Playing the birth lottery in Europe

Annaelena Valentini[a], Paolo Brunori[b,c], Francisco Ferreira[b], Pedro Salas-Rojo[b]

[a]University of Siena [b]III - London School of Economics [c]University of Florence

February, 2024

Our idea

The Inequality of Opportunity (IOp) framework often focuses on a set of inherited factors (parental background) plus sex, ethnicity,...

Most of these studies study IOp "within countries". What if we widen the focus?

We explore IOp taking Europe as a whole region, using traditional inherited factors **and** the country of origin.

Ex-ante approach to IOp (Roemer (1998), Van De Gaer (1993))

Define an outcome y and a vector c denoting *circumstances*. *Types* are groups of individuals sharing circumstances. Equality of Opportunity (EOp) is:

$$\text{EOp} \iff \forall(c, c'), \bar{y}|c = \bar{y}|c' \tag{1}$$

Estimating $\hat{y}_i$ as $\bar{y}$ for all types, and using a suitable inequality measure $I()$, then:

$$\text{IOp} = I(\hat{y}_i) \tag{2}$$

Data

Three data sources: EUSILC (EU-27, excluding Romania and Slovenia, including Norway, and Switzerland), SOEP (Germany), AND UKUS.

The wave of interest: 2019

Outcome: age-adjusted equivalised disposable household income in 2019 USD (PPP).

Circumstances at age 14: country of origin (60+17), father's and mother's education (3), father's and mother's occupation (11), sex (2).

Data: Country of Origin

We use EU-SILC information to separate natives from foreign-born.

For the latter group, we use the country of origin of the mother (available in the 2019 EUSILC wave).

- May bear mistakes.
- People may choose to migrate, but the country of origin bears much more information.
- Relevant cultural aspects (religion, language, social norms) transmitted from mothers to children (Caneva and Pozzi, 2014).

Data: Setting Europe as the region of interest

First, we pool surveys. Then use census data (EUROSTAT) to construct weights such that we:

Draw a random sample representative of population shares of residents.

Within each country, samples are representative of relative shares of countries of origin (up to 0.1% of the country's population).

N: Around 138,268 observations.

Conditional Inference Trees and Random Forests (Hothorn et al. (2006))

Recent contributions (Brunori et al., 2023b) show these algorithms suitable to estimate ex-ante IOp. Two main steps:

- Correlation test for each C and y, for some Bonferroni-adjusted p-value.
- After the most correlated C is selected, mean-test across all possibly binary partitions.

Tree: just one (main) sample

Forest: 200 random samples using the above proportions.
Issue: To make the problem finite, the split of unordered categorical factors in the second step is capped at 30 categories.

Methods: our trick (checked with Horthorn, works!)

In every node, we re-order the values of the unordered categorical C by mean y. Then:

- Run correlation test.
- If C (re-ordered) is selected, then use it as ordered categorical.

The idea: After each partition, categories will always be grouped according to their expected outcome being smaller or higher than the splitting point!

