Received: 17 December 2022 | Accepted: 17 November 2023

DOI: 10.1111/ecin.13202

ORIGINAL ARTICLE

Economic Inquiry

The volatility of survey measures of culture and its consequences

Giulio Zanella[1] | **Marina M. Bellani**[2]

[1]University of Bologna, Bologna, Italy

[2]Eni Gas e Luce S.p.A., Milano, Italy

Correspondence
Giulio Zanella.
Email: giulio.zanella@unibo.it

Abstract

Measures of cultural attitudes derived from the World Values Survey are often used in economics to investigate the relation between individual culture and aggregate output. We show (i) that these measures are constructed from underlying variables that exhibit within-country volatility and time trends; and (ii) that such lack of persistence implies fragile correlations between cultural measures and output. These points are illustrated by revisiting prominent studies in this literature.

KEYWORDS
culture, economic development, world values survey

JEL CLASSIFICATION
O12, O43, Z1

1 | INTRODUCTION

The relation between individual culture and aggregate economic outcomes is often investigated using measures of cultural attitudes derived from survey questions. Two precursors are Knack and Keefer (1997) and La Porta et al. (1997), who studied the association between trust—as measured by the World Values Survey (WVS)—and macroeconomic performance. Subsequent research has expanded the set of cultural measures constructed from the WVS beyond trust in a quest for a variety of proxies capturing richer and more complex cultural traits. Two examples are Tabellini (2010) and Gorodnichenko and Roland (2011). The first paper relates four measures of values and beliefs (trust, respect, obedience, and internal locus of control) to cross-region output in Europe; the second paper employs 16 variables to build six measures of cultural attitudes (trust, hard work and thrift, tolerance, public good provision, equality, and market orientation) and relates them to cross-country output.

Thanks to the availability of more data collected since these early papers were published, we show that the survey cultural proxies that they employed exhibit substantial within-country volatility and trends, that is, are *not* stable. Considering the nature of the survey instrument and the possibility of cultural change, this feature is not surprising and

Abbreviations: EU, European Union; GDP, Gross Domestic Product; GVA, Gross Value Added; IV, Instrumental Variable; NUTS, Nomenclature of Territorial Units for Statistics; OLS, Ordinary Least Squares; OLS-FE, Ordinary Least Squares-Fixed Effects; PC, Principal Component; PPP, Purchasing Power Parity; SD, Standard Deviation; WVS, World Values Survey.

Managing Editor: William Branch.

is not a source of concern per se. What is worrisome is that the correlation between some cultural measures and output is altered when using more recent WVS waves.[1] We emphasize that the issue that we raise is *not* about Tabellini (2010) and Gorodnichenko and Roland (2011) or their research procedures. Our point is about survey measures of values and beliefs: if these measures are sufficiently correlated with the slow-moving component of culture that researchers are interested in, then shifting measurement a few years forward should not make an appreciable difference. However, we find that it does.[2]

As Lowes (2022) points out in an up-to-date survey, measuring culture can be difficult. Survey questions are a traditional strategy that is still appealing because the data are readily available (e.g., Berggren et al., 2022; Deutsch, 2022; Desmet et al., 2022; Edwards et al., 2022; Hasnat, 2022, just to mention a few recent papers). When addressing the methodological limits of this strategy, three problems have received most of the scholarly attention: the samples' representativeness, the endogeneity of the measures themselves, and the survey questions' incentive compatibility (Falk et al., 2018; Fernández, 2008; Guiso et al., 2011). The aspect that we address—the stability of survey measures over time—has received far less attention, but is important because such measures incorporate both slow- and fast-moving components of culture (Guiso et al., 2006; Roland, 2004). The former are traits that reflect values and beliefs transmitted from one generation to the next via family or social interactions and that do not change rapidly (Acemoglu & Robinson, 2021; Bisin & Verdier, 2011); the latter can be viewed as the expressions of such traits that reflect short-term social, economic, or political dynamics and that, in contrast, may change quickly.

To be sure, the stability of survey-based cultural measures is not, per se, a necessary condition for such measures to be meaningful predictors of long-run aggregate economic outcomes. Values and beliefs do evolve in response to changes in economic, social, or political conditions; such evolution has been documented by, among others, Inglehart and Welzel (2005), Algan and Cahuc (2010), Giavazzi et al. (2013), Ananyev and Guriev (2018), and Giavazzi et al. (2019). The question is whether survey measures reflect *enough* of the persistent component of cultural attitudes. The longitudinal stability (or lack thereof) of the survey measures used to proxy for persistent cultural traits is informative about possible measurement error with respect to the targeted slow-moving component. Such error may induce spurious correlation between survey proxies and output at any particular time and/or reduce the predictive ability of plausibly exogenous instruments. If the target cultural traits are slow moving, then it is crucial to understand whether the measures are stable across waves of the survey instrument.

Focusing on the nine cultural proxies used by Tabellini (2010) and Gorodnichenko and Roland (2011), we document that they are not: there is substantial volatility across the six fully available waves of the WVS. Consider, for example, a simple measure of volatility: the within-country standard deviation of the log of conditional cultural proxies across WVS waves (a measure of the average, percentage deviation from the country mean across waves, net of changing demographic and socioeconomic characteristics); the average within-country volatility is about 30% for trust, and about 20% for respect, obedience, hard work and thrift, and equality. Decomposing the variance of the individual-level cultural measures reveals, in line with the results of Falk et al. (2018), that the average between-country variation is about 12%. However, we also find that the average between-wave variation is nearly 60% of the average between-country variation, which confirms that there is substantial volatility. Moreover, we establish that such volatility is not simply noise: most of the cultural measures exhibit significant time trends within countries. These trends are such that the country rankings change widely over time. Different reasons may explain such lack of persistence. It is possible that surveys simply capture noisy measures of the values and beliefs that the interviewer aims to elicit. It is also possible that certain beliefs, as already demonstrated in the literature, vary endogenously over time and so are not really deep cultural attitudes. In any case, the volatility of survey measures of culture may induce fragility in the measured connection between culture and output. Moreover, as demonstrated in a different context by Bond and Lang (2019), when survey questions elicit values and beliefs through ordinal scales, as is typical of many WVS questions, it may be impossible to rank two groups (or the same group over time) in terms of a particular cultural attitude. That is, it is possible that average measured culture does not vary (either across countries or across waves for a given country) while cultural attitudes do, or vice versa. In this case the connection between cultural attitudes and their survey proxies would be tenuous.

Using WVS waves that were not available at the time the research in Tabellini (2010) and Gorodnichenko and Roland (2011) was undertaken, we find that the coefficients of the linear regressions of output level (or growth) on some cultural measures either change in sign, or decline in magnitude (becoming statistically insignificant), or become imprecisely estimated. Note that this is not a matter of level of aggregation, as the same fragility is observed both when using region-level averages and country-level averages of individual cultural attitudes. Other measures, most notably trust, are instead robust. In that case, the trust measure's lack of persistence seems immaterial.[3] Applying three tests

proposed by Bond and Lang (2019), we conclude that the measurement of culture on ordinal scales may be an important part of the fragility problem. This interpretation is consistent with our finding that cultural proxies derived from binary variables are more robust than those derived from ordinal scales. Our empirical exercise reveals also that possible solutions to the problem (when present), which include employing conditional measures of culture to account for sampling variability across waves and/or using instrumental variables as filters for the slow-moving component of culture present in answers to survey questions, may not be effective in practice. For example, when using more recent WVS waves, Tabellini's (2010) historical instruments become weak. These results suggest caution when using survey information to proxy for slow-moving cultural traits. In order to provide some guidance in avoiding the perils of survey cultural measures, we discuss possible diagnostic tools and solutions that may be of use to empirical researchers.

The rest of our paper proceeds as follows: Section 2 documents the volatility of cultural measures derived from the WVS; Section 3 and Section 4 illustrate the consequences of that volatility by replicating, respectively, the analyses of Gorodnichenko and Roland (2011) and Tabellini (2010) with more recent WVS data. Section 5 concludes.

2 | VOLATILITY OF WVS CULTURAL MEASURES

When using survey questions to measure cultural attitudes in a certain geographic area, common practice is as follows. Let c_{igt} denote a variable derived from the answer given by individual i in geographic area g in survey wave t. The question's purpose is to capture a certain individual value or belief. Suppose, for example, that individual i is asked: "Generally speaking, would you say that most people can be trusted, or that you can't be too careful in dealing with people?" If i answers "most people can be trusted", then $c_{igt} = 1$ (otherwise, $c_{igt} = 0$). Or individual i may be asked to indicate on a Likert scale how much freedom of choice and control in life one has over the way life turns out, with possible answers "not at all" ($c_{igt} = 1$), "just a little" ($c_{igt} = 2$) and so on up to "a great deal" ($c_{igt} = 10$). The corresponding cultural attitude imputed to area g is constructed as

$$c_{gt} = \sum_{i=1}^{n_{gt}} w_{igt} c_{igt}, \tag{1}$$

where n_{gt} is sample size in area g during survey wave t, and w_{igt} denotes sampling weights that sum to 1 and render the average c_{gt} representative of the population in area g. If sampling weights are not available or if the sampling design is balanced, then $w_{igt} = 1/n_{gt}$.

Equation (1) defines an unconditional measure of culture. Yet researchers often work with a *conditional* measure that eliminates the influence of different socioeconomic characteristics of respondents across geographic areas (in a cross-section) or sampling differences across waves in case sampling weights are not available (in a panel) or the effects of demographic and socioeconomic evolution on measured cultural attitudes (i.e., composition effects). For example, in Tabellini (2010), the index g refers to the regions of Europe; the author constructs conditional measures of culture by projecting an individual-level cultural proxy onto a set of individual covariates $\mathbf{x}$ and regional dummies d_g via the following linear regression model,

$$c_{igt} = \beta_{0t} + \beta_{1t}\mathbf{x}_{igt} + \sum_{g=1}^{G} \beta_{2,gt} d_g + \varepsilon_{igt}, \tag{2}$$

where G is the number of geographic units. Tabellini then takes the sum of the estimated constant and the regional dummy's estimated coefficient as the conditional measure of culture in that region.[4] That is, the cultural attitude imputed to area g is constructed as

$$c_{gt} = \widehat{\beta}_{0t} + \widehat{\beta}_{2,gt}. \tag{3}$$

In either case c_{gt} will incorporate both slow- and fast-moving components of culture (Guiso et al., 2006; Roland, 2004). The former are traits that reflect values and beliefs transmitted from one generation to the next via family or social interactions and that do not change rapidly at the aggregate level; the latter can be viewed as the expressions of such traits that reflect short-term social, economic, or political dynamics and that, in contrast, may

change quickly. If c_{gt} must be used to measure slow-moving cultural attitudes, then we should want the difference $c_{gt} - c_{gt-1}$ to be small (relative to c_{gt-1}) across any two consecutive survey waves t and $t-1$. One would similarly expect c_{gt} not to contain an important time trend because cultural evolution is slow and so cultural traits should persist over relatively short time horizons. As discussed in Section 1, the stability of survey measures is not a necessary condition for them to be adequate cultural proxies in a regression framework; one reason is that the high-frequency component of c_{gt} 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* the slow-moving component of culture, which is what surveys aim to capture. Otherwise, empirical associations with economic outcomes may depend too strongly on exactly when culture is measured, and thereby exhibit statistical fragility. A related problem analyzed by Bond and Lang (2019) for happiness scales is that survey questions gauge cultural attitudes by means of ordinal scales. When there are more than two categories, computing average culture as in Equation (1) or (3) requires a researcher to cardinalize ordinal survey answers. The existence of infinite possible cardinalizations implies that average measured culture may vary across countries or over time even if there is actually no variation along these dimensions in the target average cultural attitude and vice versa.

