Internet Appendix to

Formative Experiences and Portfolio Choice: Evidence from the Finnish Great Depression*

Samuli Knüpfer
BI Norwegian Business School and CEPR

Elias Rantapuska
Aalto University

Matti Sarvimäki
Aalto University and VATT

February 15, 2016

* E-mails: samuli.knupfer@bi.no, elias.rantapuska@aalto.fi, matti.sarvimaki@aalto.fi.

1. What Is in This Internet Appendix?

Section 2 discusses the Altonji et al. (2005) and Oster (2015) methods used to assess omitted-variable bias. The tables and figures, which are arranged in the order they are referred to in the paper, appear at the end of this document. Figure IA1 plots the change in labor market conditions during the depression against the pre-depression level of labor market conditions. Table IA1 shows the results for regressions that take the pre-depression control variables from the years 1989 and 1987. Table IA2 regresses mortality on labor market conditions. Table IA3 runs logit regressions in lieu of the linear probability models. Table IA4 decomposes labor market conditions into quintile dummies. Table IA5 explains income, employment, and wealth with quintile dummies for labor market conditions. Table IA6 aggregates post-depression controls at the level of local labor markets. Table IA7 reports regressions for each of the 12 regions in the sample. Table IA8 regresses measures of portfolio risk on labor market conditions for the sample of workers who participate in the stock market. Table IA9 estimates the neighborhood regressions in the full sample without imposing the constraint on safe professions. Tables IA10 and IA11 report variants of the models used to assess effects within families.

2. Assessing Omitted-Variable Bias

Section 3.1 and 3.2 show the estimates are remarkably stable across specifications, which suggests omitted variables likely do not drive our results. Altonji et al. (2005) and Oster (2015) formalize this argument using the idea that selection on observables is informative about the likely extent of omitted-variable bias.

Applying Oster's (2015) approach to our setting, we first estimate the association between labor market conditions and investment in risky assets using a regression that includes only the region and occupation fixed effects. This regression yields a baseline estimate for the impact of labor market conditions and a baseline R^2. We compute the estimate and explanatory power also for the models that include different sets of pre-depression controls.

Oster (2015) shows the true coefficient is

$$\beta^* = \tilde{\beta} - \tilde{\delta}\frac{(\beta_0 - \tilde{\beta})(R_{max} - \tilde{R})}{\tilde{R} - R_0}. \quad \text{(A1)}$$

The estimate and explanatory power from the baseline model are denoted as β_0 and R_0, whereas the same numbers for the model with the full set of observed controls are $\tilde{\beta}$ and $\tilde{R}$. The $\tilde{\delta}$ is the relative importance of observed versus omitted variables in generating selection bias. For example, setting $\tilde{\delta}$ = 1 would mean selection on observables is as important as selection on omitted variables. The R_{max} is the maximum explanatory power of a hypothetical regression that controls for all relevant observed and unobserved factors.

Because R_{max} and δ are not known, Oster (2015) proposes a bounding approach. The first bound $\tilde{\beta}$ comes from the specification controlling for all observables. The second bound is calculated using equation (A1) by setting $\tilde{\delta}$ = 1 and R_{max} equal to the minimum of one or to the R^2 from the regression controlling for all observable factors multiplied by a factor of 1.3.[2] If the

[2] This rule of thumb is based on true randomization designs published in the *AER*, *JPE*, *QJE*, *Econometrica*, and *AEJ: Applied* in 2008-2013. Setting R_{max} to 1.3 would allow 90% of these randomized results to survive.

bounds $\tilde{\beta}$ and β^* do not include zero, the results can be considered as robust as those obtained from truly randomized experiments.

Equation (A1) illustrates the factors that influence the bound β^*. It is further away from the obtained estimate $\tilde{\beta}$ when the parameter from the full specification differs more from the baseline estimate (i.e., $\beta_0 - \tilde{\beta}$ is greater). How much the observable control variables matter in explaining the dependent variable also affects the bound (i.e., $\tilde{R}$ differs more from both R_0 and R_{max}).

