The Slave Trade and the Origins of Mistrust in Africa

A Replication Study of Nunn and Wantchekon (*American Economic Review, 2011)*

Giulio Zanella*

Journal of Comments and Replications in Economics, Volume 4, 2025-1, DOI:*10.18718/81781.40*

JEL: O12, O43, Z1
Keywords: Trust, Culture, Development, Afrobarometer

Data Availability: The STATA code and data to reproduce the results of this replication can be downloaded at JCRE's data archive (DOI: 10.15456/j1.2025049.0911138519).

Please Cite As: Zanella (2025). The Slave Trade and the Origins of Mistrust in Africa. A Replication Study of Nunn and Wantchekon (2011). *Journal of Comments and Replications in Economics*, Vol.4 (2025-1). DOI: 10.18718/81781.40

Abstract
Nunn & Wantchekon (2011) detect a long-lasting impact of the slave trade on current trust levels across ethnic groups in Africa. They use data from Afrobarometer's wave 3 to construct trust measures. While I can perfectly replicate the original OLS and 2SLS findings in this wave starting from the same raw data, using data from more recent waves which were collected after the paper was published, I fail to replicate some of the paper's central findings. Plausible explanations are discussed.

*Corresponding author: Professor of Economics, University of Bologna, Bologna, Italy, giulio.zanella@unibo.it
Declaration: Thanks to two anonymous referees and Co-Editor Marianne Saam for useful suggestions. Carlo Giuffré provided excellent research assistance.

Received February 14, 2024; Revised September 26, 2024; Accepted January 21, 2025; Published April 16, 2025

1 Introduction

Nunn & Wantchekon (2011)—henceforth, NW—study the cultural roots of Africa's underdevelopment by relating individual-level survey measures of trust collected during 2005-2006 via Afrobarometer's wave 3 to data on slave exports during 15th-19th centuries. The authors argue that this historical factor shaped a persistent culture of mistrust, which in light of the importance of trust in economic transactions provides the missing link between the slave trade and Africa's development challenges that was detected empirically by Nunn (2008).

Using data from more recent waves that were collected after NW's paper was published, I fail to replicate some of the paper's central findings. Plausible explanations are the insufficient persistence of survey measures of beliefs (Zanella & Bellani, 2024) and the use of Likert scales to gauge trust in the Afrobarometer (Bond & Lang, 2019).

I emphasize that the issue that I raise is general and is *not* about NW's research procedures (they used the best data available at the time and their analysis is impeccable). The point is the reliability of survey measures of values and beliefs in historical economics and in economic analysis at large. Although this is a well-known issue and alternative measurement strategies have been devised (Lowes, 2022 provides an up-to-date overview), survey questions are still an attractive option because the data are readily available in relatively large samples for virtually all countries, at no cost to researchers.

A different kind of replication of Nunn & Wantchekon (2011) was previously undertaken by Deconinck & Verpoorten (2013) shortly after the original paper was published. These authors employ Afrobarometer wave 4 (collected during 2008-2009), which covers additional countries, additional ethnic groups, and asks partly different trust questions. Thus, that replication uses a single, additional wave collected 3 years later to address the question of whether NW's results hold in samples that include more countries, more ethnic groups, and considering additional trust questions. My replication offers a different perspective, as it uses wave 4 and subsequent Afrobarometers waves 5–8 (collected during 2011-2021) to check whether NW's results hold in samples from the same countries, same ethnic groups, and considering the exact same trust questions that are asked up to 17 years apart. Being undertaken from different vantage points, the two replication exercises are complementary.

2 Summary of Nunn & Wantchekon (2011)

Before turning to the replication, I summarize NW in more detail. The authors are interested in the connection between historical experiences of the slave trade and current levels of trust in Sub-Saharan Africa. Using data from wave 3 of the Afrobarometer (2005-2006) across 17 countries, NW find a significant negative relation between, on the one hand, an individual's trust of relatives, trust of neighbors, trust of the elected local government council, trust of people in one's own ethnic group, or trust of people in other ethnic groups and, on the other hand, a scalar measure of the number of slaves taken from one's own ethnic group, indexed by e, between 1400 and 1900, which I denote by S_e. An individual's trust attitude is measured from answers to questions about how much one trusts certain types of people or certain institutions, on a Likert scale that ranges between 0 ("not at all") and 3 ("a lot"), with intermediate values 1 ("just a little") and 2 ("somewhat").

NW argue that historical experiences of being kidnapped and sold into slavery, even by friends and relatives, resulted in an insecure environment that generated a culture of mistrust that persists until the present. In order to shut off the possible direct effect of the slave trade on trust, a rich set of conditioning variables are included in the model: variables that characterize individual i (age, gender, ethnicity, religion, education, occupation, urban status, and living conditions), $\mathbf{X}_i$; two variables capturing the ethnic composition of one's residential district d (ethnic fractionalization and own ethnicity share), $\mathbf{X}_d$; and ethnic group-level variables that control for pre-colonial prosperity (i.e., initial conditions) and colonial rule (average malaria presence in the land historically inhabited by one's ethnic group, colonial population density, pre-colonial settlement patterns, number of jurisdictional hierarchies as a proxy for an ethnic group's political institutions, presence of railway lines on a group's land in 1911, indicators for whether European explorers traveled through that land, and contact with European missionaries during the colonial period), $\mathbf{X}_e$.

