I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 12730

The Volatility of Survey Measures of Culture and Its Consequences

Giulio Zanella
Marina M. Bellani

OCTOBER 2019

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 12730

The Volatility of Survey Measures of Culture and Its Consequences

Giulio Zanella
University of Adelaide, University of Bologna and IZA

Marina M. Bellani

OCTOBER 2019

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

The Volatility of Survey Measures of Culture and Its Consequences*

Common measures of cultural attitudes, such as those constructed from the World Values Survey, are characterized by substantial within-country volatility. This volatility is at odds with the notion of culture adopted in economics: a set of slow-moving traits that determine preferences and expectations transmitted from one generation to the next via family or social interactions. The insufficient persistence of survey proxies for such traits may compromise empirical studies of culture as a determinant of economic outcomes. We illustrate this point via a thorough replication, using the most recent WVS waves, of analyses carried out previously for regions in Europe.

JEL Classification: O12, O43, Z1

Keywords: World Values Survey, culture, development

Corresponding author:
Giulio Zanella
School of Economics
University of Adelaide
10 Pulteney Street
Adelaide SA 5005
Australia

E-mail: giulio.zanella@adelaide.edu.au

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

1 Introduction

Aggregate measures of cultural attitudes derived from answers to survey questions are often used in economics to investigate the relation between culture and macroeconomic outcomes. Two precursors of this approach 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 gross domestic product (GDP) growth or other aggregate outcomes across countries.[1] 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 attitudes. Two prominent examples are Tabellini (2010) and Gorodnichenko and Roland (2011). The former shows that four measures of values and beliefs built from the WVS (trust, respect, obedience, and internal locus of control) are associated with output across the regions of Europe; the case for causality relies on a within-country design and on historical instrumental variables. Gorodnichenko and Roland employ sixteen measures of values and beliefs from the WVS to build six variables capturing cultural attitudes (trust, hard work and thrift, tolerance, public good provision, equality, and market orientation); these attitudes are then shown to be weakly or strongly associated with GDP across countries; here causality relies on genetic markers as instruments. The set of surveys used to construct aggregate cultural proxies has also been expanded to include the General Social Survey, the European Social Survey, the Schwartz Values Survey, the Eurobarometer, the Asian Barometer, the Latinobarómetro, the Afrobarometer, and the 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.[2]

When addressing the methodological limits of using individual surveys to study culture as a determinant of economic outcomes, three problems have received most of the schol-

[1] The survey question is: "Generally speaking, would you say that most people can be trusted, or that you can't be too careful in dealing with people?" The percentage in a certain area responding that most people can be trusted is then taken as a measure of trust in that area. Knack and Keefer (1997) also consider aggregate investment as an outcome; La Porta *et al.* (1997) address measures of government effectiveness, civic participation, the success of large firms, and inflation.

[2] Guiso *et al.* (2006) and Alesina and Giuliano (2015) offer comprehensive overviews of the economic literature on culture, institutions, and economic outcomes. Guiso *et al.* (2011) and Falk *et al.* (2018) provide a rich discussion of measurement issues.

arly attention: the samples' representativeness, the endogeneity of the measures themselves (including reverse causality and omitted variables), and the survey questions' incentive compatibility (Fernández, 2008; Guiso *et al.*, 2011; Falk *et al.*, 2018). Far less attention has been devoted to another vital aspect – namely, the stability of survey measures over time. This aspect is important because such measures are meant to proxy for slow-moving, persistent cultural traits as opposed to fast-moving components that reflect economic, social, or political dynamics (Roland, 2004). La Porta *et al.* (1997) observe that the cross-country correlation between generalized trust measured through the WVS in the 1980s and in the 1990s is 0.91, but no longitudinal analysis of other cultural traits employed in the subsequent literature has been undertaken. 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), 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 via, for instance, contamination from higher-frequency components of values and beliefs. 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.

In this paper we document that these measures are *not* stable. Focusing on the nine WVS cultural proxies employed by Tabellini (2010) and Gorodnichenko and Roland (2011), we find that there is substantial volatility across the six waves of the World Values Survey.[3] Consider,

[3]The World Values Survey (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 of the WVS originates from the European Values Study (EVS), a smaller survey – conducted in about 50 European countries – that was initiated in 1981. We refer to this unified data source as the WVS. To date, six waves are available that span about three decades: waves 1 (1981–1984), 2 (1990–1994), 3 (1995–1998), 4 (1999–2004), 5 (2005–2009), and 6 (2010–2014).

for example, the *log* of the standard deviation of conditional cultural measures (so as to account for 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 10%. However, we also find that the average between-wave variation is more than half of the average between-country variation, which confirms that there is substantial volatility in the WVS measures of culture. Moreover, we establish that such volatility is not simply noise: most of the cultural measures exhibit significant time trends within countries. Different reasons may explain such lack of persistence. It is possible that surveys simply capture noisy measures of values and beliefs because individual answers are affected by unobserved factors other than the cultural attitudes that the interviewer aims to elicit. Another possibility, illustrated in a different context by Bond and Lang (2019), is that – when survey questions elicit values and beliefs through ordinal scales, as is typical of many WVS questions – the multiplicity of possible cardinalizations needed to compute aggregate measures of culture implies that the averages may change over time even when the underlying cultural attitudes are actually unchanged.

