www.d4d-ks.org
Democracy for Development
Demokraci për zhvillim
Demokratija za razvoj

Labor mobility:
A bottom-up approach

To go or not to go – that will not be the question.

Labor mobility:
A bottom-up approach

To go or not to go - that will not be the question.

Pristina, Kosovo

June 2020

Prepared by:

Gersi Gashi

Model design:

Gëzim Basha

Reviewed by:

Jehona Serhati

This document is produced by the Democracy for Development (D4D) Institute, as part of the project “Track 2 to Europeanization: A partnership approach,” supported by the Open Society Foundations. The views on this document are those of the authors’ and do not necessarily reflect those of the Open Society Foundation.

Content

Introduction 10

A voiceless Western Balkans will face a shortage of skilled workers 11

Kosovo's labour force trends Trend analysis 15

The unused potential - youth 18

Methodology 19
Survey *19*
The survey results *19*

A population that is waiting for the chance to leave 20
Youth are more prone to leaving *20*
Inability to easily find jobs is the main driver of emigration *21*
Countries where they would migrate *22*
What makes people stay *22*

Model 23
Introduction *24*
Limitations of the Axiomatic Approach *25*

The Complexity Approach 26
Limitations of the Complexity Approach *28*
Computational Limit *28*
Memory and Parallelization Limits *30*
Implementation *32*
Primary Assumptions *33*
Entities *33*
Predictive Drive *60*

Use Cases 63
Women workforce increase *63*
Wage Subsidies *65*
Migration Patterns *66*

Conclusion and recommendations 69
Policy recommendations *69*
To Kosovo Institutions *69*
The European Union *70*
Model related recommendations *70*
Government Usage *70*
Other Usage *71*

Table of Figures

Figure 1: Growth across WB region in 2019 12
Figure 2: Fiscal deficits and current account deficits 12
Figure 3: Employment and unemployment in WB countries 13
Figure 4: Kosovo Migration Patterns 14
Figure 5: Number of Kosovar first time asylum seekers and asylum seekers in the EU and Schengen zone 14
Figure 6: Kosovo emigrants in EU and Schengen Zone based on their reason of stay 15
Figure 7: Population in Working Age 15
Figure 8: Labour Force 16
Figure 9: Men and Women in the Labour Force 16
Figure 10: Participation of men and women in the Labour Force 16
Figure 11: Employment and unemployment trends 17
Figure 12: Working age population & Inactivity in total numbers 2016 - 2019 17
Figure 13: WB Population by age class, 2018 18
Figure 14: If offered the chance would you leave Kosovo? 20
Figure 15: The will to move from Kosovo based on age groups 21
Figure 16: What would be your main reason to leave Kosovo? (in %) 22
Figure 17: Countries where you would migrate to? 22
Figure 18: What changes would give you incentives to stay in Kosovo? 23
Figure 19: Chances of finding a work abroad compared to Kosovo (in %) 23
Figure 20: Cellular Automata Rule 22 logic 27
Figure 21: Cellular Automata Rule 22 development 27
Figure 22: Graph interconnectivity between agents 1-4 28
Figure 23: Number of agents increased from 10 to 40 to 100. 29
Figure 24: Use-case with 10 agents 29
Figure 25: Use-case with 10 agents fully interactive 29
Figure 26: Clusters of agents 31
Figure 27: Clusters and their identified computation zones 31
Figure 28: Entity and its concrete implementation (some entities have been hidden for clarity) 32
Figure 29: Hourly progression of the Clock entity 34
Figure 30: Daily progression of the Clock entity 34
Figure 31: Weekly progression of the clock entity 35
Figure 32: Age groups for Kosovo – 2016 37
Figure 33: Kosovo age distribution 37
Figure 34: Non-interpolated age groups 39
Figure 35: Interpolated age groups 40
Figure 36 :Strong interpolation, with heavy emphasis on young population 40
Figure 37: Well spread age distribution 41
Figure 38: Age distribution for Kosovo 47
Figure 39: Weekly workload 53
Figure 40: Weekly workload and Daily Workload 54
Figure 41: Daily workload on single load 55
Figure 42: Weekly average 55
Figure 43: Monthly average of workload 55
Figure 44: Yearly engagement of one agent 56
Figure 45: Monthly simulated data 56
Figure 46: Second month of agent data 57
Figure 47: One-year average of Fatigue 58
Figure 48: Chronic Fatigue (monthly) 58
Figure 49: Chronic Fatigue (quarterly) 59
Figure 50: Chronic Fatigue (seasonally) 59
Figure 51: Data collected from an Agent's happiness levels 59

Figure 52: A view of a linear regression model fitted to many points ... 60
Figure 53: Gaussian process predictors ... 61
Figure 54: Predictive workload ... 61
Figure 55: Confidence intervals ... 62
Figure 56: Deep learning ... 62
Figure 57: Women entering workforce ... 63
Figure 58: Unemployment Rates filtered by sex ... 64
Figure 59: Positions starting to saturate as day women are starting employment ... 64
Figure 60: Instantaneous wages per contract ... 64
Figure 61 :Tax law increase ... 66
Figure 62: Global migration patterns ... 68

Introduction

Migration is not a new phenomenon – at least in its occurrence. Yet in the recent years it has been increasing rapidly as in 2020 there are 83% more migrants compared to 2000[1]. As such migration has also gained space in the public discourse as people are on the move for economic, political, and humanitarian reasons. Migration is not concentrated on a particular region, but emigration is more evident in the Western Balkans compared to other parts of Europe. The issue is not emigration per se but it becomes an issue when there is a high rate of non-returnees. The high rate of people that never return back to the origin country, and especially if they are part of the labor force, propagates a variety of labor related policy issues.

Germany is the primary destination for Western Balkan's population to migrate to[2]. Germany, and other developed countries, are facing issues on a shrinking labor force, and it is the second highest country with migrant population, where only the US has more[3]. The demographics purport to the notion that the decline in Germany's labor force is inevitable even with the immigration of foreign labor force. The Deutsche Bank recently published a report in which a projection shows that the German labor market will have a deficiency of 6 million people compared to 2018[4]. In addition, German firms have reported that around 1.3 million vacancies are open[5]. These combined produce labor related policy issues for Germany where a status quo will inevitably lead to it having severe economic consequences. Germany has identified the decline in the labor force and the new law on skilled labor immigration is being used as a tool to combat the decline.

The immigration, both from EU and non-EU countries, to Germany has been increasing yet Germany's labour supply does not meet its demand. In the last quarter of 2018 Germany reached a peak of unfilled jobs and a record of job vacancies[6]. On the policy spectrum, in 2015, Germany opened up its labour market to the Western Balkan Six (WB6) via the Western Balkan Regulation (WBR). The new act, part of the Employment Ordinance, paved the way by not specifying skill requirements thus creating opportunities for both skilled and unskilled labour to be part of the German Labor market, always within the confinement of having a job offer. In only two years, 2016 – 2018, the German Federal Employment Agency reviewed more than 200 thousand employment applications from WB6, where the approval rate was more than 80%[7].

Since the WBR regulation came into place Kosovo has been characterized by both the highest number of applications using the WBR and the highest numbers of visas issued. Whilst emigration has been increasing Kosovo still faces high rates of unemployment ranking it with the highest unemployment rate in the region.

Besides the WBR which regulated and had the most effect on migration of skilled and unskilled labour from WB6 countries to Germany, there are also intergovernmental agreements between Germany and other countries. In the case of Kosovo, three such agreements are in place hitherto, however, the numbers of people that have benefited from these is rather dim. From 2015 – 2017, 84 people pursued vocational education, training, and work in the nursing and gastronomy sector, whilst 27 more were employed in nursing and construction fields, in addition to 30 more that left Kosovo to pursue employment in Germany in construction[8].

Germany's unemployment rate has been on a decline trend. In July 2019 the unemployment rate hit 3% which marks a record low for Germany since 1980. It maintained a stable unemployment rate around 3%, and currently

---

1 International Organization for Migration. 2019. World migration Report 2020. Retrieved from: https://publications.iom.int/system/files/pdf/wmr_2020.pdf
2 HINA. 2019. *Almost 250,000 Nationals of Western Balkan Countries Emigrated to EU in 2018.* Retrieved from: https://www.total-croatia-news.com/politics/37853-western-balkan
3 World Bank. 2019. Net Migration. Retrieved from: https://data.worldbank.org/indicator/SM.POP.NETM
4 Deutsche Bank. 2019. Marked decline in the German labour force despite substantial immigration. Retrieved from: https://www.dbresearch.com/servlet/reweb2.ReWEB?rwsite=RPS_EN-PROD&rwobj=ReDisplay.Start.class&document=PROD0000000000496399
5 Institute for Employment Research. Job Vacancy Survey. 2019. Retrieved from: https://www.iab.de/en/befragungen/stellenangebot.aspx
6 GAP Institute. 2020. The emigration of Kosovo's labor force to Germany. Retrieved from: https://www.institutigap.org/documents/38227_emigration-.pdf
7 Ibid. pg. 4.
8 Ibid. pg. 8.

in 2020, January and February both record at 3.3 and 3.4 percent respectively[9]. Job vacancy rates in 2019 were on average 3.3%[10], and them being empty is estimated to cost the economy 30 billion €[11].

In 2019 out of 62,234 visas that were approved for the WB6 countries for Germany, 13,802 were for Kosovar nationality people which is an increase of approved visas for Kosovars compared to 2018 were the number was 10,693. Only Bosnia & Hercegovina has had more approved visas in 2019, where 14,462 were approved[12].

This paper is a nonconventional approach, at least for think tanks, by D4D Institute to use agent-based models to explain how policies affect behaviors of people in regards to the labor market and mobility. The paper analyses Kosovo's context and trends in the labor force, provides case models of how labor demand-supply can be affected by new policies, and how a regional approach by the Western Balkan countries is needed.

A voiceless Western Balkans will face a shortage of skilled workers

The Western Balkans, especially the WB6[13], have had for a long time a common issue – unemployment. Unemployment in the Western Balkans ranges from 12% in Serbia and Albania, to 29% in Kosovo[14]. However, that problem will seem lightweight as a new issue is emerging, namely the massive emigration of high skilled workers. In 2019, around 230 thousand people left the Western Balkans to seek employment abroad[15]. Most of these were from Kosovo and Bosnia & Hercegovina, and albeit concrete data on their skill level does not exist it is purported that many of those were high skilled labour.

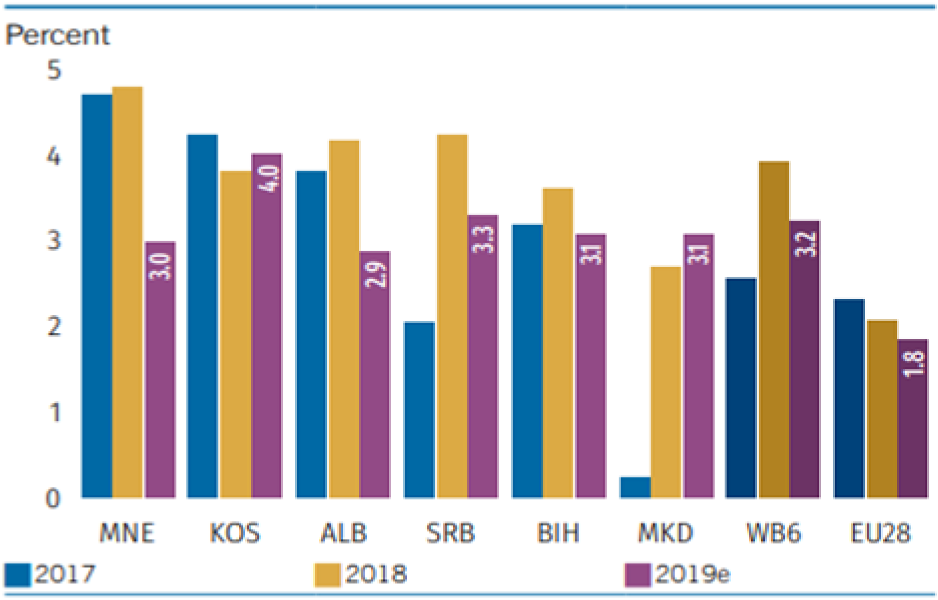
This is only the tip of the problem, as the real problem does not reside with the high number of emigrants but rather with the region's inability or more its reluctance to deal with the topic. Currently the circular migration has been trending as there are more temporary migrants than there are permanent migrants[16], at least in the developed countries. However, given that the WB6 have only recently started to discuss the idea of circular migration there is no platform or channels of communication that would allow its migrants to have connections with their country of origin. As such informal channels of communication are kept by the migrants themselves and this contributes to them detaching more from their country of origin.

The diaspora of WB6 combined amount to 4.5 million people[17]. From 1990 until 2017 the WB6 have lost one-tenth of its population. Estimations are that they will lose 14% more by 2050 and the main reason for this decline is attributed to emigration and more importantly to the emigration of young skilled workers[18]. This will leave the Western Balkans countries to deal with issues as: lack of skilled workers, labor market shifting towards low skilled jobs, and ultimately a high burden on social schemes for retirees.

---

9 Eurostat. Retrieved from: https://appsso.eurostat.ec.europa.eu/nui/submitViewTableAction.do
10 Eurostat. Retrieved from: http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=jvs_q_nace2&lang=en
11 Conolly K. (2018). Germany passes law to lure non-EU skilled workers. Retrieved from: https://www.theguardian.com/world/2018/dec/19/germany-passes-immigration-law-to-lure-non-eu-skilled-workers
12 Federal Employment Agency. (2020). Retrieved from: https://statistik.arbeitsagentur.de/nn_31954/SiteGlobals/Forms/Rubrikensuche/Rubrikensuche_Suchergebnis_Form.html?view=processForm&resourceId=210358&input_=&pageLocale=de&topicId=17350&region=&year_month=201912&year_month.GROUP=1&search=Suchen
13 WB6 is an acronym for Western Balkan Six which includes: Albania, Bosnia & Hercegovina, Kosovo, Montenegro, North Macedonia, and Serbia
14 World Bank Group. 2020. Western Balkans Labor Market Trends 2019. Pg. 11. Retrieved from: http://documents.worldbank.org/curated/en/351461552915471917/pdf/135370-Western-Balkans-Labor-Market-Trends-2019.pdf
15 Vracic A. 2020. Time for Policy Change on Western Balkans Emigration. Balkan Insight. Retrieved from: https://balkaninsight.com/2020/01/13/time-for-policy-change-on-western-balkans-emigration/
16 Economic Policy Institute. 2019. OECD highlights temporary labor migration. Retrieved from: https://www.epi.org/blog/oecd-highlights-temporary-labor-migration-almost-as-many-guestworkers-as-permanent-immigrants/
17 Berlin Institute for Population Development. 2017. Europe's Demographic Future. Retrieved from: https://www.berlin-institut.org/fileadmin/user_upload/Europas_demografische_Zukunft_2017/Europa_engl_online.pdf
18 Ibid. pg. 138

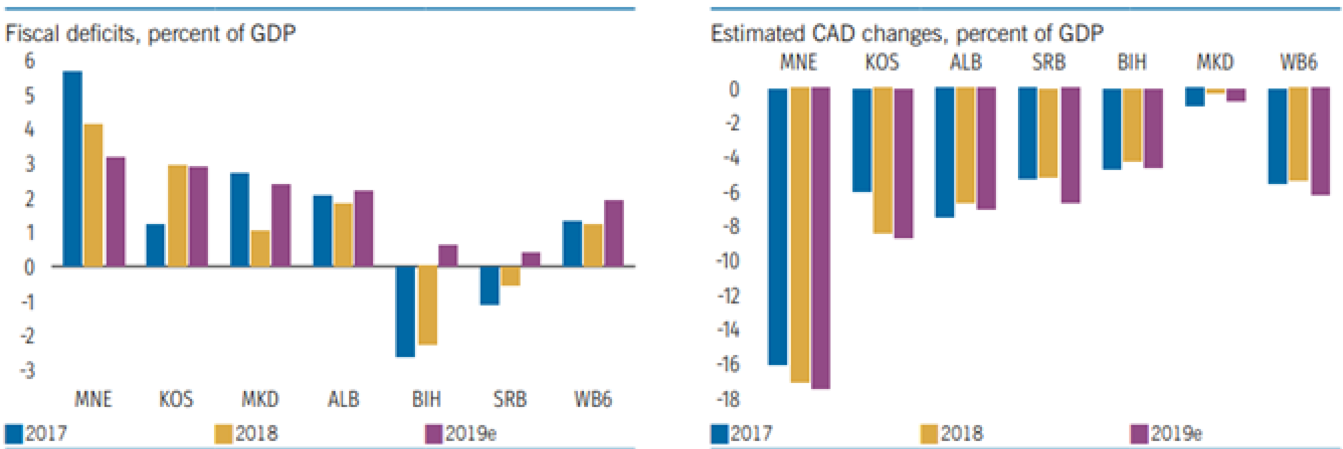
The economic growth in the Western Balkans has been slowing down mainly due to low foreign direct investments and exports. Economic growth in the WB has been estimated at 3.2% which is slower when compared to 2018 where it was at 3.9%[19]. However, the first part of 2019 saw a growth in new jobs as 150 thousand new ones had been created compared to a year earlier.



Source: National authorities and World Bank estimates.

Figure 1: Growth across WB region in 2019

The main reason attributed to the slow in economic growth is the lower contributions from exports and investments, and even a surge of public spending is combatting the issue.



Source: National authorities and World Bank estimates.

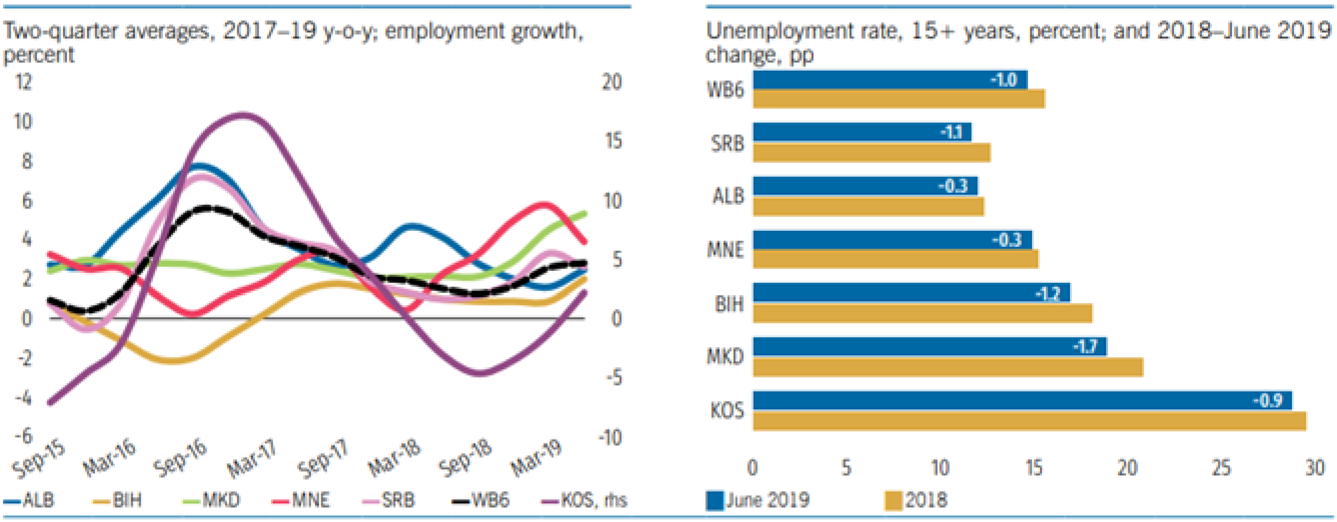
Source: National authorities and World Bank estimates.

Figure 2: Fiscal deficits and current account deficits

Whilst public investment is a good tool to boost the economy, if not accompanied by efficiency and effectiveness it will create other problems and, in this case, slow economic growth. The World Bank Report indicates that in 2019 employment rose in WB countries by 2.4% compared to 2018[20]. It is important to stress that due to the higher employment, the unemployment rate reached a record low – 15.8%. A stable and good economic model is to lower unemployment through employment. However, in B&H unemployment fell down to 15.7% partially due to increase in employment but more due to the decline in the working age population.

---

[19] World Bank. 2019. Retrieved from: http://documents.worldbank.org/curated/en/643781570478210132/pdf/Rising-Uncertainties.pdf
[20] Ibid.



Source: National statistics offices and World Bank staff estimates.
Note: June 2019 data refer to 4-quarter moving average.

Figure 3: Employment and unemployment in WB countries

Albeit the increase of jobs, and decline in unemployment, the WB countries are still facing high numbers of emigration rates. Currently, the engagement with the diaspora is left to the countries individually and each of them employ a different approach. Given that the main host for emigration from the WB is EU it ought to play a significant role in assisting the region as well. However, not a lot has been done by both parties, the EU and the WB countries, on promoting the idea of circular migration.

Kosovo's case goes on to show how unemployment can decrease whilst employment also decrease, as the two do not necessarily share a confined relation. In some cases, a drop in unemployment can attributed to high levels of emigration and inactivity, and not to a significantly improving economic climate. As emigration continues and there is a lack of a discussion in the regional level on the topic, combined with an absent and clear strategy and lines of communication with the diaspora, will most likely produce consequences for the region's economic vitality.

The only census data for Kosovo, which was conducted in 2011, went on to show that Germany (35.25%) and Switzerland (22.94%) were the primary destination of the population for emigration[21].

In the last twenty years there have been two outlier cases for Kosovo in regard to emigration. During the war in 1998-99, and in 2014-2015 period. The latter, in which more than 100 thousand people left the country, case has prompted towards more policy focus on reintegration policies but also towards informative campaigns on legal pathways on migration.

The data presented below, produced by the Ministry of Internal Affairs, as part of the publication on Kosovo's Migration Profile, shows that in the period of 2016 – 2018 there has been a considerable difference between immigrants and emigrants. Hence, the net emigration flow is on the negative, as it measures emigrants minus immigrants. The data shows that in 2018 there have been 28 thousand emigrants, which then on the migration flow displays as roughly minus 21 thousand people.

[21] Kosovo Agency of Statistics. (2015). Kosovo Migration Profile.

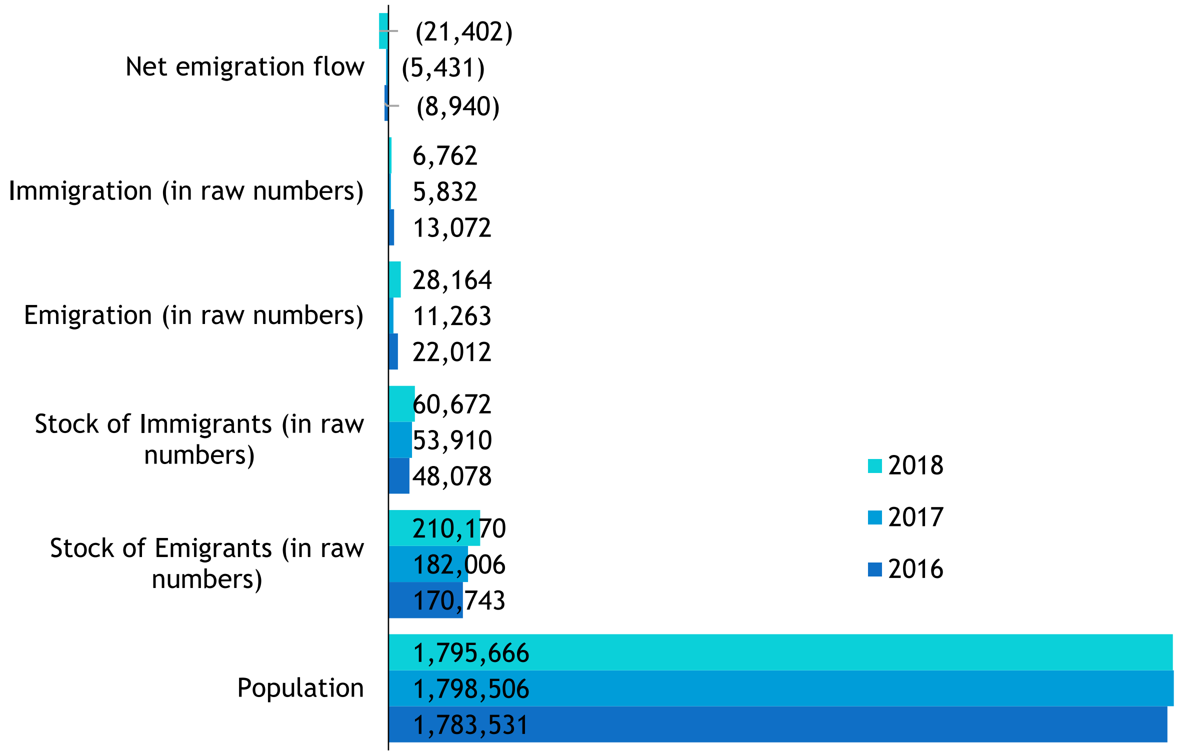


Figure 4: Kosovo Migration Patterns[22]

The number of first time Kosovar asylum-seekers in 2018 was 3,395. Besides the outlier year mentioned above, the number of Kosovar asylum seekers in EU and the Schengen Zone has been declining.

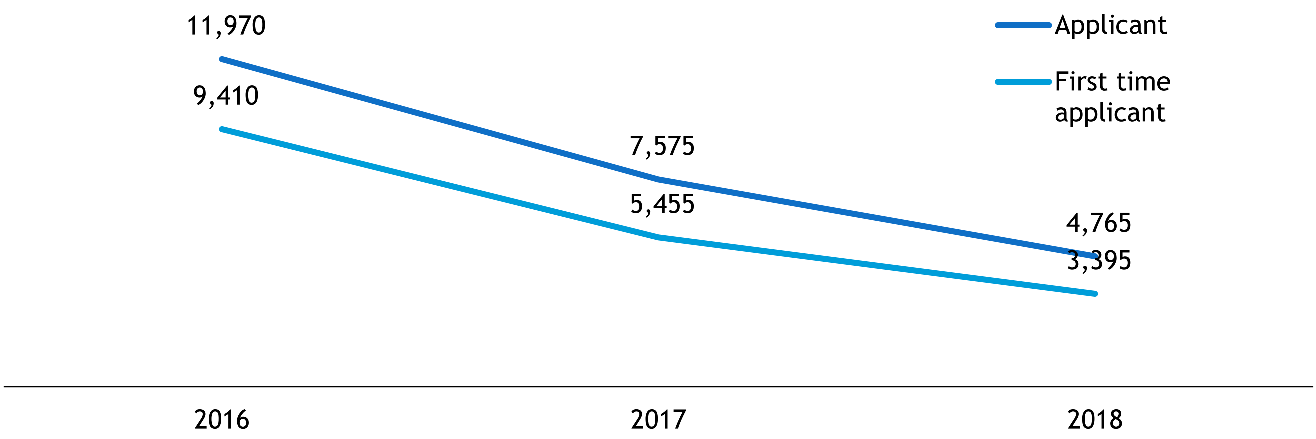


Figure 5: Number of Kosovar first time asylum seekers and asylum seekers in the EU and Schengen zone

The majority of Kosovars that got approved for their stay were for family reasons. In 2018, around 19 thousand people emigrated and got their permit due to family reasons. Notably, the number of those that got their permit accepted for work has been increasing throughout the period of 2016 – 2018.

