully controlled model.

\textsuperscript{16} As is shown in Ziebarth (2018) and Lundborg and Lindgren (2002), the accuracy of these estimations is affected by the available information such as the education about alcohol.
be correlated with a stronger belief about the ability to cope with and prevent the consequences of drinking. An increased perception of individual control might reduce the perceived importance of risk for life’s outcomes. The future risks of alcohol consumption might be underestimated if the individual control is overestimated (Slovic 1992). An increased alcohol consumption due to mis-estimated risk probabilities is a likely important explanation for the observed association above.

Potential additional explanations for an observed positive correlation between LOC and alcohol consumption include the role of being able to afford alcohol consumption, the relationship between LOC and alcohol consumption with behavior in other health domains (see e.g. Nguyen 2019), and the correlation between LOC, individual risk preferences and self-control problems as well as present-biased decision-making. However, all these possible explanations have been ruled out largely through the inclusion of earnings, household income, behavior in other health domains, willingness to take risks and patience and impulsiveness as proxies of individual time preferences in the main estimation model. Nevertheless, there is one potential factor that remains and that we will explore in the next subsection.

3.5 The (mediating) role of social activities

Based on the existing psychological literature on peer effects of alcohol consumption in adolescence (Lundborg 2006; Buonanno and Vanin 2013), a likely remaining mechanism of the association between LOC and drinking might be the link via differences in the importance of peer and networking effects. Alcohol consumption is associated with important positive effects on social networks. Drinking is common at social events and abstinence has been shown to be linked to strong negative penalties with respect to social integration (see, e.g., Leifman et al. 1995). For example, Peters and Stringham (2006) and Ziebarth and Grabka (2009) discuss the association between alcohol consumption and social networks as likely channels for their identified positive effect of alcohol consumption on earnings. As they notice, alcohol consumption remains a social norm in modern Western societies, which inevitably links drinking and the attendance of social events. Thus, moderate drinking produces social capital and can be labeled as a productive activity. In line with the argument about LOC and investment in future outcomes—which has been raised, for example, in Coleman and DeLeire (2003) and Caliendo et al. (2015)—internals are expected to invest more in social capital than externals, as they expect higher future returns from it such as a network of social support or professional contacts. This can easily be achieved by attending social gatherings and thus drinking. Hence, by default they might be more likely to drink alcohol in moderation. As opposed to excessive and uncontrolled alcohol consumption, drinking behavior that can be explained by this mechanism might be connected with less severe negative or even positive economic and medical consequences, which is why it is important to separate it from other potential explanations.

17 This argument is also largely in line with the latest literature on the association between LOC and risky investment decision which found that internal individuals are more likely to invest in risky assets and make inconsistent investment decisions (see, e.g., Salamanca et al. 2016; Pinger et al. 2018).
Table 6  Social activity determinants (OLS, outcome: # of activities conducted at least once a week)

|               | Men (1)   | Men (2)   | Women (3)  | Women (4)   |
|---------------|-----------|-----------|------------|------------|
| LOC Factor (std.) | 0.012**   | 0.016***  | 0.016***   | 0.024**    |
| (0.005)       | (0.005)   | (0.012)   | (0.012)    |            |
| Locus of control terciles (Ref.: [LOCmin, LOCp33]) |           |           |            |            |
| (LOCp33, LOCp66] | 0.013     | 0.024**   |            |            |
| (0.013)       | (0.012)   | (0.012)   |            |            |
| (LOCp66, LOCmax] | 0.028**   | 0.043***  |            |            |
| (0.013)       | (0.012)   | (0.012)   |            |            |
| Observations  | 15,988    | 15,988    | 17,122     | 17,122     |
| All controls  | ✔         | ✔         | ✔          | ✔          |

Source SOEP, 2006, 2008, 2010 waves, version 33, https://doi.org/10.5684/soep.v33, own calculations
This table displays the results of an OLS estimation with the dependent variable being the number of activities the individual conducts at least once a week. Columns (1) and (3) use the continuous LOC factor as the explanatory variable and in columns (2) and (4), LOC is included as binary indicators for the terciles of LOC (with LOC < LOCp33 being the reference group). All models include the full set of control variables in line with columns (5) and (10) of Table 4. Standard errors (in parentheses) are clustered on the individual level. *p < 0.1, **p < 0.05, ***p < 0.01

In order to check this hypothesis, we analyze whether LOC can be associated with higher levels of social activity using information on spare time activities available in the SOEP. We measure social activity with a set of ordinal variables which are based on the self-reported frequency of three social activities, namely “going out eating and drinking,” “attending social gatherings” and “visiting friends and neighbors.” For the first stage analysis of the relationship between LOC and these social activities, the activities are summarized into a continuous variable, which counts the number of activities which are conducted at least once per week. Table 6 gives the estimated effects of this analysis, estimated using a linear estimation model. As expected, LOC is associated with a higher likelihood of regular participation in these social activities.

