A Comprehensive Regression Study on the Drivers of Labour Productivity

Shevelova, Anastasia and Machukha, Ielyzaveta and Motliuk, Mark and Kulinich, Volodymyr

Kyiv School of Economics

10 July 2023

A Comprehensive Regression Study on the Drivers of Labour Productivity

July 8, 2023

Anastasiia Shevelova
ashevelova@kse.org.ua

Yelyzaveta Machukha
imachukha@kse.org.ua

Mark Motliuk
mmotliuk@kse.org.ua

Volodymyr Kulinich
vkulinich@kse.org.ua

1. Abstract[1]

Labour productivity is an essential economic indicator, offering insights into a nation's hourly economic output. Understanding a country's performance is pivotal for assessing policy effectiveness and shaping new strategies. This study aims to identify the primary determinants of labour productivity and analyze their impact. Employing data from the World Bank and ILOSTAT, the linear regression method was used for analysis to uncover significant insights. The findings reveal a positive correlation between urbanization and labour productivity, while employment in agriculture, as expected, exerts a negative influence. Furthermore, a direct relationship was observed between a country's income level and labour productivity, with higher incomes associated with increased productivity. Notably, the unemployment rate exhibits a positive association with labour productivity, and this effect intensifies as income levels decrease.

Keywords: Labour productivity, Country performance, Determinants of labour productivity, Linear regression analysis

[1]This manuscript was part of the Regression Analysis course at Kyiv School of Economics (Term VII, 2023), conducted under the curriculum designed for second-year students.

Contents

1. Abstract[2] **1**

2. Introduction **3**

3. Data and Methods Description **4**
- 3.1. Data Sources 4
- 3.2. Dependent variable 4
- 3.3. Independent variables 4
- 3.4. The model 5
- 3.5. Elaboration on independent variables 5
- 3.6. Prior expectations 6
- 3.7. Model estimation methodology 7
- 3.8. The model equation 7

4. Empirical Analysis and Results **9**
- 4.1. Visualising the relationship 9
- 4.2. Exploring corellations 10
- 4.3. Initial full model 11
- 4.4. Reduced model 12
- 4.5. Exploration of interactions or quadratic effects 13
- 4.6. F-test: determining the final model 13
- 4.7. Testing for collinearity 14
- 4.8. Testing for regression analysis assumptions 14
- 4.9. Exploring the model quality 16
- 4.10. Model interpretation 17
- 4.11. Comparing results to the expectations 18

5. Conclusions **19**

6. References **20**

7. Annexes **21**
- 7.1. Descriptive statistics of all the variables in our models 21
- 7.2. Results of our models 22

[2]This manuscript was part of the Regression Analysis course at Kyiv School of Economics (Term VII, 2023), conducted under the curriculum designed for second-year students.

2. Introduction

Labour productivity is an economic measure showing the hourly output of a country's economy. Understanding a country's performance helps estimate the efficiency of existing policies and shape new ones. Labour productivity is influenced by various economic factors such as inflation, foreign direct investment, employment distribution by economic sectors, and investments in tech, research and development, and education. The complexity and variety of input variables make it a challenging and exciting topic to analyse (Choudhry, 2009).

Our study aims to determine the most influential factors that affect labour productivity in general for the whole world, as well as separately for countries grouped by characteristics such as income levels, GDP per capita, and others. The predictors used to analyse labour productivity include urbanisation, human capital, sectoral employment distribution, foreign direct investments, inequality, unemployment rate, education, investment in research and development (R&D), investment in information and communication technology (ICT), cumulative inflation, financial sector development, and international trade.

Besides decomposing labour productivity into its shaping components, we are interested in carefully examining the policies and approaches of different countries with the hope of establishing best practices and developing potential policy interventions for the Ukrainian government.

The analysis results will contribute to an already large body of research and provide tailored advice for Ukrainian policymakers.

3. Data and Methods Description

3.1. Data Sources

The primary sources for our data collection were open datasets provided by the World Bank and ILO-STAT. The World Bank is one of the largest and most reliable providers of open data, which contributes to the authenticity and validity of our analysis.

In our research, we need the following datasets:

- Inflation, GDP deflator; (source: WorldBank)
- Trade (% of GDP); (source: WorldBank)
- Foreign direct investments; (source: WorldBank)

Literacy rate (% of adult population)- Unemployment, total (% of total labour force) (modelled ILO estimate); (source: WorldBank)

- R&D Expenditure (% of GDP); (source: WorldBank)
- ICT Exports (% of GDP); (source: WorldBank)
- Education expenditure (% of GDP); (source: WorlBank)
- Urbanisation (% of total population); (source: WorldBank)
- Human Capital Index (HCI) (scale 0-1); (source: WorldBank)
- Employment in agriculture (% of total employment) (modelled ILO estimate); (source: WorldBank)
- Country and Lending Groups; (source: WorldBank)
- Labour Productivity (GDP per hour worked, USD); (source: ILOSTAT)

3.2. Dependent variable

Our dependent variable is labour productivity. Labour productivity is a crucial economic indicator intricately tied to economic growth, competitiveness, and living standards within an economy. It is quantified as the total output volume (as represented by Gross Domestic Product, GDP) produced per unit of labour (measured as the number of employed persons or hours worked) within a given time-frame. This indicator allows us to evaluate GDP-to-labour input ratios and growth rates over time, providing invaluable insights into the efficiency and quality of human capital in the production process. Furthermore, it offers a lens to understand a given economic and social context, considering other complementary inputs and innovations used in production.

In our study, labour productivity is the outcome we aim to predict and understand better using various socioeconomic factors as predictors. We use data from ILOSTAT for this measure.

3.3. Independent variables

Our independent variables are sourced from various datasets provided by the World Bank. Each variable is measured as follows:

1. **Inflation, GDP Deflator (Annual %):** This measurement reflects the rate of price change in the economy as a whole. It is derived from the ratio of GDP in current local currency to GDP in constant local currency. A rise in the deflator indicates an increase in inflation.

