



**CONNECT
TO LEARN!**

შეუერთდი სწავლას!

The invisible tax of free knowledge

Evidence from the Wikimedia projects in the CEE region

Kiril Simeonovski



Introduction

- Free knowledge is commonly thought as, but in practice is not a pure public good.
 - while it is perfectly non-rivalrous, its non-excludability cannot be always achieved
 - in reality, it is an example of an impure public good

Introduction

- Free knowledge is commonly thought as, but in practice is not a pure public good.
 - while it is perfectly non-rivalrous, its non-excludability cannot be always achieved
 - in reality, it is an example of an impure public good
- Excludability results from limitations of its access, and it is driven by economic, institutional and social factors.
 - less people can consume it
 - less people can produce it

} economic inefficiency

Research questions

- Why is free knowledge not a pure public good?
- What are the implications of the impurity?
- How to measure the implications of the impurity?
- How big are these implications in the CEE region?

Research outline

- Definition of pure public good vs impure public good.
- Model of free knowledge as a public good.
 - ... relies on peer production in the Wikimedia movement
 - ... uses Wikimedia content as a proxy of free knowledge
- Introduction of the concept of “invisible tax”.
- Calibration using data per country and language from the Wikimedia projects in Central and Eastern Europe.
 - stylised facts on editing across countries and languages
 - invisible tax rates for countries and languages

Wikimedia vision

Imagine a world in which every single person on the planet is given free access to the sum of all human knowledge. That's what we're doing.

— Jimmy Wales

Wikimedia vision

Imagine a world in which every single person on the planet is given free access to the sum of all human knowledge. That's what we're doing.

— Jimmy Wales



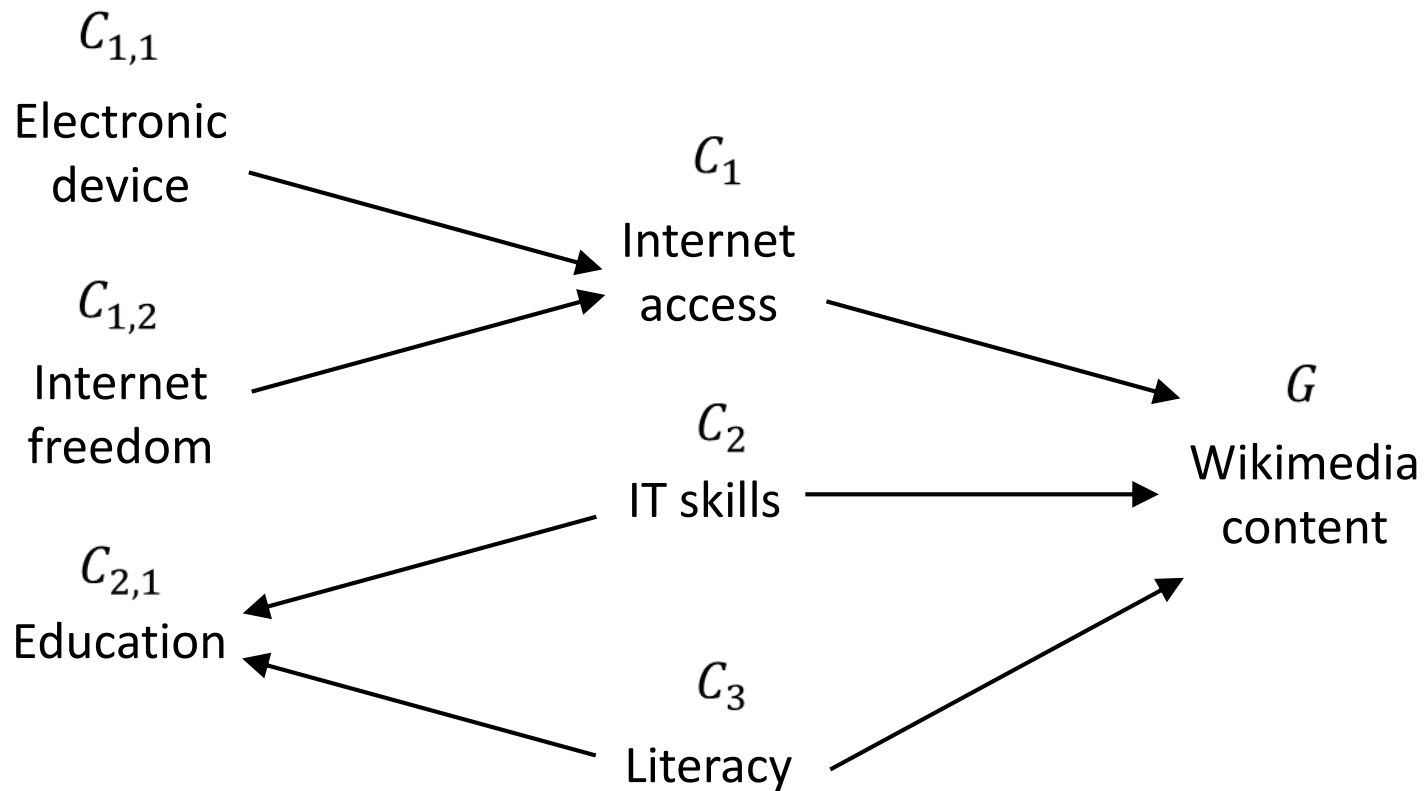
The ultimate goal is to make the content on the Wikimedia projects a pure public good.