Based on this, we investigate whether internals are simply more likely to be exposed to alcohol, as they are socially more active and outgoing, by considering social activities as a mediator in our model. For this purpose, we decompose the estimated relationship between LOC and drinking into a direct and an indirect effect via the full ordinal versions of all social activities analyzed above using the method proposed in Karlson and Holm (2011) and Breen et al. (2013) (KHB method). The KHB method allows for the comparison of estimated coefficients between two nested nonlinear probabilities models by accounting for the fact that coefficients and error variances in these models are not separately identified and coefficients, thus, cannot be directly compared.
Table 7  KHB decomposition with social activities as mediators (logit, outcome: occasional or regular drinker)

|                  | Men                     | Women                    |
|------------------|-------------------------|--------------------------|
|                  | Reduced b/(se)/[AME]    | Full b/(se)/[AME]        | Diff b/(se)/[AME]    | Reduced b/(se)/[AME]    | Full b/(se)/[AME]        | Diff b/(se)/[AME]   |
| LOC Factor (std.)| (1) 0.059** 0.040       | (2) 0.018*** 0.004       | (3) 0.048** 0.030    | (4) 0.018*** 0.004     | (5) 0.030              | (6) 0.018***    |
|                  | (0.025) (0.026)         | (0.004)                  | (0.023) (0.023)     | (0.004)                  |
| Locus of control terciles (Ref.: $LOC_{min}$, $LOC_{P33}$) |                  |                          |                        |                         |                        |                        |
| $LOC_{P33}$      | (1) 0.126** 0.096*      | (2) 0.030**              | (3) 0.050 0.018      | (4) 0.033*** (0.011)    | (5) 0.049 (0.049)       | (6) 0.011       |
|                  | (0.054) (0.054)         | (0.011)                  | (0.049) (0.049)     | (0.011)                  |
| $LOC_{P66}$      | [0.126] [0.096]         | [0.011] [0.004]          |                          |                         |                        |                        |
| $LOC_{max}$      | (1) 0.117* 0.081        | (2) 0.036***             | (3) 0.155*** 0.119** | (4) 0.036*** (0.011)    | (5) 0.055 (0.055)       | (6) 0.011       |
|                  | (0.061) (0.061)         | (0.012)                  | (0.049) (0.049)     | (0.011)                  |
| Observations     | 15,988                  | 17,122                   |                          |                         |                        |                        |
| Social Activities| ✓ ✓                     | ✗ ✓                      |                          |                         |                        |                        |
| All Controls     | ✓ ✓                     | ✓ ✓                      |                          |                         |                        |                        |

Source: SOEP 2006, 2008, 2010 waves, version 33. https://doi.org/10.5684/soep.v33, own calculations

This table displays the results of a logit estimation (as well as average marginal effects (AME) in square brackets) with the dependent variable being the binary indicator for occasional or regular drinking. Columns (1) and (4) show the reduced model which represents the specification as presented in Eq. (1) with the only difference being that residuals of the social activity variables are included as right-hand side variables. Columns (2) and (6) show the full model in which the social activity variables are included into the model as control variables. The upper and lower panels of the table represent distinct estimations in which either the continuous LOC factor or the binary indicators for the terciles of LOC (with LOC < $LOC_{P33}$ being the reference group) are included as explanatory variables. Standard errors (in parentheses) are clustered on the individual level. *p < 0.1, **p < 0.05, ***p < 0.01

between the reduced and the full model.19 It does so by augmenting the reduced model with the residuals from a regression of the mediator variable (i.e., social activity) on the key explanatory variable (i.e., LOC) and thus allows for a separation of the difference due to mediation and difference due to a rescaling with different error variances.

The results of this decomposition are reported in Table 7. The decomposition is based on the coefficients of the nonlinear estimation model and average partial effects are computed for both models and reported in square brackets. The results indicate that for both men and women, a significant share of the effect can be contributed to differences in social activities between internal and external individuals. They explain about 23.8% (30.8%) of the association between a medium (high) LOC and drinking for men and 23.2% of association between a high LOC and drinking for women. The overall effect of a high LOC drops from 2.2 (3.4) to 1.5 (2.6) percentage points for men (women). The remaining associations are only statistically significant for men with a medium LOC and women with a high LOC and we can, thus, conclude that a

19 A direct comparison of the coefficients in the reduced and full model would lead to an overestimation of the indirect effect because the error variance of the reduced model is likely to be larger due to the nesting. For a comprehensive formal explanation and implementation in STATA, see Kohler et al. (2011).
very large part of the estimated associations can be explained by differences in social activity and social networking between Internals and Externals.

4 Conclusions

Most studies in the pre-existing economic and psychological literature show that internal individuals live a healthier life. They are more likely to invest in their future health outcomes by following a healthy diet, exercising regularly, and abstaining from smoking. Although we would initially expect this to translate into drinking less or abstaining from alcohol, drinking seems to be different. We find a significant positive link between LOC and alcohol consumption. Men with a medium or high internal LOC are on average about 3.1%–3.7% more likely to be at least occasional drinkers compared to men with a low LOC. Based on this observed nonlinearity in the link between LOC and drinking, we can thus assume that among men an external LOC is linked to less drinking rather than an internal LOC being linked to more drinking. Women with a high internal LOC are even 6.5% more likely on average to be occasional or regular drinkers than women with a low internal LOC with the link being much more linear. These findings are robust to controlling for an extensive list of explanatory variables, the variation in the LOC construct, the definition of the outcome variable and they also largely pass a test for potentially omitted variables based on Oster (2019).

We argue that this finding is likely driven by the fact that the link between drinking and future outcomes is subject to uncertainty more than behavior in other health domains and that especially moderate drinking is largely associated with positive medical, economic and social outcomes. The commonly observed association between LOC and health investments is, hence, less of importance for levels of responsible alcohol consumption. As opposed to this, we suggest that internal individuals more strongly believe in or overestimate their ability to cope with and prevent the negative consequences of drinking. Thus, they might underestimate the risk associated with drinking. In a same way as external individuals overestimate the potential future risks of drinking and thus underestimate their ability to responsibly deal with it. In addition, we show that large parts of the positive relationship can be explained by differences in social behavior and investments into social networks. Internal individuals invest in social networks more strongly by being socially more active. While attending social events, meeting friends, and going out, they are more exposed to alcohol and have more opportunities to drink. A decomposition analysis—in which measures for social activities are included into the model as mediators—indicate that much of the association can be explained by this indirect effect via different levels of social interaction for both men and women.

