Czech University of Life Sciences Prague

Faculty of Economics and Management

Department of Economics

Master's Thesis

Economic growth and human capital: the case of Kazakhstan

Bc. Tolkyn Aubakirova

CZECH UNIVERSITY OF LIFE SCIENCES PRAGUE

Faculty of Economics and Management

DIPLOMA THESIS ASSIGNMENT

Bc. Tolkyn Aubakirova

Economics and Management

Thesis title

Economic growth and human capital: the case of Kazakhstan

Objectives of thesis

1. To examine the relationship between human capital and economic growth.
2. Provide a foundation for a policy guide for the chosen country.
3. Test if there is a strengthening trend of the human capital quality in Kazakhstan.
4. Check the consistency of the findings with endogenous growth theories.

Methodology

The thesis will be divided into 2 main parts.

The first one is a theoretical part that will contain the literature review of previously published work regarding the relationship between human capital and economic growth. It will be accompanied by macroeconomic theory regarding the Growth Models. The first part will also include an analysis of the human capital and economic development of Kazakhstan during the last 2-3 decades, as well as the possible factors that can influence them.

The next practical part is the main focus of the thesis. The secondary data of the last 25-30 years will be used for the construction of an econometric model. The purpose of the model will be to study the relationship between economic growth and human capital. The results will be compared with previously discussed models of other authors and provide general policy implications.

The proposed extent of the thesis

60 – 80 pages

Keywords

Human capital, economic growth, innovation, econometric analysis

Recommended information sources

Eriksson, C., Lindén, J. & Papahristodoulou, C. (2023) Human capital, innovation, and growth. International Journal of Economic Theory, 19, 343– 369. https://doi.org/10.1111/ijet.12346

Chindo, Sulaiman & Bala, Umar & Tijani, Bulama & Waziri, Salisu & Maji, Ibrahim. (2015). Human Capital, Technology, and Economic Growth: Evidence From Nigeria. SAGE Open. 5. 10.1177/2158244015615166

Pelinescu, Elena & Pauna, Carmen & Saman, Corina & Diaconescu, Tiberiu. (2020). HUMAN CAPITAL, INNOVATION AND ECONOMIC GROWTH IN THE EU COUNTRIES. Romanian journal of economic forecasting. 22. 2019.

Xu, Yunfu & Li, Aiya. (2019). The relationship between innovative human capital and interprovincial economic growth based on panel data model and spatial econometrics. Journal of Computational and Applied Mathematics. 365. 112381. 10.1016/j.cam.2019.112381.

Expected date of thesis defence

2023/24 SS – PEF

The Diploma Thesis Supervisor

doc. Ing. Irena Benešová, Ph.D.

Supervising department

Department of Economics

Electronic approval: 4. 9. 2023

prof. Ing. Lukáš Čechura, Ph.D.

Head of department

Electronic approval: 3. 11. 2023

doc. Ing. Tomáš Šubrt, Ph.D.

Dean

Prague on 28. 02. 2024

Declaration

I declare that I have worked on my master's thesis titled " Economic growth and human capital: the case of Kazakhstan" by myself and I have used only the sources mentioned at the end of the thesis. As the author of the master's thesis, I declare that the thesis does not break any copyrights.

In Prague on 27.03.2024

____

Tolkyn Aubakirova

Acknowledgement

I would like to express my gratitude to my supervisor, dear doc. Ing. Irena Benešová, Ph.D. for her valuable support, expertise and significant guidance throughout all stages of the thesis writing and studying process.

I would like to sincerely express my gratitude to all the teachers of the university for their effort and valuable knowledge.

Thanks to my family for their unconditional love and moral support, to my managers Michal and Alzbeta for the opportunity to combine a Master's degree program with a full-time job, and to my friend Kamila, who shared all up and downs of the university days with me.

I am endlessly grateful to my husband Rustem for his love, understanding and being a constant source of motivation and strength.

This support and kindness made this challenging yet successful journey possible.

Abstract

Human capital, as a factor of production, plays an essential role in the country's economic performance, while innovation has become a crucial driver of economic growth. This work aims to study the relationship between human capital, innovation, and economic growth, using the example of Kazakhstan, by conducting empirical analysis.

The thesis summarizes articles on growth theories, return on education, innovation and Research and development investment impact on the quality of human capital, together with the analysis of current human capital-related policy implications in Kazakhstan. Using a time-series of 22 years, based on statistical data from the Bureau of National Statistics in Kazakhstan, the relationship between human capital, innovation and economic growth is analysed using several regression models – the ordinary least squares model and the two-stage least squares model. The latter model was selected as the best-fit estimate due to the endogeneity problem of explanatory variables.

The empirical model identified the dependence of innovation production volumes and R&D expenditures, explaining its significant influence on the innovation level in the country. Moreover, the 2SLS model estimated the positive relationship between innovation level and economic growth through the annual GDP growth rate. Education-related variables also demonstrated a significant positive role in determining the country's economic prosperity.

The studied and estimated positive relationship between human capital, innovation and economic growth advocates for proactive policy interventions and resource allocation for development purposes.

Keywords: Human capital, innovation, economic growth, OLS model, 2SLS model.

Abstrakt

Lidský kapitál jako faktor produktivity hraje klíčovou roli v rámci ekonomické výkonnosti země, zatímco inovace se staly zásadní hybnou silou hospodářského růstu. Tato práce se zaměřuje na empirickou analýzu vztahu mezi lidským kapitálem, inovacemi a ekonomickým růstem na příkladě Kazachstánu.

Táto práce shrnuje teoretické poznatky z oblastí ekonomického růstu, návratnosti vzdělávání a inovací, zkoumá vliv investic do výzkumu a vývoje na kvalitu lidského kapitálu a současnou politiku Kazachstánu v oblasti řízení lidského kapitálu. Za pomocí 22leté časové řady, založené na statistických datech Národního statistického úřadu Kazachstánu, vztah mezi lidským kapitálem, inovacemi a ekonomickým růstem je analyzován s použitím dvou regresních modelů: metody nejmenších čtverců a metody dvoustupňových nejmenších čtverců. Posledně jmenovaná metoda byla vybrána z důvodu její vhodnosti při řešení problému endogenity vysvětlujících proměnných.

Závislost objemu inovací na výdajích do výzkumu a vývoje byla identifikována díky empirické analýze, čímž se vysvětluje jejich významný vliv na úroveň inovací v dané zemi. Kromě toho, metoda dvoustupňových nejmenších čtverců odhalila pozitivní vztah mezi úrovní inovací a ekonomickým růstem, jenž se projevil v tempu růstu ročního HDP. Proměnné související s vzděláváním též prokázaly svoji významnou pozitivní roli při stanovení ekonomické prosperity dané země.

Odhalený pozitivní vztah mezi lidským kapitálem, inovacemi a ekonomickým růstem poukazuje na nutnost proaktivního přístupu a správné alokace zdrojů pro potřeby výzkumu a vývoje.

Klíčová slova: Lidský kapitál, inovace, ekonomický růst, metoda nejmenších čtverců, metoda dvoustupňových nejmenších čtverců.

Table of contents

List of figures and tables

1. Introduction

Economic growth is one of the main macroeconomic focuses. It is clear that plenty of things we care about, such as health, education, living standards and the population's well-being, depend on the quality of economic performance. The various impacts of economic growth on human life also include its broad study from different sides. Identifying and studying its main drivers can be useful in policy formulation and their application for the prosperous development of the country.

Human capital, which consists of the population's knowledge, skills and abilities, plays an essential role in the economy's productivity. At the same time, innovation serves as a source of new ideas, processes and technologies that enhance the country's productivity and make it more competitive in the global market. The higher level of human capital reached through higher investment rates in education and healthcare positively influences technological progress and innovation (Romer, 1990). Therefore, the interconnection of human capital, innovation, and economic growth is worth deeper analysis and study, as well as possible economic stimulation through investments in human capital and innovation.

The thesis aims to study the relationship between economic growth, human capital and innovation. The impact of the latter two factors on the economic performance of the country will be investigated by the analysis of education and health levels, innovational indices and R&D investment rates. Similar research papers will be studied to understand the underlying dependence between human capital, innovation, and economic growth. The relevance of previous findings on the case of Kazakhstan will be tested by the estimation of regression models and some policy recommendations will be based on the results of empirical models.

Kazakhstan is a natural resource-rich country with a high dependence of the economy on those natural resources. This dependence creates challenges for the diversification of the economy and highlights the importance of the transition to a knowledge-based economy. Therefore, the understanding of the connection between the current human capital level and economic performance is a significant step in raising the importance of investment in innovation and human capital development.

The thesis has the following plan. The theoretical part consists of a literature review, where the existing research articles on human capital theory, innovation and human capital components are summarized, with special attention to the topics of R&D investment, return to education and health quality. A similar analysis of other country cases is investigated and the current situation of Kazakhstan in terms of human capital quality, innovation initiatives and existing policies is studied.

The practical part of the thesis is focused on econometric analysis, where the relationship between human capital, innovation and economic growth is studied by regression analysis. Two regression model types – Ordinary Least Squares and Two-Stage Least Squares models are estimated to test the hypothesis of the significant interdependence between human capital, innovation and economic growth. The models are compared and the best-fit model is used for the final interpretation of the estimated relationship and used for further policy suggestions.

2. Objectives and Methodology

2.1. Objectives

Human capital and innovation are essential tools of the state economy, and significance is rising in the era of new technologies and the Internet. The quality of these essential tools has a significant impact on the development possibilities of the country; therefore, the main objective of this thesis is to examine the relationship between human capital, innovation and economic growth in Kazakhstan. The analysis of the relationship concerns the specific features of the Central Asian region, as well as the impact of the transition period. Identifying the positive effects of human capital development and innovation should provide a basis for analyzing the relationship between human capital, innovation and economic growth.

One of the main factors of human capital that is being studied and has sufficient supporting evidence is education, so the thesis's one of partial objectives focuses on the relationship between human capital and economic growth through the impact of education and its quality. In the thesis analysis, education is considered the main estimator of human capital quality.

Innovation and human capital are interconnected source of modern profit generation and their joint impact is assumed to be a significant booster of economic growth. The thesis considers the partial objective regarding the analysis of the relationship between innovation and economic performance by assessing the actual impact of Foreign Direct Investments, patents, and innovation-related public expenditure.

The purpose of the thesis is not only to analyze the existing research related to the relationship of human capital, innovation, and economic growth but also to construct an empirical model that has to investigate the actual relationship between these factors in Kazakhstan during the last few decades. The empirical model is aimed to provide a deeper understanding of the quantitative features of studied relationships. The results of the empirical model provide a possibility to analyze the current trends and patterns in human capital, innovation, and economic growth.

One of the partial aims of the thesis is to draw policy implications based on the findings of the empirical model regarding the estimated impact of human capital and innovation factors on the economic growth of Kazakhstan. The model estimations will provide essential insights regarding the relationship between studied factors, which represents the quality of human capital, innovation efficiency and economic performance, that can be used as a basis for policy implications for efficient application of current state human capital resources and sustainable growth of the country.

2.2.Hypothesis

Human Capital and Education:

- The higher rate of expenditure on education is positively correlated with the GDP growth rate in Kazakhstan over the selected period. This hypothesis is aimed to prove the existence of human capital impact through return on education.
- The higher rate of enrolment in tertiary education is positively correlated with the GDP growth rate in Kazakhstan over the selected period.

Innovation as a driver of economic growth:

- The higher rate of innovational production out of all production is positively correlated with the growth rate in Kazakhstan over selected the period. This hypothesis aims to prove that innovation is an essential source of economic growth.
- The higher number of registered patents and the higher rate of R&D investment are positively correlated with the GDP growth rate in Kazakhstan over selected period through their positive impact on innovational production in the economy.

Capital formation and Unemployment:

- The higher rate of capital formation is positively correlated with the growth rate in Kazakhstan over selected period.

The higher rate of unemployment is negatively correlated with the growth rate in Kazakhstan over selected the period.

2.3.Methodology

The methodology of the thesis is based on a combination of theoretical and practical methods that we applied to study the current state of the relationship between human capital, innovation and economic growth in Kazakhstan. The theoretical part of the thesis applies the revision and study of existing literature that provides the theoretical base for the relationship between studied variables. The general theory examination sheds some light on existing theories of human capital, economic growth, innovation, return on education and other related factors, that provide a theoretical framework for the thesis research.

Another applied theoretical method is an examination of existing trends, patterns and policy implications related to human capital development in the selected country. This method includes the analysis of official publications, policy announcements, reports and literature with a focus on the Central Asian region and Kazakhstan.

The significant step of empirical analysis is data collection from relevant data sources such as governmental databases, official reports and publications. The possible complication of the data collection is the availability of all variables data for the chosen period. The selected time period and the list of statistical variables are specified based on the availability of data.

The exploratory data analysis is focused on descriptive characteristics of variables to understand the tendencies and distribution of key variables. The analysis is conducted based on the evaluation of descriptive statistics results. Another significant evaluation of variables includes the correlation matrix, which provides initial information about the relationship between selected indicators of the relationship between economic growth, human capital and innovation. This evaluation is also needed for the identification of possible data quality problems and multicollinearity issues. In the case of outliers or multicollinearity problems, the necessary variables transformation is applied.

The main part of empirical analysis consists of econometric models. Initially, the relationship study is done by applying the OLS regression model to the selected explanatory variables in the form of indicators of human capital, innovation, capital and labor force. The OLS regression model is used as a base for relationship examination.

The second type of econometric model is the Instrumental variable: Two-stage least squares model. The possible problem of endogeneity is solved by applying this type of regression model. The main differences between the IV-2SLS and OLS models are the selection of instrumental variables and the application of the two stages of model estimation. The best-fit model will be selected based on tests for endogeneity – Hausmann test and Durbin-Wu-Hausman tests. In the case of endogeneity presence, the estimation of the Two-stage least squares model is assumed to be a more accurate model for the estimation of relationships.

The main focus of the thesis is an interconnection of theoretical analysis results based on the existing research and publications with the estimated results of econometric models based on statistical data evaluation. The existing research results of other countries are compared with Kazakhstan's empirical analysis case results and policy implications are concluded based on identified similarities and differences.

2.4. OLS model

The regression model is aimed to examine the effect of one variable on another one with control of other possible variables. Several regression models are used in econometric analysis and Ordinary Least Squares is one of the most frequently used ones. The main aim of the model is to build the best-fit line with the minimization of the sum of squared differences between actual and predicted values, as a result, the deviation between predicted and actual values is minimized. The correctness of the applied method is examined based on several assumptions.

The first assumption is a linear relationship between dependent and independent variables, and there is no exception for independent variables. According to Fox (2015), the case of non-linear characteristics of the variables, does not provide interpretability of the relationship between dependent and independent variables due to the strict rule of unit change in the dependent variable and corresponding change in the dependent variable in OLS.

The second assumption is non-autocorrelation or independence of error terms. In the ideal case of random selection of observations, the case of residuals' autocorrelation should not be a problem, but its existence demonstrates missing essential data or variables and leads to misleading results of variables' significance. Therefore, for best OLS model estimation, the error terms must be independent from each other.

The third assumption is normality or normal distribution of residuals. In the case of not normal distribution of residuals and its concentrated tails, the estimated result of the model shows less efficiency and negatively affects the quality of the model. Moreover, the skewed distribution of residuals does not correspond to the initial characteristics of least squares of the OLS model (Fox, 2015). The normality of the model is usually tested by a graphical representation of residuals on a plot and evaluation of its distribution.

The fourth assumption is homoskedasticity or in other words constant variance of residuals. The problem of heteroskedasticity also comes against the crucial characteristic of the OLS model regarding the least-squares estimation and serves as a reason for incorrect standard error estimations. The problem detection can be done graphically by assessing the plotted residuals and predicted values relationship, as well as the conduction of an econometric test about the homoskedasticity feature of the model.

And the last main assumption to mention is the multicollinearity issue's non-existence. Multicollinearity occurs when at least two independent variables show a significant linear relationship and the change in one of them predicts almost similar change in another one. Therefore, the existence of this issue assesses the isolated effects of one independent variable on the dependent variable almost impossible. The problem of multicollinearity is usually checked by the correlation matrix and evaluation of resulted indices.

The OLS model of this thesis implies the following form:

$$Y_{it} = \alpha + \beta^{*}\Sigma X_{it} + u_{it} \quad (1)$$

Where: Y_{it} – dependent variable

ΣX_{it} – explanatory variables

u_{it} – error term

2.5. Instrumental Variable—Two-Stage Least Squares (IV-2SLS) Model

According to the similar research of impact of public expenditures on education on economic growth in North Macedonia written by Ziberi et al. (2022), Instrumental Variable—Two-Stage Least Squares was selected as the main econometric model for the assessing the relationship of human capital, innovation and economic growth in Kazakhstan. The existing similar research with public expenditure impact on economic growth demonstrates the issue of endogeneity or in other words correlation of some regressors with error term, therefore, it suggests inclusion of Instrumental variables that uncorrelated with error term. Instrumental variables are used for estimation of endogenous variables that then used for the final estimation of studied relationships.

The Two-Stage Least Squares model implies two stages of estimation, where firstly the endogenous variable is estimated by the help of instrumental variables, and this builds the predicted values of endogenous variables. The second stage considers using fitted values of endogenous variables and the exogeneous variables for model estimation of final model. This stage produces the final coefficients and estimates that being evaluated and demonstrates the estimated relationship of studied factors.