In either case, the volatility of survey measures of culture may induce fragility in the measured connection between culture and output. This claim is illustrated in a thorough replication of Tabellini (2010), a study that employs measures of culture built from the 1990–1998 waves of the WVS and then relates them either to average gross value added (GVA) during 1995–2000 across the regions of Europe or to the corresponding average growth rate between 1977 and 2000. We replicate this empirical exercise using more recent data for the same regions. In particular, we construct the same cultural measures employing the 1999–2014 waves of the WVS and then relate these measures to updated versions of output level and growth in an identical within-country design that leverages the same historical instrumental variables to address causality. If the aggregate survey measures are sufficiently correlated with the slow-moving component of culture on which the analysis focuses, then shifting measurement ten years forward should not make an appreciable difference. However, we find that it does: when employing the more recent WVS waves, the coefficients of the

linear regressions of output level (or growth) on the survey cultural measures often either decline in magnitude, becoming statistically insignificant, or become imprecisely estimated. We show that the fragility of the original results reflects the insufficient persistence, between the 1990–1998 and 1999–2014 waves of the WVS, of the cultural proxies used – a manifestation of the general volatility problem affecting indicators of culture across WVS waves. Our replication reveals also that possible solutions to this problem, 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. It is interesting that, when we employ the more recent 1999–2014 waves, the historical instruments constructed by Tabellini (2010) become weak and thereby weaken the case for causality. These results suggest caution when using survey information to proxy for slow-moving cultural traits. As an alternative, Guiso *et al.* (2006) leverage cultural aspects that are both more precisely measured and largely invariant for an individual, such as religion and ethnic background. It is also interesting that the effects of trust (as measured by the WVS and similar data sets) is robust in our replication when output level is used as a dependent variable (although not when output growth is used). In that case, there are few negative consequences to the trust measure's lack of persistence. This robustness is consistent with existing evidence: about the connection between survey-measured and experimentally observed trust (Johnson and Mislin, 2012; Falk *et al.*, 2018) and between trust surveyed in places of residence and in places of ancestral origin (Guiso *et al.*, 2006; Uslaner, 2008; Tabellini, 2008; Algan and Cahuc, 2010; Giavazzi *et al.*, 2019); about the critical role played by the ability to cooperate – for which trust is a necessary condition in a world of incomplete contracts and costly formal institutions – in explaining the behavioral gap between northern and southern Italy (Bigoni *et al.*, 2016); about how the distant past affects today's trust in Africa (Nunn and Wantchekon, 2011); and, more generally, about trust as a key component of civic capital (Guiso *et al.*, 2011).

The rest of our paper proceeds as follows: Section 2 documents the volatility of cultural measures derived from the WVS, and in Section 3 we illustrate the consequences of that volatility by replicating the analysis of Tabellini (2010) with more recent WVS data. Section 4 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 bee to careful in dealing with people?" If i answers "most people can be trusted", then $c_{igt} = 1$ (otherwise, $c_{igt} = 0$). 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. That is, the cultural attitude imputed to area g is constructed as

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

Regardless of whether the cultural measure c_{gt} is unconditional or conditional, it will incorporate both slow- and fast-moving components of culture (Roland, 2004; Guiso *et al.*, 2006). 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 waves t and $t-1$, provided these waves are close enough in time. 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.

We 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.[4] 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 and 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 characterizing individualist versus collectivist cultures and relate conceptually those traits 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 Mar-

[4] 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).

ket. 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 available waves of the WVS, to construct, as in Eqs. 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, education, social class, and health status) for 112 countries. All the averages are scaled so that they lie within the $[0, 100]$ interval.

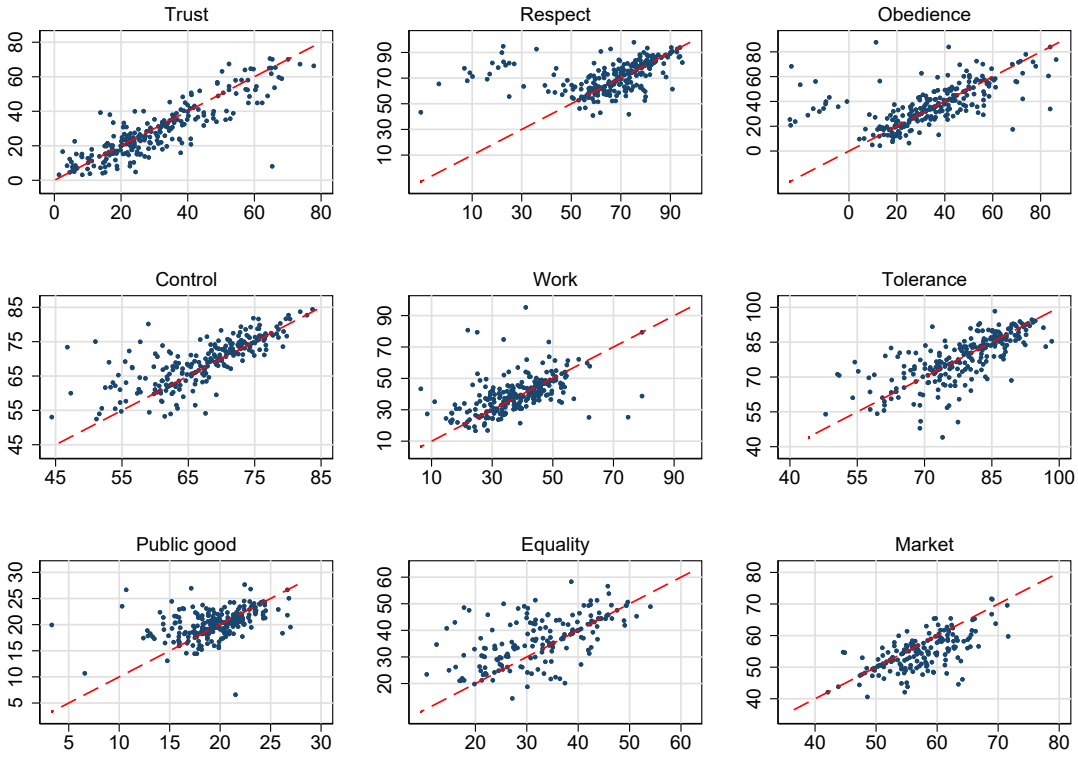
Table 1: WVS cultural measures and their rationale

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 "thift 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	Tolerant 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	Coordination culture (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	Competitive 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	Competitive culture (market orientation)

Notes: 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).