Pure vs impure public good

- A good G usually has a complementary good C so that its excludability $\eta = \eta(C, p)$ is an increasing function of the state of their complementary goods C and its price p .
- Definition: A complementary good C is
 - ... public good if every one can afford to pay to consume it
 - ... private good if some people cannot afford to pay to consume it
- Definition: A good G is
 - ... pure public good if it is free of charge and all complementary goods C are public goods (almost impossible in practice)
 - ... impure public good if it is free of charge but some complementary goods C are private goods

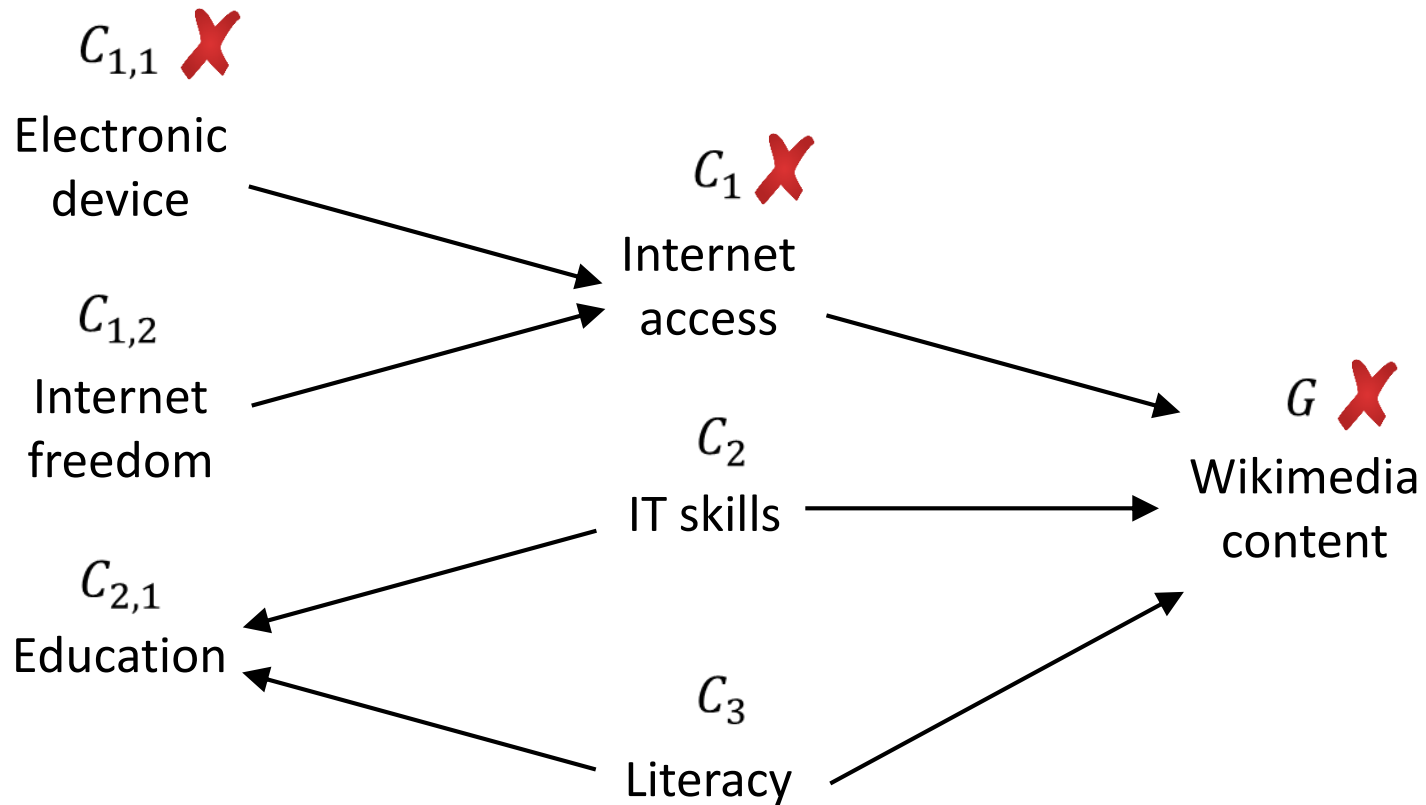
Pure vs impure public good

Example



Pure vs impure public good

Example



Free knowledge market

- Individuals derive utility from
 - consumption of free knowledge $u(v_{i,t}, G)$
 - contribution to free knowledge $S(w_{i,t}, G, F_{i,t})$
- Utility of consumption $u(v_{i,t}, G)$ increases as the leisure time spent for consuming free knowledge $v_{i,t}$ increases and the amount of available free knowledge G increases.
- Utility of contribution $S(w_{i,t}, G, F_{i,t})$ increases as the leisure time spent for contributing $w_{i,t}$ increases, decreases as the amount of available free knowledge G increases and may go in both directions depending on the social interactions with others $F_{i,t}$.
- Leisure time spent for consuming $v_{i,t}$ and contributing to free knowledge $w_{i,t}$ make up individual's total leisure time.

Free knowledge market

- Aggregate demand on the free knowledge market AD_t is the marginal benefit of the utility of all individuals.