[22] Ministry of Internal Affairs. (2018). Kosovo Migration Profile. Retrieved from: https://mpb.rks-gov.net/DocumentsShpalljet/PROFILI%20I%20LEHTE%20I%20MIGRIMIT%202018%20SPL.pdf

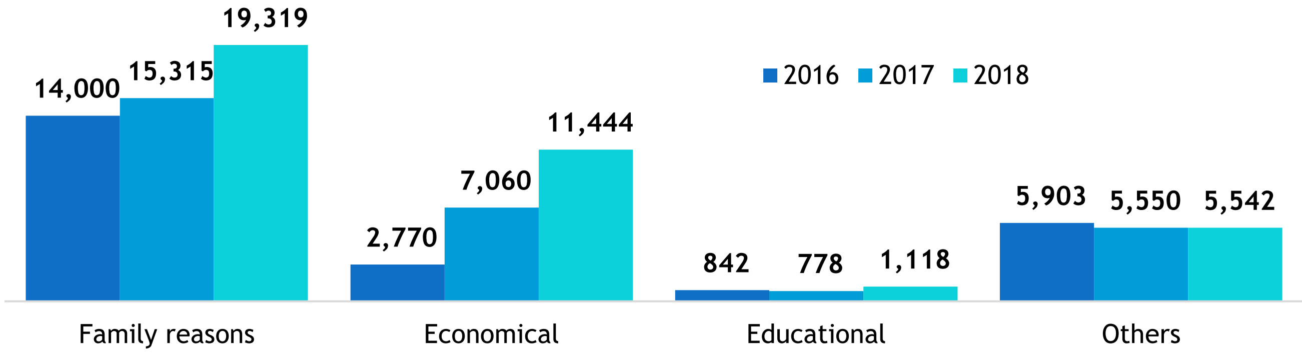


Figure 6: Kosovo emigrants in EU and Schengen Zone based on their reason of stay

Kosovo's labour force trends Trend analysis

Kosovo has not encountered any spikes that can be considered outliers in its labour force. In this part of the paper, data from the Kosovo Agency of Statistics (KAS) has been used as it is the official agency for providing statistics in Kosovo. Albeit, some of their data differ from other publications[23], for this part they have been ignored for two reasons: KAS is the official institution to produce and distribute official statistics, and the KAS reports have different time-entries thus allowing for the trend analysis.

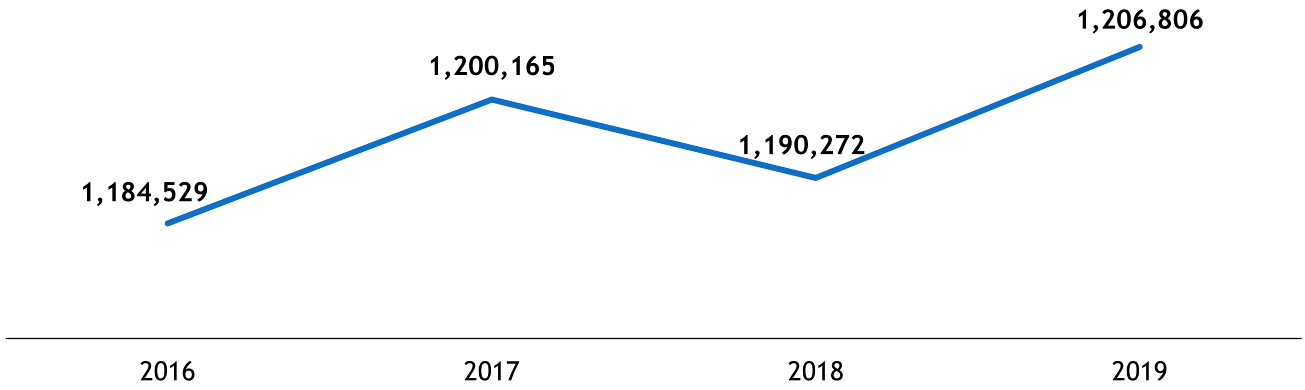


Figure 7: Population in Working Age

The results in the above graph show that from 2016 to 2019 there has been an increase of roughly 22 thousand people that have become part of the working age population.

The trend of the changes in the working age population naturally has also been followed in the labour force participation. The graph below shows that labour force has increased by 30 thousand from 2016 to 2019 and currently stands at 40% compared to the working age.

[23] Here we are referencing Millenium Foundation Kosovo's Labour Force Study publication which is retrieved here: https://millenniumkosovo.org/wp-content/uploads/2018/11/MCC-Kosovo-Labor-Force-and-Time-Use-Study-Final-Research-Report-1.pdf

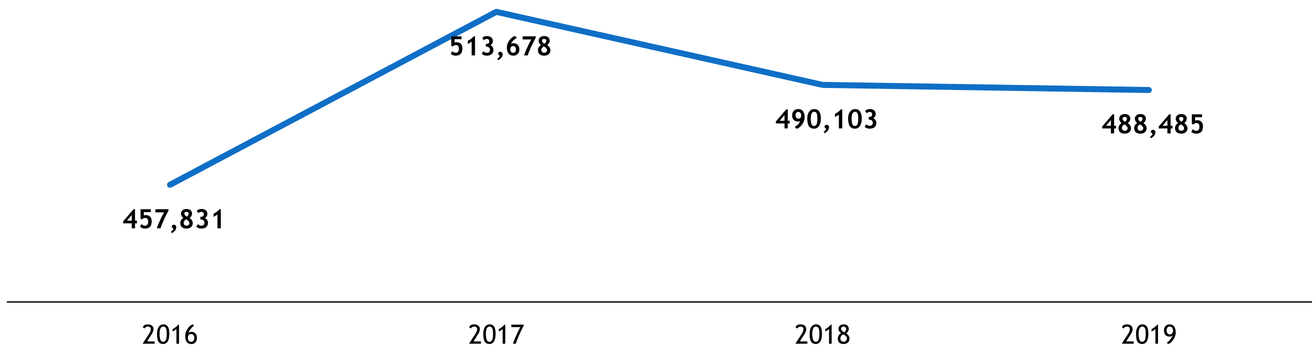


Figure 8: Labour Force

Notable differences can be seen when looking at men and women in the labour force. Men's participation in the labour force has been declining, whilst women's participation has been increasing.

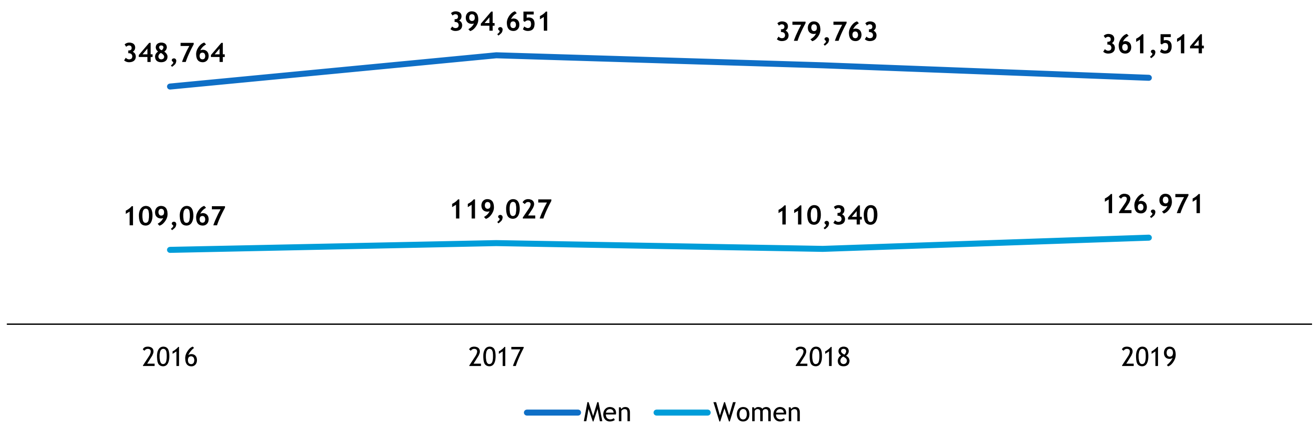


Figure 9: Men and Women in the Labour Force

The increase of women's participation in the labour force is positive, however, the alarming situation is that only 21% of women are in the active labor force. The activity rate of women has not increased in significant numbers from 2016.

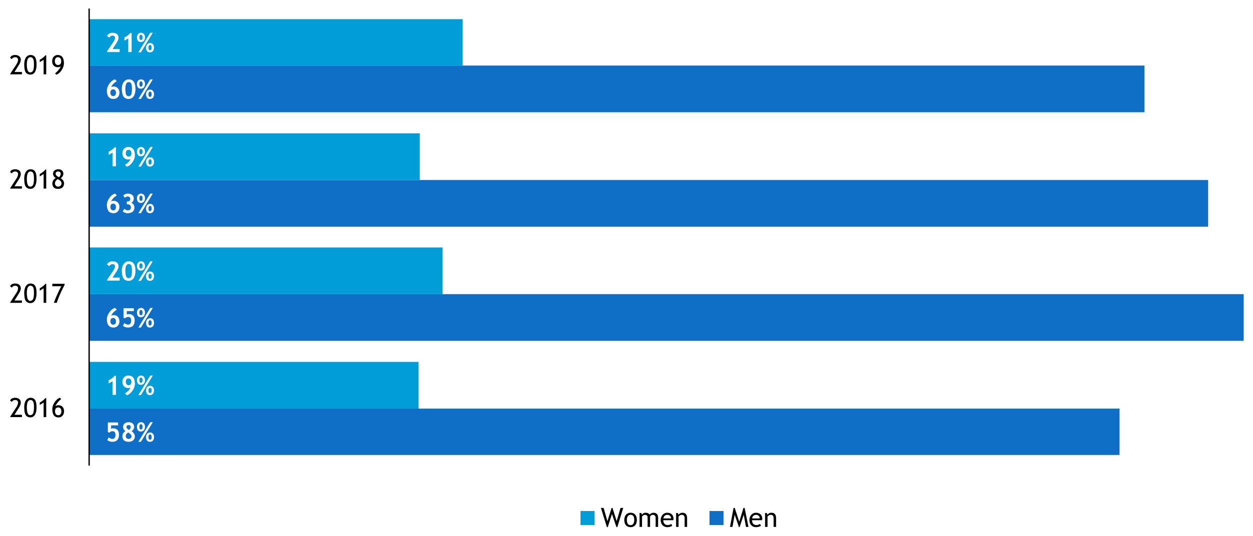


Figure 10: Participation of men and women in the Labour Force

Unemployment remains at a staggering 25.7% in Kosovo, however that is not the most important thing to note. Whilst unemployment has been decreasing throughout the years this did not show a negative relation with employment. As an example, from 2016 to 2017 employment grew by 25 thousand people, whilst unemployment at the same time grew by 30 thousand people. In principle the same occurred from 2017 to 2018 where both of these declined. In labor economics the reason why both these can have the same reaction (decline or increasing) is either people are moving from unemployed to inactive, leaving the country, or people are moving from inactive to either employment or unemployment (i.e. active).

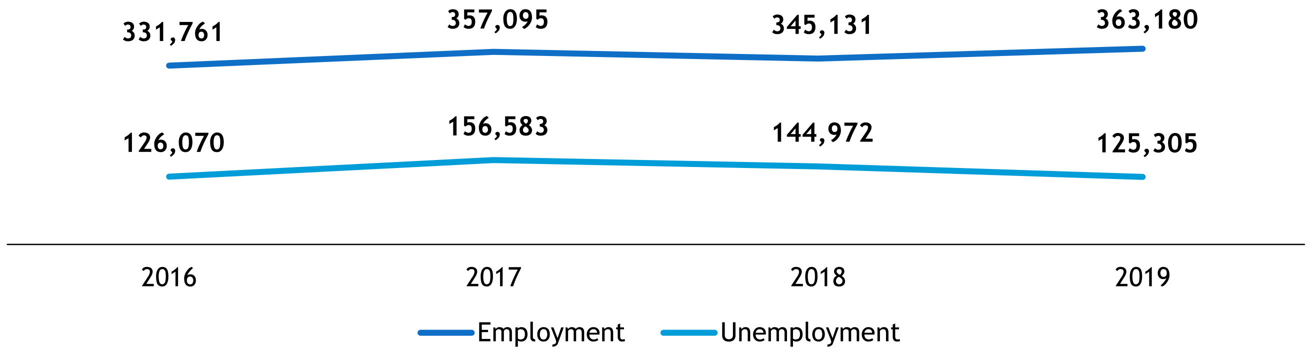


Figure 11: Employment and unemployment trends

Upon analyzing the inactivity data from 2016 – 2017 it shows that the inactivity rate dropped by 40 thousand. The drop in inactivity implies that the people that joined from inactive to active have either found employment, or are unemployed. Cross-comparing these two variables gives us a more narrowed analysis as taking the example of the 2016-2017 trend, data shows that unemployed and employment combined increased by 55 thousand, and it would be a reasonable guess to assume that the inactivity would also decrease in the range of that number however that was not the case. Inactivity as seen below dropped by 40 thousand, thus leaving 15 thousand which are not thoroughly explained. Another reasonable guess would be that the working age population has decreased, however, data again shows that that is not the case as the working age population in fact has increased by 16 thousand.

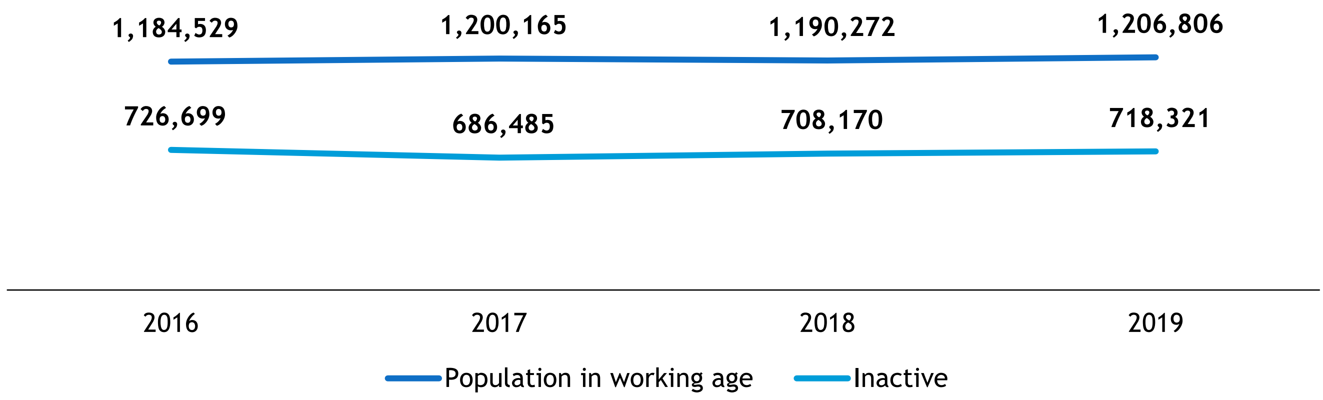


Figure 12: Working age population & Inactivity in total numbers 2016 - 2019

From 2018 – 2019 the inactivity has increased by roughly ten thousand people. The increase of the working age population contributes towards this increase in inactivity. In terms of percentage more than half of the working age population (59.5%) remains inactive.

The unused potential - youth

Kosovo has among the highest youth proportion in the World. Such an asset is very vital when discussing labour force potential. It becomes even more important when there is a trend in the EU-28 population of ageing with lower birth rates and higher life expectancy.

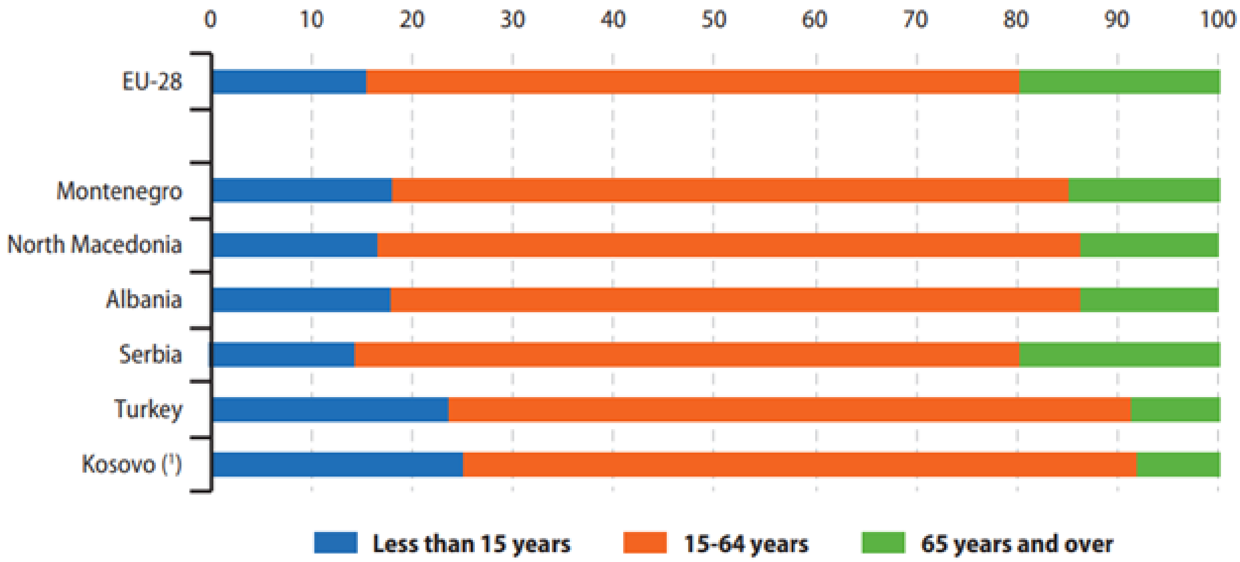


Figure 13: WB Population by age class, 2018 [24]

The figure above shows that Kosovo has a very high percentage of people in the working age and also a high percentage of people with less than 15 years who will soon enter the working age group. These parameters would allow for the country to have a comparative advantage to others. However, that does not hold true in Kosovo's case. The youth unemployment rate in Kosovo is at a staggering 49.4% in 2019, and whilst this is a very concerning issue another one is more - the NEET rate of youth[25]. NEET rate measures people that are not in education, employment, or training, and for Kosovo's youth that amounts to 32.7%[26]. The problem is not only for the young men and women as more than half of the working age population (59.4%) are not economically active, and even from those that are economically active almost one in four (25.7%) is unemployed[27].

Mildly put Kosovo has not been utilizing its potential. However, this does not mean that the potential will not be utilized by someone else. Thus, developed countries that face demographic crises will turn their eyes on countries to fulfill their labor force demands. The most affected sectors under the WBR were construction, accommodation and food services, and administrative and support services. However, notably construction, and the health sector, were granted pre-approvals also via other means besides the WBR.

Germany's demand for doctors and nurses is at 70 thousand and 500 thousand respectively[28]. Until now only a part of the doctors in Germany who are foreign come from Kosovo. On an interview for BIRN, a manager of the nursing department in a public hospital in Hassfurt, claims that "German hospitals typically pay around 6,000 euros to headhunters for every doctor or nurse placed."[29]. Further the article also makes the connection in Kosovo locally

[24] Eurostat. 2019. Key Figures on Enlargement Countries. Pg. 16. Retrieved from: https://ec.europa.eu/eurostat/documents/3217494/9799207/KS-GO-19-001-EN-N.pdf/e8fbd16c-c342-41f7-aaed-6ca38e6f709e
[25] Trading Economics. 2019. Kosovo Youth Unemployment Rate. Retrieved from: https://tradingeconomics.com/kosovo/youth-unemployment-rate
[26] Kosovo Agency of Statistics. 2019. Labour Force Survey. Pg. 10. Retrieved from: https://ask.rks-gov.net/media/5369/anketa-e-fuqis%C3%AB-pun%C3%ABtore-afp-2019.pdf
[27] Ibid
[28] Ahmetxhekaj Sh. Balkan Insight. 2019. Brain Drain: Will the last doctor in Kosovo turn out of the lights.
[29] Ibid

as one of biggest language centers that offers free German courses gets more than 3000 applications for only 140 free places.

However, that is not the entire case as other data suggest a different side of the medal. Kosovo's vocational schools in health occupations have seen an increase in enrollment. Whilst in 2016 there were 6,085 enrolled, this number increased to 8,022 in 2018/19. Further, a recent study published by GAP Institute has calculated thoroughly the risks of migration for specific occupations. In their analysis, as the Institute gathered data for unemployed people in specific occupations, and cross compared those with the job-vacancies in Germany for said occupations, the results are that administrative and support services (224%), mining and quarrying (100%), and real estate activities (100%) have the highest risk of labour migration. However, notably construction (48%) and the human health and social activities follow that list (46%).

Methodology

Survey

One survey was conducted in Kosovo in the form of an omnibus[30] using a sample of 1065 people (50.5% men and 49.5% women). The survey covered all seven (7) Kosovo regions, with respondents from age of 18-65+. The sample was stratified: (i) based on ethnicity, (ii) based on the residence, and (iii) based on the number of surveys implemented in each residence. The coverage includes both urban and rural areas. The questionnaire was built to gauge the perception of the respondents.

The demographics part of the questionnaire has allowed us to further analyze the results based on gender, age groups, education level, settlements, and employment status. **Gender** was distinguishing men and women's responses, and **settlement** delineated the urban and rural placement of the respondents. **Age groups** created grouping of age groups of: 18-24, 25-34, 35-44, 45-54, 55-64, and 65+. **Education level grouping** offers an intricate lense to the respondent's education level as it group them as follows: (i) some years of elementary school, (ii) primary school, (iii) some years of high school, (iv) high school, (v) student, (vi) university, and (vii) master or doctarate degree. On **employment level**, the grouping was done as: (i) unemployed, (ii) inactive, (iii) employed in the public sector, (iv) employed in the private sector, (v) employed from time to time, (vi) retired, (vii) housekeeper, and (viii) student.

The survey results

The survey, whose methodology was explained in detail, offers a perspective complementary to the empirical evidence. The aim of the questionnaire was to gauge perception of Kosovars in regards to labor mobility, and also dwelled upon their reason for migration. In summary the results have concluded that there is a polarization between two extremes where a majority of 55.6% said that they would leave Kosovo for a certain amount of time, out of which 31.4% said forever, but also a large number of people (40.6%) stated that they would not leave Kosovo. Analyzing the reasons for migration the survey participants were pointing to the notion that there are more job opportunities pertaining to their field of studies (22.2%), and that there are also job opportunities regardless of their academic background (26%), however also a large portion noted that family reasons (24.1%) would be the key reason for them to move out of Kosovo. This goes on to correspond to the answers given in the other question as to what changes would make them rethink their decision to leave, as 39.1% agreed that finding any kind of job with reasonable pay would make them want to stay, and 26.4% stated that they would only rethink their decision if the job they found was in their field of study and with a competitive wage.

[30] An omnibus survey is a method of quantitative marketing research where data on a wide variety of subjects is collected during the same interview. Retrieved from: https://www.surveyanalytics.com/omnibus-surveydefinition.html

A population that is waiting for the chance to leave

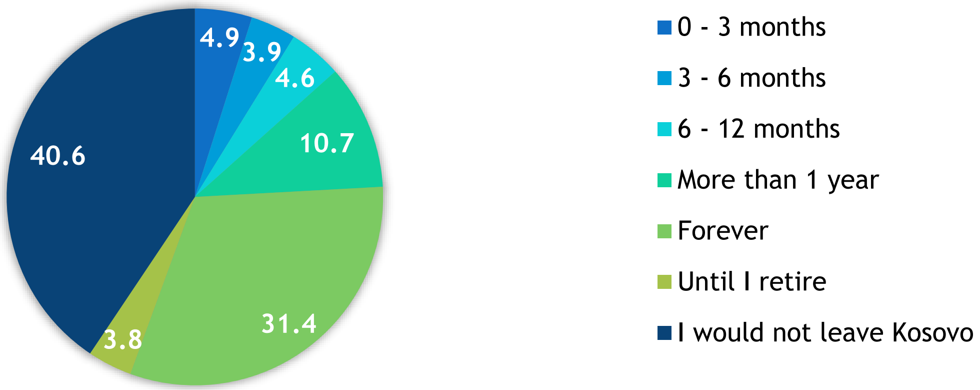


Figure 14: If offered the chance would you leave Kosovo?

The figure above shows how there are two extremes battling one-another, where the ratio of those that would leave to those that would never leave stands at 60%-40%. However, out of the 60% that would leave if given the chance, around 28.5% would only leave for a certain period and then imply that they would come back. There was no significant difference between men and women in their responses, however notable differences were observed on age, education level, and employment status.

Differences in age, education level, and employment status are explain below. To summarize, youth are more inclined to leave the country (81.6%) compared to other age-groups, whilst pertaining to the education level the highest group that would leave are students (81.4%), and based on employment level, unemployed people have the highest percentage of leaving (74.9%).

Youth are more prone to leaving

Young men and women as expected are more inclined to leave compared to other age groups. Within youth[31], a dominant 81.6% stated that they would leave Kosovo if given the chance, with 44.2% stating that they would leave forever. Only 18.4% stated that they would not leave Kosovo even if given the chance.

The more we move towards the age group of 65+ the tendency to leave if given the chance lowers as 78.8% within that age group state that they would not leave Kosovo. These data guide us on seeing that the most vulnerable groups are from 18-44, with the youth population being most prone to leaving Kosovo.

[31] Youth in this case is considered the age group 18 – 24 based on the survey conducted.

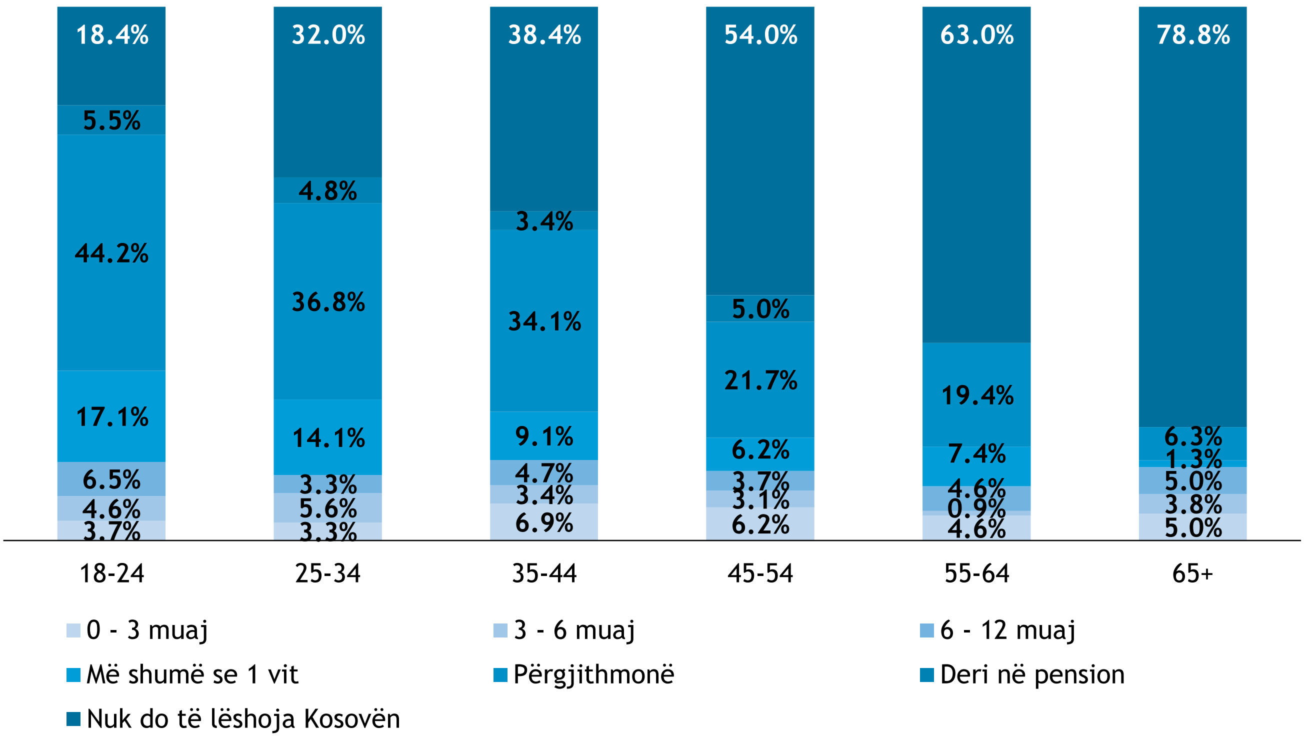


Figure 15: The will to move from Kosovo based on age groups

Inability to easily find jobs is the main driver of emigration

Out of the people that claimed that they would leave Kosovo, the highest percentage (26%) stated that the reason for leaving Kosovo is because abroad they can find a job easily, get paid well, and all that regardless of their academic background. The second answer with the highest percentage is people that would leave for family reasons where 24.1% of the respondents shared that opinion. The two other main reasons besides the ones mentioned above are linked heavily with economic wellbeing and more concretely with careers rather than just jobs. More than one in four people (22.2%) stated that their main reason for leaving would be because it is better for their profession, gives a chance to grow professionally, and there are jobs where they can express their abilities and advance them. Another 16% stated roughly the same, but only an added nuance where they perceive that their hard work for the profession that they chose is not being met with the desired wage, and conclusively the same one would result with a better situation in financial terms in another country.