To illustrate these measurement problems in a simple framework, consider a persistent, latent cultural trait $c^*_{ig} \in \mathbb{R}$. The researcher only observes survey measure c_{igt}, an ordinal variable with $S+1$ categories that is produced by an individual's reporting function,

$$c_{igt} = \begin{cases} 0 \text{ if } & c^*_{ig} + \eta_{igt} < c^1_{igt} \\ 1 \text{ if } & c^1_{igt} \le c^*_{ig} + \eta_{igt} < c^2_{igt} \\ \vdots & \vdots \\ S \text{ if } & c^S_{igt} \le c^*_{ig} + \eta_{igt}, \end{cases} \tag{4}$$

where c^s_{igt}, for $s = 1, \cdots S$, are fixed cutoff points that determine when an individual switches to a different category and η_{igt} is a random shock that alters an individual's reporting over time. In the extreme case in which each η_{igt} has a time-invariant mean and a degenerate distribution, aggregate cultural measure c_{gt} will exhibit no volatility at all (although, as demonstrated by Bond & Lang, 2019, it may still be impossible to rank countries by their average latent culture in a cross section). Otherwise, c_{gt} will typically change over time. Two possibilities are relevant to our analysis. First, η_{igt} has a time-invariant mean and positive variance, that is, individuals just report their persistent cultural attitudes noisily. If such errors are correlated (e.g., because individuals in a given area are subject to common shocks) then c_{gt} may be characterized by some volatility despite the persistence of c^*_{ig}. Moreover, the existence of multiple cardinalizations that are consistent with a given ordering produced by the survey question implies that $c_{gt} - c_{gt-1}$ may bear no connection with changes in the underlying average latent cultural trait. Second, the mean of each η^s_{igt} changes over time, in which case c_{gt} may be characterized by a time trend. In either case the cultural ranking of geographic units may vary over time. We next show how these problems manifests themselves in the data.

2.1 | The volatility problem

Consider the WVS cultural measures employed by Tabellini (2010) and by Gorodnichenko and Roland (2011), two influential papers on the cultural roots of economic development and growth.[5] Tabellini (2010) selects four raw cultural measures from survey questions about beliefs and values that (a) capture salient, persistent traits of peasants in southern Italy and (b) are related to the "backwardness" of that region (as described by Banfield, 1958; Putnam, 1993). These four measures are labeled Trust, Respect, Obedience, and Control. Gorodnichenko and Roland (2011) aggregate WVS variables to construct six indexes of cultural traits that (a) characterize individualist versus collectivist cultures and (b) are related to, respectively, innovation and coordination. These six indexes are labeled Trust (which coincides with the measure of Tabellini), Work, Tolerance (which subsumes but differs from the Tabellini's Respect measure), Public good, Equality, and Market. The resulting nine variables and the cultural traits for which they proxy are summarized in Table 1. To evaluate the persistence of these cultural measures, we use all of the fully available waves of the WVS to construct, as in Equations (1)–(3), both unconditional measures (applying sampling weights) and conditional measures (by constructing individual covariates in $\mathbf{x}_{ig}$ from survey information on age, gender, marital status,

TABLE 1 WVS cultural measures and their rationale.

Measure	Definition	Proxies for
Trust	% Thinking that most people can be trusted	Cooperative-culture (generalized trust)
Respect	% Thinking "tolerance and respect for other people" is an important quality children should learn	Tolerant-culture (generalized respect)
Obedience	% Thinking that "obedience" is an important quality children should learn	Submissive-culture (give up opportunities)
Control	Average agreement with statement "what happens in one's life depends on one's free choices"	Entrepreneurship (reap opportunities)
Work	Average of %'s thinking that "thrift and saving" and "determination and perseverance" are important qualities children should learn	Individualist-culture (hard work and thrift)
Tolerance	Average of %'s thinking that "tolerance and respect for other people" and "imagination" are important qualities children should learn, and (with negative weight) that "people of different race" or "immigrants" are undesirable neighbors	Individualist-culture (generalized-respect; innovative attitude)
Public good	Average of %'s thinking "unselfishness" is an important quality children should learn, and (with negative weight) degree "cheating on taxes" and "avoiding a fare on public transport" are justifiable	Collectivist-culture (coordination; public good provision)
Equality	Average of agreement with statement "incomes should be made more equal" and % thinking it is not fair for a more efficient and more reliable secretary to be paid more	Collectivist-culture (aversion to inequality)
Market	Average of agreements with statements "competition is good", "private ownership should be increased", "the government should take less responsibility" and of % thinking it is not fair for a more efficient and more reliable secretary to be paid more	Individualist-culture (market orientation)

Note: The first four rows summarize the World Values Survey cultural proxies employed by Tabellini (2010). The next five rows summarize the five additional World Values Survey proxies used by Gorodnichenko and Roland (2011).

education, social class, and health status) for 110 countries. The averages are scaled so that they lie within the [0,100] interval.

The volatility of these nine cultural measures is assessed from four different perspectives. First, in Figure 1 we plot the value taken by an unconditional or conditional cultural measure in wave t of the WVS (vertical axis) against the value taken by the same measure in WVS wave $t-1$ (horizontal axis). There are sizable deviations from the (dashed) 45° line; these deviations indicate that cultural proxies may change considerably from one wave to the next—even after we account for the effect of varying demographic and socioeconomic characteristics. Note that such conditioning does not necessarily reduce volatility.

Second, we quantify such variability by computing the within-country standard deviation of the log of conditional culture—that is, in analogy with the volatility measure commonly used in finance to evaluate an asset's returns. This is a convenient volatility indicator because the standard deviation of the log of a cultural measure across waves gives, approximately, the average percentage deviation of that measure from the country mean. The distribution of such indicator is shown in Figure 2 for each of the nine cultural variables. Control and Tolerance exhibit modest volatility (less than 10%, on average), but the other variables are more volatile. The means of these volatility distributions are reported in the left panel of Table 2 for both the unconditional and the conditional measures of culture. For the latter, the volatility of Trust is 30%; for Respect, Obedience, Work, and Equality it is about 20%.

Third, we follow Falk et al. (2018) in decomposing the variance of the individual-level cultural measures to gauge the relative importance of between-country and between-wave variations in the WVS cultural proxies. The R^2 from an ordinary least-squares (OLS), individual-level regression of a cultural measure on country dummies in a given wave corresponds to the between-country variation in that measure for that wave; and the R^2 from a corresponding regression on wave dummies in a given country similarly corresponds to the between-wave variation in that measure for that country.

Adding individual covariates to these regressions yields the corresponding between-country and between-wave variations for the conditional cultural measures. The respective cross-wave and cross-country averages of these R^2

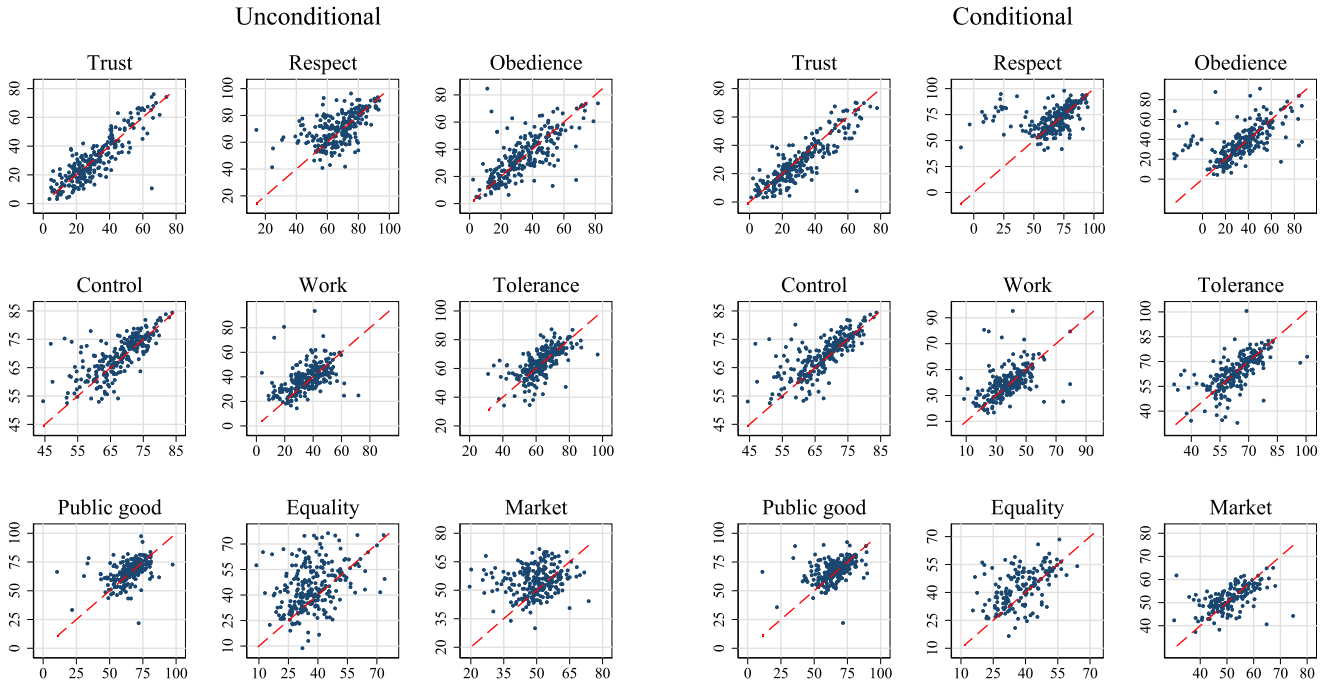

FIGURE 1 Cross-country cultural measures in WVS wave t versus wave $t-1$. A point in a scatter plot is a country in two adjacent WVS waves. The vertical axes measure the value of an unconditional (left, Equation (1)) or conditional (right, Equations (2) and (3)) cultural measure in wave t; the horizontal axes measure the corresponding value in wave $t-1$. The dashed line is the 45° line. Sample: 85 countries present in at least two consecutive waves in WVS waves (1–6), out of 110 countries present in at least one WVS wave.

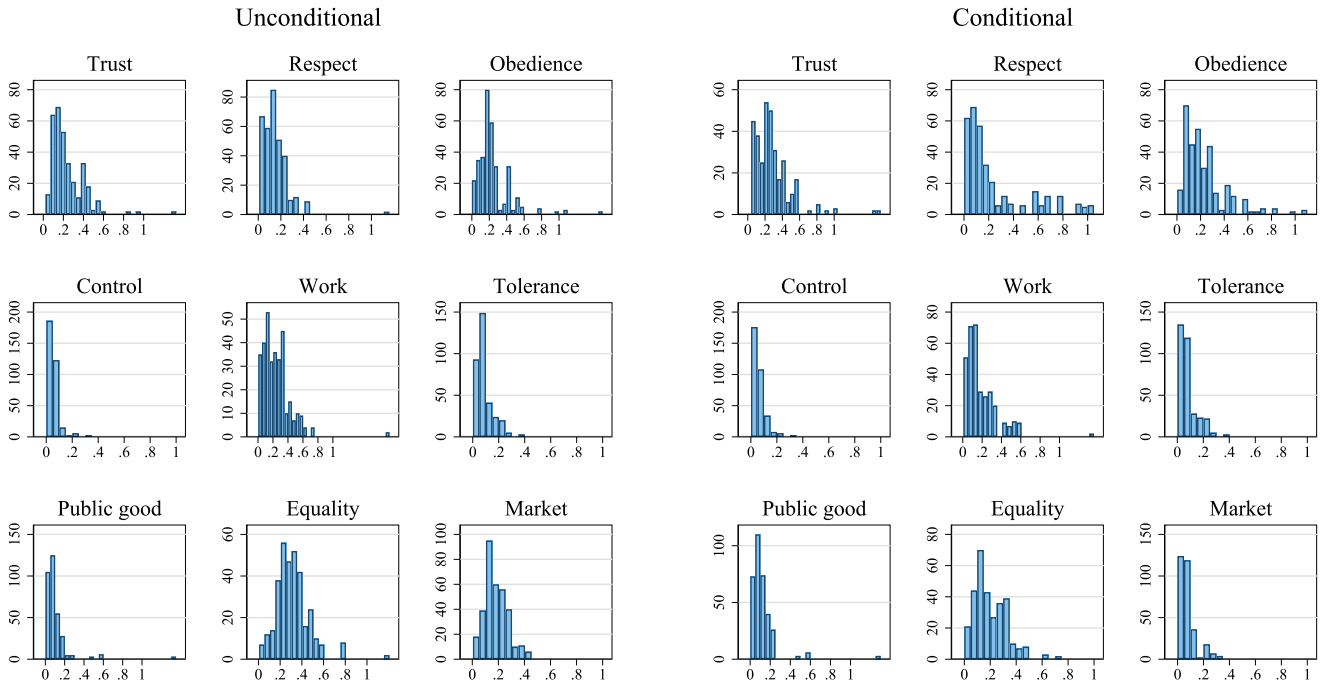

FIGURE 2 Distribution of within-country volatility of conditional cultural measures. The figure plots the distribution (absolute frequencies) of the within-country standard deviation of the log of each unconditional (left, Equation (1)) or conditional (right, Equations (2) and (3)) cultural cultural measure. Sample: 85 countries present in at least two consecutive WVS waves (1–6), out of 110 countries present in at least one WVS wave.

values are reported in Table 3. In line with the findings of Falk et al. (2018) for the preference traits measured by the Global Preference Survey, the between-country variation in the WVS conditional cultural traits ranges between 5.8% and 17.8%. The associated between-wave variation is smaller (ranging from 3.5% to 10.4%) but it amounts to between

TABLE 2 Within-country volatility and trends in WVS cultural measures.

