Risk Attitudes and Household Migration Decisions

Christian Dustmann

Francesco Fasani

Xin Meng

Luigi Minale

SUPPLEMENTARY MATERIAL

A. Online Appendix

A1. Sample selection

The fact that risk aversion is only observed for individuals who were present at home at the interview may bias our estimates if unobservables in the interview participation equation are correlated with risk aversion, conditional on observables.[33] We perform two empirical tests to address this concern. We first estimate a sample selection model using family events as instruments to identify presence at interview, and we then directly compare the distribution of risk attitudes of migrants in our estimation sample with that among migrants interviewed in urban areas.

We estimate the following sample selection model:

$$wtRisk^{*}_{ihk} = \mathbf{X'}_{ihk}A + \mathbf{W'}_{hk}B + \eta_k + e_{ihk} \quad (A.1)$$

$$int_{ihk} = 1[y^{*}_{ihk} = \mathbf{X'}_{ihk}C + \mathbf{W'}_{hk}D + \boldsymbol{Z'_{ihk}}E + \mu_k + u_{ihk} \geq 0] \quad (A.2)$$

where $wtRisk^{*}_{ihk}$ is the latent willingness to take risk and int_{ihk} is a dummy equal one if the individual i was at home at the interview (i.e. $int_{ihk} = 1$ if the latent variable $y^{*}_{ihk} \geq 0$), so that:

$$wtRisk_{ihk} = \begin{cases} wtRisk^{*}_{ihk} & if \quad int_{ihk} = 1 \\ not\ observed & if \quad int_{ihk} = 0 \end{cases} \quad (A.3)$$

The vectors $\mathbf{X'}_{ihk}$ and $\mathbf{W'}_{hk}$ collect the same observable individual-level covariates and family characteristics as in our main outcome equation (11), and $\eta_k$and μ_k are county fixed effects. The selection equation is non-parametrically identified by the variable vector $\boldsymbol{Z}_{ihk}$ that includes major events in the families of interviewees - such as illnesses and deaths - that occurred in the months immediately before or after the interview. We gather information about events from both the 2009 and the 2010 surveys, relying on the longitudinal nature of the RUMiC survey. The events recorded in the 2009 survey refer to year 2008 and, therefore, took place *before* the survey collected data on respondents' risk attitudes. The 2010 survey provides information on events that occurred *after* the 2009 survey and on their exact timing. We can

[33] Within the migrant population, individuals absent at the survey are more likely to be males, younger and less likely to be married.

thus distinguish events taking place immediately after the 2009 survey (i.e. 1-7 months after the survey) from those occurring at a later stage (i.e. 8-16 month after). These events, while arguably uncorrelated with risk attitudes, are largely unanticipated and may have induced the individual to return to the home village, or to remain longer at home, hence affecting the probability of being at home at the time of the interview. In particular, future death and illness events may increase the probability of being interviewed if they are associated with current health status of relatives and their demand for care (Giles and Mu, 2007). To test for selection, we estimate equation (A.2) using a probit model and construct the generalised residuals which we include in equation (A.1) (see Heckman 1979). A test of correlation between the unobservables determining participation and individual risk aversion corresponds to a simple t-test of whether the coefficient of the generalised residual is significantly different from zero.

We report probit estimates of the first stage in the upper panel of Table A2, where the dependent variable is the probability of being at home for the interview. The first four columns report estimates where we use as instruments death and illness events that occurred both before and after the interview. We find that past and future death and illness events are a strong predictor – both separately (columns 1 and 2) and jointly (column 3) – of migrants' presence at home at the time of the survey. Events taking place further ahead in time are still positively correlated with the probability of reporting risk attitudes, but the coefficient is about half the size of events which occurred closer to the time of the survey and is not significant (column 4). Although our preferred instruments are "illness or death" events – as they represent unexpected and unplanned family events – in the last two columns of Table A2 we include also fertility events relative to year 2008 (columns 5 and 6). The inclusion of these additional events do not affect estimated coefficients and levels of significance of death and illness events. In all specifications, we condition on individual and household controls and on county fixed effects.

In the lower panel of Table A2, we report the estimated coefficient on the generalised residuals (or inverse Mills ratios) that is included in equation A.1. For each column, the inverse Mills ratio is computed using the instrument(s) reported in the upper panel of the table. The estimated coefficient is small in magnitude (ranging from 1.4 and 2.7 percent of the average

value of wtRisk in our sample) and not statistically different from zero in any of the specifications, with a coefficient/standard error ratio that is never larger than 0.35. Thus, conditional on observables, individual risk attitudes are not correlated with unobservables that determine participation in the survey.