Formally, the estimating equation is:

$$T_{iedc} = \alpha_c + \beta \cdot S_e + \mathbf{X}_i'\mathbf{\Gamma} + \mathbf{X}_d'\mathbf{\Omega} + \mathbf{X}_e'\mathbf{\Phi} + \varepsilon_{iedc}, \tag{1}$$

where T_{iedc} denotes an individual trust measure for individual i who belongs to ethnic group e and who lives in district d of country c, and α_c is a country fixed effect. Standard errors are clustered at the ethnic group/district level. Thus, this design exploits the variability of slave trade experiences within countries that are generally characterized by identical (for all individuals in a country) formal institutions at present. NW's preferred measure of the number of slaves taken from one's own ethnic group between 1400 and 1900 is:

$$S_e = \ln\left(1 + \frac{\text{number of slaves exported}}{\text{ethnic group land area}}\right). \tag{2}$$

In order to address causality (i.e., the possibility that unobserved, persistent determinants of current trust in ε_{iedc} determined selection into the slave trade in the past and therefore correlate with S_e even after conditioning on observable individual-, district-, and ethnic-level variables), NW instrument historical explanatory variable S_e with historical distance of one's ethnic group from the coast—a plausibly exogenous variable that captures exposure to the slave trade and that I denote by Z_e—in the following 2SLS model:

$$T_{iedc} = \alpha_c + \beta \cdot \widehat{S}_e + \mathbf{X}_i'\mathbf{\Gamma} + \mathbf{X}_d'\mathbf{\Omega} + \mathbf{X}_e'\mathbf{\Phi} + \varepsilon_{iedc} \tag{3}$$

$$\widehat{S}_e = \widehat{a}_c + \widehat{b} \cdot Z_e + \mathbf{X}_i'\widehat{\mathbf{C}} + \mathbf{X}_d'\widehat{\mathbf{D}} + \mathbf{X}_e'\widehat{\mathbf{E}}, \tag{4}$$

where a "hat" denotes a predicted variable or estimated coefficient. The parameter of interest is β, which captures the causal effect of historical exposure of one's ethnic group to the slave trade on an individual's trust in certain types of people or certain institutions at present.

3 *Original findings and replication*

3.1 *Data*

NW's final data set is available at the openICPSR repository. From this data file, I take the number of slaves from one's own ethnic group, the mapping between ethnic groups reported in the Afrobarometer, Murdock's (1959) classification of such groups (this is the definition of ethnic groups used by NW and applied throughout this paper, unless otherwise noted), and a group's historical distance from the coast. For individual characteristics and trust measures, instead, I employ the Afrobarometer directly in order to check that I am able to reproduce the results of the original article by constructing cultural proxies and covariates from the original data source. Following the sample selection criteria described in the original article, I end up with a sample of 21,821 observations from Afrobarometer's wave 3 (2005-2006), versus 21,822 in NW's final data set—an irrelevant discrepancy.

For my replication, I also use the subsequent waves, which were not available when NW's research was undertaken. My analysis is primarily based on wave 5 (2011-2013)—the only one to date where at least three of the five trust questions employed by NW (namely: trust of relatives, trust of neighbors, and trust of the local council) are asked—but I also use wave 4 (2008-2009)—which contains two of the original trust questions (trust of relatives and trust of the local council)—and waves 6 (2014-2015), 7 (2016-2018), and 8 (2019-2021)—which contain only one of the five questions asked in wave 3 (trust of the local council). Table 1 compares the exact wording of the trust questions across these waves, as taken from the Afrobarometer questionnaires. There were no relevant changes in these questions between waves 3 and 8, apart from the fact that some of the wave 3 questions were not asked in subsequent waves. The response categories never changed either. [1]

In order to preserve an exact correspondence between the original results in wave 3 and my own results in these more recent waves, I select the same 17 countries and Afrobarometer ethnic groups represented in the original study. This choice comes at the cost of sample size because the support of the distribution of ethnic groups varies across waves. Thus, I am able to match Afrobarometer ethnic groups across waves 3 and 5 for 89.1% of subjects in wave 3 and 78.5% of subjects in wave 5. That is,

[1] They are always the following: Not at all (0), Just a little (1), Somewhat (2), A lot (3), Don't know/ Haven't heard/ Did not respond (9).

Table 1: Trust Questions Across Afrobarometer Waves 3 to 8

Trust Type	**Wave 3**	**Wave 4**	**Wave 5**	**Wave 6**	**Wave 7**	**Wave 8**
Trust of relatives	How much do you trust each of the following types of people: relatives?	How much do you trust each of the following types of people: relatives?	How much do you trust each of the following types of people: relatives?	not asked	not asked	not asked
Trust of neighbors	How much do you trust each of the following types of people: neighbors?	not asked	How much do you trust each of the following types of people: neighbors?	not asked	not asked	not asked
Trust of local council	How much do you trust each of the following, on a scale of 1 to 5: local council?	How much do you trust each of the following, on a scale of 1 to 5: local council?	How much do you trust each of the following, on a scale of 1 to 5: local council?	How much do you trust each of the following, on a scale of 1 to 5: local council?	How much do you trust each of the following, on a scale of 1 to 5: local council?	How much do you trust each of the following, on a scale of 1 to 5: local council?
Intra-group trust	How much do you trust each of the following types of people: members of your own group?	not asked	not asked	not asked	not asked	not asked
Inter-group trust	How much do you trust each of the following types of people: members of other groups?	not asked	not asked	not asked	not asked	not asked