	SD of log culture		Linear trend in log culture	
	Unconditional	Conditional	Unconditional	Conditional
Trust	0.261	0.313	−0.030** (0.010)	−0.084** (0.011)
Respect	0.141	0.212	0.046** (0.006)	0.102** (0.014)
Obedience	0.242	0.235	0.005 (0.012)	0.014 (0.013)
Control	0.056	0.059	0.020** (0.003)	0.024** (0.003)
Work	0.230	0.188	0.089** (0.011)	0.045** (0.010)
Tolerance	0.091	0.086	0.018** (0.004)	0.008 (0.004)
Public good	0.103	0.114	0.023** (0.006)	0.035** (0.007)
Equality	0.310	0.206	0.128** (0.011)	0.094** (0.012)
Market	0.170	0.078	−0.073** (0.007)	−0.003 (0.006)

Note: The left panel shows, for each cultural measure, the average of the standard deviation (SD) of the log of that measure (i.e., the average percentage deviation of that measure from the country mean, approximately). Both the unconditional measure (Equation 1) and the conditional measure (Equation 3) are employed. The table's right panel reports OLS estimates of β from Equation (5), with the associated robust standard errors in parentheses. Significance level: *5%; **1%. Sample: 85 countries present in at least two consecutive WVS waves (1–6), out of 110 countries present in at least one WVS wave.

34% and 103% of the between-country variation. This statistic confirms the importance of longitudinal changes in the WVS cultural proxies.

In order to better understand this variance decomposition, we report in Figure 3 the between-country variation by wave and in Figure 4 the distribution of the between-wave variation by country. The horizontal scale in these two figures is the same so as to facilitate a comparison. These figures show that while the between-country variation is larger than the between-wave variation *on average*, the former never exceeds 20% across waves while the latter may be quite large in some countries and for some cultural measures. As shown in what follows, it is the cultural traits associated with larger between-wave variation that tend to exhibit more fragility in their association with output at different points in time.

Finally, we check for the presence of within-country trends in these survey measures by means of the following regression:

$$\ln c_{gt} = \alpha_g + \beta t + \varepsilon_{gt}; \tag{5}$$

here g is a country index, α_g is a country fixed effect, and β is the common slope of a within-country linear time trend in the log of Culture. Our OLS-FE estimates of β are reported in the right panel of Table 2. Most of the cultural measures exhibit significant and often sizable trends. Net of covariates, WVS respondents' support for income equality has *increased* (on average) by 9.4% in each survey wave from a base of 41% (SD = 10.4) in the 1980s. At the same time, respondents' trust has *decreased* by an average of 8.4% from a base of 30.1% (SD = 16.6). Respect, Control, Work, and Public good have all increased, across waves, at rates that average between 2.4% and 10.2%. Equation (5) imposes a common slope across countries—so as to provide a summary measure—but the underlying trends are actually heterogeneous, which makes the consequences of the volatility of the survey measures potentially more worrisome. To appreciate this heterogeneity, we estimate Equation (5) country by country in the subset of 66 countries that are surveyed at least three times and we report in Figure 5 the distribution of the estimated slopes. It follows that country rankings in each cultural measure change over time. This fact is documented in Figure 6, which plots country rankings in each cultural measure in wave t versus wave $t-1$.

Overall, this evidence casts some doubt on the ability of WVS culture measures to serve as reliable proxies for slow-moving, persistent cultural traits. The ordinal scale problem illustrated next compounds this issue.

2.2 | The ordinal scale problem

When one uses *ordinal* individual measures from survey data to construct *cardinal* aggregate measures of culture, variations in the latter (either across countries or across waves for a country) may bear no connection with changes in the unobserved aggregate cultural traits of interest. To assess the relevance of this phenomenon, we perform for cultural attitudes measured on non-binary ordinal scales three tests proposed by Bond and Lang (2019).[6] Among the 16 raw

TABLE 3 Between-country and between-wave variability in WVS cultural measures.

	Between-country variation (Average across waves)		Between-wave variation (Average across countries)	
	Unconditional	**Conditional**	**Unconditional**	**Conditional**
Trust	0.098	0.104	0.017	0.035
Respect	0.054	0.058	0.033	0.039
Obedience	0.106	0.117	0.026	0.044
Control	0.088	0.107	0.023	0.055
Work	0.098	0.099	0.051	0.055
Tolerance	0.162	0.178	0.047	0.065
Public good	0.130	0.139	0.048	0.061
Equality	0.112	0.129	0.133	0.133
Market	0.110	0.128	0.113	0.118

Note: The table reports, for each cultural measure, the average across waves of between-country variations (left panel) and the average across countries of between-wave variation (right panel). The between-country variations are given by the R^2 from an OLS regression of an individual's cultural measure on country dummies in a given wave. When individual covariates are added, the variations are labeled "conditional". Similarly, the between-wave variations are given by the R^2 from an OLS regression of an individual's cultural measure on wave dummies in a given country. Sample: 478,939 individuals in 85 countries present in at least two consecutive WVS waves (1–6), out of 110 countries present in at least one WVS wave.

variables listed in the second column of Table 1, seven are of this kind: Control, Cheating on taxes, Avoiding a fare on public transport, Incomes should be made more equal, Competition is good, Private ownership should be increased, and Government should take more responsibility. These variables are all measured on a 1–10 scale.

The null hypothesis in the first test is that the ranking, in terms of average culture c_{gt}, between countries (or between waves for a given country) is identified without any distributional assumption, that is, without any cardinalization of the survey ordinal answers. The conditions for this nonparametric identification are quite stringent when more than two countries in a wave (or more than two waves for a given country) are compared, essentially requiring that the extreme categories (1 and 10, in this case) are not used by respondents. Formally, with reference to reporting function (4), denoting by r_{sgt} the fraction of individuals who select category $s = 0, \cdot, S$ in country g and wave t, the between-country (bc) null hypothesis is:

$$H_0^{bc} : r_{0At} = 0,\ r_{SBt} = 0, \quad \text{and} \quad r_{SBt} \geq r_{sBt} = 0 \quad \text{for} \quad s = 1, ..., S-1$$

for any two countries A and B, while the between-wave (bw) null hypothesis is:

$$H_0^{bw} : r_{0gt} = 0,\ r_{Sg\tau} = 0, \quad \text{and} \quad r_{Sgt} \geq r_{sgt} = 0 \quad \text{for} \quad s = 1, ..., S-1$$

for any two waves t and τ. Not surprisingly, as summarized in Table 4, the null is always rejected both between countries (i.e., pooling all waves) and between waves (i.e., country by country).

Given this lack of nonparametric identification, we follow Bond and Lang and we assume that η_{igt} in Equation (4) is normally distributed, so that an ordered Probit model provides the needed cardinalization. Rank identification under this normality assumption requires that the variance of η_{igt} in reporting function (4) is equal across countries (between-country test) or across waves in a given country (between-wave test). These are the null hypotheses in the second test:

$$H_0^{bc} : \mathrm{Var}(\eta_{iAt}) = \mathrm{Var}(\eta_{iBt})$$

for any two countries A and B, and

$$H_0^{bw} : \mathrm{Var}\left(\eta_{igt}\right) = \mathrm{Var}\left(\eta_{ig\tau}\right)$$

for any two waves t and τ.

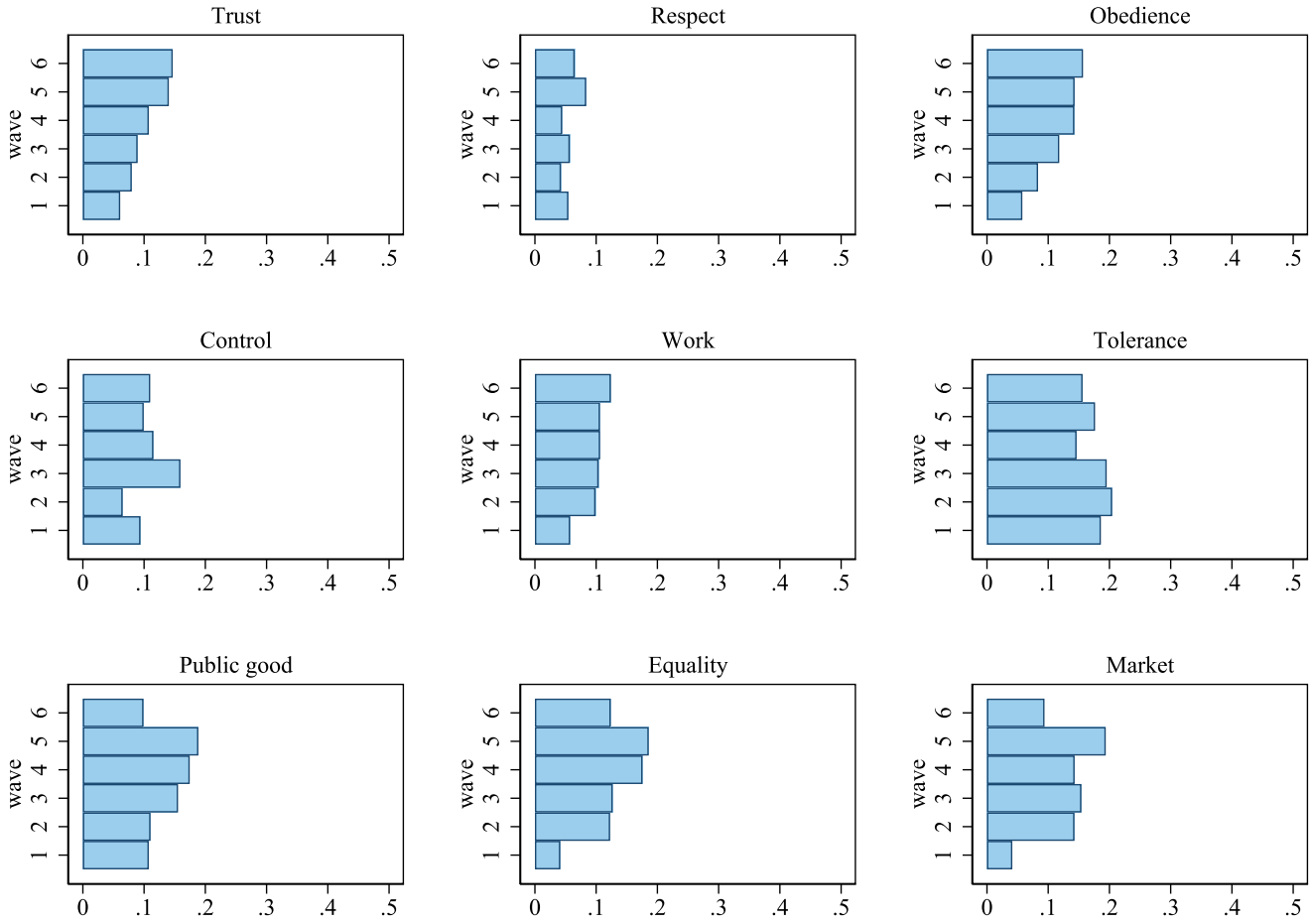

FIGURE 3 Between-country variation of conditional cultural measures, by WVS wave. The figure reports, for each cultural measure and by WVS wave, the between-country variation. This is given by the R^2 from an OLS regression of an individual's cultural measure on country dummies in a given wave. All of the cultural measures represented in the figure are conditional (see Equations (2) and (3)). Sample: 478,939 individuals in 85 countries present in at least two consecutive WVS waves (1–6), out of 110 countries present in at least one WVS wave.