We further compare the distribution of risk attitudes in our estimation sample of migrants (surveyed in rural areas) with that among migrants interviewed in urban areas (using data from the urban module of the RUMiC Survey). We regress individual willingness to take risks (wtRisk) on gender, age, age squared and years of education and we compute the residuals. Figure A 1 shows that the distribution of residual willingness to take risks among migrants from our rural household survey is slightly dominated by that of migrants from the urban sample. The vertical lines identify the average values or residual willingness to take risks for the rural (-0.337) and urban (0.041) samples[34]: the null hypothesis of the difference in means being equal to zero is rejected at the 5% level of significance. These differences in risk attitudes suggest that, in our estimation sample, we may be slightly oversampling more risk averse individuals from the population of migrants. This, however, would reduce differences in risk attitudes between migrant and non-migrant household members and, if anything, work against our main empirical findings, so that our estimates can be interpreted as lower bounds.

Note that both tests discussed in this section deal with the potential concern that the Great Recession – which disproportionately impacted on rural-urban migrants (Kong, Meng and Zhang, 2009) – created a correlation between migrants' risk attitudes and the probability of being present at the time of the interview by increasing the propensity to return home of those laid off (who may differ in unobserved traits from those who were not laid off). We further address this concern by exploring information on the latest non-farm job which is available in the survey for all migrants, regardless of whether they are present at the interview or not (and answered by proxy respondents in the latter case). We find that the vast majority of migrants (almost 75%) were employed in urban areas before returning home for the Chinese New Year

[34] The unconditional averages of willingness to take risks for migrants from the urban and from the rural sample respectively are 4.15 and 3.6.

celebrations in early 2009, and that this share is similar among migrants who were present at the time of the survey and those who were not present.

A2. Theoretical Framework – Household Migration Decisions

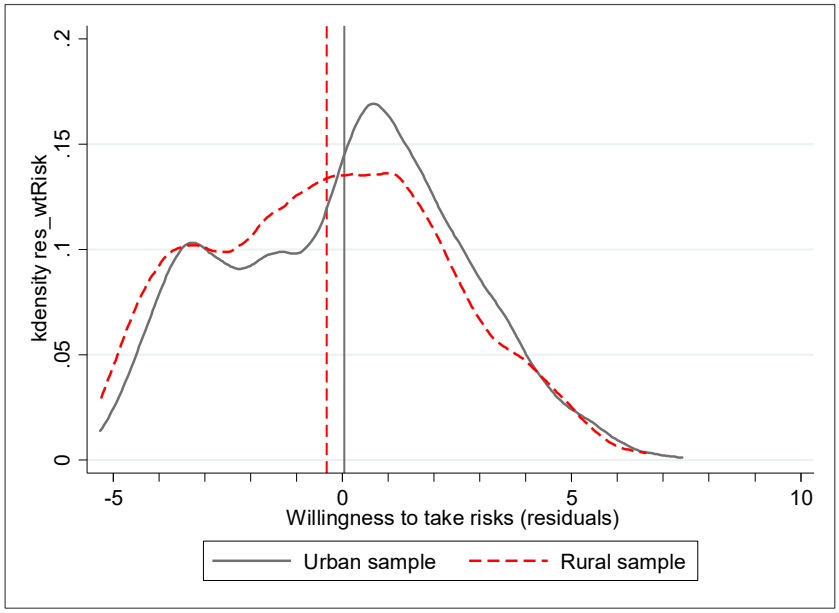
Consider two households A and B with the same within household variance in risk attitudes ($k_1^A - k_2^A = k_1^B - k_2^B$), but household B has higher average risk aversion than household A ($\bar{k}^A < \bar{k}^B$). Which household gains the more from a migration depends on the difference in utility gains:

$$\Delta U^{h_B} - \Delta U^{h_A} = \left(\bar{k}^{h_A} - \bar{k}^{h_B}\right)(\Delta V(\tilde{y}^{NM}) + \Delta V(\tilde{y}^{M}))$$

For the case where $\Delta V(\tilde{y}^{NM}) < 0$ and $\Delta V(\tilde{y}^{M}) > 0$ (i.e. a migration reduces income variance for the non-migrant, but increases it for the migrant), the sign of this expression depends on whether $\Delta V(\tilde{y}^{NM}) + \Delta V(\tilde{y}^{M})$ is positive or negative. Thus, when regressing the probability that a household sends a migrant on household's dispersion in risk aversion and household's average risk aversion, the sign of the latter is ambiguous.