It is important to note that the two mechanisms are expected to have very distinct economic and medical consequences. Whereas drinking as an investment decision might improve occupational and economic success while being related to rather moderate amounts of alcohol consumption, an underestimation of risks is potentially associated with regular drinking and the economic costs involved, e.g., through the strain that it places on individual health care expenditures and labor market perspectives. However, as excessive alcohol consumption and addiction are relatively rare, we
are unable to identify a sufficient number of individuals involved in this kind of behavior to make statements about the link between LOC and extreme forms of drinking behavior. Further disentangling the association with respect to the underlying channels is not possible with the data at hand. This might be an important path for future research.

Our paper adds interesting new aspect to the literature on behavioral implications of LOC as well as determinants of moderate levels of alcohol consumption and again supports the finding that drinking is different. It became clear that more strongly than other forms of unhealthy behavior, alcohol consumption involves multiple opposing behavioral considerations and particular degrees of uncertainty. The underlying mechanisms—social investments and/or mis-estimation of risks—have many layers and stress the individual complexity behind drinking decisions. Knowing about these specific intrinsic drivers of drinking can, e.g., crucially contribute to the efficacy of interventions with the goal of reducing habitual and dangerous alcohol consumption in the population while not adversely affecting or even promoting light and moderate drinking.

**Supplementary Information**  The online version contains supplementary material available at https://doi.org/10.1007/s00181-022-02219-3.

**Funding**  Open Access funding enabled and organized by Projekt DEAL.

**Availability of data and material**  Publicly available data.

**Code availability**  Replication code available from authors upon request.

**Declarations**

**Conflict of interest**  The authors declare that they have no conflict of interest.

**Open Access**  This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

**Appendix**

See Tables A1, A2, A3, A4 and A5.
Table A1  Sample selection and item non-response

| Step                                      | Estimation sample |
|-------------------------------------------|-------------------|
|                                           | Observations      |
|                                           | Individuals       |
| Full sample (2005–2016)                   | 516,099           |
|                                           | 100,651           |
| **Sample restriction**                    |                   |
| – Drop Younger 20, Older 70               | 318,609           |
|                                           | 63,322            |
| **Item non-response**                     |                   |
| – Drinking variable                       | 50,178            |
|                                           | 22,390            |
| – Locus of control                        | 43,176            |
|                                           | 20,340            |
| – Demographic & interview controls        | 40,078            |
|                                           | 18,035            |
| – Personality controls                    | 37,960            |
|                                           | 16,023            |
| – Education controls                      | 36,940            |
|                                           | 15,507            |
| – Labor market controls                   | 34,745            |
|                                           | 14,991            |
| – Health control                          | 33,765            |
|                                           | 14,841            |

**Source**  SOEP, 2006, 2008, 2010 waves, version 33, https://doi.org/10.5684/soep.v33, own calculations

The full sample contains all available SOEP observations between 2005 and 2016 including, e.g., children and persons without person questionnaires. The table sows the number of observations (or individuals) after the sample is reduced to all observations with valid information for all variables in the respective group.

Table A2  Summary statistics

| Demographic controls $D_{it}$ | All Mean | Men Mean | Women Mean |
|-------------------------------|----------|----------|-----------|
| Female                        | 0.52     | 0.00     | 1.00      |
| Age                           | 47.32    | 47.62    | 47.05     |
| German Nationality            | 0.94     | 0.94     | 0.94      |
| Region of Germany             |          |          |           |
| North and Central             | 0.21     | 0.21     | 0.21      |
| East                          | 0.27     | 0.27     | 0.27      |
| West                          | 0.26     | 0.26     | 0.26      |
| South                         | 0.26     | 0.26     | 0.26      |
| Number of children in HH      | 0.49     | 0.47     | 0.50      |
| Young children                |          |          |           |
| Has child under 1             | 0.02     | 0.02     | 0.02      |
| Has child 1–7 years           | 0.13     | 0.12     | 0.13      |
| Expecting child               | 0.01     | 0.01     | 0.01      |
| Married or stable partner     | 0.76     | 0.76     | 0.75      |
| Religious affiliation         |          |          |           |
| Non                           | 0.33     | 0.36     | 0.30      |
| Christian                     | 0.63     | 0.60     | 0.67      |
| Muslim                        | 0.02     | 0.02     | 0.02      |
| Other                         | 0.01     | 0.01     | 0.01      |
Table A2  continued