As reported in Table 4, when pooling all countries and waves, the null is always rejected; across waves it is instead rejected for between 39% and 58% of countries. Even when the equal variances null is not rejected, rank order identification under the normality assumption requires a common reporting function in each country, that is, that cutoff points c_i^s in (4) are invariant across individuals and waves for a given country. This is the null hypothesis of the third test. Formally:

$$H_0 : c_{igt}^s = c_{jg\tau}^s$$

for any two individuals i and j and any two waves t and τ. As explained by Bond and Lang (2019) in their Online Appendix, under the normality assumption, this hypothesis can be tested using a likelihood ratio test that contrasts estimates from an unconstrained and a constrained ordered Probit model for each country and each cultural trait, pooling all waves. In the former, the first two cutoff points c_{igt}^1 and c_{igt}^2 are normalized to 0 and 1, respectively, while the mean, variance, and cutoff points $c_{igt}^3, \ldots, c_{igt}^S$ are estimated from the data. In the constrained model, instead, the top cutoff point c_{igt}^S is constrained to be the same across waves. Considering all countries, the null is rejected between 12% and 37% of the times. Conditional on not rejecting the hypothesis that variances are equal across waves, the hypothesis that the reporting functions are also equal is rejected between 4% and 24% of the times. Thus, even under the normality assumption, researchers using ordinal scales to measure culture may confidently assume only for some countries that there is a connection between changes in survey cultural proxies and changes in the cultural traits of interest.

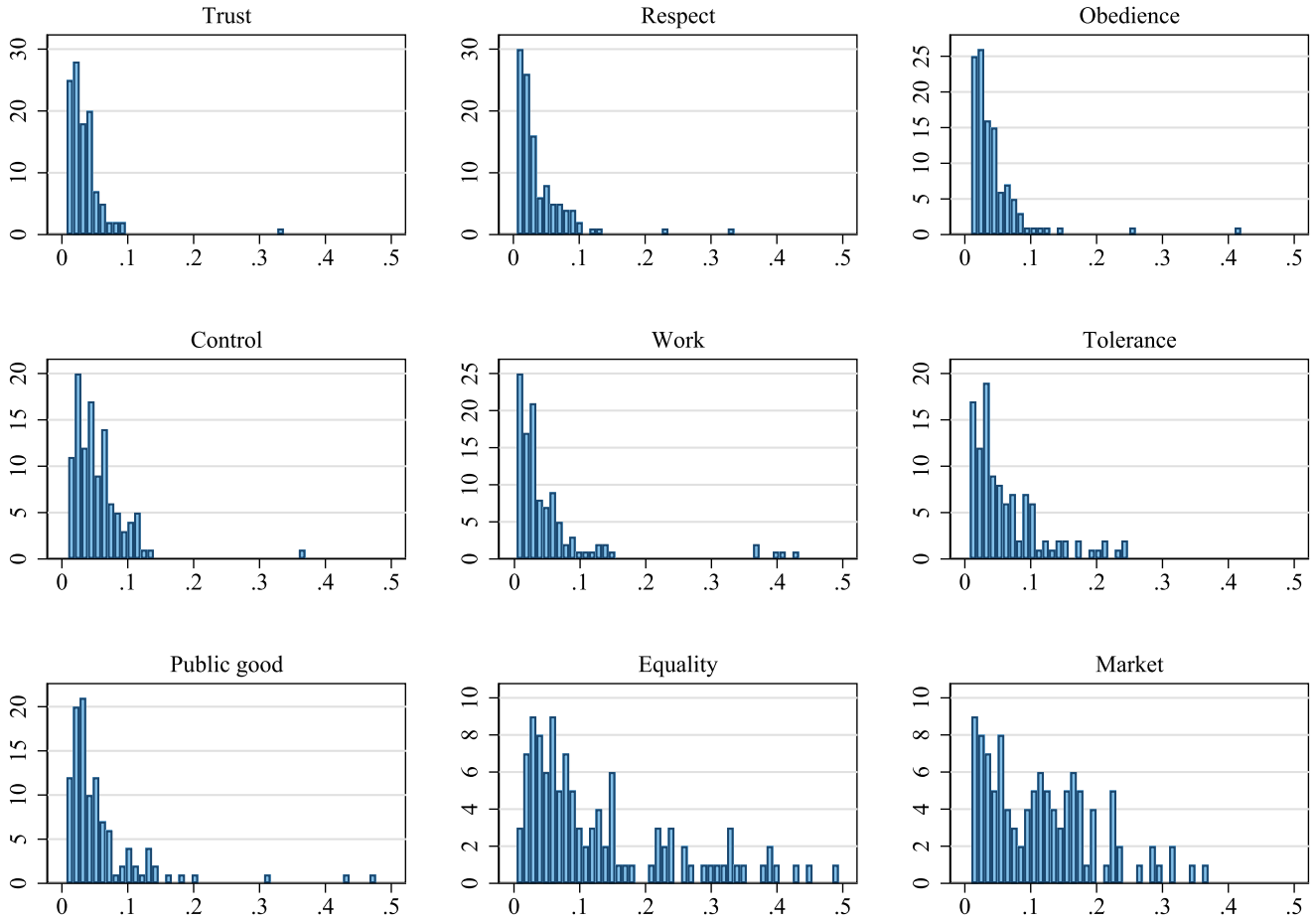

FIGURE 4 Distribution of between-wave variation of conditional cultural measures. The figure reports, for each cultural measure and by WVS wave, the distribution (absolute frequencies) of the between-wave variation across countries. For a given country, this is given by the R^2 from an OLS regression of an individual's cultural measure on wave dummies in that country. All of the cultural measures represented in the figure are conditional (see Equations (2) and (3)). Sample: 478,939 individuals in 85 countries present in at least two consecutive WVS waves (1–6), out of 110 countries present in at least one WVS wave.

We next demonstrate the possible consequences of employing non-persistent cultural measures to investigate the cultural roots of economic development by replicating the analyses of Gorodnichenko and Roland (2011) and Tabellini (2010) with more recent WVS data.

3 | CULTURE AND ECONOMIC DEVELOPMENT AROUND THE WORLD

Gorodnichenko and Roland (2011) investigate empirically which cultural traits affect cross-country per capita output. The underlying theory is that individualist cultural attitudes are conducive to higher output growth—via rewards to innovation—while collectivist cultural attitudes induce a level effect—via better coordination of economic activity—but not a growth effect because in the long term coordination has diminishing returns (Gorodnichenko & Roland, 2017). This view of culture as a persistent process that drives economic development even in the presence of poor institutions is illustrated in the following causal directed acyclic graph (Pearl, 2009):

$$\text{Past culture} \rightarrow \text{Current culture} \rightarrow \text{Output}.$$

Using different surveys, the authors construct measures that are interpreted in terms of individualist or collectivist attitudes and relate them to output in a sample of 74 countries. Although their conclusion is that only Hofstede's (2001) individualism score is a robust predictor of output differences, they report a significant correlation between, on the one hand, the WVS measures of Trust, Tolerance, and Equality and, on the other hand, real GDP per worker in 2000.

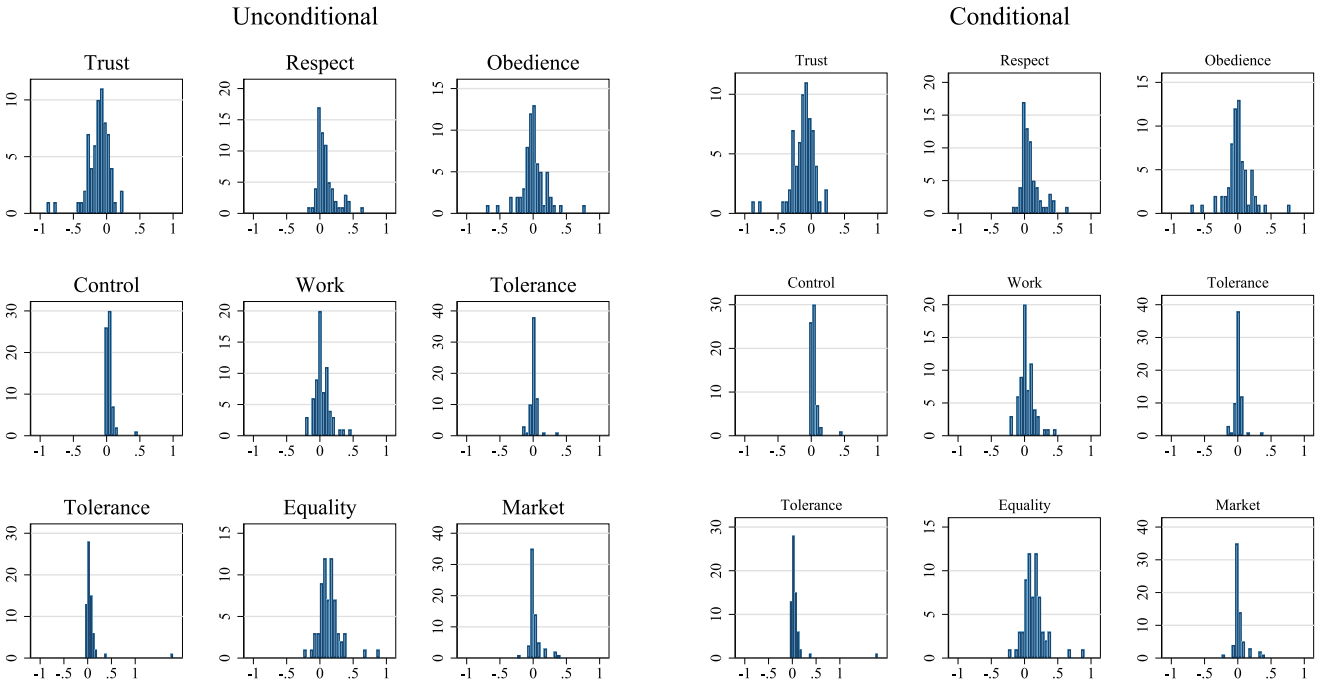

FIGURE 5 Distribution of linear trend slopes in conditional cultural measures. The figure reports, for each unconditional (left, Equation 1) or conditional (right, Equations (2) and (3)) cultural measure, the distribution (absolute frequencies) of parameter country-by-country OLS estimates of β from Equation (5). Sample: 66 countries surveyed at least three times in WVS waves (1–6), out of 110 countries present in at least one WVS wave.

