

Department of Economics

Regional Trust as a Signal of Local Trustworthiness: Evidence from Migrants

Gianluca Grimalda and Diego Marino Fages

Working Paper No. 14, 2026

Department Economics

Durham University Business School

Mill Hill Lane

Durham DH1 3LB, UK

Tel: +44 (0)191 3345200

<https://www.durham.ac.uk/business/about/departments/economics/>

© Durham University, 2026

Regional Trust as a Signal of Local Trustworthiness: Evidence from Migrants

Gianluca Grimalda
Passau University

Diego Marino Fages
Durham University

August 2026

Abstract

Beliefs about trustworthiness shape exchange, cooperation, and economic development. Kim, Putterman, and Zhang (2022) show experimentally that people draw similar inferences from observed trust and trustworthiness. We test whether these signals are similarly predictive outside the laboratory using TrustLab data on 518 first-generation migrants from 83 origin countries living in 128 regions across six OECD destination countries. A one-standard-deviation increase in native regional trust is associated with a 0.19-standard-deviation increase in migrants' expected trustworthiness, conditional on regional trustworthiness, whose coefficient is near zero. The results are consistent with migrants using local trusting behavior as a signal of trustworthiness.

JEL codes: Z13, F22, C93.

Keywords: Trust; trustworthiness; belief formation; migrants; social learning.

1 Introduction

Many economic interactions require people to assess whether others will reciprocate before reciprocity can be observed. Such beliefs sustain exchange when contracts are incomplete, shape cooperation among strangers, and determine whether mutually beneficial transactions take place. Understanding how people form expectations about others' trustworthiness is therefore central to explaining the foundations of economic exchange. Ben-Ner and Putterman (2001) conceptualize trust as a choice under vulnerability, shaped by beliefs about others' trustworthiness as well as by incentives and preferences. Trust therefore need not map one to one into actual trustworthiness, and observed trust may itself serve as a signal of it. Kim, Putterman, and Zhang (2022) provide an experimental benchmark: subjects draw the same inferences when they observe how much others trust as when they observe how trustworthy they are.

Outside the laboratory, however, these signals need not be equally available or equally credible. An act of reciprocity is direct evidence of trustworthiness, but it is often observed only by the parties to an interaction and may be difficult for outsiders to verify. Trusting behavior may be visible across a wider range of social and economic settings, but it is an indirect signal: the decision

to trust reflects beliefs about others together with risk attitudes, altruism, and a willingness to accept vulnerability. Actual trustworthiness may therefore be more informative when observed, whereas trust may be encountered more frequently. Whether the experimental equivalence of the two signals carries over to naturally occurring social environments is thus an empirical question.

We take this question outside the laboratory by studying migrants' expectations in the social environments in which they live. Migrants provide a particularly informative setting because, unlike natives, they entered the destination environment after spending part of their lives elsewhere. They are therefore exposed to local levels of trust and trustworthiness that may differ from those prevailing where their prior beliefs were formed. We use harmonized TrustLab experiments from six destination countries, covering 518 first-generation migrants and the native respondents in their destination regions. The experiment separately measures how much participants entrust to another person, how much they return when trusted, and how much they expect others to return.

We regress migrants' trust, trustworthiness, and expected trustworthiness on native regional trust and native regional trustworthiness, controlling for individual characteristics and destination-country fixed effects; a second specification adds country-of-origin fixed effects. These horse-race regressions ask which local behavior predicts migrants' outcomes more strongly. The comparison is meaningful because trust and trustworthiness do not map one to one in the data.¹ Across native regional cells, their unweighted correlation is 0.35, and regions with similar levels of trust can exhibit markedly different levels of trustworthiness.

The central result is that migrants' expected trustworthiness is associated with regional trust, not with regional trustworthiness. A one-standard-deviation increase in native regional trust is associated with a 0.19-standard-deviation increase in expected trustworthiness. The coefficient on native regional trustworthiness is close to zero. The pattern survives the inclusion of country-of-origin fixed effects: the coefficient on regional trust is 0.18, while the coefficient on regional trustworthiness remains small and statistically insignificant. Regional trust is positively, but not statistically significantly, correlated with migrants' own trusting behavior and their own trustworthiness. Regional trustworthiness is not significantly correlated with any of the migrants' measures.