|                                | All Mean | Men Mean | Women Mean |
|--------------------------------|----------|----------|------------|
| **Educational controls $E_{it}$** |           |          |            |
| Highest school degree          |           |          |            |
| No school degree               | 0.02      | 0.02     | 0.02       |
| Lower secondary school         | 0.30      | 0.32     | 0.28       |
| Intermediary school            | 0.33      | 0.29     | 0.36       |
| High school                    | 0.30      | 0.32     | 0.28       |
| Other School                   | 0.06      | 0.06     | 0.06       |
| Highest vocational degree      |           |          |            |
| No vocational diploma          | 0.16      | 0.14     | 0.18       |
| Apprenticeship                 | 0.45      | 0.47     | 0.43       |
| Higher technical college       | 0.25      | 0.24     | 0.26       |
| College or university degree   | 0.24      | 0.27     | 0.22       |
| **Labor market controls $LM_{it}$** |           |          |            |
| Net household income in KEUR   | 2.97      | 3.04     | 2.90       |
| Gross labor income in KEUR     | 1.77      | 2.43     | 1.16       |
| Occupational autonomy          |           |          |            |
| Low                            | 0.33      | 0.28     | 0.38       |
| Medium                         | 0.25      | 0.27     | 0.24       |
| High                           | 0.22      | 0.18     | 0.26       |
| Labor force status             |           |          |            |
| Employed or self-employed      | 0.67      | 0.72     | 0.62       |
| Unemployed                     | 0.06      | 0.06     | 0.05       |
| Out of the Labor Force         | 0.27      | 0.22     | 0.33       |
| **Personality controls $P_{i}$** |           |          |            |
| Conscientiousness (avg.)       | 5.89      | 5.83     | 5.96       |
| Extraversion (avg.)            | 4.81      | 4.68     | 4.93       |
| Agreeableness (avg.)           | 5.37      | 5.19     | 5.53       |
| Neuroticism (avg.)             | 3.86      | 3.61     | 4.10       |
| Openness (avg.)                | 4.48      | 4.39     | 4.57       |
| Willingness to take risk (general) (avg.) | 4.49 | 4.93 | 4.08 |
| Willingness to take health risk (avg.) | 2.96 | 3.32 | 2.63 |
| Patience (avg.)                | 6.12      | 6.11     | 6.13       |
| Impulsiveness (avg.)           | 5.11      | 4.98     | 5.24       |
| **Health controls $H_{it}$**   |           |          |            |
| Disabled                       | 0.11      | 0.13     | 0.10       |
| In bad health                  | 0.16      | 0.15     | 0.17       |
| Body Mass Index (imputed)      | 26.20     | 26.99    | 25.47      |
| Mental health score            | 50.28     | 51.26    | 49.36      |
| Physical health score          | 50.01     | 50.26    | 49.77      |
Table A2  continued

|                      | All  | Men   | Women  |
|----------------------|------|-------|--------|
|                      | Mean | Mean  | Mean   |
| Smoking              |      |       |        |
| Non                  | 0.71 | 0.67  | 0.74   |
| Light                | 0.16 | 0.16  | 0.17   |
| Heavy                | 0.13 | 0.17  | 0.09   |
| Healthy diet         |      |       |        |
| Non                  | 0.06 | 0.08  | 0.03   |
| Moderate             | 0.86 | 0.86  | 0.86   |
| Strong               | 0.08 | 0.05  | 0.11   |
| Exercise             |      |       |        |
| Non                  | 0.35 | 0.35  | 0.35   |
| Moderate             | 0.28 | 0.30  | 0.26   |
| Strong               | 0.38 | 0.35  | 0.40   |
| Interview controls I_{11} |      |       |        |
| Year 2006            | 0.34 | 0.34  | 0.34   |
| Year 2008            | 0.34 | 0.34  | 0.34   |
| Year 2010            | 0.33 | 0.33  | 0.33   |
| Day of week          |      |       |        |
| Monday               | 0.16 | 0.16  | 0.16   |
| Tuesday              | 0.17 | 0.17  | 0.17   |
| Wednesday            | 0.17 | 0.17  | 0.17   |
| Thursday             | 0.15 | 0.14  | 0.15   |
| Friday               | 0.16 | 0.16  | 0.16   |
| Saturday             | 0.13 | 0.13  | 0.13   |
| Sunday               | 0.06 | 0.06  | 0.06   |
| Interview mode       |      |       |        |
| Self completed       | 0.29 | 0.30  | 0.29   |
| Orally or with interviewer | 0.30 | 0.30  | 0.30   |
| Remote, by mail/CAPI, CAWI | 0.40 | 0.40  | 0.41   |
| Interview month      |      |       |        |
| February             | 0.35 | 0.35  | 0.35   |
| March                | 0.31 | 0.31  | 0.31   |
| April                | 0.16 | 0.16  | 0.16   |
| May                  | 0.07 | 0.07  | 0.07   |
| June                 | 0.04 | 0.05  | 0.04   |
| July                 | 0.03 | 0.03  | 0.03   |
| Other                | 0.03 | 0.03  | 0.03   |
| Observations         | 33,765 | 16,300 | 17,465 |
| Individuals          | 14,841 | 7157  | 7684   |

Source SOEP, 2006, 2008, 2010 waves, version 33, https://doi.org/10.5684/soep.v33, own calculations
Table A3  Robustness checks (logit, marginal effects, outcome: occasional or regular drinker)

