Journal of Comparative Economics 000 (2015) 1–21

Contents lists available at ScienceDirect

Journal of Comparative Economics

journal homepage: www.elsevier.com/locate/jce

Language, culture and institutions: Evidence from a new linguistic dataset☆

Lewis S. Davis*, Farangis Abdurazokzoda

Department of Economics, Union College, Schenectady NY 12308, United States

ARTICLE INFO

Article history:
Received 26 April 2014
Revised 16 October 2015
Available online xxx

JEL classification:
C80
P51
Z13

Keywords:
Language
Culture
Institutional quality
Individualism
Egalitarianism
Linguistic relativity

ABSTRACT

Davis, Lewis S., and Abdurazokzoda, Farangis—Language, culture and institutions: Evidence from a new linguistic dataset

Kashima and Kashima's (1998) linguistic dataset has played a prominent role in the economics of culture, providing the instrumental variables used in two seminal works to identify the causal effect of culture on institutional quality. However, for economists, this dataset has a number of weaknesses, including poor overlap with a key cultural dataset and reliance on sources of linguistic information of uneven quality. We address these issues by constructing a new linguistic dataset based on an authoritative source of linguistic information, the *World Atlas of Language Structures*. The resulting dataset has greater overlap with key sources of cultural information, is arguably less subject to selection bias, and provides more refined information regarding key dimensions of linguistic variation. We show that the variables in this dataset are significantly correlated with commonly used measures of individualism and egalitarianism. In addition, we reexamine the key results from the literature on culture and institutions, showing the causal relationship between culture and institutions is robust to the use of the new linguistic instruments. *Journal of Comparative Economics* **000** () (2015) 1–21. Department of Economics, Union College, Schenectady NY 12308, United States.

1. Introduction

The recent emergence of an economics of culture focused on testable hypotheses has required overcoming two empirical hurdles. The first, regarding the significant challenges involved in measuring cultural variables, has been addressed through the use of international surveys of cultural values and beliefs, the information in which may be aggregated to produce measures of national culture that correspond to key dimensions of cultural variation, such as individualism-collectivism and egalitarianism-hierarchy. The second empirical hurdle regards the identification of causal effects. As Bowles (1998) has persuasively argued, cultural values are endogenous to economic and political forces, one implication of which is that OLS regressions with cultural variables do not in general recover causal effects.

Consider, for example, Fig. 1, which provides a scatter plot and the line of best fit from regressing Voice and Accountability, a measure of democracy, on Hofstede's (2001) measure of individualism. (Both variables are described below.) The strong positive relationship displayed in the figure may indicate that culture has a causal relationship to institutions, with individualist countries

☆ Lewis Davis wishes to acknowledge the support of the Faculty Exchange Network at New York University, where he was in residence at the start of this project. The authors wish to thank Mariko Klasing, Stephen Schmidt, Claudia Williamson, and two anonymous referees for useful comments. Any remaining errors are our own.

* Corresponding author. Tel. +518-388-8395
E-mail address: davisl@union.edu (L.S. Davis).

