

Xue, Sen; Kidd, Michael P.; Le, Anh T.; Kirk, Kathy; Martin, Nicholas G.

Working Paper

The Role of Locus of Control in Education, Occupation, Income and Healthy Habits: Evidence from Australian Twins

GLO Discussion Paper, No. 371

Provided in Cooperation with:

Global Labor Organization (GLO)

Suggested Citation: Xue, Sen; Kidd, Michael P.; Le, Anh T.; Kirk, Kathy; Martin, Nicholas G. (2019) : The Role of Locus of Control in Education, Occupation, Income and Healthy Habits: Evidence from Australian Twins, GLO Discussion Paper, No. 371, Global Labor Organization (GLO), Essen

This Version is available at:

<https://hdl.handle.net/10419/200962>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

The Role of Locus of Control in Education, Occupation, Income and Healthy Habits: Evidence from Australian Twins^{*}

Sen Xue¹, Michael P. Kidd², Anh. T. Le³, Kathy Kirk⁴ and Nicholas G. Martin⁴

¹ Institute for Economic and Social Research, Jinan University,

² Royal Melbourne Institute of Technology,

³ Bank West Centre, Curtin University,

⁴ Queensland Institute of Medical Research, Brisbane, Queensland,

Corresponding author Email: michael.kidd@rmit.edu.au

Abstract: The role of non-cognitive skills in socio-economic behavior is a burgeoning research area in economics. Much interest is focused on the personality trait, locus of control, a measure of the extent to which individuals believe their fate is self-determined. The existing empirical literature generally estimates the role of locus of control via OLS. The legitimacy of the approach relies upon stability of locus of control as well as the correct specification of the model, i.e. no omitted variable bias. Recent evidence is supportive of treating locus of control as predetermined, particularly for working age individuals. However, the behavioural genetics consensus is that personality traits including locus of control have a significant heritability component. This suggests the potential for omitted variable problems associated with the prior literature's attempt to identify the impact of locus of control using cross-sectional methods. We address the issue of omitted shared family background and genetic factors using data on both monozygotic and dizygotic twins to examine the role of locus of control. Comparison of results across OLS and twins fixed effect estimators is consistent with substantial upward bias in previous estimates of the locus of the control due to omitted variable problems.

Keywords: Locus of control, twin studies, socioeconomic outcomes

JEL Classification: J24, J21, J31

^{*} The authors are grateful to Bob Gregory and Xin Meng for their constructive comments. We also thank seminar participants at Australian National University, the 2018 IESR Labour and Family Economics Workshop at Jinan University and the 2019 China Meeting of Econometric Society.

1. Introduction

Incorporating non-cognitive skills into economic analysis is an important development over the past two decades. The role of such skills in shaping individual behaviour has been examined in a number of socioeconomic contexts, including health, educational attainment, employment status, earnings and risky behaviour (see Heckman and Rubinstein, 2001; Nyhus and Pons, 2005; Groves, 2005; Heckman et al., 2006; Mueller and Plug, 2006; Flossmann et al., 2007; Heineck and Anger, 2010; Fletcher, 2013; Gensowski, 2018).

In economics, locus of control is one of the most frequently studied non-cognitive skills. This trait captures the extent to which individuals perceive success or failure in life as being self-determined (internal locus of control) versus the role of outside forces, such as chance, luck or under the control of others (external locus of control). Many studies have investigated the association between locus of control and a variety of economic behaviour. There is evidence that individuals with a stronger internal locus of control tend to: save more (Cobb-Clark et al., 2016), have a healthier lifestyle (Cobb-Clark et al., 2014), more readily adopt new technology in agricultural production (Abay et al., 2017), invest more time in parenting (Lekfuangfu et al., 2018), have higher earnings (e.g. Piatek and Pinger, 2016; Heineck and Anger, 2010; Schnitzlein and Stephani, 2016), invest in job search more intensively (e.g. Caliendo et al., 2015; McGee, 2015; McGee and McGee, 2016) and return to employment more rapidly after a health shock (Schurer, 2017) and child birth (Berger and Haywood, 2016).

A number of distinct mechanisms have been proposed in the literature underlying the positive association between (internal) locus of control and improved socioeconomic outcomes. In particular, internal locus of control is deemed to increase the perceived benefit of human capital accumulation and reduce the psychic costs of exerting effort. Thus an internal locus of control tends to increase human capital (broadly defined), and boost labour market returns via both individual and firm behavior (Cobb-Clark, 2015). On the other hand, a strong internal locus of control may increase the independence of the individual, i.e. harder to persuade, less likely to accept

advice from others and consequently less adept at successful teamwork (Avtgis, 1998). Given the potential for both positive and negative effects, the overall impact of locus of control is an empirical question and cannot be predicted *a priori*.

The literature documents a number of potential econometric issues associated with the use of OLS in estimating the role of locus of control (or more broadly, non-cognitive skills) in economic behaviour. To contextualise we focus on estimating a specification with dependent variable (y) representing some socio-economic behaviour, with a set of regressors including of key interest, locus of control and an additive error term. Locus of control may be correlated with the error term leading to biased and inconsistent estimates. Cobb-Clark and Schurer (2013) discuss in detail that the correlation may be due to: reverse causation, i.e. locus of control both determining y and in part being determined by y ; or locus of control may be measured with error, e.g. lagged locus of control used as a proxy. Cebi (2007) and Abay et al. (2017) among others point to an additional econometric issue: the potential exclusion of an omitted variable correlated with both the dependent variable (y) and locus of control.

The majority of literature side-steps all of the above issues by assuming that locus of control is predetermined and the behavioural model correctly specified. Support for this approach is either implicitly or explicitly garnered by reference to recent literature suggesting locus of control is relatively stable for the working-age population (see Specht et al., 2013; Cobb-Clark and Schurer, 2013). This has led analysis to focus on individual variation using cross-sectional methods, particularly OLS (see the discussion in Abay et al., 2017).²

A long-established literature in genetics argues that all non-cognitive/personality traits share a component that is heritable (see the review by Beauchamp et al., 2011, page 64). In particular, Miller and Rose (1982) and Pedersen et al. (1989) suggest that between 30 and 50% of variation of locus of control is attributable to genetics. Since genetic factors in turn likely affect socio-economic outcomes, this suggests that prior empirical estimates of the impact of locus of control may suffer from significant bias

² One exception is that Abay et al (2017) include the individual fixed effect estimation when they analyse the relationship between locus of control and new technology adoption in agricultural production.

due to omitted variables. One piece of indicative evidence is that Coleman and DeLeire (2003) report locus of control as a significant predictor of education attainment, whereas Cebi (2007) finds no significance once cognitive skills are included in their specification.

The literature on links between cognitive traits and genetics and between genetics and a range of individual behaviour is vast, straddling the disciplines of economics, biology, sociology, psychology, and beyond. Much of the literature uses data on twins in order to be able to control for family background and genetic variation. Our focus is specifically on contributing to the literature addressing the effect of locus of control on a set of socio-economic behaviour, i.e. educational attainment, occupation and employment status, income and healthy habits. The novelty is we attempt to control for environmental and genetic background. In analyzing the impact of locus of control we employ a twin design as a quasi-experiment to mitigate the potential omitted variable problem associated with exclusion of environment and genetic factors.

We utilise data from the Australian Twins registry. Temporarily ignoring the twins nature of the data, we employ OLS and find estimates for samples of both monozygotic (MZ) and dizygotic (DZ) twins are generally statistically significant and economically sizeable. Our OLS estimates largely replicate the empirical findings in the literature, suggesting that internal locus of control is associated with better socioeconomic outcomes. Results for the within-twins fixed effect estimates for the sample of DZ twins are similar to the OLS results, though slightly smaller in magnitude. However, the within-twins fixed effect locus of control estimates for the sample of MZ twins are both statistically and economically insignificant on all outcome variables, other than the health habit relating to exercise.

We also explore the sensitivity of our results to the potential role of measurement error, reverse causality, a within-family externality and alternative constructions of the measure of locus of control. The results are found to be relatively robust and strongly suggest that prior estimates of the locus of control are upward biased in size and significance due to omitted variable bias associated with exclusion of family

environment and genetics.

Our results refer to a limited set of outcome variables and specific population and thus may not generalise to other socioeconomic variables and populations. However, our results provide important implications for the existing economic literature. First, OLS estimates of locus of control may overestimate the effect due to omitted variables, especially omitted genetic factors. Second, comparison of our results for engaging in additional exercise and our other outcome variables (e.g. education and income) indicates that locus of control is more likely to have an impact on behaviour which is largely self-determined.

This paper is structured as follows. Section 2 describes the econometric methodology. Section 3 provides a brief description of the data including definitions of key variables of interest. Section 4 provides the main empirical results. Section 5 discusses the robustness of results, and Section 6 concludes.

2. Econometric methodology

In this paper, we model the effect of locus of control on a variety of outcome measures, y_{ij} , as follows:

$$y_{ij} = \beta_0 + \beta_1 LOC_{ij} + \beta_3 \mathbf{X}_{ij} + \beta_4 \mathbf{Z}_j + \mu_{ij}^1 + \mu_j^2 + \epsilon_{ij}, \quad (1)$$

for individual i in twin set j . The key dependent variables include level of education, labour market outcomes (i.e. occupational and employment status and income) and healthy habits (i.e. binge drinking, smoking and exercise regime). The independent variables are as follows; LOC_{ij} is the locus of control, and \mathbf{X}_{ij} is a vector of control variables which may vary across individuals within the twin set, including education, marital status and personality measures depending on the specification. In the main analysis, we exclude other personality traits from the specification, as from a policy perspective it would appear difficult to create a policy intervention which affects only the locus of control (i.e. holding other personal traits constant). The literature is divided in that some studies exclude other personal traits from the specification (e.g. Andrisani, 1977; Coleman and DeLeire, 2003; Cebi, 2007; Piatek and Pinger, 2016) whereas

others do control for other personal traits (e.g. Cobb-Clark et al., 2014; Caliendo et al., 2015; Schnitzlein & Stephani, 2016). As part of the robustness check we explore the impact of including personality measures among the covariates.³ \mathbf{Z}_j is a vector of observable twin-set characteristics (e.g. age, household size and other observed family background); μ_{ij}^1 is unobserved genetic factors which may differ across twins; μ_j^2 is the unobserved shared family background characteristics, such as early childhood experience; and ϵ_{ij} is the error term.⁴

The existence of the unobserved factors, μ_{ij}^1 and μ_j^2 , implies application of OLS to a typical household survey will lead to biased estimates. This paper utilises a twins design to mitigate this omitted variable problem. We apply the within-twins fixed effects estimator to independent samples of MZ and DZ twins and examine the role of μ_{ij}^1 and μ_j^2 in influencing the correlation between locus of control and a variety of outcome measures.

In particular, we assume that socioeconomic outcomes of each twin follows Eq. (1) ($i = 1, 2$). For the DZ twins, Eq. (1) implies the within-twins difference:

$$y_{1j} - y_{2j} = \beta_1(LOC_{1j} - LOC_{2j}) + \beta_3(\mathbf{X}_{1j} - \mathbf{X}_{2j}) + (\mu_{1j}^1 - \mu_{2j}^1) + (\epsilon_{1j} - \epsilon_{2j}), \quad (2)$$

as DZ twins share the same family background but generally have different genetic factors; and for the MZ twins, the within-twins difference is:

$$y_{1j} - y_{2j} = \beta_1(LOC_{1j} - LOC_{2j}) + \beta_3(\mathbf{X}_{1j} - \mathbf{X}_{2j}) + (\epsilon_{1j} - \epsilon_{2j}), \quad (3)$$

as MZ twins share both the same family background and genetic factors. Clearly, the within-twins fixed effect estimator in Eq. (2) still suffers from the omitted variable problem caused by μ_{ij}^1 , but the estimator in Eq. (3) does not. This feature allows us to compare the OLS and within-twins fixed effects estimates of different types of twins to

³ The results in the robustness check show that while the estimates of the impact of locus of control become slightly smaller after controlling for other personality measures, the overall pattern of the results are similar to the case we do not control for other personality measures.

⁴ Variables without subscript i are common to both twins for a particular twin set.

understand the degree of bias generated by the omission of μ_{ij}^1 and μ_j^2 . This is a similar approach to that employed by Ashenfelter and Krueger (1994) and Miller et al. (1995) in the context of estimating the rate of return to education.

While the within-twins fixed effects estimator of MZ twins purges the omitted variables μ_{ij}^1 and μ_j^2 , a potential problem remains in that $\epsilon_{1j} - \epsilon_{2j}$ may still be endogenous and generate bias (Bound and Solon, 1999; Neumark, 1999). For example, Behrman and Rosenzweig (2004) show that substantial variation of birth weight exists within MZ twin sets, and this variation also predicts schooling attainment. In addition, Sandewall et al. (2014) find that MZ twins may also differ significantly in their cognitive skills which in turn is a strong predictor of educational attainment. While the within-twins fixed effects estimator may not be perfect, we provide strong evidence in Section 5 to support that it is less biased than the OLS alternative. One other potential issue with the within-twins fixed effects estimator is that it may exacerbate the attenuation bias caused by measurement error, as the within-twins difference removes substantial variation in the “signal” but may not remove a commensurate amount of “noise”. In Section 5, we further discuss this issue and show that our results are unlikely to be driven by measurement error.⁵

3. Data

3.1 Data Description

The data employed in the current paper is from a mail survey conducted in 1988 to 1991 on the older cohort (or the Canberra sample) of twins enrolled in the Australian National Health and Medical Research Council Twin Registry (ATR hereafter). All the respondents were aged 24 and over in 1988. The survey gathered information on the respondent’s family background, socioeconomic status, such as education, employment status, occupation, income, religion, political preference, healthy habits and personality.