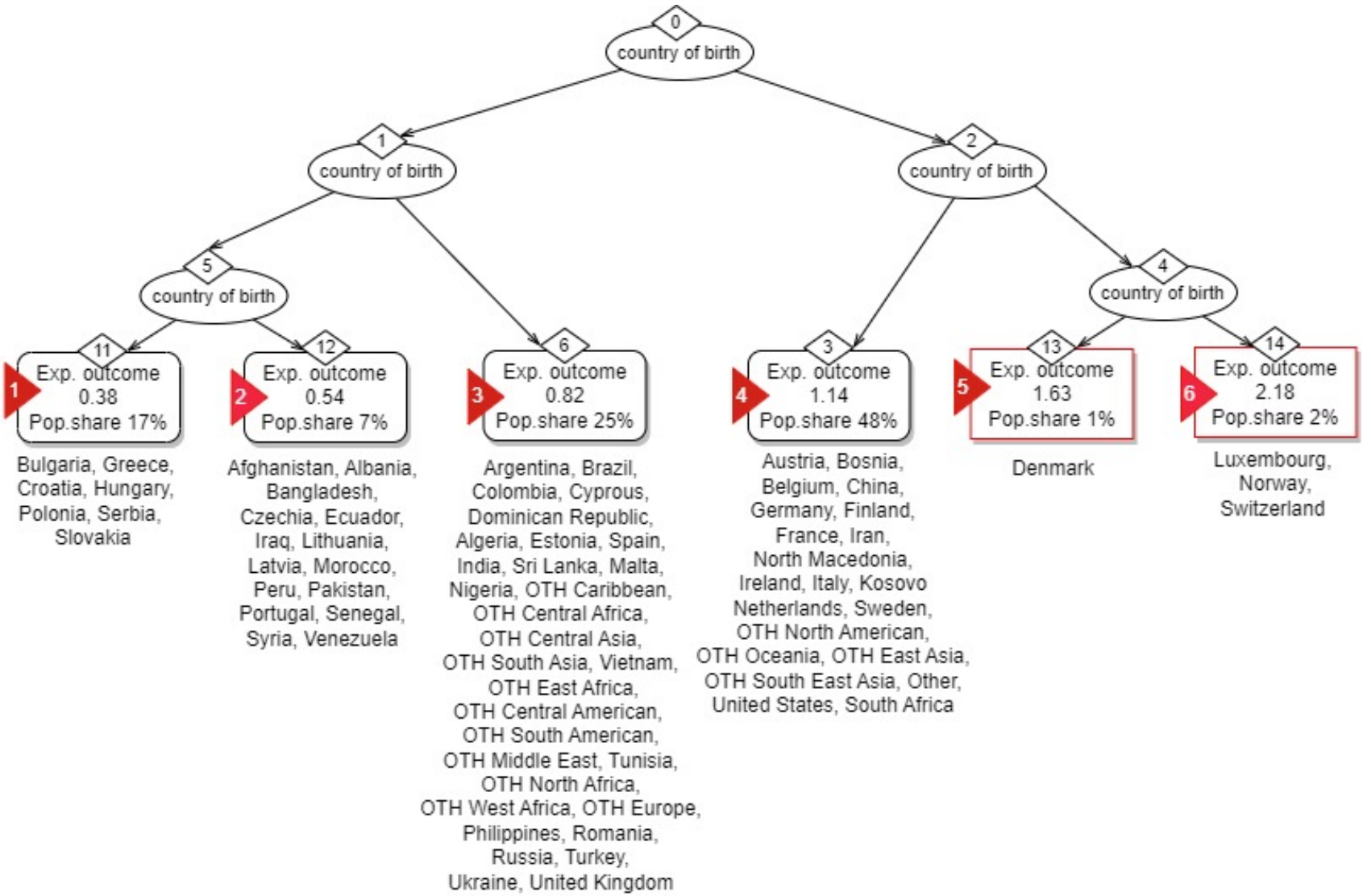
Results: IOp (Gini index)

Overall: I(y); IOp: $I(\hat{y})$; Rel. IOp: $\hat{y}/I(y)$

Overall	IOp (tree)	Rel. IOp (tree)	IOp (forest)	Rel. IOp (forest)
0.39	0.24	0.62	0.22	0.56

Source: Own elaboration, data from EU-SILC, UKUS, GSOEP. The tree is estimated with $\alpha = 0.01$, and *minbucket* $= 100$. The random forest is estimated with $\alpha = 0$, *minbucket* $= 50$, and 200 bagging subsamples of 0.1 the original sample size.

