



**HAL**  
open science

## Understanding cultural persistence and change: a replication of Giuliano and Nunn (2021)

Simone Bertoli, Melchior Clerc, Jordan Loper, Èric Roca Fernández

► **To cite this version:**

Simone Bertoli, Melchior Clerc, Jordan Loper, Èric Roca Fernández. Understanding cultural persistence and change: a replication of Giuliano and Nunn (2021). 2024. hal-04563032

**HAL Id: hal-04563032**

**<https://uca.hal.science/hal-04563032>**

Preprint submitted on 29 Apr 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Bertoli, Simone; Clerc, Melchior; Loper, Jordan; Fernández, Èric Roca

**Working Paper**

## Understanding Cultural Persistence and Change: A Replication of Giuliano and Nunn (2021)

I4R Discussion Paper Series, No. 116

**Provided in Cooperation with:**  
The Institute for Replication (I4R)

*Suggested Citation:* Bertoli, Simone; Clerc, Melchior; Loper, Jordan; Fernández, Èric Roca (2024) : Understanding Cultural Persistence and Change: A Replication of Giuliano and Nunn (2021), I4R Discussion Paper Series, No. 116, Institute for Replication (I4R), s.l.

This Version is available at:  
<https://hdl.handle.net/10419/290342>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



No. 116

DISCUSSION PAPER SERIES

## **Understanding Cultural Persistence and Change: A Replication of Giuliano and Nunn (2021)**

Simone Bertoli

Melchior Clerc

Jordan Loper

Èric Roca Fernández

*This paper received a response:*

Giuliano, Paola, and Nathan Nunn. 2024. A Note on “Understanding Cultural Persistence and Change: A Replication of Giuliano and Nunn (2021). *I4R Discussion Paper Series* #117. Institute for Replication

**April 2024**

## I4R DISCUSSION PAPER SERIES

I4R DP No. 116

# **Understanding Cultural Persistence and Change: A Replication of Giuliano and Nunn (2021)**

**Simone Bertoli<sup>1</sup>, Melchior Clerc<sup>1</sup>, Jordan Loper<sup>1</sup>, Èric Roca Fernández<sup>1</sup>**

*<sup>1</sup>Université Clermont Auvergne, Clermont/France, Centre National de la Recherche Scientifique (CNRS), Paris/France, Institut de Recherche pour le Développement (IRD), Marseille/France, and Centre d'Études et de Recherches sur le Développement International (CERDI), Clermont/France*

APRIL 2024

Any opinions in this paper are those of the author(s) and not those of the Institute for Replication (I4R). Research published in this series may include views on policy, but I4R takes no institutional policy positions.

I4R Discussion Papers are research papers of the Institute for Replication which are widely circulated to promote replications and meta-scientific work in the social sciences. Provided in cooperation with EconStor, a service of the [ZBW – Leibniz Information Centre for Economics](#), and [RWI – Leibniz Institute for Economic Research](#), I4R Discussion Papers are among others listed in RePEc (see IDEAS, EconPapers). Complete list of all I4R DPs - downloadable for free at the I4R website.

I4R Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

### Editors

**Abel Brodeur**  
*University of Ottawa*

**Anna Dreber**  
*Stockholm School of Economics*

**Jörg Ankel-Peters**  
*RWI – Leibniz Institute for Economic Research*

E-Mail: [joerg.peters@rwi-essen.de](mailto:joerg.peters@rwi-essen.de)  
RWI – Leibniz Institute for Economic Research

Hohenzollernstraße 1-3  
45128 Essen/Germany

[www.i4replication.org](http://www.i4replication.org)

# Understanding cultural persistence and change: a replication of Giuliano and Nunn (2021)\*

Simone Bertoli<sup>a</sup>, Melchior Clerc<sup>a</sup>, Jordan Loper<sup>a</sup>, and Èric Roca Fernández<sup>a</sup>

<sup>a</sup>*Université Clermont Auvergne, CNRS, IRD, CERDI, F-63000, Clermont-Ferrand*

## Abstract

Giuliano and Nunn (2021), GN henceforth, provide econometric evidence that ancestral climatic variability is negatively associated with the current importance of tradition using a variety of data sources. This replication focuses on the results that use individual-level data and identifies major discrepancies between several econometric specifications described in the article and their corresponding code. We are able to correct most of these mistakes by realigning the code with the text. Once corrections are implemented, we obtain almost invariably a smaller and non-significant coefficient for climatic variability.