⁵ The twins strategy has an advantage over the individual fixed effect model. Since the locus of control of working-age population is generally stable (Specht et al., 2013; Cobb-Clark and Schurer, 2013), the individual fixed effect model may be more vulnerable to the measurement error issue and less likely to capture the long-run effect if the time span of the data is short (also see the discussion in Abay et al., 2017). In Section 3.2 we will show that there is substantial within-twins variation in locus of control which can be used to identify the effect.

In addition and of particular importance for the current paper, information pertaining to locus of control was also collected for female respondents. This survey together with other related ATR surveys spawned a number of important studies in economics (e.g. Miller et al, 1995; Miller et al, 2001; Webbink et al, 2008; Webbink; 2011; Plug et al, 2014).

We restrict attention to female twins, as unfortunately the survey did not ask male respondents about their locus of control. In the survey, there are 1,830 complete sets of female twins. As our focus is mainly on those in the labour market, we further restrict the sample to respondents aged between 24 and 60 and exclude students and those with missing values for the covariates. This leaves us with 1,145 complete sets of female twins for the analysis, comprising 722 sets of MZ twins and 423 sets of DZ twins.⁶

3.2 Construction of Key Variables

The empirical analysis examines the impact of locus of control on a sequence of outcome variables including: level of education, occupational and employment status, individual income and healthy habits. This section describes the derivation of the key variables used in the empirical analysis.

Locus of Control

The locus of control measure is derived from the Psychological Coping Resources component of the Mastery Module proposed by Pearlin and Schooler (1978). The measure captures the extent to which the individual believes the outcomes of life events are under their own control. The measure is included in the widely-used household survey, Household, Income and Labour Dynamics in Australia (HILDA) which has generated a number of important studies relating to locus of control (e.g. Cobb-Clark and Schurer, 2013; Cobb-Clark et al., 2014; Cobb-Clark et al., 2016).

In particular, our locus of control measure is constructed based on how strongly the respondent agrees or disagrees with the following seven statements:

1. There is really no way I can solve some of the problems I have;
2. Sometimes I feel that I'm pushed around in life;

⁶ Note that the sample size differs slightly across regressions in Sections 4 and 5 due to differing numbers of missing values for each outcome variable.

3. I have little control over things that happen to me;
4. What happens to me in the future mostly depends on me;
5. I can do just about anything I really set my mind to;
6. I often feel helpless in dealing with the problems of life;
7. There is little I can do to change many important things in my life.

In the context of the current twins data, answers for each item are coded along a four-point scale, i.e. strongly agree (1), agree (2), disagree (3) and strongly disagree (4). Items 4 and 5 reflect the internal locus of control, while the others reflect an external focus. Note that these seven questions are identical to those contained in HILDA, but HILDA uses a Likert 7-point response scale. Given this difference and that this is the first twins study in economics using this measurement of locus of control, we conduct an exploratory factor analysis. Figure 1 illustrates that the information contained in the seven questions may be summarised by a single factor, with the eigenvalues of all the other factors less than one. Figure 2 further demonstrates that the seven questions cluster into two parts, according to the loadings of the first factor. Specifically, items 4 and 5 (i.e. internal locus of control) cluster together, and the other five (i.e. external locus of control) also form a cluster. This is consistent with Cobb-Clark and Schurer (2013), Cobb-Clark et al. (2014) and Caliendo et al. (2015). We also conduct a test of the internal consistency of the seven questions, the Cronbach's α is 0.73.

Given the above findings, we follow the literature in constructing two measures of locus of control. We reverse items 4 and 5 and then take the first predicted factor from the factor analysis as the first measure. For the alternative measure once again items 4 and 5 are reversed and then we take the sum of all the items (including the reversed items). The advantage of the first measure is that it contains less measurement error (Piatek and Pinger, 2016), while the alternative measure is inherently a little more transparent. For ease of interpretation, we standardise both measures to have a mean of zero and a standard deviation of one. The higher the value, the stronger the internal locus of control. Figure 3 illustrates the distributions of the two measures and illustrates the similarity between MZ and DZ twins.

To examine the validity of our data and the twins design, we calculate the within-twins correlation of locus of control for our sample and compare with those in the prior literature. Table 1 suggests that the correlations in our sample are generally comparable to those in the literature. Moreover, the sample correlation coefficient for the DZ twins is smaller than that of the MZ twins. This is consistent with the fact DZ twins share less genetic background than MZ twins. The magnitude of the correlation coefficients of the factor measure in our sample (i.e. 0.19 and 0.36) also demonstrates that there is substantial within-twins variation which may be used by the within-twin fixed effect method.⁷ We further plot the distribution of the within-twins difference in locus of control. Figure 4 shows that the distribution for the DZ twins is slightly flatter. The variances of the within-twins difference in locus of control are 1.53 and 1.33 for the DZ and MZ twins, respectively.

Outcome Variables

Outcome variables analysed in the current paper include; educational attainment, labour market related variables together with healthy habits. The education related variables are constructed based on the question asking the respondent their highest completed educational level. The potential answers include “less than 7 years’ schooling”, “8-10 years’ schooling”, “11-12 years’ schooling”, “apprenticeship, diploma, etc.”, “technical or teachers’ college”, “university first degree” and “university post-graduate training”. Following Miller et al. (1995), we first recode the above categories to provide an equivalent years of schooling: i.e. 5, 9, 11.5, 11.5, 13, 15 and 17 years of formal schooling, respectively. This is our primary continuous education variable. The other education related outcome is a dummy variable indicating whether or not the respondent attained a university degree or higher.

The labour market outcome variables capture employment and occupational status, as well as income. Specifically, the employment information is obtained from the question asking the respondent to select the most appropriate category from a mutually exclusive list: “student”, “unemployed”, “part-time”, “homemaker”, “full-time”,

⁷ The correlation coefficients of the alternative measure of locus of control are 0.18 and 0.35 for the DZ and MZ twins in our sample, respectively.

“retired” and “other”.⁸ We construct two dummy variables capturing employment status. One is whether or not the respondent works either full-time or part-time, and the other identifies only full-time employment status.

The survey asked each respondent to describe in detail “their usual and regular lifetime occupation”. These descriptors are subsequently recoded as nine major Australian Standard Classification of Occupations (ASCO) groups: “managers and administrators”, “professionals”, “para-professionals”, “tradespersons”, “clerks”, “salesman and personal service workers”, “plant and machine operators and drivers”, “labourer and related workers” and “homemakers, students and others”. From this information we generate a dummy variable defining whether or not a respondent worked as a manager, administrator or professional as the outcome variable indicating higher ranked occupational status.

The income measure is constructed based on the following survey question which asked each respondent: “Thinking of the income your family makes from all sources--salaries, investments, pensions, and other sources—approximately how much did *you* earn before tax (gross) during 1987-88?” The possible answers are “(1) none, (2) under \$5000, (3) 5000-10000, (4) 10000-15000, (5) 15000-25000, (6) 25000-35000, (7) 35000-50000, (8) over 50000.” We convert these ranges to a continuous measure by taking the mid-points of Options 2 to 7, assigning zero to Option 1 and assigning \$57,500 to Option 8. We then take the natural logarithm as the key income variable. In the following analysis, we analyse income for two distinct samples. One sample includes all respondents who recorded a positive income, and the other includes all respondents working either full or part-time. The latter captures labour income more definitively.

Finally, we construct measures for three healthy habits: whether the respondent was a smoker at the time of the survey, whether the respondent indulged in binge drinking over the past 12 months, and whether the respondent engaged in frequent exercise. The binge drinking measure is constructed based on the question “what is the greatest

⁸ Note that we exclude twin sets in which one or two twin siblings are students from the analysis.

number of drinks you have had in a single day in the past 12 months”. To be consistent with Cobb-Clark et al. (2014), we define having less than 5 drinks in a single day as avoiding binge drinking. The exercise measure is extracted from the question asking the respondent which option best describes your exercise regime : “(1) Don’t know; (2) No leisure exercise or sport; (3) Occasional exercise (2-3 times a month) or regular light gardening; (4) Regular exercise about once a week; (5) Exercise or sport activity a couple of times a week; (6) Jogging, cycling to work, or vigorous sport activity (swimming, etc.) at least 3-4 times a week; (7) Jogging, cycling to work, or vigorous sport activity/exercise daily or more often.” We define Options 6 and 7 as one, Option 1 as missing, and the others as zero. Thus, in line with Cobb-Clark et al. (2014), the constructed measure indicates if the respondents engaged in exercise three or more times per week.

3.3 Summary Statistics

Table 2 presents summary statistics for the sample analysed in the current paper. In particular, the first column shows the mean, standard deviation and sample size for the full sample, the second and third columns differentiate between the DZ and MZ twin samples, and the last column shows the difference in means between the two. The statistics suggest that 66.4% of women were employed, with 38.8% working full-time. 27.8% usually worked as a manager, administrator or professional. The average natural logarithm of income is 9.5. 78.5% did not smoke at the time of the survey, 53.9% avoided binge drinking in the past 12 months prior to the survey, and 13.3% engaged in exercise at least three times per week. The average age of respondents at the time of the survey is 38. The education distribution shows that only 1% of the sample had less than 7 years’ schooling, and the majority (55.7%) had between 8 and 12 years of schooling. 28.6% had an apprenticeship, diploma, technical or teachers’ college as their highest qualification, and only 14.7% had a university degree or above. In addition, 80.1% of the sample were married. The last column shows that the MZ and DZ twins generally have similar characteristics, although the former has more respondents with 8-10 years of schooling.

4. Econometric Results

We present estimates for our two distinct locus of control measures for each of our key outcome variables in Tables 3 and 4. With the exception of the education related outcome variables (i.e. years of schooling and university degree or above), we estimate two distinct specifications. The baseline model includes the locus of control in both OLS and within-twins fixed effect estimation together with a quadratic in age in the former. The extended model adds education and marital status for both OLS and within-twins fixed effect estimation. Comparison between these two specifications allows one to gauge the extent to which locus of control works via education attainment and marital status. For the education related outcome variables, we employ only the baseline specification, as education is generally predetermined to marriage. In each table, we illustrate results estimated separately for the samples of DZ and MZ twins. This enables one to identify the extent to which unobserved common family background and/or genetic factors generate bias in the OLS estimates.

Table 3 illustrates the results for the locus of control measure derived from factor analysis for all key dependent variables i.e. years of education, acquisition of a degree or higher, followed by labour market related outcomes and finally healthy habits. Panel 1 focusses on educational acquisition. The OLS results are generally consistent with the literature (e.g. Coleman and DeLeire, 2003 and the cases in Cebi (2007) where cognitive skill is not controlled for in the regression), suggesting that higher internal locus of control is significantly associated with higher education attainment.⁹ The within-twins fixed effect estimates for the DZ twin sample remains statistically significant with the same sign as the OLS estimates. However, for the MZ twin sample, the within-twin fixed effect estimates are statistically insignificant, and numerically small.

Specifically, the OLS results suggest that a one standard deviation increase in locus of control is associated with a significant increase of 0.432 and 0.375 years of additional

⁹ Due to data constraints, we cannot exactly replicate the regression specification in the prior literature using twins data for the education and other outcome variables. Thus we cannot directly compare coefficient sizes with the literature.

schooling and a 6.3% and 4.1% increased chance of attaining a university degree or higher for DZ and MZ twins, respectively. The within-twins fixed effect results for the DZ sample suggest an equivalent 1 s.d increase in locus of control is associated with a significant increase of 0.216 years of additional schooling and a 5.4% increased chance of attaining a university degree or higher. All the above results are statistically significant at the 5% level. On the other hand, the within-twins fixed effect results for the MZ sample drop to 0.104 years and a 1.6% increased chance, which as stated earlier are both statistically insignificant. Given these results, it is evident that the unobserved common family characteristics and genetic factors both lead to bias in the OLS estimates, with unobserved genetic factors likely playing the dominant role.

Panel 2 illustrates the results for two distinct employment status variables for both baseline and extended specifications. As Heckman et al. (2006) report, the OLS results show that a stronger internal locus of control is associated with a higher probability of being employed for females. The other results illustrate a similar pattern to the above, i.e. the within-twins fixed effect estimates for the DZ twins are statistically significant and large in size, but the within-twins fixed effects estimates for the MZ twins sample are numerically small and insignificant (i.e. -2.3% to 3.1% for both measures and specifications). In addition, the extended specification estimates are generally smaller than those of the baseline for both OLS and within-twins fixed effect estimates, suggesting the effect of locus of control works partially through education attainment and marital status.

The results for occupational status are shown in Panel 3 and differ from the above findings in two regards. First, the inclusion of education attainment and marital status substantially reduces the size and statistical significance of the locus of control estimates, suggesting important indirect pathways. Second, without controlling for the above covariates, the within-twin fixed effect estimates for the MZ sample is statistically significant at the 5% level, and the magnitude of the estimate suggests that a one standard deviation increase in locus of control is related to a 3.7% greater likelihood of attaining a higher ranked occupation. However, as pointed out above, the

impact of locus of control works mainly via educational attainment and marital status. Their inclusion reduces the size of the estimate to 2.8% and is only significant at the 10% level in this case.

Panel 4 presents the results for the income variable. The OLS results are again consistent with the literature (e.g. Heckman et al., 2006; Semykina and Linz, 2006), showing that an internal locus of control is positively correlated with (labour) income. For both the full-time sample and the employed sample, the overall pattern of results is very similar to those in Panels 1 and 2. In particular, the within-twins fixed effect estimates of the MZ sample are once again small and statistically insignificant. The estimates range between 0.006 and 0.052.