The volatility of these nine cultural measures is assessed from four different perspectives. First, in Figure 1 we plot the value taken by a conditional 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 a given cultural proxy may change considerably from one WVS wave to the next – even after we account for the effect of varying demographic and socioeconomic characteristics.[5]

Figure 1: Cross-country cultural measures in WVS wave t versus wave $t-1$



Notes: A point in a scatter plot is a country in two adjacent WVS waves. The vertical axes measure the value of a conditional cultural proxy (see Eqs. 2 and 3) 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 Word Values Survey waves (1–6), out of 112 countries present in at least one WVS wave.

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. The distribution of this volatility measure is shown in Figure 2 for each of the nine cultural variables. The Control and Tolerance variables exhibit modest volatility (less than 10%, on average), but the other variables are

[5]The same pattern is exhibited by an analogous figure that employs the *unconditional* measures.

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%.

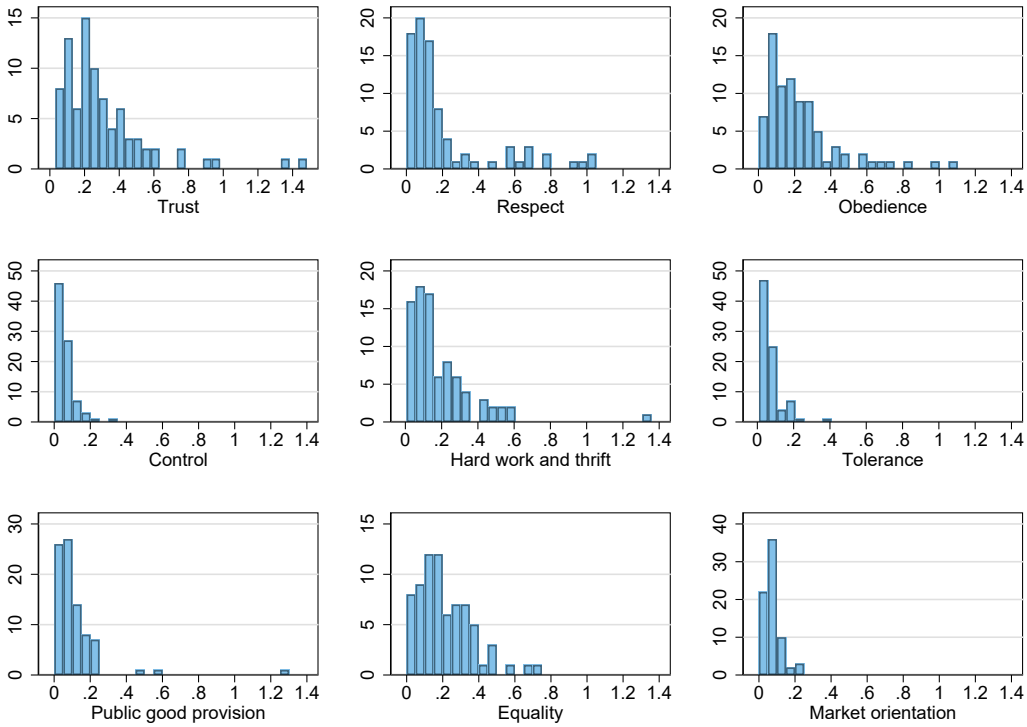
Third, we follow Falk *et al.* (2018) in decomposing the variance of the individual-level cultural measures to gauge the relative importance of between-country and between-wave variations in the WVS cultural proxies. The R^2 from an ordinary least-squares (OLS) regression of an individual's cultural 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 an OLS regression of an individual's cultural measure 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 values are reported in Table 3. In line with the findings of Falk *et al.* for the preference traits measured by the Global Preference Survey, the between-country variation in the WVS conditional cultural traits ranges between 5.7% and 13.9%. The associated between-wave variation is smaller (ranging from 3.5% to 10.4%) but it amounts (on average) to more than half the between-country variation. This statistic confirms the importance of longitudinal changes in the WVS cultural proxies.

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

$$\ln c_{gt} = \alpha_g + \beta t + \varepsilon_{gt}; \tag{4}$$

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. For instance, WVS respondents' support for income equality has *increased* (on average) by 11.6% in each survey wave from a base of 31% (SD = 8.9) in the 1980s. At the same time, respondents' market orientation has *decreased* by an average of 5.2% from a base of 63.2% (SD = 6.9). The fraction of respondents declaring that others can be trusted has also declined, by nearly 3% in each wave; however, Respect, Control, Work, Tolerance, and Public good have all increased, across waves, at rates that average between 2% and 9%.

Figure 2: Distribution of within-country volatility of conditional cultural measures



Notes: The figure plots the distribution of the within-country standard deviation of the log of each cultural measure. Sample: 85 countries present in at least two consecutive WVS waves (1–6), out of 112 countries present in at least one WVS wave. All of the cultural measures represented in the figure are conditional (see Eqs. 2 and 3).

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	0.262	0.304	–0.028*	(0.011)	–0.077**	(0.012)
Respect	0.142	0.213	0.046**	(0.007)	0.102**	(0.014)
Obedience	0.243	0.235	0.005	(0.012)	0.013	(0.013)
Control	0.056	0.059	0.020**	(0.003)	0.024**	(0.003)
Work	0.230	0.188	0.091**	(0.011)	0.047**	(0.010)
Tolerance	0.073	0.071	0.008*	(0.003)	0.001	(0.003)
Public good	0.103	0.114	0.023**	(0.006)	0.035**	(0.006)
Equality	0.298	0.216	0.116**	(0.011)	0.098**	(0.013)
Market	0.107	0.074	–0.052**	(0.004)	–0.040**	(0.004)

Notes: The left panel shows, for each cultural measure, the average of the standard deviation (SD) of the log of that measure. Both the unconditional measure (Eq. 1) and the conditional measure (Eq. 3) are employed. The table's right panel reports estimates of β from Eq. 4, 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 112 countries present in at least one WVS wave.