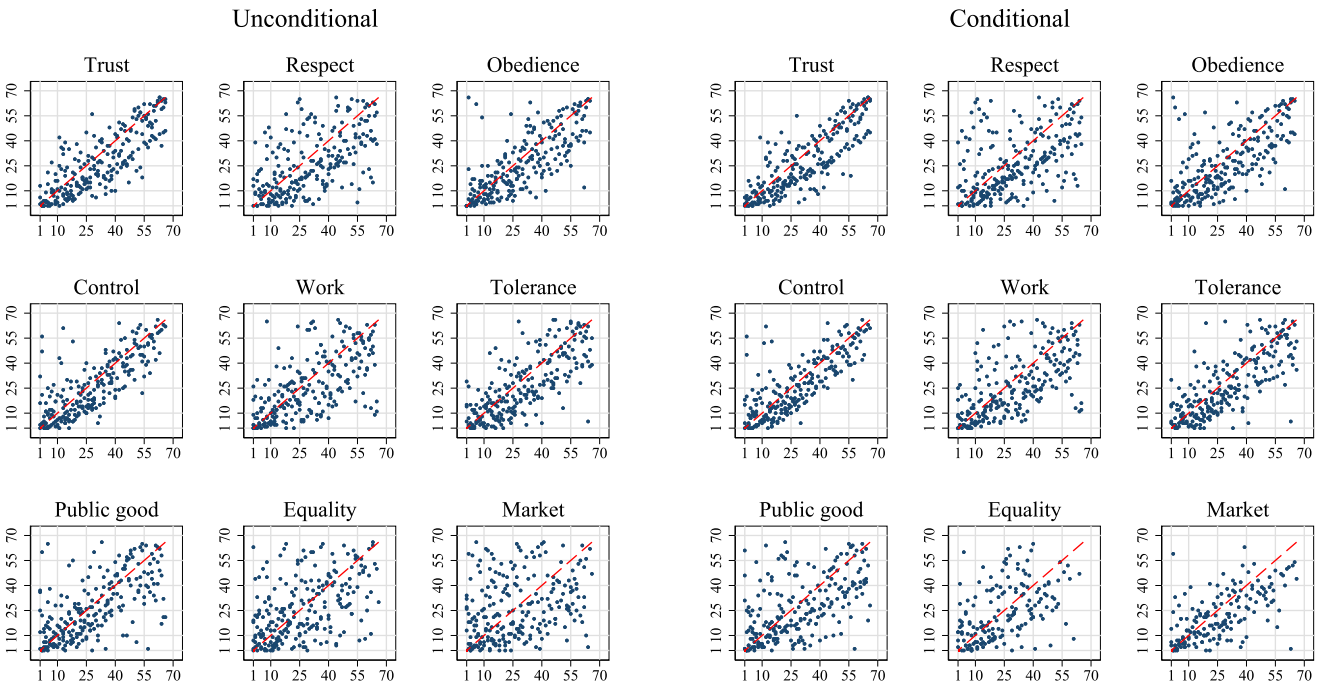

FIGURE 6 Country ranking in cultural measures in WVS wave t versus wave $t-1$. A point in a scatter plot is a country. The vertical axes measure the value of an unconditional (left, Equation 1) or conditional (right, Equations (2) and (3)) cultural proxy in waves t; the horizontal axes measure the corresponding value in waves $t-1$ (2005–2014). The dashed line is the 45° line. Sample: 85 countries present in at least two consecutive waves in WVS waves (1–6), out of 110 countries present in at least one WVS wave.

TABLE 4 Culture rank order identification tests.

	Nonparametric identific. of culture rank		Equal variances		Equal reporting: fraction reject	
	Btw-country	Btw-wave	Btw-country	Btw-wave fraction reject	Btw-wave all	Btw-wave equal var.
Control	Reject	Reject	Reject	0.58	0.18	0.11
Tax evasion	Reject	Reject	Reject	0.50	0.31	0.17
Avoiding fare	Reject	Reject	Reject	0.50	0.30	0.21
Income equal.	Reject	Reject	Reject	0.39	0.17	0.12
Competition	Reject	Reject	Reject	0.46	0.37	0.24
Private	Reject	Reject	Reject	0.42	0.15	0.12
Government	Reject	Reject	Reject	0.40	0.12	0.04

Note: The table reports, for seven WVS cultural proxies used by either Tabellini (Control) or Gorodnichenko and Roland (the remaining six variables) and measured on a 1–10 ordinal scale, the results from three tests proposed by Bond and Lang (2019). H_0 in the first test is that the culture rank order of countries either pooling all waves (between-country) or across waves for a given country (between-wave) is nonparametrically identified. The test consists of checking whether the extreme categories are used (in which case the null 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 countries 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 countries minus one. H_0 in the third test is that the reporting function is equal across countries 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 a country is surveyed minus one. See Bond and Lang's appendix for more details. Sample: 110 countries present in at least one WVS waves. In the between-wave tests, only 84 countries that appear in at least two WVS waves (and for which all the seven variables listed in the table are available) are used.

Replication files are not available, so we limit our analysis to revisiting the OLS results that employ these and other WVS measures, to show how the correlations between cultural proxies and output can be different at different points in time.

3.1 | Replication data

Data on cultural attitudes are from the EVS and the WVS (see footnote 1) and the unit of observation is a country. At the time Gorodnichenko and Roland (2011) was written, only WVS waves 1–4 (1981–2004) were available. We label these as the "old" sample, a total of 275,449 individuals in 84 countries. Notice that since Gorodnichenko and Roland (2011) was published, the WVS data (all waves) have been officially revised to clean data errors and impute missing values, which may generate discrepancies when replicating the original results. Our replication exercise adds WVS wave 5 (2005–2009), and wave 6 (2010–2014), which we refer to as the "new" sample. This comprises 235,308 individuals in 101 countries.[7] The cultural proxies used as explanatory variables are the unconditional measures of Trust, Work, Tolerance, Public good provision, Equality, and Market orientation (see Section 2).

Data on output are from version 6.3 of the Penn World Tables, which is the version used in the original article. The outcome of interest associated with the old WVS sample is real GDP per worker at PPP in 2000. We also consider this same GDP measure in 2007, which yields—consistently with the forward measure of culture in the 2005–2014 WVS waves—the corresponding output level measure to be associated with the new sample. We choose year 2007 as the terminal period to avoid contamination with the large deviations from national trends experienced by many countries from 2008 onward. After merging cultural attitudes and output data, and after keeping only countries that are present in both the old and the new sample, we obtain a cross-section of 72 countries, which is slightly smaller than Gorodnichenko and Roland's sample of 74 countries.

3.2 | Replication using the new WVS waves

Column [1] of Table 5 contains the original estimates, reproduced from column (2) of Table 1 in Gorodnichenko and Roland (2011), which derive from this estimating equation:

TABLE 5 Gorodnichenko and Roland (2011) original and replication estimates.

	[1] Original waves 1–4	[2] Replication waves 1–4	[3] Replication waves 5–6	[4] Replication waves 5–6
Trust	0.41	0.25	0.24	0.28
	(n.a.)	(0.09)**	(0.08)**	(0.08)**
	[0.10]**	[0.08]**	[0.09]**	[0.08]**
Work	−0.06	−0.17	−0.03	−0.00
	(n.a.)	(0.08)*	(0.08)	(0.07)
	[0.12]	[0.08]*	[0.09]	[0.08]
Tolerance	0.50	0.22	0.27	0.18
	(n.a.)	(0.09)**	(0.09)**	(0.08)**
	[0.12]**	[0.10]**	[0.09]**	[0.09]**
Public good	−0.07	−0.09	0.09	0.05
	(n.a.)	(0.09)	(0.09)	(0.08)
	[0.12]	[0.08]	[0.08]	[0.07]
Equality	−0.38	−0.04	0.18	0.21
	(n.a.)	(0.09)	(0.09)$^{+}$	(0.09)*
	[0.11]**	[0.08]	[0.09]$^{+}$	[0.09]*
Market	−0.21	0.22	0.12	0.11
	(n.a.)	(0.10)*	(0.09)	(0.08)
	[0.13]	[0.08]*	[0.09]	[0.08]
Countries	74	72	72	72
Dep. var	GDP 2000	GDP 2000	GDP 2000	GDP 2007

Note: Results from OLS regressions of output on unconditional cultural measures. Column [1] is reported from Gorodnichenko and Roland's (2011) Table 1, panel C, column (2). The six cultural measures are included simultaneously (after normalizing them to have a zero mean and a unit standard deviation in the estimation sample) and are derived from the 1981–2004 WVS waves in columns [1] and [2] ("old" sample), and from the 2005–2014 waves in columns [3] and [4] ("new" sample). The dependent variable is the log of real GDP per worker at PPP in 2000 in columns [1], [2], and [3]; column [4] presents the analogous measure for year 2007. Conventional standard errors in parentheses; robust in brackets. Significance levels: *5%; **1%. In column [1], (n.a.) means that conventional standard errors are "not available" in the original article.

$$y_g = \delta_0 + \sum_{m=1}^{6} \delta_{1m} c_{gm} + \epsilon_{gj}, \tag{6}$$

where y_g is the log of real GDP per worker at PPP in 2000 in country g and c_{g1}, ..., c_{g6} are the six unconditional WVS cultural proxies considered by the authors (standardized).

Column [2] reports our attempt at replicating these results in the same old sample. We computed two different sets of standard errors: conventional (reported in parentheses) and robust with respect to heteroskedasticity [in brackets]. Apart from the coefficient on Market orientation—which we find to be statistically significant and positive, in line with the theoretical prediction—and Work—which we estimate to be slightly larger in absolute value and significant—the coefficients on the cultural measures have the same sign and significance as the original ones. The important comparison is with columns [3] and [4], that is, the coefficients estimated in the new sample. The dependent variable is, alternately, GDP per worker in 2000 and 2007. These columns show that when using the more recent WVS waves in the same sample of countries, of the six cultural measures only two (Trust and Tolerance) are robustly associated with output. For Public good, and Equality, the correlations actually change in sign. The coefficient on Market drops to zero, while Equality becomes statistically significant with a sign that contradicts the theoretical prediction. Note that the three fragile measures all involve ordinal scales, while the two robust measures originate from binary questions (see Table 1). Thus, in a cross-country setting the volatility of the survey measures of culture leads to fragile correlations. We next show that a similar fragility may arise in a more refined region-level research design.

4 | CULTURE AND ECONOMIC DEVELOPMENT IN EUROPE

Tabellini (2010) investigates empirically the causal effect of culture on economic development across the regions of Europe. The underlying theory is quite different from Gorodnichenko and Roland's in that it emphasizes culture as the "missing link" between historical institutions and current output. Specifically, history (past institutions) shapes cultural attitudes, which are persistent traits transmitted from parents or society to children and which therefore affect current culture and eventually current output via the improved functioning of current informal and formal institutions. This view is illustrated as follows:

$$\begin{matrix}\text{Past}\\ \text{institutions}\end{matrix} \rightarrow \begin{matrix}\text{Past}\\ \text{culture}\end{matrix} \rightarrow \begin{matrix}\text{Current}\\ \text{culture}\end{matrix} \rightarrow \begin{matrix}\text{Functioning of}\\ \text{current institutions}\end{matrix} \rightarrow \text{Output},$$

a causal mechanism that reconciles approaches that stress either the primacy of institutions or the primacy of culture in determining economic development.[8]

Tabellini finds a significant relation between, on the one hand, Trust, Respect, Obedience, and Control in the 1990–1998 waves of the WVS and, on the other hand (a) average GVA during 1995–2000 across the regions of Europe or (b) the corresponding average growth rate between 1977 and 2000. Both OLS and instrumental variables (IV) estimates deliver associations consistent with the theory illustrated above, an interpretation that is supported by a within-country empirical design (the regions of Europe share identical formal institutions within a country) and also by two historical instruments, described in what follows, that are presumed to predict current values and beliefs and thus to filter the underlying, persistent component of culture. Replication files were made available at the author's Web page, which enable us to carry out a more thorough replication that also focuses on causality.

4.1 | Replication data

Like for Gorodnichenko and Roland (2011), we replicate the empirical exercise in question but with more recent data for the same regions. In this case the data for our replication come from three different sources. First, the WVS (including the EVS) for the cultural measures. Second, Cambridge Econometrics for the output measures. Third, the original data files provided by Tabellini (2010) for the instrumental variables.

Tabellini (2010) uses surveys of individuals in 69 regions across eight European countries (Belgium, France, Germany, Italy, Northern Ireland, the Netherlands, Portugal, and the United Kingdom) from WVS wave 2 (1990–1994) and wave 3 (1995–1998).[9] For brevity, we henceforth refer to these eight countries as "Europe". Regions are defined by the NUTS (Nomenclature of Territorial Units for Statistics) 2 level of classification yet with some further aggregation for smaller units. Our replication adds WVS wave 4 (1999–2004), wave 5 (2005–2009), and wave 6 (2010–2014)—none of which were available when Tabellini began his research. We have unified the regional codes (which may vary across waves) to ensure that we are consistently considering the same geographic entities throughout the 1990–2014 period. After discarding a few observations with missing region identifiers, our combined 1999–2014 sample ($N \approx 35{,}000$) is some 40% larger than the original 1990–1998 sample ($N \approx 25{,}000$). We also remark that, within individual regions, the size of the 1999–2014 sample is larger (on average across waves) than is the 1990–1998 sample; Spain is the lone exception. To simplify the discussion, hereafter we reference the 1990–1998 and 1999–2014 samples as (respectively) the "old" and "new" samples. Because of the official revisions in WVS data, the Spanish sample currently available in the WVS does not contain observations from the regions of Andalucia, the Basque Country, Galicia, or Valencia, which were collected in wave 3. In order to preserve a full correspondence with the old sample, we requested and received from the personnel in charge of WVS data maintenance the wave three observations for these four regions of Spain. Note also that the old sample comprises 69 regions in Tabellini (2010) but only 68 in our replication; this difference reflects that, in the revised WVS data, there is no information on cultural attitudes in the region of Alentejo (Portugal) in waves 2 and 3. Similarly, the new sample includes only 67 regions because of missing information on cultural attitudes in the regions of the Azore Islands and Madeira (Portugal) in waves 4, 5, and 6. Neither the inclusion nor the exclusion of these two peripheral and relatively small regions should have any material effects on our findings.