2. **Trade (% of GDP):** Trade is calculated as the sum of exports and imports of goods and services, represented as a share of the gross domestic product.

3. **Foreign Direct Investment (Net Inflows % of GDP):** These are net inflows of investment intended to acquire a lasting management interest (10% or more of voting stock) in an enterprise operating in an economy other than that of the investor. This measurement includes equity capital, reinvestment of earnings, and other long-term and short-term capital.Literacy rate (% of adult population)The literacy rate represents the proportion of adults in a population who possess read-

ing and writing skills, serving as a key indicator of educational attainment and human capital development

4. **Unemployment (Total % of Total Labor Force):** Unemployment refers to the portion of the labour force that is without work but is available for and seeking employment.
5. **Urbanisation (% of Total Population):** This measurement denotes the percentage of people living in urban areas as defined by national statistical offices.
6. **Human Capital Index (HCI):** The HCI computes the contributions of health and education to worker productivity. A score ranging from 0 to 1, measures the productivity of a child born today, assuming they achieve full health and complete education.
7. **Employment in Agriculture (% of Total Employment):** This refers to persons of working age who were engaged in any activity to produce goods or provide services for pay or profit in the agriculture sector.
8. **R&D Expenditure (% of GDP):** R&D expenditure refers to the amount of funds allocated to research and development activities as a percentage of a country's GDP, indicating the level of investment in innovation and technological advancement within an economy.
10. **ICT Exports (% of GDP):** ICT exports refers to the exportation of Information and Communication Technology goods and services from one country to another; it is an important component of international trade and can reflect a country's competitiveness and participation in the global digital economy.
11. **Education expenditure (% of GDP):** Education expenditure refers to the amount of money spent on education, including funding for schools, teachers, resources, and other related expenses. It reflects the investment made in the education sector to support and enhance educational opportunities and outcomes.
12. **Income group (factor):** For the current 2024 fiscal year, low-income economies are defined as those with a GNI per capita, calculated using the World Bank Atlas method, of 1,135 USD or less in 2022; lower-middle-income economies are those with a GNI per capita between 1,136 USD and 4,465 USD; upper-middle-income economies are those with a GNI per capita between 4,466 USD and 13,845 USD; high-income economies are those with a GNI per capita of 13,845 USD or more.

These independent variables have been chosen based on their theoretical and empirical relevance to labour productivity, as per economic literature and data availability.

3.4. The model

The model of is a multiple linear regression model, represented as:

$$\begin{gathered}\log(\text{labour\_productivity}) = \beta_0 + \beta_1 \cdot \text{education\_expenditure} + \beta_2 \cdot \text{literacy\_rate} + \\ + \beta_3 \cdot \text{rnd\_expenditure} + \beta_4 \cdot \text{ict\_exports} + \beta_5 \cdot \text{agriculture\_employmnet} + \\ + \beta_6 \cdot \text{urbanisation} + \beta_7 \cdot \text{trade} + \beta_8 \cdot \text{inflation} + \beta_9 \cdot \text{HCI} + \beta_{10} \cdot \text{foreign\_invest} + \\ + \beta_{11} \cdot \text{unemployment} + \beta_{12} \cdot \text{icome\_group} + \varepsilon\end{gathered}$$

This model serves our purpose by allowing us to estimate the effects of various socioeconomic factors on labour productivity.

3.5. Elaboration on independent variables

The specific variables included in our model were carefully chosen based on their relevance and importance as indicated in two research studies: "Factors influencing labour productivity in the OECD countries: Radło and Tomeczek (2022)" and "Determinants of Labor Productivity: An Empirical Investigation of Productivity Divergence (2009)".

These studies provided valuable insights into the factors most likely to influence labour productivity. In addition, when selecting our variables, we made sure to choose those for which substantial data was available across a wide range of countries. The aim was to construct a model that could provide a comprehensive and robust understanding of the factors influencing labour productivity on a global scale.

This approach ensures our model is not only grounded in theory but also applicable and useful in providing practical insights into labour productivity across different contexts and regions.

1. Inflation, GDP Deflator: Inflation can impact the cost of goods, services, and wages, all of which are important factors in productivity.
2. Trade (% of GDP): A higher trade-to-GDP ratio often suggests a more open economy with a freer exchange of goods, services, and knowledge, which could boost productivity.
3. Foreign Direct Investment (Net Inflows % of GDP): FDI can introduce new technologies and practices, boosting productivity.
4. Literacy Rate: Higher literacy rates are often linked to a better-educated workforce, higher levels of innovation, and a more informed and engaged citizenry; which positively affects labour productivity.
5. Unemployment Rate: A high unemployment rate might mean that fewer resources are being utilized, lowering overall productivity.
6. Urbanisation (% of Total Population): Urban areas can have higher productivity due to agglomeration effects.
7. Human Capital Index: Higher human capital tends to increase productivity as workers have more skills and education.
8. Employment in Agriculture (% of Total Employment): A higher share of employment in agriculture might suggest a less developed economy with lower productivity.
9. R&D Expenditure (% of GDP): Reflects the investment in research and development activities, which can drive innovation and productivity.
10. ICT Exports (% of GDP): ICT exports represent the value of information and communication technology goods and services exported, indicating the contribution of the ICT sector to economic productivity.
11. Education expenditure (% of GDP): Education expenditure signifies the investment in the education sector, which can enhance human capital and workforce skills, thereby impacting productivity.
12. Income group (factor): Countries which are classified as high-income and upper-middle-income are expected to have higher labour productivity than countries which are classified as lower-middle-income and low-income.

3.6. Prior expectations

literacy ratepositivelyindex R&D investments are expected to positively impact labour productivity.

- Higher expenditures on education are expected to positively impact labour productivity.
- Higher levels of ICT development are expected to positively impact labour productivity.Expectations about Coefficients:
- Inflation is expected to negatively impact labor productivity. (High inflation rates can erode purchasing power and increase uncertainty, leading to higher costs for businesses and reduced investment, which can ultimately hinder productivity growth.)