there are about 10% of individuals whose reported ethnic group in wave 3 is not present in wave 5 (for example, the Oshiwambo, the Bakoena, and the Msukuma), and vice versa for about 20% of individuals in wave 5 (for example, the Shona, the Wambo, and the Zezuru).[2] Since my replication is concerned with the validity of NW's results across multiple waves when considering the exact same countries and ethnic groups of the original article, the 6,863 observations from wave 5 whose Afrobarometer ethnicity is not present in wave 3 are dropped, resulting in what I henceforth refer to as the "waves 3-5 ethnic match sample", and similarly for the other, post-wave 3 waves. Despite this ethnic match constraint, the size of my wave 5 final sample ($N = 23,322$) is larger than its wave 3 counterpart ($N = 21,821$). There is a similar mismatch for wave 4 (ethnic groups are matched for 87.6% in wave 3 and 82% in wave 4; $N = 18,519$), wave 6 (ethnic groups are matched for 85.4% in wave 3 and 77.1% in wave 6; $N = 22,508$), wave 7 (ethnic groups are matched for 90.7% in wave 3 and 82.6% in wave 7; $N = 19,293$), and wave 8 (ethnic groups are matched for 76.2% in wave 3 and 66.4% in wave 8; $N = 14,387$).

Contrary to Afrobarometer's wave 3, there is no information on a respondent's district of residence in waves 5, 6, 7, or 8, and so the two district-level conditioning variables in $\mathbf{X}_d$ cannot be employed in my replications that use these more recent Afrobarometer data (and standard errors can only be clustered at the ethnic group level). In order to demonstrate that the presence of ethnic groups in wave 3 that are absent from the other waves and missing district of residence information are immaterial, I will also replicate the original results omitting district-level variables or using only wave 3 observations whose ethnic groups overlap with those represented in each of other waves (I label these smaller samples as the "wave 3, waves 3-w ethnic match" samples, for $w = \{4, 5, 6, 7, 8\}$). As it turns out, excluding district-level variables does not affect the original NW estimates at all. Similarly, wave 3 estimates derived from smaller, ethnic match samples without district information—which serve as a benchmark for the replication in each of the subsequent waves—are virtually identical to the original NW estimates.

Also note that conditioning on individual characteristics in $\mathbf{X}_i$ across waves deals with the fact that the sample may change in terms of demographic and socioeconomic characteristics because of broader changes in African societies.

3.2 Results

Results using data from waves 3 and 5 are reported in Table 2 (OLS, equation 1, corresponding to Table 3 in NW) and Table 3 (2SLS, equations 3-4, corresponding to Table 5 in NW). For each model, the original results are reproduced in column [1] using NW's replication package; my replication using wave 3 and starting from the raw data is reported in column [2]; column [3] contains my replication using wave 3 after removing the two district-level conditioning variables employed by NW and that are available in wave 3 but not in wave 5 (or subsequent waves, apart from wave 4); column [4] contains my replication using wave 3 after discarding observations whose reported ethnic group cannot be matched across waves 3 and 5, and still omitting district-level conditioning variables; finally, my replications using wave 5 or pooling waves 3 and 5 are reported, respectively, in columns [5] and [6]. For each cultural measure, estimates are produced by distinct regressions. Following NW, standard errors are clustered at the ethnic group/district level in columns [2] and [3]. In the remaining columns, they are clustered at the ethnic group level.

There is some immaterial discrepancy in the number of observations across columns [1] and [2]—estimates are virtually identical. Thus, I am able to replicate the original OLS and 2SLS findings in Afrobarometer wave 3 both using NW's final data set and estimation.do files available in their replication package while reconstructing the data set independently from the raw survey data. The estimates in column [3] are also very similar to the original ones, which establishes that conditioning or not conditioning on the two district-level controls in $\mathbf{X}_d$ (and clustering standard errors at the ethnic group/district level or only at the ethnic group level) does not make any appreciable difference. This fact is reassuring given that I cannot include district-level controls or clusters when using wave 5 or later waves 6–8.

[2]In matching ethnic denominations across waves, in addition to correcting obvious typos and minor spelling differences, I have searched for all historical and alternate spellings, possible linguistic or regional variants, and subgroup- or clan-specific names. I have also verified whether any of these correspond uniquely to ethnic descriptors present in wave 3. Consequently, the residual mismatch is primarily due to the grouping of multiple wave-3 ethnic groups that cannot be disentangled in subsequent waves, as well as a combination of sampling variability and measurement error in self-reported ethnicity across waves. Later, I demonstrate that this residual ethnic mismatch is inconsequential for my replication.