Applying equation (A1) to specifications 1 and 3 in Table 2, we get

$$\beta^* = -0.676 - 1\frac{(-0.870-(-0.676))\times(\min(1,1.3\times 0.133)-0.133)}{0.133-0.075} = -0.543.$$

Therefore, the bounds Oster (2015) suggests are –0.676 and –0.543. Because these numbers are far away from zero, omitted-variable bias is unlikely to drive our conclusions. We can also use the method to ask how much more important selection on omitted variables would have to be than selection on observable controls for the true effect to equal zero. Setting $\tilde{\delta}$ at 6.93 would yield β^* = 0, suggesting selection on omitted variables would have to be 6.9 times larger than selection on observables. This calculation is an alternative way to reach the conclusion that our estimates are robust to omitted variables.

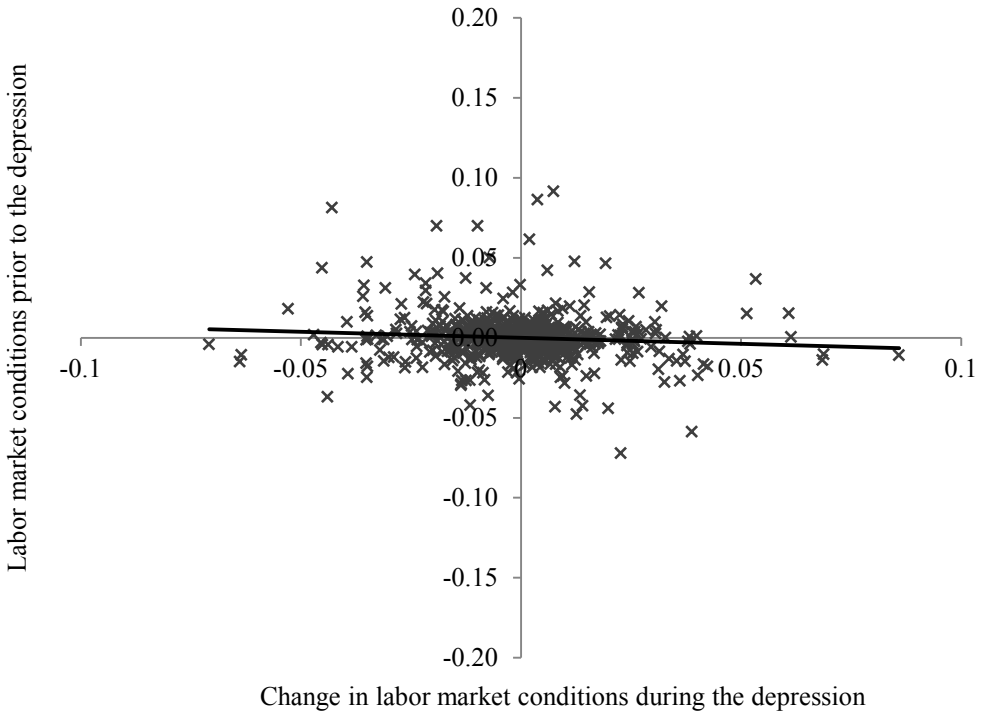

Figure IA1. Labor market conditions prior to and during the depression

For each local labor market, the graph plots the labor market conditions in 1987–90 against the change in labor market conditions in 1991–93 and 1987–90. Labor market conditions are measured as the mean share of months that workers spent in unemployment in each local labor market. The variables are demeaned by taking the residuals from a regression of labor market conditions on the region and occupation fixed effects. The line plots the fitted values from a regression of the *y*-axis on the *x*-axis. The coefficient on the *x*-variable equals –0.029 (*t*-value –0.54), and the R^2 of the regression is 0.0007.