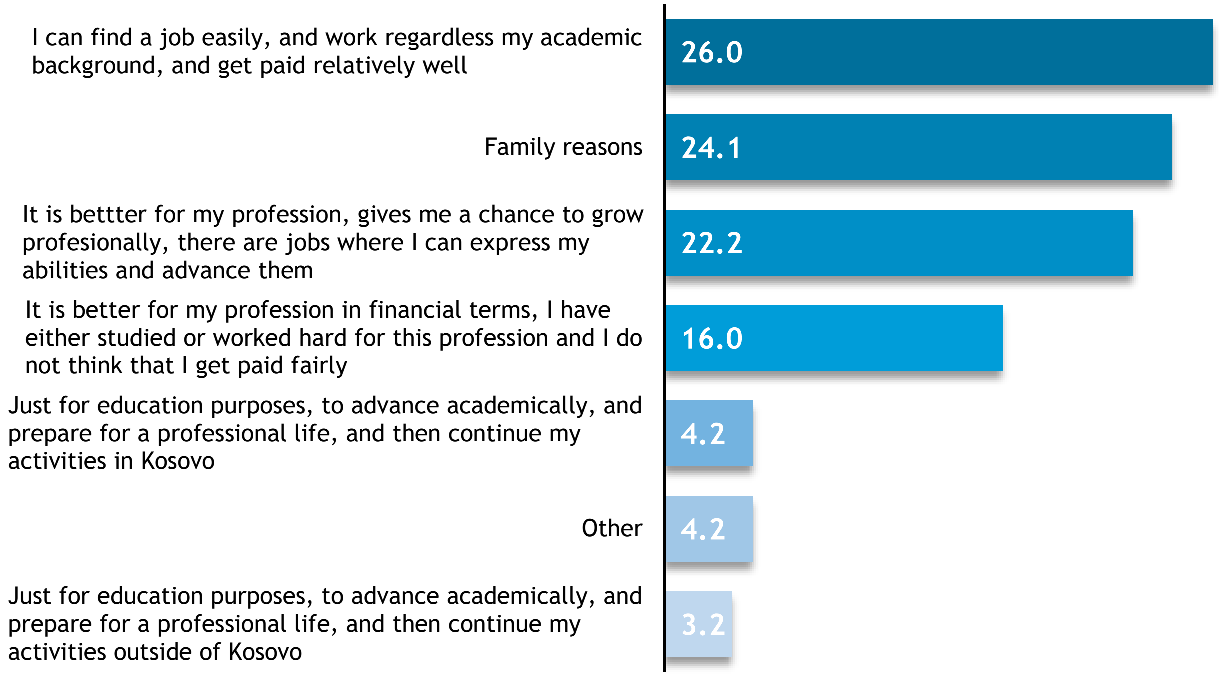


Figure 16: What would be your main reason to leave Kosovo? (in %)

Countries where they would migrate

When asking the respondents to select up to three countries that they would migrate to, the countries which stood out the most were Germany, Switzerland, and the United States. This list does not come as a surprise, especially for Germany and Switzerland, as most of Kosovo's diaspora resides in these two countries. Germany was listed as one of the countries to migrate to by a staggering 74.8% of the respondents, and Switzerland by 69.6%, with the US coming third by 26.9%. Given these choices a chapter on this policy paper has been dedicated to labor mobility from Kosovo to Germany.

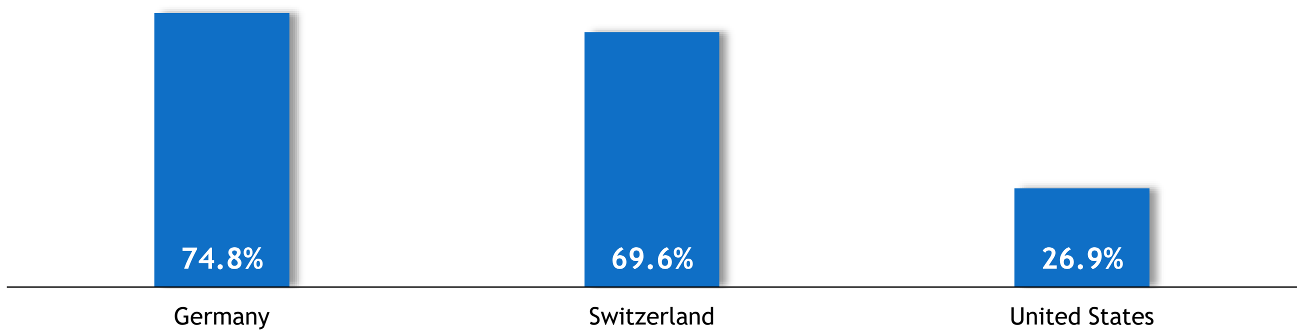


Figure 17: Countries where you would migrate to?

What makes people stay

The incentives that would make people stay in their home country ought to reflect the reason for them leaving. In this question of the survey, our intention was to explore whether there was anything beyond the reasons that the respondents mentioned when asked why they would leave.

The most notable answers are directly linked with socio-economic status. A high percentage of respondents (39.1%) responded that finding a job, any job, with reasonable pay would change their mind on leaving Kosovo. This finding purports that their main reason is not the lack of jobs in their fields, but rather a lack of jobs with a reasonable pay in general. Albeit, there can be a discussion as to what constitutes a reasonable pay, in the general discourse it can be understood as the average pay for the position that they are working on.

However, another high percentage of respondents (26.4%), stated that they would change their mind only if they get employed in their line of profession but with a competitive salary. This target group can be considered vulnerable as their answer would imply that they have an education which connects with a profession thus for a high degree can be estimated that they are part of a skilled-labour which could lead to their fast emigration. In addition, 11.5% stated that they would not leave if they found a job in their profession, regardless of the wage.

As last, a high percentage (17.5%) of people stated that they would have incentives to stay in Kosovo if there is an increase in the quality of education and then that education equates with a job placement. Again, it purports that jobs are essential for people to have incentives to stay rather than emigrate.

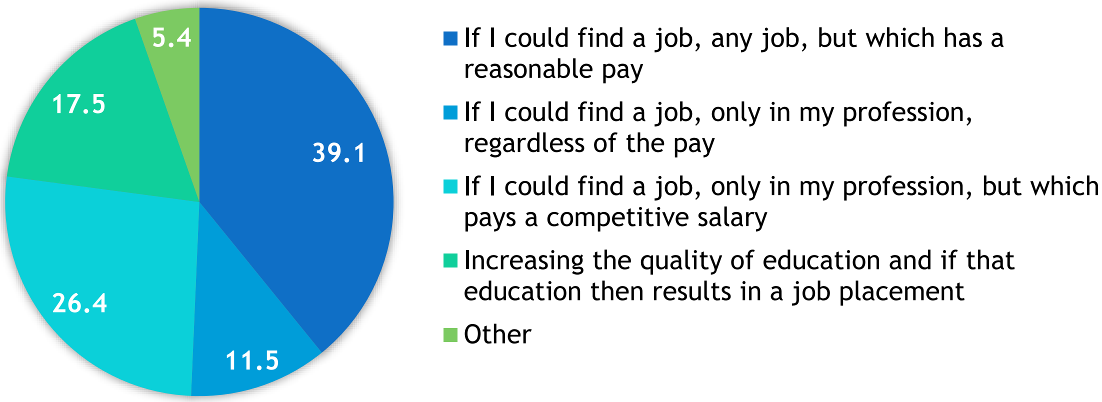


Figure 18: What changes would give you incentives to stay in Kosovo?

When asked all the respondents, also those that did not state that they would leave Kosovo, as to whether they think the chances are higher of finding a job abroad than in Kosovo, 80.8% stated that the chances are either somewhat higher or much higher. Only 6.4% respondent with lower chances, and 5.9% with the same.

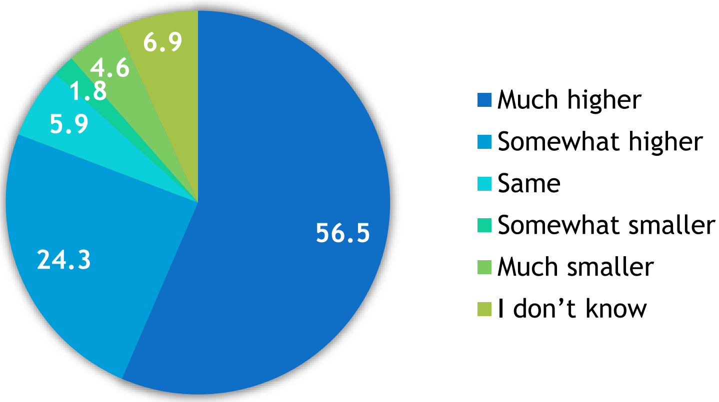


Figure 19: Chances of finding a work abroad compared to Kosovo (in %)

Model

In this section we will develop an agent-based model (ABM) that simulates the various interactions of the workforce within a country. Later we will extend this model to include the interactions between various agents internationally and model the European Union and the Western Balkans in it.

A study of this scale demands focuses on both the complexity aspect of the computation and the numerical modeling techniques. As is well known, for simulations of agents a crucial aspect of the complexity approach is to understand how interacting elements produce aggregate patterns that those elements respond to in turn (Gallegati & Richiardi, 2010). In our case, these aggregate patterns are economical in nature, whereas in the general case these may be of any variable nature. In terms of numerical modeling, we will need to focus on lowering the computational complexity and cement fast computation times as only one extra agent can cost minutes in computation. Considering that we want to discuss over larger periods of timespans, i.e. days and months there will be other issues as well which will be discussed at length later.

Introduction

As some people put it, economics is in troubled waters (Gallegati & Richiardi, 2010). Even though there is a well-established mainstream approach to working out problems in economics, its coherence is frequently questioned as to provide newer models. The current methods that are employed to analyze and further solve economic problems are remnants of older times. As Gallegati puts it, about one century after Isaac Newton published his seminal work on physics, where the motion from simple apples to planets and stars was unified under one general theory, it was inevitable that the social scientist of the time to borrow this new science and apply it to issues belonging to him. Ultimately, economists would create their own *"mechanical"* models to explain the plethora of observations that they had accumulated over many years.

Keynes, in his work, would go on to state that while politicians are slaves to economists, it is the latter (economists) who are slaves to physicists. Furthermore, the assumptions that make up classical mechanics such as reductionism and determinism would go on to formulate the basis of modern economic theory. With the onset of statistical physics, namely the era of Boltzmann and with the evolution of thermodynamics, the mathematical arsenal as wielded by physicists would now be far more advanced and even more suited to deal with the issues of economics. Ironically, even with the obvious success it harbored, by the 1970s it was widely ignored and even blatantly rejected with the new neoclassical train of thought. (Gallegati & Richiardi, 2010).

The key issue that is debated by the neoclassical school and the "old" school is that individuals do not aggregate to something beyond a simple summation – so to speak – of their behavior. Therefore, it is believed that there should be no discernible difference between the theoretical frameworks of the micro versus those of the macro. Consider a planet and its motion around the Sun, the neoclassical school goes on to claim that its motion is simply the averaged-out motion of the atoms that make that planet up. While a physicist might disagree on this on many points, the economist would place significant stress on this point only – this approach does not take into consideration the individual elements that make up larger structures do in turn react to those structures and therefore add in layers of complexity and inter-dependence.

One of the bigger issues that the entirety of the scientific school deals with, especially the mathematically inclined ones, is the concept of equilibrium. At face value, this is a very simple concept with a large practical background, but theoretically it is a very different story. In the world of physics for example, we have various equilibria, including but not limited to statistical equilibrium, quasi-equilibrium, dynamic and static equilibrium, etc. In many economic models, equilibrium simple means that there is a state in which the individuals, and their accompanying large-scale aggregated structures are engaging in economic activity where both the demand and supply are met. But let us get back to the example of the physical world in which we have a plethora of equilibria to discuss. Consider the planet which we discussed earlier. Let us assume this planet is the Earth and it is orbiting around the Sun. Now let us zoom in to the Earth and observe the motions of the components that make it up, i.e. the very living beings, mechanical objects such as cars and other associated objects in it. It is easily claimed that there is no equilibrium between the motion of the living beings and the moving objects within the planet. Cars move in different directions; airplanes fly in various directions – with many of them flying against the rotation of the earth. And yet, the Earth still moves in what looks like "perfect" orbit around the Sun. It appears, and experiment confirms this, that there is no need for local equilibrium to aggregate to large-scale equilibrium. The same applies to many fields in physics and the natural sciences, where the state of micro-scale does not necessarily correspond directly to the macro-scale. (Feller, 1957)

This view of the world inherently maintains that a planet can only be in equilibrium when it is desolate, void of life, and any moving object in it. By simple application of reductionism this would maintain that the only planet void of any moving objects is a planet with no atoms at all – i.e. a non-existent planet. Interestingly however, this very weird play is what brings the new methods to economists and ultimately solves many problems ahead.

Murray Gell-Mann, a Nobel prize winner physicist, when speaking of economics and the depth of complexity associated with it, once said "Imagine how hard physics would be if electrons could think…". In the old times, computation was severely limited to analytical solutions, simple addition and subtraction, etc. Up until World War II, even engineers were doing calculations with mechanical rulers and associated tooling. But with the advent of modern computers, large-scale calculations have become feasible, and there is almost no need to refuse using a complexity-first approach and then finally integrate Agent-based Modelling to the problem analysis.

In the following parts of this section we will discuss some limitations of the standard approach to economics, challenge how those limitations are met in the Agent-based Modelling method. After that, we will discuss limitations of our model, issues with memory, computation time and possible solutions. Ultimately, we will discuss what sort of mathematics will be used, and what we want as the end result. This will conclude the introduction part and will open way to the discussion of the practical aspects of the model that we will use, where we will continue discussing specific characteristics of our model as opposed to a general overview which we are opting to concentrate on this section.

Limitations of the Axiomatic Approach

To begin this discussion, we will first revisit one of the most important assumptions in the entirety of economic theory, namely that assumption that the demand-supply curve is one of the key elements of the axiomatic approach. This demand-supply logic does have a major flaw, in that there should be virtually zero information embedded in the price of the product that could indicate something about the quantity of said product, otherwise this would correlate demand and supply on a different manner, not that of the probability distribution (Grossman & Stiglitz, 1976).

The modern approach to explaining - economic activity - stresses on the concept that the macroscopic economical structures are formed as a result of microscopic behaviors. However, this procedure has made significant modifications to the original basis of that in the physics world. Whereas in physics, the motion of atoms does indeed aggregate to the motion of solid bodies through various laws such as that of Euler and Lagrange – the behavior of individual components in economics do not directly follow the same logic. However, let us first analyze the direct issues that plague this approach. There are two major issues with this approach,

1. The degrees of freedom problem
2. The social choice problem

In physics, and the developed fields within it, when we are faced with a problem we move from a bottom-up approach where we identify the situation and the nature of the problem. We identify the field best suited to dealing with it, and then approach a systematic identification of all constrains and degrees of freedom of that problem. Ultimately, we provide a calculated statistical probability and accuracy bounds. In economics, this approach will not work at all, and this is because the behavior of the individual elements that make up the economic structures is assumed to be rational a choice-based behavior whereas in physics the behavior of micro elements is a law-based behavior. Ultimately this makes physics and other fields associated with it acceptable to be treated with an axiomatic approach but ultimately renders economics and our processes as unusable.

The second issue as we identified is the social-choice problem, in which the micro elements of the economic world make decisions based on a collective good, with averaged-out behavior that ultimately demands reaction of the micro-scale to the macro-scale and proper rational choice. In a physical world, this is not taken into account, as the assumption is that the global structures that form are a direct result of the interplay between the microscale as submitted to laws of nature, whereas man behaves well accordingly to the laws of man.

Gallegati in his seminal work discusses the issue of the axiomatic nature of modern economics, and he voices his concern in the very fact that an axiomatic approach is generally based on un-realistic or unverifiable axioms (Gallegati & Richiardi, 2010). He goes on to discuss how supporters of this perspective claim that these abstractions are more than necessary since the real world is complicated and therefore it is simplifications like this that can lead to a better economic knowledge. Although it is agreeable that simplifications of certain interactions and parameters do indeed prove very useful on many cases, it is the complexity of the economic nature that would require more detailed analysis for better and deeper understanding.

Ultimately, according to concepts of Adam Smith and further work by Walras, Arrow and Debreu we have a situation where the axiomatic approach creates a general equilibrium which in itself has many inconsistencies.

One of the biggest problems of this general equilibrium model is that all transactions are done at the same price vector (Gallegati & Richiardi, 2010). This assumes that all buyers and sellers are adjusting to changes in a costless manner and there is an infallible *auctioneer* who is doing the arbitrage until the market settles. The problem with this logic is that no matter what one chooses to model GE, it appears that under the GE assumptions it is always the price that precedes the transactions whereas in real life it is the opposite. By this sheer fact, it is enough to discredit GE as being a scientifically correct explanation of economical phenomena (Gallegati & Richiardi, 2010).

The Complexity Approach

As discussed in the previous section, we saw that one of the bigger problems of the axiomatic approach is the GE model and the inconsistencies that arise within it. Specifically, the issue with the very limited interaction between individual components of the micro-scale. According to Schumpeter under the GE model and the neoclassical approach, we have a homogeneity among all agents of the microscale and any individual strategy will ultimately produce an error that could break the Arrow-Debreu general equilibrium model – this is a significant assumption that can be tacked appropriately with the introduction of the agent-based modeling system.

A very complex system is characterized primarily by the interaction of its smaller subcomponents. Consider a family unit with the various members that compose it. Each member spends some amount of money daily and that ultimately accounts to the daily expenditures of the family. On average, if the family is spending more than it is making it is not up to the family unit to fix this issue, it is up to the individual agents that make up the family unit. In this situation, the family unit is only an abstract clustering of a group of agents that behave individually – perhaps, and even desirable, to act upon their own individual interests. This sort of economic behavior can be seldom modeled with the neoclassical perspective, namely due to the many factors that govern the intricacies of a family. While it is true that if all agents were considered homogenous in all various parameters, and considered equal in interests and activities, it might have been possible to compute an approximate answer using models from the neoclassical approach – in reality this is far from feasible. Ultimately, to solve this issue we turn our heads towards the agent-based modeling method.

Theoretically speaking, the complexity approach - i.e. the agent-based model – does relieve GE and its associated models of the neoclassical approach by discarding the very foundations that they are built upon. Instead of asking for the deduction of an equilibrium price vector and requesting a proof of its existence, agent-based modeling does indeed make way for constructing said request from some logical algorithm or law. The very act of computing said result does indeed require that we model the various goals and individual strategies for each agent which ultimately goes on to solve one of the biggest problems that we discussed earlier, namely that of the **degrees of freedom problem**. Afterwards, it is possible to incorporate various distributions of ideas and strategies and simulate further development of the situation as to provide logical consistency to the tested hypotheses. This fact, does indeed solve the second issue of the Axiomatic Approach namely the **social-choice problem** (Epstein, 2006).

This research methodology does indeed benefit from the agent-based modeling. ABM is a methodology that on its own allows the researcher to create a rule-based, reactive simulation of various actors (agents) that behave accordingly and react to inherent patterns that are created as the simulation runs. The primary characterizations of agent-based modeling are:

1. There are a significantly large number of objects that exist within the simulation and interact with each other.
2. All the existing objects in the simulation are autonomous – which means that there is no existing top-down control over the agents
3. The outcome of their actions is numerically computed and is subject to laws of computation

These objects, as we discussed in the previous bullet points, are synonymous to agents, and they are usually referred to as agents. An agent within the simulation can be anything we want to represent - be it an atom, a metal bar, a human trader, a family or an entire government. The only request is that we select an appropriate metric and space where we model the behavior of our agents and we expect that the agents have the ability to act within that space, and also react to external stimuli ultimately formulating an interaction between agents and other objects within the simulation. Other entities such as firms, governments, policies, markets and regulatory bodies can be easily modeled as agents as well, but also as supplementary data structures that could assist to an easier computational perspective. (Gallegati & Richiardi, 2010)

In the neoclassical view, when we deal with economic problems, we approach them from the top-down perspective by assuming natural axiomatic laws and appropriately assigning those mathematical constructs to the economical behavior. In the agent-based modeling methodology we assume a bottom-up perspective where we model interactions between agents, attempting to model them as simple as possible, and ultimately, we get very complex structures out. We allow adaptation as part of the individual strategy for optimal choice, and therefore, we can notice change and approximation towards global optima.

Aggregation of these micro laws ultimately leads to larger constructs in the solution space, that more often than not provide emergent behavior and statistical regularity. Consider for example the Cellular Automata – Rule 22 – that is one of the more famous – from chaos to order – examples. Rule 22, has the following rule logic,

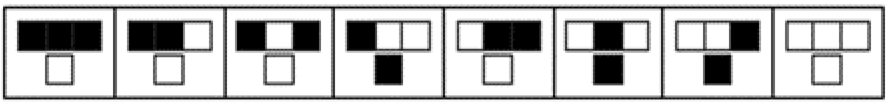


Figure 20: Cellular Automata Rule 22 logic

In this logic, we always analyze a cluster of 4 squares, and depending on their state, we can compute the next state. This very simple rule aggregates later to form the following image,

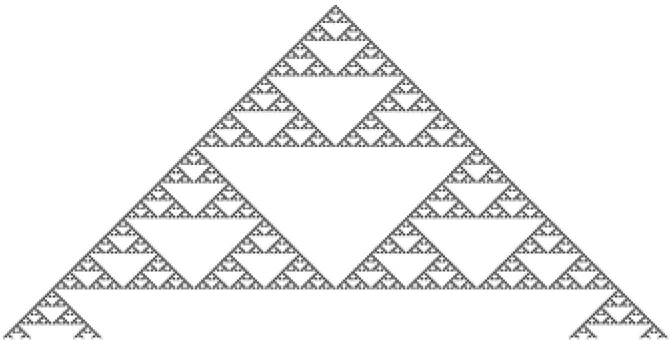


Figure 21: Cellular Automata Rule 22 development

This sort of micro-to-macro behavior is highly expected in agent-based models and other derivative work. By working from the bottom-up perspective, we can identify patterns and behaviors not only in the micro-scale, but the meso and macro-scale accordingly. This is one of the most powerful features of the agent-based modeling methodology because it provides information on all levels.

Fundamentally speaking, the agent-based research methodology is one of the most refined methodologies as it provides a reconciliation of two philosophical ideas (Gallegati & Richiardi, 2010):

1. Methodological individualism
2. Holism

Since we compute behavior of specific agents, and then aggregate said behavior to more complex structures, we can clearly see that agent-based modeling does indeed follow a methodological individualistic perspective – which in itself is a philosophical concept aimed at explaining economic and social features through analyzing its constituents individually and at that concept embedding the idea that the whole is nothing but the sum of its parts. Ironically though, the sheer ability to reduce from larger structures into smaller sub-structures does not necessarily mean that we can re-construct. For example, we have already reduced the entire universe to a composition of atoms and we have a very good grasp of the various elements that constitute the universe – yet we are unable to answer how we could construct one, or even less, how was this universe constructed in the first place. Therefore, we can easily conclude, that reductionism does not necessarily imply constructionism.

To explain how agent-based modeling instills holism in its substrate, we can look at another example. Consider a family unit with grandparents in it. Let us say that at some time $t = t_1$ we assume that one of the grandparents is ill and depending on the severity of the illness the changes to this family financial structures will vary considerably. Now let us assume that this family owns a business, and therefore these changes will reflect upon this business as well, and by logic, upon the financial situations of the employees of said business, and then by continuity to the families of said employees. This sort of nonlinear behavior is expected and very much tackled in the agent-based research methodology. While this nonlinear behavior can be easily assessed in an agent-based model, it would be near-impossible to solve analytically or from a top-down perspective. Finally, this concept of nonlinearity goes on to accumulate and form structures that would be otherwise intractable by simple mathematics. This is also one of the principal ideas of holism, that states that the whole is larger than the sum of its parts.

Limitations of the Complexity Approach

Where we discussed many of the benefits of the complexity approach, and we gave a rather broad outline to the concepts associated with it, it is the objective of this section to provide a few caveats and issues involved in designing, running and evaluating an agent-based model.

Computational Limit

One of the biggest issues that remains to this day, even with all the technological improvements, is the issue of computation. While it is true that computation has significantly improved over the last decades, we still have certain bounds to computation, and we must be careful as to not cross them. To illustrate this problem, let us assume we have 4 people, and they are trading one towards the other in a sorted fashion.

$$t_1 \to t_2 \to t_3 \to t_4$$

So basically, Trader 1 is going to trade with Trader 2, where they engage in economic activity, and then Trader 2 has access to Trader 3, which in turn has access to Trader 4. This problem, is generally represented as a graph problem like this,

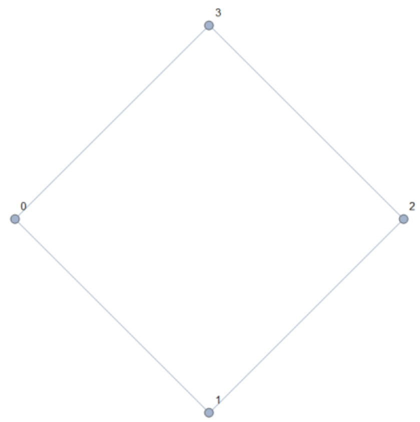


Figure 22: Graph interconnectivity between agents 1-4

This is a simple problem, where agents will interact with each other and then formulate an economic structure. If we assume the frequency of activity to be that of 1 day,

$$f(v) = 1\,day$$

And we agree to use 1 day as the minimum timespan to compute the activities between these agents, we can easily formulate a computer program that would compute these activities as follows:

- Begin day *n*
- Calculate activity for *Trader 1* with *Trader 2*
- Calculate activity for *Trader 2* with *Trader 3*
- Calculate activity for *Trader 3* with *Trader 4*
- Calculate activity for *Trader 4* with *Trader 1*
- End day *n*

There are only 4 activities here, as represented by the blue edges on the diagram above. If this problem were scaled to say 10, 40 or even 100 agents, it would scale linearly.