Results: First splits in the main tree



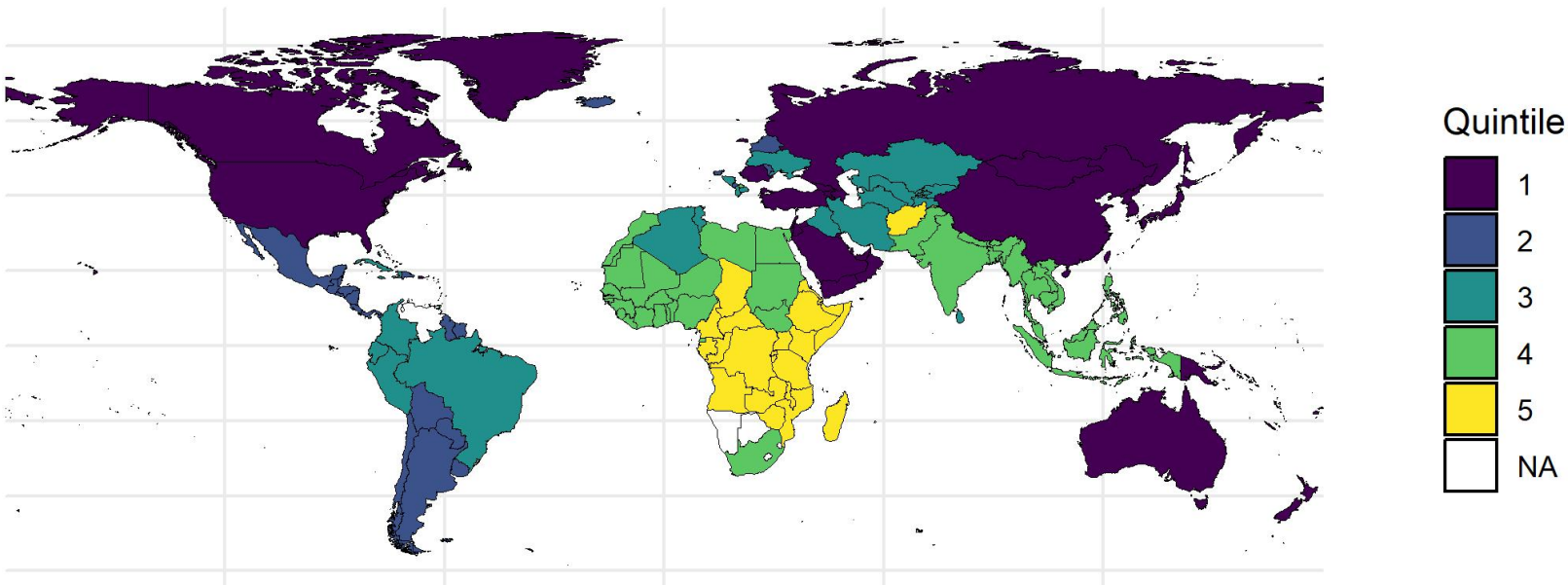
Results: Shapley Value Decomposition (as in Brunori et al. (2023a))

Circumstance	Contribution to IOp (%)
Country of origin	62.4
Father's occupation	18.2
Mother's occupation	8.8
Father's education	5.7
Mother's education	3.8
Sex	1.3

