Networks Abroad and Culture: Global Individual-Level Evidence

Riccardo Turati - Author's mail: riccardo.turati@uab.cat

Appendix – For Online Publication

A General Theoretical Framework

The general theoretical framework borrows its architecture from the discrete choice model proposed in Desmet and Wacziarg (2021). The simplified model describes a society with one identity cleavage k (in this case, connectedness) and one cultural trait (e.g., religiosity). As described in Section 2, the identity cleavage includes only two identity traits c and $-c$ (i.e., connected and not connected), and the cultural trait has only to values i and j (e.g., religious and not religious), which are unequally distribute in the society. I define i as the value hold by the majority, while j the value hold by the minority group. To simplify the description of the model, I assume that the identity groups have equal size, that time is discrete, and each trait is imperfectly transmitted across generations. Moreover, through interaction with other individuals, an agent can change her cultural value if, through the emergence of cultural innovation, one value become more socially acceptable.[34]

I can now formally describe agent's discrete choice model over the value choice at a given time t. Given her identity trait $k \in \{c, -c\}$ and cultural value $x \in \{i, j\}$, the agent decides x' by solving the following problem:

$$u(x,k) = max_{x' \in \{i,j\}} \Big\{ I(i)s^i, I(j)\frac{1}{r_k}s^j \Big\}, \tag{A-1}$$

where $I(i) = 1$ if $x' = i$ and zero otherwise, and $I(j) = 1$ if $x' = j$ and zero otherwise. The share of individuals holding value i and j in the identity group k are defined as s^i and s^j, respectively, and they are taken as given by the agent. Additionally, each agent draws a random variable r from a uniform distribution with support $[0,\ 1/\bar{r}_k]$, where $\bar{r}_k$ is specific to each identity trait. An higher $\bar{r}_k$ is related to higher likelihood to gain utility from choosing the minority value j, since it becomes more socially acceptable. Once r is drawn, the agent decides the cultural value which maximize her utility.

By having a connection abroad, connected agents have access to a broader set of information and values, which can make the minority trait j more interesting and acceptable. From a formal point of view, I then expect that $\bar{r}_c > \bar{r}_{-c}$. Hence, by acting as source of *cultural innovation*, connections abroad influence agent's decision problem, increasing the likelihood of holding the minority trait j compared to a not connected individuals, holding all the other things equal. The model can be further extended by accounting that the contribution of the connection c can be a function of the

[34] In the original Desmet and Wacziarg (2021) model, agents are sensitive to the majority value, hence if she holds the minority value, she may change over time due to conformity. However, I simplify this potential mechanism.

culture of his/her country of residence d, hence $r_{c(d)}$.

Nonetheless, the *choice of identity* trait k is not necessarily exogenous. Individuals can indeed pick up the identity trait that better fits their preferences. In that case, the discrete choice problem of an agent holding the cultural value x becomes:

$$u(x) = max\Big\{u(x, c), u(x, -c)\Big\}, \tag{A-2}$$

where $u(x, c)$ and $u(x, -c)$ are the discrete choice problem described in equation (A-1) and faced by an agent that decide to have a connection abroad or not having it.

Both cultural innovation and the choice of identity can contribute to the formation of a distinct set of cultural values associated to connected individuals. While both channels can describe the presence of a distinctive set of values and preferences associated to having a network abroad, the main testable empirical hypothesis of this study is the following one:

Hypothesis: Individuals with a network abroad are characterized by a distinct set of values and preferences compared to alike individuals.