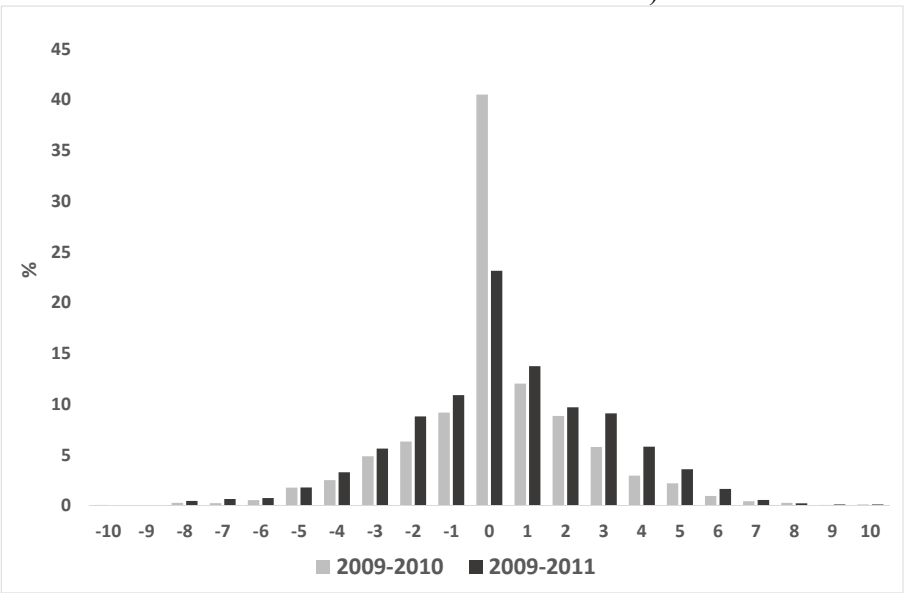
Appendix Figures

FIGURE A 1 – DISTRIBUTION OF (RESIDUAL) WILLINGNESS TO TAKE RISKS AMONG MIGRANTS IN RURAL AND URBAN SAMPLES



Note. The figure shows the distribution of residual willingness to take risks for migrants from the urban and the rural RUMiC sample, respectively. The vertical lines identify the average values or residual willingness to take risks for the rural (-0.337) and urban (0.041) samples. Residuals are obtained by regressing individual willingness to take risks on gender, age, age squared and years of education.

FIGURE A2 – DISTRIBUTION OF CHANGES IN SELF-REPORTED WILLINGNESS TO TAKE RISKS (2009, 2010 AND 2011 RUMIC-RHS WAVES)



Note: The sample is composed of 4,067 individuals from our estimating sample who reported *wtRisk* in both the 2009 and 2010 waves (grey bars) and of 2,563 individuals that reported *wtRisk* in all three waves (2009, 2010 and 2011; black bars).

Appendix Tables

Table A1. – SAMPLE OF INDIVIDUALS IN RELEVANT HOUSEHOLDS VS ENTIRE SAMPLE

	Entire sample		At least 2 individuals reporting wtRisk in the HH		
	Mean	**Std. Dev.**	**Mean**	**Std. Dev**	**p-values**
	Panel A - Individuals				
Male	0.51	0.50	0.51	0.50	0.97
Age	40.4	12.16	40.6	12.15	0.33
Married	0.82	0.38	0.83	0.37	0.04
Years of schooling	7.42	2.76	7.44	2.78	0.54
Birth order	2.20	1.31	2.16	1.28	0.06
Number of siblings	3.02	1.64	2.97	1.61	0.06
Number of child	1.49	1.08	1.47	1.06	0.09
Migrated last year	0.22	0.41	0.20	0.40	0.00
Ever migrated	0.35	0.48	0.32	0.47	0.00
Number of observations	17,658		7,808		
	Panel B - Households				
Household size	4.07	1.30	4.08	1.32	0.67
HH members aged < 16	0.58	0.74	0.57	0.73	0.33
HH members in working age	2.94	1.10	2.89	1.09	0.05
HH members aged > 60	0.24	0.52	0.34	0.61	0.00
HH head's education (years)	7.53	2.38	7.25	2.58	0.00
Plot size (Mu, 15 Mu = 1 hectare)	4.49	4.64	4.15	4.08	0.00
House value per capita (Yuan, in logs)	9.12	1.32	9.16	1.33	0.15
Number of households	6,425		2,961		

Note. The table compares characteristics of individuals in households in which more than one member in working age has reported risk attitudes with those of individuals in other households. The last column reports p-values for the test of difference in means. Source: 2009 RUMiC –RHS Survey.