The last panel reports the results for the health related habits. In contrast to the other panels, the results demonstrate that locus of control is not significantly correlated with smoking, for both OLS and within-twins fixed effect estimates. The results also suggest that locus of control is negatively correlated with binge drinking avoidance, while only the results for the DZ twins sample are statistically significant at the 10% level. The sign although somewhat counter-intuitive aligns with Cobb-Clark et al. (2014) finding, and rationale that individuals with higher internal locus of control feel more capable of controlling the effect of drinking. Finally and interestingly, the results suggest that locus of control is significantly correlated with exercise, including the within-twins fixed effect estimates for MZ twins. However, note that the within-twins fixed effect results for the MZ twins in this case are once again smaller than OLS, suggesting that unobserved genetic factors lead the OLS estimator to overestimate the effect of locus of control.

A possible explanation for the significant effect of locus of control on exercise is that engaging in exercise is largely self-determined, whereas the other outcome variables, such as education and labour market outcomes, are affected by other agents, e.g. schools, university and firms, and one can thus influence these decisions to a lesser extent. The insight from this result is that locus of control seems more likely to affect outcomes which are largely self-determined.

We repeat the above analysis in Table 4 using the alternative measure of locus of control constructed by standardising the summation of all the items.¹⁰ The results are largely consistent with those in Table 3. In sum, the results in both Tables 3 and 4 suggest that exclusion of the unobserved common household characteristics and genetic factors would yield substantial positive bias in OLS estimates. Once these unobserved factors are controlled for in the within-twins fixed effect estimation for the MZ sample, locus of control is insignificant for almost all outcome variables, and the size of the coefficient is small in magnitude. The only evidence of significance we find is that higher internal locus of control is related to more exercise. However, even in this case we still see that the magnitudes of the within-twins fixed effects estimates shrink by 28% to 36%, compared to OLS estimates for the MZ twins sample.

5. Discussion on the Robustness of the Results

In this section we conduct a set of robustness checks. In the following analysis, we focus on our measure of locus of control derived from factor analysis.¹¹

Measurement Error

One important potential econometric issue with the within-twins fixed effect estimator is the attenuation bias caused by measurement error. The econometric estimator using the within-twins difference removes part of the “signal” but may not remove a commensurate amount of “noise”, reducing the “signal to noise” ratio and thus magnifying attenuation bias. This may lead the within-twins fixed effect estimates of the MZ sample to be small and insignificant. Conventionally, the literature takes the report by the co-twin as an instrumental variable to solve the measurement error issue (e.g. Ashenfelter and Krueger, 1994; Miller et al., 1995). Unfortunately, such a report by the co-twin on locus of control is unavailable in the current survey.

We believe that the insignificance of the within-twins fixed effect estimates of the MZ sample is unlikely to be due to the measurement error problem for three distinct

¹⁰ Items 4 and 5 are reversed in the summation. See the earlier discussion in Section 3.

¹¹ The results using the alternative measure described in Section 3 are largely consistent with the results based on the measure from the factor analysis and available upon the request.

reasons. First, as claimed by Piatek and Pinger (2016), the locus of control measure derived from factor analysis has less measurement error than the alternative measure. If the factor measure does mitigate the measurement error problem, and the measurement error causes large bias in our estimation, then we would expect to see large differences in the results between Tables 3 and 4. But comparing the results across the two measures, there is very little difference.

Second, it is not unreasonable to assume that the extent of measurement error is similar between the MZ and DZ twin samples. The measured variances of the within-twins difference in locus of control are similar (i.e. 1.33 for MZ and 1.53 for DZ as discussed in Section 3.2), so the true within-twins variation of locus of control should be close between the twin samples. Given this, it is expected that the MZ and DZ twin samples suffer from attenuation bias to a similar extent. However, in Tables 3 and 4 we see that whereas the within-twins fixed effect estimates for the MZ sample are numerically small and insignificant, the fixed effect estimates for the DZ sample are large and significant. A comparison with the OLS estimates reveals the fixed effect estimates of the MZ sample drops dramatically, but the estimates of the DZ twins sample do not. This once again suggests that controlling for unobserved genetic factors, rather than measurement error, is the main reason underlying the within-twins fixed effect estimates for the MZ sample being small and insignificant.

Last, in Table 5 we repeat our analysis using a sample which is presumably less subject to the measurement error problem – the sample restricted to respondents in which education information provided by the twin and their co-twin are congruent. The results are similar to those in Tables 3 and 4. Taken together the evidence is persuasive that measurement error is unlikely to be the main reason underlying the small size and insignificance of the within-twins fixed effects estimates of the MZ sample.

Remaining Omitted Variable Bias

The within-twins difference removes the unobserved common childhood experience and family background for both DZ and MZ twins. In addition, the estimator removes genetic factors for MZ twins. However, the within-twins fixed effects

estimator may still suffer from omitted variable problems even in the MZ twins case. For example, there may exist possible competition between twins within the womb and/or subsequent differential treatment by parents. Bound and Solon (1999) and Neumark (1999) discuss the potential bias of the within-twins fixed estimator in detail. The recent literature further confirms that even within MZ twins, differences in birthweight (Behrman and Rosenzweig, 2004) and cognitive ability (Sandewall et al, 2014) exist. Given this, the within-twins fixed effect estimates may still suffer from the omitted variable problem.

It is important to note that the within-twins fixed effect estimator remains useful in tightening the upper bound of the locus of control estimates. If endogenous variation across twin sets is larger than within the twin set, then the within-twins fixed effect estimator will suffer from less bias than the OLS estimator. Further, if locus of control is positively correlated with the omitted variables, then the within-twins fixed effect estimates may be regarded as the upper bound on the locus of control effect.

We follow Ashenfelter and Rouse (1998) to test whether the endogenous variation across twin sets is larger than the associated variation within the twin set. In particular, we compute the correlations of average locus of control with average characteristics of a set of socioeconomic variables over each twin set to indicate the OLS bias. We also compute the correlations of within-twins difference in locus of control with the within-twins difference for the same set of socioeconomic variables to indicate the bias in the within-twins fixed effect estimator.

Table 6 illustrates the results for the MZ and DZ twin samples. The results show that between-twin-set correlations are generally much larger than within-twin-set correlations, especially for the MZ sample. Moreover, both the between and within-twin-set correlations for the DZ sample are larger than those for the MZ sample. This suggests that the within-twins fixed effect estimator is less biased than the OLS estimator, and the within-twins fixed effect estimator for the MZ sample is likely to be less biased than that of the DZ sample.

Inclusion of Personality Measures as Covariates

The main results reported to date exclude personal traits other than locus of control. In this section we examine robustness of our locus of control estimates to the inclusion of personality measures derived from two instruments: the Eysenck Personality Questionnaire and the Tridimensional Personality Questionnaire. The former provides personality measures on 4 scales: extraversion, psychoticism, neuroticism, and a lie scale; while the latter provides measures on 3 scales: novelty seeking, harm avoidance and reward dependence. Both instruments are widely used in psychology studies (e.g. Cloninger et al., 1991; Mor, 2010).¹²

The main rationale for the previous exclusion of such traits is from a policy perspective it would appear difficult to design a policy intervention which affects locus of control, holding all other personality traits fixed. However, as locus of control is inter-related with other personality measures, for example, internal locus of control tends to be correlated with lower neuroticism, some studies on locus of control also include other personality measures as covariates (e.g. Cobb-Clark et al., 2014; Caliendo et al., 2015; Schnitzlein & Stephani, 2016). To be comparable to these studies, we control for the personality measures in Table 7.

Table 7 shows the locus of control results after including the aforementioned personality measures in the specification. The estimates in Table 7 are generally smaller than those in Table 3, but the within-twins fixed effects estimates for the MZ sample for university-or-above education and full-time employment become statistically significant at 10% level. However, we can also see that the within-twins fixed effect estimates become statistically insignificant for both the DZ and MZ twins on the exercise measure in Table 7 after controlling for other personality measures.

Reverse Causality

In the current paper we take the locus of control measured at the survey date as the key explanatory variable of interest. This contemporaneous locus of control measure is frequently adopted in the literature (e.g. Caliendo et al., 2015; Abay et al, 2017). The advantage is it may contain less measurement error than the lagged locus of control

¹² Unfortunately, the survey does not have Big Five measures.

measure (Cobb-Clark and Schurer, 2013).

One potential problem with use of a contemporaneous measure is that it may lead to reverse causality. However, we believe this is unlikely to yield large bias in our labour market and healthy habits related estimates. This is because the literature finds that in the working-age population the locus of control is relatively stable (Specht et al., 2013; Cobb-Clark and Schurer, 2013), and our sample is made up of working age (i.e. 24 to 60) respondents. So our locus of control may be regarded as largely predetermined. Moreover, even if reverse causality exists, the bias is likely to be positive as better economic outcomes tend to lead to a stronger perception of control. Thus, our estimates may be regarded as an upper bound of the effect even if reverse causality bias exists. Regarding the education-related outcomes, the estimates may be biased due to reverse causality. But again, it may be deemed as the upper bound of the effect, as more schooling tends to strengthen the perception of control (see the discussion in Stipek, 1980).

We have one additional piece of evidence to support the argument that the insignificance of the within-twins fixed effect estimates of the MZ sample is not due to reverse causality. In Tables 3 and 4, while the within-twins fixed effect estimates for the MZ sample are small and insignificant, the estimates of the DZ twins are generally large and significant. There is no particular reason to believe that reverse causality bias would differ significantly between the MZ and DZ twin samples. Therefore, the insignificance of the fixed effect estimates of the MZ sample is unlikely to be induced by reverse causality.

Within-Family Externality

Another possibility underlying the small and insignificant fixed effect estimates for MZ twins is that the locus of control of one twin may affect the outcome of the other, i.e. a within-family locus of control externality. When a within-family externality exists, and the locus of control of the co-twin on one's outcome has the same sign as the effect of one's own locus of control, the within twins fixed effects estimator under-estimates the true effect. The insignificant results for the MZ twins sample and significant results

for the DZ twins sample may be due to a stronger within-family externality for the former.

Li et al. (2007) test the externality of communist party membership on the co-twin's income by using sibling information. They use sibling data to run two OLS regressions. In one regression, the key explanatory variable is one's own party membership without controlling for the sibling's party membership, and in the other regression the key explanatory variable is the sibling's party membership without controlling for one's own party membership. They also repeat these two regressions using twins data.

They find that in the regressions using the sibling data, the coefficient of one's own party membership is large and significant, and the coefficient of the sibling's party membership is small and insignificant. In contrast, using the twin data, both the coefficients of one's own party membership and the co-twin's party membership are large and significant. One can interpret the coefficients of the sibling's party membership as the total effect of unobserved ability and any possible family-externality. Assuming that the externality between twins and non-twin siblings is of a similar magnitude and the non-twin siblings have a larger difference in unobserved ability, Li et al. infer that the results are mainly caused by unobserved ability rather than a within-family externality. However, as Li et al. point out there is a strong caveat regarding the validity of their above assumptions. Therefore, it is difficult to identify the importance of the externality.

Information pertaining to non-twin siblings is unavailable in our survey, so the above approach is infeasible. As an alternative, we restrict our sample further to those respondents aged 35 or above to test the externality. The idea is that as people grow up, they develop their own social networks, and the externality from their twin siblings weakens.¹³ We repeat our analysis using this elder sample in Table 8. The results are similar to those in Table 3, suggesting that the within-family externality is not the main reason behind the small and insignificant results of the fixed effect estimator for the

¹³ Appendix Figure 1 shows that the mean level of the absolute within-twins difference in locus of control increases as age goes up, suggesting that twins get less similar to each other and thereby are less likely to affect each other as they grow up.

MZ twin sample.

Nonlinear Effect of Locus of Control

The insignificant results in the within-twins fixed effects estimation for the MZ sample may also be due to model misspecification. In particular, the effect of locus of control on the economic outcome variables may be non-linear. It is possible that locus of control has a positive effect below a certain threshold, and a negative role due to inflexibility and possible stubbornness thereafter; in other words, locus of control may have an inverted-U shape effect. Given this, we include a quadratic in locus of control to test for the nonlinearity in Table 9. The results indicate that the quadratic term is generally statistically insignificant in the within-twins fixed effect estimation results for the MZ sample, so the linear specification of locus of control appears appropriate.

In addition to the above robustness check, we follow Caliendo et al. (2015) in creating an alternative locus of control measure, a dummy variable equal to one for those with an internal locus is greater or equal to the median. In Appendix Table 1, we report the results estimated from this dichotomous measure, and the results show a similar pattern to our earlier reported results.

6. Concluding Remarks

In this paper, we estimate the relationship between locus of control and a set of socio-economic variables using a sample of Australian female twins. The OLS estimates for both DZ and MZ twins and the within-twins fixed effect estimates for DZ twins suggest that (internal) locus of control has a positive and significant association with education level, income, likelihood of employment and higher occupational status as well as more frequent exercise. This result is generally consistent with previous findings in the literature, an internal locus of control tending to improve socio-economic outcomes. However, our within-twins fixed effects estimates for the MZ sample are almost all statistically insignificant, and small in size. The only exception is that locus of control and exercise are positively correlated and significant at the 5% level, while the magnitude of the within-twins fixed effects estimates are smaller than the OLS estimates by 28% to 36% for MZ twins.

The key implication of our results is that OLS estimation may lead to substantial bias due to unobserved genetic factors. One caveat is our findings are based on a set of limited socio-economic outcomes and a particular population – Australian female twins. Whether our results may be generalised to other outcome variables and other populations is unknown. In particular, the existing literature suggests that the relationship between locus of control and outcome variables may differ by gender (e.g. Hansemark, 2003; Semykina and Linz, 2006).