B Additional Results

Table B-1: Connectedness and Cultural Traits - Quality of Connectedness tie

	Social Behavior		*Religiosity*		*Gender-Egalitarian*	
	OLS 2009-2012 (1)	OLS 2009-2012 (2)	LPM 2009-2012 (3)	LPM 2009-2012 (4)	OLS 2009-2011 (5)	OLS 2009-2011 (6)
$Network^{Cl}$	0.191***		0.012***		0.039***	
	(0.006)		(0.002)		(0.012)	
$FamilyNet.$		0.159***		0.010		0.086**
		(0.018)		(0.013)		(0.039)
$Interaction$	0.077***	0.157***	-0.006	0.006	0.071*	0.039
	(0.014)	(0.030)	(0.006)	(0.019)	(0.039)	(0.085)
Observations	403509	411355	384137	391893	87755	90238
Regions	2095	2095	2064	2064	736	736
Adj. R-squared	0.19	0.18	0.39	0.39	0.24	0.24
Individual Controls	✓	✓	✓	✓	✓	✓
Region-Year FE	✓	✓	✓	✓	✓	✓

Note: Author's calculations on Gallup World poll data. Standard errors are clustered at the regional level. * p<0.1, ** p<0.05, *** p<0.01. The dependent variable is: social-behavior index (col. (1)-(2)), religiosity dummy (col. (3)-(4)) and gender-egalitarian index (col. (5)-(6)). The set of individual controls includes dummies for education, gender, marital status, parental status, living in an urban area, employment status and continuous variables for age, family size and income. *Family Net.* is a dummy equal to one if the individual has a relative abroad, while $Connect^{Cl}$ is a dummy equal to one if the individual has a connection abroad who is not a relative.

C Matching Results

C.1 Matching: Methodological Discussion

In the matching literature, individuals can be matched either on the estimated probability of receiving a treatment or directly on covariates. Those matching approaches are defined as Propensity Score Matching (PSM) and Covariate Matching (CVM).[35] To deal extensively with the potential selection bias, I applied both methods in the following way.

Following the guidelines of Caliendo and Kopeining (2008), I implement PSM to compute the average cultural stance due to having a connection abroad on connected individuals. In this case the interaction term between connection location and average culture of the destination is not included,[36] so I just focus on connectedness, disregarding connection location. I first compute the individual probability of having a connection abroad (propensity score) using a probit model and a set of relevant covariates. Since the choice the of variables used to compute the propensity score is crucial, I estimate two propensity scores with statistically significant variables (Heckman *et al.*, 1998): one with a set of all relevant and statistically significant covariates (*Main model*) and the other with only a subset of truly exogenous covariates, like age and gender (*Short model*). After graphically testing whether these two models satisfy the Common Support Assumption, I perform matching through different matching algorithms. As a benchmark, I use Kernel (Epanechnikov) matching, a non-parametric matching estimator which uses a weighted average of unconnected individuals within the kernel bandwidth to construct a counter-factual outcome. I match individuals within the same region with similar propensity scores, estimated with both models (main and short). I test the quality of the matching by computing the standardized bias for each covariate (Rosenbaum and Rubin, 1985 and Sianesi, 2004) and their distribution (Heckman *et al.*, 1998 and Aakvik, 2001) after matching. Finally, I estimate the average cultural stance due to connectedness on connected people for each cultural trait with bootstrapped standard errors. As a robustness check, I estimate the same results using other matching algorithms suggested by the literature[37] and estimating the result of a placebo treatment over the control group. I also perform a sensitivity analysis of the treatment driven by a hidden bias, i.e. biased driven by selection in connectedness driven by unobserved factors, by computing the Rosenbaum Bounds (Rosenbaum, 2002). Such bounds measure how strong selection on unobserved factors should be to undermine the estimated treatment effect after matching (Aakvik, 2001 and DiPrete and Gangl, 2004).

Since I estimate the average culture due to connectedness, disregarding connection location (i.e. the interaction term in equation (2)), I rely on CVM methods to asses the robustness of the analysis once disparities in the distribution of covariates between connected and unconnected people are min-

[35] Zhao (2004) describes the main differences of the two approaches, showing through Monte Carlo experiments that these different methods do not dominate each other in terms of performance.

[36] To the best of author's knowledge, this method allows us to compute the average treatment on the treated (ATT) but implementation issues arise when interaction terms are included.