- Trade openness is expected to positively impact labor productivity. (Increased access to international markets and exposure to competition can drive efficiency and innovation, encouraging firms to improve their productivity to remain competitive.)
- Foreign Direct Investment is expected to positively impact labor productivity. (Foreign direct investment brings in capital, technology, and expertise, which can enhance productivity by introducing advanced production methods, promoting knowledge transfer, and fostering productivity spillovers in the host economy.)
- Unemployment is expected to negatively impact labor productivity. (Persistently high levels of unemployment can lead to skill erosion, reduced worker motivation, and decreased investment in training and capital equipment, all of which can hamper productivity growth.)
- Urbanization is expected to positively impact labor productivity. (Urban areas often offer agglomeration benefits, such as better access to infrastructure, services, and a larger pool of skilled workers, which can lead to increased productivity and innovation.)
- Higher human capital is expected to positively impact labor productivity. (Investments in education, training, and skills development contribute to a more capable and adaptable workforce, resulting in higher productivity levels and the ability to leverage advanced technologies effectively.)
- Higher employment in agriculture is expected to negatively impact labor productivity. (Higher employment in the agricultural sector, especially if characterized by low technology adoption and low-value-added activities, can indicate lower overall productivity levels in the economy.)
- Higher employment in services is expected to positively impact labor productivity. (The service sector encompasses a wide range of industries, including high-value-added activities such as finance, information technology, and professional services. Higher employment in these sectors often indicates a shift towards more productive and knowledge-intensive activities.)

3.7. Model estimation methodology

This model will be estimated using the Ordinary Least Squares (OLS) regression method. The main features of OLS regression analysis involve creating a model that best fits the relationship between the dependent and independent variables. It does this by minimizing the sum of the squares in the difference between the observed and predicted values of the dependent variable configured as a straight line. The OLS method provides the estimates for the intercept and slope parameters which minimize these residuals.

3.8. The model equation

The model equation we will estimate can be represented as:

$$\begin{gathered} \log(\text{labour\_productivity}) = \beta_0 + \beta_1 \cdot \text{education\_expenditure} + \beta_2 \cdot \text{literacy\_rate} + \\ + \beta_3 \cdot \text{rnd\_expenditure} + \beta_4 \cdot \text{ict\_exports} + \beta_5 \cdot \text{agriculture\_employmnet} + \\ + \beta_6 \cdot \text{urbanisation} + \beta_7 \cdot \text{trade} + \beta_8 \cdot \text{inflation} + \beta_9 \cdot \text{HCI} + \beta_{10} \cdot \text{foreign\_invest} + \\ + \beta_{11} \cdot \text{unemployment} + \beta_{12} \cdot \text{icome\_group} + \varepsilon \end{gathered}$$

Where:

- "β0" is the intercept (constant term).
- "βi" are the coefficients for the corresponding predictor variables.
- "ε" is the error term (representing the difference between the actual and predicted values of the dependent variable).

In this model, the log of labour productivity is the dependent variable, while the rest variables such as education expenditure, literacy rate, R&D expenditure, ICT exports, agricultural employment average, urbanization average, trade average, inflation average, income groups (low, lower middle, and upper middle income), foreign investment average, and unemployment average are the independent variables.

4. Empirical Analysis and Results

4.1. Visualising the relationship

We commence our analysis by visualising the relationship between the selected variables.

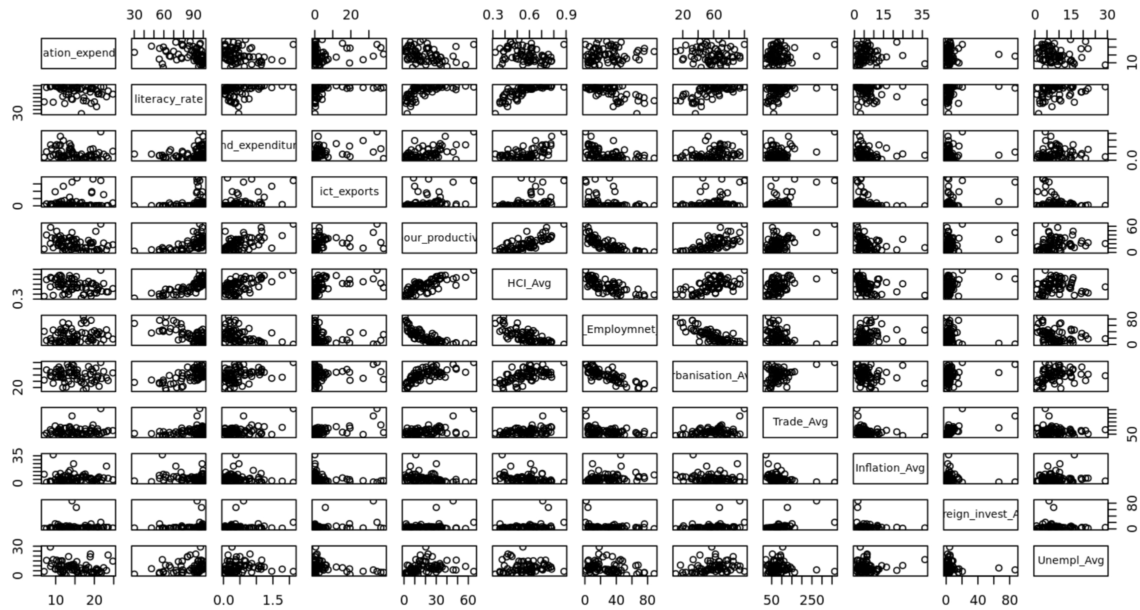

Figure 1: Plot of all the relationships between all the variables in the dataset

By plotting the relationships between all the variables in our data, as shown on Figure 1, it can be concluded that:

- Labour productivity, education expenditure, and unemployment visually do not have a strong relationship, as the points are scattered randomly without any discernible pattern.
- Labour productivity and literacy rate, human capital index, employment in agriculture, and urbanisation have a non-linear relationship, as the points form a curved pattern.
- Labour productivity and R&D expenditure and trade have a positive linear relationship, while labour productivity and inflation have a negative linear relationship.
- Observing relationships between labour productivity ICT exports, and foreign direct investments is hard to make any conclusions, as the points do form clear straightforward trends, but they are not positive, nor negative.