References

- Abay, K. A., Blalock, G., and Berhane, G. (2017). 'Locus of control and technology adoption in developing country agriculture: Evidence from Ethiopia', *Journal of Economic Behavior & Organization*, vol. 143, pp. 98-115.
- Andrisani, P. J., (1977). "Internal-external attitudes, personal initiative, and the labor market experience of black and white men", *Journal of Human Resources*, vol. 12(3), pp. 308-328.
- Ashenfelter, Orley and Krueger, Alan. (1994) "Estimates of the Economic Return to Schooling from a New Sample of Twins." *American Economic Review*, December, 84(5), pp. 1157-73.
- Ashenfelter, O., and Rouse, C. (1998). 'Income, schooling, and ability: Evidence from a new sample of identical twins', *The Quarterly Journal of Economics*, vol. 113(1), pp. 253-284.
- Avtgis, T. A. (1998). 'Locus of control and persuasion, social influence, and conformity: A meta-analytic review', *Psychological Reports*, vol. 83(3), pp. 899-903.
- Beauchamp, J. P., Cesarini, D., Johannesson, M., van der Loos, M. J., Koellinger, P. D., Groenen, P. J., Fowler, J. H., Rosenquist, J. N., Thurik, A. R., and Christakis, N. A. (2011). 'Molecular genetics and economics', *Journal of Economic Perspectives*, vol.

25(4), pp. 57-82.

Behrman, J. R., and Rosenzweig, M. R. (2004). 'Returns to birthweight', *Review of Economics and statistics*, vol. 86(2), pp. 586-601.

Berger, E. M., and Haywood, L. (2016). 'Locus of control and mothers' return to employment', *Journal of Human Capital*, vol. 10(4), pp. 442-481.

Bound, J., and Solon, G. (1999). 'Double trouble: on the value of twins-based estimation of the return to schooling', *Economics of Education Review*, vol. 18(2), pp. 169-182.

Caliendo, M., Cobb-Clark, D. A., and Uhlenhorff, A. (2015). 'Locus of control and job search strategies', *Review of Economics and Statistics*, vol. 97(1), pp. 88-103.

Coleman, M. and DeLeire, T. (2003). 'An economic model of locus of control and the human capital investment decision', *Journal of Human Resources*, vol. 38(3), pp. 701-721.

Cloninger, C.R., Przybeck, T.R. and Svrakic, D.M. (1991). 'The tridimensional personality questionnaire: U.S. normative data', *Psychological Reports*, vol. 63(3 Pt 1), pp. 1047-1057.

Cebi, M. (2007). 'Locus of control and human capital investment revisited', *Journal of Human Resources*, vol. 42(4), pp. 919-932.

Cobb-Clark, D. A. (2015). 'Locus of control and the labor market', *IZA Journal of Labor Economics*, vol. 4(1), pp. 3.

Cobb - Clark, D. A., and Schurer, S. (2013). 'Two economists' musings on the stability of locus of control', *The Economic Journal*, vol. 123(570), pp. F358-F400.

Cobb-Clark, D. A., Kassenboehmer, S. C., and Schurer, S. (2014). 'Healthy habits: The connection between diet, exercise, and locus of control', *Journal of Economic Behavior & Organization*, vol. 98, pp. 1-28.

Cobb-Clark, D. A., Kassenboehmer, S. C., and Sinning, M. G. (2016). 'Locus of control and savings', *Journal of Banking & Finance*, vol. 73, pp. 113-130.

Duffy, D. L., Mitchell, C. A., and Martin, N. G. (1998). 'Genetic and environmental

risk factors for asthma: a cotwin-control study', *American Journal of Respiratory and Critical Care Medicine*, vol. 157(3), pp. 840-845.

Fletcher, J. M. (2013). 'The effects of personality traits on adult labor market outcomes: Evidence from siblings', *Journal of Economic Behavior & Organization*, vol. 89, pp. 122-135.

Flossmann, A. L., Piatek, R., and Wichert, L. (2007). 'Going beyond returns to education: The role of noncognitive skills on wages in Germany', In European Meeting of the Econometric Society, Budapest.

Gensowski, M. (2018). 'Personality, IQ, and lifetime earnings', *Labour Economics*, vol. 51, pp. 170-183.

Groves, M. O. (2005). 'How important is your personality? Labor market returns to personality for women in the US and UK', *Journal of Economic Psychology*, vol. 26(6), pp. 827-841.

Hansemark, O. C. (2003). 'Need for achievement, locus of control and the prediction of business start-ups: A longitudinal study', *Journal of economic Psychology*, vol. 24(3), pp. 301-319.

Heckman, J. J., and Rubinstein, Y. (2001). 'The importance of noncognitive skills: Lessons from the GED testing program', *American Economic Review*, vol. 91(2), pp. 145-149.

Heckman, J. J., Stixrud, J., and Urzua, S. (2006). 'The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior', *Journal of Labor economics*, vol. 24(3), pp. 411-482.

Heineck, G., and Anger, S. (2010). 'The returns to cognitive abilities and personality traits in Germany', *Labour economics*, vol. 17(3), pp. 535-546.

Krueger, A., and Ashenfelter, O. (1992). 'Estimates of the economic return to schooling from a new sample of twins', Working Paper, National Bureau of Economic Research.

Lekfuangfu, W. N., Powdthavee, N., Warrinnier, N., and Cornaglia, F. (2017). 'Locus of control and its intergenerational implications for early childhood skill formation', *The Economic Journal*, vol. 128(608), pp. 298-329.

- Li, H., Liu, P. W., Zhang, J., and Ma, N. (2007). 'Economic returns to communist party membership: Evidence from urban Chinese twins', *The Economic Journal*, vol. 117(523), pp. 1504-1520.
- Mor, N. (2010). 'Eysenck Personality Questionnaire', *The Corsini Encyclopedia of Psychology*, pp. 1-2.
- McGee, A. D. (2015). 'How the perception of control influences unemployed job search', *ILR Review*, vol. 68(1), pp. 184-211.
- McGee, A., and McGee, P. (2016). 'Search, effort, and locus of control', *Journal of Economic Behavior & Organization*, vol. 126, pp. 89-101.
- Miller, P., Mulvey, C., and Martin, N. (1995). 'What do twins studies reveal about the economic returns to education? A comparison of Australian and US findings', *The American Economic Review*, vol. 85(3), pp. 586-599.
- Miller, J. Z., and Rose, R. J. (1982). 'Familial resemblance in locus of control: A twin-family study of the internal-external scale', *Journal of Personality and Social Psychology*, vol. 42(3), pp. 535-540.
- Miller, P., Mulvey, C., and Martin, N. (2001). 'Genetic and environmental contributions to educational attainment in Australia', *Economics of Education Review*, vol. 20(3), pp. 211-224.
- Mueller, G., and Plug, E. (2006). 'Estimating the effect of personality on male and female earnings', *ILR Review*, vol. 60(1), pp. 3-22.
- Mosing, M. A., Pedersen, N. L., Cesarini, D. Johannesson, M., Magnusson, P. K. E., Nakamura, J., Madison, G. and Ullén F. (2012) "Genetic and environmental influences on the relationship between flow proneness, locus of control and behavioral inhibition", *PLoS One*, vol7(11), e47958.
- Neumark, D. (1999). 'Biases in twin estimates of the return to schooling', *Economics of Education Review*, vol. 18(2), pp. 143-148.
- Nyhus, E. K., and Pons, E. (2005). 'The effects of personality on earnings', *Journal of Economic Psychology*, vol. 26(3), pp. 363-384.
- Pearlin, L. I., and Schooler, C. (1978). 'The structure of coping', *Journal of health and*

social behavior, vol. 19(1), pp. 2-21.

Pedersen, N. L., Gatz, M., Plomin, R., Nesselroade, J. R., and McClearn G. E. (1989). "Individual differences in locus of control during the second half of the life span for identical and fraternal twins reared apart and reared together", *Journal of Gerontology: Psychological Science*, vol. 44(4), pp. 100-105.

Piatek, R., and Pinger, P. (2016). 'Maintaining (locus of) control? Data combination for the identification and inference of factor structure models', *Journal of Applied Econometrics*, vol. 31(4), pp. 734-755.

Plug, E., Webbink, D., and Martin, N. (2014). 'Sexual orientation, prejudice, and segregation', *Journal of Labor Economics*, vol. 32(1), pp. 123-159.

Rotter, J. B. (1966). 'Generalized expectancies for internal versus external control of reinforcement', *Psychological monographs: General and applied*, vol. 80(1), pp. 1-28.

Stipek, D. (1980). 'A causal analysis of the relationship between locus of control and academic achievement in first grade', *Contemporary Educational Psychology*, vol. 5, pp. 90-99.

Sandewall, Ö., Cesarini, D., and Johannesson, M. (2014). 'The co-twin methodology and returns to schooling—testing a critical assumption', *Labour Economics*, vol. 26, pp. 1-10.

Schnitzlein, D. D., and Stephani, J. (2016). 'Locus of control and low-wage mobility', *Journal of Economic Psychology*, vol. 53, pp. 164-177.

Schurer, S. (2017). 'Bouncing back from health shocks: Locus of control and labor supply', *Journal of Economic Behavior & Organization*, vol. 133, pp. 1-20.

Semykina, A., and Linz, S. J. (2007). 'Gender differences in personality and earnings: Evidence from Russia', *Journal of Economic Psychology*, vol. 28(3), pp. 387-410.

Specht, J., Egloff, B., and Schmukle, S. C. (2013). 'Everything under control? The effects of age, gender, and education on trajectories of perceived control in a nationally representative German sample', *Developmental psychology*, vol. 49(2), pp. 353.

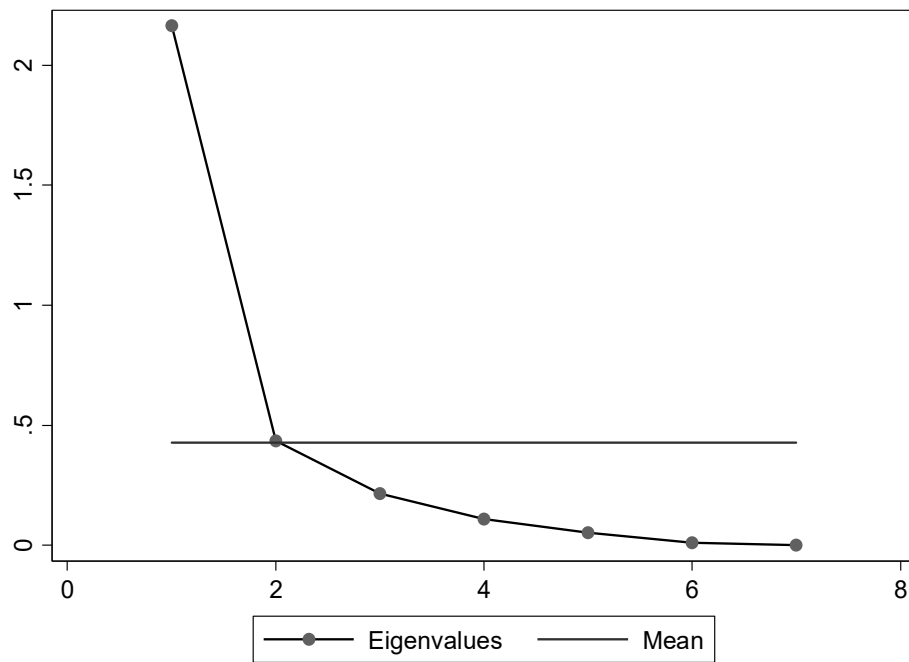
Webbink, D., Martin, N. G., and Visscher, P. M. (2008). 'Does teenage childbearing increase smoking, drinking and body size?', *Journal of Health Economics*, vol. 27(4),

pp. 888-903.

Webbink, D., Martin, N. G., and Visscher, P. M. (2011). 'Does teenage childbearing reduce investment in human capital?', *Journal of Population Economics*, vol. 24(2), pp. 701-730.