**Keywords:** cultural persistence, tradition, languages.

**JEL classification codes:** F22; Z13; N10; Q54.

---

\*Paola Giuliano and Nathan Nunn (2021), “Understanding cultural persistence and change,” *The Review of Economic Studies*, 88(4), 1541-1581; the code underlying this replication is available at: <https://www.dropbox.com/scl/fo/rcxpcp21mvouwxkwp0td/h?rlkey=ntuo54amd03isg9gpbumzh5xs&dl=0>; the Authors are grateful to Paola Giuliano and to Nathan Nunn for having exchanged with us on our replication exercise, and for their feedbacks, and they also gratefully acknowledge the comments received from Toman Barsbai, Abel Brodeur, Denis Cogneau, Yannick Dupraz, Jesús Fernández-Huertas Moraga, Francesca Marchetta, and Jérôme Valette, and the support received from the *Agence Nationale de la Recherche* of the French government through the program “*Investissements d’avenir*” (ANR-10-LABX-14-01); the usual disclaimers apply.

Simone Bertoli, UCA, CNRS, IRD, CERDI, 26 Av. Léon Blum, F63000, Clermont-Ferrand; email: simone.bertoli@uca.fr (corresponding author).

## Introduction

Giuliano and Nunn (2021), GN henceforth, present a model predicting that cultural persistence is lower the higher the intergenerational variability in climatic conditions, as this reduces the value of tradition. GN take this prediction to the data by examining how ancestral climatic variability influences the self-reported relevance of tradition, two revealed measures of its importance and measures of cultural persistence in current populations. In our replication, we analyze the code included in the replication package, comparing it with the econometric specifications described in GN, and we identify multiple instances where the two diverge for each of the six equations estimated with individual-level data. Correcting these inconsistencies has typically major implications for the size and significance of the estimated coefficients of climatic variability.

Our replication is organized in six sections, each examining one equation. We document the mistakes in each equation and report the estimates obtained when realigning the specifications and sample selection criteria with their descriptions in the text.

The original estimation sample for these equations contains many observations that should be excluded due to missing or incorrectly defined values for some variables, or because they do not meet the stated sample selection criteria. These errors artificially inflate the sample size, and correcting them necessarily leads to a reduction in the number of observations. The replication package does not include the necessary data cleaning and manipulation codes, which makes it difficult to fully understand and correct these mistakes.

We abstain from describing the mistakes in the code for equations using aggregate data.<sup>1</sup>

In what follows, the ordering of the equations reflects our assessment of the importance of rectifying the mistakes in the corresponding code or the logical relationship between two consecutive equations. All the references are based on the equation, table and page numbers in GN.

### Eq. (5) Speaking a foreign language at home

GN estimate the following equation on a sample of individuals born in the United States and reporting a foreign ancestry in the 2000 census:

---

<sup>1</sup>For instance, the dependent variable in Eq. (10), which aggregates individual-level outcomes and generates the results in Tables 8 and 10, inherits the coding errors for the dependent variable in Eq. (8).

$$I_{i,c,k}^{ForeignLang} = \alpha_k + \beta \text{Climatic Instability}_c + \mathbf{X}_c \Pi + \mathbf{X}_{c,k} \Omega + \mathbf{X}_i \Phi + \epsilon_{i,c,k} \quad (5)$$

“where  $i$  denotes an individual,  $c$  his/her ancestry, and  $k$  a location of residence (metropolitan areas). The dependent variable,  $I_{i,c,k}^{ForeignLang}$ , is an indicator that equals to one if a language other than English is spoken at home.  $\alpha_k$  denotes the inclusion of residence (i.e. metropolitan areas) fixed effects. [...] **We omit from the sample individuals from origin countries with English as an official language.**” (p. 1565, bold emphasis added).