http://dx.doi.org/10.1016/j.jce.2015.10.015
0147-5967/

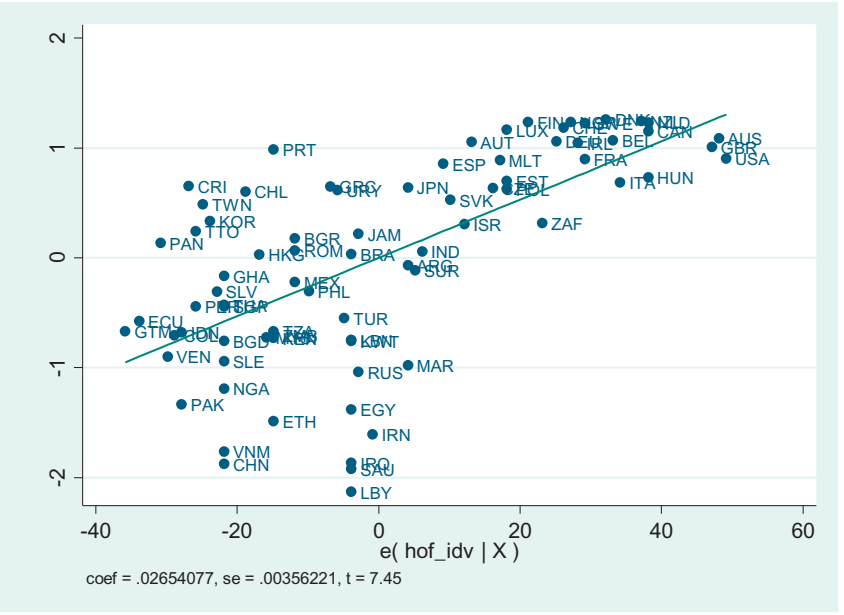

Fig. 1. Democracy and individualism across countries.

having some advantage in developing and sustaining democratic political institutions. But it could just as well reflect the influence of democratic political norms on cultural values, or even the influence of some third factor, such as economic development, that affects both political institutions and social values.

Oddly enough, linguistics has played a key role in identifying the causal effect of culture on economic and social outcomes. The linguistic relativity hypothesis, developed by Whorf (1956) and Sapir (1970), holds that the language an individual speaks plays a role in determining how he or she views the world. Kashima and Kashima (1998), hereafter *KK*, draw on this proposition to argue that the grammatical rules that govern personal pronoun use exert a causal and statistically significant effect on key dimensions of cultural variation. Licht et al. (2007) and Tabellini (2008), two seminal articles on the economics of culture, use the KK linguistic variables to identify the casual relationship between culture and institutional quality. In motivating the use of these instruments, Licht et al. (2007, p. 672) argue that the "grammar of a language may transmit and reproduce culture and social categories" and that therefore "language functions as a constraint of cultural change." In addition, Tabellini (2008, p. 273) argues that, because language evolves slowly, linguistic variables will be valid instruments for culture, even if causation flows both ways: "language is a valid instrument [for culture] … if the distant traditions that are responsible for current linguistic rules are not correlated with other unobserved determinants of the current quality of government." While additional cultural instruments have subsequently become available, the *KK* linguistic instruments have continued to play a central role in identifying the causal effect of culture on political and economic outcomes.[1]

In spite of the role it plays in the economics of culture, the *KK* dataset is not without its weaknesses. In constructing their dataset, *KK* begin with a sample that includes countries from four cultural studies, Hofstede (1980), Schwartz (1994), the Chinese Culture Connection (1987) and Smith et al. (1996), and draw on a large variety of sources in the linguistics literature as well as native speakers to identify grammatical rules for each of 39 languages spoken in 71 of the countries of their sample. This approach appears to have had two unintended consequences. First, the linguistic information used by *KK* to develop their dataset appears to have been of uneven quality. Perhaps as a result, several errors in the coding of their linguistic variables have come to light since the publication of the original paper. Kashima and Kashima (2005) address this issue, presenting an updated and corrected dataset and qualifying their previous conclusions regarding language and culture. However, since the publication of this second paper, additional errors in the coding of the linguistic variables have come to light.[2]

Second, the cultural studies used as the basis of *KK*'s sample appear to have emphasized information from "important" countries, resulting in the over-sampling of large and economically successful countries. In addition, *KK* were unable to find acceptable sources of linguistic information for several less developed countries. As a result, the dataset overlaps poorly with the World Values Survey, an important source of international data on culture and values, and may suffer from selection bias. For example, of the seven countries that *KK* include that have English as their primary languages, only one, South Africa, is currently a developing country. Furthermore, with the exception of South Africa, the English-speaking former colonies included in the sample, the United States, Canada, Ireland, New Zealand, Australia, were all settlement rather than extractive colonies, a distinction that Acemoglu et al. (2001) have shown to have important consequences for institutional development. This raises the question of

[1] Gorodnichenko and Roland (2010), Davis (2014) and Klasing (2013) respectively propose instruments for culture based on genetic distance, rainfall variation, and the cultures of neighboring countries.

[2] Private correspondence, May 14, 2012, with Emiko Kashima indicated additional revisions to the linguistic coding of Hungarian and Turkish.

Please cite this article as: L.S. Davis, F. Abdurazokzoda, Language, culture and institutions: Evidence from a new linguistic dataset, Journal of Comparative Economics (2015), http://dx.doi.org/10.1016/j.jce.2015.10.015