PB310: INDEPENDENT RESEARCH PROJECT

-Study Preregistration-

Study Title: The Effects of Cultural Distance on Labour Force Participation Rate of Immigrants in the United Kingdom.

Lead Researcher: Helena Zagala

Introduction

The economic situation of immigrants represents an important topic for psychologists, economists, and policymakers. Although findings suggest that immigrants make a positive overall fiscal contribution to their host countries (Dustmann & Frattini, 2014), there is a need for further study of the factors that could explain and predict their economic situation. One factor that has recently attracted considerable attention of this field is culture. With the recent support for the relationship between cultural diversity and innovation (Schimmelpfennig et al., 2021), social connectedness and trade flows (Bailey et al., 2021), and cultural attitudes and labour market outcomes (Algan et al, 2012), cultural variables prove to be important factors of consideration for this topic. Still, while some findings emerge, little is still known about the effects of the "cultural clash" between the immigrants' country of birth and the country of migration on their labour force participation. With the recent theoretical advancement on the acculturation and collective brain models supporting the potential for such a relationship, there is a need for exploration of ways that would allow for direct analysis of this issue.

In recent years, Muthukrishna and colleagues proposed a promising new measure for the cross-cultural psychological differences between societies – cultural distance (CF_{ST}). By stressing the importance of frequencies and variation, treating cultural attributes similarly to genetically transmitted traits, CF_{ST} offers a novel method for the identification of cultural groupings and analysis of differences on cultural dimensions. Moving beyond the study of mere mean differences, it offers a robust and theoretically justifiable method for high-resolution cultural analyses (Muthukrishna et al., 2020).

By utilizing this novel measure, this study aims to investigate if the cultural distance can explain the differences in labour force participation of immigrants in the UK. By focusing directly on the economic activity, this research seeks to contribute important insights into the factors predicting migrants' fiscal contribution. As such, it aims to provide additional predictive power for economic analyses, better understanding of barriers to full employment, and new arguments for the introduction of targeted psychological interventions.

Research question: Does the cultural distance between the country of birth and the country of migration predict the labour force participation of immigrants in the UK?

H_1: Cultural distance between the country of birth and the country of migration predicts the labour force participation of immigrants in the UK.

H_0: Cultural distance between the country of birth and the country of migration has no effect on the labour force participation of immigrants in the UK.

Methods and Measures

To answer the proposed research question, the study will use country-level secondary data. The utilized datasets used will be checked for methodological adherence to the guidelines from the 2nd edition of the BPS Code of Human Research Ethics (2021). Before study commencement, the institution's ethics review application will be submitted.

Dependent Variable

The labour force participation rate was derived from the Eurostat 2011 European Census. Eurostat collates the national censuses of over 30 European countries, offering rich, comparable records of socio-demographic statistics on the European population. As the national censuses collect the data on all residents, Eurostat offers a uniquely representative, cross-sectional database. Results of the European Census are publicly available, and there is no restriction on the usage of the aggregated data. Although the data collection methods vary between countries, all censuses are run according to the European statistical legislation. In the UK, the 2011 Census questionnaire was delivered mainly through the post.

For this study, the variables derived from the Eurostat dataset included the residents' current activity status by the country-level place of birth for the UK subset (i.e., the overall number of UK residents born in the certain country who are economically active/non-active/employed/unemployed). No individual-level data was obtained. In this dataset, economic activity is a self-reported measure compatible with the definition from the International Labour Organisation. As the economic activity data is collected only on residents above age 16, the derived aggregated data did not include data on younger participants. Further information on the data collection and variables' definition for the 2011 Census can be found in the Eurostat Quality Report (Office of National Statistics, n.d.), which has been attached to the research compendium. The data was downloaded from the European Commission website and attached to the compendium on the 3rd of November 2021.

Independent Variable

The cultural distance was based on the cultural fixation index (CF_{ST}) derived from the "Cultural Distance" dataset (Muthukrishna et al., 2020). The index is calculated as the ratio of the between-group variance to the within-groups variance on the responses to the chosen questions from the World Value Survey (WVS). The utilized questions concern culturally transmissible values, beliefs, and behaviours, with the CF_{ST} subdimensions utilizing the responses on topics ranging from financial principles to sexuality values. Importantly, contrarily to the measurement of genetic or linguistic distance, the cultural fixation index can be taken as a direct measure of culture and does not face the limitation of estimating cultural values from related proxies. The index has been shown to be highly robust even to question loss and is an effective measure of cultural differences beyond WEIRD (Western, Educated, Industrialized, Rich and Democratic) societies (Muthukrishna et al., 2020). The dataset itself is publicly available; more information on the index methodology and the variables included can be found on the Culturalytics website.