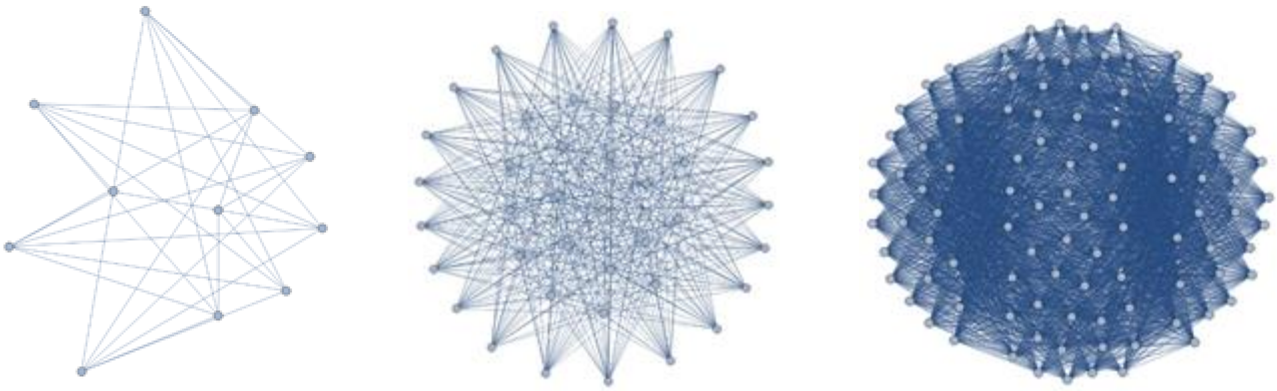


Figure 23: Number of agents increased from 10 to 40 to 100.

While it may appear that complexity is increasing severely, it is only because the number of agents increases from 10 to 40 to 100 in the last image. However, because we would have to do calculations for these agents daily, it would mean that for each day *d* we would need to do *n* calculations. This would scale linearly per day, as we can see that the computational complexity is,

$$F(n) = d \cdot n$$

Where we have d representing the number of days, and n representing the number of agents involved in the simulation. Let us assume for the sake of simplicity that each calculation costs about 1 second, for a computation of 30 days we would have 30 seconds. For a computation of 100 days, we would have around 100 seconds. It is logical to see that this linear scale is desirable in these situations. But now let us look at what happens if we increase the complexity, by allowing every agent to interact with every other agent.

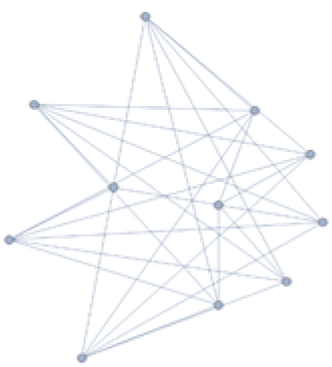


Figure 24: Use-case with 10 agents

Figure 25: Use-case with 10 agents fully interactive

So, you can see that now the number of connections has grown significantly. Earlier, with 10 agents only communicating with their next neighbor the amount of computation was significantly simpler, only 10 of such, but now we have each agent communicating with every other agent, so the scaling is quadratic, that is

$$f(n) = n \cdot n = n^2$$

Now if we acknowledge that each of these calculations must be re-run for each consecutive day, we have the following situation,

$$F(n) = d \cdot n^2$$

Now this scale is significantly larger than the first, and it can grow very fast. Let us say that, much like the other example, each calculation takes 1 second of time, then for one day, for 10 agents, we have

$$F(10) = n^2 = 10^2 = 100 seconds$$

So, it would take a significant amount of time to compute the interaction between every agent interacting with every agent. This growth will be even larger for the next value, as it scales quadratically. For 100 agents it would take somewhere around 10,000 seconds which amounts close to 2 hours and 46 minutes. And this is for only one day of calculation, for 10 days, it would take around 1 day and 3 hours. The issue here is not on the 1 second calculation time, but rather on the scaling of the interactions between agents. There are situations where this scale can be even larger. In computer science there are computations on the trillions of calculations per second, with scaling well beyond the linear scale. This sort of mechanic is very important to be kept in control when designing ABM models especially those with a large number of agents.

Memory and Parallelization Limits

Another very important limit to consider is the limit of the computer memory, issues with parallelization and how much can a problem be scaled horizontally before it would break down. In computer-science jargon, a horizontal scale means adding more computers (or processors) to work together towards solving the original problem. Parallelization by horizontal scaling is usually one of the most used methods of increasing computation efficiency.

There are two major problems associated with memory and parallelization when we apply this to an economic problem.

- The issue of memory and database scale
- The issue of parallelism

These two issues can be fatal to the scaling of the problem and depending on the nature of the analysis and research methodologies, these two problems can pose a very serious problem when working with data. While today there are data lakes and scalable databases that can sustain petabytes of data easily, the infrastructure can be very costly, and it might not prove useful enough to scale. Before we delve into further details for each of the issues, we will propose that sometimes scaling the simulation down numerically might be the proper choice. There are other solutions such as ensuring proper data types are being used such as enumerations for categorical data instead of object types which could reduce the dataset memory footprint by an order or two. However, sometimes scaling the model is not an option, then we propose analyzing cloud solutions that could prove very useful as they can run and backup data far easier than an individual can on their personal computer.

The problems with memory are generally associated with the scale of the dataset, for example in our problem we take the entirety of the European Union with some 80 million population size. If we assign each individual some metadata like an identification hash, an age, employment metadata like employment status, profession, profession type and wage and other data, in total it amounts to about 3 bytes of data per person. For 80 million people, we get a dataset of around 240 megabytes. While 240MB is not a very large number for today's infrastructure, it goes to show how fast the data grows in space. Adding only one extra piece of information to the agent, that would amount to a miniscule of 1 byte per person, the next time we serialize the dataset we will have a 320MB data. That is a huge increase in size. Therefore, it is important to notice that the memory problem is one of the issues that can plague an agent-based simulation.

The other problem as discussed earlier is the parallelization problem. This problem is a little bit more of a conceptual problem in which we look at the issue of scale-locking. Let us say we have clusters of agents like this

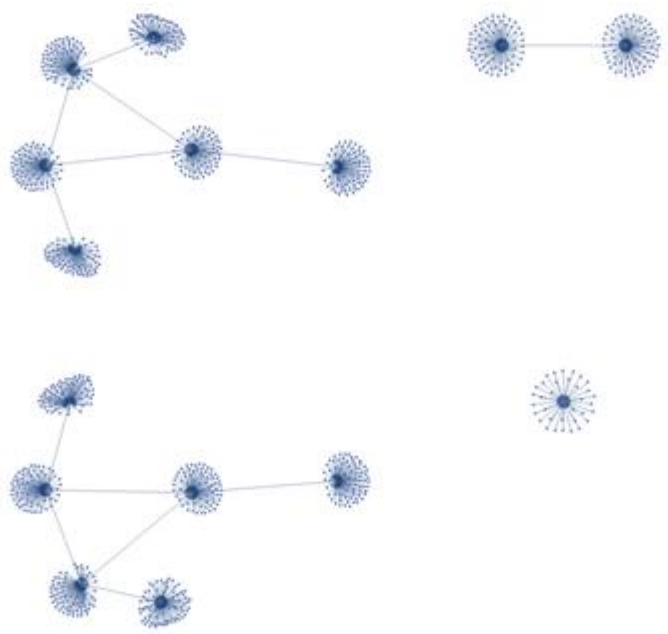


Figure 26: Clusters of agents

And let us assume that each cluster of agents is some sort of independent calculation, be it a trade exchange on various markets, economic exchanges between various countries, etc. This sort of problem is very easily scalable horizontally. By evaluating each cluster on its own, we can evaluate the larger clusters highlighted here in red, in a parallel method, and then do a release of resources and evaluate inter-cluster calculations.

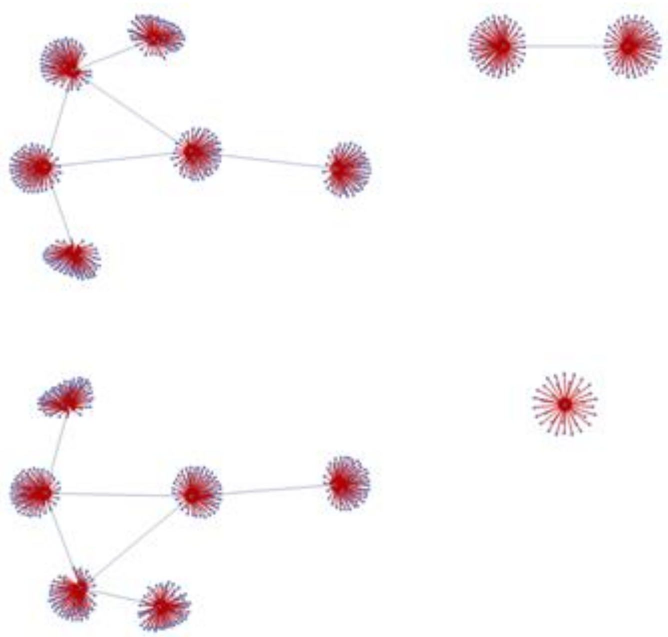


Figure 27: Clusters and their identified computation zones

However, the problem is what if we have a situation where each agent is randomly connected to another agent. In this case it will be impossible to state what sort of clusters can be made for parallelization or perform any heuristics. To combat these issues, it is important that during the setting up phase of the problem, we design the solution to be as expendable as possible. If it is possible to split activities in groups and if possible, maintain filtering views that could allow calculation to be done on smaller subsets of the data that could perhaps speed up the calculation in general.

These issues, namely the memory limits and horizontal scaling can become significant if we are working with large datasets and if the calculations are very tight i.e. if one action depends on a large number of agents, and if one action does change the behavior of many agents. If this issue does indeed become significant for the problem that you are tackling it could be a very significant design issue, that should be fixable if some amount of re-design is done.

Implementation

In this section we will discuss the specifics of the implementation of our program. Throughout the following parts we will analyze in detail the many design choices that we have made when designing our agent-based model and explain the logic behind them. We will outline the various assumptions and constraints with their appropriate relaxants.

Our primary component in the simulation is the *Entity* which is the highest, and most abstract, definition in the hierarchy of our simulation. An *Entity* serves as an interface, primarily to define various behaviors that the other entities in the simulation have, but also to provide a common ground for the logic behind. I.e. we assume that all behaviors are traceable, therefore they must be attached to some uniquely identifiable string.

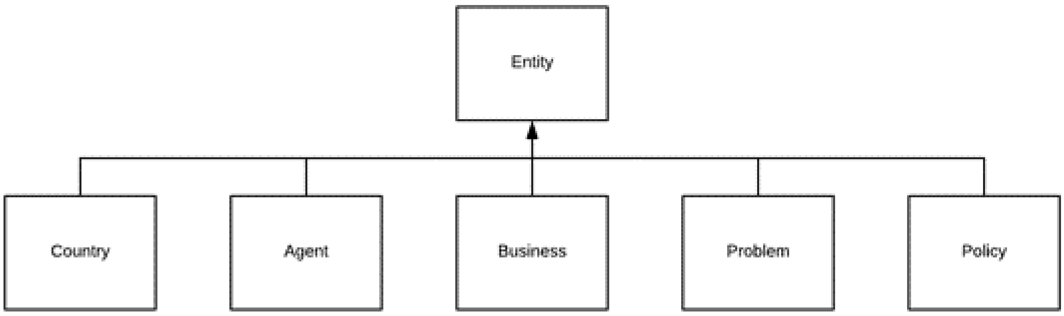


Figure 28: Entity and its concrete implementation (some entities have been hidden for clarity)

The entity can have many implementations, such as the *Agent* which represents a person, the Business which is an entity on its own, yet composed of many Agents. The *Problem* entity which represents a problem that Agents are engaged with during their work hours. These entities will be discussed in more detail in their own sections.

We designed the simulation such that the objects therein are the most important pieces of the simulation. Consider the *Business* entity for example, while it has some aggregated values on its own, it still is only a container for the *Agents*. However, the *Problem* entity is an entirely different construct. Whereas the *Business* entity is a container of many *Agents*, the *Problem* entity is an entity of its own and it represents a business problem that the *Agents* deal with, on an hourly basis (which is the smallest unit of time that we simulate on).

We will discuss about our time tracker in its own part, the *Clock* entity, which is a special kind of entity which we recommend to the implementor to be a static object, initialized only once per simulation, effectively following some form of the Singleton pattern.

Although throughout this paper we will provide various ideas, and design patterns that we recommend when implementing this simulation, keep in mind that it is up to the designer and implementor to create an optimal implementation with respect to their own situation, i.e. for an implementor who is working in an object-oriented programming language, the design might be different from someone who approaches modelling this simulation in a functional programming language.

Nevertheless, we have attempted to make the recommendations as general as possible, as to make the work of implementing this simulation as easy as possible in any paradigm. It is also very highly recommended that implementing this model is done in some form of a programming language as opposed to some ready-made software such as Excel. Due to the complexity of the data that is represented, it would be extremely difficult and very inefficient to model something like this in a program that is limited in flexibility.

Primary Assumptions

We make two primary assumptions before defining this model. The assumptions are as follows,

- All agents are rational agents
- All transactions are completely transparent

These two principles are very important to the proper functionality of the model, because first, we expect that agents act rationally, and any other implementation would be more difficult to simulate and more unrealistic in large scale behavior. Whereas the first principle is founded upon the proper functionality of the simulation and behavior of agents, the second one is founded upon a “wishful” thinking one could say. In the parameters of the model, this transparent transaction system is done for two reasons

1. Bulk of data
2. Waste of computational resources

These transactions make the bulk of our data, which means we draw conclusions and other studies upon this data; hence we want to be able to have as much as we can quantity-wise. And secondly, since agent computations do take some amount of computational resources, dropping them or hiding them from the system does take some value out of the computational process.

It is important to note, that it is possible, under the design of our model, to install a system where certain entities do hide transactions in an attempt to model incriminating behavior such as tax evasion, however at this time we do not recommend any attempts such as that, especially under the parameters of the current model because it would only cause more harm than good due to the fact that our model is currently unfit to handle these extra-ordinary circumstances.

We expect, and ultimately, we have programmed into the agent behavior, a rational choice where depending on certain variables the agent will make decisions that will ultimately work out to the best of his interest. These choices will be made by the agent depending on a few variables, but all these variables will work together to aggregate a single *Happiness* variable which will serve as the primary drive for the agent, i.e. the agent will seek to maximize the *Happiness* variable. While we have named it *happiness* here, it does not necessarily represent human happiness, it only represents an abstract perspective of happiness from the economical viewpoint. We will discuss a bit further on how we calculate happiness and how do we calculate its fluctuation during the simulation of the day.

Entities

In this section we will discuss a bit further about certain entities

Clock

The clock entity is one of the most fundamental entities to the model and simulation. It runs the time of the simulation and keeps track of changes. It is highly recommended that the clock is a static entity, following the Singleton pattern. However, it is possible to have multiple clocks running for better performance, but special techniques must be employed to synchronize these clocks. For those who want to implement a multi-threaded version of the simulation should also take care synchronizing the clocks, and also implementing a thread-safe version of the singleton clock if they choose to use a singular clock.

The clock contains many parameters that can be used. For example, the clock is linear in growth in terms of hour count. It can be used to also track day count, by modulating the linear hour count by a constant value. One could track workday count, or holidays.

All these parameters should be handled by the Clock entity.

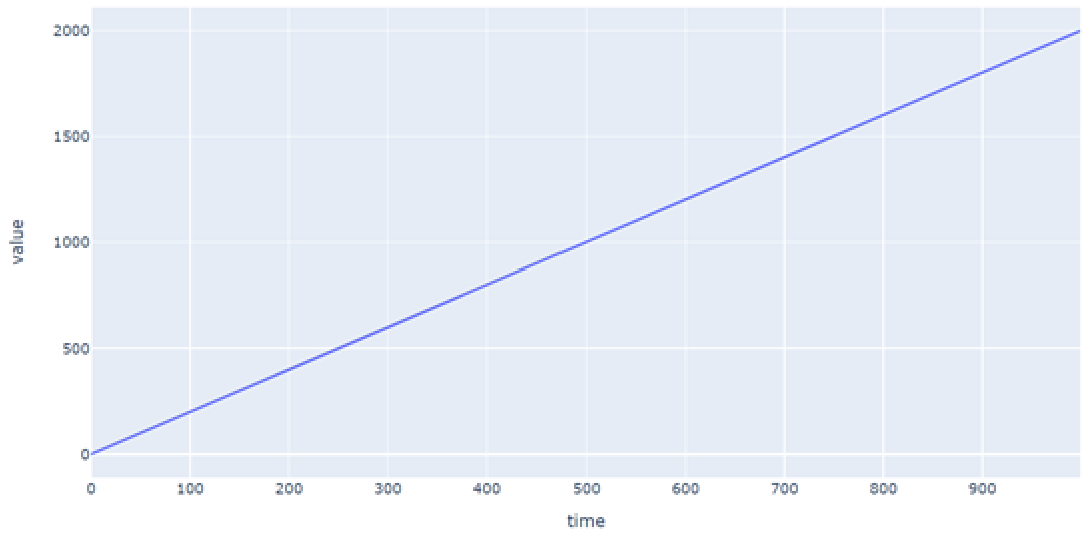


Figure 29: Hourly progression of the Clock entity

The figure shows the linear growth of the hourly count (a proportion of it). The hourly growth can also be mediated and modulated with a constant value to get daily, weekly and monthly.

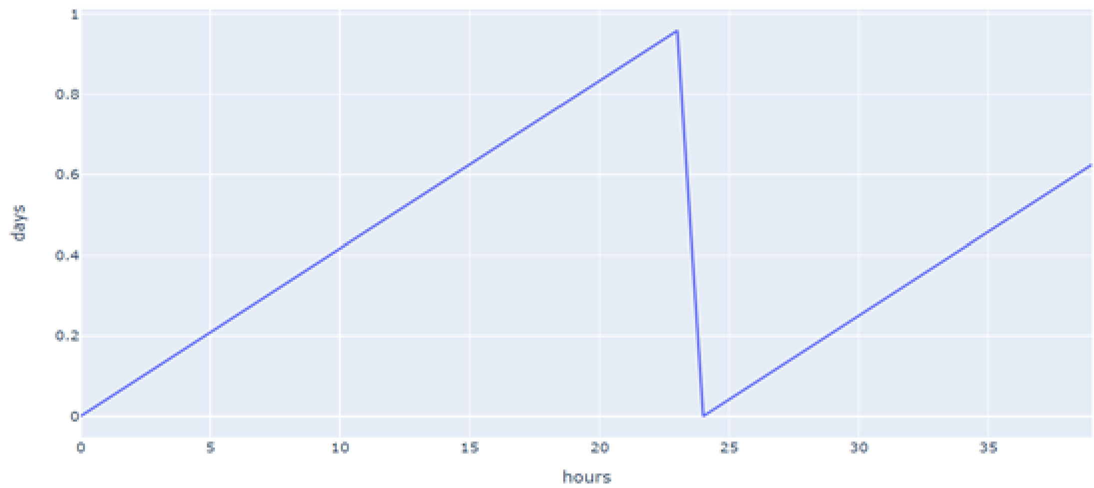


Figure 30: Daily progression of the Clock entity

We can use the daily progression of the clock entity to count days as they occur. The hours grow linearly, whereas days are divided by 24 and modulated by 1. These counters are extremely helpful when we want to pass policies later.

We also recommend another counter, to be implemented by using the Clock parameters, in order for the simulation to have information regarding weekly movements. These would be important for simulating religious weekly holidays, passing legislation that would enforce such things. The same goes for enforcing legislation for time off work that would be tracked weekly.

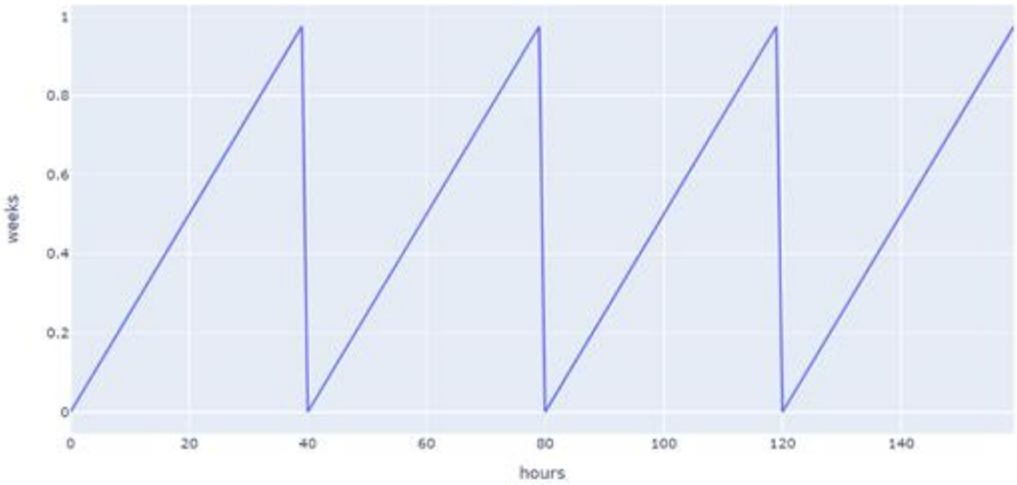


Figure 31: Weekly progression of the clock entity

Agent

A Person will represent the fundamental unit that we create to represent an element of the workforce. The *Person* represents a human being part of the 15-75 age group, capable of doing work, legally getting paid and getting into binding contracts. This person will amount a percentage of the workforce for a country and will contribute by paying taxes and respecting unequivocally other policies set up by the country.

Each person has the following general information that serve as variables for said person, these are:

- Name (identification string, 12 characters, randomly generated hash)
- Age (integer value, bounded 15 – 75)
- Geographical Metadata
- Labor Metadata
- Education
- Skill
- Capital
- Happiness
- Fatigue
- Pressure

Now we will discuss each of the following variables in length and describe any assumptions behind. When a Person is generated, we do some pre-computation to make sure that this person fits a general description. For example, a country sets forth some various bounds such as Age distribution, etc. Also depending on the Country's GDP and approximate pay structures a Person's *Capital* will be generated accordingly. Pay structures may vary throughout the simulation due to economic dynamics.

Also, the entire Labor Metadata will be generated according to information that is fed a-priori through a Country constructor.

We have designed the system like this with the aim of generating accurate representations of workforces within selected countries and ultimately within the EU. For generating people in the EU we have made sure that the variables can be located on Eurostat and by being able to identify these variables on open databases online we designed the system such as that by feeding those data to it, it would generate the agents accordingly.

These variables are generally embedded in the Geographical Metadata and Labor Metadata of the agent.

Name

To generate the Name for the person we use the universally unique identifier (UUID) algorithm to ensure the generation of a randomized name, that can be seeded beforehand for replication.

The process of generating the name goes through a specific logic, where 6 initial bits are used to represent version and variant system while the other 122 bits are used to generate uniqueness.

This ensures that all the possible combinations amount to

$$C = 2^{122}$$

Which is a very large number, and for our case, we can assume that there will be zero collisions practically speaking. Further literature could be analyzed for collision assumptions, but the general idea on one of the more conservative assumptions is that the probability to find a duplicate name is about one in a billion.

We will use these names for tracing the same individual throughout the historical update, i.e. follow a specific Person's changes to happiness, skill, migration patterns, etc.

These names can also be easily used as database keys, since they are unique and can be indexed rather easily.

Age

Age groups are among the most important variable that we want to pay attention when generating populations for the various countries. The biggest issue is that when we generate large scale populations, we will need some mechanism to generate the ages proportionally and in accord with the observed age patterns in various countries. For example, in Kosovo the average age is 29. To model this in the generating procedure we will use a normal distribution with preset means and standard deviations, namely,

$$f(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

By using this formula, we will be able to generate the actual distribution depending on the provided μ and σ variables.

These are the age groups for Kosovo, in the year 2016, which we assume to be recent enough for consideration. The distribution of the age groups is very symmetrical with respect to sexes so we can simply use the nominal values for them to generate the entire population. We do not differ the agents by sex therefore taking nominal values is more than enough to generate the entire population.

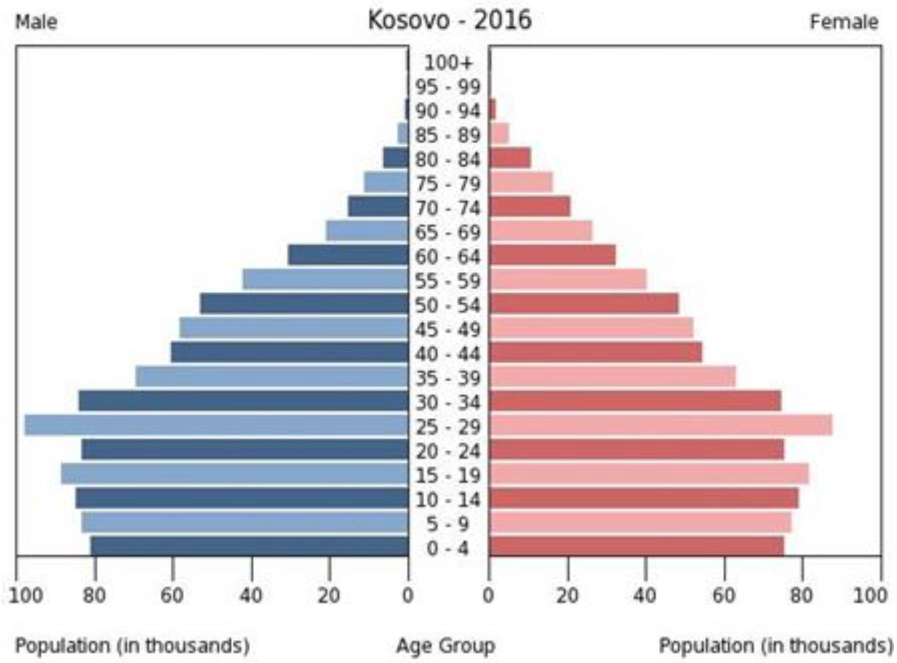


Figure 32: Age groups for Kosovo - 2016

Let us take the case for Kosovo and study the resulting age group and how the randomness engine will generate the necessary data to maintain a proper distribution according to the given data. From the graph we can see that the mean age for the Kosovar agent is about 29 years old,

$$\bar{a} = 29 years$$

And the standard deviation for this is calculated to be around

$$\sigma = 12 years$$

The resulting data for Kosovo is according to this diagram,

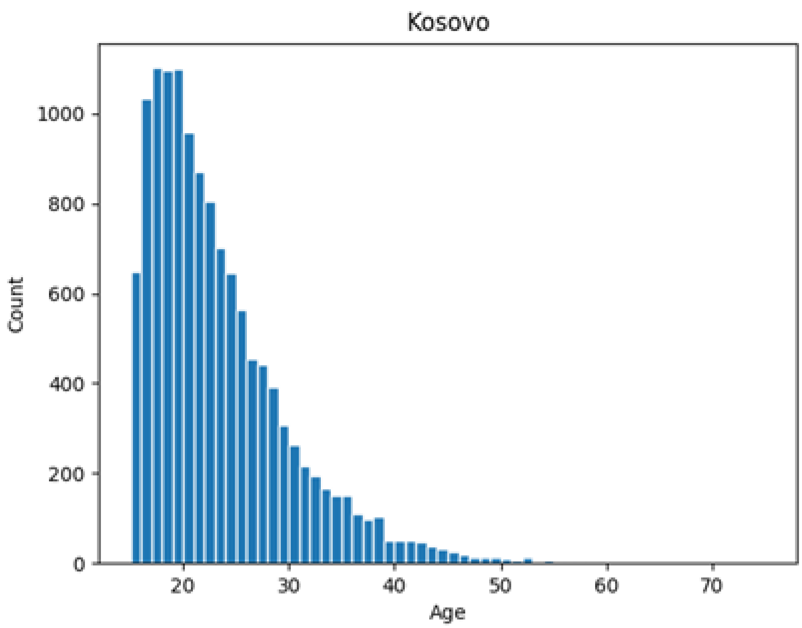


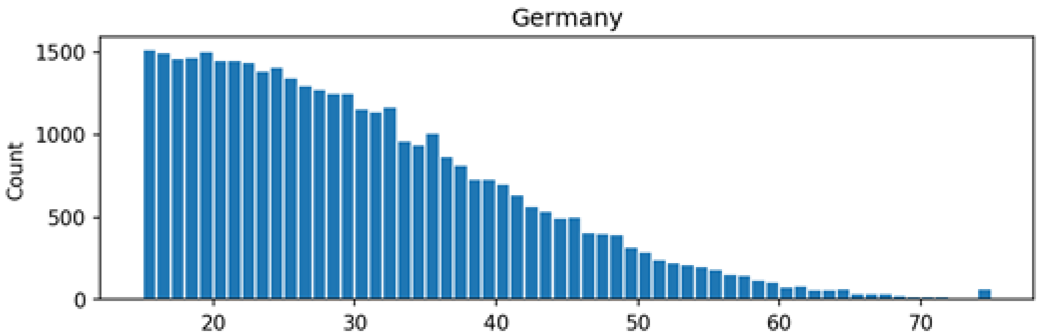
Figure 33: Kosovo age distribution

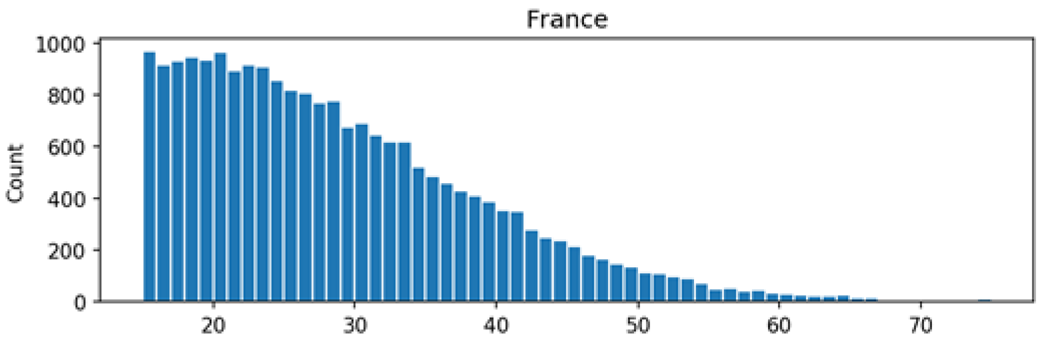
For other countries, the same logic applies, for example for Germany we have different initial values for mean and standard deviation. For many different countries we will generally calculate the mean and standard deviations due to varying factors, primarily we will use data from the internet to estimate a good fit for these parameters as to ensure we have a more proper representation of the population samples in these countries.