[37] Precisely, I compute the average cultural effect of connectedness on connected people with the following algorithms: Kernel (Normal and Uniform), Nearest Neighbour (one or five individuals, with replacement) and Radius matching. The kernel/radius bandwidth, both for the benchmark and the robustness, is around 0.05.

imized. As Imbens and Rubin (2015) point out, large distributional gaps in covariates increase the sensitivity of estimated coefficients to minor modifications in the specification. Following Ruyssen and Salomone (2018) and Docquier *et al.* (2020), I implement a design phase before the empirical analysis to create a balanced sample of individuals in terms of covariates. I match connected and unconnected individuals within the same region using the Mahalanobis Metric Matching method. This method creates a trimmed sample with an equal number of connected and unconnected individuals where distances in terms of observables are minimized. As for PSM, I test the quality of the matching by computing the standardized bias for each covariate. Estimating equation (2) on the balanced sample after implementing CVM methods shows the robustness of the estimates from the augmented approach, after I have mitigated the selection bias.

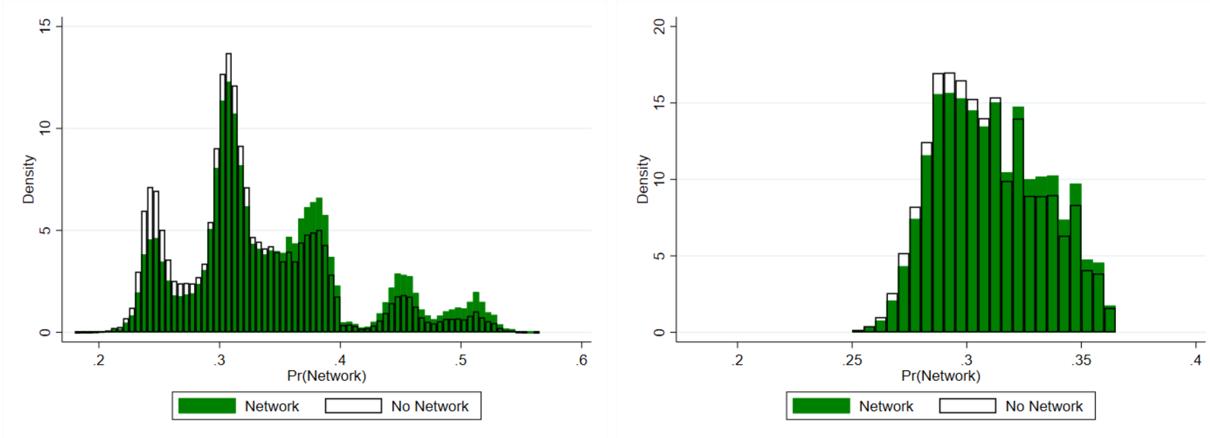
C.2 Matching: Tables and Figures

Table C-1: Propensity Score Matching Estimation

	Probit 2009-2015 (1)	Probit 2009-2015 Main (2)	Probit 2009-2015 Short (3)
Education	0.3745***	0.3774***	
	(0.0050)	(0.0050)	
Female	-0.0262***	-0.0264***	-0.0248***
	(0.0034)	(0.0034)	(0.0030)
Married	-0.1249***	-0.1149***	
	(0.0036)	(0.0038)	
Age	-0.0025***	-0.0159***	-0.0144***
	(0.0001)	(0.0018)	(0.0015)
Urban	0.1963***	0.1936***	
	(0.0038)	(0.0038)	
Unempl.	0.0404***	0.0429***	
	(0.0068)	(0.0067)	
Family Size	0.0050***		
	(0.0009)		
Child	0.0058		
	(0.0037)		
Age^2		0.0003***	0.0002***
		(0.0000)	(0.0000)
Age^3		-0.0000***	-0.0000***
		(0.0000)	(0.0000)
Observations	839707	850609	1108028
Pseudo R2	0.01	0.01	0.00

Note: Author's calculations on Gallup World poll data. * p<0.1, ** p<0.05, *** p<0.01. The dependent variable is a dummy equal to one if the individual has a reliable connection abroad. The coefficient are estimated with a Probit model.

Figure C-1: Probability Score Matching Distribution



(a) Main Model

(b) Short Model

Figure C-2: *

Note: Author's calculations on Gallup World poll Data. The Figure plots the probability density of having a reliable connection abroad using the main model (figure a) and the short model (figure b) estimated in Table C-1.