|                   | Men       | Women     |
|-------------------|-----------|-----------|
|                   | (1)       | (2)       | (3)       | (4)       |
| (A) Alternative LOC variables |           |           |           |           |
| (1) LOC Index (equal weights) |           |           |           |           |
| LOC Factor (cont.) | 0.009*    | 0.009*    |           |           |
|                   | (0.005)   | (0.005)   |           |           |
| Locus of control Terciles (Ref.: $[LOC_{min}, LOC_{P33}]$) |           |           |           |           |
| $(LOC_{P33}, LOC_{P66})$ | 0.012     | 0.012     |           |           |
|                   | (0.010)   | (0.011)   |           |           |
| $(LOC_{P66}, LOC_{max})$ | 0.020*    | 0.026**   |           |           |
|                   | (0.011)   | (0.011)   |           |           |
| Observations      | 16,300    | 16,300    | 17,465    | 17,465    |
| All controls      | ✓         | ✓         | ✓         | ✓         |
| (2) Averaged LOC  |           |           |           |           |
| LOC Factor (cont.) | 0.018***  | 0.013**   |           |           |
|                   | (0.006)   | (0.006)   |           |           |
| Locus of Control Terciles (Ref.: $[LOC_{min}, LOC_{P33}]$) |           |           |           |           |
| $(LOC_{P33}, LOC_{P66})$ | 0.023*    | 0.020*    |           |           |
|                   | (0.012)   | (0.012)   |           |           |
| $(LOC_{P66}, LOC_{max})$ | 0.039***  | 0.030**   |           |           |
|                   | (0.013)   | (0.014)   |           |           |
| Observations      | 16,300    | 16,300    | 17,465    | 17,465    |
| All controls      | ✓         | ✓         | ✓         | ✓         |
| (B) Alternative outcome variables—objective measure 2016 |           |           |           |           |
| (1) Moderate or high consumption frequency (2–4 days a month or more often) |           |           |           |           |
| LOC Factor (cont.) | 0.010     | 0.003     |           |           |
|                   | (0.007)   | (0.008)   |           |           |
| Locus of control Terciles (Ref.: $[LOC_{min}, LOC_{P33}]$) |           |           |           |           |
| $(LOC_{P33}, LOC_{P66})$ | 0.031*    | 0.015     |           |           |
|                   | (0.017)   | (0.017)   |           |           |
| $(LOC_{P66}, LOC_{max})$ | 0.022     | −0.000    |           |           |
|                   | (0.018)   | (0.018)   |           |           |
| Observations      | 4162      | 4162      | 4835      | 4835      |
| All controls      | ✓         | ✓         | ✓         | ✓         |
| (2) High consumption amount (3 or more drinks per day) |           |           |           |           |
| LOC Factor (cont.) | 0.013     | −0.000    |           |           |
|                   | (0.008)   | (0.006)   |           |           |
| Locus of control Terciles (Ref.: $[LOC_{min}, LOC_{P33}]$) |           |           |           |           |
| $(LOC_{P33}, LOC_{P66})$ | 0.060***  | −0.006    |           |           |
|                   | (0.018)   | (0.013)   |           |           |
Table A3  continued

| Table A3  continued |
|---------------------|
| **Men**            |
| *LOC_{P66}, LOC_{max}* | (1)  | (2)  | (3)  | (4)  |
| Observations       | 4162 | 4162 | 4835 | 4835 |
| All controls       | ✓    | ✓    | ✓    | ✓    |
| *(C) Stressful events as additional controls* |
| LOC Factor (cont.) | 0.011** | 0.009* | (0.005) | (0.005) |
| Locus of control Terciles (Ref.: [LOC_{min}, LOC_{P33}]) |
| *(LOC_{P33}, LOC_{P66})* | 0.027*** | 0.011 | (0.010) | (0.011) |
| *(LOC_{P66}, LOC_{max})* | 0.022* | 0.032*** | (0.012) | (0.012) |
| Job loss           | −0.023 | −0.023 | 0.024 | 0.024 |
| (0.021) (0.021)    | (0.027) (0.027) |
| Marriage           | −0.053** | −0.054** | −0.013 | −0.014 |
| (0.025) (0.026)    | (0.032) (0.032) |
| Move               | −0.065** | −0.065** | 0.061** | 0.061** |
| (0.028) (0.028)    | (0.031) (0.031) |
| Separation         | −0.001 | −0.001 | 0.089*** | 0.090*** |
| (0.026) (0.026)    | (0.025) (0.025) |
| Death of spouse    | −0.001 | −0.000 | −0.050 | −0.050 |
| (0.078) (0.079)    | (0.056) (0.056) |
| Child birth        | −0.044 | −0.044 | −0.030 | −0.031 |
| (0.035) (0.035)    | (0.040) (0.041) |
| Observations       | 16,295 | 16,295 | 17,449 | 17,449 |
| All controls       | ✓    | ✓    | ✓    | ✓    |
| *(D) Age restriction alternative—25 to 64 years* |
| LOC Factor (cont.) | 0.010** | 0.010* | (0.005) | (0.006) |
| Locus of control Terciles (Ref.: [LOC_{min}, LOC_{P33}]) |
| *(LOC_{P33}, LOC_{P66})* | 0.025** | 0.010 | (0.011) | (0.012) |
| *(LOC_{P66}, LOC_{max})* | 0.021* | 0.031** | (0.013) | (0.013) |
| Observations       | 13,412 | 13,412 | 14,629 | 14,629 |
| All controls       | ✓    | ✓    | ✓    | ✓    |
| *(E) Raw effect in unrestricted sample* |
| LOC Factor (cont.) | 0.036*** | 0.051*** | (0.004) | (0.004) |
### Table A3  continued

| Locus of control terciles (Ref.: \([\text{LOC}_{\text{min}}, \text{LOC}_{P33}]\)) | Men (1) | Women (3) | Men (2) | Women (4) |
|---|---|---|---|---|
| \((\text{LOC}_{P33}, \text{LOC}_{P66}]\) | 0.063*** | 0.082*** | (0.009) | (0.010) |
| \((\text{LOC}_{P66}, \text{LOC}_{\text{max}}]\) | 0.076*** | 0.120*** | (0.010) | (0.010) |

| Observations | 20,697 | 20,697 | 22,479 | 22,479 |

All controls ✗ ✗ ✗ ✗

Source: SOEP, 2016 wave, version 33, https://doi.org/10.5684/soep.v33, own calculations

This table displays the marginal effects estimated based on a logit estimation in line with the model in estimation Eq. (1) with the dependent variable being the binary indicator for occasional or regular drinking with the exception of Panel (B) in which the alternative outcomes based on the 2016 wave are used. Columns (1) and (3) use the continuous LOC factor as the explanatory variable and in columns (2) and (4), LOC is included as binary indicators for the terciles of LOC (with \(\text{LOC} < \text{LOC}_{P33}\) being the reference group). All model, except the raw model presented in panel (E), include the full set of control variables in line with columns (5) and (10) of Table 4. Standard errors (in parentheses) are clustered on the individual level. * \(p < 0.1\), ** \(p < 0.05\), *** \(p < 0.01\).