As mentioned in Section 2, the four cultural measures selected by Tabellini (2010) (Trust, Respect, Obedience, and Control) are motivated by their relation to the backwardness in southern Italy documented by Banfield (1958) and Putnam (1993). As the author acknowledges, there is "some unavoidable arbitrariness in this selection" (2010, p. 686). For example, Tabellini regards Obedience as a cultural trait that is *not* conducive to development because individuals in

a submissive culture are less inclined to pursue opportunities. However, Obedience is a noncognitive skill that may favor growth through better-functioning hierarchical structures such as firms or the public administration.[10] To reduce such arbitrariness, Tabellini employs the first principal components (PCs) of three different combinations of the four measures. In what follows, these combinations are labeled PC culture (first principal component of all four measures), PC culture positive (first principal component of Trust, Respect, and Control), and PC children (first principal component of Respect and Obedience). The region-level averages of the resulting seven, individual-level cultural measures (four raw measures and three PC measures) over the 1990–1998 period constitute the unconditional regional measures of culture in Tabellini (2010), computed as in Equation (1), where now the geographic index g refers to the regions of Europe. Tabellini also constructs conditional measures as described by Equations (2) and (3), while using survey information on age, gender, marital status, education, social class, and health status.[11] Moreover, the standard error of the estimated regional dummy coefficient, $\widehat{\sigma}\left(\beta_{2,g}\right)$, serves as a weight to be employed in regressions involving conditional measures. Here the wave index t is omitted because Tabellini aggregates waves to form a cross section. We follow the original article in all of these choices, and we construct unconditional and conditional cultural measures at the region level from the raw data for both the old and the new sample.

Data on output are from Cambridge Econometrics, a private company that compiles a variety of sectoral output measures for European countries at the NUTS 2 level of disaggregation; this is the same data source used by Tabellini (2010). The database made available to us spans the period 1980–2015, and it is an up-to-date version that reflects official revisions of the underlying output statistics by the national sources and Eurostat. From these data we construct average, region-level per capita GVA (at purchasing power parity, PPP) as a percentage of the 15-country European Union (EU15) in 1995–2000. This measure of output level is the one used by Tabellini (2010) and is associated with the old (1990–1998) WVS sample. We also construct this same GVA measure for the period 2001–2007 to be associated with the new sample. Both OLS and IV analyses are performed using output level as an outcome. A second output measure employed in Tabellini's IV analysis is the growth rate of per capita GVA between 1977 and 2000, which is computed as the average difference between the logs of per capita output in two consecutive years. Given the slightly different time span of the Cambridge Econometrics data base available to date, we can compute this growth rate only for the 1980–2000 period in the old sample. As a corresponding measure to be associated with the new sample, we construct the growth rate of per capita GVA between 1980 and 2007.[12] We shall use y_{gj} to denote an output measure (level or growth) for region g in country j; this measure depends on j via current national institutions.

As discussed in Section 2, the WVS cultural proxies may reflect transitory shocks or other determinants of income that are actually unrelated to the persistent cultural traits of interest. So, in order to isolate a causal effect from culture to output, Tabellini (2010) constructs two region-level historical variables to serve as excluded instruments for the cultural measures, which we denote by $\mathbf{z}_g$: (i) the Literacy rate in 1880, which captures the effect of educational history; and (ii) the first principal component of five variables measuring constraints on the executive in 1600, 1700, 1750, 1800, and 1850, which captures the impact of political history (PC Institutions). The rationale for adopting this class of instruments is as follows. Because culture is transmitted slowly from one generation to the next, variables that trace historical paths are candidate instruments for culture because they contributed to shaping norms and beliefs in the past that persist into the present.[13] Furthermore, to shut off the most obvious channels through which the excluded instruments $\mathbf{z}_g$ could affect economic development for reasons *unrelated* to culture, Tabellini also constructs two historical control variables (included instruments) that we denote by $\mathbf{q}_g$: first, the enrollment rate in primary school during 1960, which captures baseline human capital; and second, the urbanization rate in 1850, which accounts for convergence. If we use θ_j to denote country fixed effects, then the reduced-form coefficients γ_1 from the linear regression

$$y_{gj} = \gamma_0 + \gamma_1 \mathbf{z}_g + \gamma_2 \mathbf{q}_g + \theta_j + u_{gj}, \tag{7}$$

express the influence of past regional economic and political institutions on current region prosperity (as posited, e.g., by Acemoglu & Robinson, 2012). The first-stage prediction of the relevant cultural trait—namely $\widehat{c}_{gj} \equiv \mathbb{E}\left[c_g \mid \mathbf{z}_g, \mathbf{q}_g, \theta_j\right]$—is instead supposed to filter out the transitory component of the cultural proxies (including measurement error) by projecting them on the historical instruments and thus to retain only the slow-moving, persistent component of culture.

4.2 | Replication using the old WVS waves

The replication proceeds in three steps. As the first, preliminary step, we validate our variables by checking the correspondence between the cultural and output measures in Tabellini (2010) and those computed by us for the same

period (i.e., 1990–1998 for culture and 1995–2000 for output). Details are available upon requests. There are a few discrepancies that are due to retrospective data revisions in both the WVS and Cambridge Econometrics, but they are not a source of concern given that we still can—despite their presence—replicate almost perfectly the original results in Tabellini (2010). This result is achieved in the next step of our exercise, which consists of replicating the original OLS and IV results, which derive from this estimating equation:

$$y_{gj} = \delta_0 + \delta_1 c_{gj} + \delta_2 \mathbf{q}_g + \theta_j + \upsilon_{gj}; \tag{8}$$

here c_{gj} is a cultural measure, which can be either unconditional or conditional. Following Tabellini, we employ both of these measures in the OLS regressions, and use the conditional measure only in the IV regressions. All regressions that use the conditional measures of culture are weighted by $\widehat{\sigma}\left(\beta_{2,g}\right)$ (as defined in 4.1). The original results are reproduced in columns [1] and [4] of Table 6 and in the corresponding columns of Table 7. Table 6 reports OLS estimates that use output level as an outcome. Table 7 presents IV estimates that use either output level or growth as an outcome. Our replication of the original results is reported in columns [2] and [5] of these same tables. For each cultural measure, we report the estimated coefficient δ_1 from distinct regressions based on Equation (8). We again compute two different sets of standard errors: conventional (reported in parentheses) and robust with respect to heteroskedasticity [in brackets].[14] A comparison of column [2] versus [1] and of column [5] versus [4] across Tables 6 and 7 establishes that—in terms of magnitudes and statistical significance—the original pattern reported by Tabellini (2010) is robust to data revisions in both the WVS and Cambridge Econometrics databases. The pattern is evidently robust also to including, possibly, slightly different conditioning variables in $\mathbf{x}_{ig}$ and to the type of standard errors calculated. There are a few discrepancies between the original coefficients and those we derive using the old WVS waves, but they are small and—most likely—not statistically significant given the overlap between the respective 95% confidence intervals.

4.3 | Replication using the new WVS waves

After ensuring that we can replicate the original results, our third and final step is replication using the 1999–2014 waves of the WVS to construct the cultural proxies. Our results are reported in columns [3] and [6] of both Tables 6 and 7. A similar pattern emerges from the OLS estimates and from the IV estimates: of the four raw cultural measures employed by Tabellini (2010), only Trust appears to be significantly associated with output in the more recent WVS waves. As a consequence, the coefficients on the two PC measures of culture that incorporate Trust are likewise significant. Yet for Respect, Obedience, and Control as well as their PC syntheses, the estimates in the more recent waves differ from the older waves in either magnitude or statistical significance.

In the OLS regressions of output on conditional culture, the estimated coefficients for Respect and Control drop by two thirds relative to the replication in the old sample. In the IV regressions, the estimated coefficients for these two cultural measures when GVA level is the outcome increase by a factor of between 2 and 4, yet the standard errors also increase, by as much as an order of magnitude. When GVA growth is the outcome, even the coefficient for Trust decreases by two thirds and becomes imprecisely estimated.

These findings suggest that, when culture is measured in the new WVS waves, the instrumental variables predict culture less well than in the old WVS waves; in other words, the historical instruments fail to distill the slow-moving, persistent component of culture. Note that this is not a problem of the instrumental variables used—which do not change over time—or of the use of regional variation from non-representative WVS samples at the region level—the correlations also change in the cross-country setting of Gorodnichenko and Roland (2011). It is instead a direct consequence of the instability of the cultural measures. Tables 8 and 9 support this conjecture.

Table 8 reports the reduced-form estimates of the effect of the two excluded instruments on output level and growth (i.e., the estimated coefficient vector γ_1 in Equation (7) and its standard error). These reduced-form estimates are virtually identical when using the old sample (columns [2] and [5] in the table), or the new sample (columns [3] and [6]). This result is what one expects: because *relative* aggregate output moves slowly in Europe, historical instruments correlate as well with average GVA in 1995–2000 as they do with average GVA in 2001–2007. The culprit is at the first stage. Table 9 reports the *F*-statistic on the excluded instruments from these first-stage regressions. Except when Trust is instrumented, the *F*-statistic drops precipitously, and well below the conventional threshold for weak instruments. With the new WVS waves, if we adopt conventional standard errors, then the instruments cannot precisely predict Respect ($F = 0.48$); also, the *F*-statistics associated with the remaining cultural proxies (including all of the PC measures)

TABLE 6 Tabellini (2010) original and replication OLS estimates.

	[1]	[2]	[3]	[4]	[5]	[6]
	Unconditional culture			**Conditional culture**		
	Original waves 2–3	**Replication waves 2–3**	**Replication waves 4–6**	**Original waves 2–3**	**Replication waves 2–3**	**Replication waves 4–6**
Trust	0.93	0.81	1.27	0.75	0.76	1.42
	(n.a.)	(0.44)	(0.42)**	(n.a.)	(0.48)	(0.50)**
	[0.38]*	[0.38]*	[0.46]**	[0.46]	[0.44]	[0.53]*
Respect	1.64	1.39	0.65	1.79	1.45	0.49
	(n.a.)	(0.42)**	(0.65)	(n.a.)	(0.44)**	(0.68)
	[0.51]**	[0.42]**	[0.47]	[0.47]**	[0.45]**	[0.47]
Obedience	−0.93	−0.90	−0.99	−0.68	−0.75	−0.80
	(n.a.)	(0.41)*	(0.50)	(n.a.)	(0.41)	(0.53)
	[0.46]*	[0.38]*	[0.43]*	[0.48]	[0.38]	[0.40]
Control	1.36	1.06	1.06	0.88	1.14	0.48
	(n.a.)	(0.90)	(1.15)	(n.a.)	(1.03)	(1.21)
	[0.83]	[0.71]	[0.84]	[0.82]	[0.80]	[0.93]
PC culture	0.58	0.55	0.63	0.60	0.52	0.58
	(n.a.)	(0.14)**	(0.19)**	(n.a.)	(0.15)**	(0.22)*
	[0.12]**	[0.11]**	[0.17]**	[0.13]**	[0.11]**	[0.18]**
PC positive	0.71	0.59	0.60	0.74	0.57	0.52
	(n.a.)	(0.16)**	(0.21)**	(n.a.)	(0.17)**	(0.24)*
	[0.15]**	[0.14]**	[0.20]**	[0.16]**	[0.14]**	[0.23]*
PC children	0.57	0.49	0.47	0.58	0.48	0.38
	(n.a.)	(0.15)**	(0.23)*	(n.a.)	(0.16)**	(0.25)
	[0.19]*	[0.15]**	[0.18]*	[0.18]**	[0.16]**	[0.18]*
Regions	69	68	67	68	67	66
Dep. var.	GVA 95-00	GVA 95-00	GVA 01–07	GVA 95-00	GVA 95-00	GVA 01–07

Note: Results from OLS regressions of output on cultural measures (unconditional in columns [1]–[3], conditional in columns [4]–[6]). Columns [1] and [3] are reported from Tabellini's (2010) Table 2 and Table 3, respectively. Each line corresponds to a distinct regression in which the corresponding cultural measure is the main regressor of interest: from the 1990–1998 WVS waves in columns [1], [2], [4], and [5] ("old" sample), and from the 1999–2014 waves in columns [4] and [6] ("new" sample). The dependent variable is average gross value added per capita (at PPP) as a percentage of the EU15 between 1995 and 2000 in columns [1], [2], [4], and [5]; columns [3] and [6] present the analogous measure for the period 2001–2007. School, urbanization, and country fixed effects are always included as conditioning variables. Standard errors: conventional in parentheses; robust in brackets; Significance levels: *5%; **1%. In columns [1] and [4], (n.a.) means that conventional standard errors are "not available" in the original article.