Table C-2: Sample Means and Standardized Bias - Before and After Matching

	All Sample			*Matched Sample (PS-Main)*			*Matched Sample (PS-Short)*			*Matched Sample (Mahala)*			*Entropy Balance*		
	Treated (1)	Control (2)	Bias (%) (3)	Treated (4)	Control (5)	Bias (%) (6)	Treated (7)	Control (8)	Bias (%) (9)	Treated (10)	Control (11)	Bias (%) (12)	Treated (13)	Control (14)	Bias (%) (15)
Panel A - Social Behavior															
Education	0.14	0.08	11.65	0.18	0.17	0.75	0.18	0.13	9.42	0.18	0.17	1.14	0.18	0.18	-0.10
Female	0.51	0.52	-1.47	0.55	0.56	-1.22	0.54	0.55	-0.92	0.55	0.55	-0.08	0.54	0.54	0.11
Married	0.49	0.55	-9.41	0.50	0.49	1.63	0.49	0.53	-4.98	0.49	0.50	-0.27	0.50	0.50	-0.12
Child	0.58	0.59	-1.05	0.55	0.55	0.46	0.55	0.56	-1.62	0.55	0.55	-0.70	0.55	0.55	0.11
Age	37.07	38.91	-6.99	39.01	40.48	-5.58	39.00	40.21	-4.59	39.00	38.95	0.18	39.00	39.04	-0.15
Urban	0.73	0.65	11.73	0.76	0.76	-0.79	0.75	0.73	3.33	0.76	0.76	-0.36	0.76	0.76	-0.06
Unempl.	0.07	0.07	2.41	0.07	0.07	0.37	0.07	0.07	0.16	0.07	0.07	0.56	0.07	0.07	-0.02
Family Size	3.69	3.65	1.28	3.29	3.21	2.75	3.29	3.21	2.89	3.29	3.26	0.98	3.29	3.29	0.05
Income	6.24	4.47	0.45	9.73	6.85	0.70	9.74	6.57	0.80	9.84	6.78	0.74	9.75	9.78	-0.01
Panel B - Religiosity															
Education	0.14	0.08	11.52	0.18	0.18	0.80	0.18	0.13	9.62	0.18	0.17	1.18	0.18	0.18	-0.10
Female	0.51	0.52	-1.46	0.55	0.56	-1.29	0.55	0.55	-0.95	0.55	0.55	-0.13	0.55	0.55	0.11
Married	0.48	0.55	-8.66	0.49	0.48	1.77	0.49	0.53	-4.91	0.49	0.49	-0.25	0.49	0.49	-0.12
Child	0.58	0.59	-1.27	0.55	0.54	0.64	0.54	0.56	-1.45	0.54	0.55	-0.56	0.55	0.54	0.12
Age	37.27	39.04	-6.72	39.26	40.79	-5.80	39.25	40.50	-4.76	39.25	39.21	0.15	39.25	39.30	-0.16
Urban	0.73	0.65	10.86	0.75	0.76	-0.86	0.75	0.73	3.36	0.75	0.76	-0.38	0.75	0.75	-0.07
Unempl.	0.07	0.07	2.22	0.07	0.07	0.49	0.07	0.07	0.23	0.07	0.07	0.57	0.07	0.07	-0.02
Family Size	3.66	3.64	0.78	3.26	3.17	2.89	3.26	3.17	2.94	03.26	3.22	1.15	3.26	3.26	0.05
Income	6.21	4.49	0.42	9.88	6.91	0.73	9.89	6.61	0.80	9.89	6.84	0.75	9.90	9.93	-0.01
Panel C - Gender-Egalitarian															
Education	0.07	0.04	7.82	0.10	0.09	1.91	0.11	0.07	7.54	0.10	0.10	1.34	0.11	0.11	0.01
Female	0.50	0.51	-0.94	0.49	0.50	-1.14	0.50	0.51	-1.58	0.50	0.50	-0.01	0.50	0.50	-0.08
Married	0.50	0.55	-6.12	0.50	0.50	0.35	0.49	0.53	-4.89	0.49	0.50	-0.35	0.51	0.51	-0.07
Child	0.75	0.75	-0.13	0.73	0.74	-1.01	0.73	0.75	-2.85	0.73	0.74	-1.00	0.73	0.73	0.01
Age	33.43	34.41	-4.08	33.89	34.17	-1.13	33.89	33.84	0.21	34.89	33.48	1.70	34.10	34.11	-0.05
Urban	0.65	0.58	8.21	0.67	0.68	-1.25	0.67	0.65	1.71	0.67	0.68	-0.67	0.67	0.67	-0.05
Unempl.	0.08	0.07	0.95	0.08	0.08	-0.50	0.08	0.08	-0.56	0.08	0.08	1.03	0.07	0.07	0.03
Family Size	4.73	4.47	7.03	4.13	4.09	2.23	4.13	4.01	2.99	4.13	4.07	1.73	4.11	4.11	0.04
Income	2.56	2.10	2.34	3.18	2.82	1.80	3.19	2.53	3.29	3.20	3.22	-0.17	3.19	3.21	-0.07