Taken together, the findings are consistent with migrants using the trusting behavior they observe around them as a signal of how trustworthy others are. These results should not be interpreted as causal because migrants are not randomly assigned to regions, but rather as conditional associations.

The paper contributes to three related strands of literature. First, it shifts the study of trust from whether trusting choices reflect beliefs or preferences to the prior question of how those beliefs are formed. The distinction between trust, trustworthiness, and their measurement is longstanding: Glaeser et al. (2000) find that standard attitudinal trust questions predict experimental trustworthiness better than experimental trusting behavior. Related work shows that trust-game sending combines expectations with altruism, inequality aversion, reciprocity, and attitudes toward risk or vulnerability (Fehr and Schmidt, 1999; Andreoni and Miller, 2002; Fisman et al., 2007; Ashraf

¹See Ben-Ner and Putterman (2001) for a conceptual analysis of why this difference may appear.

et al., 2006; Sapienza et al., 2013; Costa-Gomes et al., 2014; Kovacs et al., 2024). We show that the local behavior most closely associated with expectations need not be local trustworthiness itself. Second, we contribute to work on social inference and norm perception by testing, in field data, how two distinct community-level signals are associated with beliefs (Kim et al., 2022; Bicchieri et al., 2021). Third, we add a belief-formation mechanism to the literature on the assimilation of trust and social preferences. Migrants’ behavior has been shown to correlate with the norms and preferences prevalent at destination (Marino Fages and Morales Cerda, 2022; Marino Fages, 2023); our results suggest that adaptation may also operate through expectations about others.

2 Data and Measurement

We use harmonized experimental and survey data from TrustLab (Murtin et al., 2018; Kovacs et al., 2024). TrustLab is an OECD-led international data-collection initiative that combines incentivized online experiments with extensive harmonized surveys administered to nationally representative samples. The regional analysis covers 518 first-generation migrants living in 128 regions across six destination countries: Germany, the United Kingdom, Italy, Japan, Luxembourg, and Slovenia. We match each migrant to the experimental trust and trustworthiness of native-born participants living in the same region. The final estimation sample includes respondents with nonmissing outcomes and baseline controls. We exclude two respondents whose reported country of birth is the non-country category “Other.”

The key behavioral measures come from the trust game (Berg et al., 1995). In TrustLab, two participants receive an initial endowment of 10 monetary units. Participant A chooses how much of this endowment to send to Participant B. The amount sent by Participant A is multiplied by three. Participant B then decides how much of their total resources to return to Participant A. TrustLab uses the strategy method to measure trustworthiness: Participant B makes 11 decisions, one for each possible amount Participant A could send.

Behavioral trust is the amount sent by Participant A. Expected trustworthiness measures the amount the respondent expects Participant B to return in the trust game. Trustworthiness is measured as the average amount the respondent returns as Participant B across the strategy-method profile.

The key regional predictors are constructed from native-born respondents in the same host country and region. Regional behavioral trust is the average amount sent by natives as Participant A. Regional trustworthiness is the average of natives’ Participant B return profile from the strategy method. We first compute native regional means in raw experimental units and then standardize each regional measure across regions within the destination country.

Trust and trustworthiness capture distinct margins of social exchange. Trust is the willingness to place resources at risk before another person’s response is known; trustworthiness is the reciprocation that follows once trust has been extended. Although the two behaviors may reinforce one another in equilibrium, neither mechanically determines the other. Figure 1 shows that this distinc-

tion is empirically consequential. Across native regional cells, behavioral trust and trustworthiness are positively but imperfectly associated: the unweighted correlation is 0.35, and the correlation weighted by the number of native respondents is 0.47. The dispersion around this relationship is substantial, with some regions combining high trust and low trustworthiness and others displaying the reverse. This imperfect comovement provides the variation that allows our horse-race regressions to distinguish which feature of the local social environment more strongly predicts migrants' outcomes.