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	0.054
Tolerance	0.108	0.116	0.030	0.044
Public good	0.130	0.139	0.048	0.061
Equality	0.101	0.116	0.099	0.104
Market	0.086	0.106	0.064	0.082

Notes: 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 112 countries present in at least one WVS wave.

These patterns suggest that the WVS measures of culture contain a substantial high-frequency component as well as pronounced time trends. As mentioned previously, these components may reflect relatively quick changes in social attitudes associated with general economic, social and political trends (Roland, 2004; Guiso *et al.*, 2006); in addition, they could reflect a more fundamental problem with using *ordinal* individual measures from survey data to construct *cardinal* aggregate measures (Bond and Lang, 2019). Either way, this evidence casts some doubt on the ability of WVS culture measures to serve as reliable proxies for slow-moving, persistent cultural traits. And nor should they; after all, the stated goal of the WVS is to "understand *changes* in the beliefs, values and motivations of people" (italics added) and not to capture their persistence. We next demonstrate the possible consequences of employing these non-persistent measures to investigate the cultural roots of economic outcomes by replicating the analysis of Tabellini (2010) with more recent WVS data.

3 Culture and economic development in Europe

Tabellini (2010) reports a statistically and economically 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 his OLS and instrumental variables (IV) estimates deliver associations consistent with theories that posit a causal effect of history (past institutions) on current culture and, in turn, of current culture on development and hence on current output via the improved functioning of current informal and formal institutions. This interpretation is supported by Tabellini's 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. In this conceptual framework, culture is a slow-moving trait that is transmitted from parents or society to children and that is central to the mechanism through which history (especially past economic and political institutions) affects the functioning of current institutions and therefore of current economic performance. This view of culture as the "missing link" is illustrated in the following causal directed acyclic graph (Pearl, 2009):

$$\begin{array}{ccccccccc} \text{past} & & \text{past} & & \textbf{current} & & \text{functioning of} & & \textbf{economic} \\ \text{institutions} & \rightarrow & \text{culture} & \rightarrow & \textbf{culture} & \rightarrow & \text{current institutions} & \rightarrow & \textbf{development.} \end{array}$$

Such a view reconciles approaches that stress either the primacy of institutions or the primacy of culture. 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).[6]

[6]These two seemingly opposed views are not at odds provided that culture and institutions co-evolve (cf. Bisin and 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.

3.1 Replication data

We replicate the empirical exercise in question but with more recent data for the same regions. 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. We shall next describe these data in more detail.

3.1.1 Cultural measures

Data on cultural attitudes are from the EVS and the WVS (see footnote 3). 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).[7] 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 ($N \approx 35000$) is some 40% larger than the original 1990–1998 sample ($N \approx 25000$). 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. Table 4 summarizes the key size statistics of the two samples. To simplify the discussion, hereafter we reference the 1990–1998 and 1999–2014 samples as (respectively) the "old" and "new" samples. Since Tabellini (2010) was published, the WVS data (all waves) have been officially revised to clean data errors and impute missing values. For this reason, 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

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

a full correspondence with the old sample, we requested and received from the personnel in charge of WVS data maintenance the wave 3 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.

Table 4: World Values Survey sample size by country

	1990-1998 sample ("old")				1999-2014 sample ("new")			
	Total size	Region size Mean	SD	Waves	Total size	Region size Mean	SD	Waves
Belgium	2,792	930.7	557.9	1	3,421	1,140.3	495.5	2
France	1,002	125.3	29.9	1	3,756	469.5	138.1	3
Germany	3,017	377.1	266.7	2	3,970	496.3	332.1	4
Italy	2,017	155.2	79.1	1	4,531	348.5	210.4	3
Netherlands	1,014	253.5	185.3	1	5,499	1,374.8	543.3	4
Portugal	1,185	197.5	115.7	1	2,553	510.6	343.4	2
Spain	11,067	737.8	869.3	2	6,298	419.9	324.9	6
UK	2,577	257.7	140.3	2	3,561	356.1	199.8	3
N. Ireland	304	304	0	1	1,500	1,500	0	2
Total	24,975				35,089			

Notes: This table shows, by country, the number of persons interviewed ("Total size" column), the average and standard deviation of the number of WVS interviews within each region ("Region size"), and the number of times such interviews were conducted ("Waves"). The same country may appear more than once, though in different years, in a WVS wave. The 1990–1998 sample results from pooling WVS waves 2 (1990–1994) and 3 (1995–1998); it is the original sample used by Tabellini (2010). The 1999–2014 sample results from pooling WVS waves 4 (1999–2004), 5 (2005–2009), and 6 (2010–2014).

As mentioned in Section 2, the four cultural measures selected by Tabellini (2010) (first four rows of Table 1) are motivated by their relation to the backwardness in southern Italy documented by Banfield (1958) and Putnam (1993). An extensive discussion of the values and beliefs captured by these measures is given in Tabellini's article. 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. According to Bowles and Gintis (1976), this factor was a prominent contributor to schooling in a traditional capitalist system. To reduce such arbitrariness, Tabellini (2010) employs the first principal components (PCs) of three different combinations of Trust, Respect, Obedience, and Control. In what follows, these combinations are labeled PC culture (first principal component of the 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 Eqs. 2 and 3, while using survey information on age, gender, marital status, education, social class, and health status.[8] Moreover, the standard error of the estimated regional dummy coefficient, $\hat{\sigma}(\beta_{2,g})$, 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 Tabellini (2010) in all of these choices – constructing unconditional and conditional cultural measures from the raw data available at the WVS data center for both the old and the new sample – in setting up a replication that is at once more narrowly defined (to validate our analysis) and also broader in the sense that it is augmented with more recent data of the same type.