The 2SLS model also has several assumptions and requirements that are needed to obtain the best-fit model with efficient estimates. The main requirement is related to identification of the model, which states that the number of instrumented variables must be more than the number of endogenous variables to be able to obtain sufficient fitted values of endogenous variables. Therefore, the identification level can be divided into 3 as under-identified, just identified and over-identified model. The situation of over-identification is not problematic as under-identification.

Another assumption of this type of econometric model is related to relevance of instruments for estimation of endogenous variables. This problem is solved by usage of theoretical base of relationship between tested factors and reference to another similar empirical research. Moreover, the problem of endogeneity should be solved by the usage of instrumental variables and the final model must demonstrate the characteristic of exogeneity.

The 2SLS model of this thesis implies the following form:

- OLS regression: $Y_1 = \beta_0 + \beta_1 Y_2 + B_2 * \Sigma X_1 + u_{it}$ (2)
- 2SLS, first-stage: $Y_2 = \gamma + \gamma_1 X_2 + \gamma_2 * \Sigma X_1 + e$ (3)
- 2SLS, second-stage: $Y_1 = \beta_0 + \beta_1 \tilde{Y}_2 + B_2 * \Sigma X_1 + u_{it}$ (4)

  Where:

Y_1 – dependent variable

Y_2 – instrumented variables

ΣX_1 – exogenous explanatory variables

X_2 – endogenous explanatory variables

u_{it} – error term

3. Literature review

3.1. Evolution of Human Capital

Human capital is a crucial economic driver. It consists of the population's knowledge, experience, and skills that bring value to the economy. It cannot be precisely quantified in accounting measures, but in the modern world of technologies and scarcity of physical resources, human capital is an essential tool for boosting productivity and increasing the value of products and services.

Human capital was first mentioned in Adam Smith‘s “The Wealth of the Nation”, where he talked about the basis of economic growth through improvements in the abilities and skills of workers. One of his main thoughts was the impact of the contribution to the working capital on the population’s income and general welfare (Smith, 1776).

The concept of human capital was known to economists for several decades, but its importance arose only during the last century. Based on the classical assumption, all measurable profit-generating products are treated as capital, even if it does not fall into usual market relations. According to Schultz (1961), this definition of capital allows us to distinguish human and "non-human" capital.

After the Second World War, economists actively consider this theory and explored the examples of different countries. Most of them considered that the primary variable was the technological progress of countries, but Solow‘s analysis in 1956 came to two conclusions based not on technological progress. His first conclusion is that the more savings, the richer the country. The second was that the larger the population, the poorer the country. Solow underestimated population growth because he did not include the concept of human capital in his conclusions (Solow, 1956).

Human beings were not excluded from the possible influencing factors of capital and its quality, but technological progress raised attention to less tangible assets such as education and skills and their interconnection with income inequality. Innovation and the wider availability of technology are other essential driver of economic growth closely connected to the quality of human capital in the country.

Globalization and the internet increased the availability of information and technology; therefore, making them more accessible for all economies worldwide. The current worldwide economy is steadily switching to knowledge-based sectors, and the importance of human capital is rising at a stronger pace.

There is a limited study on the effect of human capital on innovation and technology adaptation, therefore, on the economy's productivity in the case of developing "catching up" countries. The studies highlighted the importance of human capital for developing countries as an essential tool for increasing economy's productivity through innovations and technology integration (Nelson and Phelps, 1966).

In his study, Romer (1986) found, that any growth (in technology or capital) depended on increased knowledge and experience in the long run. Technology was an important part, but due to market competition and stagnation, knowledge made it possible to expand business and create new opportunities with greater returns. Such an influence positively affected the industry in the long term, since people could, through knowledge and experience, create something new, which would then increase the production and economy of the enterprise.

There are many visions and opinions of human capital. Becker (1964) and his work introduced the significance of investing in human capital, its differentiation, and the study of return on human capital. His work highlighted the importance of human capital in adopting technology and its further development that brings innovative discovery. Becker pointed out that capital increases productivity in all areas of the company. There are several visions of capital here, but what links them is that production depends on the different performance skills of the worker.

According to Romer (1990), developing countries‘ ability to apply the innovative progress of leading economies and use it for productivity development significantly depends on the quality of human capital. Technology and innovation play a significant role in economic growth, but are the product of human development. This fact brings the significance of human capital to the main stage.

Azariadis and Drazen (1990) conducted studies in which they also noticed the influence of human factors on economic growth, but in their work, they did not take the initial knowledge and experience, but the contribution to human capital. They took the economic situation of the two countries as a foundation, without state intervention or other external factors, where they concluded that human capital affects economic growth. A country with a significant contribution to human capital (education, health) overtook the second. The two countries had the same technological capabilities, but the human capital differed. Another important highlight of this research is that political situations or wars highly influence delays in economic development. Investment in human capital usually means investment in the education and health sphere, which supposes a payoff of current investment expenses by a future increase in economic performance. The lack of human capital-related investment is associated with the traditional approach of investing in physical capital due to visible and measurable outcomes. Even though, human capital and its quality are strongly interconnected with physical capital enlargement and the general process of development. The main target of sufficient human capital is to use the existing reserve of physical capital as efficiently as possible and maximize the capability of the equipment and labor force (Bulasheva, 2019).

The country's sustainable economic growth is significantly dependent on the accumulation of human, not physical capital. The significance of this statement has been proved by the Asian tigers that initially focused on the accumulation of human capital with a respective lack of physical capital. This economic plan helped him to maintain a significant growth rate of economic indices that exceeded the expected rates of growth.

This case of "Asian tigers" and its comparison with other countries case highlights that economic growth is not purely dependent on technological progress. There is a need for knowledgeable specialists who can apply the technological advantage to the production process. Lack of experience and absence of necessary knowledge make the appliance of technological progress less efficient or even lead to wasteful usage of the existing technological state (Korsakova, 2016).

3.2. Endogenous growth model

The endogenous growth model assumes that technological progress is not the sole driver of economic growth. The quality and volume of human capital, which is significantly influenced by investment in science, education, and health, has become an important driver of economic growth. Endogenous growth considers the relationship between the mechanisms of economic growth and the accumulation of human capital, which consists of obtaining new knowledge, technological innovations applications.

Paul Romer introduced the concept of endogenous growth and highlighted the importance of knowledge and innovation as endogenous factors in the long-run economic growth (Romer, 1990). The assumptions of the model served as the reason for the divergence of growth rates in developed and developing countries.

The introduction of the concept was followed by empirical testing of the relationship, so Barro (1991) and Kyriacou (1991) tried to estimate the empirical relationship between human capital and economic growth by conducting cross-country analysis. Barro (1991) found a significant impact of the initial level of human capital on the rate of economic growth, while Kyriacou‘s (1991) results demonstrated an insignificant impact. In both studies, different variables were used as estimates of human capital, therefore it complicates the comparison of findings.

The previous studies were concerning cross-country data analysis with different proxy variables for human capital, while Pyo (1995) tested the endogenous growth model with country-specific time series analysis. He concluded that human capital complements the physical capital and labor in production and serves as a significant variable in the estimation of income growth. Another significant finding is the different impact of human capital growth due to different initial stocks of human capital. According to Kyriacou (1991), it can be explained by the opportunity costs of getting an education in low-income countries and frame the idea that a developing country's convergence is significant only after reaching a certain stock of human capital.

Nureyev (2015) conducted a comparison of different studies of the endogenous growth model showing that the accumulation of human capital based on a particular volume of education can become a source of sustainable economic growth. His important finding is that human capital and physical capital quality can be interchangeable. The case of poor physical capital quality can be compensated by a better quality of human capital, while the high quality of physical capital can be depreciated by the low quality of human capital. South Korea and Taiwan compensated for the lack of physical capital development with better quality human capital, which is assumed to be the most important factor in the rapid growth of their economies in the 60s-80s of the previous century.

Dauda (2010) examines the relationship between human capital and economic growth in Nigeria based on the endogenous growth model. He used enrollment in primary, secondary, and tertiary education as proxies for human capital and the estimated model proved the long-term positive relationship between economic growth and human capital variables. The result of the analysis states that there is a significant relationship between enrollment rate and economic growth and model estimates were recommended as a foundation for budgetary allocation tools.

3.3. Education and returns to education

One of the possible ways of increasing human capital is education. Education influences the ability to receive, process, and use information, while this ability has an impact on the quality of performed tasks. According to Nelson and Phelps (1966), a more progressive economy that includes the most up-to-date technologies in production requires a more educated labour force to adapt to the changes and to bring innovation.

Consideration of education‘s impact on the economy is essential from two perspectives. Its benefit has a dual effect of increasing the efficiency of the individual, which is considered a private good of an individual, and also with a positive externality on society, which makes education public good (Musgrave, 1969).

The existing impact of education can be seen from 3 possible channels. Firstly, the positive impact on labour force increases productivity, therefore tasks are covered faster than before. Secondly, higher education provides conditions for the transfer and exchange of existing knowledge and technology. Thirdly, it also positively influences the creativity and thinking process of the population, which creates an opportunity for the creation of new products, processes and technologies (Grant, 2017).

There is also an important distinction between basic literacy and higher education. Basic education is necessary for existing knowledge application and skills development, while higher education is more related to innovative technology. The empirical study demonstrated that the high human capital level of the country not only leads to breakthroughs in science in technology but also makes the implication of existing knowledge from other countries smoother and more effective.

Benhabib and Spiegel (2005) studied that a lower share of the population with tertiary education negatively impacts the adaptation pace of innovation in the economy. There is another distinction between technology and innovation adaptation. In the technology adaptation case, there is less requirement for high education or specific skills, while innovation adaptation is a more complex process that requires a highly qualified labor force.

The education level has a various impact on low- and high-income countries. According to Ang et al. (2011), innovation adaptation is correlated with education only in the case of a high-income country, while for low-income countries, this correlation is statistically insignificant.

UNESCO's "Higher Education Global Data Report" demonstrated that 60% of the average income difference comes from the difference in education quality and only 40% depends on other social factors. The case of Pakistan, studied by Khan and Rehman (2012), demonstrated that proved significant positive relationship between economic performance and secondary education, which was taken as an estimated variable of human capital.

Higher education quality and better accessibility can increase the participation rate of a qualified labor force in the economy. This increases the economy‘s efficiency with the same amount of labour force. Education can provide essential knowledge for innovative breakthroughs and positively impact the economy through innovation. Another positive influence of education can be seen in the technology application. Sufficient technology skills can help to apply existing technology of other countries/spheres and help to catch up with progressive economies (Bulasheva, 2019).

The cost of education is expressed in the investment amount in the educational sphere, which considers schools, universities, training, and other educational institutions. The effectiveness of this investment is measured in labor productivity, such as GDP or income growth. According to Bulasheva’s (2019) examined case of Akmola region in Kazakhstan, there is no significant correlation between the investment amount in education and Gross Regional Product growth. However, the calculated relationship shows a positive correlation that supposes the positive impact of human capital-related investment on economic growth.

They examined the relationship between the return on education, which is one of the most proper indices for human capital, and progress in the use of technology through the technological diffusion models. As a result, the higher social investment into human capital positively impacted the dynamics of new technology adaptation in the economy. Another point of the research pointed out the positive impact of higher rates of education on existing positive externalities through innovation.

According to Nureyev (2015), the educational system and its quality cannot serve as a pure source of human capital, it requires the possibilities of applying of gained knowledge in production. Education becomes a source of development only when it can be productively used, this point becomes significant due to relatively high public expenditures in education in developing countries (from 15 to 30% of all public expenditures).

The main policy objective of the majority of the countries in the 70s – 90s of the 20th century was to enhance the coverage of education among the population, whereas the problem of its quality remained unconsidered. According to Duflo (2001), Indonesia conducted a large-scale governmental program of opening numerous elementary schools. Following the Sekolah Dasar INPRES program, 61,000 new school buildings were built in 1973-1978. The purpose of the governmental program was to increase the education coverage among elementary school-age children from 69% in 1973 to 85% in 1978. As a result of the program, the actual number of elementary school-age children provided with education reached 84% for boys and 82% for girls.

The research of Duflo (2001) highlights the importance of education quality over quantity. In the above-mentioned example of Indonesia, the number of schools doubled, while the number of teachers increased only by 43%. In many developing countries, a not fully elaborated schooling system negatively impacts human capital growth and the return on education there is significantly lower than in developed countries.

Baldacci et al. (2004), studies the relationship between public education expenditure and education performance in a sample of 120 developing countries. The result demonstrated a statistically significant correlation between government expenditure and returns on education, therefore there is an expected positive relationship between public expenditure on education and economic performance. The specific finding of the study is that the estimated impact of public expenditure is lower in countries with ineffective governments.

3.4. Health impact on economic growth

Education is not the sole influencing factor of human capital quality, there is also a significant impact of health conditions. Good health is necessary for maintaining a sufficient participation rate in education. Not sufficient health parameters sufficiently decrease the positive impact of education.

According to Schultz (2002), there is a considerable difference in cognitive abilities between pupils with low and satisfactory health states. This difference leads to a distinctive difference in initial education-related opportunities but also affects further education attainment, such as gaining tertiary education or learning specific skills. As a result, it affects the total productivity of the sector or the whole economy but also hurts the earning opportunities of the individual.

According to Taniguchi (2003), health and education are interdependent and, promote economic growth. Bad health conditions can shorten the participation period in the labor force due to lower life expectancy and enlarge the share of the population that cannot work. According to Oreopoulos and Salvanes (2011), a better education level tends to encourage people to make efficient decisions regarding health, marriage and even parenting approaches. The positive impact can be seen through a higher patience level and goal-oriented approach reached during education, which helps to act more efficiently and avoid risky behavior.

Moreover, the low health level of the population has a direct negative impact on the economy. The higher rate of population that are out of the labour force due to health conditions increases the amount of social and unemployment payments. In the case of epidemics, high general and infant mortality rates, and an inefficient health care system, there is a heavy burden on the economy due to high expenses on treatment, not health care development that might positively affect productivity in the long run.

According to Bloom and Canning (2003), higher returns on education and on-the-job training due to a healthier population promote more investment in education and training. In the case of a higher probability of illness or premature death, the investment is considered as wasted resources and decreases the incentive to invest. Moreover, a longer lifespan introduces the need for retirement savings, which positively influences the saving rate among the population.

Bloom and Canning (2003) also mentioned the impact of the lower mortality rate of infants on the economy. The lower rate of mortality among infants leads to a decreased birth rate, which allows parents to invest more in the education of children, the job market entrance age of those children is lower and brings a more qualified and educated population to the labour force in the long run.

The correlation between health and economic growth is assessed by considering the average life expectancy and average income. The correlation is expected to be driven by good nutrition and healthcare availability in the case of sufficient economic conditions. The empirical cross-country analysis showed the positive impact of higher life expectancy on further economic growth, where an additional year of life expectancy contributed to a 4% increase in GDP growth rate (Barro & Lee, 1994).

3.5. Innovation and technological advantage

The opinion of strong interdependence of economic growth and innovation has been stable for long time and its impact is evaluated through several directions. According to Geissdoerfer et al. (2018), firstly, innovation promotes breakthrough in technology that positively affects the productivity and efficiency of existing processes. Secondly, it serves as a foundation for the development of new products, processes and services with new value-added for the economy.

The relationship between innovation and economic growth can be distinguished by two hypotheses: "supply-leading" and "demand-following" (Pradhan et al., 2020). The first hypothesis supports the idea that innovation brings new products and services with positive add-value to the economy. The demand hypothesis considers that great economic performance leads to higher rates of investment in innovation to keep competitiveness in the global market. Moreover, economic growth positively influences the trade openness of the country and provides more opportunities to specialize in the sectors with a competitive advantage. The focus on the sectors with competitive advantage increases the probability of innovation and technological advances. Therefore, it is difficult to distinguish the initial driver, which leads to the idea of a two-way effect.

According to Sarangi et al. (2022), there is a country-specific feature in the estimation of the innovation and economic growth relationships, which can influence the strength of the relationship. Despite various ranges of estimated relationships, there is a clear positive correlation between innovation-related expenditures and economic growth, which highlights the need for macroeconomic policies and legal frameworks that aim to encourage innovation in the country.

The above-mentioned studies refer to the relationship between innovation and economic growth, but innovation itself is influenced by several factors. One of them is Intellectual property rights, which includes patents, copyrights and other registered trademarks. The setting of intellectual property rights regulations is a complex process and its foresight depends on the country's legislative regulation. Therefore, the well-defined mechanisms of intellectual property rights are expected to promote innovation (Gerguri and Ramadani, 2010). Maskus (2000) stated that patent numbers in open economies have a positive impact on the growth rate, while Gould and Gruben (1996) identified no strong correlation between intellectual property rights and economic growth.

The possibility of innovation use in production significantly depends on several factors. There is a requirement of preparedness and enough knowledge base among the labour force for the application of innovative technology. According to Wernerfelt (1984), the relationship between innovation and human capital can be studied in the individual firm with a combination of physical and non-physical capital. The second type of resource, which includes human capital, creates the "core competencies" of the firm, based on the collective skills and gained experience (Prahalad and Hamel, 1990).

Human resources help firms to implement technological advantages from outside in their production and use existing technologies to adapt new techniques to gain comparative advantage. (Del Canto & Gonzalez, 1999).