Table A2. - SAMPLE SELECTION

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A - Outcome = Reporting wtRisk (0/1)						
Timing of events relative to 2009 survey	Before	After	Before & after	Before & after	Before	Before & after
Illness/Death in 2008 (3-12 months before)	0.017***		0.019***	0.018***	0.017***	0.019***
	(0.007)		(0.007)	(0.007)	(0.007)	(0.006)
Illness/Death in 2009 (1-8 months after)		0.018**	0.017*	0.017*		0.016*
		(0.009)	(0.010)	(0.009)		(0.010)
Illness/Death in 2010 (9-16 months after)				0.007		
				(0.012)		
Pregnancy/Birth 2008 (3-12 months before)					0.013**	0.012*
					(0.006)	(0.007)
Observations	6,609	6,417	6,294	6,294	6,609	6,294
Pseudo R-squared	0.64	0.64	0.64	0.64	0.64	0.64
Chi2 (x,N)	4.50	2.28	8.15	8.36	7.19	10.31
Prob > chi2	0.03	0.13	0.02	0.04	0.03	0.02
Panel B - Outcome = wtRisk						
Inverse Mill's	0.050	0.070	0.053	0.061	0.036	0.038
	(0.191)	(0.196)	(0.198)	(0.198)	(0.192)	(0.199)
Observations	5,232	5,092	4,977	4,977	5,232	4,977
R-squared	0.258	0.251	0.253	0.253	0.258	0.253
F-stat (Inverse Mills)	0.07	0.13	0.07	0.09	0.03	0.04
Prob > F (Inverse Mills)	0.79	0.72	0.79	0.76	0.85	0.85

Note. Panel A of the table reports marginal effects from probit regressions of a dummy that equals one if individuals reported risk attitude during the 2009 survey on indicators for a number of major life events having occurred to them and/or their relatives *before* and *after* the 2009 survey. We define indicators for the following events or combinations of them: at least one illness or one death in 2008 (columns 1 and 3-6); at least one illness or one death in 2009 (columns 2-4 and 6); at least one illness or one death in 2010 (column 4); at least one pregnancy/birth in 2008 (columns 5-6). Events occurred in 2008 (before the 2009 survey) are measured with the 2009 survey, while events occurring in 2009 or 2010 (after the 2009 survey) are measured with the 2010 survey. The sample includes all individuals in the labour force (i.e. aged between 16 and 60 and not currently in school or disabled) who live in households in which more than one member in the labour force reported risk attitudes. Panel B reports estimates from OLS regressions of wtRisk on the estimated Inverse Mill's Ratio and other controls. For each column, the inverse Mills ratio is computed using the instrument(s) reported in the upper panel of the table. All regressions include individual controls (age, age squared, a dummy for male, years of education, a dummy for being married, relation with the HH head dummies, number of siblings, order of birth, and number of children) household controls (household size and structure (number of family members under 16, in the labour force, and older than 60); and per capita house value (in logs)) and county fixed effects. Robust standard errors are clustered at the household level and reported in brackets. *** p<0.01, ** p<0.05, * p<0.1.

Table A3. – INDIVIDUAL MIGRATION DECISION: PROBIT AND LOGIT ESTIMATES (MARGINAL EFFECTS)

	Probit				Logit			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
wtRisk	0.013***	0.004**	0.004**	0.004**	0.014**	0.004*	0.003**	0.004**
	(0.004)	(0.002)	(0.002)	(0.002)	(0.005)	(0.002)	(0.002)	(0.002)
Basic individual controls		X	X	X		X	X	X
Additional individual controls			X	X			X	X
Household controls				X				X
County fixed effects	X	X	X	X	X	X	X	X
Observations	6,332	6,332	6,103	5,992	6,332	6,332	6,103	5,992
R-squared	0.232	0.386	0.400	0.409	0.234	0.391	0.404	0.414

Note. The table shows the marginal effects derived using the probit (columns 1-4) and logit (columns 5-8) estimators of an individual indicator for migrants (in the previous year) on individual willingness to take risk (*wtRisk*) and other controls. The *wtRisk* variable measures individual willingness to take risks (decreasing with risk aversion) and has a mean of 2.57 and a standard deviation of 2.36. The basic individual controls are age, age squared, a dummy for male, and years of education; the additional individual controls are a dummy for married, a dummy for relation to head of household, order of birth, number of siblings, and number of children; and the household controls are household size and structure (number of family members under 16, in the labour force, and older than 60); and per capita house value (in logs). All regressions include 82 county fixed effects. The sample includes all individuals in the labour force (i.e. aged between 16 and 60 and not currently in school or disabled) who live in households in which more than one member in the labour force has reported risk attitudes. Robust standard errors are clustered at the household level and reported in brackets. *** p<0.01, ** p<0.05, * p<0.1.

