INSTITUTO
UNIVERSITÁRIO
DE LISBOA

How to split the coefficient of determination: an application to economic growth in the U.S.

Francisca Ferreira Neca

Mestrado em Economia Monetária e Financeira

Orientador:
Doutor José Dias Curto, Professor Catedrático,
ISCTE- Instituto Universitário de Lisboa

Co-Orientadora:
Doutora Sofia Vale, Professora Associada,
ISCTE- Instituto Universitário de Lisboa

Setembro, 2024

CIÊNCIAS SOCIAIS
E HUMANAS

Departamento Economia Política

How to split the coefficient of determination: an application to economic growth in the U.S.

Francisca Ferreira Neca

Mestrado em Economia Monetária e Financeira

Orientador:
Doutor José Dias Curto, Professor Catedrático,
ISCTE- Instituto Universitário de Lisboa

Co-Orientadora:
Doutora Sofia Vale, Professora Associada,
ISCTE- Instituto Universitário de Lisboa

Setembro, 2024

Acknowledgments

Completing my master's degree would not have been possible without all the support and trust of the people in my life. To each of you, here are my deepest and most sincere acknowledgments.

For,

My parents, Ana and Bruno, for always believing in me, investing in my future and education, and all the love they have given me throughout my life.

My grandparents, Teresa and Alfredo, for being real second parents to me.

My friends, in the city where I was born, but also to all those that life has given me during my time in Coimbra and Lisbon. My deepest thanks to each one of you.

My supervisor, who was tireless throughout this period. Tireless in his help, tireless in his support, and tireless in showing the trust he placed in me, for the challenging subject and for entrusting it to me to achieve.

My co-supervisor, who was fundamental during the process, particularly for his valuable contributions in the macroeconomic area.

Last but not least, to my best friends, in pseudonyms, Junho, Bobi, Black, Mafalda, Carolina, Amélia, and Francisco.

Resumo

A presente dissertação envolve um estudo econométrico sobre o crescimento económico dos Estados Unidos de 1970 a 2019. Foram selecionados fatores determinantes para o mesmo, nomeadamente, investimento, medido pela formação bruta em capital fixo, capital humano e a abertura comercial do país.

O objetivo do estudo pretende preencher uma lacuna na literatura económica, na quantificação do poder explicativo de cada variável independente do modelo de crescimento económico. De forma a preencher esta lacuna, foi crucial repartir o coeficiente de determinação, que por sua vez, só foi possível de alcançar através da grande contribuição da metodologia Commonality Analysis.

As principais conclusões referem-se ao facto de o investimento ser o fator, de entre os considerados, com maior poder explicativo na variância do crescimento económico. Contrariamente à teoria do crescimento endógeno, o capital humano, revelou ser uma variável com um peso bastante reduzido na explicação da variância do crescimento económico.

Adicionalmente, também foi possível analisar o impacto dos fatores previamente referidos, no crescimento económico do país, através do Método dos Mínimos Quadrados (MMQ). O impacto de os três fatores considerados no crescimento económico dos Estados Unidos é positivo, sendo apenas a formação bruta em capital fixo estatisticamente significativo.

Palavras-chave: crescimento económico, commonality analysis, coeficiente de determinação
Classificação JEL: O40; O51; C32

Abstract

This dissertation involves an econometric study of economic growth in the United States from 1970 to 2019. Determining factors were selected, namely investment, measured by gross fixed capital formation, human capital, and the country's trade openness.

The study aims to fill a gap in the economic literature by quantifying the explanatory power of each independent variable in the economic growth model. To fill this gap, it is necessary to split the coefficient of determination, which could only be achieved through the great contribution of the Commonality Analysis methodology.

The main conclusion is that investment is the factor in the model with the greatest explanatory power in the variance of economic growth. Contrary to the endogenous growth theory, human capital proved to be a variable with very little weight in explaining the variance of economic growth.

In addition, it was also possible to analyze the impact of the previously mentioned determining factors on the country's economic growth using the Ordinary Least Squares (OLS). The impact of the three considered factors on economic growth in the United States is positive, with only gross fixed capital formation being statistically significant.

Keywords: economic growth, commonality analysis, coefficient of determination

Classification JEL: O40; O51; C32

Index

Acknowledgments v

Resumo vii

Abstract ix

Introduction 1

Chapter 2. Literature Review 5

Chapter 3. Methodology 15

3.1. Theoretical overview: nonstationarity and commonality analysis 16

3.1.1. Nonstationarity 16

3.1.2. Commonality analysis 18

3.2. Data and descriptive statistics 21

Chapter 4. Empirical findings 25

Chapter 5. Conclusions 35

5.1. Limitations 35

5.2. Suggestions for further research 36

References 37

Appendix 41

Introduction

Economic growth reflects the development of a nation and is commonly measured by Gross Domestic Product, with the United States showing up at the top of the list. The U.S. economy is often considered the engine of the world economy since the "movements in U.S. economic growth appear to influence growth in other countries significantly" (Arora & Vamvakidis, 2001, p. 3).

As economies have grown more complex over time, the factors that have been investigated as drivers of economic growth have changed. The starting point was through the neoclassical theory, which emphasized the role of capital stock, labor, and technological progress as key variables in the production function. Then, the endogenous growth theory has offered an alternative view by highlighting the importance of human capital elements, such as education, and knowledge through innovation and technology.

Building on the endogenous growth theory, many studies have analyzed human capital proxies and their impact on economic growth, including the works of Islam & Alam (2023), Mohamed et al. (2021), Özdoğan Özbal (2021) and Pelinescu (2015).

Variables associated with environmental sustainability and innovation have gained relevance in recent times. This is reflected in studies that explored how Research and Development (R&D), Information and Communication Technology (ICT), and patents positively impact national economic growth namely, Ahmad & Zheng (2023), Hussain et al. (2021) and Nair et al. (2020).

In parallel, concerns about climate change and its relationship with economic growth, have resulted in important studies, such as those developed by Nwaeze et al. (2023) and Rahman et al. (2017).

Additionally, the research of Chirwa & Odhiambo (2016) demonstrate, through qualitatively reviewing, how the factors that influence economic growth differ between developed and developing economies.

The literature in the field of economic growth is a persistent attempt to explore other variables or new proxies for the same variable, new methodologies, or an unexplored country or countries to achieve a more consensual opinion of what drives or harms economic growth. With those empirical findings, the policy direction could be oriented more efficiently, to achieve sustainable economic growth development.

The studies in the literature on economic growth aim to understand the impact between the independent variables and economic growth. To achieve that, the most common methodologies used, identified in the literature, are the Generalized Method of Moments (GMM), Autoregressive distributed models (ARDL), and Ordinary Least Squares (OLS). Despite these advances in identifying the positive or negative impact of various factors on economic growth, through the previous methodologies, one question remains unanswered: what is the explanatory percentage of each variable in economic growth?

Answering this question is essential for policymakers to prioritize policies, and act more efficiently and effectively to promote economic growth. By ranking the explanatory power of each factor, policymakers can identify which factors contribute most to economic growth and focus on stimulating those areas to achieve better results.

This study conducted a time series analysis of U.S. economic growth from 1970 to 2019, with Gross Fixed Capital Formation (GFCF), Human Capital (HC), and Trade Openness (Trade) as independent variables. After the appropriate conduction of a time series analysis, the commonality analysis was applied, complemented with an extra procedure, to address the previous question.

This study aims to provide a better understanding of the explanatory power of each independent variable of the model, in the economic growth of the U.S., by partitioning the R^2 into individual contributions from each variable.

The results show that 83.52% of the variance of Gross Domestic Product (GDP) is explained by the three previous predictors, whereas 78.16% of the variance of GDP is explained only by GFCF, 5.29% only by Trade, and 0.07% only by HC. These values indicate that the most explanatory power of the variance of economic growth in the U.S. is the Gross Fixed Capital Formation. Interestingly, contrary to endogenous growth theory, human capital is not a highly explanatory predictor of economic growth.

The study is organized as follows: Chapter 2 initially presents a chronological review of the literature on economic growth, tracing its evolution. It also reviews more recent research on the variables used to explain growth. Chapter 3 begins with an overview of the theoretical overview of nonstationarity and commonality analysis, followed by descriptive statistics. Chapter 4 implements and explains in detail the model and methodology described in the previous chapter. Chapter 5 concludes with major findings, limitations of the research, and suggestions for future research.

CHAPTER 2

Literature Review

The number and regularity of studies on economic growth demonstrate its continued paramount importance to researchers and policymakers. This interest may be explained by the lack of agreement among researchers on the variables that contribute to economic growth, which is probably due to the different methodologies and data periods used, the heterogeneity of countries, and the inadequate selection of variables (Rahman et al., 2019; Rahman & Alam, 2021).

Many theories have been proposed to explain economic growth, which has differed depending on the factors the authors thought would affect economic growth and in which manners.

In the 1950s the neoclassical growth model developed by Solow (1956) posited that the output is a function of capital, labor, and technological progress, with the latter considered exogenous. The economy tends toward a steady state—an equilibrium in the long run— where it only grows if technological progress changes.

The endogenous growth theory presents a new development to the previous one by demonstrating the importance of human capital, including elements such as education or innovation, in long-term economic growth (Pelinescu, 2015). This perspective is notably advanced by the research of Romer (1990) and Lucas (1988), among others.

The model developed by Romer (1990) differs from the one developed by Solow (1956) by treating technology as an endogenous variable.

Romer (1990) defines a model with capital, labor, human capital, and an index of the level of technology as inputs. Human capital is defined as a "cumulative effect of activities such as formal education and on-the-job training" (Romer, 1990, p. 79). The implication of this model is "that an economy with a larger total stock of human capital will experience faster growth" (Romer, 1990, p. 99). Since "human capital can be privately provided and traded in competitive markets" (Romer, 1990, p. 75), it is possible to conclude "that free international trade can act to speed up growth" (Romer, 1990, p. 99).

Lucas (1988) considered two models: one emphasizes human capital accumulation through "*schooling*", and the other through "*learning-by-doing*". The first model considers human capital as "levels [that] affect current production and the way the current time allocation affects the accumulation of human capital" (Lucas, 1988, p. 17). The other model sees "human capital as a result of learning-by-doing [assuming] that the growth of human capital increases with the effort devoted to producing goods" (Lucas, 1988, p. 28). The main conclusion of his work is that human capital enhances the productivity of labor and physical capital, thereby driving economic growth.