Innovations are expected to have direct and indirect effects on the economic output. Firstly, it introduces more time/cost-efficient facilities or tools in its sphere and creates a foundation for further development. Secondly, the spillover of knowledge creates a positive externality to other spheres of the economy that might not be foreseen initially but might have a significant impact on the country's economic performance (Pritchett, 1996).

In his works, Joseph Schumpeter (1934) introduced a deep view of the interconnection between economic growth, innovation, and entrepreneurship. From his view, innovation is a necessary dragging point, that is required to break the static phase of the economy and bring dynamic development. According to (Śledzik, 2013) in the "Theory of Economic Development" and following works, Schumpeter described innovation as a main driver of development through structural changes. Based on his assumption, the creation of profit and advantage is fully dependent on innovation. Another important point of his theory is that innovation is not only focused on the introduction of a new product or new method of production but also includes the creation of a new market, the determination of new raw materials source, and even the change of industry structure that makes it more efficient and less monopolistic.

One of the most important points of Schumpeter's theory was that in the capitalistic system with plenty of producers and consumers, innovation is a foundation for profit-making business and economic development. Due to structural not quantitative changes in the sectors, innovation increases efficiency and brings profit. The ability to new business inventions is strongly dependent on the human capital stock of the firm and affects the firm's competitiveness in the market (Deakins and Whittam, 2000). This condition applies to the government as well, since the efficient policy implication and effective governance are also influenced by the quality of human capital resources of the government.

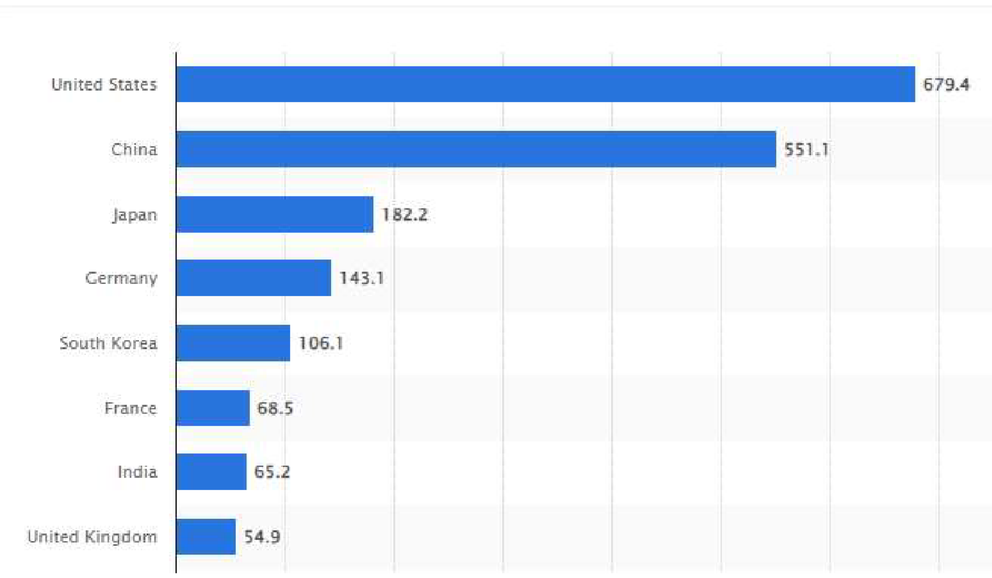
Based on the interconnection of comparative advantage of entrepreneurship and innovations, innovative production requires highly qualified professionals who can apply their technical skills and efficiently provide service. Innovations create a basis for new jobs closely related to new technologies and requiring a specific education. This shows an interrelation between innovative progress and education level that cannot lead to sufficient growth without one of the components. The main existing issue of modern economies is a lack of educated labor that can be on the same level as technological progress. This points out the need for investment in education and job training. More intensive technological growth makes this gap bigger and leads to more significant problems in labor force quality (Korsakova, 2016).

3.6.R&D

According to Pece et al (2015), the amount of investment in R&D influences the economic performance of the country through increasing the comparative advantage and competitiveness of the economy. The higher rate of investment is associated with better productivity in the focused sector, but also increases the country's overall living standards. This research also considers the inverse relationship between economic growth and innovation. The level of R&D expenditure is significantly dependent on several economic factors, such as the existence of monopoly regulation that stimulates market competition. The required institution of competition provides more incentives for businesses to invest in R&D and gain competitive advantage through innovation and related patents.

R&D is considered the main input factor of innovation and therefore an input factor into the economic growth of the country. According to Gerguri & Ramadani (2010), the relationship between R&D and innovation is not straightforward and complex and the R&D investment is used to demonstrate the policy efficiency regarding innovation. The OECD data regarding top R&D Spenders demonstrate that top spenders tend to be the innovation leaders as well (Figure 1).

Figure 1. Leading countries by gross research and development (R&D) expenditure worldwide in 2022 (in billion U.S. dollars)



Source: Statista

According to Del Canto & Gonzalez (1999), human capital is a necessary factor for the full realization of R&D functions. The empirical study of the influence of financial and intangible resources on R&D investment within a firm. The results showed that R&D investments and innovative strategies significantly depend on the quality of intangible resources such as human capital and commercial resources. Surprisingly, the availability of financial resources did not significantly correlate with the R&D rate.

A stable economic situation with foreseen inflation positively impacts investment levels in general, which positively affects the sectors of education and technology. The country's strong economic and political situation attracts more FDI which might be an efficient tool for developing productivity in different spheres. The availability of funds supports public expenditure on technology and R&D that spreads in private sectors and brings innovations to the whole economy (OECD, 2007).

The Pece et al. (2015) analysis confirmed the relationship between economic growth, innovation, and R&D investment, based on the data from Central and Eastern European countries. Their findings provided a supportive case for the endogenous growth model and highlighted the positive impact of FDI on knowledge transfer and increase of technology level.

According to Aghion and Howitt (1998), the increasing technology growth also requires an increase in R&D investment to keep the innovation level increasing or at least constant. The empirical analysis demonstrated the positive relationship between R&D investment and total factor productivity in the case of the United States.

The specific impact of R&D investment on economic growth can be distinguished between developed and developing countries. The Porter and Stern (2000) findings suggest that an increase in innovation tends to have similar effects in both developed and developing countries, while the impact of the increase in R&D investment is sufficient only for developed countries. It concludes that R&D investment is a source of innovation development in the case of developed countries, while developing countries promote their innovation through the existing tools and mechanisms of other leading countries.

4. Current state of human capital and innovation in Kazakhstan

4.1. Human capital impact in Kazakhstan

When comparing the human capital of different countries, culture, institutions, and governmental settings have a significant but sometimes misleading impact. The analysis of predefined regions allows us to isolate the effect of human capital and discover its "pure" impact. There is a lack of study regarding Kazakhstan's case of human capital and economic performance relationship.

Mukhamedzhanova (2001) described the high importance of human capital for the economic development of Kazakhstan. She mentions the potential of human capital as of the key element innovation process and in competitiveness national economy.

Turganbayev (2023) empirically tested the interconnection between human capital and economic growth, where human capital consisted of health and education indication, using the most available statistical data of independent and capitalistic Kazakhstan. The research result demonstrated an insignificant direct effect of human capital proxy variables on the economic performance of Kazakhstan, leading to the idea that the quantitative feature of physical capital and labour shows a more significant impact rather than the quality of labour force. It points out the existing issue of extractive sector dominance. Another estimation result was that the effect of human capital on the total factor productivity growth rate is significant for health variables but insignificant for education variables. The study also demonstrated that the positive impact of human capital does not come from innovation and structural change but from the implementation of existing technological breakthroughs.

Ensuring sufficient economic growth with a stable rate and optimal level is one of the strategic goals of economic policy. In the case of a relatively new economy such as Kazakhstan, the need for innovative approaches in management identifies unresolved issues regarding the most important factor of economic growth - human capital. This will allow consideration of human resources as an investment asset for individual organizations or the national economy and an investment activation factor. Technological progress, updating the product range, and expanding production are directly related to the investment process.

4.2. FDI and innovation

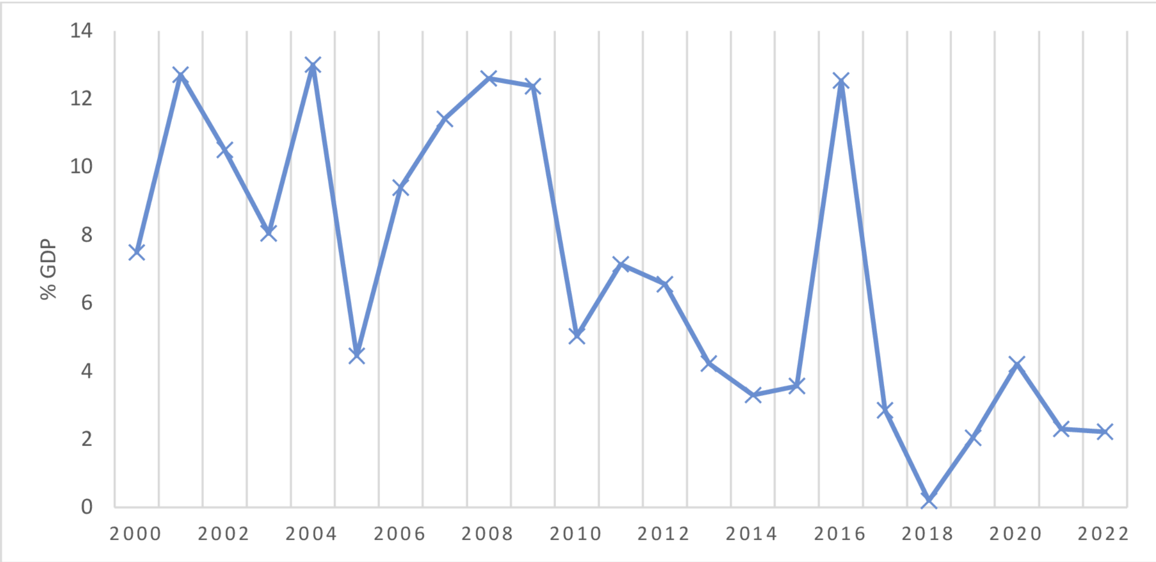
When it comes to investment, there is significance foreign direct investment (FDI), increases the economic potential of the host country, creates a new source of economic growth, ensures the expansion of exports, and creates more jobs. Foreign investment brings not only financial sources but also production and management technologies to the recipient country, which is especially important for countries with economies in transition.

Azam and Ahmed (2015) analyzed Kazakhstan's case in the study of the relationship between human capital and economic growth in 10 CIS countries. Their studies focused not only on human capital's impact on economic growth but also considered the impact of Foreign Direct Investment. The CIS countries were selected due to the recent transition period from a central-state economy to the market economy and the increased inflow of FDI after recently gaining independence. FDI brings financial resources to the recipient country, promotes large-scale projects and long-term transfer of knowledge from the donor country. Based on these characteristics, FDI has a positive spillover in skills and knowledge, which positively influences the state of human capital in the country (Kobrin, 2005).

Another study has examined the interconnection between FDI and human capital. According to Borensztein et al. (1998), the impact of FDI on economic growth is significant only in the case of sufficient technologies in the country. According to Xu (2000), FDI tends to have a positive impact on technology and, consequently on the country's economic growth only when the country has a threshold human capital stock. Azam and Ahmed's (2015) research, which included Kazakhstan's data concluded that the quality of human capital has a significant effect on the absorption of physical capital investment and therefore influences the efficient realization of FDI inflows.

The presented data (Figure 2) shows that due to the dynamics of the world, FDI inflows in Kazakhstan are characterized by instability. Despite the sharp increase in 2016, we can see that the ratio of FDI to GDP in Kazakhstan is generally decreasing. There are several reasons behind this decline, such as the unstable geopolitical situation in the region or insufficient payoffs of previous projects (Bogatyreva, 2018).

Figure 2: Foreign Direct Investment, Net Inflows (% GDP) – Kazakhstan



Source: own proceeding based on World Bank data.

FDI does not yet play a significant role in the economic development of the Eurasian Economic Union countries, primarily because of insufficient economic and legal conditions for this. Much has been done to improve business conditions in recent years, reflected in the World Bank's "Ease of Doing Business rank". According to the 2019 rating, Kazakhstan occupies the 25th position, which is higher than its neighbours, such as Russia (29th place), Uzbekistan (69th place) and Kyrgyzstan (80th place).

According to the Doing Business report (World Bank, 2016), Kazakhstan has lags in the availability of loans for business and taxation process clearness, while the protection interests of minority investors and enforcement of contracts are well established.

Optimizing bureaucratic procedures are already done based on a surprisingly good ranking in the investment climate. Further improvement of the institutional environment for attracting investments lies in the direction of developing corporate governance culture and forming the necessary institutions of a market economy. This predetermines the need to search for another factor in increasing investment activity, which can be done with the help of human capital.

4.3.Current policies and government programs

The legislative support of innovation in the case of Kazakhstan started with the adoption of the first innovation-related law – "State Support for Private entrepreneurship" in 1997. This law distinguished knowledge and technologies as an important factor of production, as well as policy implications regarding the development of innovation-supporting campaigns and centres. This law was replaced by the "Private entrepreneurship" law in 2006, which included improvement regarding the financial support of innovational activity. A significant improvement in legislative support of innovation was reached by the adoption of the "State Support of Industrial - innovative activity" law in 2012. This law reflects the requirement of lawful regulation in the innovation sphere, as well as the implementation of governmental innovation policy. The main focus of the recent innovation-related policies is a modernization of existing enterprises, the establishment of efficient enterprises and financial support of enterprises that focus on innovational activities. There is also a necessity of legal and economic instruments development that encourage small and medium-sized enterprises to be involve in innovational activity (Abdykereyeva, 2012).

The technological and innovation potential of the country can be evaluated by the share of national expenses in science and research out of the total GDP. The annual volume did not exceede 0.28% of total GDP in 2005 and dropped to 0.17%. of the total GDP in 2015. Despite the increase in total expenditure on science, Kazakhstan's parameter is 10 times less than in the average developed countries (Mukhamedzhanova & Kussayinova, 2017).

Kazakhstan's policy objectives are the creation of innovations, the modernization of existing industries and the transition from the natural resources-dependent economy to a knowledge-driven economy. Despite the set objectives, Kazakhstan has insignificant innovational potential and its modernization and development are mainly based on the application of other countries' innovational achievements (Muhamedzhanova, 2001).

The technological progress and its positive impact on the economic growth of Kazakhstan are slowed down by the existing deficit in highly qualified specialists in the ICT sphere. This issue points to the need for efficient and volumetric investment in human capital. According to the "Kazakhstan-2050" strategy, the development and strengthening of human capital is shown as one of 7 main reforms. As part of this strategy, the "Digital Kazakhstan" program for the period of 2018-2022 aimed the digitalization of existing processes and implementation of ICT in almost every sphere of the economy with a quantitative target of getting 83.2% of digital literacy among the population (Koshanov & Chulanova, 2021).

The education sector of Kazakhstan has experienced reforms that made a significant contribution to the qualitative growth of human capital. Part of these reforms are joining the Bologna Process, transferring to three-level training of specialists, international mobility, international accreditation of universities, governmental scholarships "Bolashak" and adoption of the State Program regarding the Development of Education of Kazakhstan for the period of 2011– 2020. Despite the relatively good education parameters of some education variables of Kazakhstan, the quality of secondary and higher education remains low, based on the results of the World Competitiveness Ranking (World Economic Forum, 2023).

The current state of educational spheres highlights the fact that the potential of human capital is not used effectively in Kazakhstan. According to the Ministry of Education and Science of the Republic of Kazakhstan, about 50% of pedagogical graduates are not working in the educational sphere, while on the other hand, there is a significant shortage of qualified teachers. The same problem of shortage is seen in the health care system. The coverage of doctors in hospitals in the biggest cities is 71%, and the situation in rural areas is even worse (Yesimzhanova, 2018).

Yessimzhanova (2018) concludes that social institutional expenditures in Kazakhstan, which form the basis of human capital, are lower than in average developed countries. The development of human capital stock depends on public expenditures rate on education, science and healthcare. Along with this, institutional transformation and modernization of industries need to be supported by the government and its funding.

5. Analytical part

5.1.Data

The selected dataset consists of annual data of 8 indices for 22 years, covering the period between 2000 and 2021. The statistical data is drawn from the sources on annual frequency to eliminate the possible impact of seasonal fluctuations and on a national level to avoid inconsistency due to region-specific characteristics. The latest year with data for all chosen variables is 2021; therefore, the dataset covers the period between 2000 and 2021. The annual data for some of the variables, such as rate of innovational production, patent numbers and enrolment in tertiary education, were unavailable during the procedure of statistical data collection. The availability of the data is also the reason for the starting point of time-series. The statistical data for Kazakhstan in the first 10 years of independence is inconsistent, has some missing years and is not sufficient for statistical analysis. The limited time frame of the selected dataset creates one of the model's limitations.

The dataset is presented in the form of a time-series of 22 years and all data is drawn from the Bureau of National Statistics of the Agency for Strategic planning and reforms of the Republic of Kazakhstan.

5.2.Significance of research

Understanding of relationship between human capital, innovation and economic growth provides essential insights for strategic development through investment in human capital and innovation spheres. The detailed guidance for policymakers can help with the optimal allocation of resources through efficient investment and as a result, sustainable growth. The efficient tools and decision identification in the field studied can improve Kazakhstan's competitiveness in the international arena. One of the benefits of increased competitiveness is the attraction of foreign investment and long-term relationships with partners.