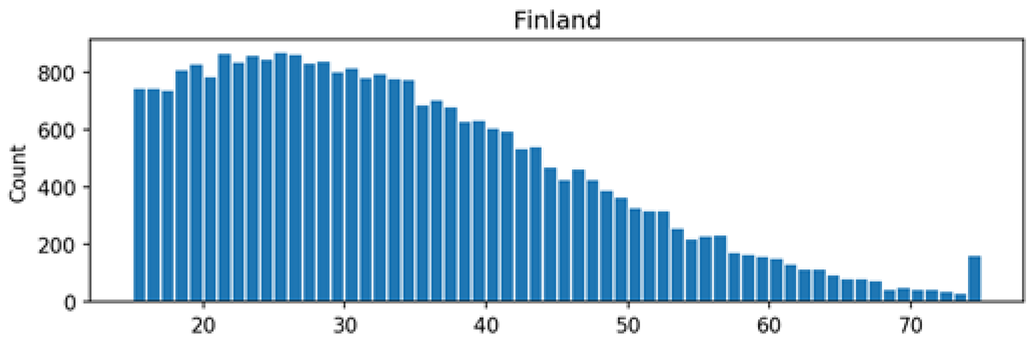
Here are a few diagrams for the age distributions of other countries, namely that of Germany, France, Finland and Sweden.

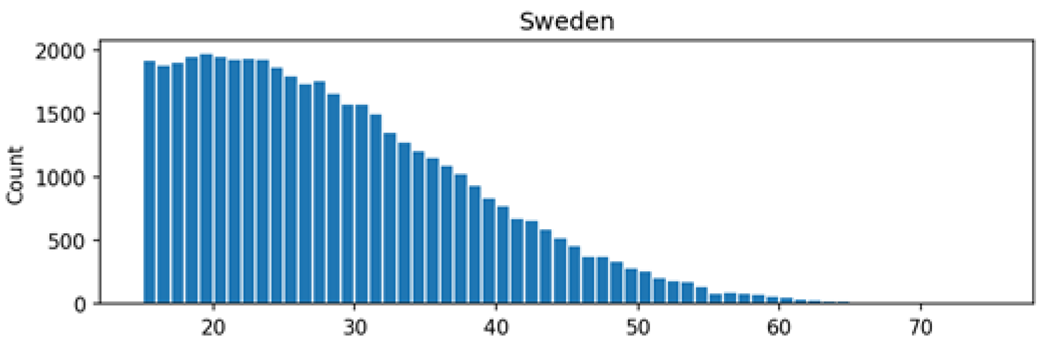
Each of these distributions have population size taken into account spread in various parameters such as a normal distribution, chi-square or sometimes uniformly. The age distribution calculations take into account various techniques, as described in the following paragraphs.

To generate the age groups for a certain country, we initially design the country's age distributions according to information with datasets such as World Bank databases, and country specific agencies of statistics.

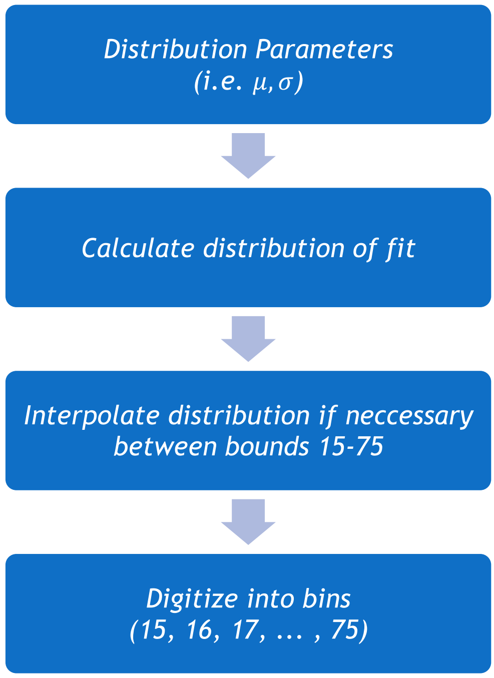








The algorithm that generates the age groups is as follows,



The distribution parameters step is done to ensure that we get correct parameters for each individual country. We know from censuses that countries vary significantly in age distributions, i.e. Kosovo has a much younger population versus Finland which has a more uniform distribution over the 30-50 range. This needs to be taken into account because we want to model dynamics between age groups and since we take age as a factor in determining probable experience levels, we want to compute workforce dynamics to be more accurate.

After we get the parameters, say for Kosovo we have $\mu = 27$ and $\sigma = 12$ we then calculate the distribution, in this case we use a normal distribution with a tighter standard deviation to maintain larger numbers in the 20-30 range, whereas we want lower numbers in the 30+ range. We do this calculation for over 1.3m agents, which we use as a foundation to create Kosovo. Remember, these are economic agents, which means they have no primary demographic data with regard to distribution on other subdivisions such as municipalities or towns. They belong only as residents and country of origin with respect to some Country, in this case Kosovo.

After the distribution is calculated, it is possible for the data to come out outside of the bounds of ages that we need. For example, calculating a normal distribution with $\mu = 40$ and $\sigma = 20$ will undoubtably result in some values well outside the minimum 15 years old age. Because we want to keep all our agents within the age range of 15-75 which is the defined age group, we must do interpolation on the data if necessary. The way this process is done is outlined as follows. When we generate the distribution, let us say we have something like this.

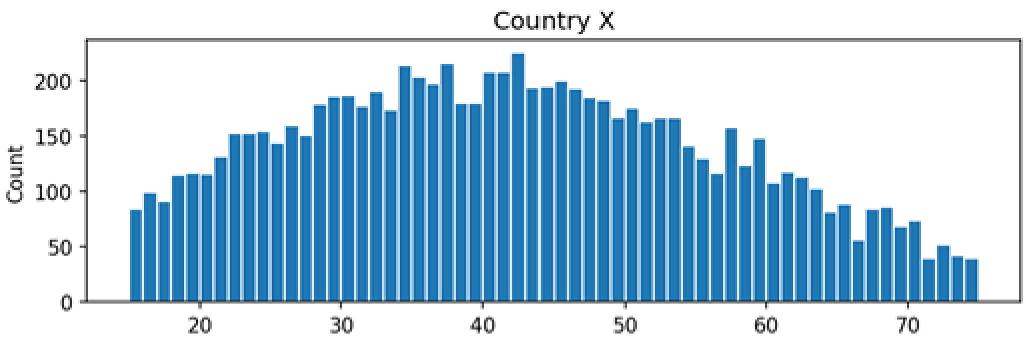


Figure 34: Non-interpolated age groups

And it is obvious that the data skips well above 75 line and is below 15 in some cases. The situation gets much worse, if the dataset is malformed in some way, such as the situation with Kosovo where we have a far stronger emphasis on the 25-30 age range.

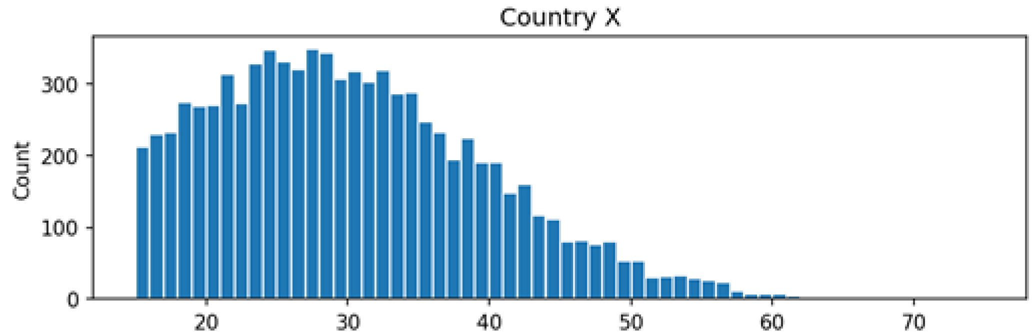


Figure 35: Interpolated age groups

This situation can be remedied by initially picking the data outside of the tolerable range and finding the minimum and maximum values in the entire distribution, then ultimately mapping those values to the tolerable minimum and maximum, basically resulting in a function that maps

$$f:\left(\min X, \max X\right) \mapsto (15,75)$$

This function f is applied over the entire distribution and it fixes the problems that we encounter initially by disproportionately assigning values to ages such as 14 or 80. Once this function is applied it will be guaranteed that all ages that appear in the distribution, the minimum value will be bound to 15 and the maximum values will be bound to 75. This interpolation will guarantee totally that we will have no values outside of these ranges. Another very interesting application of the interpolation technique that we use is that it makes sure to unify the spread of the ages towards the entire domain.

Let's look at the situation where using interpolation is very good, because it spreads the ages whereas in the un-interpolated initial distribution, we would see a very tight spread over a very narrow range. While in some cases, the latter may be more desirable, in our simulations we want to have agents of varying ages.

Figure in the next page,

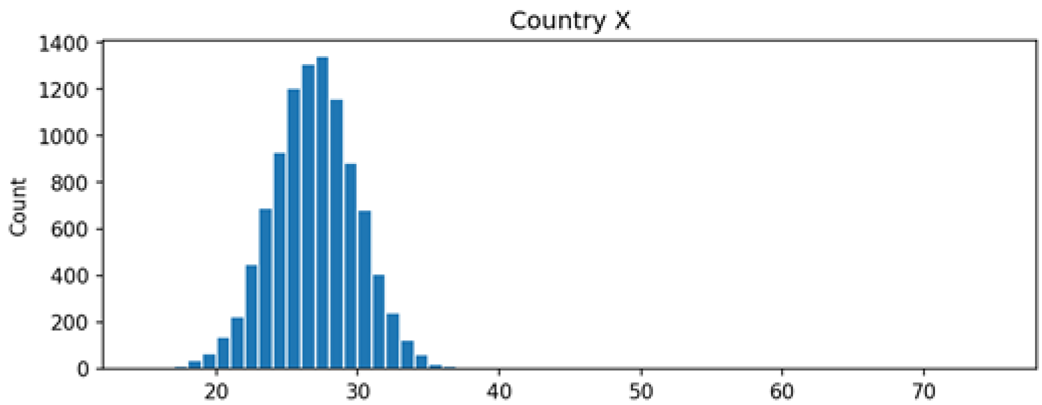


Figure 36 :Strong interpolation, with heavy emphasis on young population

As we can see here, the ages are very strongly focused over the 20-35 range, but virtually zero agents appear on the 40+ range.

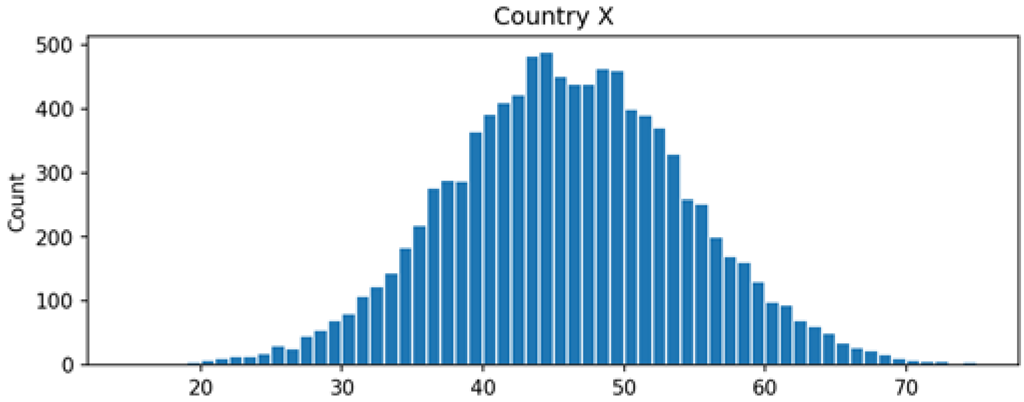


Figure 37: Well spread age distribution

This second graph is the same data, except not well spread over the entire domain. We have a small issue where the mean has moved as well, and to prevent that we cannot use a normal distribution anymore so on certain age groups we commit to other distributions such as a chi-square, beta or some other format. However, as we decided above, we must keep a domain within the

$$x \in [15,75]$$

And even a stronger assessment is that we must have at least one of each age. One of the issues that is now present in this step of the calculation is the fact that when we generate data like this with some distribution $P(x)$ we must make sure the data is being generated in an integer format, i.e.

$$x \in \mathbb{Z}$$

And this is because there are no decimal ages, e.g. 17.2 years or 25.8 years old. As a statistic, ages are reported in integers i.e. 17 years, 25 years, 30 years old, etc. To make sure all our agents also report their ages in integers we must make sure we use some sort of rounding function to achieve that. We cannot just use a rounding function directly because it is possible to round all values to a single integer, which is un-necessary and will only pose problems for us. That is why interpolation also helps in this direction before we do the rounding procedures.

To do rounding, we use a digitization process, i.e. we take numbers and bin them into discrete buckets. This binning process is designed like this,

Table 1. Table of generated and binned ages

Id	Name	Generated Age	Binned
1	2d6b3b0d-d0df-498b-b536-bdf02f038c65	17.34	17
2	7589509c-a08e-4dfa-b86c-bae7c22d83df	21.39	21
3	e4252aaa-63ea-419b-9585-87ef72ba38bf	63.57	64
4	c570c60c-110c-45e8-963d-6d38b7527e02	55.83	56
5	2ac120e4-fc1b-4351-a14a-9b9cf4627e6e	48.72	49
6	3b3f14fd-84ed-4bdd-ae5a-6758a9b846b8	48.54	49
7	3b22e5bf-df58-4306-9275-027a8f84ad54	60.52	60
8	ef20e6f6-4940-4f8e-8012-faa69fc8207a	66.53	67
9	e8e94403-07be-4f15-8f86-9e77b09311e3	41.67	42
10	46167576-fd10-46ba-99d0-8167eb2bad64	72.91	73

As can be seen in the table, a subset of the data (namely 10 agents) are shown here and their ages are generated and binned accordingly, we use rounding to bin values as is well known, that is we round to the nearest integer.

In situations like the agents in this table

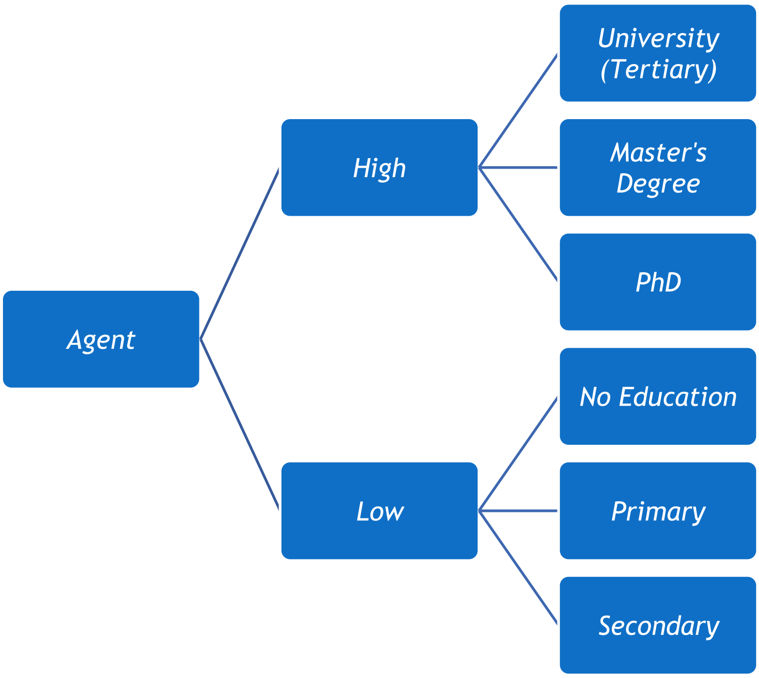
Id	Name	Generated Age	Binned
1	Agent-x	12.30	15
2	Agent-y	85.34	75

Where we have "faked" the data, so to speak, we have generated numbers intentionally out of bounds - even though the second step, basically the interpolation procedure would have taken care of this issue - we wanted to show that binning will also fix this issue by binning the values of 85.34 and 12.30 back to 75 and 15 respectively.

In conclusion the age generation process is well defined and keeps the agents within the hard bounds of 15-75 whereas the interpolation process also maintains a distribution spread well throughout the domain itself.

Education

To generate education, we want to classify agents into varying education classes.



As can be seen here in this diagram, we classify Agents according to the two primary levels of education, namely **Low** and **High**. This is done for two reasons,

1. We want to keep an aggregate variable that allows us to grab high skilled workers easily from the database (namely high and low values)
2. We want to keep the aggregate variable to a minimum (i.e. we can encode the High and Low values by numerical codes 0 and 1) and any operation on them is very computationally cheap.

While there are indeed further subdivisions, and basically the category of **No Education** should not in principle be in the same grouping as **Primary** or **Secondary** education, we reserve the categorization to be only from a technical perspective, that is even though they are named **Low** and **High**, the reader should not assume that this categorization related to the level of education, but rather the economic grouping of these education categories that will later serve for a perception index.

The No Education group maintains that agents categorized under that group have not completed successfully any level of education (be it regular education as administered by schools, but also as administered by a homeschooling process). The Primary education maintains agents that have finished the standardized education process of the K12 education (but without the high school component). The Secondary education category is the K12 + High School education completion, therefore people that complete K12 are people who have an actual High School diploma. These three groups are categorized under the Low category of education, that will be used to group them later in data analysis and processing.

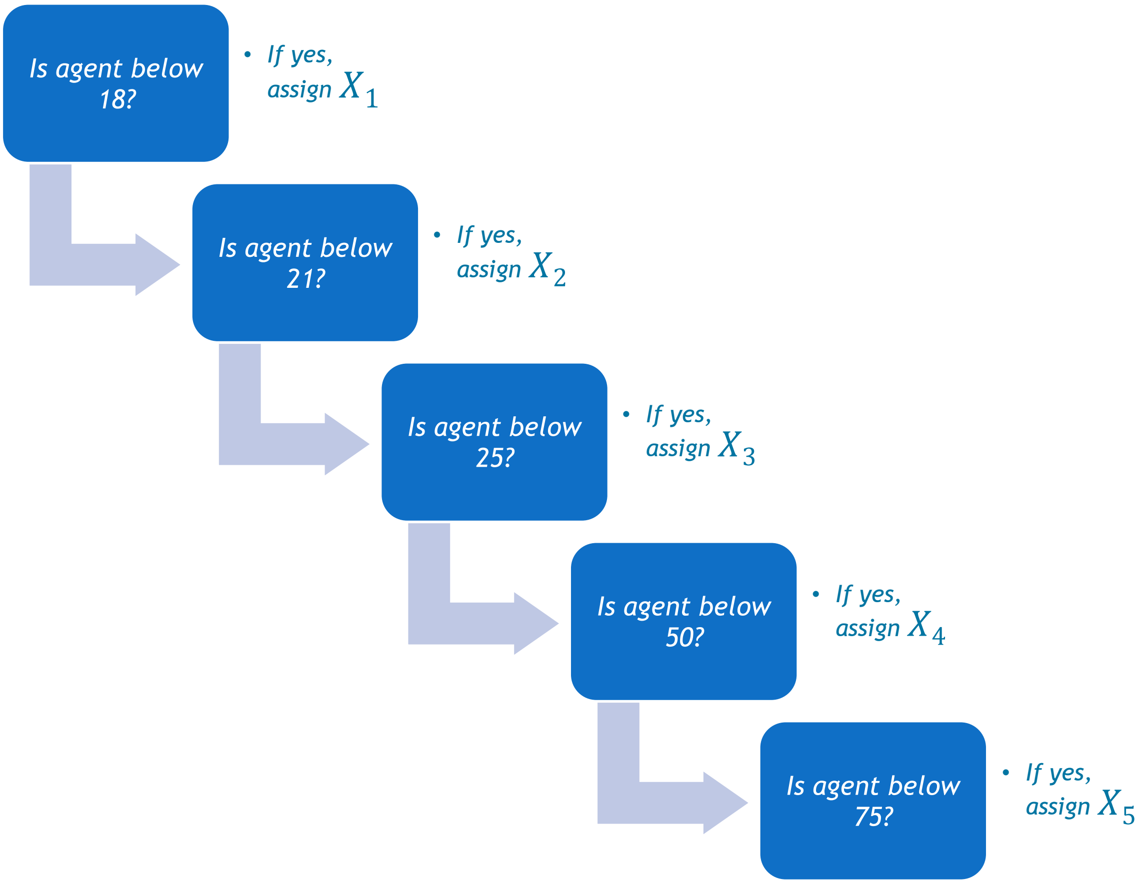
The University (Tertiary) group of Education maintains agents who have a diploma that have finished 100% of the requirements of the credits needed to graduate. In modern countries, basically within the EU, most of the workforce is located in this category. Therefore, we use this category as a degree of freedom in various distributions later throughout the implementation of the Person. The Master's degree is then naturally the continuation of 2 more years of study, and specialization in some field. A rather good chunk of agents is also located here due to data from Eurostat. Ultimately, there is a small percentage of agents that are categorized in the PhD grouping that is still in accordance with Eurostat data.

To try and stay within logical grouping we have made some sacrifices to per-country basis assignment, and have made a general agreement that we will use specific education-to-age variability, because this does ease off a lot of the computation and technical issues that appear with significant difference in education. Second, the education of the agents plays a role in the future simulations, so it becomes necessary to have a simple, tractable decision-making system that both the further simulations and the data generation processes both stand to gain. While it is important that educations are not uniformly spread throughout the entire database, it becomes less important to figure out the exact spread that would model the current European Union today, on a per-country basis. This is done for two primary reasons:

1. Sacrifice of fidelity to gain an advantage in computation
2. Risk of misrepresentation

In the reason 1, namely we choose to sacrifice model fidelity because we understand that the gains in a well-spread education that would in theory account for 100% of the situation currently in the EU does not provide benefits strong enough to outweigh the generation procedures. Second, we believe that it would be practically infeasible to find exactly correct data on the distribution of these educations, or at the very least we would have to make some decisions on our own, i.e. we would still have to agree that group A of some subset of the people, and group B – another subset – belong to the **No Education** category for example. This would constitute that we have to make a decision on our own, which in principle maintains that a probability for error exists. Due to this second issue, we can see that it is impossible for us to get 100% correct data regarding the spread of education on the agents. In conjunction with this last issue, we can correlate to reason 2, namely that of the risk of misrepresentation, where due to the probability of error from our end, or even from the data gathering techniques, it is possible to misrepresent the datasets. To make sure this doesn't happen, we have decided to use a single true value for all countries, i.e. all countries will have approximately the same distribution of educations that we have taken as granted. Furthermore, this distribution also varies by age therefore it is important to discuss the algorithms implied here.

As a standard basis, we know that there are no Tertiary graduates before the age of 21, except on some exceptional cases which we will not be taking into account here in our simulation as we want to model general case fidelity as opposed to fringe cases. Second, we know that the probability for an individual to have a Master's degree before the age of 23 is very low, therefore we assume it is zero. The same logic applies to education of PhD category where we assume people only after the age of 25 can be assigned with a PhD on average therefore, we limit that probability to almost zero in our simulation.



This diagram outlines the process we use to separate agents and classify their probability distributions namely,

$$P(x) = \{a_1, a_2, a_3, a_4, a_5, a_6\}$$

These values are designed according to this table,

Education	Probability (in numbers)	Normalized
No Education	a_1	$\frac{a_1}{a_1+a_2+a_3+a_4+a_5+a_6}$
Primary Education	a_2	$\frac{a_2}{a_1+a_2+a_4+a_4+a_5+a_6}$
Secondary Education	a_3	$\frac{a_3}{a_1+a_2+a_4+a_4+a_5+a_6}$
Tertiary	a_4	$\frac{a_4}{a_1+a_2+a_4+a_4+a_5+a_6}$
Master's Degree	a_5	$\frac{a_5}{a_1+a_2+a_4+a_4+a_5+a_6}$
PhD	a_6	$\frac{a_6}{a_1+a_2+a_4+a_4+a_5+a_6}$

And assign these distributions with values according to

$$X_1 = \{1,10,100,0,0,0\}$$

Where the following table discusses each value,

Education	Probability (in numbers)	Normalized
No Education	1	1%
Primary Education	10	9%
Secondary Education	100	90%
Tertiary	0	0%
Master's Degree	0	0%
PhD	0	0%

This table is valued for ages below 19, that is for 100 agents of Secondary Education, we will generate 1 without any education, and 10 with Primary Education.

This is done to normalize the probabilities such that 90% of the population sample is located within the Secondary Education category whereas the other 10% fall within the combined Primary and No Education categories.

Here is the table for ages 19-21 (inclusive).

Education	Probability (in numbers)	Normalized
No Education	1	0.79%
Primary Education	5	3.96%
Secondary Education	20	15.87%
Tertiary	100	79.36%
Master's Degree	0	0%
PhD	0	0%

In this table we have the introduction of the first generation of university graduates, namely those who are within the 19-21 age. While this does allow for some graduates in 20 age range, we maintain that a large group of approximately 80% will fall within the Tertiary category.

This is acceptable for our current simulation.

Education	Probability (in numbers)	Normalized
No Education	1	0.78%
Primary Education	2	1.562%
Secondary Education	5	3.9%
Tertiary	100	78.125%
Master's Degree	20	15.625%
PhD	0	0%

The table above shows the age group 21-25 where the educations are spread over Master's degrees as well. The agents can now be assigned values of Master and we have agreed to select 20 students out of 100 Tertiary as a Master's degree candidate.

This is done to spread the values a bit more evenly and is based on research about actual average spread of educations worldwide.