Figures

Figure 1: Scree plot of eigenvalues after factor



Notes: The horizontal line the mean value of the eigenvalues of all the seven factors

Figure 2: Factor loadings

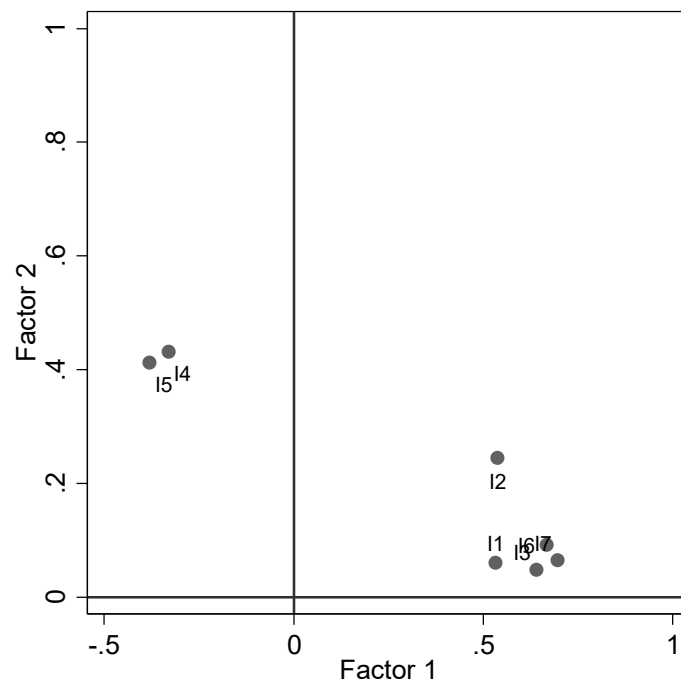


Figure 3.1: Distributions of the factor measure of locus of control

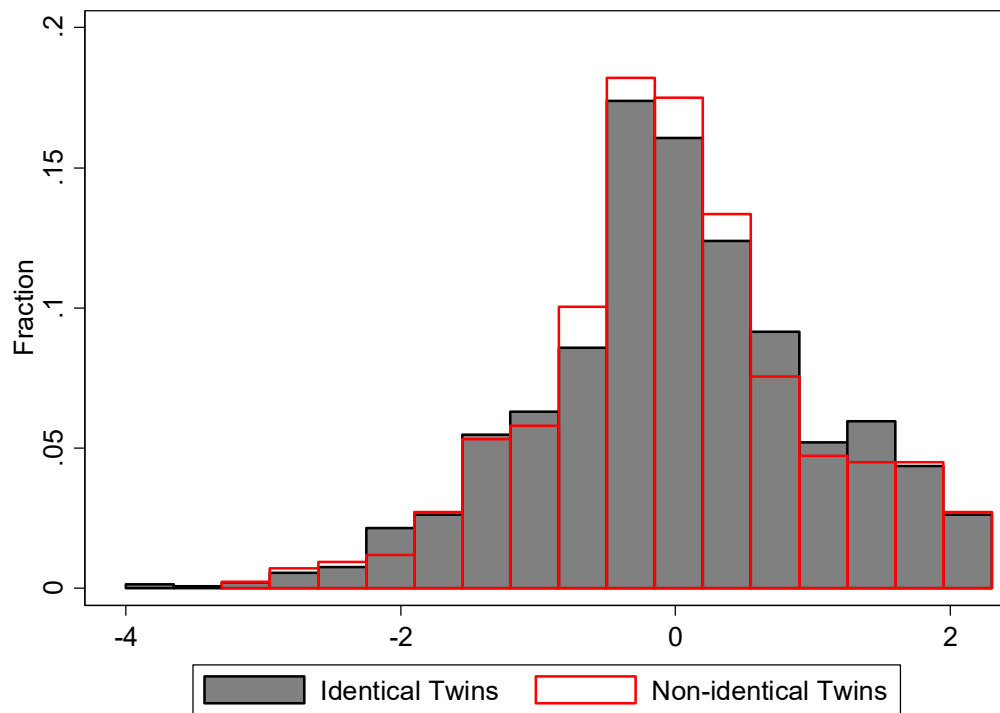


Figure 3.2: Distributions of the alternative measure

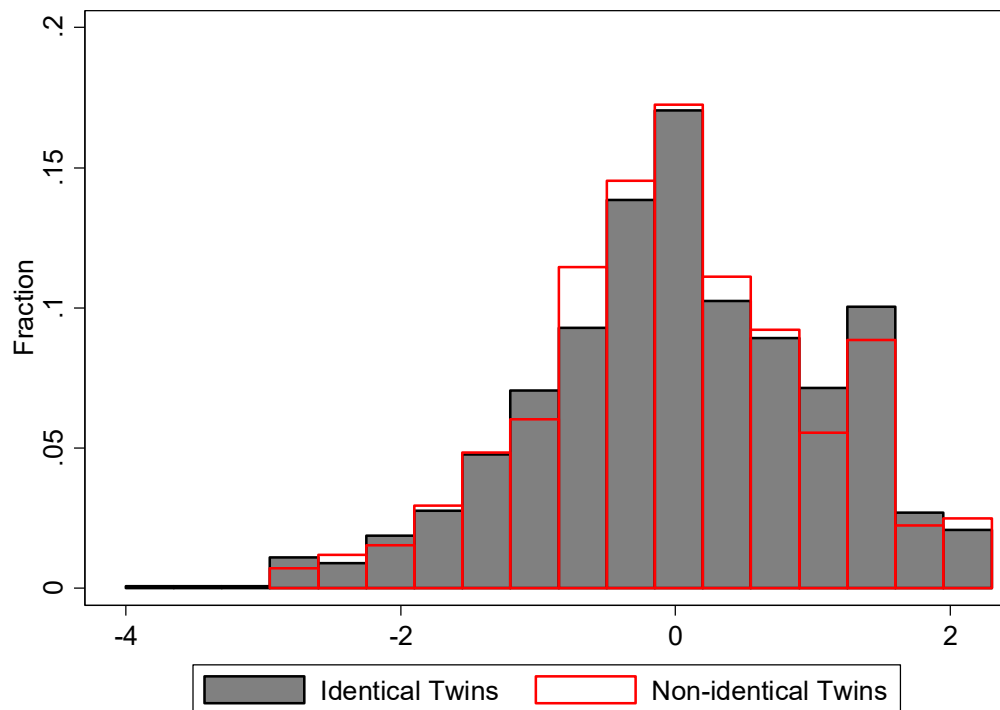


Figure 4.1: Distributions of the within-twins difference on the factor measure of locus of control

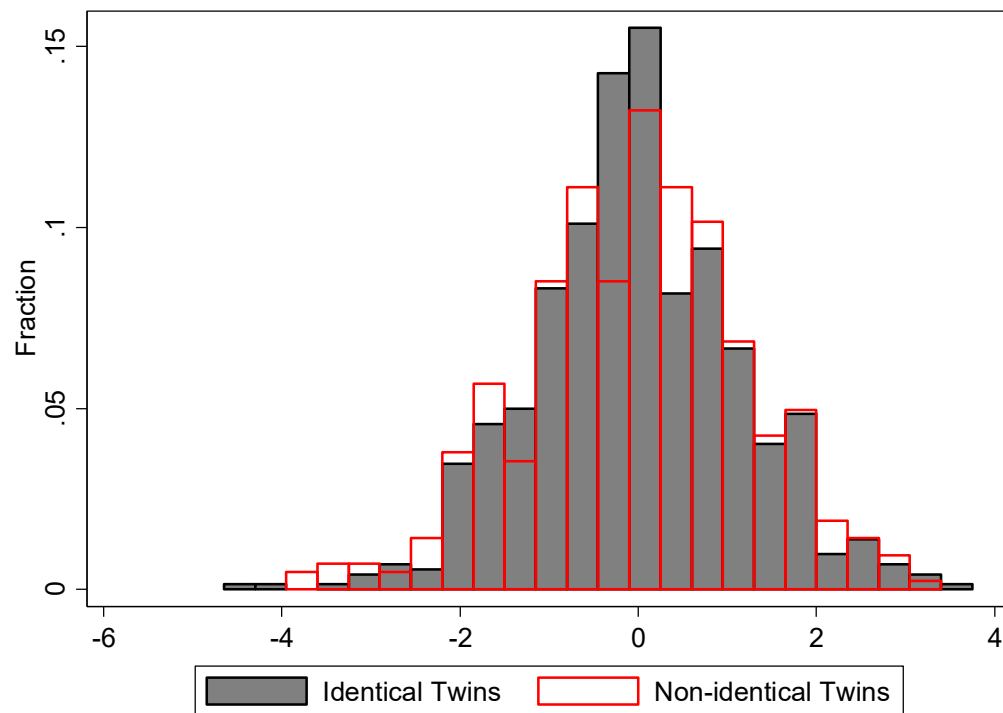
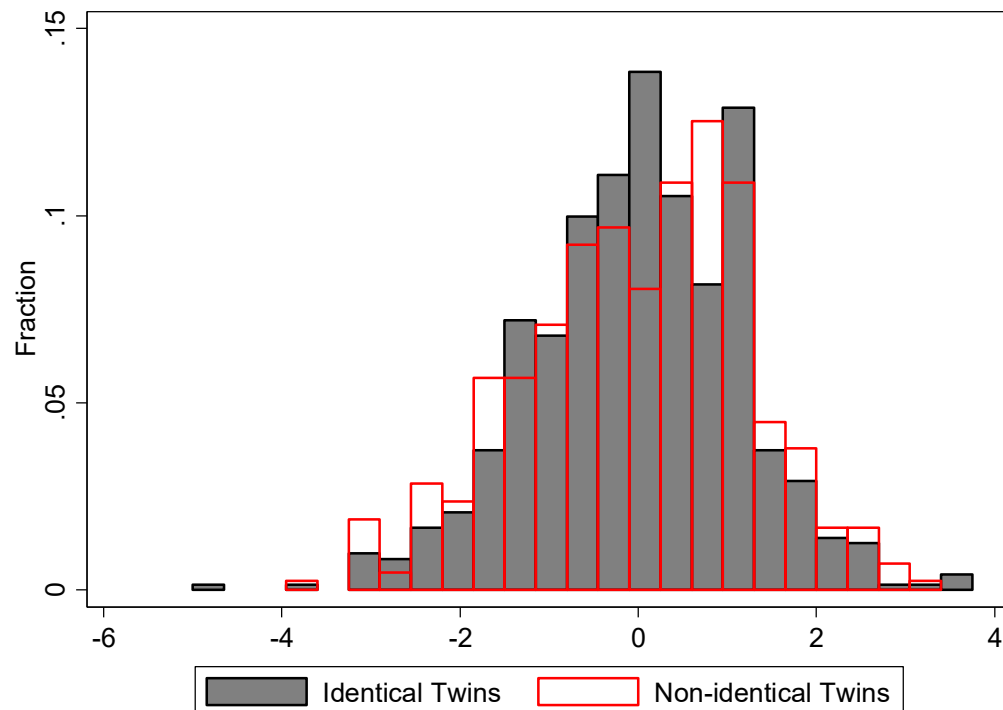


Figure 4.1: Distributions of the within-twins difference on the alternative measure of locus of control



Tables

Table 1: Within-twins correlation coefficients of locus of control

		DZ (Non-identical) Twins	MZ (Identical) Twins
Our sample		0.19 ^a	0.36 ^a
Miller & Rose (1982)		0.18 ^b	0.46 ^b
Pedersen et al. (1989)	Responsibility	0.30 ^c , 0.18 ^d	0.36 ^c , 0.30 ^d
	Life direction	0.23 ^c , 0.15 ^d	0.32 ^c , 0.30 ^d
	Luck	0.18 ^c , 0.32 ^d	0.02 ^c , 0.28 ^d
Mosing et al. (2012)		0.11 ^a	0.29 ^a

Note: Pedersen et al. (1989) report the three sub-scales of locus of control, and all the other studies report the within-twins correlations of the unified index of locus of control. The correlation coefficients of our sample are computed based on the factor measure.

Subscript a denotes the female twins sample.

Subscript b denotes the twins sample including both females and males.

Subscript c denotes the twins sample who were reared apart, including both females and males.

Subscript d denotes the twins sample who were reared together, including both females and males.

Table 2: Summary statistics

	Total (1)	DZ Twins (2)	MZ Twins (3)	DZ-MZ (4)
Full-time or part-time employed	0.664 (0.473) [2262]	0.643 (0.479) [830]	0.675 (0.468) [1432]	-0.032
Full-time employed	0.388 (0.487) [2262]	0.375 (0.484) [830]	0.396 (0.489) [1432]	-0.021
Higher ranked occupation	0.278 (0.448) [2242]	0.256 (0.437) [816]	0.290 (0.454) [1426]	-0.034*
Log annual income	9.509 (0.876) [1490]	9.501 (0.870) [544]	9.514 (0.880) [946]	-0.013
Non-smoker	0.785 (0.411) [2282]	0.798 (0.402) [842]	0.778 (0.416) [1440]	0.020
Avoid binge drinking	0.539 (0.499) [1166]	0.530 (0.500) [398]	0.543 (0.498) [768]	-0.013
Exercise 3 times or more per week	0.133 (0.340) [2240]	0.117 (0.322) [826]	0.143 (0.350) [1414]	-0.025*
LOC: the factor measure	0.000 [2290]	-0.011 (0.973) [846]	0.006 (1.016) [1444]	-0.017
LOC: the alternative measure	0.000 [2290]	-0.012 (0.975) [846]	0.007 (1.015) [1444]	-0.019
Age	38.286 (8.921) [2290]	38.296 (9.042) [846]	38.281 (8.853) [1444]	0.015
Less than 7 years' schooling	0.010 (0.0997) [2290]	0.013 (0.113) [846]	0.008 (0.0908) [1444]	0.005
8-10 years' schooling	0.313 (0.464) [2290]	0.285 (0.452) [846]	0.330 (0.470) [1444]	-0.045**
11-12 years' schooling	0.244 (0.430) [2290]	0.265 (0.441) [846]	0.232 (0.422) [1444]	0.033*
Apprenticeship, diploma, etc.	0.138 (0.345)	0.147 (0.354)	0.132 (0.339)	0.015

	[2290]	[846]	[1444]	
Technical or teachers' college	0.148	0.149	0.148	0.001
	(0.356)	(0.356)	(0.355)	
	[2290]	[846]	[1444]	
University first degree	0.091	0.093	0.089	0.004
	(0.287)	(0.291)	(0.285)	
	[2290]	[846]	[1444]	
University post-graduate degree	0.056	0.048	0.060	-0.012
	(0.230)	(0.215)	(0.238)	
	[2290]	[846]	[1444]	
Married	0.801	0.792	0.807	-0.015
	(0.399)	(0.406)	(0.395)	
	[2290]	[846]	[1444]	

Note: The summary statistics of the full sample, the MZ twins and the DZ twins are shown in Columns (1), (2) and (3), respectively. Standard errors are shown in the parentheses, and samples sizes are shown in the brackets. Column (4) shows the difference between DZ twins and MZ twins. ***Significant at 1% level; **significant at 5% level; *significant at 10% level. The significance levels are derived from t test.