3.1.2 Output measures

Data on output, which constitute the outcome in the historical experiment illustrated previously, 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;

[8] 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.

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, which yields – consistently with the forward measure of culture in the 1999–2014 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 European countries from 2008 onward. 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. 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.

3.1.3 Instrumental variables

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 one would like to examine in an empirical investigation of culture and economic development. 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. 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. 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 $\mathbf{w}_g$: 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. The variables in $\mathbf{z}_g$ and $\mathbf{w}_g$ are summarized in Table 5. We take these variables from the Tabellini (2010) data files, which are available at the author's Web page. If we use θ_j to denote country fixed effects, then the reduced-form coefficients γ_1 from the linear regression

$$y_{gj} = \gamma_0 + \gamma_1 \mathbf{z}_g + \gamma_2 \mathbf{w}_g + \theta_j + u_{gj}, \tag{5}$$

express the influence of past regional economic and political institutions on current region prosperity (as posited, e.g., by Acemoglu and Robinson (2012)). The first-stage prediction of the relevant cultural trait – namely $\hat{c}_{gj} \equiv \mathbb{E}[c_g \mid \mathbf{z}_g, \mathbf{w}_g, \theta_j]$ – 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. Therefore, the IV specification is crucial in supporting the "missing link" view of cultural traits.

3.2 Replication results

The replication proceeds in three steps. As the first, preliminary step, we illustrate in Figure 3 the correspondence between the cultural and output measures in Tabellini (2010) (vertical axis) and those computed by us for the same period (i.e., 1990–1998 for culture and 1995–2000 for output, horizontal axis). Each point in a scatter plot represents a region of Europe, so an exact replication of the original measures would produce points aligned along the (dashed) 45° line. The graphs show that the alignment, although not perfect, is quite

Table 5: Excluded and included instruments in Tabellini (2010)

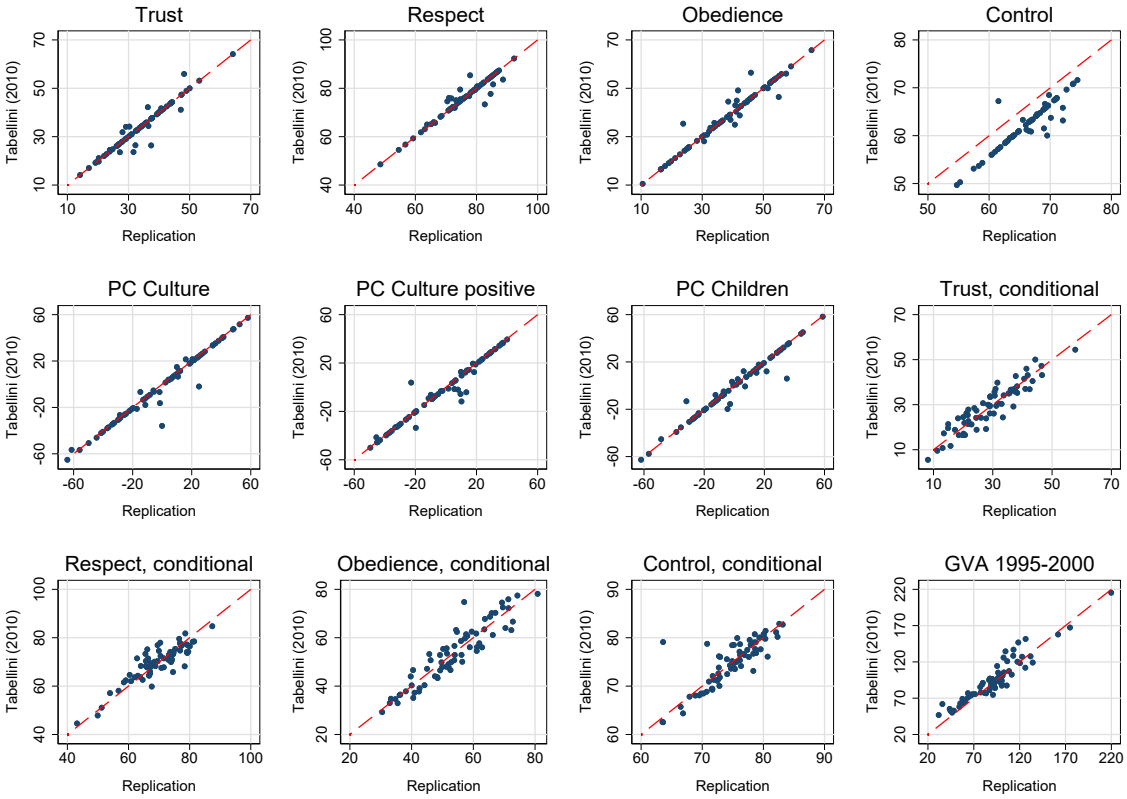
Instrument	**Definition**	**Proxies for**
Literacy	Literacy rate in 1880 (percentage of people who could read and write) in region	Educational history
PC institutions	First principal component of variables measuring constraints on the executive at different times (1600, 1700, 1750, 1800, and 1850)	Political history
School	Gross enrollment rate in primary and secondary school in 1960	Human capital
Urbanization	% of regional population living in cities with more than 30,000 inhabitants in 1850	Convergence
Fixed effects	Country fixed effects	Formal institutions

Notes: The table lists the excluded and included instruments employed by Tabellini (2010). See his historical appendix.

good. As for the unconditional cultural measures, the small discrepancies are due mainly to data revisions of the older WVS waves. The most notable discrepancy, for Control, is substantial but immaterial: the gap is roughly constant across regions and so the regression constant (or the country fixed effect) fully absorbs it.