TABLE 7 Tabellini (2010) original and replication 2SLS estimates.

	[1]	[2]	[3]	[4]	[5]	[6]
	Conditional culture			**Conditional culture**		
	Original waves 2–3	**Replication waves 2–3**	**Replication waves 4–6**	**Original waves 2–3**	**Replication waves 2–3**	**Replication waves 4–6**
Trust	4.67	4.26	4.36	0.06	0.12	0.04
	(n.a.)	(1.65)**	(1.42)**	(n.a.)	(0.06)*	(0.03)
	[1.41]**	[1.41]**	[1.15]**	[0.03]*	[0.05]*	[0.02]*
Respect	2.86	2.90	13.75	0.03	0.09	0.13
	(n.a.)	(0.92)**	(13.20)	(n.a.)	(0.04)*	(0.17)
	[0.76]**	[0.91]**	[9.70]	[0.02]	[0.04]*	[0.14]
Obedience	−5.88	−5.41	−6.27	−0.08	−0.15	−0.06
	(n.a.)	(2.75)	(3.41)	(n.a.)	(0.09)	(0.04)
	[2.19]**	[2.64]*	[2.76]*	[0.04]*	[0.08]	[0.03]*
Control	13.17	11.32	21.58	0.18	0.32	0.20
	(n.a.)	(5.54)*	(13.68)	(n.a.)	(0.21)	(0.15)
	[7.61]	[5.37]	[12.92]	[0.12]	[0.19]	[0.14]
PC culture	1.11	0.91	1.89	0.02	0.03	0.02
	(n.a.)	(0.26)**	(0.61)**	(n.a.)	(0.01)*	(0.01)
	[0.28]**	[0.26]**	[0.55]**	[0.01]**	[0.01]*	[0.01]*
PC positive	1.16	0.94	2.12	0.02	0.03	0.02
	(n.a.)	(0.27)**	(0.70)**	(n.a.)	(0.01)*	(0.01)
	[0.32]**	[0.28]**	[0.66]**	[0.01]**	[0.01]*	[0.01]*
PC children	1.40	1.18	2.75	0.02	0.03	0.03
	(n.a.)	(0.38)**	(1.43)	(n.a.)	(0.02)*	(0.02)
	[0.39]**	[0.37]*	[1.18]*	[0.01]**	[0.01]*	[0.01]
Regions	66	65	64	66	65	64
Dep. var	GVA 95-00	GVA 95-00	GVA 01–07	Growth 77-00	Growth 80-00	Growth 80-07

Note: Second-stage results from two-stage least-squares (2SLS) regressions of output on conditional cultural measures. Columns [1] and [3] are reported from Tabellini's (2010) Table 2 and Table 3, respectively. Each line corresponds to a distinct regression in which the corresponding cultural measure is the main regressor of interest: from the 1990–1998 WVS waves in columns [1], [2], [4], and [5] ("old" sample), and from the 1999–2014 waves in columns [4] and [6] ("new" sample). The dependent variable is average gross value added per capita (at PPP) as a percentage of the EU15 between 1995 and 2000 in columns [1] and [2]; the analogous measure for 2001–2007 in column [3]; the average growth rate of GVA per capita between 1977 and 2000 in column [4], and the analogous 1980–2000 and 1980–2006 growth rates in, respectively, columns [5] and [6]. The excluded instruments are Literacy and PC institutions. School, urbanization, and country fixed effects are always included as conditioning variables; also, the initial GVA level (1977 or 1980) is included in columns [5] and [6]. Standard errors: conventional in parentheses; robust in brackets. Significance level: *5%; **1%. In columns [1] and [4], (n.a.) means that conventional standard errors are "not available" in the original article.

TABLE 8 Original versus replica reduced-form estimates.

	[1] Original waves 2–3	[2] Replication waves 2–3	[3] Replication waves 4–6	[4] Original waves 2–3	[5] Replication waves 2–3	[6] Replication waves 4–6
Literacy	0.81	0.64	0.62	n.a.	0.02	0.01
	(n.a.)	(0.22)**	(0.20)**		(0.01)	(0.01)
	[0.23]**	[0.23]**	[0.22]**		[0.01]*	[0.01]*
PC instit.	7.21	6.33	6.30	n.a.	0.21	0.16
	(n.a.)	(3.29)	(3.09)*		(0.16)	(0.10)
	[4.31]	[3.48]	[3.49]		[0.14]	[0.10]
Regions	67	67	67	67	67	67
Dep. var	GVA 95-00	GVA 95-00	GVA 01–07	Growth 77-00	Growth 80-00	Growth 80-07

Note: Results from reduced-form OLS regressions. The dependent variable is: average GVA per capita (at PPP) as a percentage of the EU15 between 1995 and 2000 in columns [1] and [2]; the analogous measure for 2001–2007 in column [3]; the average growth rate of GVA per capita between 1977 and 2000 in column [4]; and the analogous 1980–2000 and 1980–2006 growth rates in (respectively) columns [5] and [6]. The two sets of coefficient reported in the table are associated with the excluded instruments (i.e., Literacy and PC institutions). School, urbanization, and country fixed effects are always included as conditioning variables and the initial GVA level (1977 or 1980) is included in columns [5] and [6]. Standard errors: conventional in parentheses; robust in brackets. Significance levels: *5%; **1%. In columns [1] and [4], (n.a.) or n.a. Mean that conventional standard errors or estimates are "not available" in the original article.

TABLE 9 Original versus replica *F*-statistics for excluded instruments.

	[1] Original waves 2–3	[2] Replication waves 2–3	[3] Replication waves 4–6
Trust	[4.84]	(4.56) [4.56]	(5.57) [8.21]
Respect	[9.29]	(7.91) [7.31]	(0.48) [1.43]
Obedience	[3.20]	(1.79) [3.84]	(1.69) [2.85]
Control	[2.40]	(2.51) [3.72]	(1.19) [1.62]
PC culture	[10.83]	(11.77) [9.92]	(5.90) [11.32]
PC positive	[17.47]	(16.85) [18.02]	(5.66) [9.21]
PC children	[6.75]	(6.52) [4.48]	(1.93) [4.63]
Regions	67	65	64
WVS waves	1990–1998	1990–1998	1999–2014

Note: The table reports the *F*-statistic for the excluded instruments from the first-stage regression associated with the 2SLS regressions reported in 7. In this first-stage regression, the dependent variable is a conditional cultural measure and the excluded instruments are the literacy rate in 1880 and the first principal component of five variables measuring constraints on the executive in 1600, 1700, 1750, 1800 an 1850. Each line corresponds to a distinct regression. School, urbanization, and country fixed effects are always included as conditioning variables. Column [1] is reproduced from Table 5 in Tabellini (2010). The *F*-statistic in parentheses corresponds to conventional standard errors; the one in brackets corresponds to robust ones.

similarly decline, to values between 1.2 and 5.9. Although historical instruments are a possible solution to the volatility problem, here this solution fails in practice.

In light of the analysis carried out in Section 2, This reduced ability of historical instruments to predict culture in the 1999–2014 WVS waves suggests that the discrepancy between the original results and our replication is due to the instability of region-level cultural proxies between the 1990s and the 2000s, itself a manifestation of the general country-level volatility problem we have documented. Table 10 quantifies that instability by reporting, in analogy with the country-level statistics in 2, the standard deviation of the log of four alternative measures of Trust, Respect, Obedience, and Control within the regions of Europe: (i) the unconditional measure; (ii) the conditional measure, which accounts for different demographic or socioeconomic characteristics of respondents within regions across the old and new samples; (iii) the predicted unconditional measures from the IV estimation's first stage; (iv) the predicted conditional measures from the IV estimation's first stage. The last two measures are motivated by the rationale for using historical

TABLE 10 Within-region volatility of culture between 1990–1998 and 1999–2014 in the WVS.

	Unconditional	Conditional	Unconditional IV-filtered	Conditional IV-filtered
Trust	0.157	0.572	0.088	0.593
Respect	0.063	0.081	0.046	0.062
Obedience	0.170	0.176	0.120	0.160
Control	0.033	0.141	0.019	0.141
Regions	66	66	64	64

Note: The table reports, for each basic cultural measure used by Tabellini (2010), the average of the standard deviation of the log of that measure in two periods: 1990–1998 and 1999–2013. Both the unconditional measure (Equation 1) and the conditional measure (Equation 3) are employed, as are their predicted values from a regression on the excluded and included instruments in addition to country fixed effects ("IV-filtered").

instruments (which is precisely to filter out components of measured cultural traits that are unrelated to the underlying persistent component). Thanks to the aggregation of five WVS waves (waves 2–6) into two periods (1990–1998 and 1999–2014), the volatility reported in Table 10 is much less than in Table 2 (which suggests a possible avenue for mitigating the problem), and the instrumental variables do indeed reduce that volatility further. Yet as our replication reveals, even modest amounts of residual volatility (such as in Respect and Control) may be enough to weaken considerably the empirical connection between culture and economic outcomes.

5 | CONCLUSIONS

We have shown that cultural measures derived from the WVS are characterized by substantial volatility and heterogeneous time trends. The possible consequences of such lack of persistence were illustrated by revisiting the analyses in Gorodnichenko and Roland (2011) at the country level and Tabellini (2010) at the region level. Correlations between culture and output change (with the notable exception of generalized trust) when fairly complex cultural traits are measured in the 2000s instead of the 1990s.

Our analysis conveys four lessons. First, it is advisable to employ simple diagnostic tools (checks or formal tests) to assess the volatility of candidate measures, like those that we have used in Section 2. These tools may help selecting cultural proxies that are more stable across waves and therefore better suited to retain the persistent component. Such pre-analytical diagnostics are facilitated by the increased longitudinal span of surveys that collect data on values and belief. When the candidate measures turn out to be volatile, it inot volatile, such as religion or ethnicity. Simple traits may not suit complex research questions, but there is a trade-off between ts advisable to pool as many survey waves as possible. It is dangerous to use only one or few waves when each wave is affected by within-sample volatility. Second, there are advantages from using simple cultural aspects that are invariant for an individual and therefore he complexity of cultural traits and the reliability of their survey measures. In the specific case of the WVS, an additional difficulty is that the stated goal of the survey is not to assess persistence but rather to "understand *changes* in the beliefs, values and motivations". A number alternatives to survey questions are discussed by Lowes (2022), and include measures derived from textual analyses of traditional community beliefs, customs, and stories (Michalopoulos & Xue, 2021), which are invariant by construction. Third, measures derived from answers to questions structured around ordinal scales (as opposed to binary questions) should be avoided, as they may exacerbate the volatility problem. Finally, while historical instrumental variables for culture are meant to address this problem by filtering out the noisy component, plausibly exogenous historical instruments may not perform equally well when applied to contemporary survey data of different vintages. Hence, caution is advised even when valid instrumental variables are available.