WVS, the source of the data used in the cultural distance database, is the largest non-commercial social science survey. As the WVS collects data on over 100 countries, its results can be taken as representative of over 94% of the world population. The survey waves are run every five years across all residents aged 18 years and older, with a minimum sample size of 1200 for each country. The survey utilizes probability sampling, ensuring nationally representative results. Where possible, the data is collected during a face-to-face interview. In some cases, responses are recorded using postal or telephone interviews (World Value Survey Association, 2020). Depending on the type of question, the WVS utilizes, among others, Likert-type, binary and categorical scales. The full questionnaires for the WVS waves underlying the derived variables are included in the appendices.

For this study, the cultural distance data was restricted to the CF_{ST} calculated based on the 2005-2009 and 2010-2014 WVS waves (Inglehart et al., 2018a; Inglehart et al., 2018b). This represented the most recent grouping for which the cultural distance for the United Kingdom was calculated. The data was downloaded directly from the Culturalytics website on the 18th of November 2021.

Control Variables

Migrant Population Size

Migrant population size was included as a control variable to adjust for the size of the migrant community (i.e., number of UK residents born in the given country). The increased migrant community size may lead to network effects, in which the utility of migrants increases with the increased size of their ethnic groups (e.g., through the accessibility of social capital and financial support, community relationships, lower language barriers, availability of guidance in the job search process, facilitated flow of information, conformity with the group rules, affinity bias etc.). The importance of this control is supported by findings that personal networks and/or ethnic concentrations have a significant effect on migrants' labour market integration (Maani et al., 2015) and occupational prestige (Mullan, 1989).

The migrant population size was derived from the Eurostat 2011 European Census and reflects the number of UK residents who were born in the given country. This variable is a part of the same dataset that was used as a source of the labour force participation; as such, the details of the data representativeness, collection method and access date match the ones described in the Dependent Variable section.

Genetic Distance

The genetic distance was controlled for to ensure that the genetic differences do not shrink the true effect sizes for the cultural distance. In the dataset, genetic distance represents the differences in frequencies of alleles in the investigated populations (i.e., how frequencies differ from the ones' one should observe with truly random mate selection). As such, it can be seen as a measure of similarity in transmitted characteristics, and it is closely related to the time since populations' last common ancestor. By including the variable in the model, we can investigate the effects of culture beyond the effects of genes. The importance of this control is related to its real-life significance in explaining genotype – and thus phenotype – differences (i.e., allows us to focus on cultural distance, moving beyond the potential effect of physical attributes/visible cues that might lead to discrimination). Moreover, as the genetic distance itself is predictive of economic outcomes (Saha & Mishra, 2020), its inclusion in the model removes its potential for yielding the third variable problem. The importance of this control is proved by the significant correlation between the genetic distance and income differences across time and populations, which is hypothesised to be caused by its negative effect on the diffusion of innovation and spread of development (Spolaore & Wacziarg, 2009).

The utilized genetic distance measure was calculated based on the human microsatellite variation (i.e., variance in the parts of DNA that have high diversity and mutation rates, which allows for analysis of more nuanced differences). The country-level distance had been obtained by matching the populations based on ethnic composition data. The details on the calculation involved can be found in Spolaore and Wacziarg (2017). This approach to genetic distance yields a measure that is highly correlated with other metrics of genetic relatedness, supporting its high reliability. The variable was downloaded and added to the compendium on the 3rd of November 2021.

Note

The remaining control variables were derived from the Varieties of Democracy dataset (V-Dem). V-Dem provides statistics on important socio-economic indicators on countries around the world, compiling publicly available data from numerous databases. All variables derived from V-Dem have been restricted to 2011 scores, to match the date of collection of the dependent variable. The whole V-Dem database has been downloaded on the 18th of November 2021, with the whole dataset, the methodology booklet and codebooks being attached to the compendium. As these procedures remained constant across variables, the following descriptions will focus solely on the motivation for this choice of controls, codebooks' identifiers for the chosen variables, variables' definitions, and their original sources.

GDP Per Capita

GDP per capita of the country of birth was included as a control variable to adjust for the between-countries differences in productivity, residents' approximate level of socioeconomic status from before migration, and potential financial support received from the country of birth. The GDP variable (*e_migdppcln* in the V-Dem dataset) was derived from the 2nd edition of the Maddison Project Database and represents the gross domestic product per capita transformed by the natural logarithm. This variable is based on benchmark comparisons of prices and is expressed in inflation-adjusted US dollars (Inklaar et al., 2018).