For the conditional measures, deviations from the 45° line are larger. This outcome must reflect an imperfect correspondence between Tabellini's (2010) and our versions of the conditioning variables employed in within-region regressions of the unconditional cultural measures on individual characteristics (i.e., vector $\mathbf{x}_{ig}$ in Eq. 2). Tabellini writes: "we have regressed each of the cultural variables described herein (including the individual principal components) on a vector of regional dummy variables, as well as on the following additional regressors: marital status, gender, the age group, a self reported social class, and two categorical variables for health condition and years of education" (p. 689). However, the exact variables used are not listed. We employed the following ones: dummies for marital status (married or cohabiting, divorced or separated, widowed, and single), a gender dummy, two dummies for the age group (whether younger than 30; whether older than 60), three dummies for self-reported social class (lower or working class, lower-middle class, upper-middle class), a dummy for poor self-reported health status, and two dummies for educational attainment (high school diploma; at least some college). To balance the samples, we created dummies

Figure 3: Replication of cultural and output measures



Notes: A point in a scatter plot is a region of Europe (eight countries, see Section 3.1). The vertical axes measure the value of a cultural or output variable in the data set of Tabellini (2010), and each horizontal axis measures the value of that same variable as replicated by us from the raw data in the old sample (1990-1999). The dashed line is the 45° line.

for missing information on gender, marital status, social class, education, and health status; we then included these indicators as additional regressors in $\mathbf{x}_{ig}$.

Finally, Figure 3 reveals deviations from the 45° line also for average GVA 1995–2000 expressed as a percentage of the EU15 average. We sought possible sources of this discrepancy by corresponding with the Cambridge Econometrics staff, who confirmed that the GVA data in the current European Regional Database *differ* from the corresponding GVA data available in the European Regional Database ten years ago because of revisions and updates in the source database maintained by Eurostat. In any case, the deviations from the 45° line observed in Figure 3 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 replication, and it ensures that the different results obtained later using the new WVS waves are only due to the volatility of the cultural measures.

3.2.1 Replication using the old WVS waves

The second step is to replicate the original OLS and IV results reported in Tabellini (2010), which derive from this estimating equation:

$$y_{gj} = \delta_0 + \delta_1 c_{gj} + \delta_2 \mathbf{w}_g + \theta_j + \upsilon_{gj}; \tag{6}$$

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 $\hat{\sigma}(\beta_{2,g})$ (as defined in Section 3.1.1). 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 6. We computed three different sets of standard errors: conventional (reported in parentheses), robust with respect to heteroskedasticity [in brackets], and clustered at the country level {in braces}. Tabellini (2010) reports only robust and clustered standard errors, which may be misleading in this context. As for robust standard errors, the small sample size (69 regions) may lead to a poor asymptotic approximation that results in smaller standard errors than conventional ones, in which case the latter would be a better guide to inference. Angrist and Pischke (2008) suggest the following rule of thumb: use "the maximum of old-fashioned and robust standard errors to avoid gross misjudgments of precision". Our replication and comparison of conventional and robust standard errors indicates that no such misjudgment was made by Tabellini. Clustering standard errors at the country level seems inappropriate in this context given the small number of clusters (eight countries) and their unbalanced nature (different number of regions within countries). Cameron and Miller (2015) show that in the presence of few unbalanced clusters[9] estimated clustered standard errors may be downward biased,

[9] Cameron and Miller (2015) point out that "there is no clear-cut definition of 'few.' Depending on the situation, 'few' may range from less than 20 clusters to less than 50 clusters in the balanced case and even more clusters in the unbalanced case." (p. 342)

again, exaggerating the estimates' precision. In our replication, standard errors clustered at the country level turn out to be larger than their conventional counterpart on most occasions.

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.

3.2.2 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. The dependent variables are updated correspondingly: 2001–2007 GVA and 1980–2007 growth.[10] Our results are reported in columns [3] and [6] of both Table 6 and Table 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 robustly associated with output. As a consequence, the two PC measures of culture that incorporate Trust are likewise robust. Yet for Respect, Obedience, and Control as well as their PC syntheses, the estimates are no longer statistically significant at conventional confidence levels.

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

[10] This approach follows the rationale given by Tabellini (2010): "Because culture is measured in the 1990s, we confine most of the analysis to the more recent period, taking the average of per capita GVA over the period 19952000." (p. 681). However, 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.

Table 6: Original vs replica OLS estimates – 1990–1998 (old) and 1999–2014 (new) samples

	Unconditional culture			Conditional culture		
	[1]	[2]	[3]	[4]	[5]	[6]
	Original	Replica	Replica	Original	Replica	Replica
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]*
	{0.53}	{0.44}$^{+}$	{0.64}*	{0.68}	{0.53}	{0.73}
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]
	{0.63}*	{0.47}*	{0.58}	{0.59}*	{0.55}*	{0.55}
Obedience	–0.93	–0.90	–0.99	–0.68	–0.80	–0.80
	(n.a.)	(0.41)*	(0.50)	(n.a.)	(0.41)$^{+}$	(0.53)
	[0.46]*	[0.38]*	[0.43]*	[0.48]	[0.39]*	[0.40]
	{0.64}	{0.41}*	{0.59}	{0.68}	{0.47}	{0.58}
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]
	{0.39}*	{0.38}*	{1.28}	{0.12}**	{0.54}*	{1.40}
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]**
	{0.17}*	{0.10}**	{0.23}*	{0.19}*	{0.13}**	{0.30}
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]*
	{0.11}**	{0.12}**	{0.30}	{0.13}**	{0.16}**	{0.38}
PC children	0.57	0.49	0.47	0.58	0.45	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]*
	{0.27}$^{+}$	{0.18}*	{0.30}	{0.28}$^{+}$	{0.20}*	{0.30}
No. regions	69	68	67	68	67	66
WVS waves	1990-1998	1990-1998	1999-2014	1990-1998	1990-1998	1999-2014
Dep. var.	GVA 95-00	GVA 95-00	GVA 01-07	GVA 95-00	GVA 95-00	GVA 01-07

Notes: 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; clustered (at the country) level in braces. Significance levels: $^{+}$10%; *5%; **1%.

