Online Appendix to Accompany

Genetic Variation in Preferences for Giving and Risk-Taking

David Cesarini

Christopher T. Dawes

Magnus Johannesson

Paul Lichtenstein

Björn Wallace

I. INTRODUCTION

In this online appendix, we provide some details on the Bayesian estimation procedure, additional information on recruitment bias, and the results for our two additional measures of risk preferences (referred to in the text as risk investment and risk assessment).

II. DETAILS ON ESTIMATION

A. Ordered Models

In the ordered models, the outcome variables are modelled under the assumption that y_{ij}^* is not directly observed. Instead, the observed variable y_{ij} is assumed to be one of $k+1$ ordered categories (0 to k):

(1) $$y_{ij} = 0 \;\; if \;\; y_{ij}^* \leq \tau_1,$$

(2) $$y_{ij} = 1 \;\; if \;\; \tau_1 < y_{ij}^* \leq \tau_2,$$

$$...$$

(3) $$y_{ij} = k \;\; if \;\; y_{ij}^* > \tau_k,$$

where τ_i is an unknown threshold parameter that is estimated as part of the model. For MZ twins, the probability of observing an outcome is given by:

(4) $$P(y_{ij}=0|A_i,C_i,X_i)=\Phi\left(\tau_1-(A_i+C_i)\right),$$

(5) $$P(y_{ij}=1|A_i,C_i,X_i)=\Phi\left(\tau_2-(A_i+C_i)\right)-\Phi\left(\tau_1-(A_i+C_i)\right),$$

$$...$$

(6) $$P(y_{ij}=k|A_i,C_i,X_i)=1-\Phi\left(\tau_k-(A_i+C_i)\right),$$

(7) $$0<\tau_1<...<\tau_k.$$

where Φ is the cumulative standard normal distribution. For DZ twins, the probability is:

(8) $$P(y_{ij}=0|A_{1i},A_{2ij}C_i,X_i)=\Phi\left(\tau_1-(A_{1i}+A_{2ij}+C_i)\right),$$

(9) $$P(y_{ij}=1|A_{1i},A_{2ij}C_i,X_i)=\Phi\left(\tau_2-(A_{1i}+A_{2ij}+C_i)\right)$$

$$-\Phi\left(\tau_1-(A_{1i}+A_{2ij}+C_i)\right),$$

$$...$$

(10) $$P(y_{ij}=k|A_{1i},A_{2ij},C_i,X_i)=1-\Phi\left(\tau_k-(A_{1i}+A_{2ij}+C_i)\right),$$

(11) $$0<\tau_1<...<\tau_k.$$

B. ADE Model

In the ADE model, we assume that,

(12) $$\chi_{ij}^{MZ}=A_i+D_i+E_{ij},$$

where A_i is the family genetic factor, D_i is the dominance deviation and E_{ij} is the individually-experienced unshared environment factor. For DZ twins the latent variable is a function of four random effects variables:

$$\chi_{ij}^{DZ} = A_{1i} + A_{2ij} + D_{1i} + D_{2ij} + E_{ij}, \tag{13}$$

In order to model a correlation of .25 in the DZ twins for the nonadditive (dominance) genetic effects we split up the dominance component, σ_D^2, into two independent parts, and assume that,

$$D \sim N\left(0, \sigma_D^2\right), \tag{14}$$

$$D_1 \sim N\left(0, \frac{1}{4}\sigma_D^2\right), \tag{15}$$

$$D_2 \sim N\left(0, \frac{3}{4}\sigma_D^2\right). \tag{16}$$

For the precision parameter associated with σ_D^2, we use a Pareto distribution with shape parameter equal to 1 and scale parameter equal to 0.001.