Table IA1

Taking pre-depression variables from the year 1989 and 1987

This table reports regressions that follow the structure of Table 3 Panel A but assign the local labor market and the pre-depression control variables based on 1989 and 1987 in lieu of 1990. The regressions exclude individuals who are not employed in the year of measuring the pre-depression controls. The *t*-values reported in parentheses are robust to clustering at the level of local labor markets. The marginal effect is the coefficient multiplied by the standard deviation of the labor-market-conditions variable.

Panel A: Regressions using 1989 pre-depression controls						
Dependent variable	Stock market participation					
	Pre-depression controls			Post-depression controls		
Specification	1	2	3	4	5	6
Labor market conditions	–0.720	–0.579	–0.563	–0.287	–0.174	–0.173
	(–5.68)	(–5.80)	(–5.85)	(–4.72)	(–3.27)	(–3.24)
Sd of labor market conditions	0.041	0.041	0.041	0.041	0.041	0.041
Marginal effect	–0.030	–0.024	–0.023	–0.012	–0.007	–0.007
Adjusted R^2	0.074	0.125	0.130	0.280	0.285	0.285
Number of observations	782,451	782,451	782,451	782,451	782,451	782,451

Panel B: Regressions using 1987 pre-depression controls						
Dependent variable	Stock market participation					
	Pre-depression controls			Post-depression controls		
Specification	1	2	3	4	5	6
Labor market conditions	–0.591	–0.452	–0.438	–0.180	–0.070	–0.069
	(–4.96)	(–5.08)	(–5.06)	(–3.30)	(–1.46)	(–1.43)
Sd of labor market conditions	0.041	0.041	0.041	0.041	0.041	0.041
Marginal effect	–0.024	–0.019	–0.018	–0.007	–0.003	–0.003
Adjusted R^2	0.073	0.124	0.129	0.279	0.284	0.284
Number of observations	733,818	733,818	733,818	733,818	733,818	733,818

Table IA2

Regressions of mortality on labor market conditions

Panel A reports regressions that indicate workers who are deceased by 2005 and therefore do not appear in the sample. Panel B reports regressions of stock market participation in which deceased workers are included in the estimation sample and assumed to not participate in the stock market. The specifications correspond to columns 1-3 in Table 2. The *t*-values reported in parentheses are robust to clustering at the level of local labor markets. The marginal effect is the coefficient multiplied by the standard deviation of the labor-market-conditions variable.

Panel A: Regressions of mortality			
Dependent variable	Indicator for deceased workers		
Specification	1	2	3
Labor market conditions	0.103	0.087	0.087
	(5.22)	(4.66)	(4.76)
Sd of labor market conditions	0.041	0.041	0.041
Mean dependent variable	0.022	0.022	0.022
Marginal effect	0.004	0.004	0.004
Adjusted R^2	0.005	0.007	0.013
Number of labor market cells	817	817	817
Number of individuals	857,335	857,335	857,335

Panel B: Regressions of stock market participation assuming no stock market participation for deceased workers			
Dependent variable	Stock market participation		
Specification	1	2	3
Labor market conditions	–0.880	–0.705	–0.684
	(–6.11)	(–5.96)	(–5.99)
Sd of labor market conditions	0.041	0.041	0.041
Mean dependent variable	0.215	0.215	0.215
Marginal effect	–0.036	–0.029	–0.028
Adjusted R^2	0.074	0.127	0.131
Number of labor market cells	817	817	817
Number of individuals	857,335	857,335	857,335

Table IA3

Results using logit estimation

This table reports regressions in columns 1-3 of Table 2 using logit estimation in lieu of a linear probability model. The coefficients are reported per one-unit change of labor market conditions and the marginal effect is per one-standard-deviation worsening of labor market conditions. Due to the large number of observations and parameters, estimation of the logit regressions that include all pre- and post-depression controls is not computationally feasible. The *z*-values in parentheses are robust to clustering at the level of local labor markets.