Also, on the scatter plot the relationships between multiple independent variables can be observed. For instance, a strong positive relationship is present between the level of human capital index and the level of urbanisation; logically urbanisation and employment in agriculture have a strong negative relationship. All of the above may lead to the problem of multicollinearity in the analysis.

4.2. Exploring corellations

	education_expenditure	literacy_rate	rnd_expenditure	ict_exports	labour_productivity	HCI	Agri Employmnet	Urbanisation	Trade	Inflation	Foreign_invest	Unempl
education_expenditure	1	-0.26	-0.19	0.12	-0.26	-0.27	0.15	-0.1	0.14	-0.02	-0.01	-0.2
literacy_rate	-0.26	1	0.26	0.26	0.6	0.78	-0.69	0.62	0.31	-0.19	0.15	0.24
rnd_expenditure	-0.19	0.26	1	0.4	0.59	0.58	-0.42	0.32	0.39	-0.18	0.08	0.07
ict_exports	0.12	0.26	0.4	1	0.31	0.37	-0.26	0.23	0.59	-0.27	0.35	-0.25
labour_productivity	-0.26	0.6	0.59	0.31	1	0.77	-0.77	0.7	0.51	-0.27	0.28	0.24
HCI	-0.27	0.78	0.58	0.37	0.77	1	-0.73	0.59	0.48	-0.33	0.28	0.05
Agri Employmnet	0.15	-0.69	-0.42	-0.26	-0.77	-0.73	1	-0.81	-0.38	0.25	-0.21	-0.27
Urbanisation	-0.1	0.62	0.32	0.23	0.7	0.59	-0.81	1	0.36	-0.17	0.25	0.26
Tradeg	0.14	0.31	0.39	0.59	0.51	0.48	-0.38	0.36	1	-0.34	0.57	-0.07
Inflation	-0.02	-0.19	-0.18	-0.27	-0.27	-0.33	0.25	-0.17	-0.34	1	-0.16	0.03
Foreign_invest	-0.01	0.15	0.08	0.35	0.28	0.28	-0.21	0.25	0.57	-0.16	1	-0.07
Unempl	-0.2	0.24	0.07	-0.25	0.24	0.05	-0.27	0.26	-0.07	0.03	-0.07	1

Table 1: The corellation matrix of all the variables

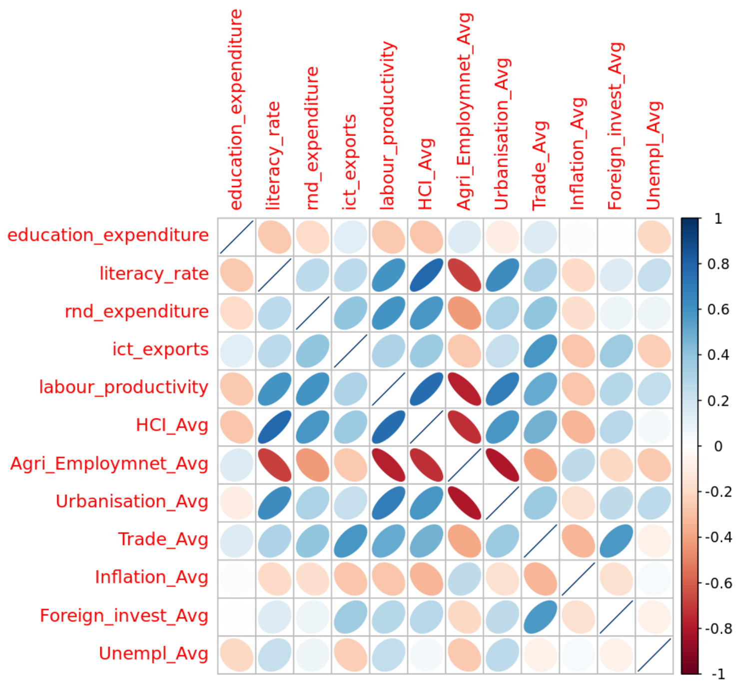

Figure 2: The corellation plot of all the variables

Table 1 shows that the correlation coefficients calculated for all the selected variables support our previous conclusion that the dependent variable is most correlated with literacy rate (0.6, positive correlation), R&D expenditure (0.59, positive correlation), HCI (0.77, positive correlation), employment in agriculture (-0.77, negative correlation), urbanisation (0.7, positive correlation), and trade (0.51, positive correlation).

Figure 2 pictures that among independent variables literacy rate is highly correlated with HCI, employment in agriculture, and urbanisation; trade is correlated with ICT exports and foreign direct investments;

4.3. Initial full model

We proceed with the analysis by fitting a multivariate regression model to our data.

```
Call:
lm(formula = log(labour_productivity) ~ ., data = data_for_model)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.51315 -0.15159  0.01463  0.16999  0.47272 

Coefficients:
                                  Estimate Std. Error t value Pr(>|t|)    
(Intercept)                      2.8455087  0.4675948   6.085 9.85e-08 ***
education_expenditure           -0.0158281  0.0085220  -1.857  0.06834 .  
literacy_rate                   -0.0019144  0.0044894  -0.426  0.67138    
rnd_expenditure                  0.0107659  0.1115582   0.097  0.92345    
ict_exports                      0.0076045  0.0046700   1.628  0.10886    
HCI_Avg                          0.5306506  0.6533693   0.812  0.42001    
Agri_Employmnet_Avg             -0.0161698  0.0034857  -4.639 2.04e-05 ***
Urbanisation_Avg                 0.0064445  0.0027922   2.308  0.02458 *  
Trade_Avg                        0.0010666  0.0009548   1.117  0.26858    
Inflation_Avg                    0.0170959  0.0063495   2.692  0.00926 ** 
Income.groupLow income          -1.1305211  0.2311790  -4.890 8.35e-06 ***
Income.groupLower middle income -0.7769811  0.1410573  -5.508 8.67e-07 ***
Income.groupUpper middle income -0.4084636  0.1039729  -3.929  0.00023 ***
Foreign_invest_Avg              -0.0037063  0.0031988  -1.159  0.25134    
Unempl_Avg                       0.0267743  0.0068806   3.891  0.00026 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2568 on 58 degrees of freedom
  (193 observations deleted due to missingness)
Multiple R-squared:  0.9339,	Adjusted R-squared:  0.918 
F-statistic: 58.55 on 14 and 58 DF,  p-value: < 2.2e-16
```

Since labour productivity is measured in money made per hour worked, applying a log function to the dependent variable is the best practice.