Again, we limit the probabilities to zero for PhD candidates because we know there are no PhD candidates below the age of 25 on average.

Education	Probability (in numbers)	Normalized
No Education	1	0.32%
Primary Education	2	0.64%
Secondary Education	5	1.6%
Tertiary	250	79.8%
Master's Degree	50	15.97%
PhD	5	1.59%

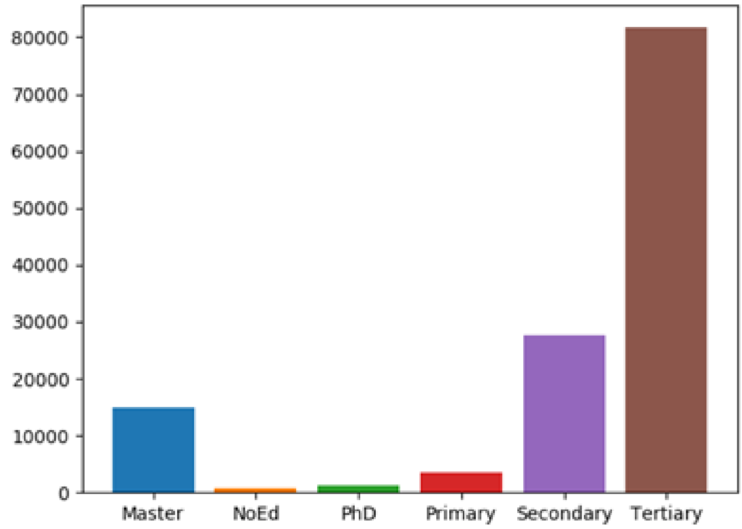
This table shows the probabilities for the age groups 25 - 50, which we assume to account for at least a large majority of the agents within the EU. This design is made to generate about 80% of the workforce as Tertiary, and the other majority is people with Master's degree. The workforce has edge cases with some 1.6% of PhD candidates and a cumulative 2.5% of other groups, i.e. No Education, Primary Education and Secondary Education.

Ultimately, to account for a discrepancy between the older generation (i.e. the 50+) and their lower numbers in higher education (Master, PhD) as opposed to their higher numbers in tertiary education - also due to differences with the Bologna process, we have made another probability distribution.

Education	Probability (in numbers)	Normalized
No Education	1	0.29%
Primary Education	2	0.59%
Secondary Education	5	1.47%
Tertiary	300	88.23%
Master's Degree	30	8.83%
PhD	2	0.58%

As can be seen the educations are directly related to age, which is generated first and is considered to be an independent variable within the scope of the simulation, although we do relate it to actual data for the distribution spreads. The educations are highly correlated to age to create a more realistic representation.

For Kosovo, for example, we get a distribution like this,



Skill

To generate the skill of a workforce agent, we make a few assumptions. The skill of an agent is a continuous numerical value in the domain,

$$s \in [0, \infty)$$

The skill is defined as the ability of the agent to do proper work, and therefore from an employer's perspective, an agent with a higher skill will be more desirable. The skill will also play a role in the 'wage ask 'factor for the agent because a higher skilled agent will also rationally perceive itself as a better suit for most workplaces (that fall within the assigned profession) and therefore ask for higher pay. This tug for balance will drive the simulation forward allowing us to see intricacies between ask and get factors.

To attempt a proper modelling of the skill value, we will use an algorithm that distributes the skill both towards the age and towards the education factors. Because we know that skill is a real, continuous variable, it is possible to use mathematical computation directly without doing distribution mechanisms to allow us to create a good spread of the skill value. To discuss the general idea, we will use Kosovo as an example throughout the algorithm explanation, but also provide equations that can be generalized further.

Initially we begin by taking the age distribution of some country, in this case, we use the distribution of Kosovo,

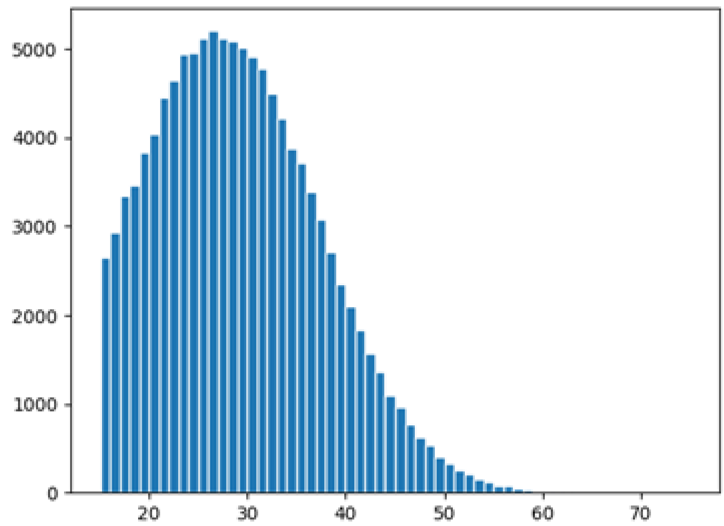


Figure 38: Age distribution for Kosovo

This age distribution is used strongly to define the next formula

$$f_a = -(x - 40)^2$$

Where f_a is the age factor, that is we expect the age of the individual to be normalized.

To map the age factor to a normalized value,

$$f_a = -(x - 40)^2$$

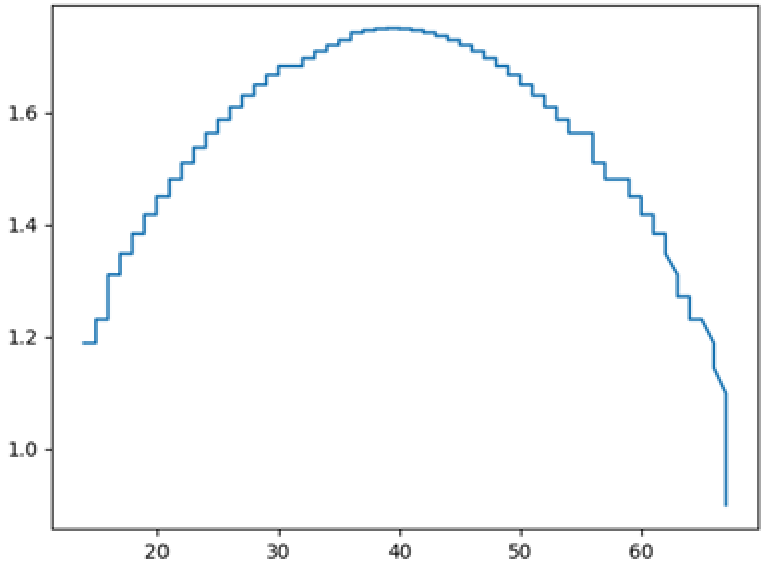
We expect that it is possible to map the minimum and maximum of f_a to varying factor that we select.

$$f_a = \begin{cases} max\, f_a \\ min\, f_a \end{cases}$$

We want to compute the vector $\boldsymbol{x}$ and then map that vector to one of our selected factor groups, like this,

$$f_a : x \mapsto y$$

Where we use the normalized value of $\boldsymbol{f_a}$ and map $\boldsymbol{x}$ to a second value $\boldsymbol{y}$.



As can be seen in this diagram the age factor f_a maximizes around 40 years old, and this fundamentally forces agents to have varying skill dependent on age. First of all, the differences are not that stark, i.e. a 25-year-old will have a factor of 1.4 whereas someone at 40 will have a factor of 1.6 which is namely a 20-percentage points difference. Although the 20 percentage points can get significantly aggregated with other factors such as education and education type. With this aggregation, it is possible to create good splits between the many groups that are located in the education category.

Our second factoring on this continuous distribution is the application of the factors for Education on the skill factor. This is important because we want the probability of high skill to be correlated to the higher levels of education, and vice-versa.

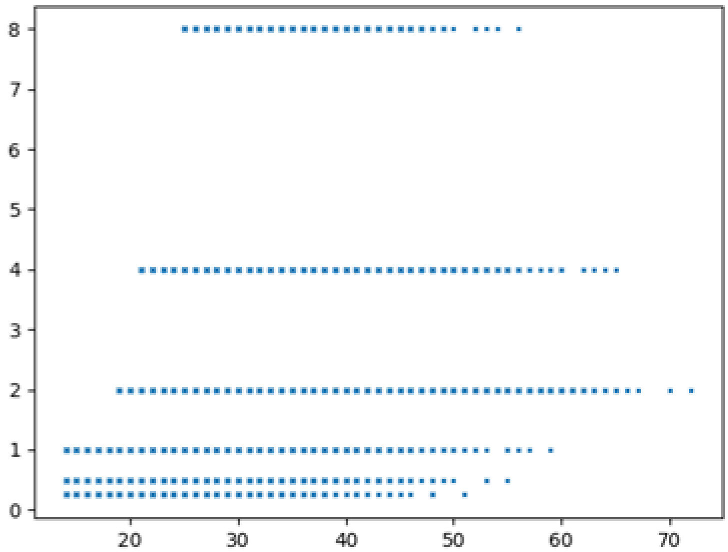
To achieve this, considering that Educations are a categorical variable, we must employ an assignment algorithm, as opposed to a mathematical multiplication. Our idea is to assign power-law in increment, i.e. we use

$$f_e = 2^{ed}$$

Where we code Educations to numerical values and order them like this,

Education	Code	Value
No Education	-2	*0.25*
Primary Education	-1	*0.5*
Secondary Education	0	*1.0*
Tertiary	1	*2.0*
Master's Degree	2	*4.0*
PhD	3	*8.0*

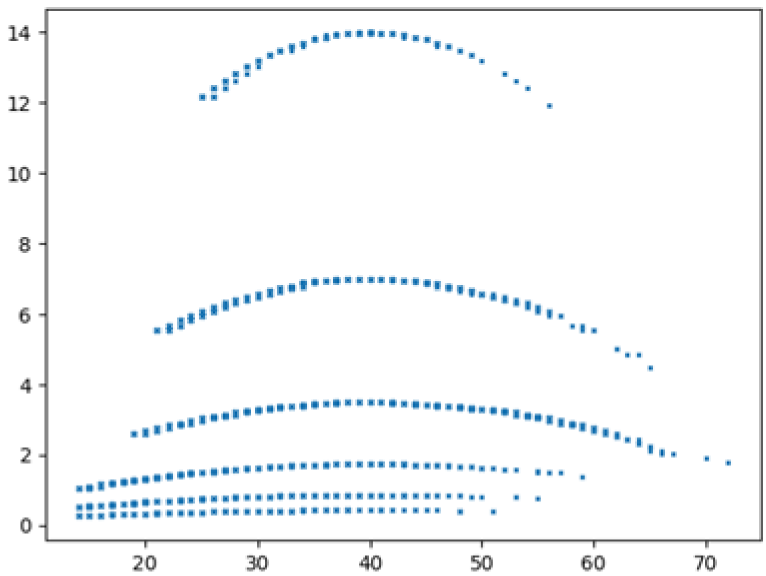
By using these values, we can now apply a mathematical assignment by using cross-tabulating the Education code, i.e. now in numerical format and using the power-law to multiply the factor. We get this distribution for Kosovo,



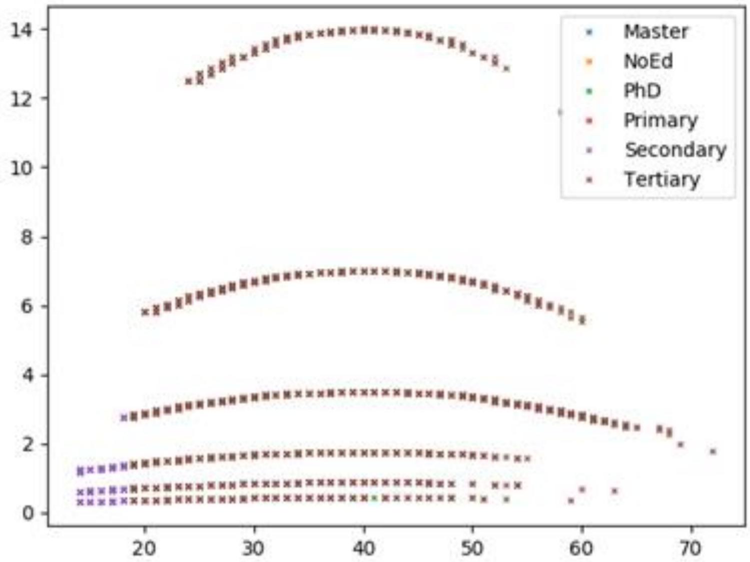
Because this factor directly is very rigid, multiplying it directly with the age factor,

$$f_s = f_a \cdot f_{ed}$$

only scales it upward, resulting in another diagram such as



Which is acceptable in some degree, however it is not yet completely defined as you can see there are very large discrepancies between the varying classes. If we look at this same graph, but with classes applied, it is obvious that there is not much coherence,



To make sure we do not get issues like this, we must balance out the skill level spread over the entire domain we will introduce a random factor that allows for better spread. This random factor can be a debatable issue, i.e. we can either use a uniform distribution over the entire population count, or we can use a normal distribution in an attempt to concentrate the workforce towards some desirable value. One factor that we need to ensure is the issue of negative skill, which is highly undesirable due to the initial condition that this skill is bound by,

$$s \in [0, \infty)$$

That is, we need positive skill only to affect the workforce. To do this, we can look at two graphs for both the uniform distribution and the normalized distribution. What needs to be clarified here, is that the user, i.e. the simulation can be run with any sort of distribution that we want, so there is a general case solution here, however for our case we want to close that degree of freedom initially and then allow plugging in of different distributions later.

Using the uniform distribution is less strong, as one could claim we are only implying a degree of uncertainty, but there are still questions to be asked, for example the PDF is

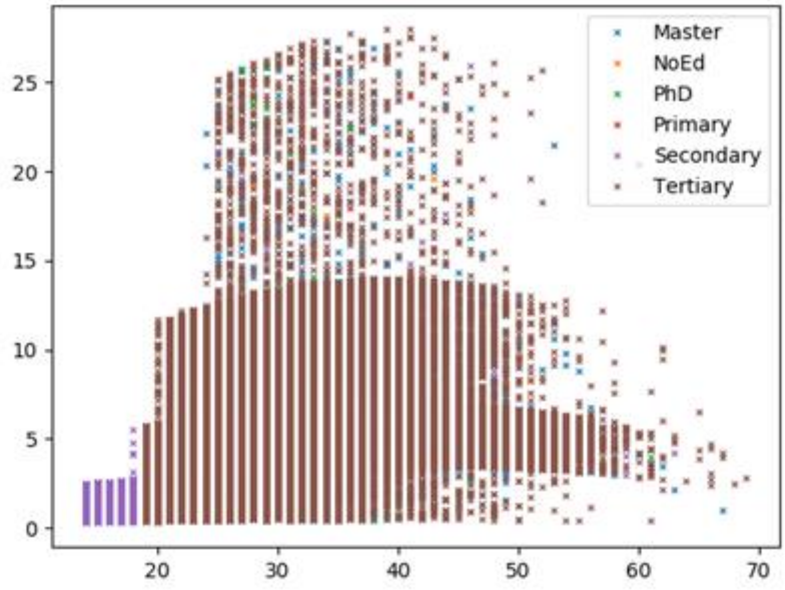
$$P(x) = \frac{1}{b - a}$$

So, one needs to figure what the factors a and b will be. This in itself poses a challenge when using the uniform distribution, albeit the challenge is not that drastic as it is possible for us to assume values between [1,2] and agree that the uniform distribution will only boost them upwards. One could instill a penalizing factor as well, i.e. by increasing the domain of the probability distribution in the

$$x \in [0.5,2]$$

That if we do this, it is possible for the workforce agents to be randomly penalized if the factor falls within the 0.5-1 range. For the 1-2 range it is boosted. This sort of randomization can be applied at any step during the simulation; however, one must make sure that the random distributions themselves make some sort of sense in the context that they are applied.

Applying a normal distribution does make more sense in some areas of the simulation, here being a prime example. This is because skill levels are observed even in censuses to be of normal distribution nature, that is a large group of the population would fall in the average skill, and some tend to be outliers in both directions. While it is important to understand that when we apply this sort of distribution, one must make sure that other factors either are boost only or penalty only because having factors on both sides of the story can cause discrepancies, i.e. one could end up with an abnormally large number of PhDs tending towards the low skill whereas most Tertiary agents would end up in the higher skill tier.



This is what happens when we apply the skill randomization factor, in an uniform method, that is within the following bounds,

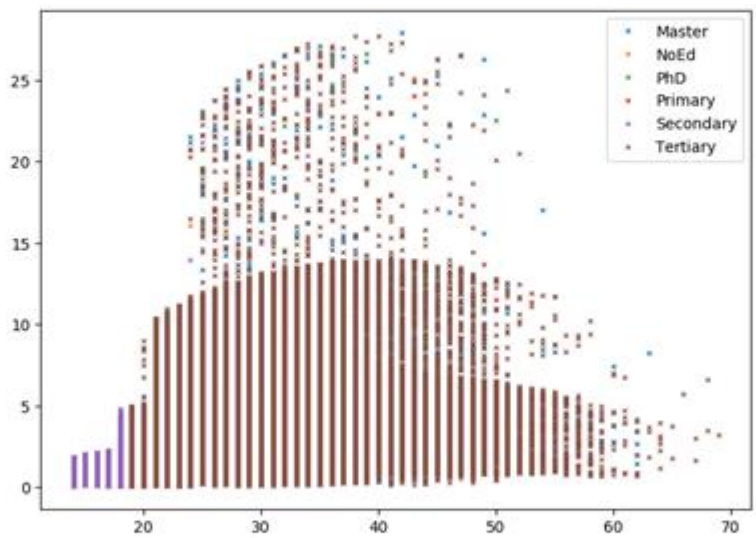
$$P(x) = \frac{1}{b-a}$$

Using like this,

$$x \in [1,2]$$

By making sure that we are only in the boosting factor, the skills themselves are guaranteed to only increase, as can be seen here. The problem is that some agents with "No Education" and "Primary Education" will decrease and then ultimately double down on this, so it is imperative that checks are made prior to database installation.

Using factors like [0.5, 2] is also possible and it results as can be seen in the next graph, with much sparser agents in the 15-25 zone in the skill level. This is done to remove abnormalities such as large groups of Primary and Secondary educations appearing in the 15-25 zone where we would naturally expect agents of PhD level and other higher education levels to appear.



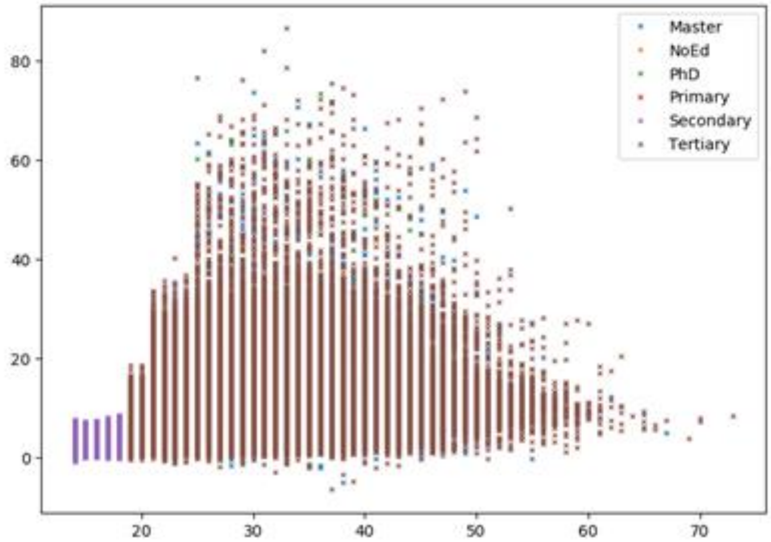
Applying a normal distribution is also desirable, especially in these cases, for example by using a mean and standard deviation of these values,

$$\mu = 3.25$$

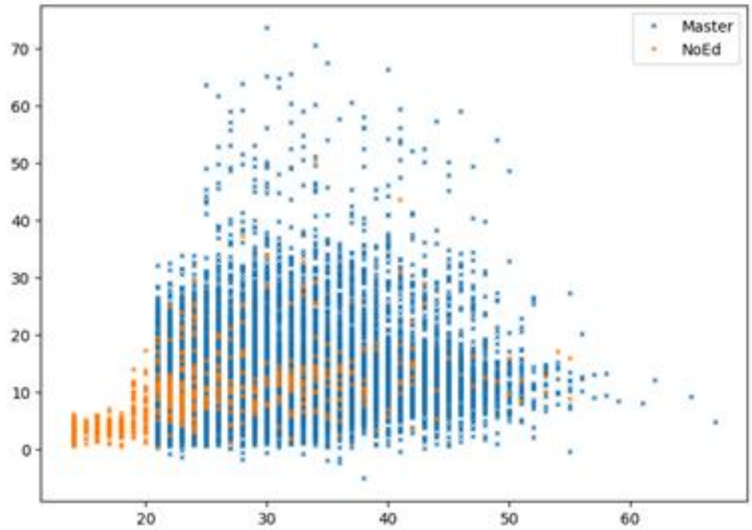
And standard deviation,

$$\sigma = 1.0$$

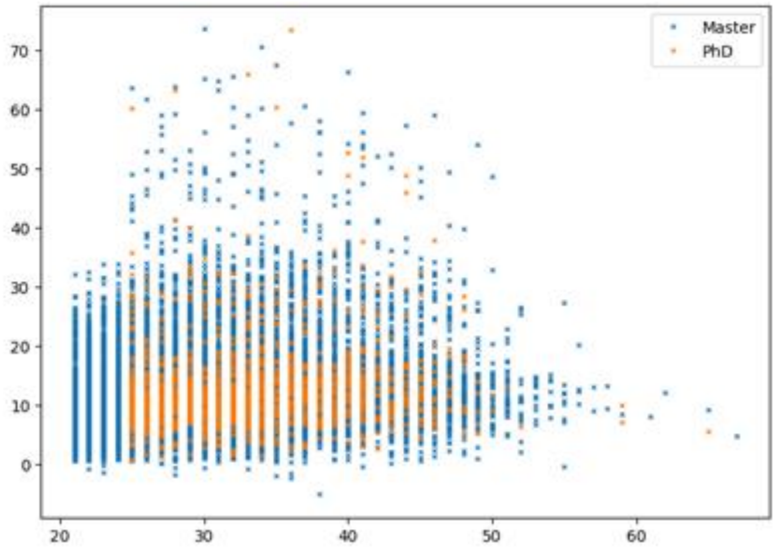
It is possible to get these values,



Although with those values it can be seen that the skill levels are scaled abnormally high in some cases,



Here we've outlined only two classes, that is the Master and No Education categories, and with using the normal distribution you can see that it is far more approachable to get Master's Degree agents in the higher tier rather than to get the No Education agents in the higher tier. One small issue with this idea, is that when we run the simulations, we need to make sure to get correct data or at least attempt to correctly randomize the spread because it is possible to penalize PhD agents and bring them to almost equal distribution with the No Education data due to initial population dynamics, as can be seen in the graph.



Pressure

Agent pressure varies widely due to what sort of work and workload is the Agent currently engaged in. For example, an agent with a medium skill could be engaged on a medium workload throughout the week. Here is a graph that shows an Agent's engagement into his workload.

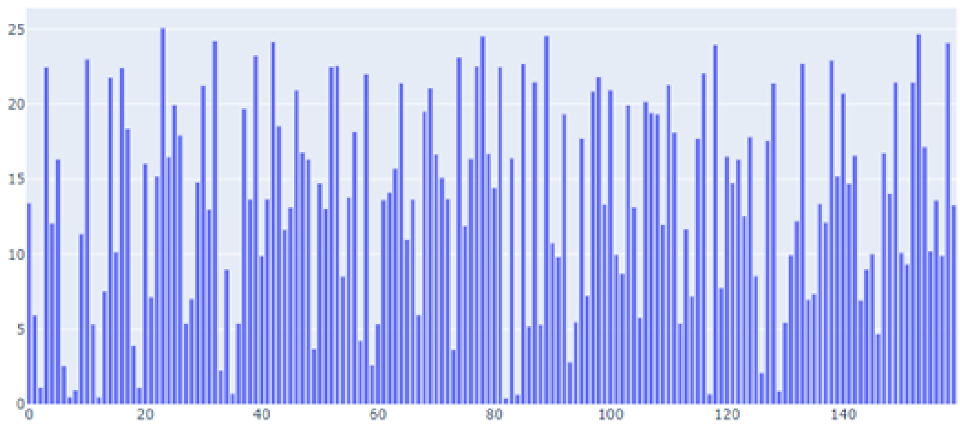


Figure 39: Weekly workload

This weekly workload can be created due to varying factors such as the business where the agent is involved, or the general happiness levels and fatigue levels the agent has.

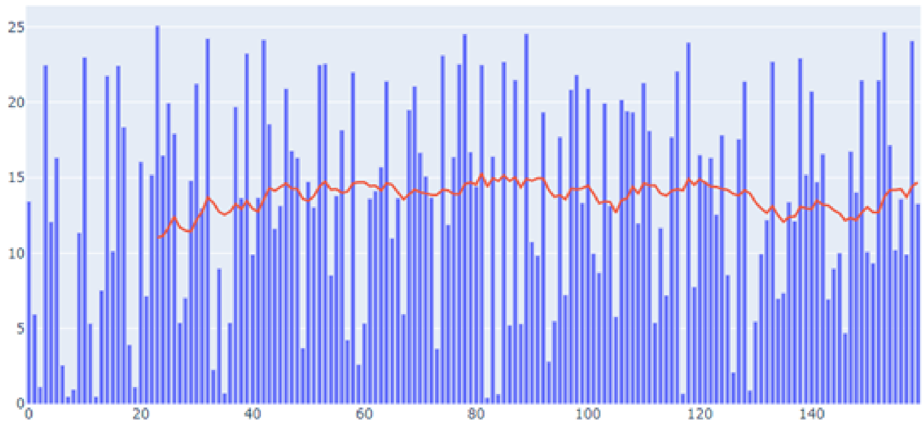
The agent has two main states

- Primary drive
- Predictive drive

Primary Drive

The primary drive for the agent, is the drive that deals with the current workload, and is dependent on the historical values of the agent workload. We define these historical values as the moving averages for hourly engagements.

For example, for agents we generally define six primary historical averages, the hourly workload – the main one. The daily workload which is the average workload spread over a day that the agent feels.



The daily workload is calculated as the moving average

$$w_d = \frac{w_1 + w_2 \ldots + w_{24}}{24}$$

Basically, it adds up 24 hours of work per day. There is a clear distinction here due to the daily workload and the true daily workload. The daily workload which we are presenting here is the workload that takes into account all daily activity, including rest, sleep and eating whereas the true daily workload would be the average over the 8-hour work period.

Figure 40: Weekly workload and Daily Workload

As we can see in the graph, we have a rolling window of the weekly workload and daily workload, represented here in the green and red graphs respectively.

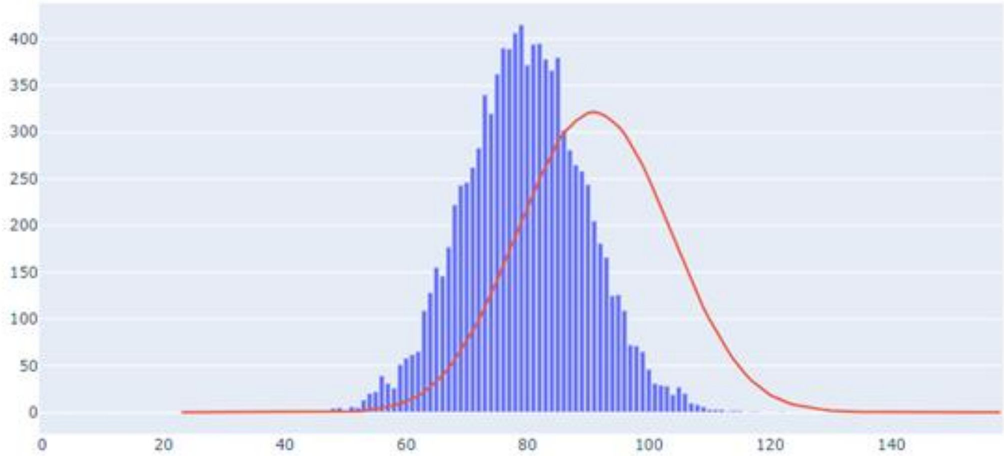


Figure 41: Daily workload on single load

The agent is generally modeled to have a lagged response to the hourly load. The same can be calculated for the weekly average. This average also responds to the weekly load, albeit it can be much slower decreasing.