Table 2: Original and Replication OLS Estimates of the Effect of the Slave Trade on Trust

	[1] Original, wave 3	[2] Replication, wave 3	[3] Replication, wave 3, no district controls	[4] Replication, wave 3, waves 3-5 ethnic match	[5] Replication, wave 5		[6] Replication, waves 3 and 5 pooled
	Slave exports 1400-1900	Slave exports 1400-1900	Slave exports 1400-1900	Slave exports 1400-1900	Slave exports 1400-1900		Slave exports 1400-1900
	Coeff. (s.e.)	Coeff. (s.e.)	Coeff. (s.e.)	Coeff. (s.e.)	Coeff (s.e.)	Diff. p-val	Coeff. (s.e.)
Trust of relatives	-0.178^{**}	-0.179^{**}	-0.176^{**}	-0.175^{**}	-0.033	0.00	-0.094^{**}
	(0.032)	(0.032)	(0.032)	(0.033)	(0.026)		(0.024)
Trust of neighbors	-0.202^{**}	-0.202^{**}	-0.198^{**}	-0.204^{**}	-0.151^{**}	0.25	-0.154^{**}
	(0.031)	(0.031)	(0.030)	(0.032)	(0.024)		(0.024)
Trust of local council	-0.129^{**}	-0.128^{**}	-0.127^{**}	-0.137^{**}	-0.025	0.01	-0.030
	(0.021)	(0.021)	(0.021)	(0.022)	(0.030)		(0.028)
Intra-group trust	-0.188^{**}	-0.188^{**}	-0.187^{**}	-0.196^{**}			
	(0.032)	(0.032)	(0.032)	(0.034)			
Inter-group trust	-0.115^{**}	-0.115^{**}	-0.118^{**}	-0.125^{**}			
	(0.030)	(0.030)	(0.030)	(0.032)			
Observations	15,905–16,709	15,886–16,688	15,886–16,688	13,787–14,595	15,525–17,875		31,411–34,583
Ethnicity clusters	146–147	146–147	146–147	121–122	118–119		147
District clusters	1,184–1,194	1,184–1,194	–	–	–		–
District controls	Yes	Yes	No	No	No		No

Notes: The table reports the OLS coefficients from the linear regression of individual trust measures on the historical variable "slave exports from Africa between 1400 and 1900" measured as in equation (2). Column [1] contains the original estimates, which correspond to Table 3 in NW. Column [2] reports my replication in the original AB wave 3 after reconstructing the cultural measures from the raw data. Column [3] removes two district-level regressors that are not available in wave 5 and subsequent waves. Column [4] is my replication in the original AB wave 3 using only ethnicities that can be matched across waves 3 and 5, and still excluding the two district-level variables. Column [5] is my replication in the more recent AB wave 5; the difference (Diff.) is with respect to the original estimates as replicated in column [4], and the p-value is from a test of the null hypothesis that the estimated coefficients are different across columns [4] and [5]. Column [6] is my replication after pooling waves 3 and 5. Each line corresponds to a distinct regression. A quadratic polynomial in age, and gender, ethnicity, religion, education, occupation, urban status, living conditions, and country dummies are always included as conditioning variables. In columns [1]–[3], district ethnic fractionalization and the fraction of own ethnicity in the district are also included. In column [6], a wave dummy is included. All regressions condition on ethnic-level "colonial controls". I report in parentheses standard errors clustered at the ethnicity/district level (columns [1]–[3]) or at the ethnicity level only (columns [4]–[6]). Significance level: * 5%; ** 1%.

Table 3: Original and Replication 2SLS Estimates of the Effect of the Slave Trade on Trust

	[1] Original, wave 3	[2] Replication, wave 3	[3] Replication, wave 3, no district controls	[4] Replication, wave 3, waves 3-5 ethnic match	[5] Replication, wave 5		[6] Replication, waves 3 and 5 pooled
	Slave exports 1400-1900 Coeff. (s.e.)	Slave exports 1400-1900 Coeff. (s.e.)	Slave exports 1400-1900 Coeff. (s.e.)	Slave exports 1400-1900 Coeff. (s.e.)	Slave exports 1400-1900 Coeff (s.e.)	Diff. *p*-val	Slave exports 1400-1900 Coeff. (s.e.)
Trust of relatives	−0.190**	−0.190**	−0.189**	−0.175**	−0.059	0.01	−0.058
	(0.067)	(0.067)	(0.068)	(0.068)	(0.086)		(0.081)
Trust of neighbors	−0.245**	−0.247**	−0.245**	−0.207**	−0.180**	0.19	−0.188**
	(0.069)	(0.070)	(0.072)	(0.068)	(0.065)		(0.072)
Trust of local council	−0.221**	−0.222**	−0.217**	−0.189**	−0.086	0.00	−0.199+
	(0.060)	(0.059)	(0.059)	(0.055)	(0.069)		(0.107)
Intra-group trust	−0.251**	−0.253**	−0.253**	−0.205*			
	(0.087)	(0.087)	(0.087)	(0.082)			
Inter-group trust	−0.174**	−0.177**	−0.174*	−0.131+			
	(0.080)	(0.080)	(0.080)	(0.073)			
First stage:							
Historical distance	−0.0014**	−0.0014**	−0.0014**	−0.0016**	−0.0013**		−0.0009**
from coast	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)		(0.0003)
F-stat	26.3	26.5	26.8	28.2	15.0		12.8
Observations	15,905–16,709	15,886–16,688	15,886–16,688	13,787–14,595	15,525–17,895		31,411–34,583
Ethnicity clusters	146–147	146–147	146–147	121–122	118–119		147
District clusters	1,184–1,194	1,184–1,194	–	–	–		–
District controls	Yes	Yes	No	No	No		No