III. REPRESENTATIVENESS

In Table A1, we compare our participants to the STAGE cohort as a whole on a number of background variables. The STAGE cohort is very large, so it is important to distinguish statistical significance from practical signficance. For health, income and employment status, we find no significant differences. We do however find that our subjects are somewhat younger than the average STAGE respondent. The difference is approximately 3.5 years for men and 1.5 years for women. We also find that participants in the experiment are less likely to be unemployed. In our experimental sample, the unemployment rate is two percentage points lower than in STAGE for women, and four percentage points lower for men. Further, marriage rates are somewhat lower, a phenomenon which is no doubt related to their lower average age. In particular, 22 % of our participating men are married, as compared to 29

% in STAGE. The corresponding figures for women are 28 % and 33 %. Participants in the experiments also, on average, have 0.25 fewer children under 18 living in their household.

While the 61 % response rate in STAGE is not alarmingly low, it merits further investigation, because STAGE respondents themselves may not be fully representative of the general population. In private correspondence with the Swedish Twin Registry, we have learnt that there are no significant differences between participants and non-participants with respect to age or birthweight. As is common in twin studies, women are overrrepresented (Lykken, McGue and Tellegen, 1980) also in STAGE, with a larger fraction of non-participants being male (58% versus 44%). Non-participants are also more likely to be diagnosed with a psychological disorder (4.4% versus 7.7%) or to have at least one parent born outside Sweden (16.1% versus 12.8 %). Participants on the other hand are more likely to have studied after high-school (41 % versus 27 %).

In Table A2, we report MZ and DZ correlations on a large number of background variables for the STAGE cohort as a whole and for our experimental sample. In general, there is no tendency for the patterns of correlations to differ between the samples.

A. Data Definitions

The data from Statistics Sweden is for the year 2005 and includes income excluding capital income (förvärvskinkomst), marital status and years of education. Unlike the STAGE data, the data from Statistics Sweden is not self-reported but registry based.

Researchers interested in the variables in STAGE are advised to contact the Swedish Registry, which maintains a web-based (but password protected) database with variable definitions.

IV. ADDITIONAL RESULTS FOR RISK

In Tables A3 and A4 we report ACE model results for risk investment and risk assessment. Since the correlations we observe for risk assessment are significantly outside the permissible space of correlations, we also estimate an ADE model for risk assessment, see Table A5. The DIC model selection criterion suggests that the ADE model better fits the data. Histograms and scatterplots for risk investment and risk assessment are reported in Figures A1 and A2.

V. REFERENCES

Liang, K. and S. L. Zeger "Longitudinal data analysis using generalized linear models," *Biometrika,* 73 (1986), 13-22.

Lykken, David T., Matthew K. McGue, and Auke Tellegen, "Recruitment bias in twin research: The rule of two-thirds reconsidered," *Behavior Genetics*, 17 (1986), 343–362.

VI. ADDITIONAL TABLES AND FIGURES

Table A1.
Comparison of Experimental Sample to STAGE Cohort.

	Men			Women		
	Sample	STAGE	p-value	Sample	STAGE	p-value
Age	33.03	36.66	<0.01	35.29	36.57	<0.01
Education	13.69	12.50	<0.01	13.67	12.78	<0.01
Income	243524	269764	0.11	196591	195289	0.84
Employed Full Time	0.72	0.78	0.24	0.52	0.54	0.46
Unemployed	0.02	0.06	0.02	0.04	0.06	<0.01
Self-Employed	0.09	0.14	0.07	0.04	0.05	0.23
On Sickleave	0.02	0.02	0.93	0.03	0.04	0.43
Government Employee	0.28	0.22	0.10	0.44	0.50	0.02
Health	1.74	1.85	0.11	1.92	1.96	0.25
Marital Status	0.22	0.29	0.03	0.28	0.33	0.02
Number of Children	0.60	0.81	0.02	0.75	1.01	<0.01

Notes. Education refers to years of education. Income is the sum of wage income, taxable transfers and income from own company for the year 2005 (in SEK). Employment information was gathered when the subject responded to the STAGE questionnaire. Health is self-reported on a scale from 1 to 5. Marital status is a dummy variable taking the value 1 if the subject is married. Number of children is number of children under 18 living in the respondent's household in the year 2005. We utilized adjusted Wald tests for equality taking into account non-independence within twin families (Liang and Zeger 1986).

Table A2.
Correlations in Experimental Sample and STAGE.