The specific focus of the thesis on education and innovation can provide a basis for educational and innovational policies, which are essential for the diversification of the economy and the sustainable development of all spheres of the economy.

5.3. Variables overview

Observations for eight variables were collected based on similar published research papers, theoretical background and data availability. One variable (***GDP growth***) is considered the main dependent variable, while the other seven are explanatory. The empirical part of the thesis consists of two types of models:

- The ordinary least squares model is where all explanatory variables are considered exogenous explanatory variables.
- Two-Stage Least Squares model, where one variable (***Innovation production***) is instrumented and considered endogenous explanatory variable, while two variables (Patent numbers and R&D expenditures) are instrumental. The rest of the explanatory variables are considered as exogeneous explanatory variables.

Dependent variable:

- **GDP growth** (*GDP*)– this variable is selected as the most suitable estimator of economic growth for this study. It is not the only available estimator of economic growth for a selected country and period, but based on the examples of other studies, this variable works as the most frequently used estimator. The GDP growth rate is expressed in annual percentage change on constant KZT prices (KZT – local currency of Kazakhstan). The GDP is calculated by expenditure method and the growth rate has not been adjusted to population size due to the consideration of one country and the same absence of population adjustment for explanatory variables.

Explanatory variables:

- **Share of innovational production (% GDP)** (*innovation_prod*) – this variable represents the total country's production related to innovational technologies and innovational manufacturing demonstrated as a percentage of GDP in the observed year. This variable is aimed to explain the significance of innovational production in the total production of the country and study the impact of changes in volumes on the growth of the economy.
- **Government expenditure on education (% GDP)** (*educ_expend*)– this variable represents public expenditure related to education, educational institutions, linked

subsidies and transfers, demonstrated as a percentage of total GDP in the observed year. The study of the relationship between education expenditure and GDP growth is aimed to identify the possible impact of investment flows into education on economic growth through return on education and higher literacy level of the population.

- **Share of R&D expenditure (%GDP)** *(RD_expend)* - this variable represents public expenditure related to Research and Development, linked subsidies and transfers, demonstrated as a percentage of total GDP in the observed year. Research and Development activities are aimed at developing new innovational technologies or production or improving the existing ones. Improvement in any sector of the economy is expected to have an impact on the overall performance of the economy.
- **Patent numbers (in thousands)** *(patents)* – this variable demonstrates the number of annually registered patents on the territory of Kazakhstan and represented in thousands. Patent number is used as an estimator of successfulness of development in new technologies and reforms in production processes, therefore also expected to be correlated with changes in economic performance.
- **Gross capital formation (% GDP)** *(CapForm)* – this variable identifies total annual investment in fixed assets by government, households and businesses as a percentage of total GDP in the observed year. It includes expenditures on additional physical capital stock and tangible assets. Based on theoretical background, gross capital formation is an essential predictor of economic growth because this kind of investment positively influences the productive capacity of the economy, leads to job creation and related improvements increases the competitiveness of the economy in the global arena.
- **Unemployment rate (% of the total labour force)** *(Unemployment)* – This variable is expressed as a percentage of the population that is currently unemployed but considered as part of labour force due to their availability for work and being in the process of looking for a job. There are several types of unemployment, such as cyclical, structural and frictional that have different reasons and impacts on economic conditions, but the higher rate of total unemployment rate is associated with stagnation in economic growth.
- **Enrolment in tertiary education** *(Enrollment)* - this variable is demonstrated as the ratio of the population that enrolled in the given year for tertiary education out of the

total population that belongs to the age group that statistically corresponds to the tertiary education level, expressed as a percentage. This variable is aimed to provide the overview of tertiary education coverage that is essential for innovational development. The consideration of tertiary education is crucial due to the fact that the general literacy of the population is a basic characteristic of a modern economy, therefore higher education is needed to gain relative competitiveness and improvements in production.

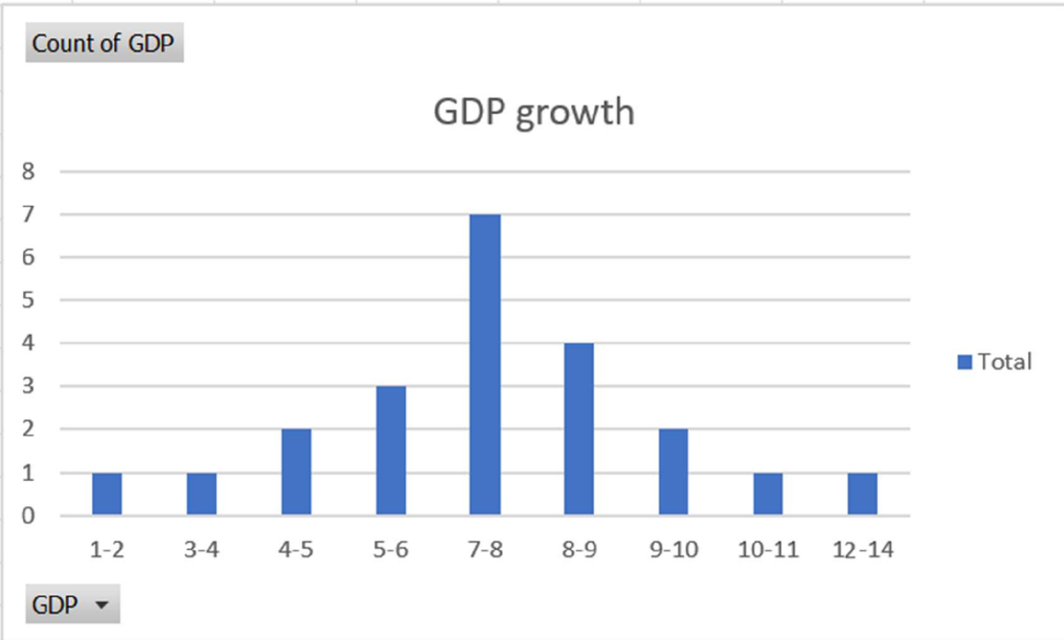
5.4.Variables distribution

The normal distribution of variables must be checked before proceeding with model estimation. Firstly, the normality assumption allows the application of statistical tests and interpretation of their estimation. In the case of a not normally distributed variable, the accuracy of estimation might be lower. Especially when it comes to linear regression models, the estimators are the most efficient only in the case of normal distribution.
The normal distribution of variables is a desired feature for linear regression models; however, the real-world observations do not always provide the normal distribution pattern for all variables. Therefore, it is necessary to check the distribution before model estimation to detect severe deviation from normality and apply necessary variable transformation if it is needed.

Two main possible deviations from normal distribution are skewness and kurtosis of data. In the case of skewness, there is an asymmetry of data distribution around the mean. The observations might be concentrated on the left or right tail, where the mean significantly varies from the median of a variable. The significance of skewness influences the predictability power of statistical tests and estimations. Kurtosis implies deviation of peak from normal a deviation to a higher peak or flatter peak of observations. Sharper or flatter peaks, even in the case of observations symmetry, can also negatively influence the efficiency of estimations.

To check the normality of the variables distribution, the distribution of each variable's observation was constructed in Excel.

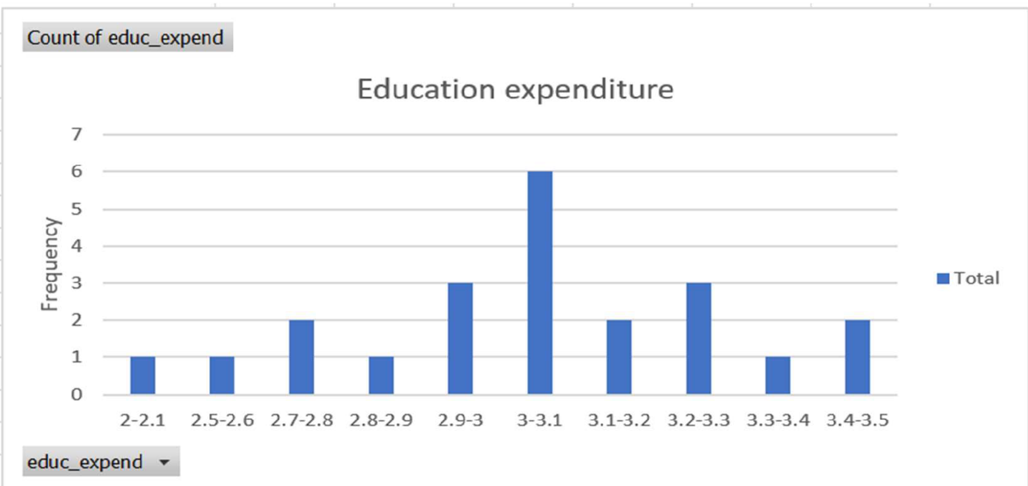
Figure 3. Distribution of "GDP growth" variable



Source: own proceeding based on Bureau of National Statistics of Kazakhstan

The dependent variable – ***GDP growth*** shows the features of a normal distribution with a mean of around 7-8 percentage points (Figure 3). The observations construct an almost perfect bell-shaped normal distribution that does not require any transformation and shows a feature of the efficient estimator.

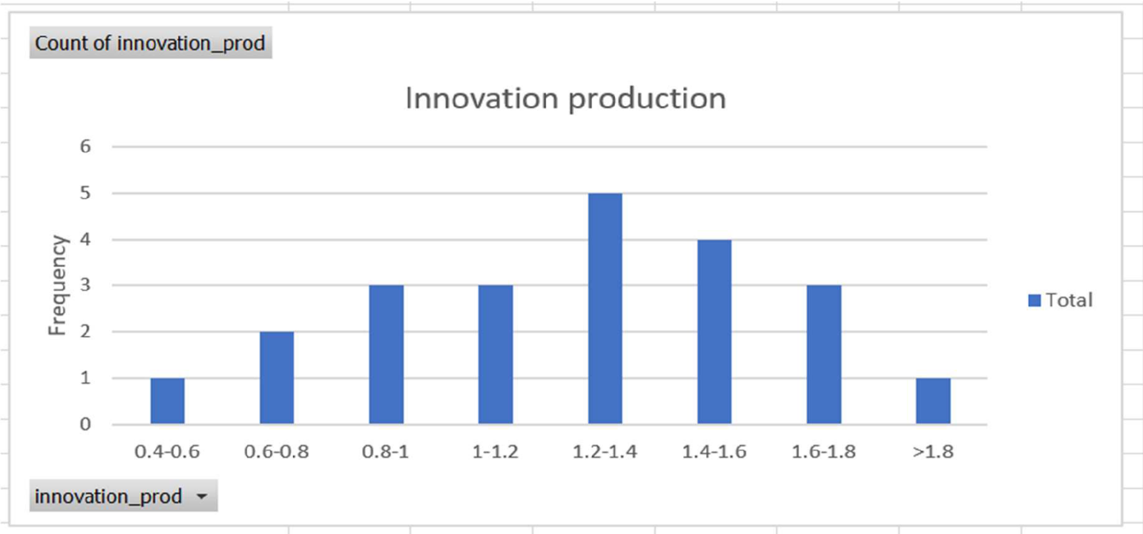
Figure 4. Distribution of "Education expenditure" variable



Source: own proceeding based on Bureau of National Statistics of Kazakhstan

The first evaluated explanatory variable is ***Education expenditure***. The observations are concentrated around the mean of 3-3.1 percentage points and the rest observations are almost symmetrically distributed from both sides of the mean (Figure 4). However, the tails from both sides do not perfectly fluctuate around the mean and some more extreme values show higher frequency, as an example, there are more observations for the 3.4-3.5 percentage points group rather than the 3.3-3.4 percentage points group. The variable demonstrates not perfect normal distribution, but the deviation is not significant, which allows the further usage variable without variable transformation.

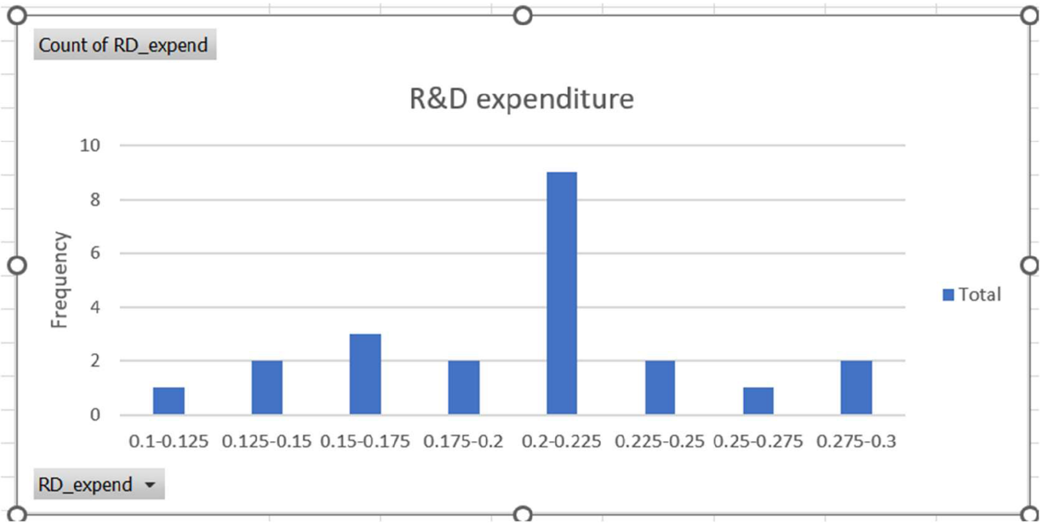
Figure 5. Distribution of "Innovation production" variable.



Source: own proceeding based on Bureau of National Statistics of Kazakhstan

The ***Innovation production*** variable shows a bell-shape distribution, but the mean of the observations is around 1.2-1.4 percentage points group, which is not the perfect middle of the covered values range (Figure 5). It means that the observation is slightly right skewed; however, the degree of skewness is insignificant and allows the consideration of the original form of the variable for model estimation, no variable transformation is required for this case.

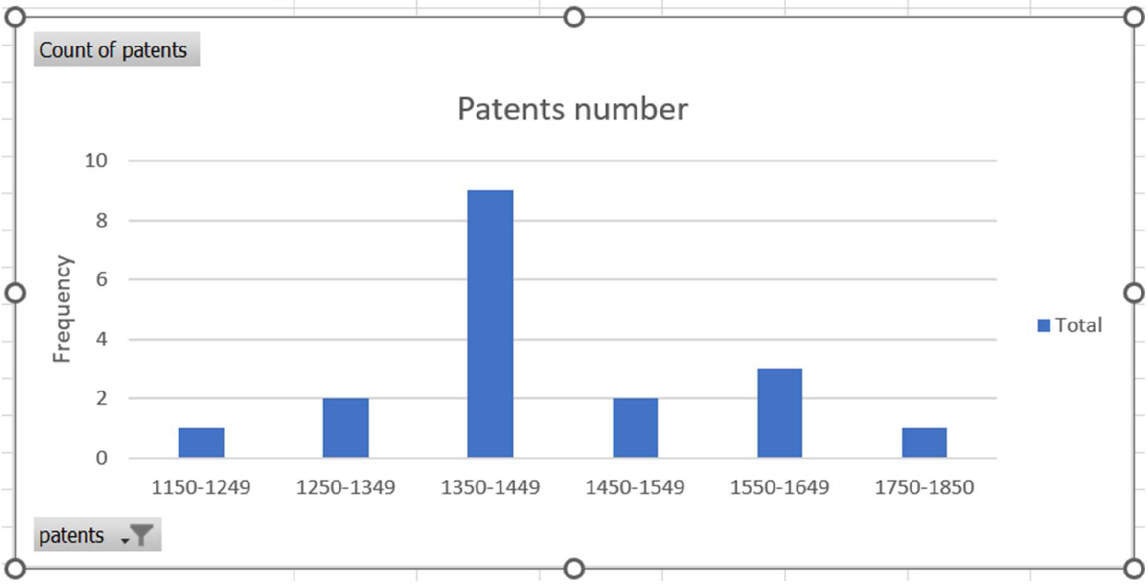
Figure 6. Distribution of "R&D expenditure" variable



Source: own proceeding based on Bureau of National Statistics of Kazakhstan

The ***R&D expenditure*** shows an almost symmetrical distribution of observations around the mean, but there is a sharper peak of observations on the mean and more flattened tails from both sides (Figure 6). Even if it looks like a bell-shaped distribution sharper peak, the difference in absolute frequency is not vital, therefore, the ***R&D expenditure*** variable continues to be used in its original form without considering the issue of kurtosis.