Dependent variable	Stock market participation		
Specification	1	2	3
Labor market conditions	–0.455	–0.296	–0.275
	(–3.68)	(–3.53)	(–3.37)
Sd of labor market conditions	0.041	0.041	0.041
Marginal effect	–0.019	–0.012	–0.011
Pseudo-R^2	0.071	0.121	0.126
Number of observations	838,881	838,881	838,881

Table IA4

Breaking labor market conditions down into quintiles

This table decomposes the labor-market-conditions variable in Table 2 into quintile dummies (bottom quintile omitted). Coefficients for each quintile dummy are reported, and the *t*-values in parentheses are robust to clustering at the level of local labor markets.

Dependent variable	Stock market participation					
Specification	Pre-depression controls			Post-depression controls		
	1	2	3	4	5	6
Labor market conditions						
2nd quintile	0.008	0.007	0.006	0.006	0.004	0.004
	(0.75)	(1.07)	(1.06)	(1.20)	(0.89)	(0.81)
3rd quintile	0.001	0.005	0.005	0.006	0.005	0.004
	(0.11)	(0.72)	(0.73)	(1.14)	(1.05)	(0.81)
4th quintile	–0.024	–0.014	–0.015	–0.006	–0.004	–0.005
	(–1.86)	(–1.46)	(–1.61)	(–0.93)	(–0.68)	(–0.73)
Top quintile	–0.055	–0.036	–0.036	–0.018	–0.013	–0.013
	(–4.05)	(–3.56)	(–3.65)	(–2.53)	(–1.90)	(–2.04)
Adjusted R^2	0.075	0.128	0.133	0.282	0.287	0.291
Number of observations	838,881	838,881	838,881	838,881	838,881	838,881

Table IA5

Explaining income, employment, and wealth with quintile dummies for labor market conditions

This table reports regressions that follow the structure of Table 3 but break down labor market conditions into quintile dummies (bottom quintile omitted). The *t*-values reported in parentheses are robust to clustering at the level of local labor markets. The marginal effect is the coefficient multiplied by the standard deviation of the labor-market-conditions variable.

Dependent variable	Labor income			Months unemployed			Net worth		
Specification	1	2	3	4	5	6	7	8	9
Labor market conditions									
2nd quintile	0.024	0.015	0.014	0.110	0.102	0.096	0.009	0.016	0.014
	(0.86)	(1.37)	(1.36)	(1.93)	(2.31)	(2.29)	(0.23)	(0.74)	(0.66)
3rd quintile	–0.007	–0.003	–0.003	0.194	0.176	0.161	–0.022	0.005	0.003
	(–0.26)	(–0.28)	(–0.31)	(3.19)	(3.63)	(3.51)	(–0.52)	(0.19)	(0.13)
4th quintile	–0.051	–0.035	–0.040	0.338	0.298	0.278	–0.129	–0.064	–0.066
	(–1.50)	(–2.04)	(–2.42)	(4.71)	(5.27)	(5.06)	(–2.05)	(–1.56)	(–1.62)
Top quintile	–0.113	–0.075	–0.082	0.491	0.410	0.392	–0.288	–0.160	–0.160
	(–3.03)	(–4.00)	(–4.54)	(6.81)	(6.87)	(6.73)	(–3.99)	(–3.37)	(–3.43)
Pseudo-R^2	0.253	0.509	0.525	0.096	0.193	0.214	0.044	0.246	0.248
Number of observations	838,881	838,881	838,881	838,881	838,881	838,881	838,881	838,881	838,881

Table IA6

Aggregating post-depression controls at the level of local labor markets

This table reports regressions that follow the structure of columns 4-6 in Table 2 but aggregate the post-depression variables at the level of local labor markets. The aggregation takes the average of each post-depression variable in the region-sector-occupation cell and then decomposes each variable into decile dummies. The *t*-values reported in parentheses are robust to clustering at the level of local labor markets. The marginal effect is the coefficient multiplied by the standard deviation of the labor-market-conditions variable.