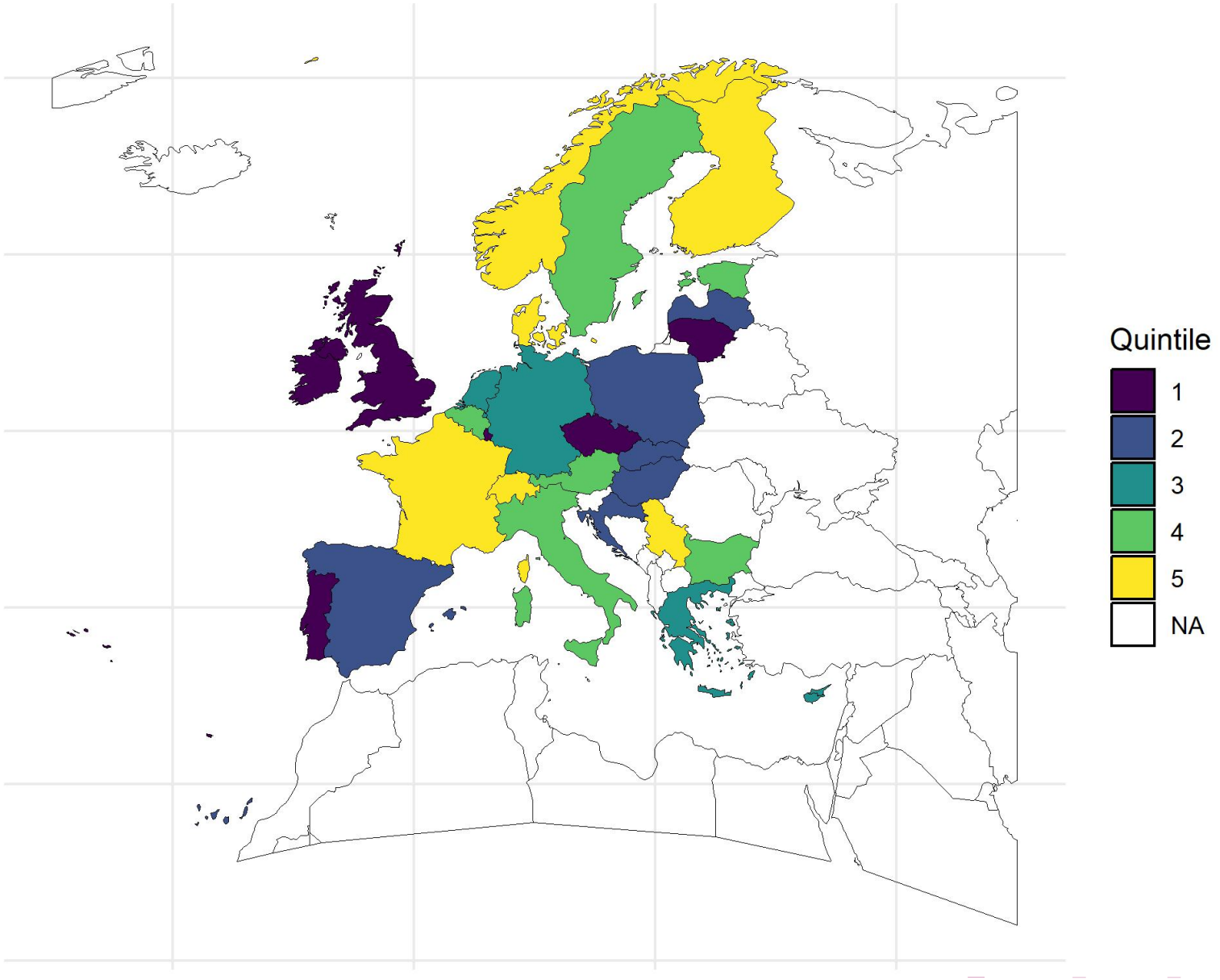
Source: Own elaboration, data from EU-SILC, UKUS, GSOEP. The Shapley value decomposition is estimated with a random forest procedure, with $\alpha = 0.1$, *minbucket* $= 50$ and 100 bagged repetitions sized 0.1 the original sample size.

Results: Migration Premium (I)