	Exp. Sample		STAGE	
	MZ	DZ	MZ	DZ
Education	0.68	0.43	0.69	0.45
Income	0.69	0.58	0.59	0.45
Employed Full Time	0.25	0.21	0.28	0.29
Self Employment	0.30	0.30	0.33	0.25
On Sickleave	–0.04	-0.02	0.18	0.06
Government Employee	0.35	0.21	0.29	0.25
Health	0.46	-0.04	0.33	0.17
Marital Status	0.33	0.33	0.39	0.27
Number of Children	0.51	0.44	0.53	0.38

Notes. Education refers to years of education. Income is the sum of wage income, taxable transfers and income from own company for the year 2005 (in SEK). Employment information was gathered when the subject responded to the STAGE questionnaire. Health is self-reported on a scale from 1 to 5. Marital status is a dummy variable taking the value 1 if the subject is married. Number of children is number of children under 18 living in the respondent's household in the year 2005.

Table A3.
Results of the ACE model and its nested submodel for Risk Investment
95% credible intervals within parentheses.

		Model			
		ACE	AE	CE	E
	A	0.19 (0.01, 0.34)	0.29 (0.20, 0.39)	-	-
Continuous	C	0.10 (0.00, 0.26)	-	0.24 (0.15, 0.33)	-
	E	0.71 (0.62, 0.81)	0.71 (0.62, 0.80)	0.76 (0.67, 0.85)	1.00 (1.00-1.00)
DBar		3683	3670	3734	3988
pD		216.4	224.5	180.9	2.0
DIC		3900	3894	3915	3990
	A	0.22 (0.02, 0.38)	0.32 (0.21, 0.42)	-	-
Ordered	C	0.10 (0.01, 0.27)	-	0.26 (0.17, 0.35)	-
	E	0.68 (0.59, 0.79)	0.68 (0.58, 0.80)	0.74 (0.65, 0.83)	1.00 (1.00-1.00)
DBar		2375	2367	2431	2677
pD		221.7	226.8	182.1	4.99
DIC		2597	2593	2614	2682

Notes. A is the genetic contribution; C is the common environment contribution; E is the unique environment contribution.

DBar: Deviance.

pD: Effective number of parameters.

DIC: Bayesian Deviance Information Criterion.

Table A4.
Results of the ACE model and its nested submodel for Risk Assessment
95% credible intervals within parentheses.

		Model			
		ACE	AE	CE	E
	A	0.29 (0.14, 0.41)	0.35 (0.25, 0.44)	-	-
Continuous	C	0.05 (0.00, 0.17)	-	0.25 (0.17, 0.34)	-
	E	0.65 (0.56, 0.75)	0.65 (0.56, 0.75)	0.75 (0.66, 0.84)	1.00 (1.00-1.00)
DBar		3466	3455	3578	3844
pD		253.5	257.9	187.1	2.00
DIC		3719	3713	3765	3846
	A	0.33 (0.19, 0.45)	0.38 (0.28, 0.48)	-	-
Ordered	C	0.05 (0.00, 0.17)	-	0.28 (0.19, 0.36)	-
	E	0.62 (0.53, 0.72)	0.62 (0.53, 0.72)	0.72 (0.64, 0.81)	1.00 (1.00-1.00)
DBar		3474	3471	3604	3877
pD		279.7	279.4	204.7	9.86
DIC		3753	3751	3809	3897

Notes. A is the genetic contribution; C is the common environment contribution; E is the unique environment contribution.

DBar: Deviance.

pD: Effective number of parameters.

DIC: Bayesian Deviance Information Criterion.

Table A5.
Results of the ADE model and its nested submodel for Risk Assessment 95% credible intervals within parentheses.

		Model
		ADE
	A	0.05 (0.00-0.14)
Continuous	D	0.33 (0.19-0.44)
	E	0.63 (0.54-0.73)
DBar		3424
pD		275,8
DIC		3700
	A	0.04 (0.22-0.48)
Ordered	D	0.37 (0.22-0.48)
	E	0.59 (0.00-0.15)
DBar		3432
pD		301,1
DIC		3733

Notes. A is the genetic contribution; D is the dominance deviation; E is the unique environment contribution.