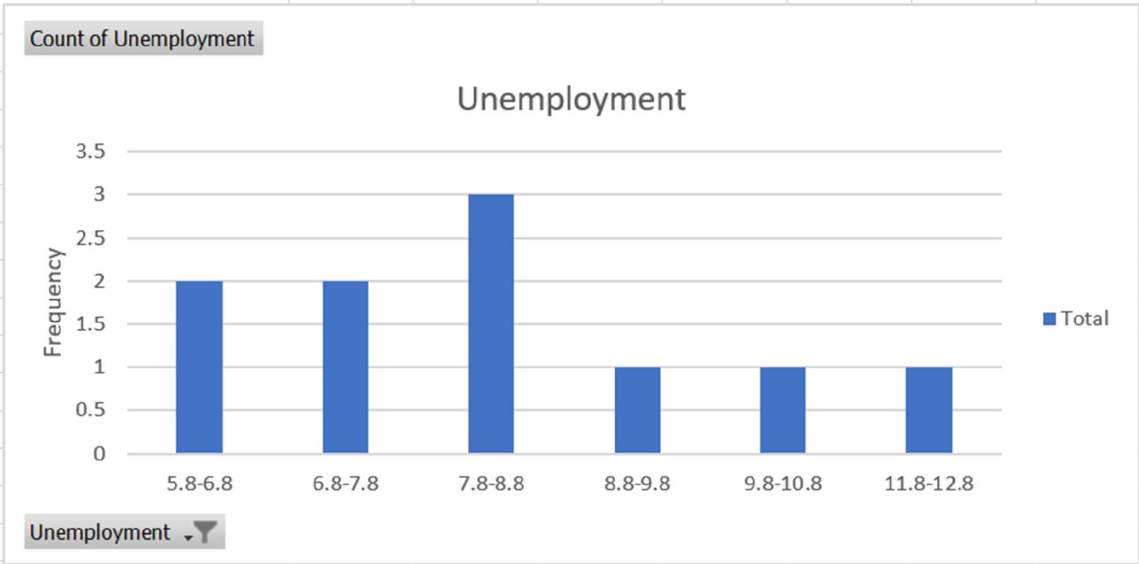
Figure 7. Distribution of "Patent numbers" variable



Source: own proceeding based on Bureau of National Statistics of Kazakhstan

The ***Patent*** variable demonstrates some deviations from the perfect normal distribution, due to the slight negative skewness of observations and the kurtosis feature due to a higher peak around the mean (Figure 7). However, observations are only lightly left-skewed and the sharper peak is not significant in an absolute frequency manner, so the variable transformation is not an important requirement.

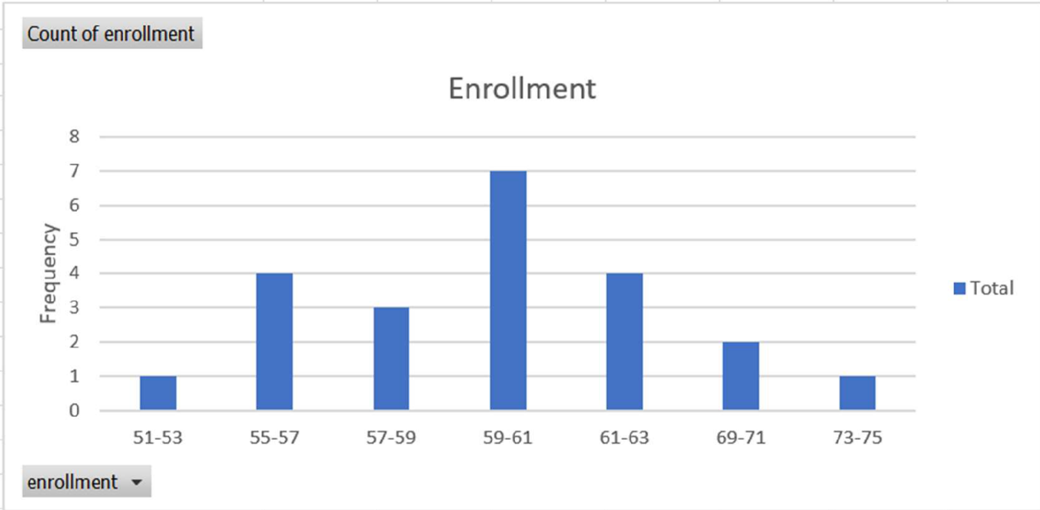
Figure 8. Distribution of "Unemployment" variable



Source: own proceeding based on Bureau of National Statistics of Kazakhstan

The ***Unemployment*** variable is an exact case of observations collected from the real world, where the observations do not always demonstrate perfect normal distribution. The data is left-skewed with more observations on the left tail (Figure 8). The skewness is not considered extremal; therefore, the transformation of the variable is not considered as well.

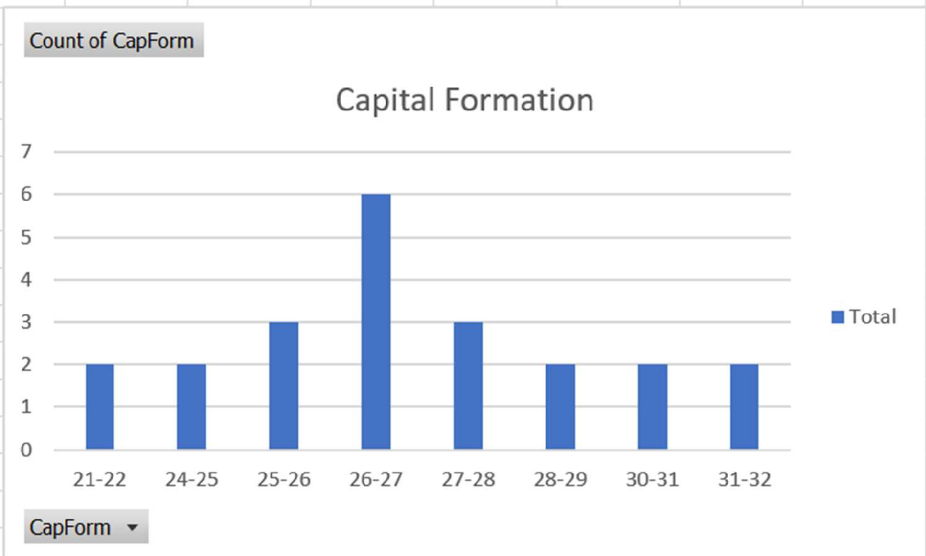
Figure 9. Distribution of "Enrolment" variable



Source: own proceeding based on Bureau of National Statistics of Kazakhstan

The ***Tertiary education*** variable shows a symmetrical distribution around the mean and smooth peak (Figure 9). The group of 55-57 percentage points has a larger frequency compared to the 55-59 percentage points group, which deviates the distribution from a perfect bell shape but does not dramatically worsen the normal distribution of the variable. No variable transformation is applied.

Figure 10. Distribution of "Capital Formation" variable



Source: own proceeding based on Bureau of National Statistics of Kazakhstan

The ***Capital Formation*** variable demonstrates well-defined bell shape with a bit of left-skewness that does not significantly influence the normal distribution of the variable (Figure 10). As a result, the variable distribution is considered normal and proceeded to be used in its original form.

The graphical representation of variables distribution tested the normality of variables distribution, proved the absence of severe cases of skewness and kurtosis and provided evidence for non-requirement of variable transformation.

5.5.Correlation matrix

The correlation matrix is an essential method to evaluate the interdependency between the selected list of variables and might give the information about relationship between dependent and independent variables even before constructing the regression model. It is even more important for the evaluation of the relationship between dependent variables and the identification of possible multicollinearity problems. Therefore, it is necessary to use it for diagnostic purposes before proceeding with the regression model and interpretation of its results.

Figure 11. Correlation matrix of variables

	GDP	Innovation	Education	R&D	patents	CapForm	Unempl	Enrollment
GDP	1.0000							
Innovation	0.2414	1.0000						
Education	0.4355	0.7532	1.0000					
R&D	0.7132	0.6854	0.5665	1.0000				
patents	0.5852	0.4549	0.3115	0.6220	1.0000			
CapForm	0.6833	0.3709	0.3543	0.6983	0.4561	1.0000		
Unemployme	0.7623	-0.1172	0.0902	0.4229	0.2571	0.4539	1.0000	
Enrollment	0.0966	0.7506	0.7063	0.4101	0.0465	0.3756	-0.1723	1.0000

Source: own proceeding based on Bureau of National Statistics of Kazakhstan

The correlation matrix includes our main dependent variable (***GDP growth***) and a selected list of explanatory variables that aim to explain the quality of human capital and the socio-economic situation. The correlation matrix is calculated in the Gretl application and presented in a convenient Excel table (Figure 11).

The interpretation of results can be divided into 2 parts. The first one is the evaluation of interdependence significance between a dependent variable and selected explanatory variables, which is aimed to show the relevance of each explanatory variable for model estimation. In the case of extremely low correlation, the explanatory variable can be excluded from model estimation due to expected low explanatory power. The second part is to identify the possible problem of multicollinearity between dependent variables. In the case of multicollinearity issue (correlation index is higher than 0.8 in absolute terms), a highly correlated variable can be transformed through the usage of its logarithmic form, first difference form or even excluding it from the model estimation.

GDP growth variable shows a relatively significant correlation with the majority of selected dependent variables. The highest correlation is shown with the ***Unemployment rate*** (0.7623), ***R&D expenditure*** (0.7132) and ***Capital Formation*** (0.6833) variables. Some variables such as ***Patent numbers*** (0.5852), ***Education expenditure*** (0.4355) and ***Innovation production*** (0.2414) show relatively moderate correlation, therefore they can be for model estimation and their relevance for model interpretation can be evaluated in the further steps of empirical model estimation. The only variable with a relatively low correlation index is ***Tertiary education enrolment*** (0.0966) variable and based on the general rules of correlation interpretation, the variable is expected to have low explanatory power toward the selected dependent variable. However, one of the regression models used in the evaluation of human capital and economic growth is the Two-Stage Least Squares model, where ***Tertiary education enrolment*** variable will be used as an Instrument for the Instrumented variable (Innovation production). In this case, it is necessary to evaluate the correlation index between the ***Tertiary education enrolment*** and ***Innovation Production***. The correlation index is 0.7506, which is relatively high and represents ***Tertiary education enrolment*** variable as an important explanatory variable for the estimation of the ***Innovation production*** value.

It was stated that ***Innovation production*** is used as an instrumented variable in the Two-Stage Least Squares model, therefore it is essential to evaluate the correlation of this variable with instruments used in the first-stage model estimation. The ***Innovation production*** variable is significantly or moderately correlated with all instruments, except the ***Unemployment*** variable. It is an expected result, since ***Unemployment*** has a more significant influence on the economic situation of the country (which is reflected in a relatively high correlation index with the ***GDP growth*** variable), rather than ***Innovation production*** levels. Based on the first part of the correlation matrix interpretation, it is not necessary to exclude any of the explanatory variables due to their low interpretation power.

The second part of the evaluation is to identify possible cases of multicollinearity among explanatory variables. The issue of multicollinearity exists when the the correlation matrix is higher than 0.8 or lower than -0.8. Based on the result of correlation matrix, there is no visible case of correlation higher than 0.8. ***Education expenditure, Tertiary education enrolment*** and ***R&D expenditure*** show a relatively high correlation with the ***Innovation production*** variable but based on the usage of the ***Innovation production*** variable as an instrument in the Two-Stage Least Squares model, therefore the higher correlation is positively accepted. Other than these cases of high correlation, there ais also a relatively moderate correlation between ***R&D expenditure*** and ***Patents*** (0.622) and ***R&D expenditure*** and ***Capital Formation*** (0.6983). This might be linked to some interdependence of registered patent numbers on the investment in the R&D sphere and dependence of availability of funds for R&D on the Capital Formation of the country. Another case of relatively high correlation is detected between ***Education expenditure*** and ***Tertiary education enrolment*** (0.7063), which is also might be linked to the existing dependence of enrolment rates on investment levels in education spheres. These identified high correlation indices provide the basis for a more sophisticated regression model, where highly correlated variables are used as instruments and instrumented variables for the Instrumental model. However, their correlation indices are lower than 0.8 and the case of multicollinearity is not identified, so all variables can be processed further for empirical model estimation without variables transformation.

Both evaluation steps of correlation matrix provided evidence that all selected independent variables are relevant for planned model estimation and can be used for next steps of empirical model.

5.6.Summary statistics

All variables are presented as percentages, except the Patent number variable, expressed in thousands of units (Table 1). The abbreviation of variables is included in the Variables overview chapter. The average value calculation results show that the mean GDP growth rate for the selected years is 6.33 percentage points, which is quite a pleasant index for developing countries. The average value of education expenditure is 3.02 percentage points, while the R&D expenditure is relatively low and only 0.2 percentage points.

Table 1. Summary statistics of variables

	Mean	Median	Minimum	Maximum
GDP	6.3318	5.4000	1.1000	13.500
innovation_prod	1.2177	1.2450	0.49000	1.7100
educ_expend	3.0212	3.0591	2.0607	3.4570
RD_expend	0.20237	0.21118	0.11654	0.29281
patents	1486.4	1407.0	1193.0	1824.0
CapForm	27.132	26.443	23.828	33.901
Unemployment	6.6933	5.5800	4.8000	12.750
Enrollment	60.355	59.870	51.660	73.180

	Std. Dev.	C.V.	Skewness	Ex. kurtosis
GDP	3.4577	0.54608	0.18373	-0.90735
innovation_prod	0.34877	0.28641	-0.44478	-0.85623
educ_expend	0.30803	0.10196	-1.2544	2.4862
RD_expend	0.045770	0.22617	0.033144	-0.32419
patents	168.60	0.11343	0.35575	-0.85764
CapForm	2.6629	0.098146	0.96902	0.25289
Unemployment	2.1800	0.32571	1.2175	0.77970
Enrollment	4.9958	0.082773	1.0881	1.0452

	5% perc.	95% perc.	IQ range	Missing obs.
GDP	1.1150	13.080	5.5250	0
innovation_prod	0.52000	1.6980	0.68750	0
educ_expend	2.1400	3.4554	0.35807	0
RD_expend	0.11844	0.29143	0.050701	0
patents	1204.7	1811.7	245.75	0
CapForm	23.840	33.545	2.9613	0
Unemployment	4.8075	12.402	3.2450	0
Enrollment	52.284	72.755	4.4675	0

Source: own proceeding based on Bureau of National Statistics of Kazakhstan

This shows a relatively low pattern of human capital and innovation-related expenditure. The ***Innovation production***’s mean value is 1.22 percentage points, which is also might be linked to relatively low investment volumes in this field. The ***Patent number***’s average is 1486 thousand, which is hard to be interpreted due to the absence of other country’s indicators to see the difference and do a comparison. Moreover, the number of patents can be dependent not only on the quality and amount of innovation but also on legislative procedures in the country regarding intellectual rights and patent registration. Another 3 variables demonstrate relatively positive mean values. The ***Capital Formation***'s mean is 27.13 percentage points and the ***Unemployment rate***'s average is 6.7 percentage points only. These indices serve as evidence of sufficient factors for economic growth from a socio-economic point of view.

The median values of the variables do not significantly vary from the mean of the variables, which serves as another proof of the normal distribution of the variables without extreme cases of skewness. The variables with fewer features of skewness in the graphical plot of distribution also demonstrate less difference between mean and median values.

The minimum and maximum values of the variables show the range of observation coverage. The higher range of observations suggests more fluctuations of variables. As an example, the range of the ***R&D expenditure*** is quite narrow, which makes its observations more stable during the selected period. On the other hand, the ***GDP growth*** values have shown more volatile results in the selected period with a range between 1 and 13 percentage points. The variability of the dependent variable might be efficient for examining its dependence on the selected explanatory variables.

The standard deviation indicates the spread of observations from the mean. The standard deviation of the variables is different and based on the mean values of the variables. Therefore, the variables with higher values in general demonstrate larger standard deviations that cannot be compared with each other.

There is a separate estimation of skewness and kurtosis that have been evaluated in the section on variable distribution. As a result, the variables transformation was not considered and relatively low indices of skewness and kurtosis in summary statistics provide support for the interpretation of the non-existence of significant skewness and kurtosis problem among the selected variables.

Another important evaluation is a comparison of minimum/maximum values of variables with their 5%/95% percentiles. The small difference between them leads to the conclusion of the non-existence of outliers that could influence the minimum and maximum indices and that negatively influence the quality of the estimations. The estimation line of linear regression might be significantly influenced by odd values, leading to unexpectedly unusual dependent variables estimated based on the usual observations of dependent variables.

The last important point of summary statistics is to check for missing observations in the sample of each variable. The missing observation can have a negative influence on the interpretability and estimation of the model and therefore requires some value modification or change of time series. The selected variables do not have missing observations that lead to no requirement of missing value replacement.

5.7. OLS model results

According to the stated methodology of the study, to analyze the linear relationship between economic growth, human capital and innovation-related variables, the Ordinary Least squares model is constructed with the assumption of exogeneity and consideration of all independent variables as exogenous explanatory variables. Based on the assumption, the OLS model contains 1 dependent variable and 7 explanatory variables, where the human capital, innovation and socio-economic variables are regressed to estimate their correlation with economic growth.

The following part presents the obtained results and estimations of the OLS model with the selected dataset. The general form of the OLS model equation is the following:

$$Y_t = \alpha + \beta * \Sigma X_{it} + u_t \qquad (1)$$

Where:

Y_t – dependent variable

ΣX_{it} – explanatory variables

u_t – error term

The replacement of the general form of the selected list of variables transforms the equation into this form:

$$\mathbf{GDP_t = \beta_0 + \beta_1 * innovation + \beta_2 * eduexp + \beta_3 * RDexp + \beta_4 * patents + \beta_5 * capform + \beta_6 * unempl + \beta_7 * enrolment + e_t}$$

where *t* means the time period between 2000-2021

The model estimations were estimated using Grelt software; the results are presented in Table 2.