In the opinion of Ruttan (1998), the most significant contribution of the research of Romer and Lucas "has been their endogenization of human capital formation" (Ruttan, 1998, p. 7).

Mankiw et al. (1992) extended the Solow model by incorporating human capital into production function, with the assumption that technology is exogenous. The authors identify four main implications from their study, which both align with and diverge from Solow's original conclusions. A significant takeaway from their research is the recommendation for future studies to focus on understanding why the exogenous variables in the Solow model differ so widely across countries.

Additionally, the Mankiw, Romer, and Weil model assumes that the technology used to produce human capital is the same as that used to produce physical capital. Klenow & Rodríguez-Clare (1997) critique this assumption, referencing Kendrick's (1976) research that "presents evidence that the technology for producing human capital is more intensive in labor than is the technology for producing other goods" (Klenow & Rodríguez-Clare, 1997, p. 83).

The three models developed by Solow, Romer, and the Augmented Solow Model by Mankiw, Romer, and Weil are still being utilized in recent papers on economic growth as demonstrated by the works of Matousek & Tzeremes (2021), Nguyen & Doytch (2022), Rahman et al. (2017), Rahman & Alam (2021) and Trejos & Barboza (2015). All these researchers have extended these foundational models by transforming their functions or incorporating additional variables to provide a deeper understanding of economic growth.

In contrast, other studies have examined how emerging factors, such as climate change and technological advancements, impact economic growth, often alongside traditional variables. The theory of market value illustrates this approach by emphasizing "the influence of intangible assets such as research and development, patents, intellectual capital on the market value of companies and also on their development, leading ultimately to economic growth overall national, regional or global" (Pelinescu, 2015, p. 185).

Supporting this perspective, Nair et al. (2020) found a positive long-run relationship between Research and Development (R&D) activities, Information and Communication Technology (ICT) infrastructure development, and economic growth in OECD countries. Similarly, Olaoye et al. (2021) identified a positive relationship between R&D expenditure and economic growth in four African countries. Additionally, Ahmad & Zheng (2023) observed that R&D expenditure and patents are linked to the business cycle for thirty-six OECD economies, with the relationship being pro-cyclical: during the boom phase, positive shocks in these variables lead to higher economic growth.

Several recent studies show a growing interest in ICT topic, in fact, according to Vu et al. (2020), 208 academic papers studied the link between ICT and economic growth from 1991 until 2018. For instance, Hussain et al. (2021) found evidence that showed a positive impact in the short and long run of ICT penetration on the economic growth of South Asian countries.

According to the analysis of Nguyen & Doytch (2022), which covered 43 economies (26 developed and 17 emerging market economies), the short-term effects of total patents and ICT patents on real GDP per capita growth are not statistically significant. However, while total patents do not affect economic growth over the long term, ICT patents do have a significant impact. Furthermore, it was discovered that advanced economies experienced a greater impact on economic growth from total patents compared to emerging economies.

Dahmani et al. (2022) tried to understand the impact of ICT investment and trade openness on Tunisia's economic growth. This study is noteworthy for its analysis of the variable's effects at the sector level. A cross-section ARDL was applied and the results in the short and long run are similar, the estimations of the coefficients of those two variables are positive and statistically significant. However, by looking at the long-run estimates at the sector level of trade openness, the results have a mixed effect depending on the sector. Authors state that this "can be reversed by structural reforms aimed at improving access to investment, human capital, innovation capabilities, and competitiveness and diversifying the economy" (Dahmani et al., 2022, p. 2330). Additionally, ICT investment does not significantly affect productivity at the sectoral level, with a few exceptions. This is explained by the authors since certain industries have not made the organizational or financial changes necessary to incorporate new technologies.

The study of Trejos & Barboza (2015) also focused on the relationship between trade openness and output growth in twenty-three Asian countries. The main conclusion of this study was that there was evidence that “[refutes] the argument that the Asian economies miracle was driven by increased trade openness” (Trejos & Barboza, 2015, p. 120). Through the Error Correction Model (ECM), the results for the long run show that an increase in trade openness increases output growth with statistical significance, only in two Asian countries, showing that this result is an exception to the major conclusion.

The article by Zaman et al. (2021) is similar to Dahmani et al. (2022) in that, both studies explore trade-related factors and technology's impact on economic growth, where Zaman et al. (2021) studied the trade openness and Information Technology (IT) exports for Belt and Road Initiative (BRI) countries using a two-step system Generalized Method of Moments (GMM). The findings demonstrate that, at a significant level of 1%, gross capital formation and foreign direct investment (FDI) have a positive impact on economic growth. On the other hand, trade openness and IT exports have a negative impact on economic growth with only IT exports being statistically significant.

The study developed by Qayyum & Zaman (2019) stands out due to its focus on understanding the relationships between various economic variables using Granger causality, in addition to the conventional interpretation of estimation coefficients. For Pakistan, their analysis through cointegration reveals that, in the long run, the relationships between gross capital formation and labor force with GDP are negative, whereas trade openness has a positive relationship with GDP. Granger causality tests indicate a unidirectional causality from economic growth to trade openness and from the labor force to economic growth.

Chirwa & Odhiambo (2016) identified the problem of the lack of agreement regarding the elements that promote economic growth, which was earlier brought up by Rahman et al. (2019) and Rahman & Alam (2021). Beyond that, they notice a gap in the literature that does not distinguish the drivers of economic growth between developing and developed countries. For that reason, "by qualitatively surveying from previous empirical studies" (Chirwa & Odhiambo, 2016, p. 41) they were able to fill that gap. The findings show that many research conclusions show that "physical capital is largely positive and significantly associated with economic growth" (Chirwa & Odhiambo, 2016, p. 39) in developing and developed countries, apart from a few exceptions. The same cannot be concluded with the same certainty for human capital development, since the authors found a lot of incongruences in the results of this factor for both developed and developing countries. Additionally, variables related to trade such as trade openness have a positive and significant impact on economic growth in developed and developing economies. To conclude, according to this study, focusing on developed countries, the main drivers of economic growth are physical capital, technological factors, human capital, and proxies related to trade.

Still within the scope of studies that stand out for their distinction between countries' income levels, the study by Batrancea et al. (2021) covers 34 African countries, which have the particularly of include four types of economies (low-income, lower-middle-income economies, upper-middle-income economies, and high-income economies). This paper aims to understand the impact of imports, exports, foreign direct investment inflows and outflows, gross domestic savings, and gross capital formation on economic growth using a panel data analysis from 2001 to 2019.

The authors created two models with cross-section fixed effects. The first model has all the independent variables that I mentioned above except the imports and the second one replaces the imports for exports, ceteris paribus. The results of the first model with and without time-fixed effects show that the only variable statistically significant is imports, which positively affects the GDP rate. The second model, whether with or without time-fixed effects, leads to the same conclusion: exports and capital are statistically significant and positively impact GDP.

Also using panel data analysis, Rahman & Alam (2021) for the world's 20 biggest economies studied the effects of energy use, trade, capital, labor, foreign direct investment, and human capital on economic growth in the data period between 1980 and 2018.

After the estimation of the parameters by using the dynamic ECM with the panel pooled mean group, the conclusions were that all the independent variables have a positive and significant impact on economic growth in the long run. The conclusion for the short-run effects is a little different: energy use, trade, and capital have a positive and statistically significant impact on economic growth, but labor and human capital have a negative. An interesting result is that the variable human capital has opposite signs in the short and long run, however, according to the authors it is due to the "benefit from human capital development on economic growth is not immediate" (Rahman & Alam, 2021, p. 6). Specifically, for the United States, the estimation for the coefficient in the short run of trade is negative and has a significant impact, against previous results.

Similarities exist between the research conducted by Thaddeus et al. (2024) and Rahman & Alam (2021) since both studies were able to comprehend the short and long-term effects of the variables. This study considers several explanatory variables to explain the economic growth of Cameroon country from 1970 to 2018, which are reflected in the good fit of the model proved by the high value of R^2.

Specifically, for the variables trade openness, gross capital formation, and human capital the Autoregressive Distributed Lag (ARDL) results show that all have a positive relationship with economic growth in the long run. The authors justify their previous trade openness finding by stating that the "government supports more exports than imports, resulting in long-term economic growth" (Thaddeus et al., 2024, p. 10). Additionally, the positive relationship between the variable used as a proxy for investment and economic growth shows that the country by its "political and economic environment [attracts] capital investments" (Thaddeus et al., 2024, p. 12). The results for the short run for the previous three variables mentioned remained positive except for the human capital.

As has already been briefly mentioned, the increased awareness of climate change in recent years is reflected in several research that include factors relevant to this phenomenon. Because of this, there is a huge increase in interest in learning how climate change affects economic growth.

This worry is reflected in the study of Nwaeze et al. (2023) studying the 12 top tourist countries of the EU from 1995 to 2018 with five different models. Given the study that will be conducted, our attention will be on the third model, where the dependent variable is GDP per capita and the independent variables are CO2 emissions, energy use, international tourism, and trade openness, without diminishing the significance of the other four models.

A panel ARDL was conducted to achieve the effects in the short and long run of the variables, and the results show that all of them have a positive impact on economic growth in the short run, but only trade openness and tourism have a statistically significant impact. In the long run trade openness and energy use have a positive and statistically significant relationship with the GDP, opposite to CO2 emissions which have a negative but insignificant impact on GDP.

This study produced a curious conclusion about the relationship between trade, CO2 emissions, and economic growth. In the initial phase, an increase in trade increases pollution (from the increase of commodity production and energy use) but also encourages the development of cleaner technologies between countries. In this way, promotes the sustainability of the environment as well as economic growth.

Rahman et al. (2017) is a perfect example of a study that allies the neoclassical theory with emerging factors. The authors extended the neoclassical growth model adding trade openness, population growth, and CO2 emissions. The study was applied to 6 countries (3 major developed and 3 major emerging countries) with an extension of Vector Autoregression (VAR) as an ECM.

Analyzing GDP as a dependent variable the results show that trade openness, labor, and capital impact positively economic growth for the 3 scenarios (3 developed countries, 3 developing countries, and 6 countries), all statistically significant but with different levels. Population growth has a positive impact on economic growth with statistical significance for 3 developing countries and 6 countries. Additionally, the CO2 emissions have a positive impact on economic growth for the 3 scenarios, but none is statistically significant. Such as Qayyum & Zaman (2019), this article also analyses the Granger causality.

The endogenous growth theory has evolved to highlight the importance of human capital and it is a variable that is still currently under debate in the economic literature due to its complexity (Chirwa & Odhiambo, 2016; Teles, 2005).