Dependent variable	Stock market participation		
Specification	Post-depression controls		
	1	2	3
Labor market conditions	–0.362	–0.350	–0.322
	(–4.97)	(–4.74)	(–3.87)
Sd of labor market conditions	0.041	0.041	0.041
Marginal effect	–0.015	–0.014	–0.013
Adjusted R^2	0.134	0.134	0.137
Number of observations	838,881	838,881	838,881

Table IA7

Regressions run for each of the 12 regions

This table repeats the regressions in Table 2 for each of the 12 regions. Within each region, the labor market conditions are calculated for each occupation. The set of controls corresponds to column 3 in Table 2, excluding the region and occupation fixed effects. The *t*-values reported in parentheses are robust to clustering at the occupation level. The marginal effect is the coefficient multiplied by the standard deviation of the labor-market-conditions variable.

Panel A: Regions 1 to 6						
Dependent variable	Stock market participation					
Region number	1	2	3	4	5	6
Labor market conditions	–0.785	–0.567	–0.603	–0.566	–0.531	–0.456
	(–8.09)	(–7.20)	(–7.13)	(–7.28)	(–5.52)	(–4.76)
Mean participation rate	0.332	0.160	0.194	0.217	0.178	0.186
Sd of labor market conditions	0.035	0.041	0.039	0.041	0.041	0.042
Marginal effect	–0.028	–0.023	–0.024	–0.023	–0.022	–0.019
Adjusted R^2	0.125	0.076	0.096	0.093	0.092	0.080
Number of labor market cells	69	70	69	68	68	67
Number of observations	253,376	106,224	81,302	77,433	59,158	51,855

Panel B: Regions 7 to 12						
Dependent variable	Stock market participation					
Region number	7	8	9	10	11	12
Labor market conditions	–0.416	–0.379	–0.450	–0.545	–0.476	–0.394
	(–5.41)	(–5.15)	(–6.25)	(–6.30)	(–5.50)	(–5.44)
Mean participation rate	0.138	0.148	0.150	0.179	0.122	0.136
Sd of labor market conditions	0.042	0.040	0.044	0.041	0.044	0.043
Marginal effect	–0.018	–0.015	–0.020	–0.023	–0.021	–0.017
Adjusted R^2	0.079	0.077	0.087	0.077	0.079	0.079
Number of labor market cells	69	67	68	68	67	67
Number of observations	49,240	40,001	33,663	31,912	30,443	24,274

Table IA8

Regression of portfolio risk conditional on participation

This table restricts the sample to workers who participate in the stock market and estimates the impact of labor market conditions on the return volatility, beta, and idiosyncratic risk of their financial portfolios. Volatility is based on 36 observations of monthly portfolio returns. Beta comes from regressions of the monthly portfolio return on the euro-denominated MSCI Europe index or the euro-denominated MSCI World index, whereas idiosyncratic risk measures the share of the total portfolio variance not attributable to the variance in the index. The *t*-values reported in parentheses are robust to clustering at the level of local labor markets. The marginal effect is the coefficient multiplied by the standard deviation of the labor-market-conditions variable.

Panel A: Pre-depression controls					
Dependent variable	Volatility	Beta, MSCI Europe	Idiosyncratic risk, MSCI Europe	Beta, MSCI World	Idiosyncratic risk, MSCI World
Specification	1	2	3	4	5
Labor market conditions	0.133	0.995	0.132	1.042	0.049
	(2.85)	(3.62)	(1.19)	(4.00)	(0.46)
Sd of labor market conditions	0.035	0.035	0.035	0.035	0.035
Mean dependent variable	0.183	1.247	0.613	1.109	0.689
Marginal effect	0.005	0.034	0.005	0.036	0.002
Adjusted R^2	0.085	0.108	0.019	0.092	0.022
Number of labor market cells	785	785	785	785	785
Number of individuals	184,246	184,246	184,246	184,246	184,246