By looking at the full model, which includes all predictors, we can see that the R-squared is high, meaning that our model explains the relationship between variables in 91.1% of cases. So far with the 0.05 level of significance significant predictors that affect the response variable are education expenditure urbanisation, employment in agriculture, inflation, trade, unemployment and level of income. To be accurate and keep only statistically significant predictors, we used the backwards selection.

4.4. Reduced model

```
Call:
lm(formula = log(labour_productivity) ~ ., data = data_for_model_reduced)

Residuals:
    Min      1Q  Median      3Q     Max 
-0.5113 -0.2195  0.0011  0.2019  0.6600 

Coefficients:
                                 Estimate Std. Error t value Pr(>|t|)    
(Intercept)                      3.077940   0.245557  12.535  < 2e-16 ***
Agri_Employmnet_Avg             -0.018727   0.003575  -5.238 1.82e-06 ***
Urbanisation_Avg                 0.006492   0.002918   2.224 0.029548 *  
Income.groupLow income          -1.099295   0.198397  -5.541 5.66e-07 ***
Income.groupLower middle income -0.824891   0.118772  -6.945 2.03e-09 ***
Income.groupUpper middle income -0.451769   0.093198  -4.847 7.94e-06 ***
Unempl_Avg                       0.023971   0.006405   3.743 0.000384 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2761 on 66 degrees of freedom
Multiple R-squared:  0.9131,	Adjusted R-squared:  0.9052 
F-statistic: 115.5 on 6 and 66 DF,  p-value: < 2.2e-16
```

After applying the backward selection procedure, we are keeping the following predictors: employment in agriculture, urbanisation, level of unemployment, and level of income.

4.5. Exploration of interactions or quadratic effects

After checking every possible interaction and quadratic effect, we found only one interaction between the level of income and unemployment variables.

The reference category for our factor variable is the high-income group.

```
Call:
lm(formula = log(labour_productivity) ~ . + Unempl_Avg:Income.group,
    data = data_for_model_reduced)

Residuals:
     Min       1Q   Median       3Q      Max
-0.51061 -0.17442  0.01991  0.18365  0.52771

Coefficients:
                                            Estimate Std. Error t value Pr(>|t|)
(Intercept)                                 3.205693   0.280530  11.427  < 2e-16 ***
Agri_Employmnet_Avg                        -0.015856   0.003369  -4.706 1.43e-05 ***
Urbanisation_Avg                            0.006565   0.002736   2.399  0.01940 *
Income.groupLow income                     -1.760085   0.280432  -6.276 3.58e-08 ***
Income.groupLower middle income            -1.203596   0.232513  -5.176 2.51e-06 ***
Income.groupUpper middle income            -0.553611   0.191563  -2.890  0.00528 **
Unempl_Avg                                  0.005734   0.017600   0.326  0.74565
Income.groupLow income:Unempl_Avg           0.081920   0.025961   3.156  0.00246 **
Income.groupLower middle income:Unempl_Avg  0.039588   0.024439   1.620  0.11025
Income.groupUpper middle income:Unempl_Avg  0.008806   0.019020   0.463  0.64495
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2534 on 63 degrees of freedom
Multiple R-squared:  0.9301,  Adjusted R-squared:  0.9201
F-statistic: 93.11 on 9 and 63 DF,  p-value: < 2.2e-16
```

4.6. F-test: determining the final model

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
model_initial	58	3.823748	NA	NA	NA	NA
model_reduced	66	5.030526	-8	-1.2067779	2.288106	0.033367724
model_final	63	4.045698	3	0.9848278	4.979409	0.003840625

Table 2: F-test results comparing the initial, reduced and final models

By taking a look at Table 2, the F-test confirms that our final model is better than the initial (full) one because our final model has the lowest p-value, which is lower than the level of significance of 0.05, we accept the alternative hypothesis and keep the final model for further analysis.

H_0 (Null hypothesis): all the regression coefficients (except the intercept) are zero, meaning that the predictor variables do not have any significant effect on the response variable.

H_1 (Alternative hypothesis): at least one of the regression coefficients is not zero, indicating that the predictor variables collectively have a significant effect on the response variable.

4.7. Testing for collinearity

```
> vif(model_final)
Agri_Employmnet_Avg                5.33049069261892
Urbanisation_Avg                   3.17421093568276
Income.groupLow income             8.72331010514207
Income.groupLower middle income    11.4170957771371
Income.groupUpper middle income    9.72230877119998
Unempl_Avg                         10.318019442093
Income.groupLow income            Unempl_Avg           4.12447383392239
Income.groupLower middle income   Unempl_Avg           8.02489472860826
Income.groupUpper middle income   Unempl_Avg           17.4845781685745


> vif(model_reduced)
Agri_Employmnet_Avg                5.05728397436518
Urbanisation_Avg                   3.04179289241486
Income.groupLow income             3.67857590990215
Income.groupLower middle income    2.51000321438684
Income.groupUpper middle income    1.93885839105324
Unempl_Avg                         1.15126504973655
```

To check whether there is a problem of collinearity in our model, the Variance Inflation Factor (VIF) test is used. In the final model, some VIF-values are higher than 5, hence it may indicate the problem of multicollinearity; however, it is a result of introducing the interaction between unemployment and level of income.

To prove it we check the reduced model, and after finding out that there is no problem with collinearity, we keep all the predictors.