A seminal contribution in this area is Barro (1991), who empirically demonstrated for 98 countries, the critical role of human capital (using school-enrollment rates as proxy) in driving economic growth. As Barro found “given the level of initial per capita GDP, the growth rate is substantially positively related to the starting amount of human capital. Thus, poor countries tend to catch up with rich countries if the poor countries have high human capital per person (in relation to their level of per capita GDP), but not otherwise” (Barro, 1991, p. 437).

Furthermore, as demonstrated in the study of Matousek & Tzeremes (2021) there are a lot of proxies that can be used to study human capital.

Özdoğan Özbal (2021) used proxies for human capital, such as higher education expenditures, the enrollment rate in higher education, and the Human Development Index. The results show that public spending on higher education is a relevant tool for increasing human development in OECD countries since an increase in higher education expenditures leads to a positive response in higher education attendance and human development index. However, the study also found that “per capita national income reacted in the direction of a decrease” (Özdoğan Özbal, 2021, p. 190) due to factors such as uncoordinated university education with market demands. The author suggests that to address this, higher education must better align with the needs of the labor market. The global conclusion is that human capital positively impacts economic growth as long as higher education is of high quality and is aligned with the demands of the labor market. Specifically, this article stands out from the others by applying a variance decomposition in a VAR model.

The study of Pelinescu (2015) investigated the impact of human capital on economic growth in EU member states from 2000 to 2012, using proxies such as patents (innovative capacity), employees with secondary education (qualification), and education expenditure. The results show that patents and secondary education positively and significantly affect economic growth, consistent with existing literature. However, education expenditure had a negative and unexpected impact on economic growth, explained by the heterogeneity of the countries in the study.

Another proxy for human capital was studied by Mohamed et al. (2021). The authors attempt to understand the impact of human capital, using government spending on health and education as proxies for human capital, on the economic growth of Egypt (1995-2018) with an ARDL approach. The results show that both previous variables are not statistically significant, with government spending on health with a negative sign. This last result may be due to the “lack of enough capacity in Egypt to utilize the productivity of human capital efficiently” (Mohamed et al., 2021, p. 77) and inadequate data in the study, weak health sector governance, and corruption.

Islam & Alam (2023) employed the identical human capital proxies as Mohamed et al. (2021) in their analysis of economic growth for Bangladesh from 1990 to 2019.

The findings of an ARDL technique demonstrate that health expenditure has a beneficial long-term influence on economic growth but in the short run has not. The findings regarding education spending indicate that, while it has a positive short-term impact on economic growth, the long-term relationship is the opposite. The authors attributed this to the fact that "expenditure on education has been unstable and declining, and low compared to that of health expenditures over the years" (Islam & Alam, 2023, p. 3025). They also highlighted that economic well-being is hindered by distorted labor markets and the inability to create employment opportunities for literate adults. Despite high growth rates, job creation remains a challenge, particularly for youth, due to the presence of foreign workers without proper legal documentation.

The scope of research on economic growth appears almost limitless, driven by the constant emergence of new gaps in the literature. These gaps may pertain to unexplored countries, variables, or methodologies. As discussed, scholars have examined a wide array of variables grounded in both neoclassical and endogenous growth theories, as well as factors related to intangible assets and climate change.

Besides that, as seen in the previous paragraphs, the conclusions about how a variable impacts economic growth are not homogenous, and a possible reason for this inconsistency was already given by (Rahman et al., 2019; Rahman & Alam, 2021).

However, a significant gap in the literature persists: beyond determining whether a variable positively or negatively affects economic growth and assessing its statistical significance, there is a lack of understanding about the magnitude of these effects. Questions such as "Which variable explains the most economic growth?" or " What is the explanatory percentage of each variable for economic growth?" remain largely unanswered.

CHAPTER 3

Methodology

The main objective of economic growth studies is to understand the relationship between potential drivers of economic growth and economic growth itself, often distinguishing between the short and long-term effects. Thus, the methodologies are relatively consistent.

The main approaches to achieve these, are the Generalized Method of Moments (GMM), Autoregressive distributed models (ARDL), and Ordinary Least Squares (OLS) along with their adaptations for more complex data structures. For instance, the GMM is utilized in both its two-step system, in studies of Nguyen & Doytch (2022) and Zaman et al. (2021) and its one-step system, exemplified in Rahman et al. (2019). The methodology using an ARDL was applied in the studies of Dahmani et al. (2022), Nwaeze et al. (2023), Rahman & Alam (2021) and Thaddeus et al. (2024). Variants of OLS were applied such as pooled least squares by Pelinescu (2015), the Fully Modified Ordinary Least Squares (FMOLS) by Hussain et al. (2021) and both FMOLS and Dynamic Ordinary Least Squares (DOLS) by Nair et al. (2020) and Rahman & Alam (2021). Additionally, some researchers enhance their analyses with an Error Correction Model (ECM) namely Trejos & Barboza (2015), Rahman et al. (2017), Nair et al. (2020), Hussain et al. (2021) and Thaddeus et al. (2024).

The methodology that will be applied could be divided into two parts: first, a simple Ordinary Least Squares (OLS) approach to time series data, followed by the application of commonality analysis. The OLS approach will provide the estimations of the coefficients, allowing us to determine the impact of each variable on the economic growth of the U.S. This part of the analysis follows conventional methods and does not introduce new techniques. In the second part, applying the commonality analysis (CA) and with my work development will be able to split the R^2 into the portions of each variable in the analysis to answer the primary objective of this work.

The dependent variable to study the economic growth of the U.S. is Gross Domestic Product and the independent variables are Gross Fixed Capital Formation, Human Capital, and Trade openness.

In the literature on economic growth, it is consensual to measure economic growth with real GDP per capita. The GFCF represents the capital variable, and its importance has been noticed in several studies, namely Ahmad & Zheng (2023), Batrancea et al. (2021), Dahmani et al. (2022), Nguyen & Doytch (2022), Qayyum & Zaman (2019), Rahman et al. (2017, 2019), Rahman & Alam (2021), Thaddeus et al. (2024), Trejos & Barboza (2015) and Zaman et al. (2021). For human capital, the proxy used was inspired by the study of Özdoğan Özbal (2021), using an index of human capital, but according to Penn World Table 9.0 (Feenstra et al., 2015). Trade openness was utilized as an independent variable of economic growth studies, namely by Dahmani et al. (2022), Qayyum & Zaman (2019), Rahman & Alam (2021), Thaddeus et al. (2024), Trejos & Barboza (2015) and Zaman et al. (2021). Since almost all these studies were applied to developing economies, an additional explanation for including this variable in the model of economic growth of the U.S. has to be given.

Initially, the model included the variable for trade openness because Portugal was one of the countries analyzed. However, after narrowing the study to the United States, I chose to keep the trade openness variable instead of replacing it with another, since the results obtained with this specification were quite interesting, as will be explained in the chapter on empirical findings.

Based on the literature, it has been more appropriate to have a variable as a proxy for the technology. Although I recognized this limitation, the results are bigger than this and it is logical to maintain this variable since as mentioned earlier, the U.S is an engine of the world economy, one of the reasons is that it is a global trading partner (Arora & Vamvakidis, 2001). Thus, it is already recognized that the U.S. leads to growth in other countries, so it is important to investigate if the country also has a self-benefit in trading.

Additionally, based on the literature, all variables are expected to have a positive impact on economic growth, which means that the OLS estimates of the coefficients are expected to have positive signs.

3.1. Theoretical overview: nonstationarity and commonality analysis

3.1.1. Nonstationarity

A simple definition of time series data is given by Stock & Watson (2011, p. 525) as "data for a single entity (person, firm, country) collected at multiple time periods".

In a time series regression, "the assumption that the future will be like the past" (Stock & Watson, 2011, p. 526) is very important to forecast the future and is formalized by the concept of stationarity.

The importance of having a stationary time series is related to the problem of having spurious regression, Brooks (2008, p. 319) points out that:

> "*If two variables are trending over time, a regression of one on the other could have a high R^2 even if the two are totally unrelated. So, if standard regression techniques are applied to non-stationary data, the end result could be a regression that 'looks' good under standard measures (significant coefficient estimates and a high R^2), but which is really valueless. Such a model would be termed a spurious regression".*

However, some time series are not stationary, due to two main reasons referred by Stock & Watson (2011). The first one is because series can have trends a "persistent long-term movement of a variable over the time" (Stock & Watson, 2011, p. 526) and the other one is due to breaks, related to the fact that "the population regression can be unstable over the time" (Stock & Watson, 2011, p. 526). With formal statistical procedures that detect these breaks and trends, we can adjust the model, making the series stationary.

Stock & Watson (2011) refer that trends can be stochastic (random and varies over time) or deterministic (nonrandom function of time) but it is more common for economic time series to have stochastic trends than deterministic ones. If the time series has a unit root, meaning a stochastic trend, after differencing once, the series does not have a trend and then is stationary.

The nonstationary could be tested with the Augmented Dickey-Fuller test (Dickey & Fuller, 1979).This test estimates the regression model:

$$\Delta x_{it} = \beta_0 + \alpha t + (1-\rho)x_{i,t-1} + \sum_{j=2}^{p} \gamma_p \Delta x_{i,t-j+1} + \varepsilon_t \qquad (1)$$

Where β_0 is the constant term, αt is the deterministic linear time trend, $(1-\rho)x_{i,t-1}$ denotes the stochastic trend and ε_t the errors. The symbol Δ denotes the first difference and the summation term $\sum_{j=2}^{p} \gamma_p \Delta x_{i,t-j+1}$ captures the autocorrelation of the first difference of the variable on the left side of the equation with p order. This last term is the result of the Augmented Dickey-Fuller test, to an AR(p), to achieve white noise errors. Thus, the "lag length, p, remains to be determined" (Greene, 2002, p. 644). The more appropriate value of p is determined by minimizing the value of the information's criteria, the Bayesian information criteria (BIC), also called the Schwarz information criterion (SIC), or the Akaike information criteria (AIC).