### Table A4 Descriptive analysis—alcohol consumption 2016

| How often do you drink alcohol? | Men | Women |
|---|---|---|
| Mean | Mean | Mean | Mean | Mean | Mean |
| **Every day** | 0.11 | 0.11 | 0.11 | 0.03 | 0.03 | 0.04* |
| 4–6 days a week | 0.12 | 0.10 | 0.13*** | 0.05 | 0.05 | 0.06 |
| 2–3 days a week | 0.25 | 0.23 | 0.27*** | 0.17 | 0.15 | 0.19*** |
| 2–4 days a month | 0.24 | 0.22 | 0.25** | 0.24 | 0.24 | 0.25 |
| Once a month or less | 0.17 | 0.18 | 0.16** | 0.30 | 0.30 | 0.30 |
| **Never** | 0.12 | 0.15 | 0.09*** | 0.20 | 0.23 | 0.17*** |

| When you drink, how many drinks do you consume per day? |
|---|---|---|---|---|---|---|
| 1–2 drinks | 0.59 | 0.60 | 0.59 | 0.79 | 0.79 | 0.79 |
| 3–4 drinks | 0.29 | 0.29 | 0.30 | 0.17 | 0.16 | 0.17 |
| 5–6 drinks | 0.08 | 0.08 | 0.08 | 0.03 | 0.03 | 0.03 |
| 7–9 drinks | 0.02 | 0.02 | 0.03 | 0.01 | 0.01 | 0.01 |
| 10+ drinks | 0.01 | 0.01 | 0.01 | 0.00 | 0.00 | 0.00 |

Source: SOEP, 2016 wave, version 33, https://doi.org/10.5684/soep.v33, own calculations

Individuals are grouped into Internals and Externals based on whether their LOC is lower/equal or higher than the median. Significance stars refer to the significance level of a \(t\)-test for mean equivalence between externals and internals: * \(p < 0.1\), ** \(p < 0.05\), *** \(p < 0.01\).
Table A5  Relative degree of selection test (Oster (2019)) (OLS, outcome—moderate or regular drinking)

|                         | Baseline effect (SE), \([R^2]\) | Controlled effect (\(\hat{\beta}\)) (SE), \([R^2]\) | Identified set \([\hat{\beta}, \hat{\beta}^*]\) | \(\hat{\delta}\) for \(\beta = 0\) given \(R_{max}\) |
|-------------------------|----------------------------------|--------------------------------------------------------|---------------------------------|----------------------------------|
| **Men**                 |                                  |                                                        |                                 |                                 |
| Locus of control Terciles (Ref.: \([LOC_{min}, LOC_{P33}]\)) |                                  |                                                        |                                 |                                 |
| \((LOC_{P33}, LOC_{P66})\) | 0.067*** (0.011)                  | 0.028*** (0.009)                                        | [0.014, 0.028]                  | 1.88                             |
| \((LOC_{P66}, LOC_{max})\) | 0.081*** (0.011)                  | 0.026** (0.012)                                         | [0.001, 0.026]                  | 1.05                             |
| Observations            | 16,300                           |                                                        |                                 |                                 |
| Health                  | ×                                 | ✓                                                       |                                 |                                  |
| \(R_{max}\)             | 0.095                             |                                                        |                                 |                                  |
| Locus of control Terciles (Ref.: \([LOC_{min}, LOC_{P33}]\)) |                                  |                                                        |                                 |                                 |
| \((LOC_{P33}, LOC_{P66})\) | 0.038*** (0.011)                  |                                                        | [0.027, 0.038]                  | 3.14                             |
| \((LOC_{P66}, LOC_{max})\) | 0.037** (0.012)                  |                                                        | [0.017, 0.037]                  | 1.72                             |
| Observations            | 16,300                           |                                                        |                                 |                                 |
| Health                  | ×                                 | ✓                                                       |                                 |                                  |
| \(R_{max}\)             | 0.072                             |                                                        |                                 |                                  |
| **Women**               |                                  |                                                        |                                 |                                  |
| Locus of control terciles (Ref.: \([LOC_{min}, LOC_{P33}]\)) |                                  |                                                        |                                 |                                 |
| \((LOC_{P33}, LOC_{P66})\) | 0.067*** (0.010)                  | 0.013                                                  | [−0.007, 0.013]                 | 0.66                             |
| \((LOC_{P66}, LOC_{max})\) | 0.115*** (0.012)                  | 0.030**                                                 | [−0.007, 0.030]                 | 0.82                             |
| Observations            | 17,465                           |                                                        |                                 |                                 |
| Health                  | ×                                 | ✓                                                       |                                 |                                  |
| \(R_{max}\)             | 0.141                             |                                                        |                                 |                                  |
| Locus of control terciles (Ref.: \([LOC_{min}, LOC_{P33}]\)) |                                  |                                                        |                                 |                                 |
| \((LOC_{P33}, LOC_{P66})\) | 0.028 *** (0.011)                 |                                                        | [0.013, 0.028]                  | 1.81                             |
| \((LOC_{P66}, LOC_{max})\) | 0.051*** (0.012)                  |                                                        | [0.022, 0.051]                  | 1.64                             |
Table A5 continued

|                          | Baseline effect (SE), $[R^2]$ | Controlled effect ($\hat{\beta}$) (SE), $[R^2]$ | Identified set $[\hat{\beta}, \beta^*]$ | $\delta$ for $\beta = 0$ given $R_{max}$ |
|--------------------------|------------------------------|-----------------------------------------------|----------------------------------------|------------------------------------------|
| 
| Observations            | 16,300                       |                                               |                                        |                                          |
| Health                   | ✓                            |                                               |                                        |                                          |
| $R_{max}$                | 0.107                        |                                               |                                        |                                          |
| Interview                | ✓                            | ✓                                             |                                        |                                          |
| Demographics             | ✓                            |                                               |                                        |                                          |
| Personality              | ✓                            |                                               |                                        |                                          |
| Education                | ✓                            |                                               |                                        |                                          |
| Labor Market             | ✓                            |                                               |                                        |                                          |