Figure 42: Weekly average

This is the weekly load for the agent that has only the first week with high load, but then the average falls. The red line represents the daily average load, which can be seen that it never hits the peak load as it is dragged down by the previous values, but it reflects the workload rather strongly.

We hypothesize that agents (i.e. workers, employees) follow this same pattern in their daily operations, that is their hourly engagement does average out per day to some value, and we assume that the agents make decisions based on these lagged values as opposed to making decisions based promptly.

Figure 43: Monthly average of workload

Here is the same workload, simulated over a month, and the agent response. We have made the modelling decisions to make a correlation between the long term effects of workload and the fatigue of the agent, as to try to simulate a more realistic response to the workload engagement of the agents.

Here is a yearly engagement of an agent, and our simulation shows the various months.

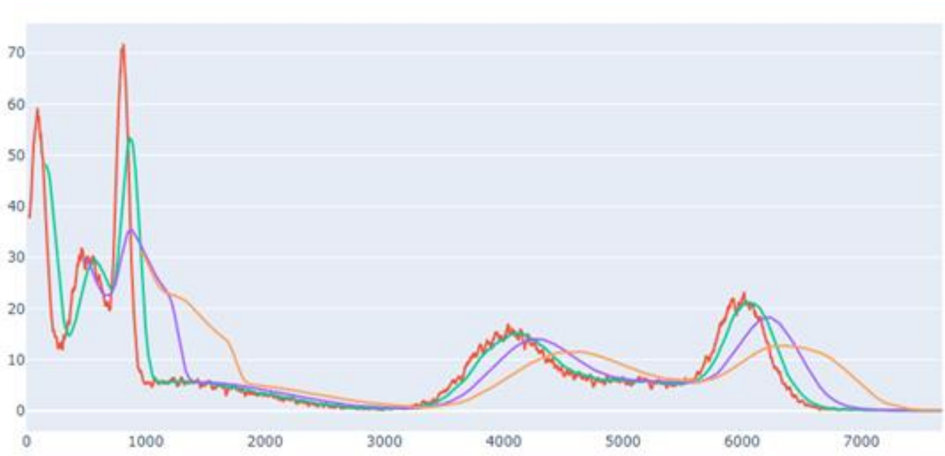


Figure 44: Yearly engagement of one agent

The yearly engagement of this agent has been simulated with various problem sizes, and issues. The agent responds well, attempting to solve problems, but an increased workload can be seen early in the year (the red line) which shows the daily workload.

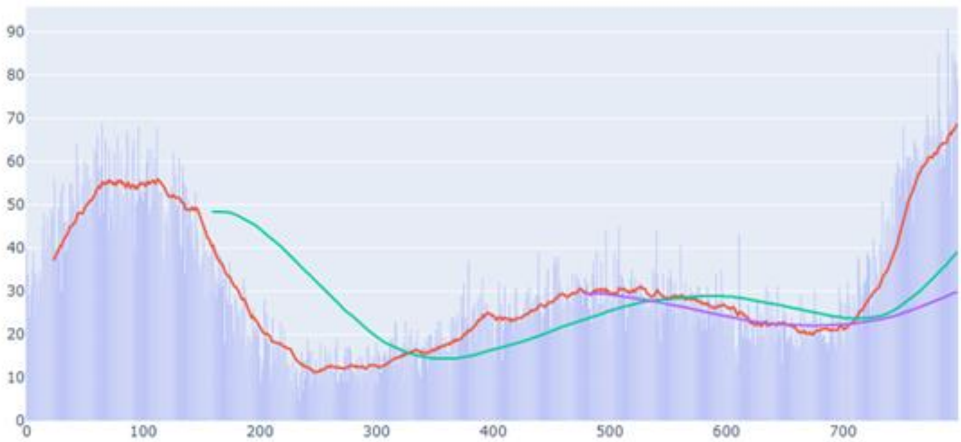


Figure 45: Monthly simulated data

The first month is engaged and it can be also seen by the values reflected by the green and purple lines.

The agent has responded as best as the agent can (rational behavior argument) and has received help from other agents that are co-operating from the same business.

In this version of the model, we encourage the agents to perform proper cooperation and transparent cooperation in respect to the business they are engaged in contracts with, however we are recommending that non-transparent cooperation is also implemented in a future version as it may play a role in some areas.

This is due to response variables given by the agent, especially in cases where the agent may receive help from colleagues but does in practice take all the credit for this help. This may translate to more capital gain for the agent.

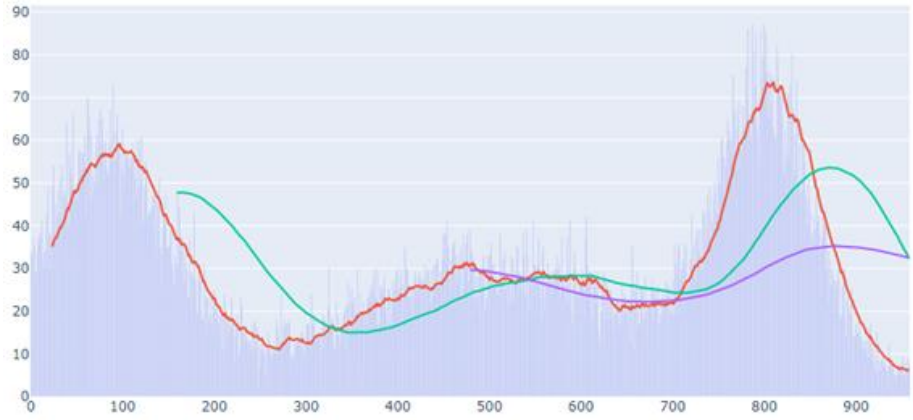


Figure 46: Second month of agent data

The agent has two variables that are correlated strongly with the historical values, namely the Agent Happiness and the Agent Fatigue.

Here is a table of variables we recommend to correlate to the various workloads for a proper and realistic modelling of the agent.

Variables	Correlation	Historical Value
Happiness	MEDIUM	*Quarterly*
Fatigue	STRONG	*Quarterly, Annual and Half-Yearly*
Capital	LOW	*Weekly*
Skill	STRONG	*Daily*

The happiness variable should be related only on a medium strength to the Quarterly workload because we suppose that Agents generally make rational choices based on other factors therefore a medium correlation should be good enough considering that other factors such as capital itself, correlation to health variables, correlation to situations that occur politically (i.e. policies, business dynamics) could strongly influence an Agent's happiness and cause him/her to move forward with migration logic or perhaps quit a job or change jobs.

We also recommend that Happiness is strongly related to the wage and wage differences between agents of the same profession and the same location.

We recommend the agent to have a very strong, emphasis on strong, connection between Fatigue and the 3 long-term averages namely the Quarterly, Half-Yearly and Annual workload.

Here is a graph of the correlation between fatigue and workload over a year period, simulated with a Single Agent simulation.

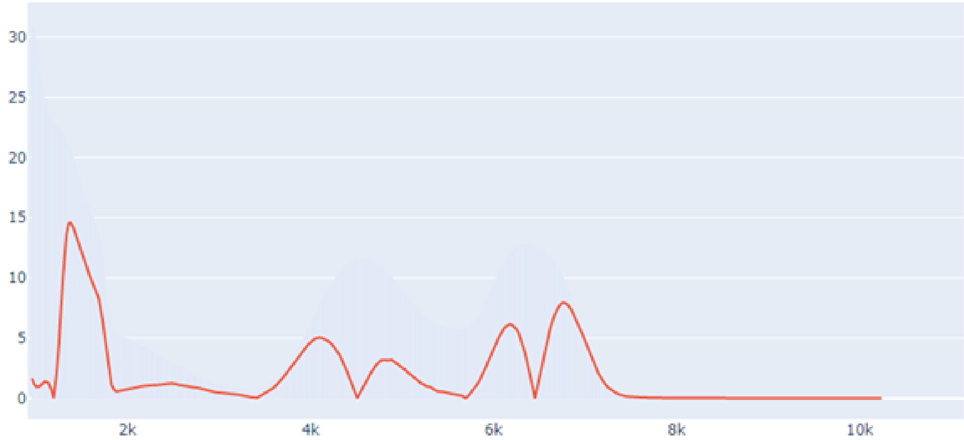


Figure 47: One-year average of Fatigue

We expect the person to feel fatigue on a continuous matter, however we do not expect negative fatigue therefore we correct the simulation by adding absolute value requirements to this situation.

The agent has an historical value to fatigue, namely the Historical Fatigue, which we calculate by taking the moving average of the generated fatigue over some rolling window. We recommend a minimum of

$$w = 640 \, hours$$

According to our simulations, this is a good value for the fatigue modelling. This minimum is significant because it makes sure that it weeds out spikes in fatigue that can be attributed to cases where transitive periods happen in businesses or countries that could shift the workload and stress of the agent.

Therefore, we plan correctly and introduce a smoothed variable for the agent, namely the average fatigue that continually models the changes and maintains a good representation of the long-term effects that fatigue has on the agent.

This is used to further model many other dynamics that drive the agent as there is strong correlation between fatigue and will to continue working, a strong factor that can push an agent to make decisions regarding movement (i.e. migration.

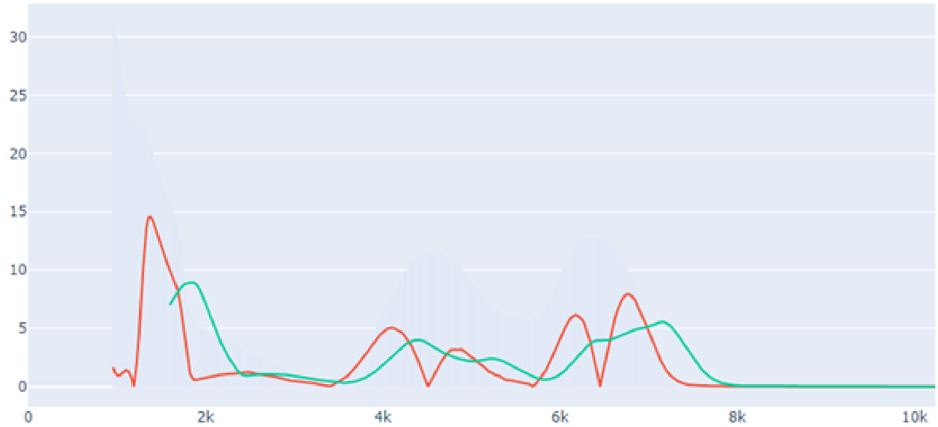


Figure 48: Chronic Fatigue (monthly)

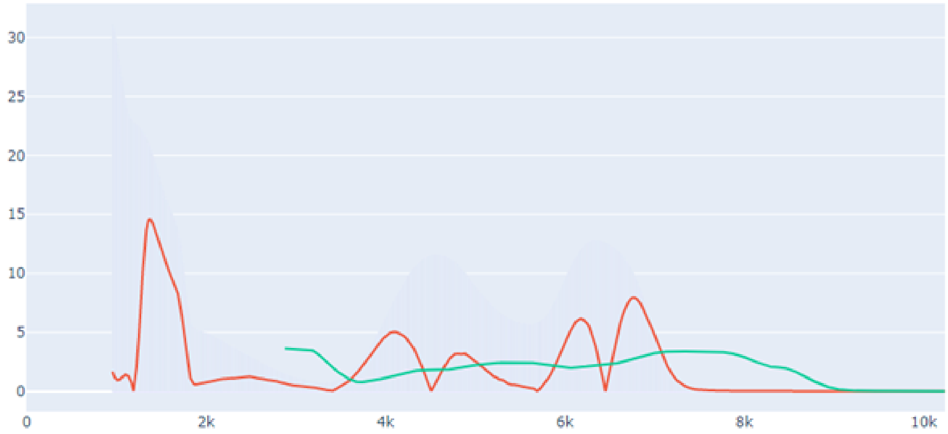


Figure 49: Chronic Fatigue (quarterly)

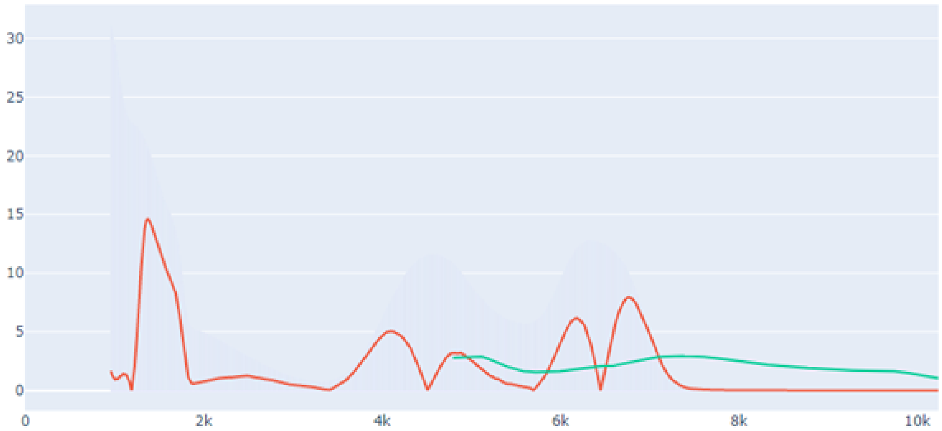


Figure 50: Chronic Fatigue (seasonally)

The images above all describe the various chronic fatigue that the agent feels. If the agent is subjected to strong chronic fatigue, it will severely underperform in daily tasks, have higher chance of damaging health parameters that are assigned to the agent and therefore causing potential moves to less stressful work or less workload.

Variables	Correlation	Historical Value
Happiness	STRONG	*Quarterly*
Health	STRONG	*Seasonally*

This table represents strong correlations between health parameters that are affected directly by the seasonal chronic fatigue, whereas happiness may be more directly affected, and more frequently by the Quarterly value.

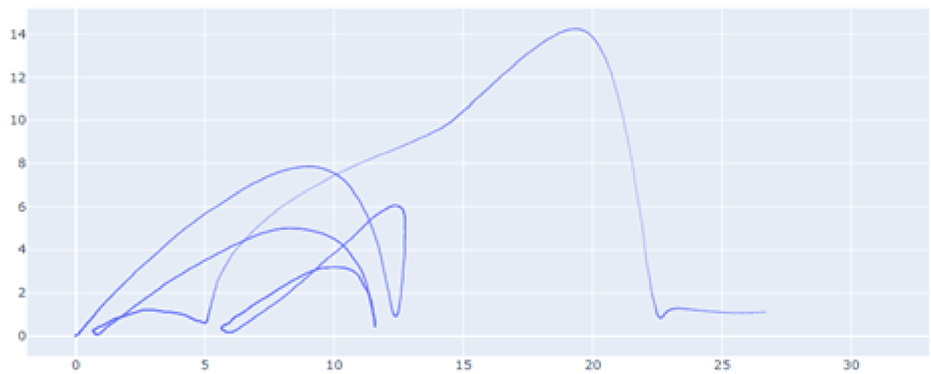


Figure 51: Data collected from an Agent's happiness levels

The happiness levels can be visualized as scatter points on a graph, as they move up and down in phase space.

In this graph, the x-direction represents happiness, whereas the y-direction represents intensity and correlation between the chronic fatigue and happiness.

If the agent is subjected to continuous, long-term chronic fatigue, the agent is more likely to dislike the current position that he/she is in and look for better options, perhaps even sacrificing wage or other parameters.

Predictive Drive

In this model we have also included the feature of predictive drive for Agents, which are rational agents. The predictive drive is a collection of calculations that the agent does on an hourly basis, predicting the next workload. We do this in a few methods, primarily relying on linear regression models

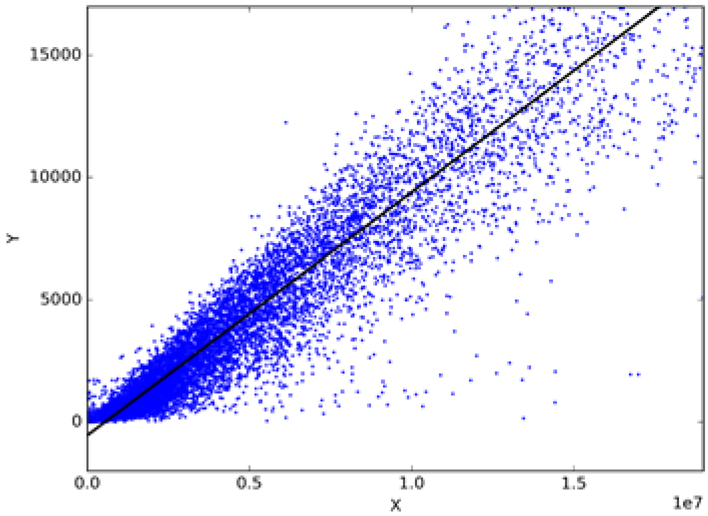


Figure 52: A view of a linear regression model fitted to many points

These regression models give the Agent the edge in making a better decision to how best to allocate his/her skills.

For example, if an Agent has a lot of work in one company, but is getting paid the same as other agents in his/her vicinity and they are working much less (chronic fatigue, and chronic workload) as opposed to the agent in hand – then due to the predictive drive the agent may make changes to his/her own circumstances.

The predictive drive is also used in various other methods, primarily in Agents who have a better skill, we recommend fitting Gaussian Process models which more realistically can predict an agent's workload. Since Agents can generally sustain a year's worth of work without predictive capacities, the Agent can start making predictions for the next season workload, and offer better terms.

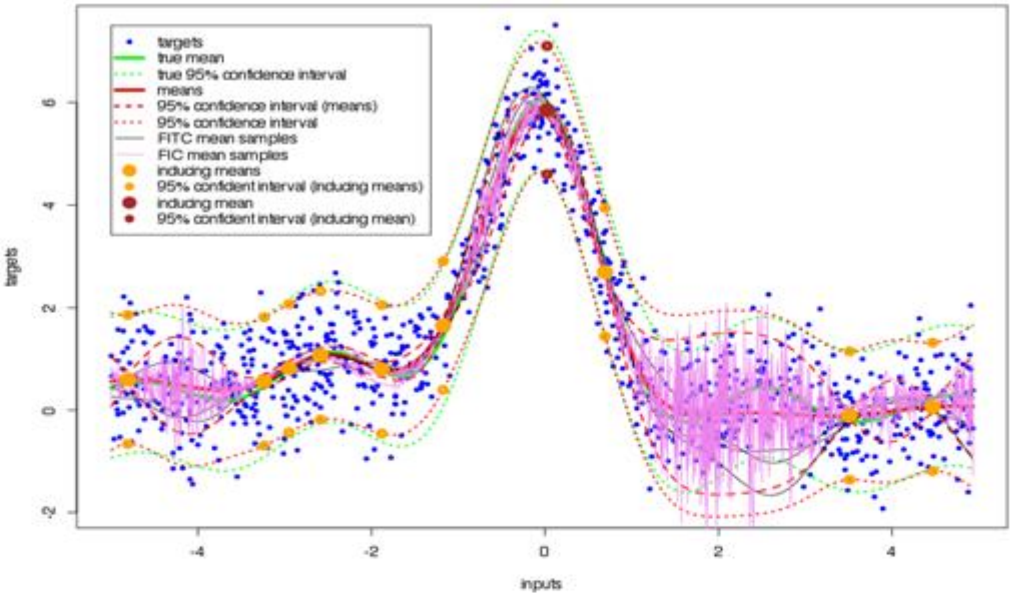


Figure 53: Gaussian process predictors

Consider for example a repeated workload such as this one displayed in the figure below.

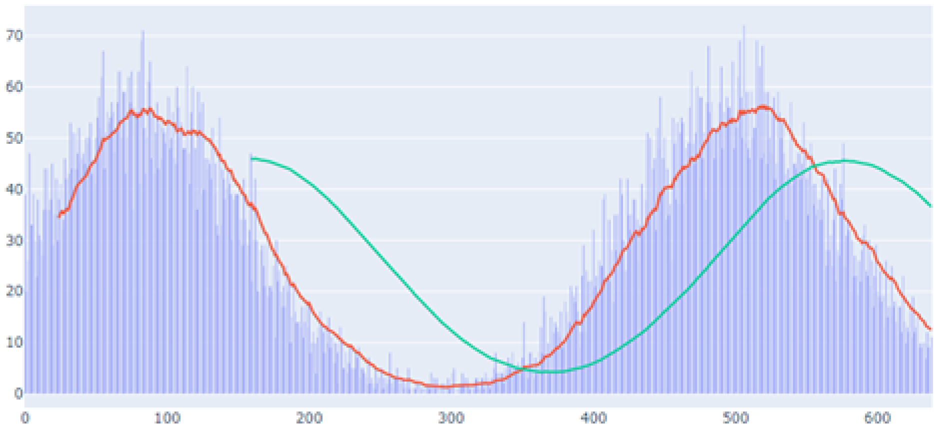


Figure 54: Predictive workload

This means that workload increases strongly in the middle of the month, while the starts and ending of the months are very low.

The agent may use Gaussian predictors to predict the high workload that is coming, and feel more prepared to handle it. If the workload is not predicted correctly however, there is strong correlation between hits to chronic fatigue and wrong predictions.

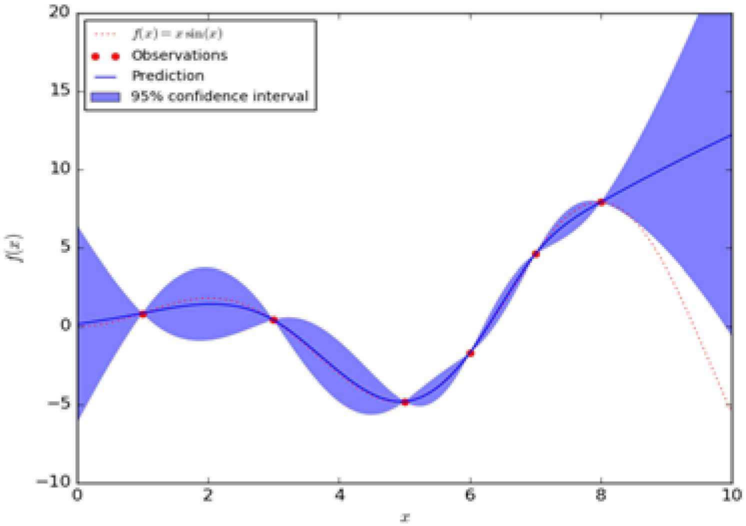


Figure 55: Confidence intervals

We attempt to give confidence intervals to agents who predict various workloads as can be seen in this diagram (a representation). The more the workload is within these confidence intervals, the predictive strength increases and the agent's pressure and chronic fatigue decrease.

By decreasing these values the agent subsequently performs much better in their hourly tasks, which are ultimately aggregated to a better weekly, monthly or yearly performances.

Better performances then lead to higher skill gain, better employment, less movement in work clusters and ultimately a stronger economy for the business, and by continuation a stronger economy for the country.

The longer the agent works in a company, the better are the chances that the agent predicts the coming workload, therefore we have a stronger correlation between longevity, dedication and strong work-ethic.

There is also the possibility to use Deep Learning and Neural Networks to model these predictive capacities.

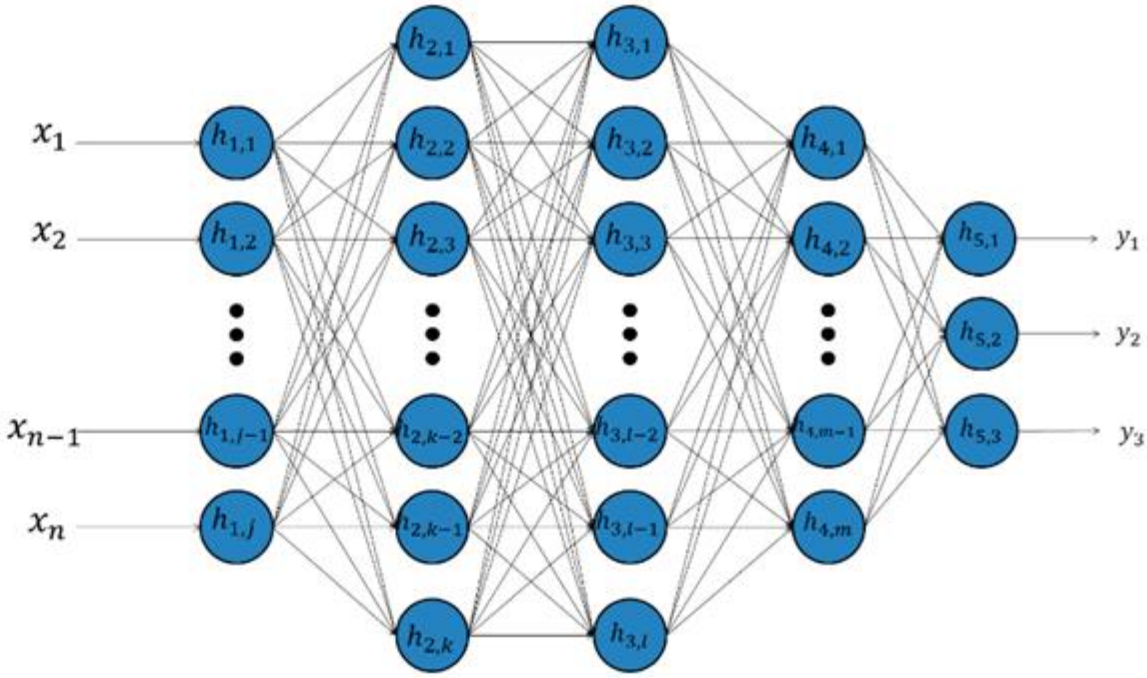


Figure 56: Deep learning

However, these are not recommended for computational resources as training and fitting them takes a lot of time.

There is also the risk of over-fitting deep learning models, that cause strong failure rates among predictive capacities.

We recommend that prediction is generally handled by very general models such as the Liner Regressors, Support Vector Machines or Gaussian Predictors.

Use Cases

In the following part we will look at some use cases that are simulated by a partial implementation of the model discussed earlier. These use-cases are generated by applying specific policies and simulation cases to the run-time engine.

Women workforce increase

In this first simulation, we force the number of active women to increase by 10%, artificially, by increasing their will to search for employment, and investigate what happens afterwards.

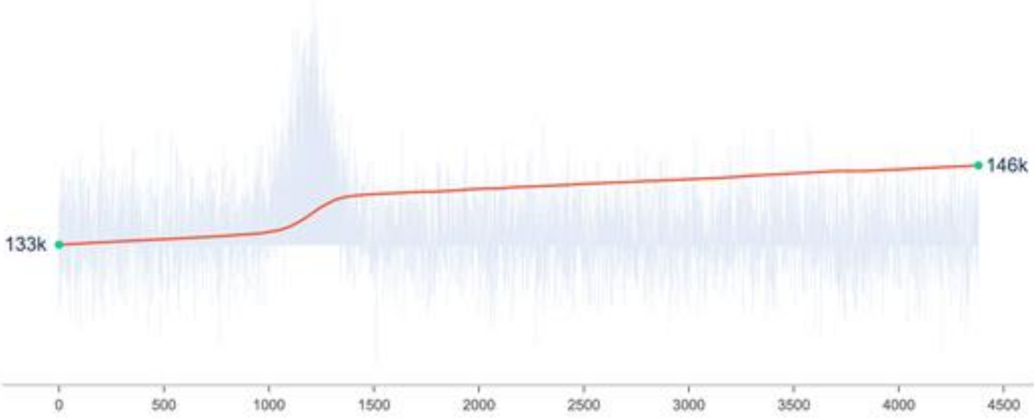


Figure 57: Women entering workforce

In this graph, we see women as they enter the workforce after an artificial imperative around

$$day = 40$$

The graph shows women at an initial ~133k active workforce, both employed and unemployed, cumulatively. The graph shows two main values,

- The active flow rate
- The average cumulative sum of active women

The blue bars show on average a daily rate of change, i.e. women entering the active cluster (when the bars are above the zero line – positive). The values below the zero-line show women exiting the active market on average.