Note: Author's calculations on Gallup World poll data. Columns (3), (6), (9), (12) and (15) reproduce the standardized bias suggested by Rosenbaum and Rubin (1985). The standardized bias is computed as follow before and after the matching procedure: $SB_{Bef}(X) = 100 * \frac{\overline{X_1} - \overline{X_0}}{\sqrt{[V_1(X)+V_0(X)]0.5}}$ and $SB_{Aft}(X) = 100 * \frac{\overline{X_{1M}} - \overline{X_{0M}}}{\sqrt{[V_1(X)+V_0(X)]0.5}}$. The Entropy Balance statistics reported in columns (13) to (15) are based on the new weights associated developed by the Entropy Balance method.

Figure C-3: Matching Results - Robustness to fake treatment on control group

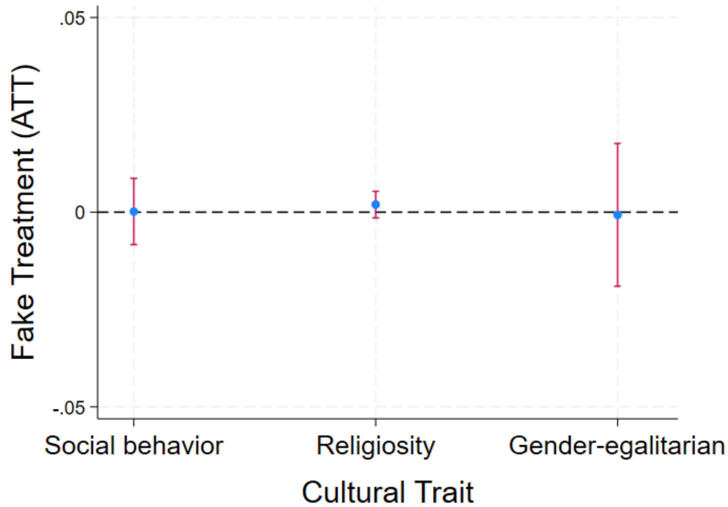


Figure C-4: *

Note: Author's calculations on Gallup World poll Data. The Figure plots the average effect of a fake treatment drawn from an uniform distribution and randomly assigned over the unconnected individuals belonging to the control group on three different cultural traits (Social Behavior, Religiosity and Gender-Egalitarian) and the interval of confidence at 99% level. The propensity score matching is done using Kernel Epanechnikov matching. Standard errors are bootstrapped.