- Aggregate supply on the free knowledge market AS_t is the total amount of available free knowledge.
- The market equilibrium is where aggregate demand meets aggregate supply ($AD_t = AS_t$).

Free knowledge market

- Aggregate demand on the free knowledge market AD_t is the marginal benefit of the utility of all individuals.
- Aggregate supply on the free knowledge market AS_t is the total amount of available free knowledge.
- The market equilibrium is where aggregate demand meets aggregate supply ($AD_t = AS_t$).
- But free knowledge is an impure public good and not all individuals have access to it.
- Implications:
 - Less individuals derive utility ($AD_t^Z < AD_t$).
 - Less individuals are likely to contribute ($AS_t^Z < AS_t$).
 - The market equilibrium will be reached at lower level ($AD_t^Z = AS_t^Z$).

Invisible tax of free knowledge

- Definition: The invisible tax reflects the lower supply of free knowledge as a result of excludability and rivalry, and it can be calculated as

$$\tau_t = \frac{G_t^Q}{G_t}$$

where $G_t^Q = G_t - G_t^Z$ is the lower supply of free knowledge.

Invisible tax of free knowledge

- Definition: The invisible tax reflects the lower supply of free knowledge as a result of excludability and rivalry, and it can be calculated as

$$\tau_t = \frac{G_t^Q}{G_t}$$

where $G_t^Q = G_t - G_t^Z$ is the lower supply of free knowledge.

- Why to call it an “invisible tax”?

Invisible tax of free knowledge

- Definition: The invisible tax reflects the lower supply of free knowledge as a result of excludability and rivalry, and it can be calculated as