### Eq. (5) Coding mistakes

1. The sample used to estimate Eq. (5) includes natives with ancestry in seven distinct countries or territories (out of 84) in which English is an official language.<sup>2, 3</sup>
2. The estimating sample includes 950,490 obs. (28.4%) for which the value of the location of residence  $k$  is “Not identifiable or not in an MSA”. A unique value of the fixed effect  $\alpha_k$  is incorrectly assigned to these observations, bundling them together. Moreover, the vector  $\mathbf{X}_{c,k}$ , includes a variable measuring the “fraction of those living in the same metropolitan area who are first-generation immigrants of the same ancestry” (p. 1565); this variable is incorrectly defined for these observations.

### Eq. (5) Relevance of fixing the coding mistakes

The coefficient  $\beta$  of Climatic Instability <sub>$c$</sub>  stands at -0.447 ( $p$ -value=0.010) in Col. (1) of Table 5. Fixing the two mistakes substantially alters the estimated coefficient: excluding observations corresponding to ( $i$ ) natives whose ancestry is a country where English is an official

<sup>2</sup>The seven countries or territories are Canada (71,315 obs., 2.1%), Puerto Rico (36,844 obs., 1.1%), Philippines (16,031 obs., 0.5%), India (5,127 obs., 0.2%), Samoa (1,062 obs., 0.03%), Pakistan (712 obs., 0.02%), and Eritrea (24 obs., 0.0007%); we identify countries and territories where English is an official language by cross-checking [Wikipedia contributors \(2023\)](#) and [Central Intelligence Agency \(2021\)](#), retaining only the cases in which both data sources report English as an official language.

<sup>3</sup>The 2000 census allows respondents reporting two ancestries; drawing on the data from 5% sample of the 2000 census ([Ruggles et al., 2023](#)), the most common foreign ancestry among natives aged 5 and above is German (1,446,921 obs.); among them, 673,611 (46.6%) also report a second ancestry, with the most frequent cases being “Irish” (192,584 obs., 13.3%) and “English” (122,363 obs., 8.5%); GN do not mention whether they only use the first ancestry, or if they also drop the natives with a second ancestry in a country in which English is an official language.

language, we obtain a coefficient equal to -0.008, which is not significant ( $p$ -value=0.967).<sup>4</sup> Dropping the observations at point ( $i$ ) and also those corresponding to ( $ii$ ) natives with unknown residence location, yields a coefficient equal to -0.003 ( $p$ -value=0.985).

### Eq. (6) Speaking a foreign language at home (living with parents)

“We estimate whether the intergenerational transmission of the tradition of speaking one’s ancestral language is affected by climatic instability of the ancestral country [...]:

$$\begin{aligned} I_{i,c,k}^{ForeignLang} = & \beta_1 \text{Climatic Instability}_c + \beta_2 I_{i,c,k}^{ParentForeignLang} + \\ & + \beta_3 I_{i,c,k}^{ParentForeignLang} \times \text{Climatic Instability}_c + \\ & + \mathbf{X}_c \Pi + \mathbf{X}_{c,k} \Omega + \mathbf{X}_i \Phi + \alpha_k + \epsilon_{i,c,k} \end{aligned} \quad (6)$$

[...]  $I_{i,c,k}^{ParentForeignLang}$ , which is an indicator variable that equals one if either the father or the mother speaks a foreign language at home.” (p. 1567, bold emphasis added).

### Eq. (6) Coding mistakes

Eq. (6) inherits the two mistakes described above with respect to Eq. (5), and its code also includes one additional mistake:

3. The sample split is done on the basis of the variable `relate`, describing the relationship of individual  $i$  to the household head. Thus, the subsample of individuals “Living with parents” includes all the individuals who are children of the household head, while the subsample “Not living with parents” includes all other individuals. For three-generation households, the split is incorrect when, for instance, a grandchild of the household head co-resides with at least one of her parents, or when the household also includes a parent of the household head; in these two cases, the grandchild and the household head are incorrectly classified by GN as “Not living with parents”. The correct split of the sample and the exact identification of a co-residing parent is possible by using the variables `momloc` and `poploc`, which are available from IPUMS USA (Ruggles et al., 2023).<sup>5</sup>

<sup>4</sup>Dropping just the natives with a Canadian ancestry already produces a non-significant value for  $\beta$  (-0.259,  $p$ -value=0.207).