Education

Education was included as a control variable to adjust for the level of and approach to education in migrants, and the effect of education on the attitude towards host society and ease of acculturation (De Vroome et al., 2014). As no direct data on the level of migrants' education was available, the level of education in the country of birth was included as its proxy. The education variable is available in the V-Dem dataset as *e_peaveduc* and is originally derived from the dataset offered by the Clio Infra project (2021). The variable represents the average years of education among citizens older than 15 in each country. Although this dataset derives the year-specific data based on cross-analysis of other databases, rather than the primary data collections, all data for the 2010 benchmark was derived from the Central Statistical Agencies, supporting high reliability of the 2011 scores.

Population Size

The population size of the country of birth was included in the analysis to control for the effects of the increased number of potential migrants (as the size of the pool of potential migrants might affect the profile of the candidates who ultimately manage to migrate, leading to potential candidate selection and unobserved clustering). It also allows to control for migrants' familiarity with novel cultures – as bigger countries tend to represent an assembly of smaller communities, they may make migrants more accustomed to cultural differences and differing populations.

The total population size (identified as *e_wb_pop* in the V-Dem codebook) comes from the World Bank Development Indicators (WBDI; World Bank, 2019). WBDI offers a representative, compiled source of global development data on national, regional, and global levels. The offered population size variable is derived from official sources, including national censuses, United Nations reports and other international databases. Although the data quality may depend on the statistical systems of participating countries, WBDI can be still regarded as an exceptionally reliable database for such a metric.

Regional Identifier

The regional identifier was included as a control variable to adjust for the influence of geographical factors on immigrants' situation and decrease the probability of occurrence of the Simpson's paradox (i.e., to ensure that potential group clustering does not lead to misleading results by concealing the true trend). The variable reflects a categorical score for the geographic region in which a country is located (e.g., West Asia, Southern Europe etc.). Identified as *e_regiongeo* in the V-Dem, it has its source in the database from the United Nations Statistics Division (2013). As this level of clustering might have been too exclusive for this research, which might have led to a loss of some of the between-country variation, the variable was further transformed into continent-level categorization.

Confirmatory Hypotheses and Analysis Plan

The relationship between cultural distance and labour force participation will be analysed using multiple regression analysis. The model will include one dependent (labour force participation), one independent (cultural distance), and six control variables (GDP per capita distance, genetic distance, migrant population size, education distance, population size distance, regional identifier).

The labour force participation rate will be calculated by dividing the number of employed respondents by the number of to economically active respondents for each country.

The education, GDP per capita and population size variable will be transformed to education, GDP and population distance by subtracting the country score from the UK score. The regional identifier will be transformed from regional to continent level by clustering the groups, as described in the previous section.

The dataset will be checked for the occurrence and distribution of missing data using the MCAR test from Jamshidian and colleagues (2014). As recommended by Cole (2008), the cases with more than 10% of data missing will be removed from the analysis. Depending on the distribution of the data missingness, the missing cases will be replaced using either scale mean imputation, or multiple imputation, in line with the guidelines from Schafer and Graham (2002). The potential univariate and multivariate outliers will be identified and removed according to the guidelines of Tabachnick and Fidell (2013), and Osborne (2013).

Following the data cleaning procedures, the linear model with the beforementioned variables will be built. According to the guidelines from Best and Wolf (2014), the model will be checked for assumptions of linearity, homoscedasticity, normality of residuals, autocorrelation, and multicollinearity. The significance of the main effects will be further checked with 5000 bootstrapped resamples, as per the recommendations from Banjanovic and Osborne (2016). This approach was chosen due to its effectiveness in controlling for Type 1 error. If the confirmatory analysis finds a significant relationship before the main variables, the effect will be tested for robustness by multiverse analysis.

As the study utilized secondary data, with no direct control over sample sizes, an a-priori power analysis will be conducted. However, given that the analysis included several controls, and utilized large, representative datasets, the data should provide sufficient power to detect true effects. Still, to ensure that the study has met the recommended 80% statistical power, a post-hoc power analysis will be performed (Cohen, 1988)

The confirmatory analysis follows the pilot analyses run by Michael Muthukrishna in 2019. The results were never officially published, did not include data preprocessing, and were not extended by any further exploratory analyses. To ensure full attribution, the whole analysis run by Michael Muthukrishna is attached to the compendium in the Analysis folder.