4.8. Testing for regression analysis assumptions

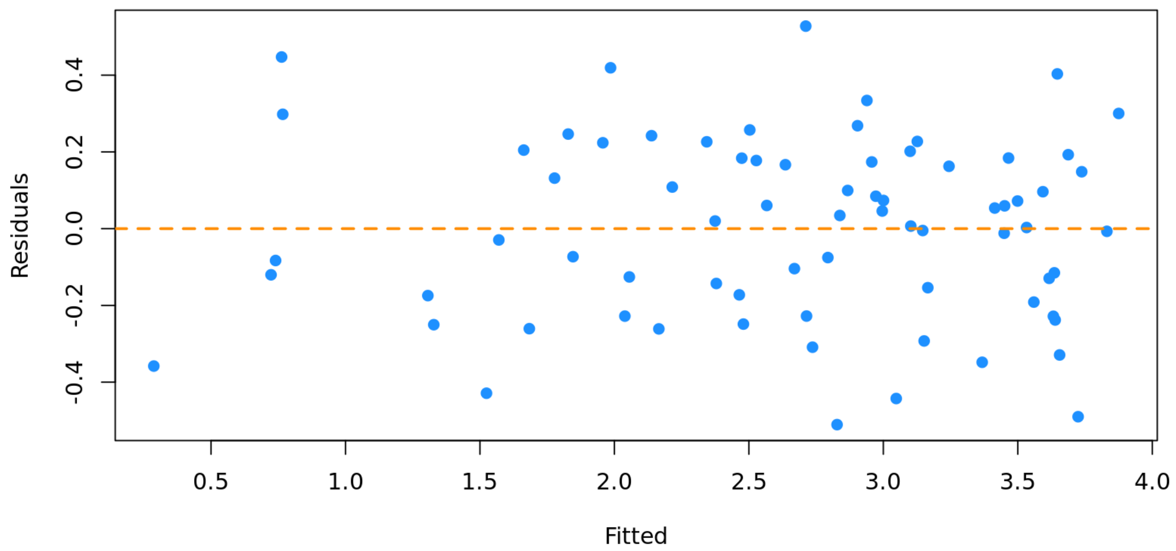

Figure 3: The fitted & residuals plot of the final model

From Figure 3 the following assumptions can be confirmed:

- The assumption of linearity is true since at any fitted value, the mean of the residuals is about 0.
- The assumption of constant variance in residuals is also true since at every fitted value, the spread of the residuals is roughly the same.

```
	studentized Breusch-Pagan test

data:  model_final
BP = 5.2761, df = 9, p-value = 0.8096
```

To check the assumption of constant variance in residuals analytically we used the Breusch-Pagan test, which has the following hypothesises:

• H0: Homoscedasticity. The residuals have constant variance about the true model.

• H1: Heteroscedasticity. The residuals have non-constant variance about the true model.

The p-value of the Breusch - Pagan test statistic is 0.809, which is higher than the level of significance of 0.05. Therefore we accept the null hypothesis, indicating that the residuals have constant variance.

- Normality in residuals distribution

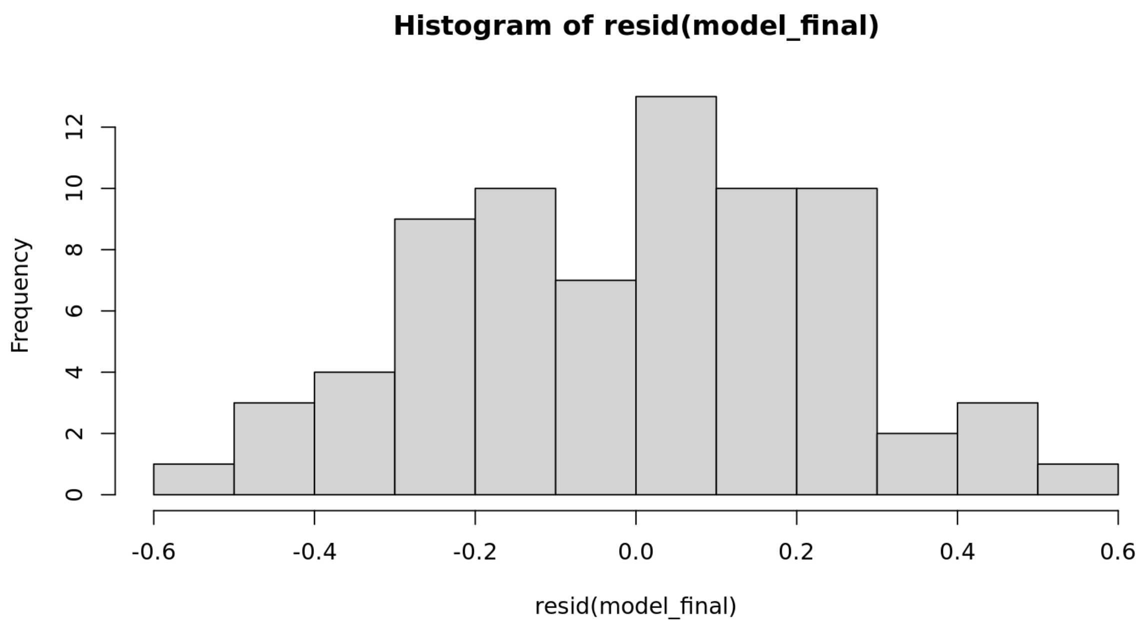

Figure 4: Histogram of residuals of the final model

Due to the small number of observations, we cannot make clear conclusions based on the histogram of residuals. Figure 4 mostly follows the bell-shaped pattern, however, there are some unexpected bends.