Table 2. Model results of the OLS model (exogeneity case)

Model 1: OLS, using observations 2000-2021 (T = 22)
Dependent variable: GDP

	coefficient	std. error	t-ratio	p-value	
const	−18.3447	7.15341	−2.564	0.0225	**
innovation_prod	−2.27903	2.31386	−0.9849	0.3414	
educ_expend	6.38535	2.23321	2.859	0.0126	**
RD_expend	17.8559	15.7579	1.133	0.2762	
patents	0.00314290	0.00292286	1.075	0.3004	
CapForm	0.397806	0.207608	1.916	0.0760	*
Unemployment	−0.547990	0.217604	−2.518	0.0246	**
enrollment	−0.241663	0.171745	−1.407	0.1812	

Mean dependent var	6.331818	S.D. dependent var	3.457688
Sum squared resid	27.74440	S.E. of regression	1.407744
R-squared	0.889494	Adjusted R-squared	0.834242
F(7, 14)	16.09862	P-value(F)	0.000011
Log-likelihood	−33.76856	Akaike criterion	83.53711
Schwarz criterion	92.26545	Hannan-Quinn	85.59325
rho	0.297717	Durbin-Watson	1.308362

Excluding the constant, p-value was highest for variable 2 (innovation_prod)

Source: own proceeding based on Bureau of National Statistics of Kazakhstan

To proceed with an evaluation of obtained coefficients, it is required to assess the overall significance of the model. The model significance evaluation shows that the p-value of the F-test is 0.000011, which allows to rejection null hypothesis of model insignificance and accept the alternative hypothesis that at least one independent variable has a significant relationship with the studied dependent variable (Table 2). Moreover, the R-squared of the model is 0.8895, which is a relatively high result and demonstrates that almost 89% of outcomes can be predicted by the model.

The model is evaluated for significance and positive results of evaluation allow specification of the relationship of each variable separately. The estimated equation of the model is the following:

GDP = -18.35 – 2.28* innovation + 6.39 *eduexp + 17.86* RDexp + 0.0031* patents + 0.398* capform - 0.548*unempl – 0.242*enrolment

Four explanatory variables (including the constant) demonstrate statistical significance, where three of them (including the constant) are significant on a 5% significance level and 1 variable is significant on a 10% significance level.

The constant variable is negative, and its coefficient is -18.35. It means that in the case of zero values of all other explanatory variables, the ***GDP growth*** is expected to be -18.35 percentage points. The negative sign of the constant is not against the theoretical background since there is no expected case of zero public expenditure, zero unemployment and zero capital formation, therefore the constant coefficient is considered acceptable.

The coefficient of ***Innovation production*** is -2.28 and a relatively high p-value shows its statistical insignificance. The negative coefficient is not in line with the theoretical background, but insignificance makes economic interpretation not needed. The possible reason for this variable's insignificance can be related to a significant correlation with other explanatory variables that does not allow for a test of the pure effect of innovation production level on economic growth rates.

The ***Education expenditure*** variable shows positive coefficient of 6.39 and statistically significant, the result is in line with economic theory regarding positive impact of public education expenditures on economic growth through the higher rates of return on education. Education is a crucial feature of human capital; therefore, the significance of education related input variable is essential and provides further support for larger amounts of investment into education. So 1 percentage point increase in ***Education expenditure*** is associated with 6.39 percentage points increase in ***GDP growth*** rate, ceteris paribus.

The next 2 explanatory variables (***R&D expenditures*** and ***registered patent numbers***) are aimed to explain the quality of innovation in the country and test its influence on economic performance of Kazakhstan. The estimated coefficient of ***R&D expenditures*** is 17.86 and patent number's coefficient is 0.0031. Both variables have positive coefficient that correspond to expect positive relationship between innovation and economic growth, however, they are not statistically significant. The possible reason might be related to correlation with ***Innovational***

production variable; therefore, the estimated model is not able to express the pure effect of each innovation related variable.

The socio-economic variables such as ***Capital formation*** and ***Unemployment rate*** show statistical significance and proves their existing impact on economic condition of the country. The ***Capital formation's*** estimated coefficient is 0.398 and it is positive, which is associated with expected positive impact of investment level on economic performance. So, 1 percentage point increase in ***Capital formation*** rate is associated with increase in ***GDP growth*** rate by 0.398 percentage points, ceteris paribus. The ***Unemployment rate*** is considered as statistically significant variable, the estimated coefficient is -0.578. The negative relationship between economic growth indicator and unemployment rate is expected due to lower productivity in case of higher rates of unemployed population. As a result, 1 percentage point increase in ***Unemployment rate*** is associated with 0.578 percentage points decrease in ***GDP growth*** rate.

The ***Tertiary enrolment*** variable is statistically insignificant and has negative estimated coefficient (-0.24). The theoretical background suggests that higher amount of enrolment into tertiary education provides positive impact on the quality of human capital and as a result, on the economic performance of the country. However, the impact of enrolment can be insignificant due to lagged impact that cannot be fully considered due to relatively small sample size or due to correlation with other explanatory variables.

5.8.Two stage least squares model

The existing research related to human capital and economic growth, as an example study on North Macedonia (Ziberi, 2022), assumes the existence of an endogeneity issue when the equation considers several input and output indicators of one area. In this study, the possible issue of the problem is associated with innovation-linked variables. The ***Innovation production*** variable is considered as the output index of the current quality of innovation support in the country, while ***Patent numbers*** and ***R&D expenditures*** act more as input factor that identifies governmental interest in the development of innovation and their associated resource investment. Moreover, the results of the OLS model estimation also demonstrated an insignificant impact of all innovation-related variables, which supports the idea of possible endogeneity issues due to innovation variables.

To check the presence of an endogeneity issue and test its possible solution method, the Instrument Variables-Two stage least squares model is estimated by considering ***R&D expenditure*** and ***Patent numbers*** as instruments of ***Innovation production*** variable, which is considered an endogenous explanatory variable.

The endogeneity issue is considered an assumption, and it is further tested by comparing the results of the models. In the case of proven issue of endogeneity, the Two-Stage Least Squares model is considered more accurate and has predictable power, therefore, its results will be used for interpretation and policy implications.

The general form of the Two-Stage Least Squares model transformed to model equations by the application of selected variables:

- OLS regression: $\mathbf{GDP_t = \beta_0 + \beta_1 * innovation + \beta_2 * eduexp + \beta_3 * capform + \beta_4 * unempl + \beta_5 * enrolment + e_t}$

The OLS equation in this model estimation is used for comparison of estimation without consideration of Instrument variables. The OLS equation of the Two-Stage Least Squares model is distinguished from the previously mentioned OLS equation without consideration of the endogeneity problem by not including ***R&D expenditures*** and ***Patent numbers*** variables. ***Innovation production*** is the only innovation-related variable in the equation. Based on the theory of Instrumental Variables (Katchova, 2013), in the case of endogeneity and endogenous explanatory variable, regression of ***Innovation production*** by instrument variables is expected to change the estimated relationship between an endogenous explanatory variable and dependent variable significantly. The significant difference in the estimated coefficients also supports the presence of an endogeneity problem.

- 2SLS, first-stage: $\mathbf{Innovation = \gamma_0 + \gamma_1 * eduexp + \gamma_2 * RDexp + \gamma_3 * patents + \gamma_4 * capform + \gamma_5 * unempl + \gamma_6 * enrolment + e_t}$

The first stage of Two-Stage Least Squares is aimed at predicting the values of an endogenous explanatory variable by regressing it on instrumental variables. For this model, the ***Innovation production*** variable is regressed by all exogenous variables, including instrumental variables that were not regressed for the previous OLS model (***R&D expenditures*** and ***Patent numbers***). The actual estimation process is very similar to the Ordinary Least Squares method but with specifying necessary exogenous and instrumental variables. Based on this stage predicted values of ***Innovation production*** are saved.

- 2SLS, second-stage: **GDP= β_0 + β_1* innovation_hat + β_2 *eduexp + β_3* capform + β_4*unempl + β_5*enrolment + e_t**

During the second stage of the Two-Stage Least Squares model, the main focus of the study, the relationship of economic growth, human capital and innovation is estimated. The selected list of variables is similar to the previously stated OLS regression equation, the main difference is that values of the endogenous explanatory variable – ***Innovation production*** are replaced by predicted values from the first stage. This replacement assumes that the impact of the other two innovation-related explanatory variables is expressed by the values of the ***Innovation production*** variable.

To proceed with the interpretation of the model‘s results, it is necessary to make sure that equations are identified. The purpose of this evaluation is to make sure that the number of instrumental variables that have been omitted from the final equation is at least in the same amount as endogenous explanatory variables. There are 3 possible options of identification such as just-identified, over-identified and under-identified model. The main concern is the under-identified model, where instrumental variables are less than endogenous variables and this leads to the situation of an indefinite amount of possible solutions that cannot provide precise estimates.

The identification formula applied for equations is the following:

$$\mathbf{k^{**} \geq g^{*} - 1} \quad (5)$$

where,

- **k**** - predetermined variables in other equations. k** can also be expressed as a difference between all independent variables of the model and predetermined variables in the evaluated equation (k-k*)
- g* - endogenous variables in the evaluated equation.

The evaluation of both variables demonstrates:

- 1st equation: **Innovation** = $\gamma_0 + \gamma_1$* **eduexp** + γ_2 ***RDexp** + γ_3* **patents** + γ_4 * **capform** + γ_5***unempl** + γ_6***enrolment** + e_t

k** = k-k* = 7 – 6 = 1

g*= 1(only ***Innovation production*** variable)

k**(1) > g*(1) - 1, which means that the 1st equation of Two-stage least squares model is over-identified.

- 2nd equation: **GDP**= $\beta_0 + \beta_1$* **innovation_hat** + β_2 ***eduexp** + β_3* **capform** + β_4***unempl** + β_5***enrolment** + e_t

k** = k-k* = 7-5 = 2

g*=2 (***GDP*** and ***Innovation production***)

k**(2) > g*(2) – 1, which means that the 2nd equation of Two-stage least squares model is over-identified.

Based on the results of 2 equations, the model is considered over-identified. In this case, the reduced form of the model does correspond to several structural forms of equation, therefore Two-stage least squares method is used as one of the efficient methods of estimating.

The model estimations were estimated by Grelt software, and the results of the initial OLS estimation part of the Two-stage least squares model are presented in Table 3.

Table 3. Model results of the OLS model (as part of the Two-stage Least squares model)

gretl: model 2

File Edit Tests Save Graphs Analysis LaTeX

Model 2: OLS, using observations 2000-2021 (T = 22)
Dependent variable: GDP

	coefficient	std. error	t-ratio	p-value	
const	−13.0549	5.27039	−2.477	0.0248	**
innovation_prod	0.502374	1.53696	0.3269	0.7480	
educ_expend	7.58561	2.11708	3.583	0.0025	***
CapForm	0.612612	0.161912	3.784	0.0016	***
Unemployment	−0.619231	0.202445	−3.059	0.0075	***
enrollment	−0.412699	0.130942	−3.152	0.0062	***

Mean dependent var	6.331818	S.D. dependent var	3.457688
Sum squared resid	32.90843	S.E. of regression	1.434147
R-squared	0.868926	Adjusted R-squared	0.827965
F(5, 16)	21.21370	P-value(F)	1.53e-06
Log-likelihood	−35.64620	Akaike criterion	83.29240
Schwarz criterion	89.83865	Hannan-Quinn	84.83450
rho	0.253418	Durbin-Watson	1.339218

Excluding the constant, p-value was highest for variable 2 (innovation_prod

Source: own proceeding based on Bureau of National Statistics of Kazakhstan

Generally, an applied method for this part of the Two-stage least squares method is very similar to the OLS model with an assumption of exogeneity. The only difference is connected with selected independent variables for regression, the currently evaluated OLS model does not contain ***R&D expenditures*** and ***Patent numbers*** variables due to their classification as Instrumental variables in the Two-stage least squares model (Table 3). The construction OLS model aims to compare its results with a final estimation of the Two-stage least squares, which will be presented as estimates of 2^{nd} equation. The only difference between these 2 equations estimates is the values of ***Innovation production*** variables. The OLS part considers actual values of the variable, while the final Two-stage least squares part considers predicted values based on the estimation of the Instrumental variables.

To proceed with an evaluation of the obtained coefficients of the initial OLS model, the overall quality and significance must be evaluated. The model significance evaluation shows that the p-value of the F-test is extremely small ($1.53E^{-6}$), which provides evidence for null hypothesis rejection regarding the model insignificance and supports the alternative hypothesis that at least one independent variable has a significant relationship with the studied dependent variable. The next evaluated indicator is the R-squared of the model, which demonstrates the result of 0.8689, which means that almost 87% of the dependent variable's variance can be explained by the difference in the explanatory variables. The model-related results are very similar to the results of the previously estimated OLS model that included Instrumental variables as exogenous explanatory variables.

The estimated equation of the model is the following:

GDP = -13.05 + 0.5* innovation + 7.58 *eduexp + 0.61* capform - 0.62*unempl – 0.41*enrolment

Almost all explanatory variables and constants, except the ***Innovation production*** variable demonstrate statistical significance, where all statistically significant variables (excluding the constant) are significant on a 1% significance level, which is relatively better than the general OLS model, the highest significance level among statistically significant variables was only 5%.

The constant variable is negative and statistically significant on a 5% significance level, and its coefficient is -13.05. It means that in the case of zero values of all other explanatory variables, the ***GDP growth*** is expected to be -13.05 percentage points. As was already mentioned before, the negative sign of the constant is not against the theoretical background due to the almost impossibility of the case in the real world, where all included explanatory variables are zero, in the real observations the impact of other variables cannot be isolated.

The only statistically insignificant variable is ***Innovation production***, which means that the exclusion of other innovation-related variables did not solve the issue of the variable's insignificance. The estimated coefficient is 0.5024 but the main improvement of the previously estimated coefficient in the exogenous OLS model is that the coefficient is positive, which makes it consistent with economic theory regarding the positive relationship between innovation and economic performance. It is assumed that the ***Innovation production*** variable is an endogenous variable and that there is an omitted impact of innovation-related variables on the endogenous variable. The assumptions of endogeneity and corruption of the estimates due to it will be tested by estimation of the 2^{nd} equation of the Two-stage least squares and following comparison with this OLS estimates.

The ***Education expenditure*** variable shows a positive coefficient of 7.58 and is statistically significant at a 1% level, the result is in line with economic theory regarding the positive impact of public education expenditures on economic growth through the higher rates of return on education and also does not sufficiently vary from the previously estimated OLS model, where the coefficient was 6.39. The exclusion of R&D and patents-related variables increased the estimated coefficient and decreased the p-value, which demonstrates an even stronger relationship between education expenditure and economic growth. The significance of the ***Education variable*** supports the idea about the significance of education in the economy and respectively significance of human capital. Based on this model, a 1 percentage point increase in ***Education expenditure*** is associated with a 7.58 percentage point increase in ***GDP growth*** rate, ceteris paribus.

The socio-economic variables (***Capital formation*** and ***Unemployment rate***) show stable statistical significance at a 1% level and provide evidence for their existing impact on economic performance. The ***Capital formation's*** estimated coefficient is 0.61 and it is positive, which is associated with the expected positive impact of investment level on economic performance and the estimated impact is even higher than in the previous OLS model. So, a 1 percentage point an increase in ***Capital formation*** rate is associated with increase in ***GDP growth*** rate by 0.61 percentage points, ceteris paribus.

The ***Unemployment rate*** is considered as a statistically significant variable even at a lower significance level (1% significance level in comparison with a 5% significance level in the previously estimated model), the estimated coefficient is -0.62. The negative relationship between the economic growth indicator and the unemployment rate is expected due to lower productivity in case of higher rates of unemployed population. As a result, a 1 percentage point increase in the ***Unemployment rate*** is associated with a 0.62 percentage point decrease in the ***GDP growth*** rate.

Another sufficient improvement in the exclusion of Instrumental variables is related to the tertiary enrolment variable, which is showing statistical significance at 1% level compared to the previous model but the estimated coefficient is still negative (-0.41). As mentioned before, the estimated result is contradictory to an economic theory about the positive relationship between a higher amount of enrolment into tertiary education and the quality of human capital, respectively with the economic growth of the country. It is still assumed that the impact of enrolment can be incorrectly estimated due to its lagged impact on economic indicators.

The next estimated model is based on the 1st stage of the Two-stage least squares model and is presented in Table 4 below:

Table 4. Results of the 1st stage of the Two-stage Least squares model

```
Model 2: OLS, using observations 2000-2021 (T = 22)
Dependent variable: innovation_prod

                    coefficient    std. error    t-ratio   p-value
  ----------------------------------------------------------------
  const             -1.65998       0.673405      -2.465    0.0263   **
  educ_expend        0.0420123     0.248962       0.1687   0.8682
  RD_expend          4.25205       1.37353        3.096    0.0074   ***
  patents            0.000492122   0.000300386    1.638    0.1222
  CapForm           -0.0319594     0.0216470     -1.476    0.1605
  Unemployment      -0.0343431     0.0226050     -1.519    0.1495
  Schoolenrollment   0.0373757     0.0165576      2.257    0.0393   **

Mean dependent var   1.217727   S.D. dependent var   0.348766
Sum squared resid    0.370145   S.E. of regression   0.157087
R-squared            0.855094   Adjusted R-squared   0.797132
F(6, 15)             14.75261   P-value(F)           0.000016
Log-likelihood       13.71729   Akaike criterion    -13.43458
Schwarz criterion   -5.797281   Hannan-Quinn        -11.63546
rho                  0.170483   Durbin-Watson        1.619042

Excluding the constant, p-value was highest for variable 3 (educ_expend)
```

Source: own proceeding based on Bureau of National Statistics of Kazakhstan

The first stage of the model is aimed to calculate the predicted values of endogenous variables by the usage of Instrumental variables. The endogenous variable is the ***Innovation variable*** and the instrumental variables are ***R&D expenditures*** and ***Patents.*** The 1st stage equation also includes exogenous explanatory variables such as ***Education expenditure, Capital Formation, Unemployment and Tertiary education enrolment.*** The regression estimates the impact of selected variables on Innovation production level.