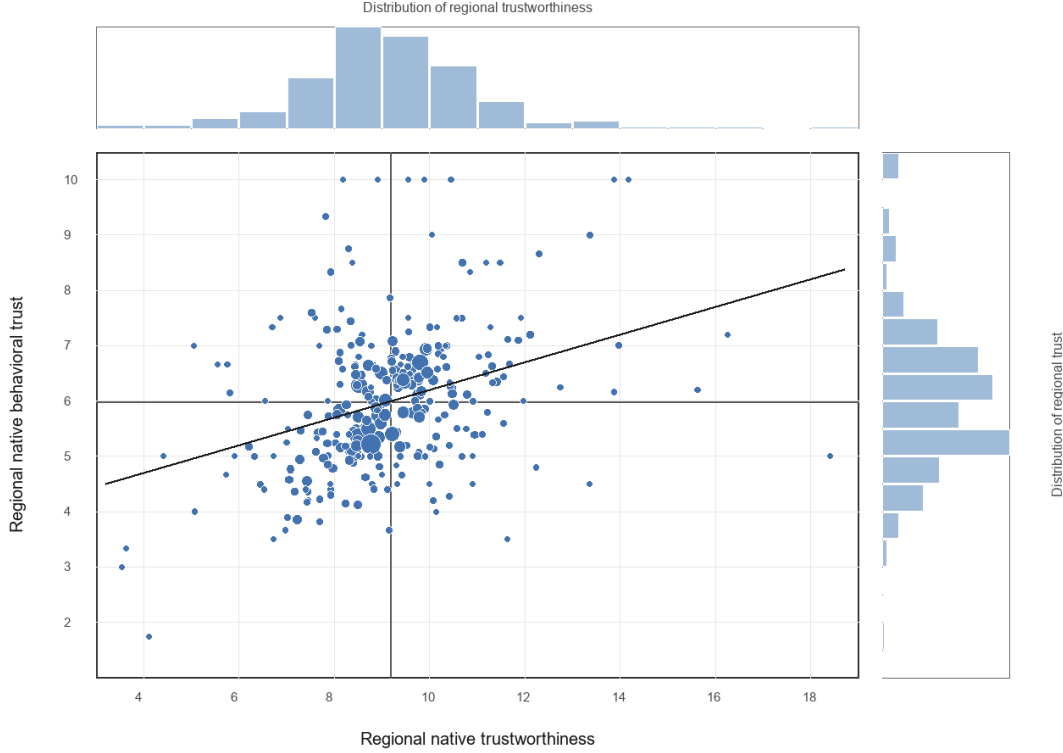


Figure 1: Native regional behavioral trust and trustworthiness.

Note: Each point represents a native regional cell, with marker size increasing in the number of native respondents. The solid black line is the bivariate least-squares fit; the gray reference lines indicate the unweighted means. The marginal histograms show the distributions of regional native trustworthiness and behavioral trust. The sample contains 272 regional cells in six countries. The unweighted correlation is 0.35; weighting by the number of native respondents gives a correlation of 0.47.

3 Empirical Specification

For first-generation migrant i , residing in region r of destination country c and born in country o , we estimate:

$$Y_{irc} = \alpha + \beta_1 \text{RegTrustworthiness}_{rc} + \beta_2 \text{RegTrust}_{rc} + X'_i \gamma + \delta_c + \lambda_{o(i)} + \varepsilon_{irc}. \quad (1)$$

The outcome Y_{irc} is alternatively the migrant’s expected trustworthiness, behavioral trust, or trustworthiness. $RegTrust_{rc}$ and $RegTrustworthiness_{rc}$ denote the within-country standardized native regional averages of trust and trustworthiness. The coefficients β_1 and β_2 therefore compare the association of each regional behavior with the migrant outcome while holding the other fixed. The vector X_i includes age, a female indicator, education fixed effects, and log personal income; δ_c denotes destination-country fixed effects. In specifications with country-of-origin fixed effects, $\lambda_{o(i)}$ absorbs differences associated with the migrant’s country of birth. The term ε_{irc} is the individual-level error term. Because the regional predictors are shared by migrants living in the same destination region, standard errors are clustered at that level.