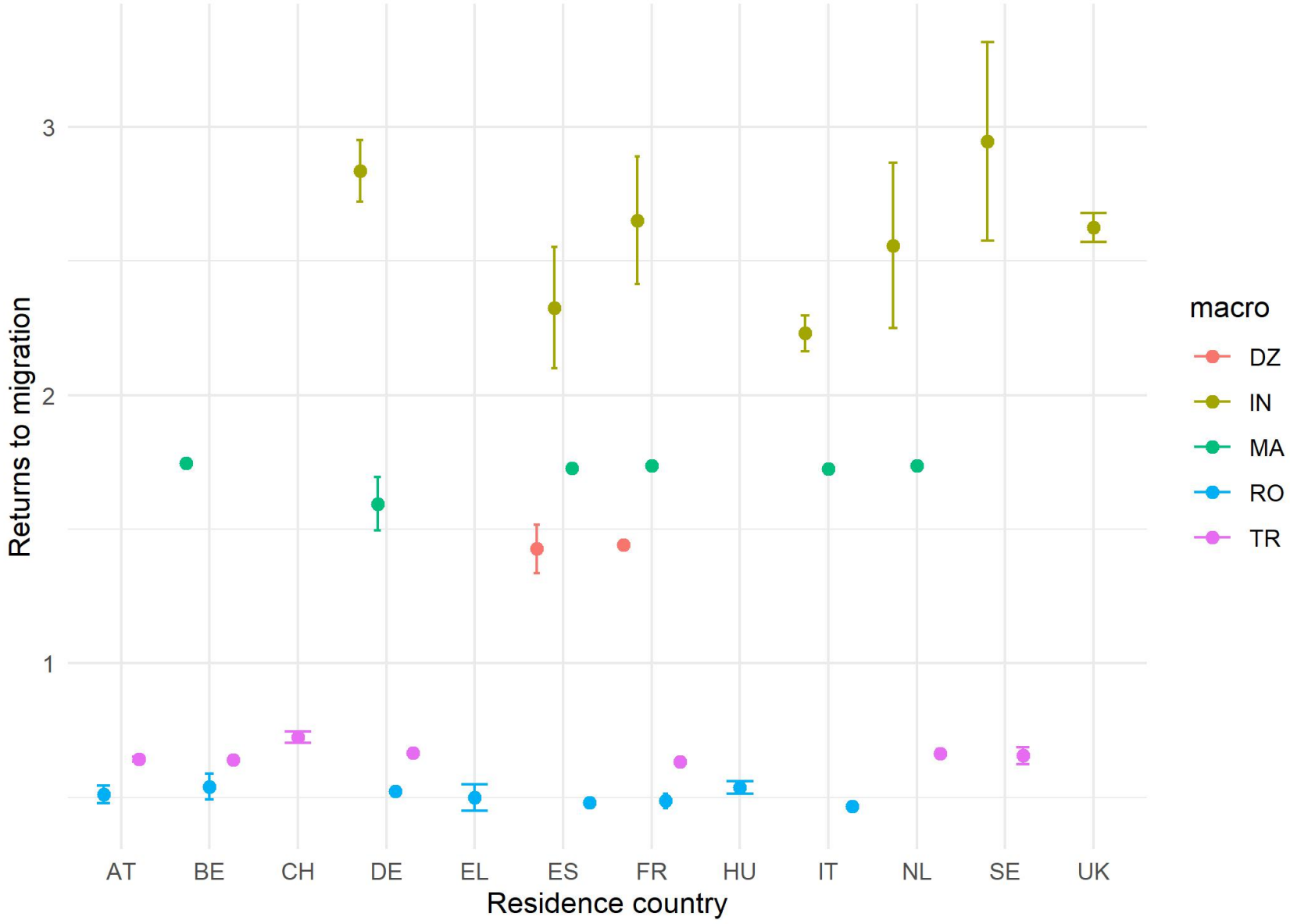
Very simple exercise: $y_{ret} = \hat{y}_c / GDPpc_c$