If $\alpha = \gamma_p = 0$ the previous regression is reduced to $\Delta x_{it} = \beta_0 + (1-\rho)x_{i,t-1} + \varepsilon_t$, a random walk with drift. The null hypothesis of ADF test is $\rho = 1$, which means that the series contains a unit root, is rejected if $DF_\tau < c$, where c are the critical values proposed by MacKinnon (1996) and DF_τ is the t-statistic proposed by Dickey-Fuller and it is calculated:

$$DF_\tau = \frac{\hat{\rho}-1}{se(\hat{\rho})} \qquad (2)$$

The Phillips–Perron test (Phillips & Perron, 1988) "gives the same conclusions as the ADF tests" (Brooks, 2008, p. 330) and "are similar to ADF tests, but they incorporate an automatic correction to the DF procedure to allow for autocorrelated residuals" (Brooks, 2008, p. 330). The PP test estimates the same regression used in the ADF test (1) and both the critical values and the null hypothesis are the same as the ADF test.

Unlike the two previous tests, the null hypothesis of the KPSS test (Kwiatkowski et al., 1992) is that the time series is stationary or trend stationary against the alternative of the existence of a unit root. They consider that a time series (x_{it}) is decomposed into:

$$x_{it} = \alpha t + z_t + \varepsilon_t \qquad (3)$$

$$z_t = z_{t-1} + u_t \qquad (3.1)$$

Where αt is the deterministic trend, z_t is a pure random walk, ε_t is a stationary error term and $u_t \sim iid(0, \sigma_u^2)$. The null hypothesis of stationary is given by $\sigma_u^2 = 0$.

The test statistic of the KPSS Lagranger multiplier test is:

$$LM = T^{-2} \frac{\sum_{t=1}^{T} S_t^2}{\hat{\sigma}_e^2} \qquad (4)$$

Where $\hat{\sigma}_e^2$ are the residuals variance from the regression (3) and $S_t = \sum_{t=1}^{T} e_t , t = 1,2, \ldots, T$.

The ZA test (Zivot & Andrews, 1992) allows us to test the stationarity with structural changes. The null hypothesis is a unit root process without structural break against the alternative hypothesis of a trend-stationary with possible structural change occurring at an unknown point in time.

3.1.2 Commonality analysis

Commonality analysis "is a procedure for decomposing R^2 in multiple regression analyses into the percentage of variance in the dependent variable associated with each independent variable uniquely and the proportion of explained variance associated with the common effects of predictors" (Seibold & McPhee, 1979, p. 355).

This simple definition is a result of several research (Creager, 1971), however, the three main references about this topic are Newton & Spurrell (1967), Wisler (1968) and Mood (1971).

The starting point of all of them is the multiple regression which attempts to reveal relationships between a dependent variable (y) and a set of regressor variables $(x_1, x_2, x_3 \dots, x_p)$. Thus, the model of a multiple regression:

$$y_i = b_0 + b_1 x_{1i} + b_2 x_{2i} + b_3 x_{3i} + \cdots + b_p x_{pi} + e_i \quad (5)$$

Newton & Spurrell (1967) mentioned the two objectives of using multiple regression, the predictive and the operational. The first one "is to obtain an equation, which will predict values of y from values of $(x_1, x_2, x_3 \dots, x_p)$ with predefined accuracy" (Newton & Spurrell, 1967, p. 52). The operational "wishes to identify those variables which are important in controlling the process and which independently have as large effects upon the residual sum of squares as possible" (Newton & Spurrell, 1967, p. 52).

Furthermore, they discussed how a combination of the square of the multiple correlation coefficient and the significance of the estimates of each coefficient significance can result in a "better" equation. The researchers must look for more than beta weights to understand the relative importance of the predictor variables (Nimon & Reio, 2011; Seibold & McPhee, 1979; Zientek & Thompson, 2006).

According to Zientek & Thompson (2006) there are two mistakes that authors usually make when they want to understand the relative importance of the predictor variables but are faced with multicollinearity: interpreting only regression beta weights and using the stepwise method.

The stepwise method and its problems were identified by authors such as Jernstedt (1980), Zientek & Thompson (2006), Warne (2011), and Ray-Mukherjee et al. (2014). The commonality analysis emerges as an alternative to this method, not resulting in its problems because it has the "advantage of producing the same results for a given set of predictors-regardless of the order in which the predictors are entered into the model" (Amado, 2003 cited in Warne, 2011, p. 314).

In terms of the first mistake, presented by Zientek & Thompson (2006), Nimon & Reio (2011, p. 331) state that "when predictor variables are correlated, a beta weight may not be representative of a predictor variable's contribution to the regression equation". For this reason, Zientek & Thompson (2006) suggest looking at the structure coefficients too. A structure coefficient "is a bivariate correlation between a predictor and the predictive criterion resulting from the regression model" (Nimon & Reio, 2011, p. 331), in other words, are "Pearson correlation coefficients between given predictors and the $\hat{y}$" (Zientek & Thompson, 2006, p. 300). When squared "identify how much variance is common between a predictor and $\hat{y}$" (Ray-Mukherjee et al., 2014, p. 322).

"Structure coefficients are independent of collinearity among variables and have the additional property of ranking independent variables based on their contribution to the regression effect" (Kraha et al., 2012, cited in Ray-Mukherjee et al., 2014, p. 322).

However, looking at these coefficients could not be enough, in the presence of a suppressor, which could be defined as "independent variables that by themselves have very little impact on the dependent variables. However, when combined with other independent variables, suppressors can improve the predictive power of other independent variables in the regression equation" (Warne, 2011, p. 314).

Wisler (1968) clarifies that the variation in a dependent variable could be split into three parts:

1. The part that can be attributed to the regressor variables individually
2. The part that can be attributed to the regressor variables as a group
3. The residual variation which is unexplained by the regression

The 1. and 2. together represent the explained sum of squares (ESS) and 3. is the residual sum of squares (RSS). The sum of these two components is the total sum of squares (TSS).

Thus, the square of the multiple correlation coefficient (R^2) is the proportion of the total variation which has been accounted for by the regression:

$$R^2 = \frac{ESS}{TSS} \tag{6}$$

The focus of the commonality analysis is precisely the second part, mentioned above, meaning, the nonunique part. Wisler (1968, p. 4) explains that "if a set of data vectors are mutually orthogonal then they are uncorrelated with one another and the unique sums of squares will add up to the total explained sum of squares". However, it is usual that variables are correlated with each other (multicollinearity), this leads to the failure of orthogonality of the vectors resulting in the non-unique sums of squares being different from zero.

This nonunique portion is called commonalities and is the difference in the sum of squares which may be associated with two or more variables. If we consider a pair of variables is called second-order commonalities, three variables are a third-order commonality, and so on. The value of these commonalities could be negative, exemplified in the example of Wisler (1968), and the geometric explanation is given by Newton & Spurrell (1967).

Negative commonalities "can occur in the presence of suppression or when some of the correlations among predictor variables have opposite signs" (Ray-Mukherjee et al., 2014, p. 322, cited in Pedhazur, 1997). These authors mentioned that the magnitude of suppression could be examined by looking at negative commonality coefficients. Calculating these coefficients can be done quite easily in RStudio, with the aid of the Nimon et al. (2008) research.

Mood (1971) relates the correlation between two variables and commonality saying that if that correlation is high, the unique parts of each variable decrease and the common part between them increases.

Newton & Spurrell (1967), Wisler (1968) and Mood (1971) have come to the same conclusion: if there are k number of variables, they will have $2^k - 1$ number of commonalities coefficients. Mood (1971) went deeper saying that will be k uniques parts, $\frac{k(k-1)}{2}$ parts in common to two variables, $\frac{k(k-1)(k-2)}{6}$ parts in common to three variables and so on.

Mood (1971) draws attention to the fact that the larger the values of k, the smaller the values of the parts and therefore the more difficult they are to interpret. For a model with 7 predictors, the number of commonalities increased to 127 (Ray-Mukherjee et al., 2014). For this reason, he advises to apply this method to up to three variables. If it is not possible, Wisler (1968) and Mood (1971) suggest grouping variables to reduce the number of k when is conceptually and empirically possible (Seibold & McPhee, 1979).

The steps and methodology commonality analysis will be explained in greater detail in the Empirical Findings chapter. That section will present the approach used to conduct the empirical analysis, as well as the procedures followed to achieve the results.

3.2. Data and descriptive statistics

This time series study on the economic growth of the United States covers the period from 1970 to 2019, chosen based on the availability of observations for all variables. As previously explained, the dependent variable is Gross Domestic Product per capita (GDP), while the independent variables are Gross Fixed Capital Formation per capita (GFCF), Human Capital (HC), and Trade openness (Trade).

A specific justification is required for the GFCF per capita variable. To the best of my knowledge, no existing database provides statistics on GFCF per capita for the United States. Therefore, in this study, GFCF per capita was calculated (using Microsoft Excel) by dividing the GFCF in constant local currency by the total population for each year, thereby obtaining the per capita values.

To preserve some consistency in the units of measurement, both GDP and GFCF are expressed in constant local currency, specifically in US dollars. So, the local currency is US dollars and in real value, which means that is adjusted to inflation which provides a more accurate understanding of the economic reality.

Table 3.2.1 below lists the variables in the study, along with their corresponding definitions, units, and sources. All variables have an annual frequency, so seasonality is not a concern in this analysis.

Table 3.2.1. Data for the econometric study

Variable	Definition[1]	Unit	Source
GDP	GDP per capita is gross domestic product divided by midyear population	Constant local currency	World DataBank
GFCF	Includes land improvements (fences, ditches, drains, and so on); plant, machinery, and equipment purchases; and the construction of roads, railways, and the like, including schools, offices, hospitals, private residential dwellings, and commercial and industrial buildings.	Constant local currency	World DataBank
Trade	Trade is the sum of exports and imports of goods and services measured as a share of gross domestic product	% of GDP	World DataBank
HC	Human capital index, based on years of schooling and returns to education		Penn World Table 10.01

[1] Definition of the respective database.

Population	Total population is based on the de facto definition of population, which counts all residents regardless of legal status or citizenship.	Number of people	World DataBank

Source: Author's elaboration using Penn World Table 10.01 and World Bank (World Development Indicators) information

Table 3.2.2 below, displays descriptive statistics of the four series for the U.S. The correlation matrix of the variables is provided in Appendix A.

Table 3.2.2. Descriptive statistics

	GDP	GFCF	HC	Trade
Mean	42465	8129	3.49	21.59
Median	41409	7417	3.51	21.67
Minimum	25231	3974	3.06	10.76
Maximun	61331	13340	3.75	30.84
Std.dev.	10894.82	2748.15	0.19	5.36
Skewness	0.02	0.16	-0.44	-0.08
Kurtosis	-1.41	-1.36	-0.82	-0.85
Kolmogorov-Smirnov Test	1***	1***	0.99888***	1***

Source: Author's output via RStudio using Penn World Table 10.01 and World DataBank data (***) indicates statistical significance at the significance level of 1%

The decision to use the Kolmogorov-Smirnov test to test the normality of the sample is because it is more appropriate given the sample size of 50 observations (Mishra et al., 2019). The null hypothesis that sample distribution is normal is rejected with a p-value<0.01 for all four series.