The red line shows the cumulative

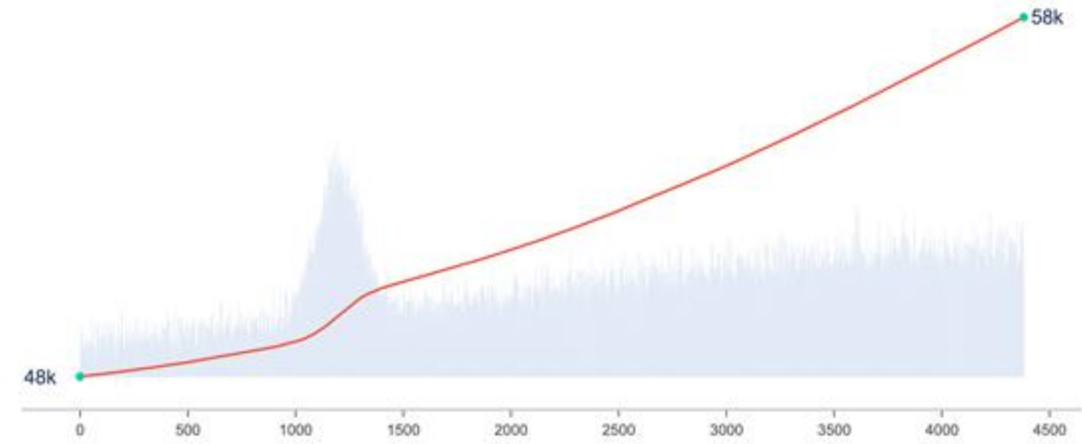


Figure 58: Unemployment Rates filtered by sex

The unemployment rates increase dramatically, this causes a decrease in wages as supply increases. Many of the new women that enter the market, but also some of the earlier agents are willing to accept new contracts with lower wages due to sudden influx of new hires.

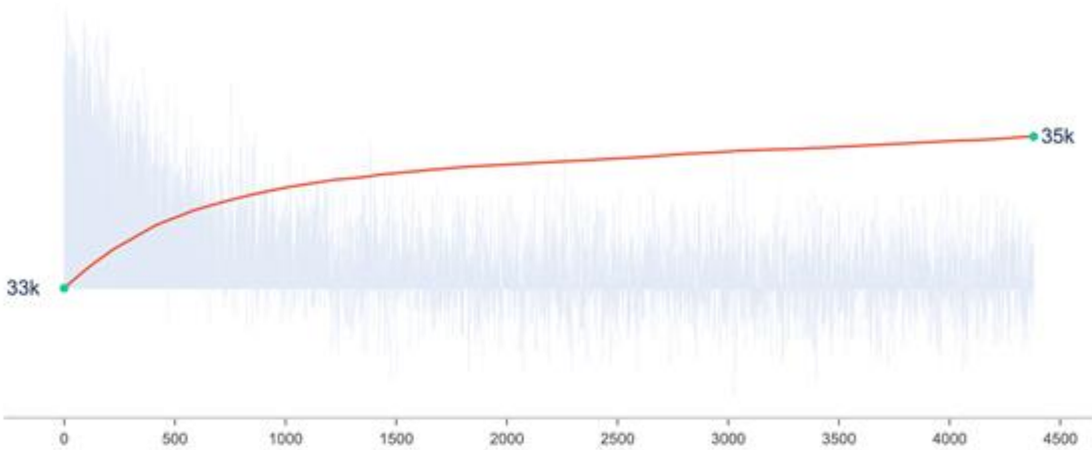


Figure 59: Positions starting to saturate as day women are starting employment

The graph above shows a saturating job market, as 33k contracts were signed by women earlier (the Kosovo case), whereas at the end of the six month period a new ~2k contracts are signed.

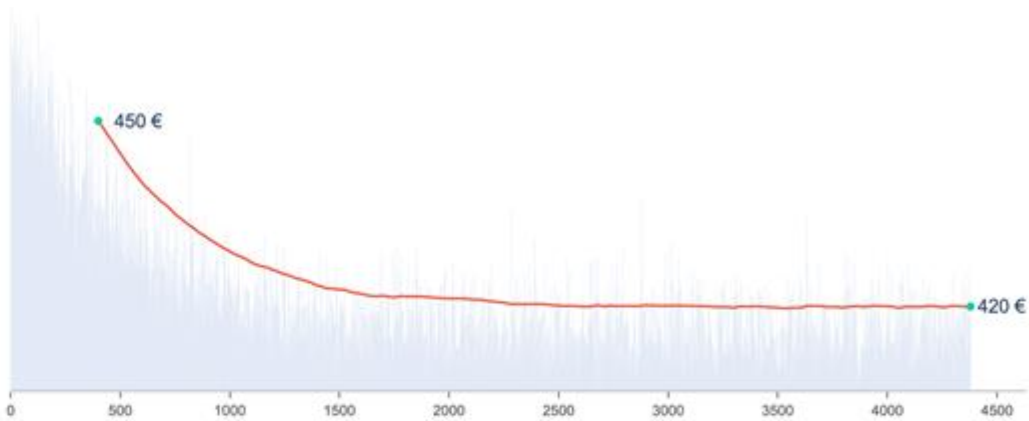


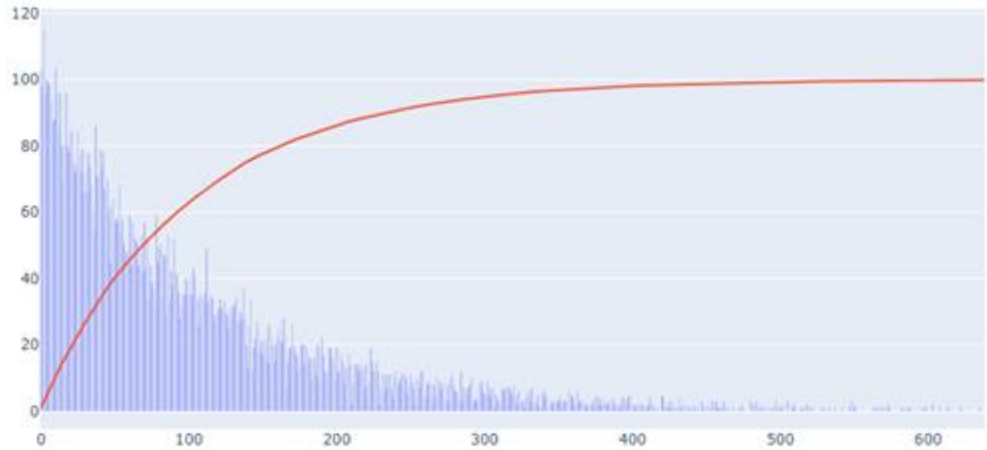
Figure 60: Instantaneous wages per contract

After an average 2k empty positions are filled, companies notice a growth of interest in new jobs and the increase in the job supply if not affected by the job demand will result in a decrease in wages.

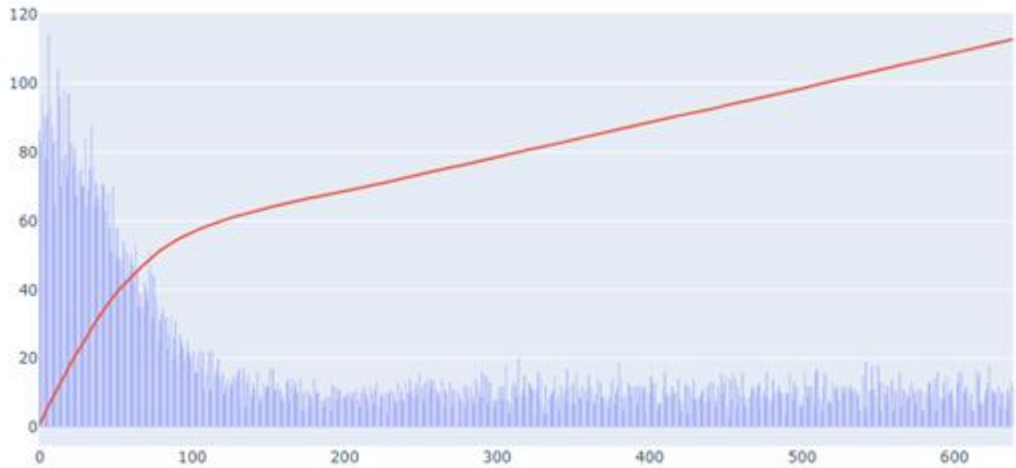
Wage Subsidies

If we artificially subsidize wages, i.e. we assume for the sake of the simulation that the country has money to do this continuously then our model predicts the following simulation parameters,

If this were to occur, we would immediately see a very fast saturation of the job market, as companies start offering more jobs,



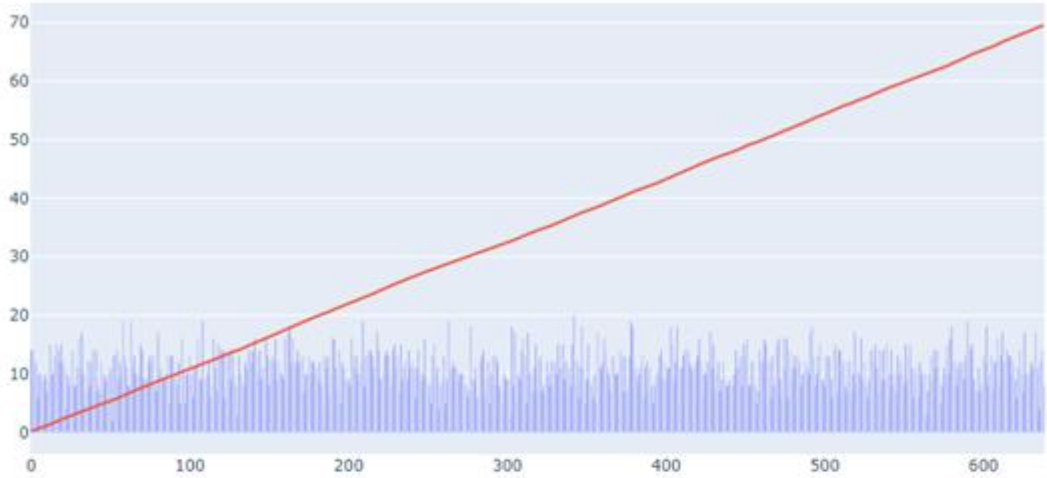
As can be seen in the picture, companies immediately want to attract new talent because now they can pay a smaller wage due to subsidizing policies. However, this very quickly fills up the market as agents are looking to get employed start getting employment and a strong agreement exists between agents of different regions that there is good condition for business.



Agents start joining the workforce quickly, assuming that they are going to get employed with new policies.

This causes strong employment immediately, with a fall-off effect, however we see that proper good employment continues.

Naturally in this case, we also see an increase in supply of workers, uniformly however throughout the year.



The supply has linear growth.

Migration Patterns

This use case, we see movements in between countries, where policies are enacted to increase taxation laws.

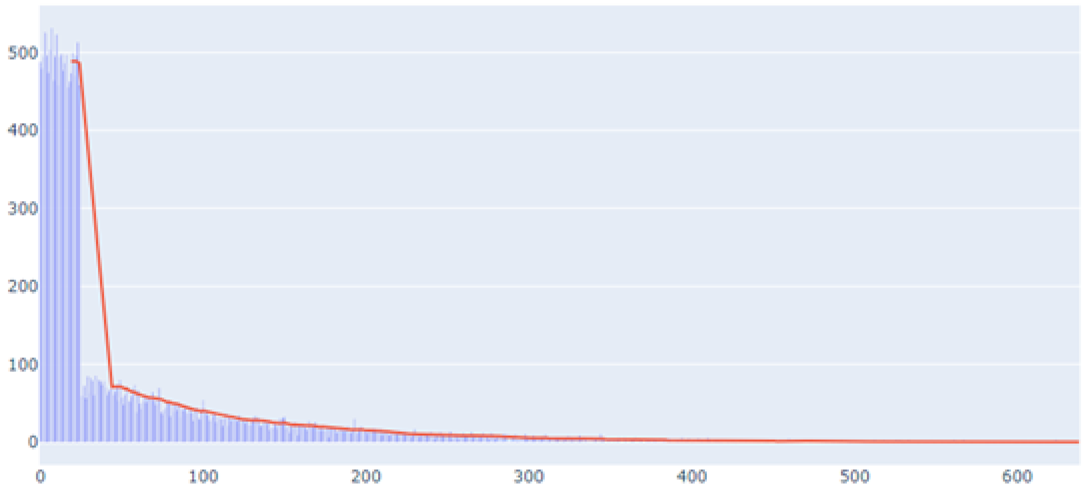
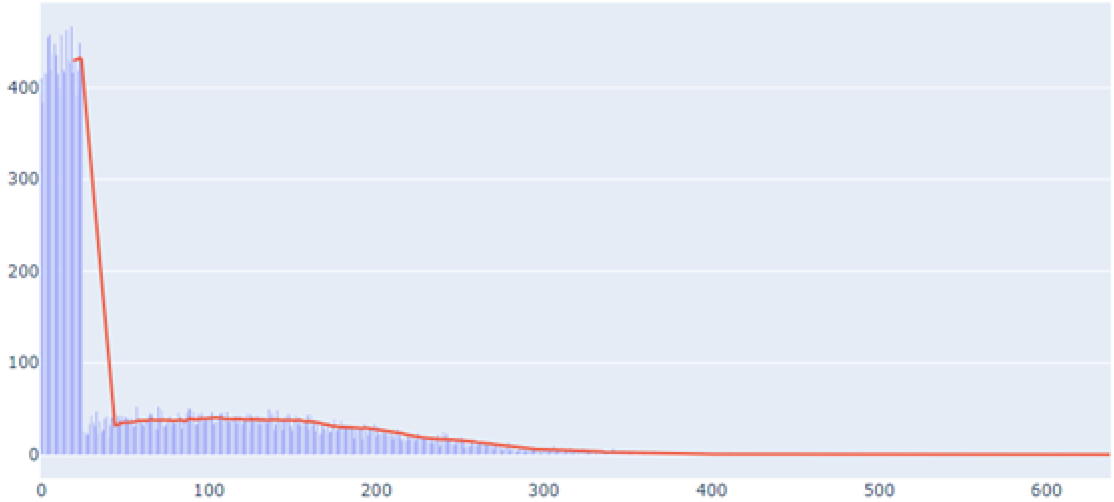


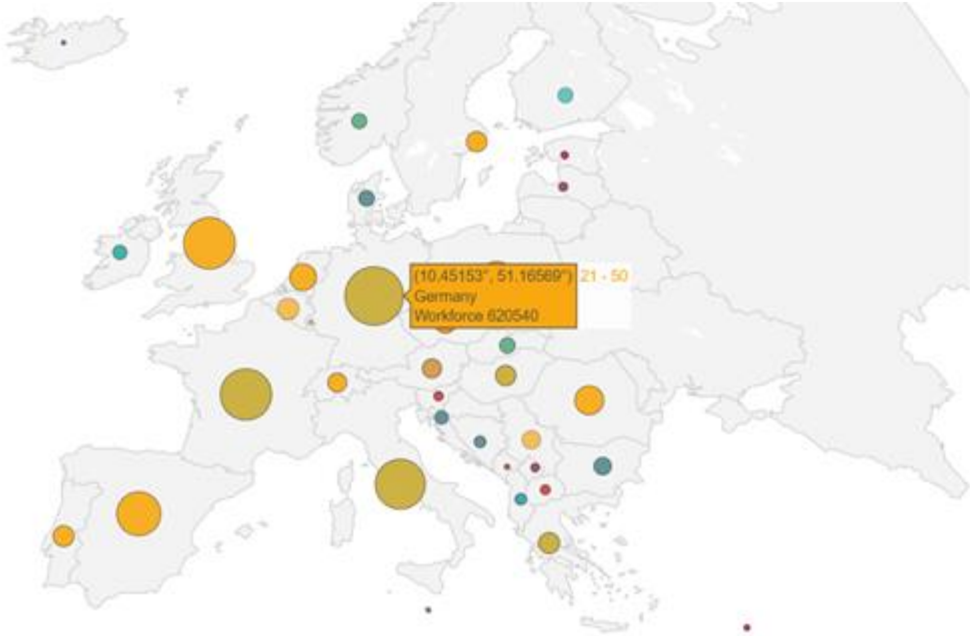
Figure 61 :Tax law increase

Once taxation laws are increased we see a severe drop in employment as companies generally prefer higher profits and will not sacrifice profit for effectiveness.

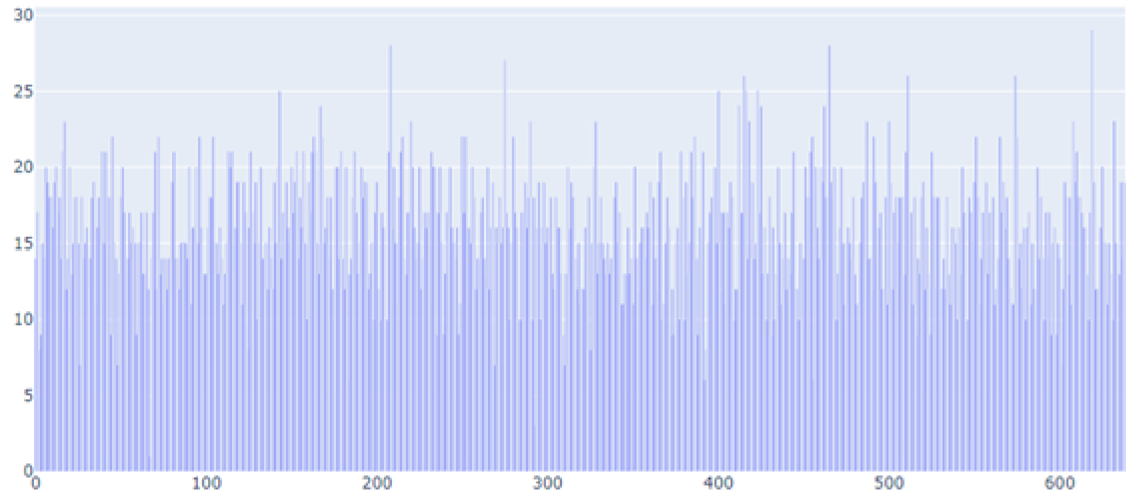
Due to the fact that companies drop employees, we see a surge in unemployment.



The results can be seen that at around day 200 and 300 we see a lowering of the unemployment levels in a country (in this case Kosovo). This may appear interesting, as we can see migration patterns between Kosovo and other countries



In this map we see the workforce sizes that are being simulated (they have been scaled down for computability purposes). These workforces are under the same dynamics.



Average cases of contracts signed globally are represented here in this chart. The simulation shows that even though a law has been passed to increase taxation in one country, the effect is rather small due to the small population size (i.e. workforce scale).

Migration patterns can be visualized and seen here.

Figure 62: Global migration patterns

A stronger line can be seen between Kosovo and other countries accordingly because the model simulated that barriers would loosen attracting both unskilled and skilled workers thus causing the workforce to move beyond the country.

Conclusion and recommendations

The survey conducted by D4D provided data that 60% of the population would leave Kosovo if given the chance. This is a concern as it shows also the level of unsatisfaction with the country's development. However, within that high proportion of people that would leave, 28.4% of those stated that they would only leave for a certain amount of time and come back. Such information provides basis for opening up the discussion on what are their incentives to come back, and also dwell on the reason of those that would not come back.

One thing that stands out the most when Kosovo is mentioned is its young population. Contemporary, having a high youth population contributes vastly to development. However, in Kosovo's case the youth's potential is not utilized. Youth unemployment stands at 49.4%. In addition, youth that are neither employed, nor in education or training comprise 32.7% of the youth population.

Kosovo's working age population holds a ratio of 1-1 of men and women, yet women have a 21% of activity rate thus another under-utilized potential. Labor supply and demand not matching has been a persisting issue in Kosovo. Economic growth has been used as one of the key points for politicians to shift the discourse, however, economic growth was not necessarily followed with improvements in wellbeing. If the gains from economic growth are not utilized, which may occur due to not well-informed financial decisions, corruption, and low return of profit investments, then it cannot be concluded that it has contributed towards an increase in wellbeing.

This paper has taken Kosovo as a main point of study, however, has also touched upon the issue of labor mobility in the Western Balkan 6 countries. A regional approach combined also with individual strategies is the key to utilizing labor mobility to the region's development. Given that the main part of the paper also introduced an agent-based model to project labor mobility, the recommendations are divided into two parts: (a) policy recommendations, and (b) model-related recommendations.

Policy recommendations

To Kosovo Institutions

- Initiate the discussion on migration. Currently migration has been given a bad reputation and is being avoided as a topic in the WB countries. Migration is not bad per se, but when avoided can become critical for the country's development. Therefore, it is recommended to open up the discussion by each country on migration via different communication tools, and focus heavily on circular migration.

- Create mechanisms that create incentives (e.g. fiscal policies) for investment, and building long-lasting networks with diaspora

- Provide easy to use platforms for diaspora to voluntary sign up where they input their data, namely their professions and line of work, which can then be used to devise policies on brain gain and investments. CSOs which have experience with diaspora must be utilized to facilitate the process of diaspora engagement and diaspora profiling.

- Create linkages between private sector companies for staff exchanges via private-public funding.

- Systemic Labour Market analysis: local labour market's demand and supply

- Informed emigration - strengthen the role and capacities of all institutions that provide information on labour emigration and mobility (i.e. Employment Agencies - PES, Universities - through Career centers, private providers - higher education institutions?)

The European Union

- The EU should aim to create new ways of using IPA funds to provide WB countries assistance to link with their diaspora as a means of funneling FDI. IPA funds ought to have a regional component in order to incentivize WB countries on a regional level to tackle their shrinking demographics via diaspora utilization.

- Create new programs that promote circular migration in order to create triple-win situations where both the host country and the country of origin would benefit but also the migrants themselves. Specific programs that target high skilled labor in WB countries for circular migration which provides EU countries with short-term labor but also grants WB countries an improved labor upon their return.

Model related recommendations

Government Usage

Using the model, especially by government-related entities (i.e. institutions, advisors to government, commissions, etc.) is highly recommended as a means to achieving data driven and mathematical decision making. The model can serve as a tool to evaluate decisions, provide support for hypotheses and validate existing policies. Today, almost universally, all stakeholders in government decision making processes have access to a wide spectrum of data that can be highly utilized for creating accurate representations of their economic situation. The model, as described in this paper can coherently use these datasets that are available to the aforementioned entities and create a more accurate depiction of the situation, ultimately strengthening the accuracy of the simulation and helping government make better decisions.

One of the ways governments entities can utilize the model is to define a problem they want to solve, create various hypotheses and finally create a quantifiable result that can be simulated. Once the hypotheses and the results are defined, it is possible to simulate the data and see if the hypothesis leads to actual result values. To try and concretely describe this process, let us take a simple example. We assume that government is planning on increasing taxes for businesses of some specific sector, and they want to know the effect this new policy would have on the nation. This problem has various endpoints we need to assess. Initially the government would be highly interested in understanding the effect of this policy on tax collection and national income. The second factor would be workforce dynamics, i.e. how many jobs would be cut or gained as a result of the policy. Ultimately, economic variables such as trade balance, GDP and if applicable various indices such as manufacturing or sales would change.

Ideally, the government would first collect current data regarding sector size (i.e. workforce participation disseminated by age, sex, education and other demographical data) and assess the sector's overall role in the national income. This assessment can be translated to mathematical statements which can be input into the model and then evaluated for each session. Various other parameters such as GDP and trade balance can be tracked in the simulation, which in turn can be compared to real-world data to get a more accurate picture of expected changes.

Increasing taxes, generally will lead to significant job market changes which can be tracked by the simulation, and these results are easily quantifiable thus can be used on optimization studies or comparative studies. The model can ultimately become a function that can be used on a grander study to optimize tax rates for industries or sectors.

It is important for governments to supplement the simulation data with real-world data in an attempt to increase the accuracy. It is also highly recommended for the government to work together with individuals or entities serving as consultants with a mathematical or statistics background that could interpret and interpolate results from the simulation (model) and accurately map them to real world data as sometimes results from the simulation are heavily scaled due to computational complexity issues.

Other Usage

The model is also recommended to be used by entities that are not of government affiliation. For example, NGOs, especially think tanks, can benefit highly from running simulation on a smaller scale with the model, as they can provide significantly strong and reasonable recommendations for governments and other entities that are involved in the decision-making process. NGOs can also use the model to validate data they collect from surveys and use the model in a proactive approach where they can target certain study areas and supplement the simulation data with surveyed data.

As a concrete case, assume an NGO is working towards analyzing migration patterns for the entire workforce of a country. This case is ideal to be analyzed and simulated with the model. The NGO would feed the data to the model of the current scale and size of the workforce and make the model scale appropriately for faster calculations or maintain normalization for slower calculations but better accuracy. Ultimately the NGO can run the analysis for some period of time and assess movement between countries in the simulation. It is possible to run the simulation in a one-to-one country relationship (i.e. Kosovo and Germany) but also it is possible to run it in a many-to-many relationship where many countries interact with each other. At this point the model can handle - to some degree of accuracy, but is highly discouraged as very specific conditions must be met for this case to work accurately - the simulation of one-to-many where one country is calculated against many targets and possible numerical values can be obtained but the accuracy needs to be supplemented with real-world data to get a good picture.

Policy papers, researchers and academics who are working in these socio-economic fields can also gain from the model, as it is compact enough to be run on a single computer (i.e. personal laptop) and can be run for various scenarios that can lead to hypothesis validation, data driven assessments and ultimately serve as a helpful tool in providing supporting arguments to some hypothetical situation under study.

Using the model as a hobbyist or for personal understanding can be done, but generally is discouraged as the margin for error is large in some scenarios where human factor is needed to correct for.

Katalogimi në botim – (CIP)

Biblioteka Kombëtare e Kosovës "Pjetër Bogdani"

331.103(496.51)(047)

Labor mobility : A bottom-up approach To go or not to go – that will not be the question / Prepared by: Gersi Gashi – Prishtinë : Demokraci për Zhvillim, 2020. – 69 f. : ilustr. 21 cm.

1. Gashi, Gersi

ISBN 978-9951-608-92-3

The Democracy for Development (D4D) Institute was established in April 2010 by a group of analysts who were increasingly worried that the state-building exercise had neglected democracy.

D4D’s vision is to promote an active and educated citizenry that participates fully in the public space and utilises the public arena of representation and decision-making to deliberate and build consensus over resource allocation that is efficient, smart, long-term, and that brings about equitable development.

D4D influences specific policy, promotes a cross-sectorial approach to problem solving, and tackles the institutional routine of decision-making by recommending incremental improvements and operates with maximum effectiveness in order to fully promote Kosovo’s stabilization and democratic development.

For more information about D4D's activities please visit our website: **www.d4d-ks.org**.

Supported by:

OPEN SOCIETY
FOUNDATIONS

ISBN 978-9951-608-92-3

9 789951 608923