Table 7: Original vs replica IV estimates – 1990–1998 (old) and 1999–2014 (new) samples

	[1]	[2]	[3]	[4]	[5]	[6]
	Original	Replica	Replica	Original	Replica	Replica
Trust	4.67	4.26	4.36	0.06	0.12	0.04
	(n.a.)	(1.81)**	(1.42)**	(n.a.)	(0.06)$^{+}$	(0.03)
	[1.41]**	[1.55]*	[1.15]**	[0.03]*	[0.06]$^{+}$	[0.02]*
	{1.73}*	{1.99}*	{1.67}**	{0.04}	{0.10}	{0.02}**
Respect	2.86	2.90	13.75	0.03	0.09	0.13
	(n.a.)	(1.01)**	(13.20)	(n.a.)	(0.05)$^{+}$	(0.17)
	[0.76]**	[1.00]**	[9.70]	[0.02]$^{+}$	[0.04]*	[0.14]
	{0.60}**	{1.42}*	{11.36}	{0.02}$^{+}$	{0.06}	{0.13}
Obedience	–5.88	–4.83	–6.27	–0.08	–0.14	–0.06
	(n.a.)	(2.33)*	(3.41)$^{+}$	(n.a.)	(0.08)	(0.04)
	[2.19]**	[2.19]*	[2.76]*	[0.04]*	[0.07]$^{+}$	[0.03]*
	{1.90}*	{3.32}	{3.49}$^{+}$	{0.03}$^{+}$	{0.12}	{0.03}$^{+}$
Control	13.17	11.32	21.58	0.18	0.32	0.20
	(n.a.)	(6.08)$^{+}$	(13.68)	(n.a.)	(0.21)	(0.15)
	[7.61]$^{+}$	[5.90]$^{+}$	[12.92]	[0.12]	[0.21]	[0.14]
	{6.53}$^{+}$	{4.00}*	{15.00}	{0.15}	{0.23}	{0.16}
PC culture	1.11	0.91	1.89	0.02	0.03	0.02
	(n.a.)	(0.29)**	(0.61)**	(n.a.)	(0.01)*	(0.01)
	[0.28]**	[0.29]**	[0.55]**	[0.01]**	[0.01]*	[0.01]*
	{0.39}*	{0.43}*	{1.04}$^{+}$	{0.01}*	{0.02}	{0.01}
PC positive	1.16	0.94	2.12	0.02	0.03	0.02
	(n.a.)	(0.29)**	(0.70)**	(n.a.)	(0.01)*	(0.01)
	[0.32]**	[0.30]**	[0.66]**	[0.01]**	[0.01]*	[0.01]*
	{0.37}*	{0.37}*	{1.16}$^{+}$	{0.01}	{0.02}	{0.01}
PC children	1.40	1.30	2.75	0.02	0.04	0.03
	(n.a.)	(0.52)*	(1.43)$^{+}$	(n.a.)	(0.02)$^{+}$	(0.02)
	[0.39]**	[0.52]*	[1.18]*	[0.01]**	[0.01]*	[0.01]$^{+}$
	{0.48}*	{0.77}*	{1.66}$^{+}$	{0.01}$^{+}$	{0.03}	{0.02}
Regions	66	65	64	66	65	64
WVS waves	1990-1998	1990-1998	1999-2014	1990-1998	1990-1998	1999-2014
Dep. var	GVA 95-00	GVA 95-00	GVA 01-07	growth 77-00	growth 80-00	growth 80-07

Notes: 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 (see Table 5). 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; clustered at the country level in braces. Significance level: $^{+}$10%; *5%; **1%.

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. 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 Eq. 5 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 the regions of Europe, historical instruments correlate as well with average GVA during 1995–2000 as they do with average GVA in 2001–2007.

Table 8: Original versus replica reduced-form estimates

	[1]	[2]	[3]	[4]	[5]	[6]
	Original	Replica	Replica	Original	Replica	Replica
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]*
	{0.23}*	{0.19}**	{0.17}**		{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]$^{+}$
	{4.42}	{4.07}	{4.37}		{0.18}	{0.13}
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

Notes: 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; see Table 5). 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; clustered at the country level in braces. Significance levels: $^{+}$10%; *5%; **1%.

In fact, the culprit is at the first stage – that is, when regressing the cultural measures on the instruments. 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) 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.

Table 9: Original versus replica F-statistics for excluded instruments

	Original	Replica	Replica
Trust	4.84	(4.56) [4.56] {9.98}	(5.57) [8.21] {5.82}
Respect	9.29	(7.91) [7.31] {16.42}	(0.48) [1.43] {1.26}
Obedience	3.20	(2.58) [3.84] {2.91}	(1.69) [2.85] {3.52}
Control	2.40	(2.51) [3.72] {26.68}	(1.19) [1.62] {1.39}
PC culture	10.83	(11.77) [9.92] {10.70}	(5.90) [11.32] {3.85}
PC positive	17.47	(16.85) [18.02] {27.80}	(5.66) [9.21] {2.64}
PC children	6.75	(4.37) [4.48] {6.97}	(1.93) [4.63] {2.87}
Regions	66	65	64
WVS waves	1990-1998	1990-1998	1999-2014

Notes: The table reports the F-statistic for the excluded instruments from the first-stage regression associated with the 2SLS regressions reported in Table 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 (see Table 5). Each line corresponds to a distinct regression. School, urbanization, and country fixed effects are always included as conditioning variables. The F-statistic in parentheses corresponds to conventional standard errors; the one in brackets corresponds to robust standard errors; and the one in braces corresponds to standard errors clustered at the country level.