Notes: The table reports 2SLS coefficients from the linear regression of individual trust measures on the historical variable "slave exports from Africa between 1400 and 1900", measured as in equation (2). This variable is instrumented with distance from one's ethnic group from the coast. Column [1] contains the original estimates, which correspond to Table 5 in NW. Column [2] reports my replication in the original AB wave 3 after reconstructing the cultural measures from the raw data. Column [3] removes two district-level regressors that are not available in wave 5 and subsequent waves. Column [4] is my replication in the original AB wave 3 using only ethnicities that can be matched across waves 3 and 5, and still excluding the two district-level variables. Column [5] is my replication in the more recent AB wave 5; the difference (Diff.) is with respect to the original estimates as replicated in column [4], and the *p*-value is from a test of the null hypothesis that the estimated coefficients are different across columns [4] and [5]. Column [6] is my replication after pooling waves 3 and 5. Each line corresponds to a distinct regression. A quadratic polynomial in age, and gender, ethnicity, religion, education, occupation, urban status, living conditions, and country dummies are always included as conditioning variables. In columns [1]–[3], district ethnic fractionalization and the fraction of own ethnicity in the district are also included. In column [6], a wave dummy is included. All regressions condition on ethnic-level "colonial controls". I report in parentheses standard errors clustered at the ethnicity/district level (columns [1]–[3]) or at the ethnicity level only (columns [4]–[6]). Significance level: * 5%; ** 1%.

The key comparison is between the estimates in columns [3] and [4]. Of the three trust variables that are used by NW and that are also observed in wave 5, only trust of neighbors is robustly associated with slave exports. For trust of relatives or trust of the local council, the estimated coefficients (both when using OLS and when using 2SLS) drop substantially relative NW's estimates in wave 3 and become statistically indistinguishable from zero. Note that for these two variables the estimates across waves 3 and 5 are statistically different. Also note in column [5] that pooling waves 3 and 5 (in this case a wave dummy is added as a conditioning variable in the regressions) does not fix the problem: the insignificant effect of the slave trade on trust of relatives and trust in the local council persists in the pooled 2SLS and OLS estimates, respectively.

As summarized in a more compact way in Table 4, a similar pattern emerges when using waves 7 and 8, or, for the 2SLS part, waves 4 and 6, or when pooling waves 3–8, as far as the same trust questions used by NW are available in these more recent waves—trust of relatives and trust of the local council are available in wave 4, and only trust of the local council is available in waves 6, 7, and 8. Specifically, when using OLS, trust of relatives remains robustly associated with the slave exports measure in wave 4 or when pooling waves 4–8; but not when the explanatory variable is instrumented for. Similarly, trust of the local council is significantly correlated with slave exports in waves 4 and 6, but such correlation vanishes (and its sign flips in waves 7 and 8) when using 2SLS. Thus, for the trust measure that can be consistently compared across all waves, the 2SLS results vanish altogether when an Afrobarometer wave other than wave 3 is used, and even when waves 3–8 are pooled.
Like in column [3] of Tables 2 and 3, Table 4 reports (in the "Wave 3" panel at the top of the table) benchmark estimates in wave 3 after discarding observations whose reported ethnic group cannot be matched across waves 3 and either wave 4, 6, 7, or 8, and omitting district-level conditioning variables. Since information on district of residence is available for wave 4, the bottom panel of Table 4 reports results for wave 4 excluding or including district-level variables. This comparison confirms that they have little influence on the results.

The bottom line of the replication summarized in Tables 2, 3, and 4 is that some of the central findings in NW are not robust to employing Afrobarometer data of different vintages. I next provide explanations for why this may happen.

4 *Explanations*

The econometric specification and the RHS variables are identical across my regressions that use different Afrobarometer waves, and the underlying samples are homogeneous in terms of geographic distribution and ethnic groups. These facts suggest that the culprit of the fragility of some of NW's results across Afrobarometer wave 3 and the more recent waves is the LHS variable, i.e., the survey measures of trust. If such measures are sufficiently correlated with the slow-moving component of unobservable trust levels that NW are interested in, then shifting measurement a few years forward should not make any important difference. Why it does, instead? I offer two explanations: (i) the volatility of Afrobarometer trust measures; (ii) the Afrobarometer survey of trust on a Likert scale. These explanations are corroborated by the analysis in a companion paper to this one (Zanella & Bellani, 2024), where it is demonstrated that two prominent studies that use the World Values Survey to derive measures of cultural attitudes are subject to the same problem of possible replication failure across survey waves that I illustrate here.