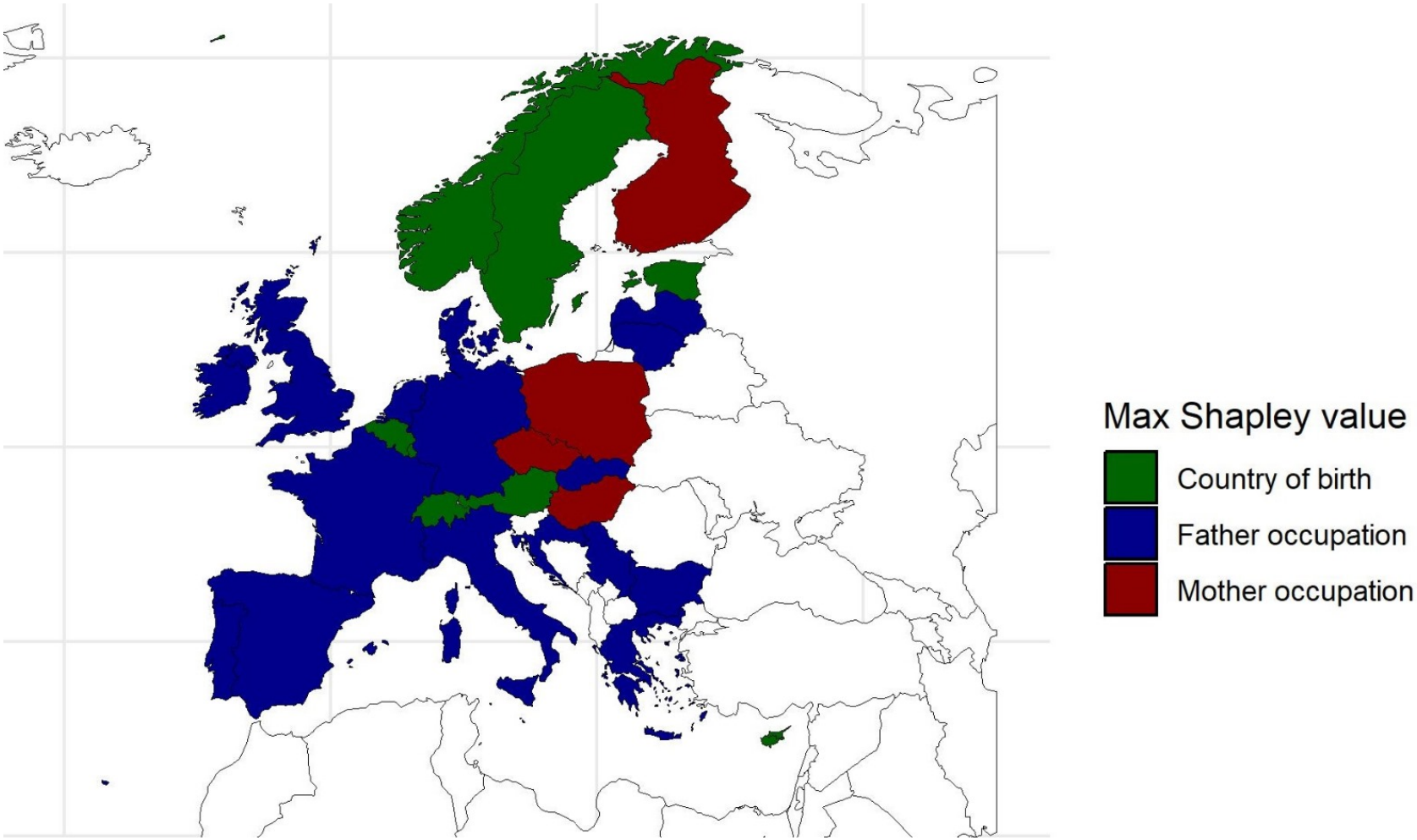
Results: Migration Premium (II)



Results: Migration Premium (III)



Results: Heterogeneity in the DGP: Main circumstance from country-specific Shapley



Final remarks

There is always a way to trick an algorithm!

IOp in Europe, as a region, is huge: 22 Gini points, around 56% of overall inequality. Far from convergence.

Country of origin explains a remarkable share (around 62% of IOp)

Migration premium is quite heterogeneous across countries:

- By regions: SSA and South-Eastern Asia are especially benefited.
- In Europe: the Nordic and central (France, Italy, Switzerland,...) also get relatively more.

Comments to: p.salas-rojo@lse.ac.uk

Playing the birth lottery in Europe

Annaelena Valentini[a], Paolo Brunori[b,c], Francisco Ferreira[b], Pedro Salas-Rojo[b]

[a]University of Siena [b]III - London School of Economics [c]University of Florence

February, 2024

References I

Brunori, P., Ferreira, F. H., and Salas-Rojo, P. (2023a). Inherited inequality: a general framework and an application to south africa.

Brunori, P., Hufe, P., and Mahler, D. (2023b). The roots of inequality: Estimating inequality of opportunity from regression trees and forests. *The Scandinavian Journal of Economics*, 125(4):900–932.

Caneva, E. and Pozzi, S. (2014). The transmission of language and religion in immigrant families: a comparison between mothers and children. *International Review of Sociology*, 24(3):436–449.

Hothorn, T., Hornik, K., and Zeileis, A. (2006). Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical statistics*, 15(3):651–674.

Roemer, J. E. (1998). Equality of opportunity.

References II

Van De Gaer, D. (1993). Equality of opportunity and investment in human capital.