ACKNOWLEDGMENTS

We are grateful to Jaime Díez-Medrano for help in obtaining the older region-level WVS data and for data clarifications as well as to Alberto Bisin, Umair Khalil, Andrea Moro, Guido Tabellini, Firmin Doko Tchatoka, and especially Fabio Bertolotti and Andrea Hamaui for useful suggestions and comments.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in openicpsr.org at https://doi.org/10.3886/E194661V3, reference number 194661 (Zanella & Bellani, 2023).

ORCID

Giulio Zanella https://orcid.org/0000-0001-8963-2351

ENDNOTES

[1] The WVS (Inglehart et al., 2018) is a nationally representative survey conducted in about 100 countries using the same questionnaire and aimed at understanding "changes in the beliefs, values and motivations of people throughout the world." The current version originates from the European Values Study (EVS), a smaller survey conducted in about 50 countries that was initiated in 1981. We refer to this unified data source as the WVS. To date, six waves are available: waves 1 (1981–1984), 2 (1990–1994), 3 (1995–1998), 4 (1999–2004), 5 (2005–2009), and 6 (2010–2014). Wave 7 (2017–2022) is still partly in progress.

[2] The set of surveys used to construct cultural proxies has also been expanded beyond the WVS to include: General Social Survey, European Social Survey, Schwartz Values Survey, Eurobarometer, Asian Barometer, Latinobarómetro, Afrobarometer, German Socio Economic Panel. Falk et al. (2018) implement the Global Preference Survey to measure, by way of experimentally validated survey questions, individual traits that can be regarded as cultural. The problem that we analyze may apply to all of these.

[3] The robustness of trust is consistent with a large body of agreeing evidence about the connection between survey-measured and experimentally observed trust (Falk et al., 2018; Johnson & Mislin, 2012) and between trust surveyed in places of residence and in places of ancestral origin (Algan & Cahuc, 2010; Giavazzi et al., 2019; Guiso et al., 2006; Tabellini, 2008; Uslaner, 2008).

[4] The WVS is not a panel, so individual fixed effects cannot be includes in Equation (2) or the other individual-level regressions employed in the paper.

[5] Tabellini (2010) builds on his 2007 Presidential Address to the European Economic Association (Tabellini, 2008), and Gorodnichenko and Roland (2011) extends Gorodnichenko and Roland (2017).

[6] For binary variables, Bond and Lang's problem does not arise because the share of individuals in one of two categories is always non-parametrically identified.

[7] Following the two original papers considered in the present study, we construct the "new" sample from pooling multiple WVS waves, thereby collapsing the country (or region) panel to a cross section.

[8] The former view is epitomized by Acemoglu and Robinson (2012), who argue that "countries differ in their economic success because of their different institutions" (p. 73); the latter view, by Landes (1998): "If we learn anything from the history of economic development, it is that culture makes all the difference" (p. 516). These two seemingly opposed views are not at odds provided that culture and institutions co-evolve (cf. Bisin & Verdier, 2017). Giavazzi et al. (2019) show that, for immigrants facing a different institutional environment in the United States, by the fourth generation many cultural traits have converged to the local norm. This evidence of "cultural mobility" seems to contradict evidence of "cultural persistence" (Guiso et al., 2016). A possible reconciliation is that culture is more likely to persist in the absence of major shocks such as radical institutional changes or contact with a different cultural majority. The coefficients estimated by Giavazzi *et al*. imply that, following such shocks, most traits converge to new cultural norms within a few generations. Guiso et al. (2006), Alesina and Giuliano (2015), and Lowes (2022) offer comprehensive overviews of the economic literature on culture, institutions, and economic outcomes.

[9] Wave 1 (1981–1984) of the WVS does not contain enough regional identifiers for a region-level analysis.

[10] According to Bowles and Gintis (1976), this factor was a prominent contributor to schooling in a traditional capitalist system.

[11] Conditioning is paramount in this region-level analysis because WVS sampling weights are available only at the country level. The WVS is representative at the national level but not necessarily at the regional one, so conditioning removes some of the sampling imbalances over time.

[12] Like in the replication of Gorodnichenko and Roland (2011), we reach the same conclusions when limiting ourselves to 1995–2000 GVA and 1977–2000 growth as outcomes—as one would expect given the persistence of *relative* output across the regions of Europe.

[13] Bisin and Moro (2021) provide a comprehensive overview of the use and interpretation of valid historical instrumental variables for endogenous persistent processes such as culture or institutions.

[14] Clustering standard errors at the country level would be inappropriate in this context given the small number of clusters (8 countries) and their unbalanced nature (different number of regions within countries), see Cameron and Miller (2015).

REFERENCES

Acemoglu, D. & Robinson, J.A. (2012) *Why nations fail: the origins of power, prosperity, and poverty*. Crown Business.

Acemoglu, D. & Robinson, J.A. (2021) *Culture, institutions and social equilibria: a framework*. National Bureau of Economic Research. Working Paper 28832.

Alesina, A. & Giuliano, P. (2015) Culture and institutions. *Journal of Economic Literature*, 53(4), 898–944. Available from: https://doi.org/10.1257/jel.53.4.898

Algan, Y. & Cahuc, P. (2010) Inherited trust and growth. *The American Economic Review*, 100(5), 2060–2092. Available from: https://doi.org/10.1257/aer.100.5.2060

Ananyev, M. & Guriev, S. (2018) Effect of income on trust: evidence from the 2009 economic crisis in Russia. *The Economic Journal*, 129(619), 1082–1118. Available from: https://doi.org/10.1111/ecoj.12612

Banfield, E.C. (1958) *The moral basis of a backward society*. The Free Press.

Berggren, N., Ljunge, M. & Nilsson, T. (2022) Immigrants' tolerance and integration into society. IFN Working Paper No. 1447.

Bisin, A. & Moro, A. (2021) LATE for history. In: Bisin, A. & Federico, G. (Eds.) *The Handbook of historical economics*, 269–296. Elsevier.

Bisin, A. & Verdier, T. (2011) *The economics of cultural transmission and socialization, volume 1 of Handbook of Social Economics*. North-Holland, 339–416.

Bisin, A. & Verdier, T. (2017) *On the joint dynamics of culture and institutions*. mimeo.

Bond, T.N. & Lang, K. (2019) The sad truth about happiness scales. *Journal of Political Economy*, 127(4), 1629–1640. Available from: https://doi.org/10.1086/701679

Bowles, S. & Gintis, H. (1976) *Schooling in capitalist America*. Basic Books.

Cameron, A.C. & Miller, D.L. (2015) A practitioner's guide to cluster-robust inference. *Journal of Human Resources*, 50(2), 317–372. Available from: https://doi.org/10.3368/jhr.50.2.317

Desmet, K., Ortuño Ortín, I. & Wacziarg, R. (2022) Endogenous partitions, working paper.

Deutsch, F. (2022) Political Participation across cultures. In: *The Oxford Handbook of political Participation*. Oxford University Press.

Edwards, J.A., Romero, A. & Naanwaab, C. (2022) Income inequality, the World Values Survey, and the interaction of cultural dimensions. *Applied Econometrics and International Development*, 22, 5–24.

Falk, A., Becker, A., Dohmen, T., Enke, B., Huffman, D. & Sunde, U. (2018) Global evidence on economic preferences. *Quarterly Journal of Economics*, 133(4), 1645–1692. Available from: https://doi.org/10.1093/qje/qjy013

Fernández, R. (2008) Culture and economics. *The New Palgrave Dictionary of Economics*, 1–8, 1229–1236.

Giavazzi, F., Petkov, I. & Schiantarelli, F. (2019) Culture: persistence and evolution. *Journal of Economic Growth*, 24(2), 117–154. Available from: https://doi.org/10.1007/s10887-019-09166-2

Giavazzi, F., Schiantarelli, F. & Serafinelli, M. (2013) Attitudes, policies, and work. *Journal of the European Economic Association*, 11(6), 1256–1289. Available from: https://doi.org/10.1111/jeea.12061

Gorodnichenko, Y. & Roland, G. (2011) Which dimensions of culture matter for long-run growth? *The American Economic Review*, 101(3), 492–498. Available from: https://doi.org/10.1257/aer.101.3.492

Gorodnichenko, Y. & Roland, G. (2017) Culture, institutions, and the wealth of nations. *The Review of Economics and Statistics*, 99(3), 402–416. Available from: https://doi.org/10.1162/rest_a_00599

Guiso, L., Sapienza, P. & Zingales, L. (2006) Does culture affect economic outcomes? *The Journal of Economic Perspectives*, 20(2), 23–48. Available from: https://doi.org/10.1257/jep.20.2.23

Guiso, L., Sapienza, P. & Zingales, L. (2011) *Civic capital as the missing link: Handbooks in economics*. North-Holland, 417–480.

Guiso, L., Sapienza, P. & Zingales, L. (2016) Long-term persistence. *Journal of the European Economic Association*, 14(6), 1401–1436. Available from: https://doi.org/10.1111/jeea.12177

Hasnat, B. (2022) Culture, institution, and policy response to COVID-19. *Journal of Economic Issues*, 56(2), 624–631. Available from: https://doi.org/10.1080/00213624.2022.2066446

Hofstede, G. (2001) *Culture's consequences: Comparing values, behaviors, institutions and organizations across nations*. Sage Publications.

Inglehart, R., Haerpfer, C., Welzel, C., Kizilova, K., Diez-Medrano, J., Lagos, M. et al. (2018) *World Values Survey: all rounds - country-pooled datafile version, technical report*. Madrid: JD Systems Institute.

Inglehart, R. & Welzel, C. (2005) *Modernization, cultural change, and democracy: the human development sequence*. Cambridge University Press.

Johnson, N.D. & Mislin, A. (2012) How much should we trust the World Values Survey trust question? *Economics Letters*, 116(2), 210–212. Available from: https://doi.org/10.1016/j.econlet.2012.02.010

Knack, S. & Keefer, P. (1997) Does social capital have an economic payoff? A cross-country investigation. *Quarterly Journal of Economics*, 112(4), 1251–1288. Available from: https://doi.org/10.1162/003355300555475

La Porta, R., Lopez-de-Silanes, F., Shleifer, A. & Vishny, R.W. (1997) Trust in large organizations. *The American Economic Review*, 87, 333–338.

Landes, D.S. (1998) *The wealth and poverty of nations*. Norton.

Lowes, S. (2022) Culture in historical political economy. In: Jenkins, J.A. & Rubin, J. (Eds.) *The Oxford Handbook of Historical Political Economy*. Oxford University Press.

Michalopoulos, S. & Xue, M.M. (2021) Folklore. *Quarterly Journal of Economics*, 136(4), 1993–2046. Available from: https://doi.org/10.1093/qje/qjab003

Pearl, J. (2009) *Causality*. Cambridge University Press.

Putnam, R. (1993) *Making democracy work: civic traditions in modern Italy*. Princeton University Press.

Roland, G. (2004) Understanding institutional change: fast-moving and slow-moving institutions. *Studies in Comparative International Development*, 38(4), 109–131. Available from: https://doi.org/10.1007/bf02686330

Tabellini, G. (2008) Presidential address: institutions and culture. *Journal of the European Economic Association*, 6(2-3), 255–294. Available from: https://doi.org/10.1162/jeea.2008.6.2-3.255

Tabellini, G. (2010) Culture and institutions: economic development in the regions of Europe. *Journal of the European Economic Association*, 8(4), 677–716. Available from: https://doi.org/10.1111/j.1542-4774.2010.tb00537.x

Uslaner, E.M. (2008) Where you stand depends upon where your grandparents sat: the inheritability of generalized trust. *Public Opinion Quarterly*, 72(4), 725–740. Available from: https://doi.org/10.1093/poq/nfn058

Zanella, G. & Bellani, M. (2023) *Data and code for "The volatility of survey measures of culture and its consequences"*. Inter-university Consortium for Political and Social Research [Distributor. Available from: https://doi.org/10.3886/E194661V3

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Zanella, G. & Bellani, M.M. (2024) The volatility of survey measures of culture and its consequences. *Economic Inquiry*, 62(2), 675–697. Available from: https://doi.org/10.1111/ecin.13202