Table 3: Various estimates based on the factor measure of locus of control

VARIABLES	Panel A: DZ (Non-identical) Twins				Panel B: MZ (Identical) Twins			
	OLS		FE		OLS		FE	
	Baseline (1)	Extended (2)	Baseline (3)	Extended (4)	Baseline (5)	Extended (6)	Baseline (7)	Extended (8)
<i>Panel 1: Educational Measures</i>								
Years of schooling	0.432*** [0.087]		0.216** [0.102]		0.375*** [0.071]		0.104 [0.063]	
Observations	846		846		1,444		1,444	
R-squared	0.128		0.014		0.117		0.004	
Number of twin sets			423				722	
University degree or above	0.063*** [0.012]		0.054*** [0.017]		0.041*** [0.011]		0.016 [0.011]	
Observations	846		846		1,444		1,444	
R-squared	0.049		0.027		0.044		0.003	
Number of twin sets			423				722	
<i>Panel 2: Employment Measures</i>								
Full-time or part-time employment	0.065*** [0.017]	0.045*** [0.017]	0.055** [0.024]	0.049* [0.025]	0.043*** [0.012]	0.037*** [0.012]	-0.023 [0.019]	-0.016 [0.019]
Observations	830	830	830	830	1,432	1,432	1,432	1,432
R-squared	0.035	0.119	0.012	0.068	0.033	0.089	0.002	0.031
Number of twin sets			415	415			716	716
Full-time employment	0.074*** [0.017]	0.050*** [0.015]	0.057** [0.022]	0.049** [0.023]	0.053*** [0.013]	0.046*** [0.013]	0.018 [0.019]	0.031 [0.019]
Observations	830	830	830	830	1,432	1,432	1,432	1,432
R-squared	0.044	0.231	0.014	0.129	0.041	0.169	0.001	0.067

Number of twin sets			415	415			716	716
<i>Panel 3: Occupation Outcome</i>								
Higher ranked occupation	0.077***	0.026*	0.061***	0.033	0.050***	0.015	0.037**	0.028*
	[0.015]	[0.014]	[0.024]	[0.020]	[0.013]	[0.010]	[0.016]	[0.015]
Observations	816	816	816	816	1,426	1,426	1,426	1,426
R-squared	0.036	0.352	0.019	0.232	0.034	0.361	0.008	0.157
Number of twin sets			408	408			713	713
<i>Panel 4: Income Measures</i>								
Ln income -full sample	0.167***	0.125***	0.150***	0.118**	0.096***	0.074***	0.048	0.052
	[0.036]	[0.035]	[0.056]	[0.058]	[0.029]	[0.027]	[0.041]	[0.040]
Observations	544	544	544	544	946	946	946	946
R-squared	0.048	0.181	0.028	0.1	0.035	0.145	0.003	0.048
Number of twin sets			272	272			473	473
Ln income -employed sample	0.171***	0.119***	0.139**	0.108*	0.098***	0.081***	0.007	0.006
	[0.046]	[0.043]	[0.064]	[0.064]	[0.031]	[0.028]	[0.042]	[0.042]
Observations	330	330	330	330	618	618	618	618
R-squared	0.055	0.216	0.03	0.084	0.055	0.16	0	0.021
Number of twin sets			165	165			309	309
<i>Panel 5: Healthy Habits Measures</i>								
Not current smoker	0.006	-0.001	-0.023	-0.022	-0.006	-0.016	-0.014	-0.019
	[0.016]	[0.015]	[0.018]	[0.017]	[0.012]	[0.012]	[0.012]	[0.012]
Observations	842	842	842	842	1,440	1,440	1,440	1,440
R-squared	0.030	0.064	0.004	0.027	0.008	0.042	0.001	0.020
Number of twin sets			421	421			720	720
Avoid binge drinking	-0.049*	-0.048*	-0.071*	-0.065*	-0.020	-0.028	-0.003	-0.009
	[0.027]	[0.027]	[0.037]	[0.037]	[0.017]	[0.017]	[0.023]	[0.023]

Observations	398	398	398	398	768	768	768	768
R-squared	0.031	0.035	0.018	0.029	0.031	0.051	0.000	0.016
Number of twin sets			199	199			384	384
Doing exercise 3 times or more per week	0.034***	0.033***	0.048**	0.048**	0.041***	0.039***	0.027**	0.028**
	[0.012]	[0.012]	[0.020]	[0.020]	[0.010]	[0.010]	[0.014]	[0.014]
Observations	826	826	826	826	1,414	1,414	1,414	1,414
R-squared	0.018	0.032	0.016	0.036	0.019	0.026	0.006	0.016
Number of twin sets			413	413			707	707

Notes: Standard errors are clustered at twin set level. ***Significant at 1% level; **significant at 5% level; *significant at 10% level. In addition to locus of control, the baseline OLS specification includes constant term, age and age square; the extended OLS specification includes constant term, age, age square, education attainment dummies and marriage dummy; the baseline FE specification includes no other covariate, and the extended FE specification includes education attainment dummies and marriage dummy.

Table 4: Various Estimates based on the Alternative Measure of Locus of Control

VARIABLES	Panel A: DZ (Non-identical) Twins				Panel B: MZ (Identical) Twins			
	OLS		FE		OLS		FE	
	Baseline	Extended	Baseline	Extended	Baseline	Extended	Baseline	Extended
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<i>Panel 1: Educational Measures</i>								
Years of schooling	0.410***		0.205**		0.349***		0.098	
	[0.085]		[0.099]		[0.071]		[0.063]	
Observations	846		846		1444		1444	
R-squared	0.125		0.013		0.113		0.004	
Number of twin sets			423				722	
University degree or above	0.061***		0.051***		0.037***		0.016	
	[0.012]		[0.016]		[0.011]		[0.011]	
Observations	846		846		1444		1444	
R-squared	0.048		0.025		0.042		0.003	
Number of twin sets			423				722	
<i>Panel 2: Employment Measures</i>								
Full-time or part-time employment	0.071***	0.051***	0.059**	0.051**	0.041***	0.035***	-0.017	-0.011
	[0.017]	[0.016]	[0.023]	[0.024]	[0.012]	[0.012]	[0.019]	[0.019]
Observations	830	830	830	830	1432	1432	1432	1432
R-squared	0.038	0.121	0.015	0.069	0.032	0.088	0.001	0.03
Number of twin sets			415	415			716	716
Full-time employment	0.080***	0.055***	0.064***	0.053**	0.050***	0.044***	0.018	0.029
	[0.016]	[0.015]	[0.022]	[0.023]	[0.013]	[0.013]	[0.019]	[0.019]
Observations	830	830	830	830	1432	1432	1432	1432
R-squared	0.048	0.233	0.018	0.131	0.04	0.168	0.001	0.067

Number of twin sets			415	415			716	716
Panel 3: Occupation Outcome								
Higher ranked occupation	0.075***	0.026*	0.062***	0.035*	0.044***	0.012	0.037**	0.029*
	[0.016]	[0.014]	[0.024]	[0.020]	[0.013]	[0.010]	[0.016]	[0.015]
Observations	816	816	816	816	1426	1426	1426	1426
R-squared	0.035	0.352	0.02	0.233	0.032	0.36	0.008	0.158
Number of twin sets			408	408			713	713
Panel 4: Income Measures								
Ln income -full sample	0.166***	0.122***	0.141**	0.106*	0.088***	0.067**	0.036	0.040
	[0.036]	[0.035]	[0.055]	[0.057]	[0.029]	[0.027]	[0.041]	[0.040]
Observations	544	544	544	544	946	946	946	946
R-squared	0.048	0.181	0.026	0.097	0.033	0.144	0.002	0.046
Number of twin sets			272	272			473	473
Ln income -employed sample	0.171***	0.117***	0.133**	0.100	0.093***	0.079***	-0.003	-0.005
	[0.046]	[0.043]	[0.063]	[0.064]	[0.031]	[0.028]	[0.043]	[0.043]
Observations	330	330	330	330	618	618	618	618
R-squared	0.056	0.215	0.029	0.082	0.053	0.159	0	0.021
Number of twin sets			165	165			309	309
Panel 5: Healthy Habits Measures								
Not current smoker	0.004	-0.003	-0.024	-0.023	-0.007	-0.016	-0.015	-0.020
	[0.015]	[0.015]	[0.017]	[0.017]	[0.012]	[0.012]	[0.012]	[0.012]
Observations	842	842	842	842	1,440	1,440	1,440	1,440
R-squared	0.030	0.064	0.005	0.028	0.009	0.042	0.002	0.020
Number of twin sets			421	421			720	720
Avoid binge drinking	-0.043	-0.041	-0.064*	-0.057	-0.019	-0.026	0.002	-0.004
	[0.027]	[0.027]	[0.037]	[0.037]	[0.017]	[0.017]	[0.022]	[0.022]

Observations	398	398	398	398	768	768	768	768
R-squared	0.029	0.033	0.015	0.026	0.031	0.051	0.000	0.016
Number of twin sets			199	199			384	384
Doing exercise 3 times or more per week	0.035***	0.034***	0.043**	0.043**	0.044***	0.042***	0.028**	0.029**
	[0.011]	[0.012]	[0.019]	[0.019]	[0.010]	[0.010]	[0.014]	[0.014]
Observations	826	826	826	826	1,414	1,414	1,414	1,414
R-squared	0.019	0.033	0.013	0.034	0.021	0.028	0.006	0.016
Number of twin sets			413	413			707	707

Notes: Standard errors are clustered at twin set level. ***Significant at 1% level; **significant at 5% level; *significant at 10% level. In addition to locus of control, the baseline OLS specification includes constant term, age and age square; the extended OLS specification includes constant term, age, age square, education attainment dummies and marriage dummy; the baseline FE specification includes no other covariate, and the extended FE specification includes education attainment dummies and marriage dummy.

Table 5: Measurement Error: Estimates based on the factor measure of locus of control using the sample with correct cross-report on education

VARIABLES	Panel A: DZ (Non-identical) Twins				Panel B: MZ (Identical) Twins			
	OLS		FE		OLS		FE	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<i>Panel 1: Educational Measures</i>								
Years of schooling	0.426***		0.155		0.418***		0.084*	
	[0.126]		[0.099]		[0.087]		[0.044]	
Observations	444		444		934		934	
R-squared	0.147		0.011		0.164		0.007	
Number of twin sets			222				467	
University degree or above	0.033**		0.015		0.042***		0.004	
	[0.017]		[0.015]		[0.013]		[0.010]	
Observations	444		444		934		934	
R-squared	0.025		0.003		0.063		0	
Number of twin sets			222				467	
<i>Panel 2: Employment Measures</i>								
Full-time or part-time employment	0.075***	0.058***	0.077**	0.076**	0.043***	0.034**	-0.024	-0.021
	[0.024]	[0.022]	[0.035]	[0.034]	[0.015]	[0.015]	[0.025]	[0.025]
Observations	436	436	436	436	926	926	926	926
R-squared	0.041	0.13	0.022	0.113	0.035	0.102	0.002	0.045
Number of twin sets			218	218			463	463
Full-time employment	0.059**	0.039*	0.076**	0.070**	0.052***	0.042***	0.013	0.023
	[0.024]	[0.022]	[0.033]	[0.033]	[0.016]	[0.015]	[0.024]	[0.025]
Observations	436	436	436	436	926	926	926	926
R-squared	0.033	0.253	0.025	0.216	0.051	0.192	0.001	0.07

Number of twin sets			218	218			463	463
Panel 3: Occupation Outcome								
Higher ranked occupation	0.079***	0.035*	0.071**	0.025	0.046***	0.003	0.043**	0.034*
	[0.022]	[0.020]	[0.034]	[0.027]	[0.016]	[0.013]	[0.021]	[0.018]
Observations	434	434	434	434	928	928	928	928
R-squared	0.038	0.377	0.024	0.339	0.04	0.432	0.012	0.263
Number of twin sets			217	217			464	464
Panel 4: Income Measures								
Ln income -full sample	0.068	0.059	0.119	0.119	0.096***	0.074**	0.011	0.013
	[0.054]	[0.052]	[0.086]	[0.085]	[0.036]	[0.034]	[0.054]	[0.056]
Observations	264	264	264	264	618	618	618	618
R-squared	0.013	0.168	0.016	0.116	0.046	0.158	0	0.058
Number of twin sets			132	132			309	309
Ln income -employed sample	0.089	0.074	0.141	0.123	0.100**	0.071*	-0.015	-0.031
	[0.069]	[0.061]	[0.093]	[0.089]	[0.040]	[0.036]	[0.054]	[0.056]
Observations	150	150	150	150	388	388	388	388
R-squared	0.018	0.234	0.032	0.172	0.059	0.182	0	0.041
Number of twin sets			75	75			194	194
Panel 5: Healthy Habits Measures								
Not current smoker	-0.009	-0.016	-0.007	-0.006	-0.009	-0.023	-0.004	-0.009
	[0.021]	[0.020]	[0.024]	[0.024]	[0.015]	[0.014]	[0.015]	[0.015]
Observations	442	442	442	442	932	932	932	932
R-squared	0.029	0.070	0.000	0.027	0.004	0.046	0.000	0.021
Number of twin sets			221	221			466	466
Avoid binge drinking	-0.094**	-0.097***	-0.093*	-0.096*	-0.021	-0.034	0.006	0.003
	[0.037]	[0.037]	[0.047]	[0.049]	[0.021]	[0.021]	[0.029]	[0.029]

Observations	198	198	198	198	506	506	506	506
R-squared	0.055	0.067	0.033	0.056	0.038	0.071	0.000	0.028
Number of twin sets			99	99			253	253
<hr/>								
Doing exercise 3 times or more per week	0.041*** [0.014]	0.043*** [0.015]	0.069** [0.027]	0.069** [0.027]	0.044*** [0.013]	0.044*** [0.013]	0.030* [0.017]	0.035** [0.017]
Observations	438	438	438	438	912	912	912	912
R-squared	0.022	0.038	0.032	0.087	0.021	0.027	0.007	0.014
Number of twin sets			219	219			456	456

Notes: Standard errors are clustered at twin set level. ***Significant at 1% level; **significant at 5% level; *significant at 10% level. In addition to locus of control, the baseline OLS specification includes constant term, age and age square; the extended OLS specification includes constant term, age, age square, education attainment dummies and marriage dummy; the baseline FE specification includes no other covariate, and the extended FE specification includes education attainment dummies and marriage dummy.