Source: SOEP, 2006, 2008, 2010 waves, version 33, [https://doi.org/10.5684/soep.v33](https://doi.org/10.5684/soep.v33), own calculations

Columns (1) and (2) display the estimation results of a linear estimation with the binary indicator for occasional or regular drinking as dependent variable and the binary indicators for the terciles of LOC (with LOC $< LOCP_{33}$ being the reference group) as explanatory variables. In the baseline models in column (1), only interview characteristics and time fixed effects are included as control variables, while the controlled models in column (2) include the full set of control variables. The full set of control variables is varied between the two panels for each gender with the upper panel including health controls and the lower panel not including health controls in the controlled setting. Standard errors (in parentheses) are clustered on the individual level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. $R_{max}$ is set to 1. $\delta$ and reported in the bottom row of each panel. Column (3) reports the identified set, which is bounded below by $\hat{\beta}$ and above by $\beta^*$ at $R_{max}$ and $\delta = 1$. Column (4) shows the value of $\delta$ that would produce $\beta = 0$ given the values of $R_{max}$

References

Andersen HH, Mühlbacher A, Nübling M, Schupp J, Wagner GG (2007) Computation of standard values for physical and mental health scale scores using the SOEP version of SF-12v2. Schmollers Jahrbuch J Appl Soc Sci Stud

Bennett P, Norman P, Murphy S, Moore L, Tudor-Smith C (1998) Beliefs about alcohol, health locus of control, value for health and reported consumption in a representative population sample. Health Educ Res 13(1):25–32

Brant R (1990) Assessing proportionality in the proportional odds model for ordinal logistic regression. Biometrics 46(4):1171–1178

Breen R, Karlson KB, Holm A (2013) Total, direct, and indirect effects in logit and probit models. Socio Methods and Res 42(2):164–191

Buddelmeyer H, Powdthavee N (2016) Can having internal locus of control insure against negative shocks? Psychological evidence from panel data. J Econ Behav Organ 122:88–109

Buonanno P, Vanin P (2013) Bowling alone, drinking together. Empir Econ 44(3):1635–1672

Caliendo M, Cobb-Clark DA, Hennecke J, Uhlendorff A (2019). Locus of control and internal migration. Reg Sci Urban Econ 79

Caliendo M, Cobb-Clark DA, Obst C, Seitz H Uhlendorff A (2020). Locus of control and investment in training. J Hum Resour (forthcoming)

Caliendo M, Cobb-Clark DA, Uhlendorff A (2015) Locus of control and job search strategies. Rev Econ Stat 97(1):88–103

Chatterji P, Dave D, Kaestner R, Markowitz S (2004) Alcohol abuse and suicide attempts among youth. Econ Hum Biol 2(2):159–180