Exploratory Analyses

With culture being a complex, multifaceted phenomenon, the further exploratory analysis will focus on investigating the relative importance of the different dimensions of culture (e.g., beliefs on law, social relations, financial values) on labour force participation rate. Such analysis will not involve additional datasets, only requiring access to supplementary variables from the Cultural Distance database (i.e., subdimensions of cultural distance). The relative importance of the dimensions will be checked using elastic net regression.

Further exploratory tests will extend the analysis beyond the migrant population in the UK, investigating if the potential relation between cultural distance and labour force participation remains true across countries. If so, it will be tested if the between-countries differences in the effect of cultural difference on labour force participation can be explained by between-countries differences in social expenditure. In the first step, this analysis will involve following identical steps as the confirmatory analysis, i.e., checking the relation between cultural distance and labour force participation in countries other than the UK. This will require downloading country-specific data from the beforementioned sources and will not involve additional databases. In the next steps, between-country differences in social expenditure will be used to try to explain the potential between-country differences in the main effect. As social spending can be taken as a proxy of financial support from the government, this analysis will ultimately aim to investigate if the availability of financial aids for migrants increases the chance of entering the labour force in spite of cultural differences.

The social expenditure data will be derived from the publicly available Social Expenditure Database (SOCX; OECD, 2021). To control for differences in countries wealth, the social expenditure will be defined as the percentage of GDP allocated to public social spending.

References

Algan, Y., Bisin, A., & Verdier, T. (2012). Introduction: perspectives on cultural integration of immigrants. *Cultural integration of immigrants in Europe*, 1-48.

Bailey, M., Gupta, A., Hillenbrand, S., Kuchler, T., Richmond, R., & Stroebel, J. (2021). International trade and social connectedness. *Journal of International Economics*, *129*, 103418.

Banjanovic, E. S., & Osborne, J. W. (2016). Confidence intervals for effect sizes: Applying bootstrap resampling. *Practical Assessment, Research, and Evaluation*, *21*(1), 5.

Best, H., & Wolf, C. (2014). Regression Analysis: Assumptions and Diagnostics. In H. Best & C. Wolf (Eds), *The SAGE Handbook of Regression Analysis and Causal Inference*. (83-110). https://doi.org/10.4135/9781446288146

British Psychological Society. (2021). Code of Human Research Ethics. Retrieved from: https://www.bps.org.uk/news- and-policy/bps-code-human-research-ethics-2nd-edition-2014.

Clio-Infra (2021). *Clio-Infra Project*. Database. Available at: www.clio-infra.eu.

Cohen, J. (1988) *Statistical Power Analysis for the Behavioral Sciences*, 2nd edn. Lawrence Erlbaum, New Jersey.

Cole, J. (2008). How to deal with missing data conceptual overview and details for implementing two modern methods. In Osborne, J. (Ed.), *Best practices in quantitative methods* (pp. 214-238). SAGE Publications, Inc., https://www.doi.org/10.4135/9781412995627

Coppedge, M., Gerring, J., Knutsen, C. H., Lindberg, S. I., Teorell, J., Marquardt, K. L., ... & Wilson, S. (2021). V-Dem Methodology v11.1". Varieties of Democracy (V-Dem) Project.

Coppedge, M., Gerring, J., Knutsen, C. H., Lindberg, S. I., Teorell, J., Altman, D., ... & Ziblatt, D. (2021). "V-Dem Codebook v11.1" Varieties of Democracy (V-Dem) Project.

Coppedge, M., Gerring, J., Knutsen, C. H., Lindberg, S. I., Teorell, J., Altman, D., ... & Ziblatt, D. (2021). V-Dem Dataset v11.1"Varieties of Democracy (V-Dem) Project. https://doi. org/10.23696/vdemds21.

De Vroome, T., Martinovic, B., & Verkuyten, M. (2014). The integration paradox: Level of education and immigrants' attitudes towards natives and the host society. *Cultural Diversity and Ethnic Minority Psychology*, *20*(2), 166.

Didenko, D., Földvári, P. & van Leeuwen B. (2012). A Dataset on Human Capital in the Former Soviet Union Area: Sources, Methods, and First Results. Working Paper 0035. Utrecht: Utrecht University.