Table 6: Remaining omitted variable bias: Between and within family correlations of the factor measure of locus of control and other variables

	Panel 1: MZ twins		Panel 2: DZ twins	
	Between-twin-set	Within-twin-set	Between-twin-set	Within-twin-set
Own years of schooling	0.237***(P<0.001)	0.065*(P=0.079)	0.272***(P<0.001)	0.118**(P=0.015)
Spousal years of schooling	0.169***(P<0.001)	0.027(P=0.563)	0.180***(P=0.004)	0.131**(P=0.035)
University degree or above	0.165***(P<0.001)	0.055(P=0.138)	0.206***(P<0.001)	0.164***(P=0.001)
Spousal University degree or above	0.147***(P=0.001)	0.057(P=0.212)	0.092(P=0.137)	0.159***(P=0.010)
Full-time or part-time employment	0.186***(P<0.001)	-0.045(P=0.229)	0.158***(P=0.001)	0.112**(P=0.023)
Spousal full-time or part-time employment	0.098**(P=0.032)	0.040(P=0.386)	0.070(P=0.255)	0.079(P=0.198)
Full-time employment	0.176***(P<0.001)	0.036(P=0.338)	0.194***(P<0.001)	0.119**(P=0.015)
Spousal full-time employment	0.100**(P=0.028)	0.033(P=0.475)	0.117*(P=0.057)	0.026(P=0.674)
Higher ranked occupation	0.143***(P<0.001)	0.089***(P=0.017)	0.202***(P<0.001)	0.135***(P=0.006)
Higher ranked occupation of Spouse	0.125***(P=0.006)	-0.021(P=0.651)	0.085(P=0.171)	-0.007(P=0.915)
Ln income	0.170***(P=0.001)	0.051(P=0.270)	0.218***(P<0.001)	0.167***(P=0.006)
Spousal ln income	0.162***(P=0.001)	0.086*(P=0.084)	0.231***(P<0.001)	0.016(P=0.813)
Smoking	-0.016(P=0.674)	-0.039(P=0.298)	0.037(P=0.451)	-0.065(P=0.183)
Spouse was smoking	0.039(P=0.387)	-0.014(P=0.758)	-0.013(P=0.836)	-0.055(P=0.365)
Doing exercise	0.149***(P=0.001)	0.076***(P=0.043)	0.099***(P=0.045)	0.128***(P=0.009)
Spouse was doing exercise	0.127***(P=0.005)	0.045(P=0.325)	0.107*(P=0.081)	0.129***(P=0.036)

Notes: ***Significant at 1% level; **significant at 5% level; *significant at 10% level. Between-twin-set correlation are correlations of average twin-set locus of control (average of the twins) with average twin-set characteristics, and within-twin-set correlations are correlations of the within-twin-set differences in locus of control with within-twin-set differences in other characteristics.

Table 7: Adding personality measures: Various estimates based on the factor measure of locus of control

VARIABLES	Panel A: DZ (Non-identical) Twins				Panel B: MZ (Identical) Twins			
	OLS		FE		OLS		FE	
	Baseline (1)	Extended (2)	Baseline (3)	Extended (4)	Baseline (5)	Extended (6)	Baseline (7)	Extended (8)
<i>Panel 1: Educational Measures</i>								
Years of schooling	0.277*** [0.104]		0.224* [0.116]		0.319*** [0.083]		0.115 [0.073]	
Observations	812		812		1,400		1,400	
R-squared	0.172		0.037		0.179		0.012	
Number of twin sets			406				700	
University degree or above	0.047*** [0.015]		0.053*** [0.019]		0.042*** [0.012]		0.022* [0.012]	
Observations	812		812		1,400		1,400	
R-squared	0.104		0.047		0.084		0.012	
Number of twin sets			406				700	
<i>Panel 2: Employment Measures</i>								
Full-time or part-time employment	0.049** [0.020]	0.036* [0.019]	0.035 [0.027]	0.025 [0.028]	0.051*** [0.015]	0.048*** [0.016]	-0.014 [0.021]	-0.008 [0.022]
Observations	798	798	798	798	1,388	1,388	1,388	1,388
R-squared	0.059	0.136	0.033	0.098	0.054	0.106	0.008	0.036
Number of twin sets			399	399			694	694
Full-time employment	0.053*** [0.019]	0.033* [0.018]	0.051* [0.027]	0.039 [0.027]	0.054*** [0.016]	0.051*** [0.015]	0.024 [0.023]	0.038* [0.023]
Observations	798	798	798	798	1,388	1,388	1,388	1,388
R-squared	0.078	0.251	0.022	0.137	0.063	0.182	0.017	0.081

Number of twin sets			399	399			694	694
<i>Panel 3: Occupation Outcome</i>								
Higher ranked occupation	0.046**	0.010	0.065**	0.035	0.045***	0.011	0.040**	0.028*
	[0.018]	[0.016]	[0.026]	[0.023]	[0.015]	[0.012]	[0.018]	[0.016]
Observations	790	790	790	790	1,384	1,384	1,384	1,384
R-squared	0.058	0.347	0.033	0.237	0.075	0.377	0.047	0.192
Number of twin sets			395	395			692	692
<i>Panel 4: Income Measures</i>								
Ln income -full sample	0.146***	0.114***	0.151**	0.103	0.107***	0.100***	0.051	0.059
	[0.042]	[0.041]	[0.064]	[0.065]	[0.036]	[0.034]	[0.046]	[0.045]
Observations	522	522	522	522	922	922	922	922
R-squared	0.067	0.192	0.047	0.133	0.055	0.160	0.023	0.066
Number of twin sets			261	261			461	461
Ln income -employed sample	0.094*	0.062	0.091	0.050	0.081**	0.080**	0.020	0.022
	[0.051]	[0.050]	[0.075]	[0.075]	[0.039]	[0.036]	[0.049]	[0.049]
Observations	314	314	314	314	598	598	598	598
R-squared	0.115	0.252	0.079	0.150	0.079	0.177	0.033	0.053
Number of twin sets			157	157			299	299
<i>Panel 5: Healthy Habits Measures</i>								
Not current smoker	-0.005	-0.013	-0.020	-0.024	0.003	-0.007	-0.007	-0.013
	[0.018]	[0.018]	[0.020]	[0.020]	[0.014]	[0.014]	[0.014]	[0.014]
Observations	810	810	810	810	1,396	1,396	1,396	1,396
R-squared	0.059	0.095	0.017	0.040	0.072	0.099	0.015	0.030
Number of twin sets			405	405			698	698
Avoid binge drinking	-0.023	-0.022	-0.064	-0.058	0.004	-0.005	-0.001	-0.010
	[0.032]	[0.032]	[0.043]	[0.042]	[0.020]	[0.020]	[0.028]	[0.028]

Observations	386	386	386	386	750	750	750	750
R-squared	0.110	0.117	0.049	0.062	0.120	0.135	0.022	0.040
Number of twin sets			193	193			375	375
Doing exercise 3 times or more per week	0.026*	0.025*	0.026	0.027	0.033***	0.032***	0.018	0.021
	[0.014]	[0.014]	[0.023]	[0.023]	[0.012]	[0.012]	[0.016]	[0.016]
Observations	796	796	796	796	1,370	1,370	1,370	1,370
R-squared	0.030	0.041	0.038	0.056	0.038	0.045	0.012	0.024
Number of twin sets			398	398			685	685

Notes: Standard errors are clustered at twin set level. ***Significant at 1% level; **significant at 5% level; *significant at 10% level.

In addition to locus of control, the baseline OLS specification includes constant term, age, age square and personality measures; the extended OLS specification includes constant term, age, age square, education attainment dummies, marriage dummy and personality measures; the baseline FE specification includes personality measures, and the extended FE specification includes education attainment dummies, marriage dummy and personality measures.

Table 8: Within-family externality: Estimates based on the factor measure of locus of control using the sample with aged 35 above

VARIABLES	Panel A: DZ (Non-identical) Twins				Panel B: MZ (Identical) Twins			
	OLS		FE		OLS		FE	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<i>Panel 1: Educational Measures</i>								
Years of schooling	0.465***		0.235*		0.286***		0.088	
	[0.118]		[0.135]		[0.093]		[0.083]	
Observations	498		498		848		848	
R-squared	0.082		0.016		0.083		0.003	
Number of twin sets			249				424	
University degree or above	0.050***		0.054***		0.024**		0.017	
	[0.014]		[0.020]		[0.012]		[0.013]	
Observations	498		498		848		848	
R-squared	0.04		0.033		0.037		0.004	
Number of twin sets			249				424	
<i>Panel 2: Employment Measures</i>								
Full-time or part-time employment	0.073***	0.056***	0.065**	0.063*	0.030*	0.024	-0.032	-0.024
	[0.022]	[0.021]	[0.032]	[0.032]	[0.015]	[0.015]	[0.025]	[0.025]
Observations	486	486	486	486	836	836	836	836
R-squared	0.081	0.147	0.02	0.081	0.067	0.095	0.004	0.031
Number of twin sets			243	243			418	418
Full-time employment	0.072***	0.046**	0.060**	0.051*	0.035**	0.032**	0.014	0.027
	[0.020]	[0.019]	[0.028]	[0.029]	[0.016]	[0.016]	[0.025]	[0.025]
Observations	486	486	486	486	836	836	836	836
R-squared	0.044	0.196	0.017	0.128	0.025	0.124	0.001	0.073
Number of twin sets			243	243			418	418

<i>Panel 3: Occupation Outcome</i>								
Higher ranked occupation	0.087***	0.034**	0.099***	0.065**	0.033**	0.011	0.024	0.02
	[0.019]	[0.017]	[0.030]	[0.026]	[0.015]	[0.013]	[0.022]	[0.021]
Observations	484	484	484	484	834	834	834	834
R-squared	0.053	0.362	0.05	0.249	0.03	0.34	0.004	0.16
Number of twin sets			242	242			417	417
<i>Panel 4: Income Measures</i>								
Ln income -full sample	0.211***	0.187***	0.210***	0.179**	0.100**	0.073**	0.035	0.040
	[0.044]	[0.041]	[0.074]	[0.073]	[0.040]	[0.036]	[0.058]	[0.058]
Observations	316	316	316	316	536	536	536	536
R-squared	0.072	0.206	0.056	0.147	0.031	0.159	0.001	0.041
Number of twin sets			158	158			268	268
Ln income -employed sample	0.216***	0.185***	0.217**	0.215**	0.121***	0.102**	-0.004	0.000
	[0.059]	[0.053]	[0.087]	[0.086]	[0.046]	[0.040]	[0.063]	[0.065]
Observations	200	200	200	200	358	358	358	358
R-squared	0.079	0.22	0.063	0.17	0.032	0.171	0	0.021
Number of twin sets			100	100			179	179
<i>Panel 5: Healthy Habits Measures</i>								
Not current smoker	0.000	-0.004	-0.033	-0.032	-0.010	-0.014	-0.013	-0.017
	[0.017]	[0.017]	[0.020]	[0.022]	[0.014]	[0.014]	[0.016]	[0.016]
Observations	496	496	496	496	848	848	848	848
R-squared	0.017	0.052	0.010	0.031	0.003	0.029	0.001	0.026
Number of twin sets			248	248			424	424
Avoid binge drinking	-0.060*	-0.066*	-0.065	-0.069	-0.018	-0.027	-0.018	-0.021
	[0.036]	[0.037]	[0.049]	[0.047]	[0.022]	[0.022]	[0.029]	[0.029]
Observations	190	190	190	190	408	408	408	408

R-squared	0.037	0.066	0.017	0.103	0.018	0.035	0.002	0.045
Number of twin sets			95	95			204	204
Doing exercise 3 times or more per week	0.031**	0.030**	0.046**	0.048**	0.029***	0.028**	0.014	0.016
	[0.012]	[0.013]	[0.020]	[0.021]	[0.011]	[0.011]	[0.017]	[0.017]
Observations	486	486	486	486	822	822	822	822
R-squared	0.019	0.033	0.020	0.060	0.010	0.016	0.002	0.020
Number of twin sets			243	243			411	411

Notes: Standard errors are clustered at twin set level. ***Significant at 1% level; **significant at 5% level; *significant at 10% level. In addition to locus of control, the baseline OLS specification includes constant term, age and age square; the extended OLS specification includes constant term, age, age square, education attainment dummies and marriage dummy; the baseline FE specification includes no other covariate, and the extended FE specification includes education attainment dummies and marriage dummy.