DBar: Deviance.

pD: Effective number of parameters.

DIC: Bayesian Deviance Information Criterion.

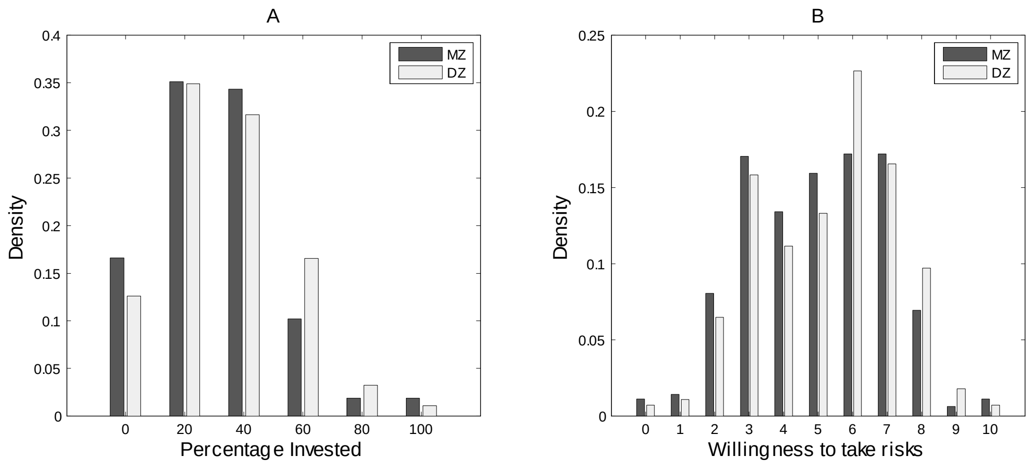

FIGURE A1.
Panel A: Risk investment (% invested), by zygosity.
Panel B: Risk assessment (0-10 scale), by zygosity.

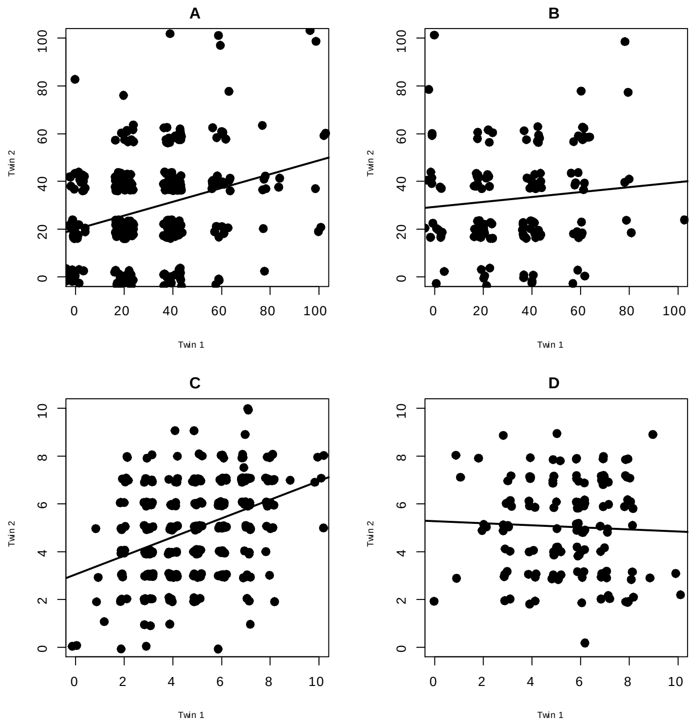

FIGURE A2. Scatterplots jittered for expositional clarity.
Panel A. Scatterplot for risk investment, percent donated, MZ twins.
Panel B. Scatterplot for risk investment, percent donated, DZ twins.
Panel C. Scatterplot for risk assessment, 0-10 scale, MZ twins.
Panel D. Scatterplot for risk assessment, 0-10 scale, DZ twins.