The initial step of estimate evaluation is the overall significance of the model. The model significance evaluation shows that the p-value of the F-test is 0.000016, which provides the argument for null hypothesis rejection, in other words the hypothesis about the insignificance of any variables in the model, therefore the alternative hypothesis about the statistical significance of at least one independent variable is accepted. Another important result is R-squared, which shows 0.855 or 85.5% of ***Innovation production*** outcomes can be predicted by the model.

The estimated equation of 1st stage of the Two-stage least squares model is the following:

Innovation = -1.66 + 0.042* eduexp + 4.25 *RDexp + 0.0005* patents – 0.032 * capform – 0.034*unempl + 0.037*enrolment

When it comes to the statistical significance of estimated coefficients, only 2 independent variables and constant demonstrate significance on 1% and 5% significance levels.

The constant variable is negative, and its coefficient is -1.66. The result suggests that zero values of all other explanatory variables predict that innovation production is expected to be -1.66 percent of total annual GDP. The negative sign of the constant is not in line with the expectation since the production level of any sphere cannot be negative. However, the absolute value of the constant is relatively small which be related to the impossible case of zero impact of other variables in the equation.

The focus of the coefficient evaluation is concentrated on 2 Instrumental variables - ***R&D expenditures*** and ***Patents.*** These variables are regressed to ***Innovation production*** to predict its values based on the assumption of the interdependence of these variables and the endogeneity issue of ***Innovation production*** can be solved by the application of these Instrumental variables. Regression demonstrates the statistical significance of ***R&D expenditures*** at a 1% significance level in predicting the values of ***Innovation production*** and the statistical insignificance of the second instrumental variable – ***Patents.*** The estimated coefficient of ***R&D expenditures*** is 4.25, while the p-value is quite low – 0.0074. This result concludes the significance ***of R&D expenditure*** in defining the effectiveness of innovation production and allows the consideration of ***R&D investment*** as the input variable of innovation level in the country, while ***Innovation production***'s share in total production can be considered as the output variable of innovation level in the country. Therefore, a 1 percentage point increase in ***R&D expenditure*** is associated with a 4.25 percentage points increase in ***Innovation production*** out of total GDP, ceteris paribus.

The second Instrumental variable's (***Patents***) coefficient is 0.0004. The estimated coefficient is positive and in line with economic theory that assumes the positive impact of patent registration on the development of innovations and as a result on innovation production. However, the p-value of the estimate is 0.12, which makes it statistically insignificant concerning innovation production estimation. The insignificance of the variable does not allow us to interpret the relationship between innovation and patent registration. The possible reason for insignificance can be related to the existence of omitted variables that define the annual number of patents such as the legislative system, the protection of intellectual property and the financial motivation behind patent registration. This result provides a foundation for another analysis of the dependence of patents on institutional norms in the country.

The exogenous explanatory variables are also included in the 1st stage of the Two-stage least square model and their relationship with ***Innovation production*** can be evaluated. The education expenditure variable shows a positive coefficient of 0.042 but this estimated coefficient is statistically insignificant, concluding that there is no clear significant relationship between ***Innovation production*** and ***Education expenditures***. Based on the economic theory higher education levels should positively influence all innovational and development spheres of the country. However, considering the small sample size for time series and expected delayed positive impact of education investments on the innovational sphere, the selected dataset might be not able to detect the expected impact. Therefore, the insignificant relationship between these 2 variables falls into the possible results options and might be further studied based on other time series samples with wider time ranges.

Surprisingly, another education-related variable – ***Tertiary education enrolment*** demonstrates a statistically significant (at a 5% significance level) and a positive coefficient. The coefficient is 0.037, which suggests that a higher rate of enrolment in higher education is associated with higher volumes of innovation production. ***Tertiary education enrolment*** can be considered as a variable that identifies the current level of education quality, therefore its impact on innovation development is less lagged and can explain the interconnection between education and innovation. Therefore, a 1 percentage point increase in ***Tertiary education enrolment*** is associated with a 0.037 percentage points increase in ***Innovation production*** out of total GDP, ceteris paribus.

The socio-economic variables such as ***Capital formation*** and ***Unemployment*** show statistical insignificance in relation to innovation production. Both variables have negative coefficient and high p-values that makes the explanation of their relationship with innovation production inconsistent and erroneous.

The result of 1st stage provides evidence of the interdependence of ***Innovation*** and ***R&D*** variables and calculates the predicted values of ***Innovation production*** that are required for the 2nd stage of the Two-stage Least squares model.

The 2nd stage of the Two-stage least squares model is estimated and presented in Table 5 below:

Table 5. Results of the 2nd stage of the Two-stage Least squares model

```
Model 5: TSLS, using observations 2000-2021 (T = 22)
Dependent variable: GDP
Instrumented: innovation_prod
Instruments: const educ_expend RD_expend patents Unemployment Enrollment
  CapForm

                    coefficient   std. error   t-ratio   p-value
  ---------------------------------------------------------------
  const              -9.02255      6.19287     -1.457    0.1645
  innovation_prod     6.88911      3.12638      2.204    0.0416   **
  educ_expend         6.32618      2.38998      2.647    0.0176   **
  Unemployment       -0.693722     0.219361    -3.162    0.0060   ***
  Enrollment          0.441038     0.139405     3.164    0.0060   ***
  CapForm             0.556581     0.174903     3.182    0.0058   ***

Mean dependent var   6.331818   S.D. dependent var   3.457688
Sum squared resid    36.59191   S.E. of regression   1.512281
R-squared            0.854629   Adjusted R-squared   0.809200
F(5, 16)             19.34573   P-value(F)           2.84e-06
rho                  0.296873   Durbin-Watson        1.282406
```

Source: own proceeding based on Bureau of National Statistics of Kazakhstan

This estimated equation is aimed to study the relationship between variables of our interest (human capital, innovation and economic growth) with the assumption of endogeneity of some explanatory variables. The main difference between this final model stage and the initially estimated relationship in the form of OLS is that the initial observations of the ***Innovation production*** variable that were gathered from the real-world population are replaced by the predicted values based on the previously explained 1^{st} stage of the model.

The first evaluation contains the evaluation of the overall significance of the model and its comparison with the initial OLS model. The model significance evaluation shows that the p-value of the F-test is $2.84E^{-6}$, which is an extremely low value and allows to confidently reject the null hypothesis of model insignificance and accept the alternative hypothesis that at least one independent variable has a significant relationship with the studied dependent variable (Table 6). The p-value of the initial OLS model without endogeneity consideration was already very low, so no drastic difference in this characteristic The R-squared of the model is 0.8546, which is a relatively high result and demonstrates that almost 85% of outcomes can be predicted by the model, the R-squared of the initial OLS model with the similar list of independent variables was 86%, which also confirms almost same results of overall model significance between the Ordinary Least Squares and Two-Stage Least Squares models.

The estimated equation of the 2^{nd} stage of the Two-Stage Least Squares model is the following:

GDP = -9.02 + 6.89 * innovation + 6.33 *eduexp + 0.56* capform - 0.69*unempl + 0.44*enrolment

All explanatory variables (excluding the constant) demonstrate statistical significance, where 3 of them are significant on a 1% significance level and 2 variables are significant on a 5% significance level. The main improvement of the Two-Stage Least Squares model compared to the Ordinary Least Squares model is that the variable with assumed endogeneity issue – ***Innovation production*** is finally showing statistical significance on a 5% level.

The constant variable is negative, its coefficient is -9.02 but, in this model, it shows statistical insignificance. As mentioned before, the negative sign of the constant is not against the economic theory due to the impossibility of case of zero public expenditure, zero unemployment and zero capital formation or in other words isolated impact of the constant. The statistical insignificance of the constant can be interpreted that in the theoretical case of zero of explanatory variables constant won't significantly differ from zero.

The improved variable (***Innovation production***) shows a statistically significant result, which points to the improvement of the endogeneity issue and its correction by the replacement of Instrumental variables' predicted values. So, the final estimated coefficient of ***Innovation production*** is 6.89 and it is positive. The positive coefficient is in line with the theoretical background about the positive impact of innovational breakthroughs on performance in various spheres of the economy and reached statistical significance in interpreting the innovation variable and economic growth relationship possible. Based on the estimated coefficient it can be concluded that a 1 percentage point increase in ***Innovation production*** is associated with a 6.89 percentage points increase in ***GDP growth*** rate, ceteris paribus.

The ***Education expenditure*** variable keeps showing a positive coefficient regarding its impact on economic growth, so the final coefficient is 6.33 and is statistically significant, the result is in line with economic theory regarding the positive impact of public education expenditures on economic growth through the higher rates of return on education. The ***Innovation variable*** was considered as the main source of endogeneity and has shown several possible relationship estimations, while the education-related variable showed statistical significance over all estimated types of models, therefore sufficient relationship with economic growth can be concluded. Based on the results of the Two-Stage Least Squares model, a 1 percentage point increase in ***Education expenditure*** is associated with a 6.33 percentage point increase in ***GDP growth*** rate, ceteris paribus.

The socio-economic variables such as ***Capital formation*** and ***Unemployment rate*** show statistical significance even on a 1% significance level and prove that their impact on the economy is proved by the Two-Stage Least Squares model as it was already seen in the initial Ordinary Least squares model with the same list of explanatory variables. The ***Capital formation***'s estimated coefficient is 0.56 and it is positive, which is associated with an expected positive impact of investment level on economic growth. So, a 1 percentage point increase in ***Capital formation*** rate is associated with an increase in ***GDP growth*** rate by 0.56 percentage points, ceteris paribus.

The ***Unemployment rate*** is considered as a statistically significant variable, the estimated coefficient is -0.69. The negative relationship between the economic growth indicator and the unemployment rate is in line with economic theory, which suggests that higher rates of unemployed population lead to lower productivity of the economy and as a result lower rates of economic growth. The final estimated coefficient demonstrates that a 1 percentage point increase in the ***Unemployment rate*** is associated with a 0.69 percentage point decrease in ***GDP growth*** rate, ceteris paribus.

The ***Tertiary education enrolment*** variable, which is considered another important variable expressing the quality of education in the country, is statistically significant and has a positive estimated coefficient (0.44). Based on the economic theory higher amount of enrolment into tertiary education has to provide a sufficient positive impact on the quality of human capital and as a result, on the economic performance of the country. The previously estimated model with assumed endogeneity demonstrated negative and statistically insignificant coefficients that were concluded as a result of the small sample size and lagged impact of the studied variable on the dependent variable. Fortunately, the endogeneity elimination attempt led to obtaining of statistically significant and positive coefficient of the ***Enrolment*** variable. The estimated the ***Enrolment*** coefficient of the Two-Stage Least Squares model demonstrates that a 1 percentage point increase in ***Tertiary education enrolment*** is associated with a 0.44 percentage points increase in ***GDP growth*** rate, ceteris paribus.

Table 6. Comparison of the OLS and 2SLS models:

	OLS model for ***GDP***	2SLM: 1st stage for ***Innovation production***	2'SLM: 2nd stage for ***GDP***
Constant	-13.055**	-1.659**	-9.022
Innovation_prod	0.502	-	6.889**
Educ_expend	7.586***	0.042	6.326**
RD_expend	-	4.252***	-
Patents	-	0.0004	-
CapForm	0.613***	-0.032	0.556***
Unemployment	-0.619***	0.037	-0.694***
Enrolment	-0.413***	-0.034**	0.441***

Source: own proceeding based on Bureau of National Statistics of Kazakhstan

The main difference between the OLS and 2SLS is related to the endogenous explanatory variables (***Innovation production***), which initially demonstrated a statistically insignificant relationship with the ***GDP growth***, started to show a statistically significant coefficient concluding that a 1 percentage point increase in ***Innovation production*** is associated with a 6.889 percentage points increase in ***GDP growth*** rate. Even considering the estimated coefficient, without its statistical significance, the coefficient was underestimated with only a 0.502 percentage points increase compared to 6.889 percentage points of 2SLS estimates.

The rest explanatory variables show stable statistical significance in both models. The ***Education expenditure*** variable's impact is lower in the case of the 2SLS model, where a 1 percentage increase is associated with a 6.326 percentage points increase in ***GDP growth***, compared to a 7.586 percentage points increase in the case of the OLS model.

The impact of ***Capital Formation*** is also lowered in the 2SLS sample, while ***Unemployment***'s impact on the contrary is estimated higher in the results of the Two-Stage Least Squares model. The surprising impact of the replacement of the endogenous variable's observations by predicted values of 1st stage can be seen in the estimated coefficient of the ***Tertiary education enrolment variable***. In the results of the 2SLS model, the variable shows a positive correlation with the ***GDP growth***, compared to the initial results of the OLS model, which makes the 2SLS estimated coefficient in line with the economic theory.

5.9. Endogeneity test

The aim of the Two-Stage Least Squares model was the assumption of endogeneity issue in the initial OLS model related to the omitted impact of ***R&D expenditures*** and ***Patents*** on the values of ***Innovation production***. To test the correctness of the assumption and the efficiency of relationship estimation by the Two-Stage Least Squares model the existence of endogeneity must be tested.

There are several ways of endogeneity testing, The first one is the Hausman test for endogeneity, which tests the difference between the OLS and 2SLS estimates and tests if they are significant. If significant differences are detected, then the model includes the case of endogeneity.

H_0: *There is no significant difference between OLS and 2SLS models.*

H_1: *There is a significant difference between OLS and 2SLS models.*

The result of the Hausman test is presented in Table 7.

Table 7. Hausman test result.

```
Hausman test -
  Null hypothesis: OLS estimates are consistent
  Asymptotic test statistic: Chi-square(1) = 3.92387
  with p-value = 0.0476051
```

Source: own proceeding based on Bureau of National Statistics of Kazakhstan

The p-value of the test is 0.0476 (Table 7), which is lower than 0.05 and which suggests that the null hypothesis regarding the insignificant differences between OLS and 2SLS models can be rejected, and the alternative hypothesis can be accepted at a 5% significance level. This concludes that there is a proven case of endogeneity and the Two-Stage Least Squares model is required to obtain efficient and statistically significant estimates of the relationship between human capital, innovation and economic growth.

Another way of testing the endogeneity existence is the Durbin-Wu-Hausman test for endogenous regressors. This test aims to identify if endogenous explanatory variables in the model are truly endogenous, and their estimation requires a more sophisticated model than the Ordinary Least Squares model.

To test the model for endogeneity, after the first stage of the Two-Stage Least Squares model, the residuals of this regression are saved and included in the final stage regression as a separate variable. The estimated coefficient of the residual is evaluated, and it gives information about the endogeneity in the final regression model.

In the actual case of the endogenous explanatory variable, the estimated coefficient of the residuals has not to be equal to zero and show statistical significance. The 1st stage residuals are saved as ***uhat11*** and its estimated coefficient is presented in Table 8.

Table 8. Residual coefficient estimates for the Durbin-Wu-Hausman test.

Model 14: OLS, using observations 2000-2021 (T = 22)
Dependent variable: GDP

	coefficient	std. error	t-ratio	p-value	
const	-9.02255	5.58765	-1.615	0.1272	
educ_expend	6.32618	2.15641	2.934	0.0103	**
Unemployment	0.693722	0.197923	3.505	0.0032	***
Enrollment	-0.441038	0.125781	-3.506	0.0032	***
CapForm	0.556581	0.157810	3.527	0.0031	***
innovation_prod	2.55921	1.92864	1.327	0.2044	
uhat11	4.25205	1.37353	3.096	0.0074	***

Mean dependent var	6.331818	S.D. dependent var	3.457688
Sum squared resid	27.92737	S.E. of regression	1.364487
R-squared	0.888766	Adjusted R-squared	0.844272
F(6, 15)	19.97506	P-value(F)	2.31e-06
Log-likelihood	-33.84086	Akaike criterion	81.68172
Schwarz criterion	89.31902	Hannan-Quinn	83.48084
rho	0.304912	Durbin-Watson	1.296180

Excluding the constant, p-value was highest for variable 2 (innovation_prod

Source: own proceeding based on Bureau of National Statistics of Kazakhstan

The estimated coefficient of residuals shows a coefficient of 4.25 (Table 8), which is different from zero and statistically significant at a 1% level. This result corresponds to the assumption of endogeneity of the estimated variable in the 1st stage of the Two-stage Least Squares model, therefore ***Innovation production*** variable is an endogenous explanatory variable as it was assumed in the beginning. Moreover, it confirms that the Ordinary Least Squares model cannot be used for precise estimation of the studied relationship and the Two-Stage Least Squares model regression is a more accurate estimation.

5.10. Weak Instrumental variables

The significance and relevance of the selected Instrumental variables can be also tested by evaluating the weaknesses of the instruments. In the case of a just-identified model with one instrumental variable corresponding to 1 endogenous variable, their correlation and its significance works as an indicator of instrumental variable significance. In the case of more than one instrumental variable for one endogenous variable, the partial F-statistic of the 1st stage of the Two-stage least squares model is used to test the joint significance of instrumental variables and the result less than 10 points to weak instrumental variables. The Gretl application automatically provides the results of the Weak instrumental variables test as a part of the Two-Stage Least squares model estimation and the results for a previously estimated model are presented in Table N.