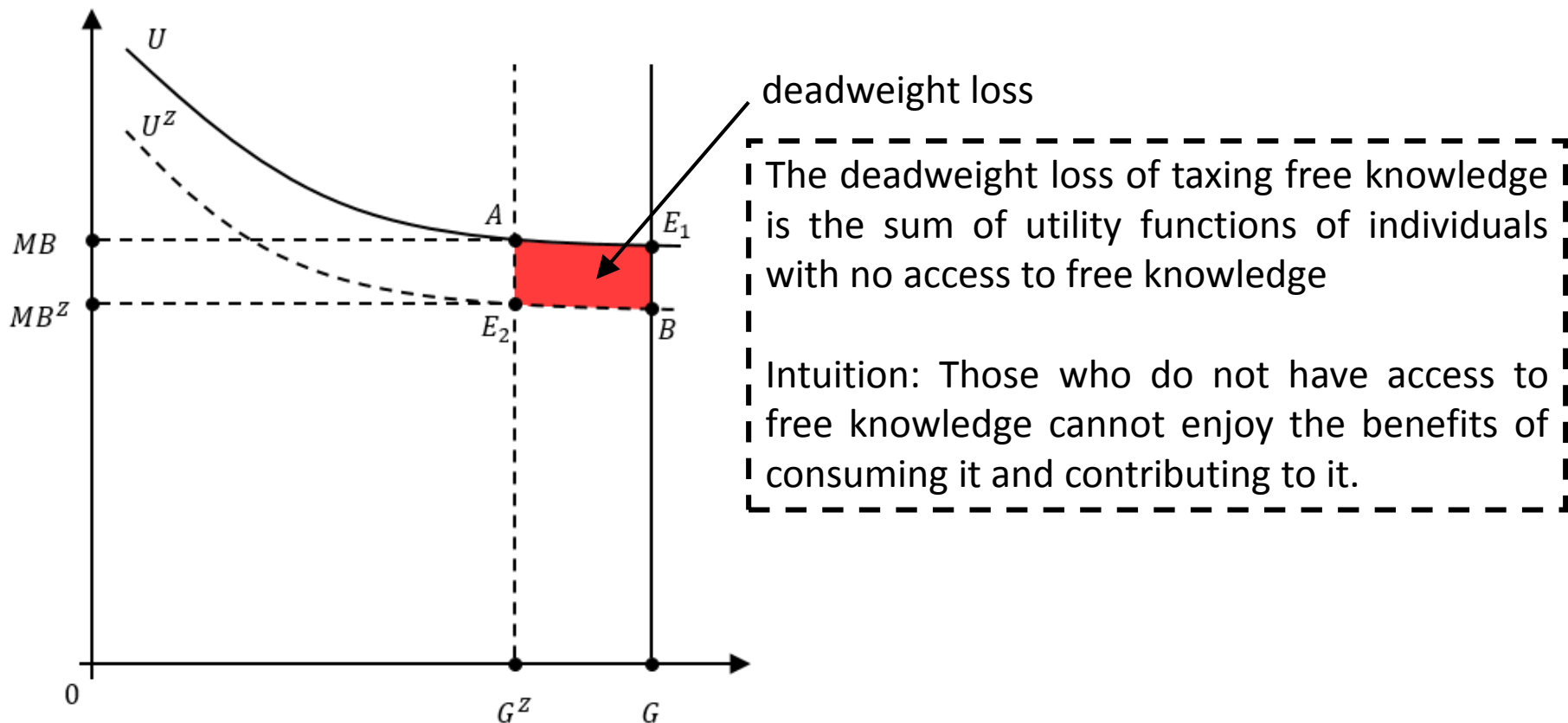
$$\tau_t = \frac{G_t^Q}{G_t}$$

where $G_t^Q = G_t - G_t^Z$ is the lower supply of free knowledge.

- Why to call it an “invisible tax”?
 - in public economics, a tax is an amount levied to support production and provision of public goods
 - in microeconomics, a tax is a source of economic inefficiency, which results in lower supply and demand (deadweight loss)
 - it is invisible because there is no monetary payment

Invisible tax of free knowledge

- Supply and demand shifts as a result of taxing free knowledge.



Calibration

- Page views —————> consumption of free knowledge (aggregate demand)
- Page edits —————> production of free knowledge (aggregate supply)
- Data on page views and page edits across states/territories were obtained from the Wikimedia Foundation's databases.
 - Missing data on page edits for many states and territories, including Azerbaijan, Belarus, Czech Republic, Kazakhstan, Russia and Turkey.
- As a result of the missing data on page edits, the calibration was not done for some languages, including Russian, Turkish, Czech and others.

Calibration

- Annual elasticities of page edits estimated with the quadratic regression

$$Page\ edits\ per\ capita_i = \alpha + \beta_1 Share\ of\ Internet\ users_i + \beta_2 Literacy\ rate_i + \beta_3 Literacy\ rate_i^2 + \varepsilon$$

- Aggregating page edits per country using the formula

$$\sum_{b=1}^2 \sum_{l=1}^L \overline{Page\ edits}_b \cdot \overline{Number\ of\ editors}_{b,l}$$

- Average page edits per buckets (5 to 99 edits and 100 or more edits) and average number of editors (intervals of ten) were calculated using simple interval means and normalisation to the aggregate number of page edits.