<sup>5</sup>These variables provide the identifiers of the co-residing mother and father; see, for instance, <https://>



### Eq. (6) Incomplete replication package

The replication package’s incompleteness hinders our ability to fully address the three errors pertaining to Eq. (6). Specifically, the absence of individual identifiers in the replication data prevents us from merging it with the raw census data, which contain the variables `momloc` and `poploc`. The two mistakes that can be fixed are those shared with Eq. (5). Correcting these errors produces an estimated coefficient  $\beta_1$  for climatic instability, which is non-significant at conventional confidence levels in Cols. (2) to (7) of Table 5, except for Col. (3) where the  $p$ -value is 0.050. The coefficient  $\beta_3$  of the interaction term in Eq. (6) remains statistically significant in Cols. (6) and (7), but these estimates are still based on a split of the sample that is inconsistent with the description given in the text.

### Eq. (8) Indigenous populations speaking a traditional language

For a subsample of individuals who report a Native American tribal affiliation drawn from the 1930, 1990 and 2000 censuses,

“[the] estimating equation is:

$$I_{i,e,k}^{NativeLang} = \alpha_k + \beta \text{Climatic Instability}_e + \mathbf{X}_e \Pi + \mathbf{X}_i \Phi + \epsilon_{i,e,k} \quad (8)$$

where  $i$  denotes an individual,  $e$  his/her ethnic group, and  $k$  a location of residence (metropolitan area). The dependent variable,  $I_{i,e,k}^{NativeLang}$ , is an indicator that equals to one if the individual reports speaking an Indigenous language at home. The specification includes location (i.e. metropolitan areas) fixed effects,  $\alpha_k$ . Thus, the variation used to estimate  $\beta$  is across individuals from different Native American ethnic groups, but living in the same location. [in footnote 30] **The 1930, 1990 and 2000 US Censuses ask the following question: ‘Does the person speak a language other than English at home?’ If yes, the person indicates which language.**” (p. 1571, bold emphasis added).

### Eq. (8) Coding mistakes

1. The questionnaire of the 1930 census did not include the two questions necessary to define the dependent variable in Eq. (8).<sup>6</sup> The sample used in Col. (1) of Table 7 to es-

---

usa.ipums.org/usa-action/variables/MOMLOC#availability\_section (last accessed on May 31, 2023); the use of these variables would also, for instance, avoid incorrectly identifying a step-father as the father of the child, something that can occur using `relate`.

<sup>6</sup>See [https://usa.ipums.org/usa-action/variables/LANGUAGE#availability\\_section](https://usa.ipums.org/usa-action/variables/LANGUAGE#availability_section) (last accessed on May 31, 2023).

timate Eq. (8) includes 11,468 obs. (9.0%) drawn from this census, and corresponding to 83 Native American ethnic groups. Until September 2021, IPUMS USA incorrectly coded the variable `language` for the 1930 census, using the information on the mother tongue.<sup>7</sup> However, the question about the mother tongue should have been asked in 1930 only to foreign-born individuals,<sup>8</sup> even though some enumerators probably failed to respect this restriction.<sup>9</sup> The replication file reveals that the dependent variable is always equal to 0 in 1930 for 68 of these 83 ethnic groups, and two groups (Navajo and Hopi) account for the great majority (1,860 out of 1,951) of the observations for which the dependent variable in Eq. (8) is equal to 1 in 1930.<sup>10</sup> This, in turn, strongly suggests that the question on the mother tongue was not systematically asked to all Native Americans in the 1930 census.<sup>11</sup>