The coefficients of skewness and kurtosis are very useful for studying the shape of the probability distribution. Analyzing Table 3.2.2, we can conclude that all series exhibit flat tails (platykurtic[2]). Specifically, GDP and GFCF are right-skewed (positive skewness) while HC and Trade are left-skewed. These observations about the shape of the probability distribution for the four series are consistent with the results of the Kolmogorov-Smirnov test, which indicates that they are not normally distributed.

[2] Values of kurtosis less than 3.

CHAPTER 4

Empirical findings

The necessity of having a stationary time series was discussed in the methodology section, along with the tests used to identify the presence of unit roots. Accordingly, the Augmented Dickey-Fuller (ADF) test, Phillips-Perron (PP) test, Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test, and Zivot-Andrews (ZA) test were applied. Specifically, considering the research covers the period from 1970 to 2019, it is important to note that 2008 was an unusual year for the U.S. economy due to the financial crisis, this underscores the relevance of the ZA test, which has the advantage of detecting potential structural breaks that might be associated with disturbances like that.

The results from those four tests were ambiguous for all series in levels, as summarized in Table 4.1 below, with detailed explanations provided in Appendix B from Table B.1. to Table B.4.

Table 4.1. Summary of the test results of stationarity

	ADF	**PP**	**KPSS**	**ZA**
GDP	Nonstationary	Nonstationary	Stationary	Stationary
GFCF	Nonstationary	Nonstationary	Stationary	Stationary
HC	Nonstationary	Stationary	Nonstationary	Stationary
Trade	Nonstationary	Nonstationary	Stationary	Nonstationary

Source: Author’s elaboration

As demonstrated in Table 4.1, the test results of stationary were ambiguous. Therefore, the first differences of the series in logarithmic form were calculated to address these unclear results, and the same four tests were conducted after these modifications. The results of the four tests are in Appendix B from Table B.5 to Table B.8.

The series of $\Delta \log GFCF$ and $\Delta \log Trade$ are stationary according to all tests. The series $\Delta \log GDP$ is stationary according to all tests except the ADF test. This result is inconsistent with the ADF test result of the original series, where nonstationarity was attributed to a stochastic trend. There are three reasons why the ADF test tends "to accept the null of unit root more frequently than is warranted" (Gujarati & Porter, 2009, p. 759). Considering this, we will assume that the result of the ADF test to series $\Delta \log GDP$ reflects the lower power of this test since the remaining tests are in accordance.

The tests of the presence of unit root for the series $\Delta \log HC$ are still ambiguous. Since the ZA test results demonstrate that the series is stationary, we would confirm if the potential structural breakpoint at position 35 suggested by that test occurred and if it is the reason why these series are still nonstationary for ADF and PP tests. Since we suspect when the point is, the Chow test is adequate. With the Chow test (Chow, 1960) it is possible to conclude "if in fact there is no structural change" (Gujarati & Porter, 2009, p. 257). The null hypothesis of no break or structural change is rejected when the value of the F ratio is higher than the critical F value, or the p-value is lower than 0.05. The results of the Chow test to test if there is a break at position 35 in series $\Delta \log HC$ shows that the null hypothesis of no structural change is not rejected with a p-value of 0.5525. The output of this test is presented in Table B.9 in Appendix B.

Given the findings from the KPSS and ZA tests, along with the Chow test's confirmation that no structural break exists, it is reasonable to conclude that the series $\Delta \log HC$ is stationary. Nonetheless, we will confirm this assumption by the informal methods, by looking at the autocorrelation function of the series after computing the logarithm and first differences, visible in Figure B.1 in Appendix B. It is possible to conclude with more certainty, by the ACF of the series that its behavior seems a stationary series because the ACF decays fast to zero.

After this analysis, it is reasonable to assume that all series are stationary after the appropriate transformations. Therefore, the model that will be estimated is:

$$\Delta \log GDP_t = \beta_0 + \beta_1 \Delta \log GFCF_t + \beta_2 \Delta \log HC_t + \beta_3 \Delta \log Trade_t + u_t \qquad (7)$$

The model estimate is expressed below in Table 4.2.

Table 4.2. Model estimated by Ordinary Least Squares

Variable	Estimate	Standardized	Std. Error	T value	Pr(> |t|)	
Intercept	0.008715	NA	0.002224	3.919	0.0003	***
$\Delta \log GFCF$	0.338171	0.902882	0.023849	14.179	<2e-16	***

$\Delta \log Trade$	0.012357	0.044323	0.019060	0.648	0.5201
$\Delta \log HC$	0.198921	0.029664	0.444797	0.447	0.6569
Multiple R-squared: 0.8352					
Adjusted R-squared: 0.8243					

Source: Author's output via RStudio

(***) indicates statistical significance at the significance level of 1%

Interpreting the estimations of the unstandardized coefficients of the regression, 0.34 is the elasticity of the growth rate of GDP in relation to the growth rate of GFCF; 0.012 is the elasticity of the growth rate of GDP in relation to the growth rate of Trade, and 0.20 is the elasticity of the growth rate of GDP in relation to the growth rate of HC. This means that, on average, if the GFCF increases by 1%, the GDP increases by 0.34%, ceteris paribus. Additionally, if Trade increases by 1%, the GDP increases by 0.012%, ceteris paribus and if HC increases by 1% the GDP increases by 0.20%, ceteris paribus. The expected value of $\Delta \log \text{GDP}$ is 0.87%, when X=0, ceteris paribus. Only the term intercept and $\Delta \log GFCF$ are statistical significative at the level of 1%, the remain variables are not statistical significative. All the variables have a positive relationship with economic growth which aligns with the expectations formulated earlier.

For GFCF, the positive impact on economic growth is aligned with the authors who also studied the impact of proxies of investment, namely Zaman et al. (2021) and Batrancea et al. (2021). Also, a positive relationship was confirmed in the short and long run by Rahman & Alam (2021) and Thaddeus et al. (2024). The study of Qayyum & Zaman (2019) was the only one, identified in the literature review, that found a negative relationship in the long run, which is the opposite of the empirical findings of this study.

The positive relationship between trade openness and economic growth in both the short and long run was found by Dahmani et al. (2022), Rahman & Alam (2021), Thaddeus et al. (2024) and Nwaeze et al. (2023), aligned with my previous findings. Specifically, despite the fact of the positive results in the paper of Rahman & Alam (2021), for the U.S. country, the relationship in the short run was negative. The same positive sign was found by Rahman et al. (2017) and Qayyum & Zaman (2019) for the opposite of the findings of Trejos & Barboza (2015) and Zaman et al., 2021) who found a negative impact in the long run.

For human capital and its proxies, the studies of Rahman & Alam (2021) and Thaddeus et al. (2024) show a positive relationship between human capital and economic growth in the long run but negative in the short run. This study's results align with the results of these authors in the long run. Additionally, Özdoğan Özbal (2021) findings show that, overall, human capital has a positive relationship with economic growth but with some limitations related to other factors.

The commonality analysis was applied to the previous model, to decompose the R^2 value of 83.52%, into unique and common components. These components were calculated using the formulas offered by Seibold & McPhee (1979), with detailed computational steps provided in Table C.1 in Appendix C.

As illustrated in Table C.1 (Appendix C), all possible R^2 combinations were required and were retrieved via RStudio. However, with the assistance of the SAS PROC RSQUARE procedure, these computations could be simplified (Murthy, 1994; Rowell, 1991).

Using Microsoft Excel, the formulas in Table C.1 were put into practice, and the outcomes are shown in Table 4.3 below.

Table 4.3. Commonality analysis results

U ($\Delta \log GFCF$)	0.7361
U ($\Delta \log Trade$)	0.0015
U ($\Delta \log HC$)	0.0007
C ($\Delta \log GFCF$, $\Delta \log Trade$)	0.0972
C ($\Delta \log GFCF$, $\Delta \log HC$)	0.0179
C ($\Delta \log Trade$, $\Delta \log HC$)	0.0015
C ($\Delta \log GFCF$, $\Delta \log Trade$, $\Delta \log HC$)	-0.0197
R^2	0.8352

Source: Author's output via Microsoft Excel

The sum of all of the unique and common components should be equal to the squared multiple correlation (R^2) of the regression model with all independent variables (Murthy, 1994; Rowell, 1991; Wisler, 1968). The results in Table 4.3 indicate that the verification was successful.

The R^2 of model the is 0.8352 which means that our three predictors explained 83.52% of the total variation in $\Delta \log GDP$. This high value of R^2 reflects a good fit of the model.

By interpreting the components of commonality analysis is possible to conclude that GFCF uniquely explains 73.61% of the variance of $\Delta \log GDP$. The variables Trade and HC uniquely explain 0.15% and 0.07% of the variance of $\Delta \log GDP$, respectively. A similar interpretation could be given as the unique contribution of the variable GFCF: the unique contribution of the variable GFCF, to the proportion of total dependent variable variance explained is 73.61%.

From the unique effects results analyzed above, it is possible to conclude that GFCF stands out from the others with a huge explanatory power of the variance of economic growth of the U.S., meaning that it is the dominant factor in predicting this country's economic growth (in this model). Since the GFCF is statistically significant at the level of 1% while the remaining variables are not, this conclusion through the OLS model is consistent with CA results.

Focusing on the common effects, the common variance of the model shared by GFCF and Trade stands out by their value, which is higher than the unique effects of Trade or HC and is almost 10%. The two remaining second-order commonalities are the common variance of the model shared by GFCF and HC which is nearly 2% and the common variance of the model shared by Trade and HC with a value near zero.

The common variance "occurs when two or more independent variables explain the same proportion of dependent variable variance and this variance would be included in the R^2 if the other variable(s) were eliminated from the multiple regression" (Warne, 2011, p. 316). This possibility is explained by Wisler (1968) saying that "if a set of data vectors are mutually orthogonal then they are uncorrelated with one another and the unique sums of squares will add up to the total explained sum of squares" (Wisler, 1968, p. 4). However, it is usual that variables are correlated with each other (multicollinearity), this leads to the failure of orthogonality of the vectors resulting in the non-unique sums of squares being different from zero. These nonunique portions are called commonalities and are the differences in the sum of squares which may be associated with two or more variables.