Calibration

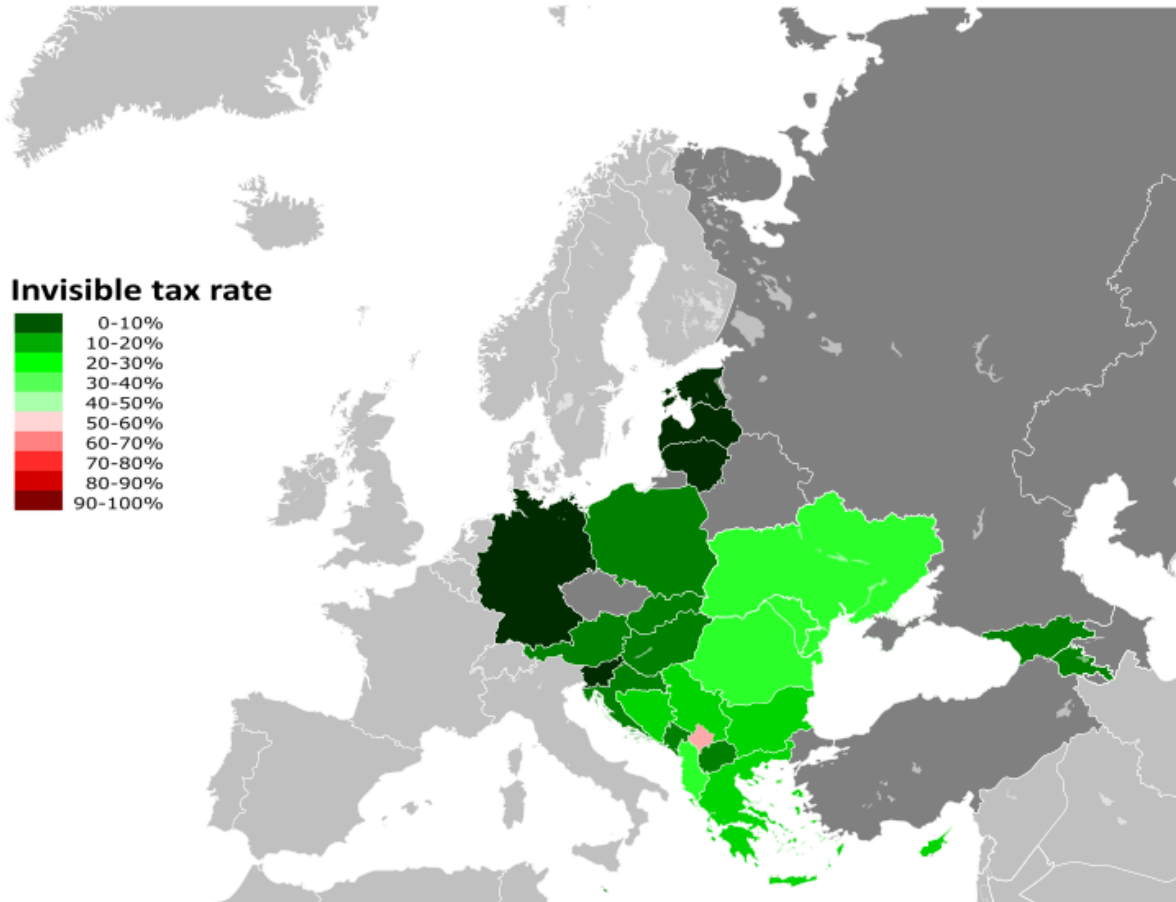
- Annual elasticities of page edits per capita:

Variable	2021	2022
<i>Share of Internet users</i>	0.3282** (0.1438)	0.2960*** (0.1096)
<i>Literacy rate</i>	-1.5652*** (0.3959)	-1.8668*** (0.4200)
<i>Literacy rate</i> ²	1.2425*** (0.3699)	1.4713*** (0.3374)
Intercept	0.3507*** (0.1256)	0.4394*** (0.1377)
Number of observations	144	144
Vertex	63.0%	63.4%

Notes: Robust standard errors are reported in parentheses. Symbols ***, ** and * denote statistical significance at the level of 1%, 5% and 10%, respectively.

- Using the elasticities to calculate the potential maximum of edits made and Wikipedia articles created.