2. The two questions on language used to define  $I_{i,e,k}^{NativeLang}$  are asked only to individuals aged 5 and above in the 1990 and 2000 censuses.<sup>12</sup> The estimation sample includes 10,364 obs. (8.1%) corresponding to children aged 0 to 4, for which  $I_{i,e,k}^{NativeLang}$  is always equal to 0, while it should have been missing.
3. The question on employment status, which are used to define an element of the vector  $\mathbf{X}_i$  in Eq. (8), is only asked to individuals aged 16 and above in the 1990 and in the 2000 census;<sup>13</sup> the estimation of Eq. (8) treats the observations for which `empstat` is not available as a distinct value of this categorical variable, differently from what is stated at p. 1572, where the Authors write that labour-force-status fixed effects take three values: employed, unemployed, and outside of the labour force.<sup>14</sup> The sample

---

<sup>7</sup>See [https://usa.ipums.org/usa-action/revisions#revision\\_02\\_09\\_2021](https://usa.ipums.org/usa-action/revisions#revision_02_09_2021) (last accessed on October 23, 2023); we are grateful to Paola Giuliano and Nathan Nunn for pointing this out to us.

<sup>8</sup>See [https://usa.ipums.org/usa-action/variables/MTONGUE#description\\_section](https://usa.ipums.org/usa-action/variables/MTONGUE#description_section) (last accessed on October 23, 2023).

<sup>9</sup>Email exchange with Paola Giuliano and Nathan Nunn, who had contacted IPUMS USA on this point.

<sup>10</sup>The share of observations where the dependent variable is equal to 1 stands at 91.0 and 99.2 percent respectively for Navajo and Hopi in 1930.

<sup>11</sup>Notice that the replication file only includes the dependent variable, and not the language associated to each individual.

<sup>12</sup>See [https://usa.ipums.org/usa-action/variables/LANGUAGE#universe\\_section](https://usa.ipums.org/usa-action/variables/LANGUAGE#universe_section) (last accessed on June 6, 2023).

<sup>13</sup>See [https://usa.ipums.org/usa-action/variables/EMPSTAT#universe\\_section](https://usa.ipums.org/usa-action/variables/EMPSTAT#universe_section) (last accessed on October 23, 2023).

<sup>14</sup>Notice that this differs from what GN do when estimating Eqs. (5) and (6) with the data from the

used in Col. (1) of Table 7 to estimate Eq. (8) includes 35,316 obs. (27.6%) with a missing labor force status.

4. The estimation sample includes 81,371 obs. (63.6%) for which the location  $k$  is “Not identifiable or not in an MSA”.<sup>15</sup> Hence, individuals for whom the variable `metarea` takes this value are effectively bundled together, as they share the same location fixed-effect  $\alpha_k$ . Thus, the statement “the variation used to estimated  $\beta$  is across individuals from different Native American ethnic groups, but living in the same location” (p. 1571) is factually incorrect.

### Eq. (8) Relevance of fixing the coding mistakes

The coefficient of Climatic Instability<sub>e</sub> stands at -1.097 ( $p$ -value=0.004) in Col. (1) of Table 7, which brings Eq. (8) to the data. We jointly correct the four mistakes described above excluding observations (*i*) drawn from the 1930 census, (*ii*) for which the location of residence is unknown, and (*iii*) corresponding to individuals aged 0 to 15. This reduces the sample from 128,005 to 34,092 obs., with an ensuing drop in the number of distinct ethnic groups from 83 to 19.<sup>16</sup> These corrections produce an estimated coefficient for Climatic Instability<sub>e</sub> which stands at -0.364 ( $p$ -value=0.192).<sup>17</sup>

---

2000 census, as the estimation sample is restricted to individuals aged 16 and above, for whom the variable `empstat` is available; Paola Giuliano and Nathan Nunn wrote to us that the choice in Eq. (8) is driven by the fact that the variable `empstat` is coded also for individuals aged below 16 in the 1930 census, a data source that is not used for Eqs. (5) and (6).

<sup>15</sup>In particular, this occurs for 10,966 out of 11,468 obs. drawn from the 1930 census.

<sup>16</sup>The reduction in the number of distinct ethnic groups is exclusively due to dropping the observations from the 1930 census, which includes a much more detailed coverage of Native American tribes.