Dustmann, C., & Frattini, T. (2014). The fiscal effects of immigration to the UK. *The economic journal*, *124*(580), F593-F643

Földvári, P., & van Leeuwen, B. (2014). Educational and Income Inequality in Europe, ca. 1870–2000'. *Clio-metrica*, 8(3): 271–300. https://doi.org/10.1007/s11698-013-0105-3

Inglehart, R., Haerpfer, C., Moreno, A., Welzel, C., Kizilova, K., Diez-Medrano, J., ... & Puranen, B. (2018a). World values survey: Round Five - Country-Pooled Datafile. *Madrid: JD Systems Institute*, 12. doi.org/10.14281/18241.7

Inglehart, R., Haerpfer, C., Moreno, A., Welzel, C., Kizilova, K., Diez-Medrano, J., ... & Puranen, B. (2018b). World values survey: Round six-country-pooled datafile version. *Madrid: JD Systems Institute*, 12. doi.org/10.14281/18241.8

Inklaar, R., de Jong, H, Bolt, J., & van Zanden, J, (2018). Rebasing 'Maddison': new income comparisons and the shape of long-run economic development. GGDC Research

Memorandum GD-174, Groningen Growth and Development Centre, University of Groningen.

Jamshidian, M., Jalal, S., & Jansen, C. (2014). MissMech: An R Package for Testing Homoscedasticity, Multivariate Normality, and Missing Completely at Random (MCAR). *Journal of Statistical Software, 56*(6), 1 - 31. doi:http://dx.doi.org/10.18637/jss.v056.i06

Maani, S. A., Wang, X., & Rogers, A. (2015). Network Effects, Ethnic Capital and Immigrants' Earnings Assimilation: Evidence from a Spatial, Hausman-Taylor Estimation. Chicago

Mitchell, B. (1998a). *International Historical Statistics: Africa, Asia and Oceania, 1750–1993*, 3rd edition. Basingstoke: Macmillan.

Mitchell, B. (1998b). *International Historical Statistics: Europe, 1750–1993*, 4th edition. Basingstoke: Macmillan.

Mitchell, B. (1998c). *International Historical Statistics: The Americas 1750–1993*, 4th edition. Basingstoke: Macmillan.

Mullan, B. P. (1989). The impact of social networks on the occupational status of migrants. *International migration (Geneva, Switzerland)*, *27*(1), 69-86.

Muthukrishna, M., Bell, A. V., Henrich, J., Curtin, C. M., Gedranovich, A., McInerney, J., & Thue, B. (2020). Beyond Western, Educated, Industrial, Rich, and Democratic (WEIRD) Psychology: Measuring and Mapping Scales of Cultural and Psychological Distance. *Psychological Science*, *31*(6), 678–701. https://doi.org/10.1177/0956797620916782

OECD (2021), Social Expenditure Database. https://www.oecd.org/social/expenditure.htm

Office of National Statistics (n.d.). Eurostat Quality Report. Retrived from: https://www.ons.gov.uk/file?uri=/census/2011census/2011ukcensuses/ukandeuropecensuscomparison/variables/eurostatqualityreport03tcm77358400.pdf

Osborne, J. W. (2013*). Best practices in data cleaning: A complete guide to everything you need to do before and after collecting your data*. Sage. Chicago

Saha, A. K., & Mishra, V. (2020). Genetic distance, economic growth and top income shares: evidence from OECD countries. *Economic Modelling*, *92*, 37-47.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7*(2), 147–177. https://doi.org/10.1037/1082- 989X.7.2.147

Schimmelpfennig, R., Razek, L., Schnell, E., & Muthukrishna, M. (2021). Paradox of diversity in the collective brain. *Philosophical Transactions of the Royal Society B*, *377*(1843), 20200316.

Spolaore, E. & Wacziarg, R. (2017). Ancestry and Development: New Evidence

Spolaore, E., & Wacziarg, R. (2009). The diffusion of development. *The Quarterly journal of economics*, *124*(2), 469-529.

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics*. Boston: Pearson Education.

UNESCO (n.d.). *Statistical Yearbook (1963–1999)*. Paris: UNESCO.

United Nations Statistics Division, (2013). Methodology. Retrived from: *https://unstats.un.org/unsd/methodology/*

van Leeuwen, B., van Leeuwen-Li, J. & Földvári, P. (2011). Regional Human Capital in Republican and New China: Its Spread, Quality and Effects on Economic Growth. MPRA Working Paper 43582. Munich: University Library of Munich.

van Leeuwen, B., van Leeuwen-Li, J. & Földvári, P. (2012). Education as a Driver of Income Inequality in Twentieth-Century Africa. MPRA Working Paper 43574. Munich: University Library of Munich.

World Bank. (2019). World Bank Development Indicators. https://databank.worldbank.org/source/world-development-indicators

World Value Survey Association. (2020). World Value Survey: Fieldwork and Sampling https://www.worldvaluessurvey.org/wvs.jsp