Panel B: Post-depression controls					
Dependent variable	Volatility	Beta, MSCI Europe	Idiosyncratic risk, MSCI Europe	Beta, MSCI World	Idiosyncratic risk, MSCI World
Specification	1	2	3	4	5
Labor market conditions	0.090	0.765	0.091	0.830	0.022
	(1.94)	(2.92)	(0.79)	(3.36)	(0.21)
Sd of labor market conditions	0.035	0.035	0.035	0.035	0.035
Mean dependent variable	0.183	1.247	0.613	1.109	0.689
Marginal effect	0.003	0.026	0.003	0.029	0.001
Adjusted R^2	0.149	0.156	0.037	0.134	0.034
Number of labor market cells	785	785	785	785	785
Number of individuals	184,246	184,246	184,246	184,246	184,246

Table IA9

Estimates of neighborhood influences in full sample

This table estimates regressions in Table 8 without imposing the restriction on safe professions. The *t*-values reported in parentheses are robust to clustering at the level of local labor markets. The marginal effect is the coefficient multiplied by the standard deviation of the labor market conditions.

Analysis	Falsification exercise	Post-depression labor market outcomes and wealth accumulation			Risky investment
Dependent variable	Stock market part.	Labor income	Months unemployed	Net worth	Stock market part.
Specification	1	2	3	4	5
Neighbor's labor market conditions	–0.038	–0.692	5.608	–0.463	–0.320
	(–0.46)	(–6.72)	(19.37)	(–2.57)	(–2.28)
Mean dependent variable	0.070	26.750	11.247	16.088	0.220
Sd of labor market conditions	0.014	0.014	0.014	0.014	0.014
Marginal effect	–0.001	–0.0095	0.077	–0.006	–0.004
Pseudo-R^2 / Adjusted R^2	0.041	0.528	0.215	0.250	0.138
Number of individuals = 84,409					

Table IA10

Robustness checks on sibling analyses

This table estimates regressions in column 5 of Panel A in Table 7 using samples of individuals whose sibling is younger or older than the individual. The *t*-values reported in parentheses are robust to clustering at the level of local labor markets. The marginal effect is the coefficient multiplied by the standard deviation of the labor market conditions.

Dependent variable	Stock market participation	
Specification	Sibling younger than individual	Sibling older than individual
	1	2
Personal labor market conditions	–0.657	–0.460
	(–5.16)	(–5.03)
Sibling's labor market conditions	–0.123	–0.141
	(–1.76)	(–2.01)
Mean dependent variable	0.224	0.215
Sd of personal labor market conditions	0.041	0.041
Sd of siblings' labor market conditions	0.042	0.041
Marginal effect for personal conditions	–0.027	–0.019
Marginal effect for siblings' conditions	–0.005	–0.006
Adjusted R^2	0.138	0.136
Number of individuals	223,639	239,813

Table IA11

Robustness checks on intergenerational effects

Columns 1 and 2 in this table estimate regressions in column 5 of Panel B in Table 7 using samples of children born in 1978-82 (younger) and in 1972-77 (older). Column 3 replaces the father's labor market conditions and control variables with those measured for the mother. The *t*-values reported in parentheses are robust to clustering at the level of local labor markets. The marginal effect is the coefficient multiplied by the standard deviation of the labor market conditions.

Dependent variable	Stock market participation		
Specification	Younger children	Older children	Mother's labor market experiences
	1	2	3
Father's labor market conditions	–0.291	–0.153	
	(–4.62)	(–2.34)	
Mother's labor market conditions			–0.064
			(–0.82)
Mean participation rate	0.129	0.177	0.154
Sd of labor market conditions	0.046	0.046	0.028
Marginal effect	–0.013	–0.007	–0.002
Adjusted R^2	0.089	0.085	0.086
Number of individuals	199,173	206,359	343,995