<sup>17</sup>We also obtain a non-significant coefficient for climatic variability when we remedy the four errors in lines 153 and 155, corresponding to Cols. (2) and (3) in Table 7. Still, the sample split between individuals “Living with parents” and “Not living with parents” is incorrect, as discussed above with respect to Eq. (6).<sup>18</sup>

## Eq. (9) Speaking a traditional language (living with parents)

“We again explore the transmission of this cultural practice [speaking an Indigenous language at home] from parents to children by estimating:

$$\begin{aligned} I_{i,e,k}^{NativeLang} = & \alpha_k + \beta_1 \text{Climatic Instability}_e + \\ & + \beta_2 I_{i,e,k}^{ParentNativeLang} + \beta_3 I_{i,e,k}^{ParentNativeLang} \times \text{Climatic Instability}_e + \\ & + \mathbf{X}_e \Pi + \mathbf{X}_i \Phi + \epsilon_{i,e,k} \end{aligned} \quad (9)$$

where all variables are defined as in Eq. (8) and  $I_{i,e,k}^{ParentNativeLang}$  is an indicator variable that equals one if individual  $i$ 's parents speak an indigenous language at home.” (p. 1572).

Eq. (9) is estimated for individuals drawn from the 1930, 1990 and 2000 censuses reporting an affiliation to a Native American tribe and “living with parents” (p. 1572).

## Eq. (9) Coding mistakes

Eq. (9) inherits the four coding mistakes described above for Eq. (8),<sup>19</sup> plus an additional coding mistake shared with Eq. (6):

5. The split of the sample and the identification of the co-resident parents for individuals living with parents is done on the basis of the variable `relate`; the remarks related to the fourth mistake for Eq. (6) also apply here.

## Eq. (9) Relevance of fixing the coding mistakes

Fixing the first four mistakes for Cols. (4)–(7) of Table 7 yields a non-significant coefficient  $\beta_1$  for `Climatic Instabilitye`, and non-significant values for the coefficient  $\beta_3$  of the interaction term (except in Col. (6), -1.041,  $p$ -value=0.047, where the estimation sample consists of 1,791 obs., compared to the 25,794 obs. in GN). However, these estimates are still based on a code that is inconsistent with the original paper, as the variable `relate` does not allow correctly splitting the sample and identifying the co-resident parent(s).

## Eq. (9) Incomplete replication package

The replication package does not include the code to transition from the raw data to the sample used in Table 7, and the replication data do not include the individual identifiers that

<sup>19</sup>The first three mistakes also extends to the definition of the variable  $I_{i,e,k}^{ParentNativeLang}$  in 1930.

would allow merging them with the underlying raw data. This prevents us from assessing the relevance of the fourth mistake described above, and of simultaneously fixing the four mistakes related to Eq. (9).

## Eq. (2) Importance of tradition

GN write that, when estimating Eq. (2) using data from the World Values Survey (WVS):

“After matching **respondents’ self-reported mother tongue from the WVS** with ethnicity from the *Ethnographic Atlas*, we estimate:

$$\text{Tradition}_{i,e,c} = \alpha_c + \beta_2 \text{Climatic Instability}_e + \mathbf{X}'_i \Phi + \mathbf{X}'_e \Omega + \epsilon_{i,e,c} \quad (2)$$

where  $i$  denotes an individual who is a member of the historical ethnic group  $e$  and lives in country  $c$ . Tradition $_{i,e,c}$  is the self-reported importance of tradition, measured on an 1 – 6 integer scale and increasing in the importance of tradition.” (p. 1557, bold emphasis added).

## Eq. (2) Coding mistake

1. The WVS does not provide information on mother tongue, but just on the language that the respondent normally speaks at home.<sup>20</sup> Because respondents are typically presented with a short list of languages from which to choose from, a non-negligible share report speaking “other” languages that are not separately recorded in the data.

## Eq. (2) Relevance of fixing the mistake

Eq. (2) produces the results presented in Table 2 in GN. The coefficients for climatic variability in Cols. (1) and (2) of Table 2, differing in terms of control variables, stand at -0.839 ( $p$ -value=0.002) and at -0.624 ( $p$ -value=0.034), respectively.<sup>21</sup> The replication package does not include the code used to match the respondents’ language in the WVS to an ethnic group in the *Ethnographic Atlas*.