In what follows, I focus again on Afrobarometer wave 5 because it maximizes the overlap with the trust questions asked in wave 3 (see Table 1). Given that the Afrobarometer consists of repeated cross-sections, the volatility aspect can be assessed by considering average trust within the same ethnic groups e across waves $w = 3$ and $w = 5$, denoted $\bar{T}_{ew}$. If the survey trust measures capture slow-moving cultural attitudes that were shocked by the slave trade experience and that were subsequently transmitted across generations via family or social interactions until the present, then $\bar{T}_{e3}$ and $\bar{T}_{e5}$ should be quite similar. These two quantities are contrasted in Figure 1, where a circle represent an ethnic group and its size is proportional to the average size of that ethnic group across the two waves. In this figure, two alternative statistics are employed. The first is an unconditional weighted average (top row) that uses within-country sampling weights. This measure may be unsatisfactory because after applying such weights, the Afrobarometer is representative at the country level but not necessarily at the ethnic-group level. Moreover, like in the inferential analyses of the previous section, the sample may change in terms of demographic and socioeconomic characteristics, reflecting broader changes in African societies. Thus, I

Table 4: Replication Estimates of the Effect of the Slave Trade on Trust, Waves 4, 6, 7, and 8

	[1]	[2]	[3]	[4]		
Dependent variable:	Trust of relatives	Trust of local council	Trust of relatives	Trust of local council		
	OLS	OLS	2SLS	2SLS		
	Slave exports 1400-1900	Slave exports 1400-1900	Slave exports 1400-1900	Slave exports 1400-1900	N	ethnic clusters
Wave 3						
original	-0.178^{**}	-0.129^{**}	-0.190^{**}	-0.221^{**}	15,905–	147
	(0.032)	(0.021)	(0.094)	(0.060)	16,709	
w3-w4 match	-0.167^{**}	-0.142^{**}	-0.192^{*}	-0.251^{**}	13,933–	127
	(0.040)	(0.025)	(0.092)	(0.069)	14,711	
w3-w6 match	–	-0.145^{**}	–	-0.203^{**}	13,258	112
	–	(0.023)	–	(0.066)		
w3-w7 match	–	-0.134^{**}	–	-0.200^{**}	14,136	125
	–	(0.024)	–	(0.057)		
w3-w8 match	–	-0.142^{**}	–	-0.167^{*}	11,448	104
	–	(0.030)	–	(0.068)		
Wave 4,	-0.164^{**}	-0.059^{**}	-0.105	-0.083	12,717–	112
w3-w4 match	(0.032)	(0.022)	(0.081)	(0.076)	13,547	
Wave 4,	-0.164^{**}	-0.033	-0.186^{**}	-0.044	7,927–	107
+district vars	(0.031)	(0.033)	(0.067)	(0.084)	8,498	
Wave 6,	–	-0.061^{*}	–	-0.042	16,647	110
w3-w6 match	–	(0.025)	–	(0.106)		
Wave 7,	–	-0.042	–	0.151	11,954	112
w3-w7 match	–	(0.048)	–	(0.180)		
Wave 8,	–	0.009	–	0.072	10,827	90
w3-w8 match	–	(0.030)	–	(0.116)		
Waves 4–8	-0.099^{**}	-0.036^{+}	-0.046	-0.047	48,130–	147
pooled	(0.022)	(0.020)	(0.071)	(0.060)	82,556	

Notes: The table reports the OLS and 2SLS coefficients from the linear regression of individual trust measures used by NW and also available in Afrobarometer waves 4, 6, 7, and 8 on the historical variable “slave exports from Africa between 1400 and 1900” measured as in equation (2). For each of these waves, a benchmark estimate is provided in the top panel, which is obtained in the original wave 3 using only ethnicities that can be matched across waves 3 and either wave 4, 6, 7, or 8. District-level information is not included because it is not available in waves 5, 6, 7, or 8. However, it is available in wave 4, for which results from a second specification (Wave 4 + district vars) that includes such variables as conditioning covariates are reported. A quadratic polynomial in age, and gender, ethnicity, religion, education, occupation, urban status, living conditions, and country dummies are always included as conditioning variables. When pooling waves 3–8, wave dummies are included. All regressions condition on ethnic-level “colonial controls”. I report in parentheses standard errors clustered within ethnic groups. Significance level: $^{+}$ 10%; * 5%; ** 1%.

also employ a second, conditional average (bottom row) that is built following Tabellini (2010), where a similar problem arises using region-level averages of cultural attitudes across European countries. Conditional trust measure that eliminates the influence of different characteristics of respondents across ethnic groups or waves are constructed by projecting an individual’s reported trust in wave w onto the set of individual covariates $\mathbf{X}_i$, country dummies $\mathbf{C}_i = [c_{1i}, c_{2i}, \cdots, c_{Ci}]$, where C is the number of countries and $c_{ji} = 1$ if individual i belongs to ethnic group j, and zero otherwise, and ethnic group dummies $\mathbf{G}_i = [g_{1i}, g_{2i}, \cdots, g_{Gi}]$, where G is the number of ethnic groups and $g_{ei} = 1$ if individual i belongs to ethnic group e, and zero otherwise, via the following linear regression model:

$$T_{iew} = \alpha_w + \mathbf{X}_i'\mathbf{\Gamma}_w + \sum_{j=1}^{C} \eta_{jw} c_{ji} + \sum_{e=1}^{G} \theta_{ew} g_{ei} + \epsilon_{iew}. \tag{5}$$

The sum of the estimated constant and the ethnic group dummy’s estimated coefficient is a conditional measure of culture for ethnic group e in wave w, i.e., $\bar{T}_{ew} = \hat{\alpha}_w + \hat{\theta}_{ew}$.