Table A4.– INDIVIDUAL MIGRATION DECISION: FULL SPECIFICATION

	Migrated last year				Ever migrated			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
wtRisk	0.014***	0.005***	0.005***	0.005***	0.030***	0.014***	0.014***	0.014***
	(0.0018)	(0.0019)	(0.0019)	(0.0019)	(0.0025)	(0.0027)	(0.0028)	(0.0028)
Male dummy		0.060***	0.081***	0.086***		0.135***	0.132***	0.132***
		(0.0064)	(0.0145)	(0.0147)		(0.0087)	(0.0178)	(0.0179)
Age		-0.021***	-0.005	-0.004		-0.020***	0.005	0.003
		(0.0040)	(0.0051)	(0.0053)		(0.0046)	(0.0059)	(0.0060)
Age squared*100		0.015***	0.001	-0.000		0.009*	-0.015**	-0.013**
		(0.0044)	(0.0055)	(0.0056)		(0.0052)	(0.0063)	(0.0063)
Years of education		0.001	0.001	0.001		0.002	0.002	0.002
		(0.0014)	(0.0014)	(0.0014)		(0.0020)	(0.0020)	(0.0020)
Married		-0.032	-0.044	-0.055*		-0.013	-0.032	-0.035
		(0.0274)	(0.0289)	(0.0286)		(0.0291)	(0.0305)	(0.0310)
Order of birth			-0.002	-0.003			-0.002	-0.002
			(0.0028)	(0.0027)			(0.0043)	(0.0043)
# of siblings			0.002	0.001			0.003	0.003
			(0.0027)	(0.0027)			(0.0040)	(0.0040)
# of children			-0.002	0.001			-0.008	-0.005
			(0.0059)	(0.0057)			(0.0079)	(0.0080)
# HH members below age 16				0.009				0.008
				(0.0059)				(0.0083)
# HH members in work force				-0.005				-0.011*
				(0.0040)				(0.0059)
# HH members above age 60				-0.004				0.009
				(0.0094)				(0.0130)
Ln (p.c. house value)				-0.002				0.002
				(0.0034)				(0.0049)
Relationship with HH head dummies			X	X			X	X
County fixed effects	X	X	X	X	X	X	X	X
Observations	6,332	6,332	6,103	5,992	6,280	6,280	6,052	5,946
R-squared	0.187	0.288	0.305	0.310	0.148	0.273	0.288	0.292

Note. The table reports estimates from LPM regressions of a dummy for individual migration status on individual willingness to take risk (*wtRisk*) and other controls. The migration status dummy equals one if the individual migrated for working reasons in the year before the interview (columns 1–4) or has ever migrated for working reasons (columns 5–8). The *wtRisk* variable measures individual willingness to take risks (decreasing with risk aversion) and has mean of 2.57 and a standard deviation of 2.36. The basic individual controls are age, age squared, a dummy for male, years of education, and a dummy for married; the additional individual controls are: relation with HH head dummies, order of birth, number of siblings, and number of children; and the household controls are household size and structure (number of family members under 16, in the labour force, and older than 60); and per capita house value (in logs). All regressions include 82 county fixed effects. The sample includes all individuals in the labour force (i.e. aged between 16 and 60 and not currently in school or disabled) who live in households in which more than one member in the labour force has reported risk attitudes. Robust standard errors are clustered at the household level and reported in brackets. *** p<0.01, ** p<0.05, * p<0.1.