4 Results

Columns (1) and (2) of Table 1 examine migrants’ expected trustworthiness. The table reports estimates of Equation 1. All specifications include individual controls and destination-country fixed effects, while the even-numbered columns additionally include country-of-origin fixed effects. In column (1), the coefficient on regional trust is 0.193 (standard error 0.066), implying that a one-standard-deviation increase in native regional trust is associated with a 0.19-standard-deviation increase in migrants’ expected trustworthiness. The coefficient on regional trustworthiness is close to zero at -0.010 (standard error 0.062). Adding country-of-origin fixed effects leaves this pattern largely unchanged: the coefficient on regional trust is 0.174 (standard error 0.071), while that on regional trustworthiness is -0.016 (standard error 0.071). The results are unchanged when we exclude Japan, which contributes only five migrants, and when we exclude origin countries represented by a single migrant in the estimation sample.

Table 1: Regional trust and migrants’ outcomes.

	Exp. trustworthiness		Behavioral trust		Trustworthiness	
	(1)	(2)	(3)	(4)	(5)	(6)
RegTrust	0.193*** (0.066)	0.174** (0.071)	0.092 (0.058)	0.061 (0.060)	0.090 (0.064)	0.071 (0.070)
RegTrustworthiness	-0.010 (0.062)	-0.016 (0.071)	-0.024 (0.059)	-0.033 (0.063)	-0.038 (0.064)	-0.016 (0.077)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Host-country FE	Yes	Yes	Yes	Yes	Yes	Yes
Country-origin FE	No	Yes	No	Yes	No	Yes
Observations	518	518	518	518	518	518
R-squared	0.044	0.220	0.066	0.234	0.045	0.177

Note: Coefficients are standardized-effect estimates. Regional predictors are native-born regional averages, standardized within host country. Controls include age, female respondent, education fixed effects, and log personal income. Standard errors, clustered by destination region, are in parentheses. Inference is similar when clustering by country of birth or two-way by country of birth and destination region. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Columns (3)–(6) consider migrants’ own behavior. Regional trust is positively associated with both behavioral trust and trustworthiness, but the estimates are smaller and statistically insignificant. Regional trustworthiness is also unrelated to either outcome. The association with the local social environment is therefore concentrated in migrants’ expectations about others rather than in their own trust-game behavior.

5 Discussion

Experimental evidence shows that people infer cooperativeness from both trust and trustworthiness when both are observable (Kim et al., 2022). In the field, however, the environment provides more complex signals. Trusting behavior may be publicly visible, while the actual reciprocation of trust may be harder to observe. The results of this study suggest that the more visible local behavior is the stronger predictor of migrants’ beliefs.

Using experimental data on migrants in six OECD countries, we find that migrants living in regions with higher native trust report higher expected trustworthiness, conditional on native regional trustworthiness.

The main caveat is that the regressions are observational. Even in specifications with country-of-origin and destination-country fixed effects, migrants’ choice of destination region remains endogenous. The scope of the data nevertheless gives the analysis broad empirical reach. The estimation sample includes 518 first-generation migrants born in 83 countries and living in 128 regions across six destination countries. It therefore combines variation along three dimensions: migrants’ origin

environments, the national settings to which they move, and the local social environments they encounter within those destinations. The estimates do not rely on a single migration corridor or on comparisons between a small number of national cultures.² Destination-country fixed effects absorb average differences across the six destinations, and specifications with country-of-origin fixed effects additionally absorb average differences across the 83 origin countries; the regional predictors exploit variation in native trust and trustworthiness across 128 local contexts. The evidence thus speaks to a relationship observed across a broad range of migrant backgrounds and regional environments rather than to the experience of one migrant group in one destination.

The main contribution is to show that, although trust and trustworthiness are positively correlated across regions, their relationship is far from perfect and migrants' expectations about trustworthiness are predicted by regional trust rather than by regional trustworthiness. This pattern is more consistent with belief formation based on observable local behavior than with a direct adjustment of social preferences or a best response to the local level of trustworthiness.