Invisible tax across states



CEE average:

- 20.0% (2022)
- 23.3% (2021)

Global average:

- 55.5% (2022)
- 56.9% (2021)

Lowest:

- Estonia (3.0%)
- Latvia (4.5%)
- Lithuania (5.2%)

Highest:

- Kosovo (59.6%)
- Moldova (38.9%)
- Romania (33.6%)

Invisible tax across languages

CEE average:

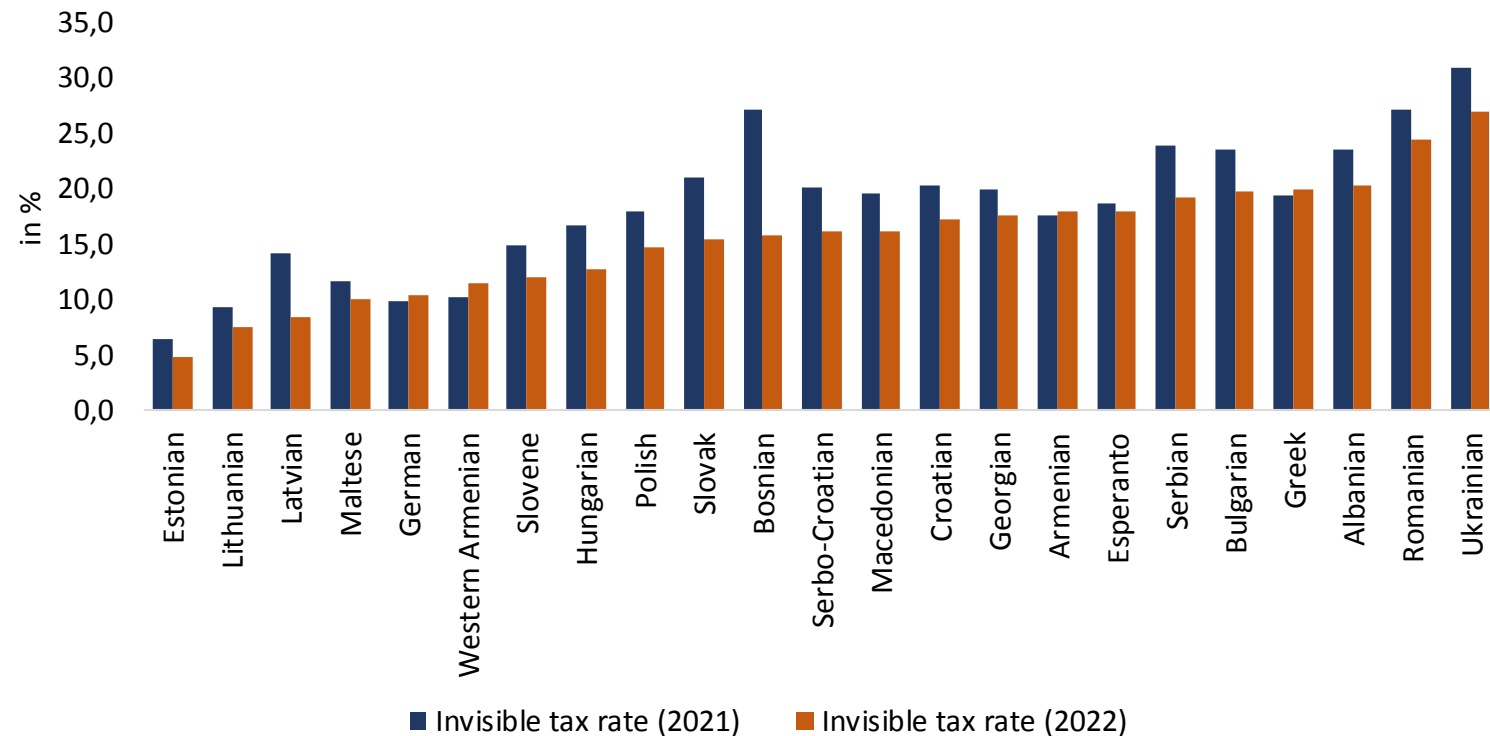
- 18.5% (2022)
- 15.5% (2021)

Lowest:

- Estonian (4.8%)
- Lithuanian (7.5%)
- Latvian (8.5%)

Lowest:

- Ukrainian (26.9%)
- Romanian (24.4%)
- Albanian (20.4%)



Future research

- Re-calibration of the model with more precise and more granular data.
 - WMF Analytics Team will start publishing data on total edits soon.
- Calculating other measures of economic inefficiency:
 - Loss of new articles created.
 - Loss of content quality.
- Extending the model to study the effect of other variables.
 - Engagement of the local communities in offline activities.
 - Studying social interactions and productivity of contributors.