To evaluate the implications of mistaking the language spoken at home for the mother tongue, one may naturally focus on the subset of observations where the two are more likely

<sup>20</sup>In that sense, [Becker \(forthcoming\)](#) writes that “the WVS does not record any useful information on respondents’ ethnicity. However, the WVS often records the language a respondent speaks at home” (p. 12).

<sup>21</sup>The remaining four data columns in Table 2 correspond to different (and slightly more detailed) data sources on ethnic groups.

to diverge: respondents of immigrant origin. Thus, we re-evaluate the estimates in Table 2 but dropping the observations that correspond to these respondents. First, we define them based on maternal foreign-born status, i.e., respondents with a mother born in a country other than where they are surveyed.<sup>22</sup> To overcome the constraint represented by the absence of individual identifiers and of the language in the replication data, and of the code associating languages to ethnic groups, we combined the raw data from the WVS with the replication data for Table 2 to infer the association between a respondent's language and an ethnic group (we refer here to our work on `Table2_A.dta`). Unfortunately, the version of the WVS used in the original paper is no longer available on-line,<sup>23</sup> and our code might fail to correctly identify the association of each respondent with an ethnic group.

To assuage this concern, we first use our entire reconstructed sample, which yields results that are very close to those in the original paper. Furthermore, results remain stable when we restrict the sample to country-wave pairs for which we can identify respondents of immigrant origin. This reassures us about the quality of our replication data and make us confident that any changes we may observe by dropping individuals whose mother is foreign-born are due exclusively to the latter. Once we drop the small proportion of respondents of immigrant origin (7.5% of the sample), the coefficients become -0.576 ( $p$ -value=0.076) and -0.541 ( $p$ -value=0.136). Similar results emerge when (i) we repeat the analysis for the remaining data columns in Table 4, using the files `Table2_B.dta` and `Table2_C.dta`, or (ii) we adopt different definitions of respondents of immigrant origin, or when we combine (i) and (ii).<sup>24</sup>

## Eq. (4) Within-group marriage

“[The] sample comprises all married women in all waves of the March Supplement of the Current Population Survey (CPS) with **at least one** parent who was born outside the US. [...]

$$I_{i,c,k}^{IngroupMarriage} = \alpha_k + \beta_3 \text{Climatic Instability}_c + \mathbf{X}_c \Pi + \mathbf{X}_{c,k} \Omega + \mathbf{X}_i \Phi + \epsilon_{i,c,k} \quad (4)$$

<sup>22</sup>The WVS does not provide information on the country of birth, and (surprisingly enough) the number of country-wave pairs that allow identifying foreign-born respondents is much smaller than the number of those that allow identifying respondents with a foreign-born mother.

<sup>23</sup>Bahrain belongs to the sample used in GN, but it is omitted from the versions of the WVS that are currently available on-line; for our purposes, this is immaterial because all respondents in Bahrain belong to the same ethnic group in the replication data, and the inclusion of country dummies  $\alpha_c$  in Eq. (2) implies that these observations provide no identifying variability.

<sup>24</sup>These additional results are available at <https://www.dropbox.com/scl/fo/rcxpcp21mvouwxbkwp0td/h?rlkey=ntuo54amd03isgbumzh5xs&dl=0>, in the folder for Eq. (2).

[the sample includes] individuals who were born in the US, but whose **parents are immigrants who were born outside the US** [...] The outcome of interest,  $I_{i,c,k}^{IngroupMarriage}$ , is an indicator variable equal to one if **an individual's spouse was born in origin country  $c$  or if his or her mother or father was born in country  $c$ .**" (p. 1562, bold emphasis added).