There are sizable deviations from the (dashed) $45°$ line in Figure 1—regardless of how averages are computed—which indicate that trust measures may change substantially across the two waves. Such variability can be quantified by computing the within-country standard deviation of the log of a trust measure—a convenient volatility indicator because the standard deviation of the log of a trust measure across the two waves gives, approximately, the average percentage deviation of that measure from the ethnic group mean. Based on this indicator, the volatility is about 17% (31%) for unconditional (conditional) trust of relatives; about 19% (30%) for unconditional (conditional) trust of neighbors; about 25% (35%) for unconditional (conditional) trust of the local council. Note that for the purposes of the specific exercises in this section, I use the subset of ethnic groups that are present both in wave 3 and in wave 5.

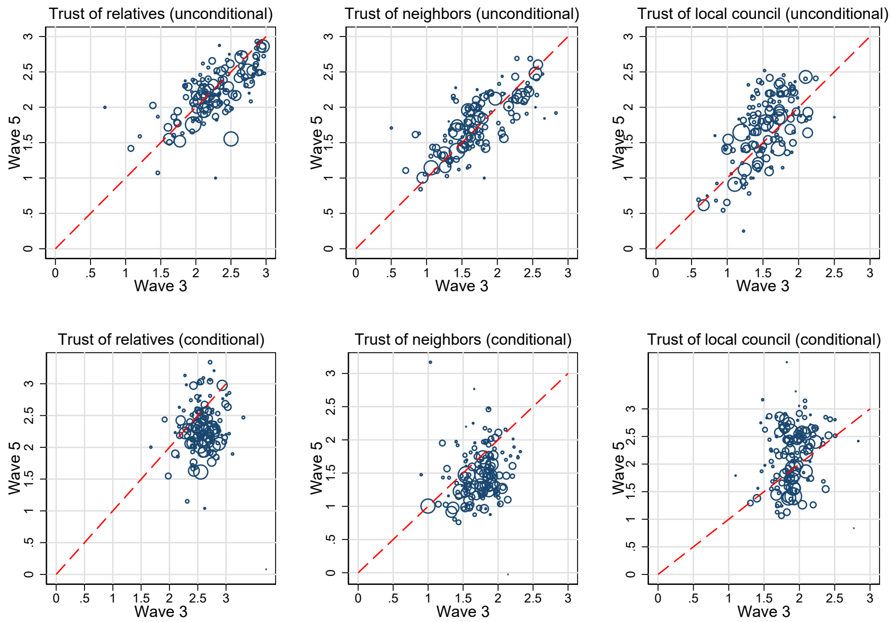

Figure 1: Average, Ethnicity-level Trust Measures in Afrobarometer Wave 3 versus Wave 5

Notes: The figure plots average trust measures within ethnic groups in Afrobarometer wave 3 (horizontal axis) and wave 5 (vertical axis). A circle is an ethnic group, and its size is proportional to an ethnic group's size. The dashed line is the $45°$ line. Unconditional measures (top row) are constructed as weighted averages, using sampling weights. Conditional measures (bottom row) are constructed as the sum of the estimated constant and the ethnic group dummy's estimated coefficient from a linear regression of an individual's reported trust onto individual covariates, country dummies, and ethnic group dummies. Sample: 41,863 individuals from pooled Afrobarometer waves 3 and 5, in ethnic groups that are present in both waves.

As already noted in Zanella & Bellani (2024), the stability of survey measures of trust is not necessary for them to be meaningful dependent variables in equation (3)—the high-frequency component may just act as noise on top of the low-frequency one. The question is whether such measures retain *enough* of the latter to prevent the noise from dominating. This may be why NW's results exhibit statistical fragility across waves.

A second possible reason relates to a problem analyzed by Bond & Lang (2019). The Afrobarometer gauges specific trust by means of discrete 0–3 Likert scales. Since there are more than two categories, computing average trust (either explicitly like in Figure 1 or implicitly when characterizing the conditional mean of trust in a regression framework like equation 3) requires a cardinalization of ordinal survey answers. The existence of infinite possible cardinalizations implies that average measured trust may vary across ethnic groups or across waves even if there is actually no variation along these dimensions in the latent trust variables and vice versa. To assess the relevance of this phenomenon in the Afrobarometer, I perform three tests proposed by Bond & Lang (2019). Table 5 reports the results. The null hypothesis in the first test is that the ranking, in terms of average trust $\bar{T}_{ew}$, between ethnic groups (or between waves for a given group) is identified without any distributional assumption, i.e., without any cardinalization of the survey ordinal answers. The conditions for this nonparametric identification are quite stringent when more than two ethnic groups in a wave are compared, essentially requiring that the extreme categories (0 and 3, in this case) are not used by respondents. Not surprisingly, the null is always rejected, both between groups (i.e., pooling the two waves) and between waves (i.e., ethnic group by ethnic group).