From Table 4.3 the third-order commonality differs from the others due to its negative value. Authors such as Frederick (1999) as cited in Capraro & Capraro (2001, p. 20) have suggested that "negative commonalities should be interpreted as zero", while others argue that "negative partitions are possible, although they are illogical because we are partitioning the R^2" (Zientek & Thompson, 2006, p. 305) or "counterintuitive since the result could be taken to mean that … predictor variables have in common the ability to explain less than 0% of the variance" (Thompson, 1985, as cited in Rowell, 1991, p. 13).

However, other authors defend those negative commonalities “can occur in the presence of suppression or when some of the correlations among predictor variables have opposite signs” (Pedhazur, 1997, as cited in Ray-Mukherjee et al., 2014, p. 322).

Beckstead (2012) compiled the definitions of suppressor proposed by different authors to provide a synthesis of the various perspectives found in the literature. Thus, there are three types of suppression: classic suppression, which occurs when the “correlation between the predictor and criterion is zero” (Beckstead, 2012, p. 226), negative suppression, when the “suppressor has a very small but positive correlation with the criterion” (Beckstead, 2012, p. 226) and reciprocal suppression “which occurs when the predictor and suppressor have positive correlations with the criterion but are negatively correlated with one another” (Beckstead, 2012, p. 226).

However, these suppressor variables suppress or remove the irrelevant variance, and thus improve the R^2 (Ray-Mukherjee et al., 2014; Zientek & Thompson, 2006). The “irrelevant variance is the variance shared with another predictor and not with the dependent variable, and hence, it does not directly affect R^2” (Pedhazur, 1997, as cited in Ray-Mukherjee et al., 2014, p. 322).

Since the only negative commonality is associated with the third-order commonality, it is plausible that any of the three variables could be a suppressor. There are several ways to identify a suppressor, one of those is revealed “when it has a large beta coefficient in association with a disproportionally small structure coefficient that is close to zero” (Ray-Mukherjee et al., 2014, p. 323). From Table 4.2, it is possible to conclude that the variable HC approximates the most of a suppressor variable, comparing estimations of both coefficients. The other way is the “difference in signs between the beta weight and structure coefficient” (Nimon & Reio, 2011, p. 337), also looking at Table 4.2 this does not occur with any variable. We cannot definitively say from these two approaches if HC is a suppressor variable.

As explained earlier, negative commonalities can occur not only due to suppressor variables but also when correlations among predictor variables have opposite signs. From Table C.2 in the Appendix, it is evident that this is the case, as seen with the variable HC, which is negatively correlated with GFCF, while the correlation between GFCF and Trade is positive. Therefore, the negative value in the third-order commonality is better explained by the opposite correlations between these variables.

Based on the analyses conducted, we can conclude that the presence of opposite signs correlations between the variables adequately justifies the negative third-order commonality value, and the suppression hypothesis does not hold strongly in this case. While the variable HC shows some characteristics of a suppressor, the evidence from the beta coefficients and structure coefficients does not definitively support this conclusion. Thus, the most plausible explanation for the observed negative commonality lies in the opposite correlations, ruling out suppression as a justification.

Despite the major conclusion that CA allows us to retrieve which will be explained later, the objective is to allocate the R^2 value to each variable individually. The commonality analysis was a great driving force to achieve the primary objective since now it is only necessary to find a way to split the common components.

For now, the R^2 can be split into unique and common components, using only CA, like this:

$$R^2 = U(x_1) + U(x_2) + U(x_3) + C(x_1, x_2) + C(x_1, x_3) + C(x_2, x_3) + C(x_1, x_2, x_3) \quad (8)$$

The values of the components of the equation above are expressed in Table 4.3.

The common components result from the problem of multicollinearity, as explained before. The values of the variance inflation factor (VIF) were computed to know if the multicollinearity may be problematic and the results, displayed in Table C.3 in the Appendix, show that the regressors are moderately correlated, with VIF values around 1, thus are not problematic.

Having said that, the thought is if a common variance is the variance that is shared by two or more variables, we cannot assume that they contribute to that variance in the same proportion, meaning 50% of one variable and 50% of the other variable. Because of this, we need to know the proportion of the variable $x_1$and x_2 in the common variance. However, we can know what is the proportion that each variable has in the common variance, like this:

$$Proportion\ of\ x_1\ in\ C(x_1, x_2) = \frac{Variance(x_1)}{Variance(x_1) + Variance(x_2)} \quad (9)$$

Multiply the last result by the common variance:

$$Unique\ contribution\ of\ (x_1)\ in\ C(x_1, x_2) = Proportion\ of\ x_1\ in\ C(x_1, x_2) C(x_1, x_2) \quad (10)$$

We have the amount of unique portion of the variable x_1 inside the common variance between the variable $x_1$and x_2. The formulas (9) and (10) were applied replacing the variable x_1 for the variable x_2 and for all second-order commonalities with the respective variables.

For the third-order commonalities, the denominator of formula (9) was replaced by the sum of all three variances of variables, maintaining the same logic after this nuance.

This logic was applied for all commonalities, using for the effect, Microsoft Excel, the results are presented in Table 4.4 below, with the values of variance of each variable present in Table C.4. in Appendix C.

Table 4.4. Splitting the common variance

Common variance		Proportion		Unique contribution
C($\Delta \log GFCF$, $\Delta \log Trade$)	0.0972	GFCF	0,3565	0,0346
		Trade	0,6435	0,0625
C($\Delta \log GFCF$, $\Delta \log HC$)	0.0179	GFCF	0,9969	0,0179
		HC	0,0031	0,0001
C($\Delta \log Trade$, $\Delta \log HC$)	0.0015	Trade	0,9983	0,0015
		HC	0,0017	0,0000
C($\Delta \log GFCF$, $\Delta \log Trade$, $\Delta \log HC$)	-0.0197	GFCF	0,3561	-0,0070
		Trade	0,6428	-0,0126
		HC	0,0011	0,0000

Source: Author’s output via Microsoft Excel

Summing the results of the unique contributions of Commonality Analysis and the unique contributions after the application of the logic explained above, it is possible to allocate the amount of R^2 into each independent variable considered in the model. The results are displayed in Table 4.5 below.

Table 4.5. Allocating the value of R^2 into each independent variable

Variable	R^2
GFCF	0,7816
Trade	0,0529
HC	0,0007
Total	0.8352

Source: Author's elaboration

The results from Table 4.5 show that 83.52% is the variance of GDP that is explained by the 3 predictors. Additionally, from that 83.52%, 78.16% of the variance of GDP is explained by GFCF, 5.29% by Trade, and 0.07% by HC.

The GFCF, besides a positive and statistically significant impact on U.S. economic growth, is the key driver in this model of economic growth. The investment namely in equipment and machinery improves efficiency in productivity, leading to higher economic growth in the U.S. In another way, investment in the construction of infrastructure creates employment during that period but also after, creating jobs in those infrastructures, namely, schools, hospitals, and commercial buildings. The U.S. government such as private companies should be promoting investment in fixed assets, to lead to economic growth.

The variable trade openness and human capital have a comparatively lower value in the variance of GDP. Additionally, their impact is positive but not statistically significant. As explained before, the variable trade openness was maintained in the model due to this curious result: human capital explains less the economic growth of the U.S. rather than trade openness (0.07% for HC versus 5.29% for Trade). This unexpected result and a contradiction to endogenous growth theory may be due to several reasons.

One of the reasons is that the GFCF variable includes the construction of schools and hospitals. From the literature review, papers like Islam & Alam (2023) and Mohamed et al. (2021) used government spending on health and education as proxies for human capital, for this reason, maybe some of the explanatory power on economic growth that could be captured by human capital was absorbed by variable GFCF. Another reason, related to the first, is the choice of the variable as a proxy for human capital, maybe another proxy could have led to other results. Given that we are discussing the U.S. economic growth, which is a highly innovative economy, perhaps patents, as utilized in the Pelinescu (2015) research would be a better option.

Although human capital does not exhibit the expected explanatory power on economic growth, as suggested by its importance in the literature, we cannot overlook the fact that trade openness plays, a comparatively higher role in explaining the economic growth of the U.S. The result suggests that the economic growth of the U.S. is more driven by factors related to trade openness than by returns on education. However, this does not necessarily diminish the long-term importance of human capital but rather highlights the important role of international trade in a highly developed and diversified economy like the U.S.

Additionally, compared to the OLS model, even though human capital has a positive impact, but not statistically significant, on economic growth, meaning HC increases by 1% the GDP increases by 0.20%, ceteris paribus, this variable has little explanatory power in the variance of GDP. This conclusion is a reflection of the power of splitting R^2 and the conclusions that can be retrieved.

CHAPTER 5

Conclusions

This study made it possible to answer the question: "What is the explanatory percentage of each variable in U.S. economic growth?". The analysis revealed that gross fixed capital formation explains 78.16% of the variance of GDP, trade openness explains 5.29% of the variance of GDP and 0.07% is the explanatory percentage of human capital in the variance of GDP.

These results provide important information for U.S. policymakers, however, these conclusions should be taken with caution because trade and human capital variables are not statistically significant.

The main factor driving economic growth, among the variables analyzed, is investment (measured by gross fixed capital formation), suggesting that policies to stimulate investment should continue to be prioritized.

On the other hand, the low explanatory power of human capital in economic growth indicates that this factor should be more stimulated, for example, through greater investment and qualification in education.

The relatively high value of the explanatory percentage of trade openness indicates that the U.S. should continue promoting international agreements.

Additionally, the impact of gross fixed capital formation, human capital and trade openness in the U.S. economic growth is positive, with only gross fixed capital formation statistically significant.

5.1. Limitations

The first limitation is related to the fact that the application of commonality analysis is recommended to be limited to three or four predictors. Consequently, this study was restricted to the three explanatory variables, gross fixed capital formation, human capital, and trade openness, for this reason, the reduced number of variables leads to the reduction of valuable information to explain economic growth.

The second limitation refers to the results of the tests for the presence of unit roots. The stationarity tests, especially for human capital time series, presented ambiguous results, raising doubts about its true nature. Although complementary methodologies such as the Chow test and the Autocorrelation Function (ACF) were applied to confirm stationarity, it is not possible to guarantee with complete certainty that the series is stationary.

The third limitation relates to the commonality analysis, we have to take into account that the results of uniqueness contribution results are specific to that set of predictors in the model, if it is deleted or added variables, the results may be different (Pedhazur, 1997, as cited in Ray-Mukherjee et al., 2014), for this reason, the main conclusions cannot be generalized.