- Q-Q plot

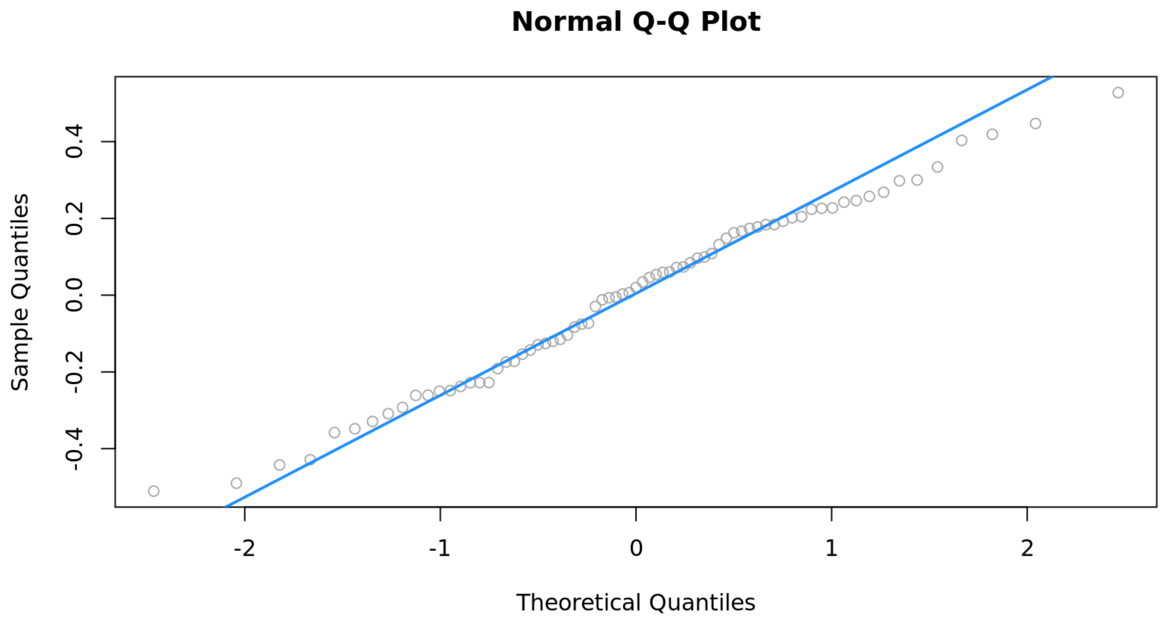

Figure 5: Q-Q plot of residuals of the final model

The Q-Q plot pictured in Figure 5 suggests that most points closely follow a straight line, hence we can formerly assume that residuals are distributed normally.

```
  Shapiro-Wilk normality test

data:  resid(model_final)
W = 0.98583, p-value = 0.5884
```

However, to be sure of our assumption we used the Shapiro-Wilk test. The p-value for W-statistic is larger than the significance level of 0.05, we accept the null hypothesis, meaning that the residuals are normally distributed, indeed.

4.9. Exploring the model quality

According to R-squared our model explains the relationship between variables in 91.1% of cases. Also, in the model, there are seven points of large leverage and four points with large residuals.

There are five influential points in our model, that stand for Burundi (9), Mauritius (40), Nepal (46), Sudan (66), and Uruguay (72). After removing them from our model, we can be sure that they do not influence coefficients enough to remove them from the final model.

Therefore, we keep the model:

$$\log(\text{Labour\_productivity}) = 3.205 – 0.015 \cdot \text{Agri\_Employmnet\_Avg} + 0.007 \cdot \text{Urbanisation\_avg} +$$
$$+ 0.006 \cdot \text{Unemployment\_avg} - 1.76 \cdot \text{Low\_income} – 1.203 \cdot \text{Lower\_Middle\_income} –$$
$$- 0.553 \cdot \text{Upper\_Middle\_icome} + 0.082 \cdot \text{Low\_income} \cdot \text{Unemployment\_avg} +$$
$$+ 0.039 \cdot \text{Lower\_Middle\_income} \cdot \text{Unemployment\_avg} +$$
$$+ 0.008 \cdot \text{Upper\_Middle\_income} \cdot \text{Unemployment\_avg}$$

4.10. Model interpretation

To interpret the coefficients correctly we exponentiate them:

The fitted regression line for high-income countries:

$$\text{Labour\_productivity} = 24.65 - 1.015 \cdot \text{Agri\_Employmnet\_Avg} + 1.007 \cdot \text{Urbanisation\_avg} + \\ +1.006 \cdot \text{Unemployment\_avg}$$

1. For high-income countries, labour productivity is 24.65 USD per unit of labour, if all the predictors equal zero.
2. For every one-per-cent increase in unemployment, the estimated average value of labour productivity increases by 0.6%, holding other predictors constant.

The fitted regression line for upper-middle-income countries:

$$\text{Labour\_productivity} = 22.91 - 1.015 \cdot \text{Agri\_Employmnet\_Avg} + 1.007 \cdot \text{Urbanisation\_avg} + \\ +1.014\ \text{Unemployment\_avg}$$

1. For upper-middle-income countries, labour productivity is 22.91 USD per unit of labour, if all the predictors equal zero.
2. For every one-per-cent increase in unemployment, the estimated average value of labour productivity increases by 1.4%, remaining other predictors constant.

The fitted regression line for lower-middle-income countries:

$$\text{Labour\_productivity} = 21.32 - 1.015 \cdot \text{Agri\_Employmnet\_Avg} + 1.007 \cdot \text{Urbanisation\_avg} + \\ +1.046 \cdot \text{Unemployment\_avg}$$

1. For lower-middle-income countries, labour productivity is 21.32 USD per unit of labour, if all the predictors equal zero.
2. For every one-per-cent increase in unemployment, the estimated average value of labour productivity increases by 4.6%, remaining other predictors constant.

The fitted regression line for low-income countries:

$$\text{Labour\_productivity} = 18.84 - 1.015 \cdot \text{Agri\_Employmnet\_Avg} + 1.007 \cdot \text{Urbanisation\_avg} + \\ +1.092 \cdot \text{Unemployment\_avg}$$

1. For low-income countries, labour productivity is 18.84 USD per unit of labour, if all the predictors equal zero.
2. For every one-per-cent increase in unemployment, the estimated average value of labour productivity increases by 9.2%, remaining other predictors constant.

For all income groups:

1. For every one-per-cent increase in employment in agriculture, the estimated average value of labour productivity decreases by 1.5%, holding other predictors constant.
2. For every one-per-cent increase in urbanisation, the estimated average value of labour productivity increases by 0.7%, holding other predictors constant.