6 Acknowledgements

Refine.ink was used to check the paper for consistency and clarity.

References

- Andreoni, J. and Miller, J. (2002). Giving according to garp: An experimental test of the consistency of preferences for altruism. *Econometrica*, 70(2):737–753.
- Ashraf, N., Bohnet, I., and Piankov, N. (2006). Decomposing trust and trustworthiness. *Experimental Economics*, 9(3):193–208.
- Ben-Ner, A. and Putterman, L. (2001). Trusting and trustworthiness. *Boston University Law Review*, 81(3):523–551.
- Berg, J., Dickhaut, J., and McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior*, 10(1):122–142.
- Bicchieri, C., Dimant, E., and Xiao, E. (2021). Deviant or wrong? the effects of norm information on the efficacy of punishment. *Journal of Economic Behavior & Organization*, 188:209–235.
- Costa-Gomes, M. A., Huck, S., and Weizsäcker, G. (2014). Beliefs and actions in the trust game: Creating instrumental variables to estimate the causal effect. *Games and Economic Behavior*, 88:298–309.

²The destination countries also span markedly different institutional and cultural settings: the liberal British welfare regime, the conservative-corporatist systems of continental Europe, Italy's Southern European model, Slovenia's post-socialist institutional trajectory, and Japan's East Asian productivist regime (Esping-Andersen, 1990; Ferrera, 1996; Holliday, 2000). They cover four of the eight cultural areas identified by Inglehart and Welzel (2005)—English-speaking, Protestant Europe, Catholic Europe, and Confucian—and, through Japan, include a non-Western setting (Henrich et al., 2010).

- Esping-Andersen, G. (1990). The three worlds of welfare capitalism. *Polity*.
- Fehr, E. and Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics*, 114(3):817–868.
- Ferrera, M. (1996). The ‘southern model’ of welfare in social europe. *Journal of European social policy*, 6(1):17–37.
- Fisman, R., Kariv, S., and Markovits, D. (2007). Individual preferences for giving. *American Economic Review*, 97(5):1858–1876.
- Glaeser, E. L., Laibson, D. I., Scheinkman, J. A., and Soutter, C. L. (2000). Measuring trust. *The Quarterly Journal of Economics*, 115(3):811–846.
- Henrich, J., Heine, S. J., and Norenzayan, A. (2010). Beyond weird: Towards a broad-based behavioral science. *Behavioral and brain sciences*, 33(2-3):111.
- Holliday, I. (2000). Productivist welfare capitalism: Social policy in east asia. *Political studies*, 48(4):706–723.
- Inglehart, R. and Welzel, C. (2005). Modernization, cultural change, and democracy. *The human development sequence*.
- Kim, J., Putterman, L., and Zhang, X. (2022). Trust, beliefs and cooperation: Excavating a foundation of strong economies. *European Economic Review*, 147:104166.
- Kovacs, R., Dunaïski, M., Galizzi, M. M., Grimalda, G., Hortala-Vallve, R., Murtin, F., and Putterman, L. (2024). The determinants of trust: Findings from large, representative samples in six oecd countries. *Economica*, 91(364):1521–1552.
- Marino Fages, D. (2023). Migration and trust: Evidence on assimilation from internal migrants. NICEP Working Paper 2023-08, Nottingham Interdisciplinary Centre for Economic and Political Research.
- Marino Fages, D. and Morales Cerda, M. (2022). Migration and social preferences. *Economics Letters*, 218:110773.
- Murtin, F., Fleischer, L., Siegerink, V., Aassve, A., Algan, Y., Boarini, R., Gonzalez, S., Lonti, Z., Grimalda, G., Schmidt, U., Hortala-Vallve, R., Kim, S., Lee, D., Putterman, L., and Smith, C. (2018). Trust and its determinants: Evidence from the TrustLab experiment. OECD Statistics Working Papers 2018/02, OECD Publishing, Paris.
- Sapienza, P., Toldra-Simats, A., and Zingales, L. (2013). Understanding trust. *The Economic Journal*, 123(573):1313–1332.