Chiteji N (2010) Time preference, noncognitive skills and well being across the life course?: do noncognitive skills encourage healthy behavior? Am Econ Rev 100:200–204
Cobb-Clark DA, Kassenboehmer SC, Schurer S (2014) Healthy habits: the connection between diet, exercise, and locus of control. J Econ Behav Organ 98:1–28
Cobb-Clark DA, Kassenboehmer SC, Sinning MG (2016) Locus of control and savings. J Bank Finance 73:113–130
Cobb-Clark DA, Schurer S (2013) Two economists’ musings on the stability of locus of control. Econ J 123(570):358–400
Coleman M, DeLeire T (2003) An economic model of locus of control and the human capital investment decision. J Hum Resour 38(3):701–721
Corrao G, Bagnardi V, Zambon A, La Vecchia C (2004) A meta-analysis of alcohol consumption and the risk of 15 diseases. Prev Med 38(5):613–619
Cotti C, Dunn RA, Tefft N (2014) Alcohol-impaired motor vehicle crash risk and the location of alcohol purchase. Soc Sci Med 108:201–209
Cuesta MB, Budría S (2015) Income deprivation and mental well-being: the role of non-cognitive skills. Econ Hum Biol 17:16–28
Francesconi M, James J (2019) Liquid assets? the short-run liabilities of binge drinking. Econ J 129(621):2090–2136
Goebel J, Grabka MM, Liebig S, Kroh M, Richter D, Schröder C, Schupp J (2019) The German socioeconomic panel (SOEP). Jahrbücher für Nationalökonomie und Statistik 239(2):345–360
Grønbæk M (2009) The positive and negative health effects of alcohol- and the public health implications. J Internal Med 265(4):407–420
Grossman M (1972) On the concept of health capital and the demand for health. J Political Econ 80(2):223–255
Grossman M (2000) The human capital model. In: Culyer AJ, Newhouse JP (eds) Handbook of health economics, vol 1, chapter 7. Elsevier B.V., pp 347–408
Heckman JJ, Stixrud J, Urzua S (2006) The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. J Labor Econ 24(3):411–482
Hennecke J (2020) The independent woman—locus of control and female labor force participation. AUT Economics working paper series, 2020/03
Jones AS, Richmond DW (2006) Causal effects of alcoholism on earnings: estimates from the NLSY. Health Econ 15(8):849–871
Källmén H (2000) Manifest anxiety, general self-efficacy and locus of control as determinants of personal and general risk perception. J Risk Res 3(2):111–120
Karlson KB, Holm A (2011) Decomposing primary and secondary effects: a new decomposition method. Res Soc Stratif Mobil 29(2):221–237
Kohler U, Karlson KB, Holm A (2011) Comparing coefficients of nested nonlinear probability models. Stata J 11(3):420–438
Lassi G, Taylor AE, Mahedy L, Heron J, Eisen T, Munafò MR (2019) Locus of control is associated with tobacco and alcohol consumption in young adults of the Avon longitudinal study of parents and children. R Soc Open Sci 6(3)
Leifman H, Kühlhorn E, Allebeck P, Andréasson S, Romelsjö A (1995) Abstinence in late adolescence—ancestors to and covariates of a sober lifestyle and its consequences. Soc Sci Med 41(1):113–121
Lundborg P (2006) Having the wrong friends? Peer effects in adolescent substance use. J Health Econ 25(2):214–233
Lundborg P, Lindgren B (2002) Risk perceptions and alcohol consumption among young people. J Risk Uncertain 25(2):165–183
Macdonald Z, Shields MA (2004) Does problem drinking affect employment? Evidence from England. Health Econ 13(2):139–155
Mangiavacchi L, Piccoli L (2018) Parental alcohol consumption and adult children’s educational attainment. Econ Hum Biol 28:132–145
Marcus J, Siedler T (2015) Reducing binge drinking? The effect of a ban on late-night off-premise alcohol sales on alcohol-related hospital stays in Germany. J Public Econ 123:55–77
McGee AD, McGee P (2016) Search, effort, and locus of control. J Econ Behav Organ 126(2):89–101
Mendolia S, Walker I (2014) The effect of noncognitive traits on health behaviours in adolescence. Health Econ 23(9):1146–1158
Nguyen H (2019) How does alcohol access affect transitional adults’ healthy dietary behaviors? Econd Hum Biol 35:82–95
Nowicki S, Iles-Caven Y, Gregory S, Ellis G, Golding J (2018) Stability of, and associations between, parent and child locus of control expectancies. Front Psychol 9
Oster E (2019) Unobservable selection and coefficient stability: theory and evidence. J Bus Econ Stat 37(2):187–204
Ours JV (2004) A pint a day raises a man’s pay; but smoking blows that gain away. J Health Econ 23(5):863–886
Peters BL, Stringham E (2006) No booze? You may lose: Why drinkers earn more money than nondrinkers. J Labor Res 27(3):411–421
Piatek R, Pinger P (2016) Maintaining (locus of) control? Data combination for the identification and inference of factor structure models. J Appl Econom 31(4):734–755
Pinger P, Schäfer S, Schumacher H (2018) Locus of control and consistent investment choices. J Behav Exp Econ 75(11537):66–75
Preuss M, Hennecke J (2018) Biased by success and failure: how unemployment shapes locus of control. Labour Econ 53:63–74
Ronksley PE, Brien SE, Turner BJ, Mukamal KJ, Ghali WA (2011) Association of alcohol consumption with selected cardiovascular disease outcomes: a systematic review and meta-analysis. Br Med J 342(7795):479
Rotter JB (1966) Generalized expectancies for internal versus external control of reinforcement. Psychol Monogr Gen Appl 80(1):1–28
Sacks JJ, Gonzales KR, Bouchery EE, Tomedi LE, Brewer RD (2015) 2010 National and state costs of excessive alcohol consumption. Am J Prev Med 49(5):e73–e79
Salamanca N, De Grip A, Fournage D, Montizaan R (2016) Locus of control and investment in risky assets. IZA discussion paper Series, no, p 10407
Sayed BA, French MT (2016) To your health!: re-examining the health benefits of moderate alcohol use. Soc Sci Med 167:20–28
Schurer S (2017) Bouncing back from health shocks: locus of control and labor supply. J Econ Behav Organ 133:1–20
Sjoberg L (2000) Factors in risk perception. Risk Anal 20(1):1–11
Sloan FA, Eldred LM, Guo T, Xu Y (2013) Are people overoptimistic about the effects of heavy drinking? J Risk Uncertain 47(1):93–127
Slovic P (1992) Perception of risk: reflections on the psychometric paradigm. In: Golding D, Krimsky S (eds) Social Theories of risk, pp 117–152
SOEP (2017) Socio economic panel data: data for years 1984–2016, version 33. https://doi.org/10.5684/soep.v33
Specht J, Egloff B, Schmukle SC (2013) Everything under control? The effects of age, gender, and education on trajectories of perceived control in a nationally representative german sample. Dev Psychol 49(2):353
Steptoe A, Wardler J (2001) Locus of control and health behaviour revisited: a multivariate analysis of young adults from 18 countries. Br J Psychol 92:659–672
Stürmer T, Hasselbach P, Amelang M (2006) Personality, lifestyle, and risk of cardiovascular disease and cancer: follow-up of population based cohort. Br Med J 332(7554):1359–1362
World Health Organization (2018) Global health observatory data repository—global information system on alcohol and health—patterns of consumption. https://apps.who.int/gho/data/node.main.A1044?lang=en. Accessed 23 April 2020
Ziebarth NR (2018) Biased lung cancer risk perceptions: smokers are misinformed. Jahrbücher fur Nation-alökonomie und Statistik 238(5):395–421
Ziebarth NR, Grabka MM (2009) In vino pecunia? The association between beverage-specific drinking behavior and wages. J Labor Res 30(3):219–244

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.