Table 9: Nonlinear effect: Estimates based on the factor measure of locus of control

VARIABLES	Panel A: DZ (Non-identical) Twins				Panel B: MZ (Identical) Twins			
	OLS		FE		OLS		FE	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<i>Panel 1: Educational Measures</i>								
Years of schooling								
Linear term	0.429***		0.217**		0.388***		0.102	
	[0.087]		[0.101]		[0.071]		[0.067]	
Quadratic term	-0.069		-0.004		0.085**		-0.005	
	[0.061]		[0.068]		[0.041]		[0.036]	
Observations	846		846		1,444		1,444	
R-squared	0.13		0.014		0.12		0.004	
Number of twin sets			423				722	

University degree or above								
Linear term	0.063***		0.054***		0.043***		0.017	
	[0.012]		[0.016]		[0.011]		[0.012]	
Quadratic term	0.007		0.004		0.016***		0.001	
	[0.008]		[0.010]		[0.006]		[0.007]	
Observations	846		846		1,444		1,444	
R-squared	0.05		0.028		0.049		0.003	
Number of twin sets			423				722	
<i>Panel 2: Employment Measures</i>								
Full-time or part-time employment								
Linear term	0.065***	0.045***	0.057**	0.051**	0.044***	0.037***	-0.022	-0.016
	[0.017]	[0.017]	[0.023]	[0.024]	[0.012]	[0.012]	[0.018]	[0.018]

Quadratic term	-0.005	-0.005	-	-	0.005	0	0.003	0
	[0.012]	[0.011]	[0.014]	[0.014]	[0.008]	[0.008]	[0.011]	[0.011]
Observations	830	830	830	830	1,432	1,432	1,432	1,432
R-squared	0.035	0.119	0.029	0.085	0.033	0.089	0.002	0.031
Number of twin sets			415	415			716	716
<hr/>								
Full-time employment								
Linear term	0.074***	0.050***	0.057**	0.049**	0.056***	0.048***	0.017	0.029
	[0.017]	[0.015]	[0.022]	[0.023]	[0.013]	[0.013]	[0.019]	[0.020]
Quadratic term	0.018*	0.013	-0.002	-0.004	0.021**	0.012	-0.004	-0.007
	[0.010]	[0.009]	[0.013]	[0.013]	[0.008]	[0.008]	[0.011]	[0.011]
Observations	830	830	830	830	1,432	1,432	1,432	1,432
R-squared	0.047	0.232	0.014	0.129	0.045	0.17	0.001	0.068
Number of twin sets			415	415			716	716

Panel 3: Occupation Outcome

Higher ranked occupation								
Linear term	0.077***	0.026*	0.061***	0.033	0.052***	0.016	0.039**	0.030**
	[0.015]	[0.014]	[0.023]	[0.020]	[0.013]	[0.010]	[0.016]	[0.015]
Quadratic term	-0.007	-0.005	0.009	0.003	0.016*	0.005	0.008	0.007
	[0.010]	[0.008]	[0.014]	[0.011]	[0.008]	[0.007]	[0.010]	[0.010]
Observations	816	816	816	816	1,426	1,426	1,426	1,426
R-squared	0.037	0.352	0.02	0.233	0.037	0.361	0.009	0.158
Number of twin sets			408	408			713	713

Panel 4: Income Measures

Ln income -full sample								
Linear term	0.168***	0.125***	0.156***	0.124**	0.093***	0.071***	0.049	0.055

	[0.037]	[0.035]	[0.057]	[0.058]	[0.029]	[0.027]	[0.041]	[0.040]
Quadratic term	-0.021	-0.019	-0.046	-0.045	0.035*	0.031*	-0.029	-0.034
	[0.022]	[0.020]	[0.031]	[0.031]	[0.019]	[0.017]	[0.022]	[0.021]
Observations	544	544	544	544	946	946	946	946
R-squared	0.049	0.182	0.035	0.107	0.038	0.148	0.005	0.051
Number of twin sets			272	272			473	473

Ln income -employed sample								
Linear term	0.177***	0.126***	0.168**	0.138**	0.092***	0.076**	0.016	0.017
	[0.050]	[0.045]	[0.068]	[0.067]	[0.033]	[0.030]	[0.042]	[0.042]
Quadratic term	-0.021	-0.027	-0.064*	-0.066*	0.021	0.019	-0.035	-0.042*
	[0.031]	[0.026]	[0.037]	[0.036]	[0.024]	[0.022]	[0.025]	[0.025]
Observations	330	330	330	330	618	618	618	618
R-squared	0.057	0.218	0.045	0.1	0.056	0.161	0.005	0.029
Number of twin sets			165	165			309	309

Panel 5: Healthy Habits Measures

Not current smoker								
Linear term	0.006	-0.001	-0.023	-0.022	-0.007	-0.017	-0.016	-0.021*
	[0.016]	[0.015]	[0.017]	[0.017]	[0.012]	[0.012]	[0.012]	[0.012]
Quadratic term	-0.023**	-0.023**	-0.001	-0.001	-0.010	-0.011	-0.007	-0.006
	[0.010]	[0.009]	[0.010]	[0.010]	[0.007]	[0.007]	[0.007]	[0.007]
Observations	842	842	842	842	1,440	1,440	1,440	1,440
R-squared	0.037	0.070	0.004	0.027	0.010	0.043	0.003	0.021
Number of twin sets			421	421			720	720

Avoid binge drinking								
Linear term	-0.045	-0.044	-0.069*	-0.062*	-0.021	-0.029	-0.001	-0.007
	[0.027]	[0.028]	[0.036]	[0.036]	[0.017]	[0.018]	[0.022]	[0.022]

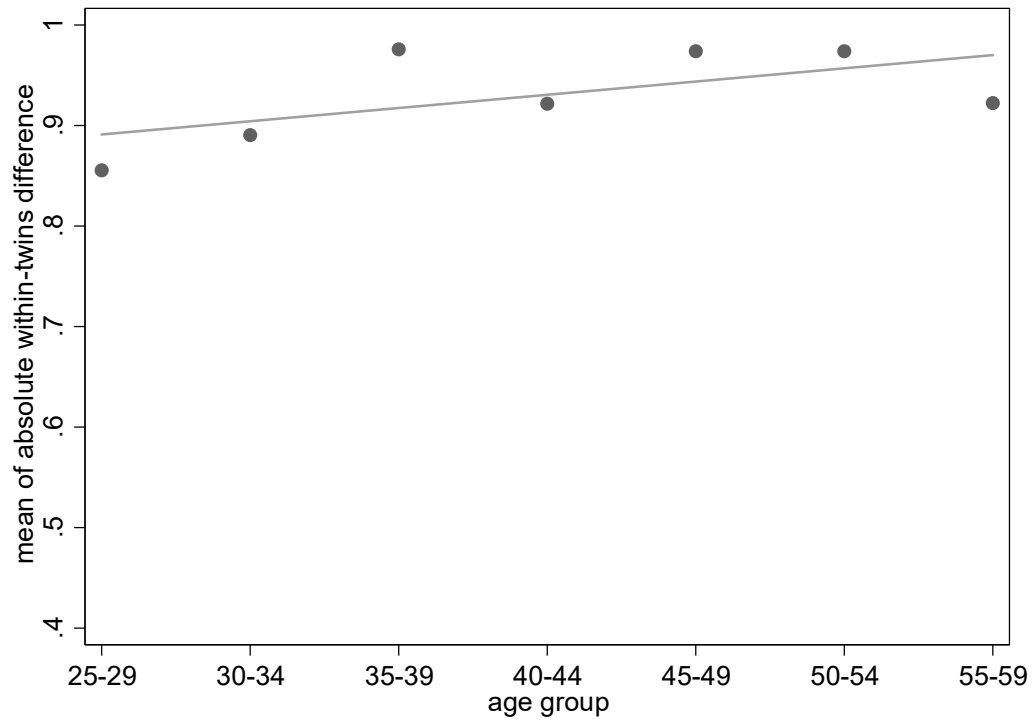
Quadratic term	-0.014	-0.011	-0.008	-0.010	-0.001	-0.003	0.012	0.013
	[0.020]	[0.020]	[0.025]	[0.025]	[0.010]	[0.010]	[0.013]	[0.013]
Observations	398	398	398	398	768	768	768	768
R-squared	0.032	0.036	0.018	0.030	0.031	0.051	0.002	0.019
Number of twin sets			199	199			384	384

Doing exercise 3 times or more per week								
Linear term	0.034***	0.033***	0.048**	0.048**	0.042***	0.040***	0.026*	0.028*
	[0.012]	[0.012]	[0.020]	[0.019]	[0.010]	[0.010]	[0.015]	[0.015]
Quadratic term	0.001	0.000	-0.001	-0.001	0.009	0.009	-0.003	-0.003
	[0.008]	[0.008]	[0.011]	[0.011]	[0.006]	[0.006]	[0.009]	[0.009]
Observations	826	826	826	826	1,414	1,414	1,414	1,414
R-squared	0.018	0.032	0.016	0.036	0.021	0.028	0.006	0.016
Number of twin sets			413	413			707	707

Notes: Standard errors are clustered at twin set level. ***Significant at 1% level; **significant at 5% level; *significant at 10% level. In addition to the linear and quadratic terms of locus of control, the baseline OLS specification includes constant term, age and age square; the extended OLS specification includes constant term, age, age square, education attainment dummies and marriage dummy; the baseline FE specification includes no other covariate, and the extended FE specification includes education attainment dummies and marriage dummy.

Appendix Figure

Appendix Figure 1: The relationship between the mean of the absolute within-twins difference in locus of control and age



Appendix Table 1: Various estimates based on the dichotomous measure of locus of control

VARIABLES	Panel A: DZ (Non-identical) Twins				Panel B: MZ (Identical) Twins			
	OLS		FE		OLS		FE	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<i>Panel 1: Educational Measures</i>								
Years of schooling	0.804***		0.454***		0.649***		0.173	
	[0.159]		[0.173]		[0.135]		[0.123]	
Observations	846		846		1,444		1,444	
R-squared	0.125		0.018		0.109		0.003	
Number of twin sets			423				722	
University degree or above	0.120***		0.097***		0.078***		0.016	
	[0.026]		[0.030]		[0.020]		[0.024]	
Observations	846		846		1,444		1,444	
R-squared	0.047		0.025		0.043		0.001	
Number of twin sets			423				722	
<i>Panel 2: Employment Measures</i>								
Full-time or part-time employment	0.093***	0.055*	0.055	0.037	0.060**	0.050**	-0.067*	-0.056
	[0.033]	[0.032]	[0.044]	[0.045]	[0.025]	[0.025]	[0.037]	[0.036]
Observations	830	830	830	830	1,432	1,432	1,432	1,432
R-squared	0.027	0.114	0.004	0.060	0.028	0.085	0.005	0.033
Number of twin sets			415	415			716	716
Full-time employment	0.087***	0.037	0.033	0.010	0.075***	0.066***	0.031	0.049
	[0.034]	[0.032]	[0.043]	[0.045]	[0.027]	[0.026]	[0.039]	[0.038]
Observations	830	830	830	830	1,432	1,432	1,432	1,432
R-squared	0.030	0.223	0.001	0.119	0.035	0.164	0.001	0.066
Number of twin sets			415	415			716	716

Panel 3: Occupation Outcome

Higher ranked occupation	0.124***	0.021	0.074*	0.009	0.106***	0.049**	0.051*	0.041
	[0.033]	[0.029]	[0.043]	[0.039]	[0.025]	[0.020]	[0.030]	[0.028]
Observations	816	816	816	816	1,426	1,426	1,426	1,426
R-squared	0.028	0.349	0.008	0.227	0.036	0.362	0.004	0.155
Number of twin sets			408	408			713	713

Panel 4: Income Measures

Ln income -full sample	0.271***	0.176**	0.157	0.077	0.198***	0.166***	0.147*	0.154*
	[0.076]	[0.075]	[0.111]	[0.115]	[0.059]	[0.056]	[0.084]	[0.083]
Observations	544	544	544	544	946	946	946	946
R-squared	0.036	0.172	0.008	0.085	0.035	0.147	0.007	0.052
Number of twin sets			272	272			473	473
Ln income -employed sample	0.318***	0.187**	0.224*	0.137	0.166***	0.147***	0.076	0.079
	[0.093]	[0.092]	[0.126]	[0.130]	[0.059]	[0.056]	[0.088]	[0.089]
Observations	330	330	330	330	618	618	618	618
R-squared	0.049	0.208	0.021	0.074	0.050	0.158	0.003	0.024
Number of twin sets			165	165			309	309

Panel 5: Healthy Habits Measures

Not current smoker	0.012	-0.003	0.005	0.008	-0.013	-0.031	-0.012	-0.019
	[0.028]	[0.028]	[0.031]	[0.032]	[0.024]	[0.024]	[0.025]	[0.024]
Observations	842	842	842	842	1,440	1,440	1,440	1,440
R-squared	0.030	0.064	0.000	0.024	0.009	0.042	0.000	0.018
Number of twin sets			421	421			720	720
Avoid binge drinking	-0.123**	-0.121**	-0.184***	-0.174**	-0.042	-0.061*	-0.026	-0.039
	[0.051]	[0.051]	[0.068]	[0.069]	[0.036]	[0.036]	[0.044]	[0.044]
Observations	398	398	398	398	768	768	768	768

R-squared	0.037	0.041	0.036	0.045	0.031	0.051	0.001	0.018
Number of twin sets			199	199			384	384

Doing exercise 3 times or more per week	0.061**	0.057**	0.072*	0.071*	0.081***	0.077***	0.032	0.033
	[0.024]	[0.024]	[0.037]	[0.038]	[0.020]	[0.021]	[0.029]	[0.029]
Observations	826	826	826	826	1,414	1,414	1,414	1,414
R-squared	0.016	0.030	0.010	0.031	0.018	0.026	0.002	0.012
Number of twin sets			413	413			707	707

Notes: Standard errors are clustered at twin set level. ***Significant at 1% level; **significant at 5% level; *significant at 10% level. In addition to locus of control, the baseline OLS specification includes constant term, age and age square; the extended OLS specification includes constant term, age, age square, education attainment dummies and marriage dummy; the baseline FE specification includes no other covariate, and the extended FE specification includes education attainment dummies and marriage dummy.