Given the lack of nonparametric identification, I follow Bond & Lang and I assume that the random shock that alters an individual's reporting over time is normally distributed, so that an ordered Probit model provides the needed cardinalization. Rank identification under this normality assumption requires that the variance of the random shocks are equal across ethnic groups (or across waves in a given group), which is the null hypothesis in the second test. As reported in Table 5, when pooling all ethnic groups and waves, the null is always rejected; across waves it is instead rejected for between 68% and 80% of the groups. Even when the equal variances null is not rejected, rank order identification under the normality assumption requires a common reporting function, i.e., that the cutoff points that determine when an individual switches from one category of the Likert scale to the next are invariant across waves for a given ethnic group. This is the null hypothesis of the third test. This null is rejected for between 84% and 91% of the groups. Conditional on not rejecting the hypothesis that variances are equal across waves, the hypothesis that the reporting functions are also equal is rejected between 77% and 87% of the ethnic groups in the sample.

Table 5: Afrobarometer Trust Rank Order Identification Tests

	Nonparam. identific. of culture rank		Equal variances		Equal reporting: fraction reject	
	between-ethnic group	between-wave	between-ethnic group	btw-wave, fraction reject	btw-wave all	btw-wave equal var.
Trust relativ.	Reject	Reject	Reject	0.80	0.84	0.77
Trust neigh.	Reject	Reject	Reject	0.78	0.86	0.78
Trust council	Reject	Reject	Reject	0.68	0.91	0.87

Notes: The table reports, for three Afrobarometer trust measures (0-3 ordinal scale) available both in waves 3 and 5, the results from three tests proposed by Bond & Lang (2019). H_0 in the first test is that the trust rank order of ethnic groups either pooling waves (between-ethnic group) or across waves for a given group (between-wave) is nonparametrically identified. The test consists of checking whether the extreme categories are used (in which case H_0 is rejected) or not. H_0 in the second test is that, conditional on reporting errors that are normally distributed, the error variance is equal across ethnic groups or across waves. The test statistic is a χ^2, with degrees of freedom equal to the number of categories minus three, times the number of ethnic groups minus one. H_0 in the third test is that the reporting function is equal across ethnic groups or across waves. The test statistic is again a χ^2, with degrees of freedom equal to the number of categories minus three, times the number of instances an ethnic groups is surveyed minus one. Sample: 41,028 individuals from pooled Afrobarometer waves 3 and 5, in ethnic groups that are present in both waves.

5 *Conclusions*

While I perfectly replicate Nunn & Wantchekon's (2011) OLS and 2SLS findings in Afrobarometer wave 3 starting from the same raw survey data used by these authors, their central results are considerably weakened when individual trust measures across Sub-Saharan ethnic groups are computed using more recent Afrobarometer waves instead of wave 3. This problem, which could not manifest itself until more Afrobarometer waves containing the same trust measures used by NW were collected, casts doubts on the reliability of survey measures of cultural attitudes in economic analysis—especially when values and beliefs are gauged using Likert scales. Although the economics profession is well aware of this issue and is exploring ways to circumvent it (for example, Falk et al., 2018 develop experimentally-validated survey measures of trust and other preference traits using samples that span 76 countries), survey measures from sources such as the Afrobarometer or the World Values Survey are still appealing owing to their widespread availability at virtually no cost to researchers. This replication study suggests avoiding variables measured on Likert scales and, in any case, using simple diagnostic tools such as volatility tests and rank order identification tests (described in more detail in Zanella & Bellani, 2024, a companion paper to this one) to validate survey measures of trust or values and beliefs in general before employing them.

References

Bond, T. N. & Lang, K. (2019). "The Sad Truth about Happiness Scales". *Journal of Political Economy*, 127(4). DOI: 10.1086/701679.

Deconinck, K. & Verpoorten, M. (2013). "Narrow and Scientific Replication of 'The slave Trade and the Origins of Mistrust in Africa'". *Journal of Applied Econometrics*, 28(1): 166–169. DOI: 10.1002/jae.2309.

Falk, A., Becker, A., Dohmen, T., Enke, B., Huffman, D. & Sunde, U. (2018). "Global Evidence on Economic Preferences". *The Quarterly Journal of Economics*, 133(4): 1645–1692. DOI: 10.1093/qje/qjy013.

Lowes, S. (2022). "Culture in Historical Political Economy". In **Jenkins, J. A. & Rubin, J.**, editors, "The Oxford Handbook of Historical Political Economy", Oxford University Press. DOI: 10.1093/oxfordhb/9780197618608.013.47.

Murdock, G. P. (1959). Africa: Its Peoples and Their Culture History. McGraw-Hill Book Company.

Nunn, N. (2008). "The Long-term Effects of Africa's Slave Trades". *The Quarterly Journal of Economics*, 123(1): 139–176. DOI: 10.1162/qjec.2008.123.1.139.

Nunn, N. & Wantchekon, L. (2011). "The Slave Trade and the Origins of Mistrust in Africa". *American Economic Review*, 101(7): 3221–3252. DOI: 10.1257/aer.101.7.3221.

Tabellini, G. (2010). "Culture and Institutions: Economic Development in the Regions of Europe". *Journal of the European Economic Association*, 8: 677–716. DOI: 10.1111/j.1542-4774.2010.tb00537.x.

Zanella, G. & Bellani, M. M. (2024). "The Volatility of Survey Measures of Culture and its Consequences". *Economic Inquiry*, 62(2): 675–697. DOI: 10.1111/ecin.13202.