Table A 5 - INDIVIDUAL MIGRATION DECISION: VILLAGE CONTROLS AND ALTERNATIVE CLUSTERING

	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: Migrated last year					
wtRisk	0.005**	0.005**	0.006**	0.009***	0.006*	0.009*
	(0.0023)	(0.0023)	(0.0023)	(0.0033)	(0.0030)	(0.0049)
migr_rate_village			0.092*		0.092	
			(0.0512)		(0.0562)	
Observations	5,992	5,902	5,902	5,992	5,902	5,992
R-squared	0.310	0.312	0.312	0.409	0.312	0.409
	Panel B: Ever migrated					
wtRisk	0.014***	0.014***	0.014***	0.016***	0.014***	0.016***
	(0.0033)	(0.0033)	(0.0033)	(0.0038)	(0.0029)	(0.0042)
migr_rate_village			0.151*		0.151*	
			(0.0882)		(0.0809)	
Observations	5,946	5,856	5,856	5,946	5,856	5,946
R-squared	0.292	0.298	0.299	0.444	0.299	0.444
Individual and HH controls	X	X	X	X	X	X
Village controls		X	X		X	
FE	County	County	County	Village	County	Village
Clustering SE	Village	Village	Village	Village	County	County

Notes. The table reports estimates from LPM regressions of a dummy for individual migration status (migrated last year in panel A and ever migrated in panel B) on individual willingness to take risk (*wtRisk*) and other controls. *Individual and HH controls*: age, age squared, a dummy for male, years of education, a dummy for married, relation with HH head dummies, order of birth, number of siblings, and number of children; household size and structure (number of family members under 16, in the labour force, and older than 60) and per capita house value (in logs). *Village controls*: total population, share of employment by sector (agriculture, manufacture, construction and services), average yearly income, presence of a school in the village. In columns 3 and 5 we condition on the village migration rate. County fixed effects are included in columns 1-3 and 5; village fixed effects are included in columns 4 and 6. Standard errors are clustered at the village level (approximately 580 clusters) in columns 1-4 and at the county level in columns 5-6 and are reported in brackets: *** p<0.01, ** p<0.05, * p<0.1.

Table A6.– MIGRATION DURATION AND RISK ATTITUDES (RUMIC URBAN SURVEYS)

	OLS	Fixed Effects
	(1)	(2)
Years since first migration	0.006	-0.002
	(0.004)	(0.007)
Individual controls	X	X
Year and city dummies	X	X
Individual fixed effects		X
Observations	22,208	22,208
R-squared	0.132	0.013

Note. This table tests the relationship between attitude toward risks and the length of migration experience. It reports estimates of wtRisk on years since first migration for a sample of migrants living in urban areas. Individual controls are age, age squared, a dummy for male, years of education, a dummy for married and the number of cities the individual has migrated to. OLS estimates are presented in column 1 while estimates including individual fixed effects are presented in column 2. The sample is an unbalanced panel of rural-urban migrants living in urban areas coming from six waves (from 2008 to 2013) of the urban module of the RUMiC Survey. Robust standard errors are clustered at the household level and reported in brackets. *** p<0.01, ** p<0.05, * p<0.1.

Table A 7 - ACROSS HOUSEHOLD MIGRATION DECISION - SHARE OF MIGRANTS

		(1)	(2)	(3)	(4)
HH_avg_wtRisk		0.005**	0.001		
		(0.002)	(0.002)		
HH_range_wtRisk			0.009***		
			(0.002)		
HH_max_wtRisk				0.008***	0.012***
				(0.002)	(0.003)
HH_oth_wtRisk				-0.005	-0.010**
				(0.003)	(0.005)
HH controls		X	X	X	X
Labour force age range:	**16-60**	X	X	X	
	16-50				X
Observations		2,961	2,961	2,961	2,189
R-squared		0.281	0.285	0.284	0.374

Note. The table reports estimates from OLS regressions of the share of household members in the labour force sent as migrants on different household-level measures of willingness to take risks and other controls. The variables HH_avg_wtRisk and HH_range_wtRisk measure the average and the range of willingness to take risks in the household, respectively. The variables $HH\_\max\_wtRisk_{hk}$ and $HH\_\text{oth}\_wtRisk_{hk}$ measure the risk preferences of the individual with the highest willingness to take risks in the household and the average risk attitudes among the other household members, respectively. Columns 1 and 2 replicate columns 3 and 4 in Table 8; while columns 3 and 4 replicate columns 4 and 8 in Table 9. *HH controls*: household size and structure (number of family members under 16, in the labour force, and older than 60); per capita house value (in logs); size of the family plot; and the years of education and age of the head of household. All specifications include 82 county fixed effects. In columns 1–3, the age bracket for workers to be considered part of the labour force is 16–60; in column 4 it is 16–50. The sample includes all households in which at least two individuals have reported risk attitudes, and at least one of these is in the labour force (i.e. aged between 16 and 60 and not currently in school or disabled). Robust standard errors are reported in brackets. *** p<0.01, ** p<0.05, * p<0.1.