There is a gap in the literature regarding the application of this methodology to time series data. It is unclear whether a minimum number of observations is required, which presents a challenge for studies like mine.

Another limitation, and, in my opinion, the most concerning is the possibility of negative common variance in commonality analysis. The justifications given by the authors and utilized in my study could be not satisfactory for all investigators and, thus, require careful reconsideration in future research.

5.2. Suggestions for further research

For further research, it will be interesting to replace the proxy for human capital with others, to determine whether the low contribution of human capital to economic growth persists.

Additionally, due to the interesting result of the relatively high weight of trade openness in explaining the economic growth of the U.S., it would be helpful to investigate its impact by sector, which could inform more targeted policy interventions.

This research was a step to aware the researchers of the importance and the conclusions that will be able to take when we split the R^2. I recognize the advantages but also the limitations of using commonality analysis to achieve that objective. However, the most skeptical researchers could find a way to overcome the commonality analysis. If it is found, it is possible, to have a model of economic growth with several variables and determine the weight of each one in R^2.

Additionally, although this methodology was used in the context of economic growth, it might also be utilized in other fields where the decomposition of the R^2 is relevant.

References

Ahmad, M., & Zheng, J. (2023). The Cyclical and Nonlinear Impact of R&D and Innovation Activities on Economic Growth in OECD Economies: a New Perspective. Journal of the Knowledge Economy, 14, 544–593. https://doi.org/10.1007/s13132-021-00887-7

Arora, V., & Vamvakidis, A. (2001). The Impact of U.S. economic growth on the rest of the world: How much does it matter? (IMF Working Paper No. WP/01/119). International Monetary Fund

Barro, R. J. (1991). Economic Growth in a Cross Section of Countries. The Quarterly Journal of Economics, 106(2), 407–443.

Batrancea, L., Rathnaswamy, M. M., & Batrancea, I. (2021). A Panel Data Analysis of Economic Growth Determinants in 34 African Countries. Journal of Risk and Financial Management, 14(6), 1–15. https://doi.org/10.3390/jrfm14060260

Beckstead, J. W. (2012). Isolating and Examining Sources of Suppression and Multicollinearity in Multiple Linear Regression. Multivariate Behavioral Research, 47(2), 224–246. https://doi.org/10.1080/00273171.2012.658331

Brooks, C. (2008). Introductory Econometrics for Finance (Second Edition).Cambridge University Press.

Capraro, R. M., & Capraro, M. M. (2001). Commonality Analysis: Understanding Variance Contributions to Overall Canonical Correlation Effects of Attitude Toward Mathematics on Geometry Achievement. Multiple Linear Regression Viewpoints, 27(2).

Chirwa, T. G., & Odhiambo, N. M. (2016). Macroeconomic determinants of economic growth: A review of international literature. South East European Journal of Economics and Business, 11(2), 33–47. https://doi.org/10.1515/jeb-2016-0009

Chow, G. C. (1960). Tests of Equality Between Sets of Coefficients in Two Linear Regressions. Econometrica, 28(3), 591–605.

Creager, J. A. (1971). Orthogonal and Nonorthogonal Methods for Partitioning Regression Variance. American Educational Research Journal, 8(4). https://doi.org/10.2307/1162134

Dahmani, M., Mabrouki, M., & Ben Youssef, A. (2022). ICT, trade openness and economic growth in Tunisia: what is going wrong? Economic Change and Restructuring, 55(4), 2317–2336. https://doi.org/10.1007/s10644-022-09388-2

Dickey, D. A., & Fuller, W. A. (1979). Distribution of the Estimators for Autoregressive Time Series With a Unit Root. Journal of the American Statistical Association, 74(366), 427–431. https://doi.org/10.1080/01621459.1979.10482531

Feenstra, R. C., Inklaar, R., & Timmer, M. P. (2015). Human capital in PWT 9.0.

Feenstra, R. C., Inklaar, R., & Timmer, M. P. (2019). Penn World Table versão 10.01. University of Groningen. https://www.rug.nl/ggdc/productivity/pwt/

Frederick, B. N. (1999). Partitioning variance in the multivariate case: A step-by-step guide to canonical commonality analysis. In B. Thompson (Ed.), *Advances in social sciences methodology* (Vol. 5, pp. 305-318). Stamford, CT: JAI Press.

Greene, W. H. (2002). Econometric analysis (Fifth Edition). Prentice Hall.

Gujarati, D. N., & Porter, D. C. (2009). Basic Econometrics (Fifth Edition). McGraw-Hill Education.

Hussain, A., Batool, I., Akbar, M., & Nazir, M. (2021). Is ICT an enduring driver of economic growth? Evidence from South Asian economies. Telecommunications Policy, 45(8). https://doi.org/10.1016/J.TELPOL.2021.102202

Islam, M. S., & Alam, F. (2023). Influence of Human Capital Formation on the Economic Growth in Bangladesh During 1990–2019: an ARDL Approach. Journal of the Knowledge Economy, 14(3), 3010–3027. https://doi.org/10.1007/s13132-022-00998-9

Jernstedt, G. C. (1980). Commonality analysis: Partitioning variance for multivariate prediction. Educational and Psychological Measurement, 40(3). https://doi.org/10.1177/001316448004000317

Kendrick, J. W. (1976). The Formation and stocks of total capital. NBER.

Klenow, P. J., & Rodríguez-Clare, A. (1997). NBER Macroeconomics Annual 1997 (B. S. Bernanke & J. Rotemberg, Eds.; Vol. 12, pp. 73–114). MIT Press.

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. How sure are we that economic time series have a unit root? Journal of Econometrics, 54(1–3), 159–178. https://doi.org/10.1016/0304-4076(92)90104-Y

Lucas, R. E. (1988). On the mechanics of economic development. Journal of Monetary Economics, 22, 3–42.

Mankiw, N. G., Romer, D., & Weil, D. N. (1992). A contribution to the empirics of economic growth. The Quarterly Journal of Economics.

Matousek, R., & Tzeremes, N. G. (2021). The asymmetric impact of human capital on economic growth. Empirical Economics, 60, 1309–1334. https://doi.org/10.1007/s00181-019-01789-z

Mishra, P., Pandey, C. M., Singh, U., Gupta, A., Sahu, C., & Keshri, A. (2019). Descriptive statistics and normality tests for statistical data. Annals of Cardiac Anaesthesia, 22(1), 67–72. https://doi.org/10.4103/aca.ACA_157_18

Mohamed, R. A., Abd El-aziz, A. I., Ramadan, H. N., Abd El-sayed, M. H., & Emam, H. A. (2021). Impact of Human Capital on Economic Growth in Egypt: An ARDL Approach. European Journal of Economics, Finance and Administrative Sciences.

Mood, A. M. (1971). Partitioning Variance in Multiple Regression Analyses as a Tool For Developing Learning Models. American Educational Research Journal, 8(2), 191–202. https://doi.org/10.3102/00028312008002191

Murthy, K. (1994). Commonality analysis for the regression case. Texas A&M University

Nair, M., Pradhan, R. P., & Arvin, M. B. (2020). Endogenous dynamics between R&D, ICT and economic growth: Empirical evidence from the OECD countries. Technology in Society, 62. https://doi.org/10.1016/j.techsoc.2020.101315

Newton, R. G., & Spurrell, D. J. (1967). A Development of Multiple Regression for the Analysis of Routine Data. Journal of the Royal Statistical Society. Series C (Applied Statistics), 16(1), 51–64. https://doi.org/10.2307/2985237

Nguyen, C. P., & Doytch, N. (2022). The impact of ICT patents on economic growth: An international evidence. Telecommunications Policy, 46(5). https://doi.org/10.1016/j.telpol.2021.102291

Nimon, K., Lewis, M., Kane, R., & Haynes, R. M. (2008). An R package to compute commonality coefficients in the multiple regression case: An introduction to the package and a practical example. Behavior Research Methods, 40(2). https://doi.org/10.3758/BRM.40.2.457

Nimon, K., & Reio, T. G. (2011). Regression commonality analysis: A technique for quantitative theory building. Human Resource Development Review, 10(3), 329–340. https://doi.org/10.1177/1534484311411077

Nwaeze, N. C., Okere, K. I., Ogbodo, I., Muoneke, O. B., Ngini, I. N. S., & Ani, S. U. (2023). Dynamic linkages between tourism, economic growth, trade, energy demand and carbon emission: evidence from EU. Future Business Journal, 9. https://doi.org/10.1186/s43093-023-00193-5

Olaoye, I. J., Ayinde, O. E., Ajewole, O. O., & Adebisi, L. O. (2021). The role of research and development (R&D) expenditure and governance on economic growth in selected African countries. African Journal of Science, Technology, Innovation and Development, 13(6), 663–670. https://doi.org/10.1080/20421338.2020.1799300

Özdoğan Özbal, E. (2021). Dynamic effects of higher education expenditures on human capital and economic growth: an evaluation of OECD countries. Policy Reviews in Higher Education, 5(2), 174–196. https://doi.org/10.1080/23322969.2021.1893125

Pelinescu, E. (2015). The impact of human capital on economic growth. Procedia Economics and Finance , 22, 184–190. https://doi.org/10.1016/s2212-5671(15)00258-0

Pedhazur, E. (1997). Canonical and discriminant analysis, and multivariate analysis of variance. In *Multiple Regression in Behavior Research.* (3rd ed., pp. 924-979). Fort Worth, TX: Harcourt Brace

Phillips, P. C. B., & Perron, P. (1988). Testing for a unit root in time series regression. Biometrika, 75(2). https://doi.org/10.1093/biomet/75.2.335

Qayyum, A., & Zaman, K. (2019). Dynamic Linkages between International Trade, Gross Fixed Capital Formation, Total Labor Force and Economic Growth: Empirical Evidence from Pakistan. Economica, 15(1), 191–202.

Rahman, M. M., & Alam, K. (2021). Exploring the driving factors of economic growth in the world's largest economies. Heliyon, 7(5). https://doi.org/10.1016/j.heliyon.2021.e07109

Rahman, M. M., Rana, R. H., & Barua, S. (2019). The drivers of economic growth in South Asia: evidence from a dynamic system GMM approach. Journal of Economic Studies, 46(3), 564–577. https://doi.org/10.1108/JES-01-2018-0013

Rahman, M. M., Saidi, K., & Ben Mbarek, M. (2017). The effects of population growth, environmental quality and trade openness on economic growth: A panel data application. Journal of Economic Studies, 44(3), 456–474. https://doi.org/10.1108/JES-02-2016-0031

Ray-Mukherjee, J., Nimon, K., Mukherjee, S., Morris, D. W., Slotow, R., & Hamer, M. (2014). Using commonality analysis in multiple regressions: A tool to decompose regression effects in the face of multicollinearity. Methods in Ecology and Evolution, 5(4). https://doi.org/10.1111/2041-210X.12166

Romer, P. M. (1990). Endogenous Technological Change. Journal of Political Economy, 98(5), 71–102.