4.11. Comparing results to the expectations

As expected, urbanisation does positively increase labour productivity. Employment in agriculture also leads to our expectations and negatively impacts labour productivity.

The level of income of the country affects the productivity of labour respectively; meaning that the higher the level of income, the higher the productivity.

Surprisingly, the unemployment rate positively affects the productivity of labour, and its effect increases, as the level of income decreases. It can be explained by the following idea: high wages lead firms to substitute labour with capital. This leads to increasing unemployment and to increase productivity since the workers who are still employed become more productive (Bräuninger, 2002).

5. Conclusions

Achieving the Goal: Insights from the Model

Our model offers valuable insights into the factors influencing labour productivity. We have successfully identified variables such as urbanization, employment in agriculture, level of income, and the unemployment rate that significantly affect productivity levels. By understanding these factors, we have made progress in achieving our goal of comprehending the determinants of labour productivity.

Policy Recommendations

Based on our research findings, we have formulated a set of policy recommendations which can be tailored to address Ukraine's unique circumstances:

- Implementing urban development initiatives is vital to enhance urbanization. It can improve access to essential services and employment opportunities in urban areas, attracting people from rural regions and stimulating economic growth.
- Supporting the transition of the agricultural sector towards more technologically advanced and efficient practices can boost productivity and improve agricultural outputs.
- Promoting economic reforms that facilitate income growth and reduce income disparities is crucial for Ukraine. These reforms can include measures to attract foreign investment, streamline regulations, and enhance the business climate.
- Providing training programs, supporting entrepreneurship, and creating a favourable environment for job creation, ensuring the nation is equipped with the necessary competencies to contribute effectively to the economy.

By adopting these recommendations adapted to Ukraine's context, the country can work towards achieving higher productivity levels, economic growth, and overall development.

Future Directions for Research

Considering all things mentioned above, there are several promising directions for future research in the field of productivity analysis:

- Exploring the causal relationships between the identified variables and labour productivity would provide a deeper understanding of their impact.
- Considering additional factors such as education, technological advancements, and industry-specific variables would broaden the analysis and capture a more comprehensive picture of productivity determinants.
- Conducting comparative studies across countries could identify best practices and valuable lessons from high-productivity nations.
- Utilizing longitudinal data to analyze trends and changes in productivity over time would offer valuable insights into productivity dynamics.

Pursuing these avenues of research will contribute to a more robust understanding of productivity and inform evidence-based strategies and policies for enhancing productivity levels.

6. References

- Bräuninger, M., & Pannenberg, M. (2002). Unemployment and productivity growth: an empirical analysis within an augmented Solow model. Economic Modelling, 19(1), 105–120.
- Choudhry M. T. (2009). Determinants of Labor Productivity: An Empirical Investigation of Productivity Divergence. University of Groningen, The Netherlands.
- Radlo, M., & Tomeczek, A. (2022). Factors Influencing Labor Productivity in Modern Economies: a Review and Qualitative Text Analysis. WSEAS Transactions on Environment and Development, 18, 291–314.

7. Annexes

7.1. Descriptive statistics of all the variables in our models

	N	Mean	St. Dev.	Min.	Max.
education_expenditure	73	15.034	4.258	7.034	24.794
literacy_rate	73	86.760	15.338	30.171	99.855
rnd_expenditure	73	0.478	0.424	0.035	2.120
ict_exports	73	4.682	9.107	0.006	38.179
labour_productivity	73	19.565	13.915	0.932	65.061
HCI_Avg	73	0.570	0.124	0.320	0.883
Agri_Employmnet_Avg	73	27.362	20.465	0.736	88.435
Urbanisation_Avg	73	58.754	19.445	10.915	100.000
Trade_Avg	73	84.549	50.922	21.135	363.118
Inflation_Avg	73	6.518	5.658	0.877	36.305
Foreign_invest_Avg	73	6.383	12.566	0.257	86.951
Unempl_Avg	73	8.487	5.451	0.704	29.145

Annex 1: Discriptive statistics of all the variables in the full model

7.2. Results of our models

	model_initial	model_reduced	model_final
education_expenditure	−0.016* (0.009)		
literacy_rate	−0.002 (0.004)		
rnd_expenditure	0.011 (0.112)		
ict_exports	0.008 (0.005)		
HCI_Avg	0.531 (0.653)		
Agri_Employmnet_Avg	−0.016*** (0.003)	−0.019*** (0.004)	−0.016*** (0.003)
Urbanisation_Avg	0.006** (0.003)	0.006**, (0.003)	0.007** (0.003)
Trade_Avg	0.001 (0.001)		
Inflation_Avg	0.017*** (0.006)		
Income.groupLow income	−1.131*** (0.231)	−1.099*** (0.198)	−1.760*** (0.280)
Income.groupLower middle income	−0.777*** (0.141)	−0.825*** (0.119)	−1.204*** (0.233)
Income.groupUpper middle income	−0.408*** (0.104)	−0.452*** (0.093)	−0.554*** (0.192)
Foreign_invest_Avg	−0.004 (0.003)		
Unempl_Avg	0.027*** (0.007)	0.024** (0.006)	0.006 (0.018)
Income.groupLow income:Unempl_Avg			0.082*** (0.026)
Income.groupLower middle income:Unempl_Avg			0.040 (0.024)
Income.groupUpper middle income:Unempl_Avg			0.009 (0.019)
Constant	2.846*** (0.468)	3.078*** (0.246)	3.206*** (0.281)
Observations	73	73	73
R2	0.934	0.913	0.930
Adjusted R2	0.918	0.905	0.920
Residual Std. Error	0.257(df = 58)	0.276(df = 66)	0.253(df = 63)
F Statistic	$58.545 *** (\text{df} = 14; 58)$	$115.518 *** (\text{df} = 6; 66)$	$93.110 *** (\text{df} = 9; 63)$

Note: *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$

Annex 2: Detailed comparison of key properties of the initial model, reduced model and the final model