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 cultural proxies between the 1990s and the 2000s, itself a manifestation of the general volatility problem we have documented. Table 10 quantifies that instability by reporting, in analogy with the country-level statistics in Table 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 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.

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	0.089	0.593
Respect	0.063	0.076	0.046	0.066
Obedience	0.170	0.150	0.120	0.133
Control	0.033	0.105	0.019	0.102
Regions	66	66	64	64

Notes: 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 (Eq. 1) and the conditional measure (Eq. 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").

4 Conclusions

This study has established that cultural measures derived from the WVS – and frequently employed to investigate the connection between culture and macroeconomic outcomes – are characterized by substantial volatility and time trends. We illustrated the possible consequences of such lack of persistence via a thorough replication of the empirical results in Tabellini (2010), which weaken considerably (with the notable exception of generalized trust if output level is the dependent variable) when culture is measured in the 2000s instead of the 1990s. Roland (2004) and Guiso *et al.* (2006) make a useful distinction between slow-moving and fast-moving components of culture, where the latter originate in economic, social, and political dynamics. In light of the general volatility problem affecting survey measures that we have documented, one interpretation of our findings is that the cultural proxies selected by Tabellini (2010) are unstable over 10–15 years because to some extent they reflect *short*-run dynamics or trends in respondents' perceptions. Recall that the stated goal of the WVS is not to assess persistence but rather to "understand *changes* in the beliefs, values and motivations of people" (italics added).

An additional source of volatility is that aggregating ordinal scales (which are often used in the WVS) generates cardinal measures of culture that are comparable neither over time nor across geographic areas. Historical instrumental variables for culture are meant to address this issue by filtering out the noisy component in these unstable proxies, but the instruments employed by Tabellini (2010) turn out to be weak when applied to survey data of a more recent vintage. Hence, caution is advised when using survey questions to proxy for slow-moving cultural traits – even when plausibly valid instrumental variables are available. Guiso *et al.* (2006) seek to avoid this problem by choosing "cultural aspects like religion and ethnic background that can largely be treated as invariant over an individual's lifetime" (p. 24). It seems that some of the WVS proxies violate this requirement in a way that conditional measures or instrumental variables cannot accommodate. Cultural economics is an active research area. Although the last two decades have witnessed impressive advances, measurement once again lags behind theory. Our paper should be of use to empirical researchers in this field, which clearly needs robust measures of cultural traits.

References

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

Alesina, A., Giuliano, P., 2015, Culture and institutions, Journal of Economic Literature 53, 898–944.

Algan, Y., Cahuc, P., 2010, Inherited trust and growth, American Economic Review 100, 2060–92.

Ananyev, M., Guriev, S., 2018, Effect of income on trust: evidence from the 2009 economic crisis in Russia, The Economic Journal 129, 1082–1118.

Angrist, J. D., Pischke, J.-S., 2008, Mostly harmless econometrics: an empiricist's companion, Princeton University Press.

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

Bigoni, M., Bortolotti, S., Casari, M., Gambetta, D., Pancotto, F., 2016, Amoral familism, social capital, or trust? The behavioural foundations of the Italian northsouth divide, The Economic Journal 126, 1318–1341.

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.

Bowles, S., Gintis, H., 1976, Schooling in capitalist America, Basic Books.

Cameron, A. C., Miller, D. L., 2015, A practitioners guide to cluster-robust inference, Journal of Human Resources 50, 317–372.

Falk, A., Becker, A., Dohmen, T., Enke, B., Huffman, D., Sunde, U., 2018, Global evidence on economic preferences, The Quarterly Journal of Economics 133, 1645–1692.

Fernández, R., 2008, Culture and economics, The New Palgrave Dictionary of Economics: Volume 1–8 1229–1236.

Giavazzi, F., Petkov, I., Schiantarelli, F., 2019, Culture: persistence and evolution, Journal of Economic Growth 117–154.

Gorodnichenko, Y., Roland, G., 2011, Which dimensions of culture matter for long-run growth?, American Economic Review 101, 492–98.

Gorodnichenko, Y., Roland, G., 2017, Culture, institutions, and the wealth of nations, Review of Economics and Statistics 99, 402–416.

Guiso, L., Sapienza, P., Zingales, L., 2006, Does culture affect economic outcomes?, Journal of Economic Perspectives 20, 23–48.

Guiso, L., Sapienza, P., Zingales, L., 2011, Civic Capital as the Missing Link: Handbooks in Economics, 417–480, North-Holland.

Guiso, L., Sapienza, P., Zingales, L., 2016, Long-term persistence, Journal of the European Economic Association 14, 1401–1436.

Inglehart, R., C. Haerpfer and, A. M., Welzel, C., Kizilova, K., Diez-Medrano, J., M. Lagos, P. N., Ponarin, E., Puranen, B., 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, 210–212.

Knack, S., Keefer, P., 1997, Does social capital have an economic payoff? a cross-country investigation, The Quarterly Journal of Economics 112, 1251–1288.

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.

Nunn, N., Wantchekon, L., 2011, The slave trade and the origins of mistrust in Africa, American Economic Review 101, 3221–3252.

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, 109–131.

Tabellini, G., 2008, Presidential Address: Institutions and culture, Journal of the European Economic Association 6, 255–294.

Tabellini, G., 2010, Culture and institutions: economic development in the regions of Europe, Journal of the European Economic Association 8, 677–716.

Uslaner, E. M., 2008, Where you stand depends upon where your grandparents sat: The inheritability of generalized trust, Public Opinion Quarterly 72, 725–740.