Table 9. Weak Instrumental variables test result.

```
Weak instrument test -
  First-stage F-statistic (2, 15) = 10.1421
  Critical values for desired TSLS maximal size, when running
  tests at a nominal 5% significance level:

     size      10%      15%      20%      25%
    value    19.93    11.59     8.75     7.25

  Maximal size may exceed 15%
```

Source: own proceeding based on Bureau of National Statistics of Kazakhstan

The estimated result is 10.1421 (Table 9), which is higher than 10 and it concludes that even in the case of statistical insignificance of the ***Patents*** variable regarding the regression of ***Innovation production***, the selected pair of instrumental variables (***R&D expenditure*** and

Patents) demonstrate joint significance and reject the existence of weak instrumental variables problem.

5.11. Model limitations and possible improvements

Despite considerably efficient estimates of the final Two-Stage Least Squares model and their relevance in explaining the relationship between human capital, innovation and economic growth, the model has several weak points and limitations.

The first main limitation is related to sample size. The source of the data is the Bureau of National Statistics of the Republic of Kazakhstan and the data for the majority of the variables are available only from 2000, therefore it limits the possible time series to the range of two decades only. Another limitation of the data sources is the unavailability of some variables data for Kazakhstan, which has been used in similar studies of the relationship between human capital and economic growth. A possible improvement is exploring other variables that are used in unexplored research related to human capital, innovation and economic growth relationships.

Another important limitation is that the GDP growth rate is selected as the most precise available measure of economic performance, it is important to mention that several variables can be used as well. The GDP growth rate does not express the inequality rate, sustainability of production, or contribution to the welfare of the population and especially in the case of high inequality, it cannot fully demonstrate the quality of life in the country. The possible improvement is similar to model estimation with the usage of alternative indicators of economic performance such as the Human development index and Average income per capita. The omitted quality of life can be controlled by supplementary analysis of socio-economic indicators such as the Gini index, poverty rate and environmental sustainability.

In the case of Instrumental variables, the ***Patents*** coefficient demonstrated statistical insignificance in the regression of ***Innovation production***. The possible reason behind the insignificance issue can be related to omitted variables. The patent numbers are not only related to the number of innovations but also hardly dependent on the intellectual property rights and legislative system in the country. These factors were not considered in the model, which might be a reason for instrumental variable insignificance. The available solution is the addition of an explanatory variable that will estimate the quality of the legislative system towards the protection of intellectual property rights.

Last but not least, the possible improvement is related to the sophistication model with the addition of another endogenous explanatory variable related to education. In this model, education expenditure and tertiary education enrolment were included as exogenous explanatory variables. However, it is possible to estimate of quality of education level with the help of Instrumental variables that will give a deeper understanding of the relationship between education level and economic performance.

6. Discussion

Becker (1964) emphasized that human capital contributes to increased productivity and enhances the development of innovative technologies. Romer (1990) and Azariadis and Drazen (1990) proved that the ability of developing countries to converge with technological progress and increase productivity depends on the quality of human capital.

Currently, the importance of human capital for the economic development of Kazakhstan is recognized, however, the empirical studies on this topic is limited. As for developing countries after the transition period from a centrally planned to a market economy, FDI is a significant factor in economic growth, and its efficient use tends to be dependent on the quality of human capital. Despite this significant interdependence of physical and human capital utilization, Kazakhstan's expenditure on science and research remains relatively low compared to developed countries. It highlights the need for an increase in public expenditures on education, science, and healthcare to support innovation and promote economic growth.

Referring to research on human capital and economic growth in North Macedonia (Ziberi et al., 2022), besides the OLS regression of the relationship between human capital, innovation and economic growth, the Two-Stage Least Squares model was estimated, with considering ***innovation productio***n as endogenous explanatory variables and ***R&D expenditures*** and ***patents*** as instrumental variables. The model demonstrates the significance of R&D investment and innovation production relationship and the non-significance of patents regarding the innovation production level.

Aghion and Howitt (1998), provided evidence that an increase in R&D investment is significant for keeping the innovation level increasing or at least constant. Their analysis demonstrated the positive relationship between R&D investment and total factor productivity in the case of the United States, which corresponds to the finding of this thesis regarding the positive impact of R&D investment on innovation and on economic growth through innovation production. The result of patent insignificance corresponds to the mentioned by Gould and Gruben (1996) complexity of intellectual property rights regulations and resulting no strong correlation between patents and economic growth.

In the 2SLS model, where the endogeneity issue of ***innovation production*** is considered, the significance of this variable in explaining GDP growth is proved. The estimated significant relationship between innovation and economic growth in the case of Kazakhstan corresponds to the results of Geissdoerfer et al. (2018). He studied that innovation plays a crucial role in supporting economic growth by promoting technological development, and enhancing productivity and also serves as a catalyst for the development of new products, processes, and services. There is a clear positive correlation between innovation-related expenditures and economic growth, highlighting the importance of policies and frameworks to encourage innovation. Moreover, innovations require a skilled workforce with the ability to implement new technologies efficiently, emphasizing the importance of human capital in driving innovation and economic growth. This corresponds to the result of the joint significance of education and innovation variables in the final Two-Stage Least Squares model, emphasizing the interconnection and joint impact of improvements in education and innovation level on the economic performance of the country.

Considering the positive and significant coefficient of education expenditure in both models, the findings of Benhabib and Spiegel (2005) regarding the negative impact of a lower share of the population with tertiary education on the pace of innovation integration in the economy, therefore stagnate the innovation-related growth. They clarified that innovation adaptation is a complex process that requires a highly qualified labor force; therefore, the estimated positive impact of education expenditure and tertiary education enrolment is expected.

Referring to the significance of the education variable, the government has to prioritize investment in education, with a special focus on secondary and tertiary education, which are a more neglected part of education reforms in Kazakhstan. Policymakers should increase governmental and create incentives for private R&D spending, in the form of tax credits or governmental grants. Moreover, innovation can be supported not only but raising the funds but also by eliminating unnecessary bureaucratic barriers and strengthening intellectual property rights protection.

7. Conclusion

The thesis aims to study and estimate the relationship between human capital, innovation and economic growth in Kazakhstan. Human capital and innovation are significant pillars of economic development; therefore, their research can benefit developing countries' policy framework construction.

The empirical analysis aimed to estimate and study the relationship between human capital, innovation and economic growth based on the dataset of Kazakhstan for the period of 2000-2021. The analysis was done using the estimation of two regression models – The ordinary Least Squares and Two-Stage Least Squares models.

The OLS model was estimated without considering the endogeneity problem of explanatory variables. Overall, the model demonstrated a strong explanatory power, and ***education expenditure,*** which was selected as a proxy of human capital, showed a strong and positive relationship with GDP growth. On the other hand, the innovation-related variables showed an insignificant relationship with GDP growth, assuming the endogeneity issue of the ***innovation production*** variable.

The Two-Stage Least Squares model was estimated, considering innovation production as endogenous explanatory variables and R&D expenditures and patents as instrumental variables. The estimated model proves the strong positive relationship between innovation production and R&D investment, while patents did not significantly affect innovation. In the final model, the endogeneity issue of innovation production is considered, demonstrating a statistically significant and positive coefficient, suggesting its significance in explaining GDP growth. Human capital-related (education expenditure, tertiary education enrolment) and socio-economic (capital formation, unemployment rate) variables demonstrate statistical significance toward economic growth.

The endogeneity tests, such as the Hausmann and Durbin-Wu-Hausman tests, were conducted, proving the endogeneity issue; therefore, the Two-Stage least squares model is selected as a best-fit model, and its results correspond to the initial hypotheses about a positive significant relationship between human capital, innovation and economic growth in the case of Kazakhstan.

The findings of this empirical study can be used as a basis for further analysis and policy suggestions, which can contribute to an increase of Kazakhstan's human capital potential, create an incentive for innovation development and encourage sustainable economic growth.

References

Abdykereyeva, G. (2012). Innovation activity in Kazakhstan at the present stage and prospects for its development. *Regional Bulletin of the East, 3.* 73-80.

Aghion, P., & Howitt, P. (1998). Endogenous growth theory. Cambridge, Massachusetts: The MIT Press.

Ang, J. B., Madsen, J. B., & Rabiul Islam, M. (2011). The effects of human capital composition on technological convergence. *Journal Macroeconomics, 33*(3), 465–476.

Azam, M., & Ahmed, A. M. (2015). Role of Human Capital and Foreign Direct Investment in Promoting Economic Growth: Evidence from Commonwealth of Independent States. *International Journal of Social Economics, 42*(2), 98–111.

Azariadis, C., & Drazen, A. (1990). Threshold Externalities in Economic Development. *The Quarterly Journal of Economics, 105*(2), 501–526.

Baldacci, E., Clements, B., Gupta, S., & Qiang, C. (2004). Social Spending, Human Capital, and Growth in Developing Countries: Implications for Achieving the MDGs. *Washington, DC: IMF.*

Barro, R. (1991). Economic growth in a cross section of countries. *Quarterly Journal of Economics, 56.*

Barro, R. and Lee, J., (1994). Sources of Economic Growth, *Carnegie-Rochester Conference Series on Public Policy, 40*, 1–46

Becker, G. S. (1962). Investment in Human Capital: A Theoretical Analysis. *Journal of Political Economy, 70*(5), 9–49.

Benhabib, J., & Spiegel, M. (2005). Human Capital and Technology Diffusion. *Handbook of Economic Growth, 1*, 935–966.

Bloom, D., & Canning, D. (2003). Health as Human Capital and its Impact on Economic Performance. *The Geneva Papers on Risk and Insurance. Issues and Practice, 28*(2), 304–315.

Bogatyreva, V. (2018). Human capital is a factor in increasing investment activity and economic growth. *Economic Bulletin of Donbass, 1*(51), 124–136.

Borensztein, E., Gregorio, J. and Lee, J.-W. (1998). How does foreign direct investment affect economic growth?. *Journal of International Economics, 45*(1), 115-135.

Bulasheva, A. (2019). Impact assessment of investments in education on the development of human capital and its influence on economic growth. B*ulletin of Karaganda University. Economy Series, 93*(1), 41–48.

Dauda, R. (2010). Role of Human Capital in Economic Development: An Empirical Study of Nigerian Case. *Oxford Business and Economics Conference Program.*

Deakins, D., & Whittam, G. (2000). Business start-up: theory, practice and policy. *Enterprise and Small Business: Principles, Practice and Policy*, 115–131.

Del Canto, J. D., & Gonzalez, I. S. (1999). A resource-based analysis of the factors determining a firm's R&D activities. *Research Policy, 28*(8), 891–905.

Doing Business 2016. *The World Bank.* Retrieved from Doing Business.

Duflo, E. (2001). Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment. *American Economic Review, 91*(4), 795–813.

Fox, J. (2016). Applied regression analysis and generalized linear models (3rd ed.). *Thousand Oaks, CA: Sage publications.*

Geissdoerfer, M., Vladimirova, D., & Evans, S. (2018). Sustainable business model innovation: A review. *Journal of Cleaner Production, 198*, 401–416.

Gerguri, S., & Ramadani, V. (2010). The Impact of Innovation into the Economic Growth. University Library of Munich, Germany, MPRA Paper.

Gould, D. M., & Gruben, W. C. (1996). The role of intellectual property rights in economic growth. *Journal of Development Economics, 48*, 323–350.

Grant, C. (2017). The Contribution of Education to Economic Growth. *Brighton: Institute of Development Studies.*

Kibayeva A., Tussupova L., Gize R. (2022). Method of human capital assessment in the digital economy of Kazakhstan. *Bulletin of "Turan" University*, *2*. 261-267.

Kobrin, S. (2005), The determinants of liberalization of FDI policy in developing countries: a cross-sectional analysis, *Transnational Corporations, 14*(1), 67-103.

Korsakova, E. (2016). The accumulation of human capital and economic growth. *Successes of modern science, 3*(12), 154–156.

Koshanov, A., & Chulanova, Z. (2021). Approaches to assessing the state of human capital (based on the example of Kazakhstan). *Bulletin of Moscow University. Economics, 4*. 49-72.

Kyriacou, G. (1991). Level and growth effects of human capital. *C. V. Starr Center Working Paper, 91*(26). New York University.

Lutz Hendricks. (2002). How Important Is Human Capital for Development? Evidence from Immigrant Earnings. *American Economic Review, 92*(1), 198–219.

Maskus, K. E. (2000). Intellectual property rights and economic development. Prepared for the series Beyond the Treaties: A Symposium on Compliance with International Intellectual Property Law, organized by Fredrick K. Cox, International Law Center at Case Western Reserve University, Cleveland, OH, 18.

Muhamedzhanova, A. (2001). Human capital of Kazakhstan: formation, condition, use. *Almaty: Gylym*, 276.

Mukhamedzhanova A., Kussayinova A. (2017). The state of human capital in the field of Science and Innovations in Republic of Kazakhstan. *Bulletin of KazNU. Economic series, 1*. 119.

Musgrave, R. (1969). The Theory of Public Finance. *New York: McGraw-Hill Book Company*.

Nelson, R. R., & Phelps, E. S. (1966). Investment in Humans, Technological Diffusion, and Economic Growth. *The American Economic Review, 56*(1/2), 69–75.

Nureyev, R. (2015). The economy of development: the market economy establishing models. *NIC INFRA-M*. 640.

OECD. (2007). Innovations and growth: Rationale for an innovation strategy. *3-29*.

Oreopoulos, P., & Salvanes, K. (2011). Priceless: The nonpecuniary benefits of schooling. *J. Econ. Perspect., 25*, 159–184.

Pece, A. M., Simona, O. E. O., & Salisteanu, F. (2015). Innovation and Economic Growth: An Empirical Analysis for CEE Countries. *Procedia Economics and Finance*, *26*, 461–467.

Porter, M., & Stern, S. (2000). Measuring the 'Ideas' Production Function:

Evidence from International Patent Output, *NBER Working Paper, 7891,*

Cambridge: Massachusetts: National Bureau of Economic Research.

Pradhan, R. P., Arvin, M. B., Nair, M., & Bennett, S. E. (2020). The dynamics among entrepreneurship, innovation, and economic growth in the Eurozone countries. *Journal of Policy Modeling, 42*(5), 1106–1122.

Prahalad, C. K., & Hamel, G. (1990). The core competence of the corporation. *Harvard Business Review, 90*(3), 79–91.

Pritchett, L. (1996). Where has all the education gone? *World Bank Policy Research Working Paper*, (1581).

Pyo, H. K. (1995). A time-series test of the endogenous growth model with human capital. In *Growth Theories in Light of the East Asian Experience* (pp. 229-245). University of Chicago Press.

Romer, P. (1990). Endogenous technical change. *Journal of Political Economy, 98*, 71–102.

Romer, P. M. (1986). Increasing Returns and Long-Run Growth. *Journal of Political Economy, 94*(5), 1002–1037.

Sarangi, A. K., Pradhan, R. P., Nath, T., Maradana, R. P., & Roy, H. (2022). How Does Innovation Affect Economic Growth? Evidence from G20 Countries. *The Indian Economic Journal, 70*(1), 8-21.

Schultz, T. W. (1961). Investment in Human Capital. *The American Economic Review, 51*(1), 1–17.

Schultz, T. P. (2002). Wage gains associated with height as a form of health human capital. *Yale Economic Growth Center Discussion Paper*, (841).

Schumpeter, J. (1934). Harvard Economic Studies *Theory of Economic Development*. Harvard Economic Studies 46.

Śledzik, K. (2013). Schumpeter's View on Innovation and Entrepreneurship. *SSRN Electronic Journal*.

Smith, A. (1776). An inquiry into the nature and causes of the wealth of nations (1976 edition). Chicago: University of Chicago Press.

Solow, R. (1956). A Contribution to the Theory of Economic Growth. *The Quarterly Journal of Economics, 70*(1), 65-94.

Statista. (2023). Leading countries by gross research and development (R&D) expenditure worldwide in 2022, Top countries by R&D expenditure 2022. Retrieved from Statista.

Taniguchi, M. (2003). Internet Metaphors Matter. *New Directions for Teaching and Learning,* 13-22.

Turganbayev, Y. (2023). The Effect of Human Capital on Economic Growth: Evidence from Kazakh Regions. *Economy of Regions, 19*(2), 385–396.

UNESCO. (2022). *Higher education global data report.* UNESCO.

Wernerfelt, B. (1984). A resource-based view of the firm. *Strategic Management Journal, 5*, 171–180.

World Bank. (2016). Doing Business 2016: Measuring Regulatory Quality and Efficiency. Washington, DC: World Bank.

World Bank. (2019). Ease of doing business rank. Retrieved from World Bank.

World Economic Forum. (2023). World competitiveness ranking. Retrivied from World Economic Forum.

Xu, B. (2000). Multinational enterprises, technology division, and host country productivity growth. *Journal of Development Economics, 62*(2), 477-493.

Yesimzhanova, S. (2018). Human capital in Kazakhstan: level of development and quality. *Human capital and professional education, 1*(25). 4-12.

Ziberi, B. F., Rexha, D., Ibraimi, X., & Avdiaj, B. (2022). Empirical Analysis of the Impact of Education on Economic Growth. *Economies*, 10(4), 89.