#### Eq. (4) Coding mistakes

1. The estimation samples in the four data columns of Table 4 include individuals that reside in metropolitan areas that are not separately identified, e.g., 6,335 obs. (18.4%) in Col. (1). The control variable measuring "the fraction of the population in the same metropolitan area as the individual who are first or second-generation immigrants from the same country of origin" (p. 1562) cannot be meaningfully defined for these observations. Similarly, the statement that "[b]y examining individuals who live in the same location, we are able to hold constant the contemporaneous local environment, while examining the effects of an individual's ancestral environment" (p. 1564) is factually incorrect.
2. The dependent variable in Eq. (4) is created in a portion of the code not included in the replication package, and its definition requires information on the country of birth of the spouse, and of the parents-in-law. The March Supplement of the CPS provides these pieces of information exclusively for individuals co-residing with their spouses,<sup>25</sup> but the estimation samples in Table 4 include married individuals not co-residing with the spouse, e.g., 1,007 obs. (2.7%) in Col. (1), with 180 of these having the dependent variable equal to 1. Thus, it is not possible to create the dependent variable in a way consistent with its description provided at p. 1562 for married individuals whose spouse is absent.
3. The code corresponding to Eq. (4) restricts the sample to natives with at least one parent who is born abroad, e.g., 13,840 obs. (38.4%) in Col. (1) of Table 4 correspond to individuals with a native parent. The quote above includes two contradictory descriptions of the sample selection criteria, which are both inconsistent with the code. The first part of the quote above imposes the restriction that at least one parent of

---

<sup>25</sup>See, for instance, User Note 1 in Appendix G of the Codebook of the 1994 March Supplement of the CPS, available at: <https://cps.ipums.org/cps/resources/codebooks/cpsmarapr94.pdf> (last accessed: June 25, 2023).

individual  $i$  is foreign-born, with no reference to the fact that individual  $i$  should be a native or not. Conversely, the second part of the quote above requires that both parents must be born abroad, and that the individuals included in the sample must be born in the United States.

#### Eq. (4) Relevance of fixing the coding mistakes

The estimated values of the coefficient for climatic variability in the four columns of Table 4 stand at -0.274 ( $p$ -value=0.081) and at -0.492 ( $p$ -value=0.007) for married women, and at -0.102 ( $p$ -value=0.457) and -0.250 ( $p$ -value=0.093) for married men, respectively. When we drop married men or women ( $i$ ) for whom the metropolitan area is not identified, and ( $ii$ ) who do not co-reside with their spouse, the estimated coefficients that we obtain are -0.330 ( $p$ -value=0.092) and -0.528 ( $p$ -value=0.013) for married women, and -0.146 ( $p$ -value=0.381) and -0.294 ( $p$ -value=0.099) for married men. When we also drop ( $iii$ ) individuals with a native parent,<sup>26</sup> the estimated coefficients that we obtain are -0.343 ( $p$ -value=0.156) and -0.476 ( $p$ -value=0.078) for married women, and -0.091 ( $p$ -value=0.666) and -0.123 ( $p$ -value=0.579) for married men, i.e., only one coefficient is significant at the 10 percent confidence level.

### Concluding remarks

We are unable to reproduce the results obtained by [Giuliano and Nunn \(2021\)](#) with individual-level data. Correcting the various coding and factual mistakes considerably weakens the evidence of a negative association between ancestral climatic variability and the importance of tradition.

### References

BECKER, A. (forthcoming): “On the Economic Origins of Restricting Women’s Promiscuity,” *The Review of Economic Studies*.

---

<sup>26</sup>We are unable to extend the sample to also include foreign-born individuals with at least one foreign-born parent, consistently with the first part of the quote from p. 1562 reported above, as we do not know whether the replication files, which also include observations corresponding to foreign-born individuals, include all of them.



CENTRAL INTELLIGENCE AGENCY (2021): “The World Factbook 2021,” [Online; accessed 20-June-2023].

GIULIANO, P. AND N. NUNN (2021): “Understanding cultural persistence and change,” *The Review of Economic Studies*, 88, 1541–1581.

RUGGLES, S., S. FLOOD, M. SOBEK, D. BROCKMAN, G. COOPER, S. RICHARDS, AND M. SCHOUWEILER (2023): “IPUMS USA: Version 13.0 [dataset],” Minneapolis, MN: IPUMS, <https://doi.org/10.18128/D010.V13.0>.

WIKIPEDIA CONTRIBUTORS (2023): “List of official languages by country and territory — Wikipedia, The Free Encyclopedia,” [Online; accessed 20-June-2023].