C.3 Matching: Sensitivity Analysis

Following the sensitivity analysis proposed by Rosenbaum (2002), I test the robustness of the estimates cultural to unobserved factors after matching methods. Matching methods manage to eliminate bias driven by selection into observable characteristics, by minimizing the difference between connected and unconnected individuals. However, as for linear methods, they are not robust against "hidden bias" (DiPrete and Gangl, 2004): unobserved factors that affect simultaneously individual culture and conncetedness. Even though Section 4.3 shows that selection on unobservables is a minor concern, I decide to follow the matching literature (Aakvik, 2001 and DiPrete and Gangl, 2004) and compute the Rosenbaum bounds of the estimates.[38] Assumed a certain level of hidden bias (presented with the variable γ), such approach allows us to compute the bounds of average cultural stance due to connectedness, once I assume that hidden bias is causing an over-estimation of the relationship (MH^+) and under-estimation of the relationship (MH^-). To give an intuition behind the value of γ, a value of $\gamma = 1$ is associated with no hidden-bias, while $\gamma = 1.5$ implies that individuals with the same characteristics differ in their odds of having a connection by a factor of 1.5. Following Becker and Caliendo (2007), I use Mantel and Haenszel (1959) test statistic to compute the Rosenbaum bounds after PSM. Table C-3 presents the results of the sensitivity analysis on social behavior (panel A), religiosity (panel B) and gender-egalitarian views (panel C). Column

[38] Due to the high amount of regions in the analysis, the available Stata packages have issues to compute Rosenbaum Bounds over an high number of strata. For this reason, for this sensitivity analysis I perform the matching disregarding the geographical location of individuals.

(1) reports the level of hidden bias (γ). Since the estimates are always positive, then the main concern is related to an over-estimation of the relationship due to hidden bias. For this reason I report the p-value related to the over-estimated (MH^{+}) bound in column (2). Columns (3) and (4) report the bounds while column (5) the confidence interval. The critical level of γ at which the estimated positive estimates of having a connection abroad after matching should be questioned is between 1.85 and 1.90 for social behavior, between 1.20 and 1.25 for gender-egalitarian views and between 1.15 and 1.20 for religiosity. It is important to recall that this approach allow us to evaluate the robustness of the results in the *worst-case scenario*. For instance, concerning the estimate on religiosity, the cultural stance of connectedness would include zero if unobserved variables influence the odds ratio of having a connection abroad between connected and unconnected people with equal characteristics by 1.15 and influence the cultural trait. If unobserved factor influence only individual connectedness and not individual culture, then the confidence interval should not include zero. Nonetheless, those results suggest that the distinctive cultural stance of connected individuals is less robust on religiosity, compared to gender-egalitarian views and social behavior.

Table C-3: Sensitivity Analysis - Rosenbaum Bounds

	Gamma (γ)	p-value	MH^{+}	MH^{-}	C.I.
	(1)	(2)	(3)	(4)	(5)
Panel A - Social Behavior					
	1.00	<0.001	0.159	0.159	[0.158, 0.160]
	1.80	<0.001	0.010	0.608	[0.005, 0.609]
	1.85	<0.001	0.003	0.613	[0.002, 0.615]
	1.90	0.341	0.001	0.616	[-0.009, 0.617]
	1.95	0.999	-0.021	0.617	[-0.052, 0.618]
Panel B - Religiosity					
	1.00	<0.001	0.227	0.227	[0.227, 0.228]
	1.10	<0.001	0.224	0.234	[0.223, 0.235]
	1.15	<0.001	0.219	0.237	[0.217, 0.238]
	1.20	0.102	0.206	0.237	[-0.048, 0.239]
	1.25	1.00	-0.075	0.240	[-0.085, 0.241]
Panel C - Gender-Egalitarian					
	1.00	<0.001	0.217	0.217	[0.211, 0.231]
	1.10	<0.001	0.183	0.272	[0.161, 0.292]
	1.15	<0.001	0.156	0.304	[0.137, 0.311]
	1.20	0.013	0.117	0.313	[0.006, 0.318]
	1.25	0.780	-0.003	0.321	[-0.011, 0.332]

Note: Author's calculations on Gallup World poll data. Column (1) presents the difference in odds of having a connection abroad between matched connected and unconnected individuals (hidden bias). Column (2) present the significance level associated to the over-estimated bound (MH^{+}). Columns (3) and (4) show the over-estimated and under-estimated bounds due to hidden bias. Column (5) shows the confidence interval of the estimated bounds.