Rowell, R. K. (1991). Partitioning predicted variance into constituent parts: how to conduct commonality analysis. Texas A&M University.

Ruttan, V. W. (1998). The new growth theory and development economics: A survey. Journal of Development Studies, 35(2), 1–26. https://doi.org/10.1080/00220389808422562

Seibold, D. R., & McPhee, R. D. (1979). Commonality Analysis: A method for decomposing explained variance in multiple regression analyses. Human Communication Research, 5(4). https://doi.org/10.1111/j.1468-2958.1979.tb00649.x

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

Stock, J. H., & Watson, M. W. (2011). Introduction to Econometrics.

Teles, V. K. (2005). The role of human capital in economic growth. Applied Economics Letters, 12(9), 583–587. https://doi.org/10.1080/13504850500077013

Thaddeus, K. J., Ngong, C. A., Nebong, N. M., Akume, A. D., Eleazar, J. U., & Onwumere, J. U. J. (2024). Selected macroeconomic determinants and economic growth in Cameroon (1970–2018) "dead or alive" an ARDL approach. Journal of Business and Socio-Economic Development, 4(1), 1–19. https://doi.org/10.1108/jbsed-05-2021-0061

Thompson, B. (1985). Alternate methods for analyzing data from education experiments. Journal of Experimental Education.

Trejos, S., & Barboza, G. (2015). Dynamic estimation of the relationship between trade openness and output growth in Asia. Journal of Asian Economics, 36, 110–125. https://doi.org/10.1016/j.asieco.2014.10.001

Vu, K., Hanafizadeh, P., & Bohlin, E. (2020). ICT as a driver of economic growth: A survey of the literature and directions for future research. Telecommunications Policy, 44(2). https://doi.org/10.1016/J.TELPOL.2020.101922

Warne, R. T. (2011). Beyond multiple regression: Using commonality analysis to better understand r2 results. Gifted Child Quarterly, 55(4). https://doi.org/10.1177/0016986211422217

Wisler, C. E. (1968). On partitioning the explained variation in a regression analysis. National Center for Educational Statistics

World Bank. (2019). World Development Indicators. https://data.worldbank.org/

Zaman, M., Pinglu, C., Hussain, S. I., Ullah, A., & Qian, N. (2021). Does regional integration matter for sustainable economic growth? Fostering the role of FDI, trade openness, IT exports, and capital formation in BRI countries. Heliyon, 7(12). https://doi.org/10.1016/j.heliyon.2021.e08559

Zientek, L. R., & Thompson, B. (2006). Commonality analysis: Partitioning variance to facilitate better understanding of data. Journal of Early Intervention, 28(4), 299–307. https://doi.org/10.1177/105381510602800405

Zivot, E., & Andrews, D. W. K. (1992). Further evidence on the great crash, the oil-price shock, and the unit-root hypothesis. Journal of Business & Economic Statistics, 10(3), 251–270. https://doi.org/10.1080/07350015.1992.10509904

Appendix

Appendix A

Table A.1. Correlation matrix

	GDP	**GFCF**	**HC**	**Trade**
GDP	1			
GFCF	0.9865^{***}	1		
HC	0.9732^{***}	0.9409^{***}	1	
Trade	0.9342^{***}	0.8962^{***}	0.9561^{***}	1

Source: Author's output via RStudio using Penn World Table 10.01 and World DataBank data

(***) indicates statistical significance at the significance level of 1%

Appendix B

The results of the ADF test for the series in levels of GDP, GFCF, HC, and Trade are presented in Table B.1 below.

Table B.1. Test ADF for series in levels

H_0: series contains a unit root							
Model		Without drift and trend	With drift and without trend		With drift and trend		
		Tau 1	Tau 2	Phi 1	Tau3	Phi 2	Phi 3
GDP	Test value	2.5452	-0.2531	3.4392	-2.2029	4.2342	2.4442
GFCF		1.158	-1.099	1.6168	-2.7396	3.3098	3.7545
HC		-0.317	-4.0176	9.3413	-0.5117	5.5182	7.1614

Trade		0.6174	-1.6302	1.7077	-1.8379	1.6834	2.138
Critical values							
1%		-2.62	-3.58	7.06	-4.15	7.02	9.31
5%		-1.95	-2.93	4.86	-3.50	5.13	6.73
10%		-1.61	-2.60	3.94	-3.18	4.31	5.61

Source: Author's output via RStudio

Note: The critical values are those proposed by MacKinnon (1996) for the sample size of 50 and the length of the automatic lag was based on the Schwarz information criterion (SIC).

It is possible to conclude from Table B.1, that for the series in levels of GDP, GFCF, and Trade the nonstationary possibly comes from a stochastic trend, by the significant levels of 10%, 5%, and 1%. This conclusion is possible since the value of phi 2 is lower than the critical values, not rejecting the null hypothesis that there is no drift or deterministic trend. Additionally, to confirm the existence of a unit root of these tree series, we interpret, for each one, the value of the test from tau 1, which is higher than the critical values, for all significance levels, for this reason, the null hypothesis is not rejected for the existence of a unit root for GDP, GFCF, and Trade. For this reason, differencing once each series should be enough to make them stationary.

ADF test results for the HC series in levels reveal that, at the significance level of 1%, the regression to estimate is without drift and trend. At this significance level, the nonstationary comes from a stochastic trend. However, for higher significance levels, the nonstationary (proved by the value of tau 3) comes from a deterministic trend (the phi's values are higher than critical values of 5% and 10% significance levels). ADF test concludes that all series are nonstationary.

Table B.2. Test PP for series in levels

H_0: series contains a unit root					
Model		Without trend and short lag	Without trend and long lag	With trend and short lag	With trend and long lag
GDP	Test value	0.228	0.4567	-2.4777	-2.024
GFCF		-0.0919	0.2287	-2.3946	-1.8419
HC		-5.2051	-4.4133	-3.9854	-3.8782
Trade		-1.9051	-1.9291	-2.7531	-2.4328
Critical values					

1%	-3.568111	-4.154028
5%	-2.921459	-3.502455
10%	-2.598313	-3.180404

Source: Author’s output via RStudio

The PP test results (Table B.2) demonstrate that without trend and with both short and long lags the conclusions are the same: the series in levels of GDP, GFCF, and Trade are nonstationary for all significance levels. However, for the same model, but for HC series in levels, the test rejected the null hypothesis of the existence of a unit root for all significance levels, meaning the series is stationary.

To test the presence of a time trend, a new model was added, taking that account, for short and long lags. The findings show that the series in levels of GDP, GFCF, and Trade, did not reject the null hypothesis, for all significance levels. The same conclusion was retrieved for the HC series at the significance level of 1%, also with the trend and with short and long lags, however for higher significant levels, the series is stationary.

To conclude, apart from a few exceptions, the PP test results showed that all series, except HC, are nonstationary.

Table B.3. Test KPSS for series in levels

H_0: series do not contain a unit root			
Model		Short	Long
GDP	Test value	0.0973	0.0833
GFCF		0.071	0.0757
HC		0.232	0.1577
Trade		0.0644	0.0849
Critical values			
10%		0.119	
5%		0.146	
2.5%		0.176	
1%		0.216	

Source: Author’s output via RStudio

The KPSS test does not reject the null hypothesis (the test value is lower than the critical values) for series in levels of GDP, GFCF, and Trade considering both short and long lags and for all main significance levels. For HC series in levels, the test for short lags shows that is nonstationary while with long lags this also verifies for 5% and 10% of significance levels. The KPSS test concludes that, except HC series, the remaining are stationary.

Table B.4. Test ZA for series in levels

H_0: series contains a unit root			Potential break point at position
GDP	Test value	-6.2849	48
GFCF		-5.9035	48
HC		-75.4907	40
Trade		-3.3386	49
Critical values			
1%		-5.57	
5%		-5.08	
10%		-4.82	

Source: Author's output via RStudio

The test values of the ZA test are lower than the critical values, of all significant levels, for all series in levels except the Trade series. This means that all series except trade are stationary.

Table B.5. Test ADF after computing the first differences

H_0: series contains a unit root		
Model		Without drift and trend
$\Delta \log GDP$	Test value	-1.4024
$\Delta \log GFCF$		-2.6103
$\Delta \log HC$		-1.5269
$\Delta \log Trade$		-4.407
Critical values		
1%		-2.62
5%		-1.95
10%		-1.61

Source: Author's output via RStudio

Table B.6. Test PP after computing the first differences

H_0: series contains a unit root			
Model		Without trend and short lag	Without trend and long lag
$\Delta \log GDP$	Test value	-5.0927	-4.9855
$\Delta \log GFCF$		-4.3832	-4.1006
$\Delta \log HC$		-2.2839	-2.3021
$\Delta \log Trade$		-6.3021	-6.2985
Critical values			
1%		-3.571174	
5%		-2.92277	
10%		-2.599003	

Source: Author's output via RStudio

Table B.7. Test KPSS after computing the first differences

H_0: series do not contain a unit root

Model		Short	Long
$\Delta \log GDP$	Test value	0.0427	0.0954
$\Delta \log GFCF$		0.0399	0.0819
$\Delta \log HC$		0.1743	0.1161
$\Delta \log Trade$		0.0664	0.0946
Critical values			
10%		0.119	
5%		0.146	
2.5%		0.176	
1%		0.216	

Source: Author's output via RStudio

Table B.8. Test ZA after computing the first differences

H_0: series contains a unit root			Potential break point at position
$\Delta \log GDP$	Test value	-4.9238	29
$\Delta \log GFCF$		-7.3079	37
$\Delta \log HC$		-3127.991	35
$\Delta \log Trade$		-6.6464	45
Critical values			
1%		-5.57	
5%		-5.08	
10%		-4.82	

Source: Author's output via RStudio

Table B.9. Chow test to $\Delta \log \mathrm{HC}$ series

M-fluctuation test	
$\boldsymbol{f(\mathbf{efp}) = 1.0945}$	$\mathrm{p-value} = 0.5525$

Source: Author's output via RStudio

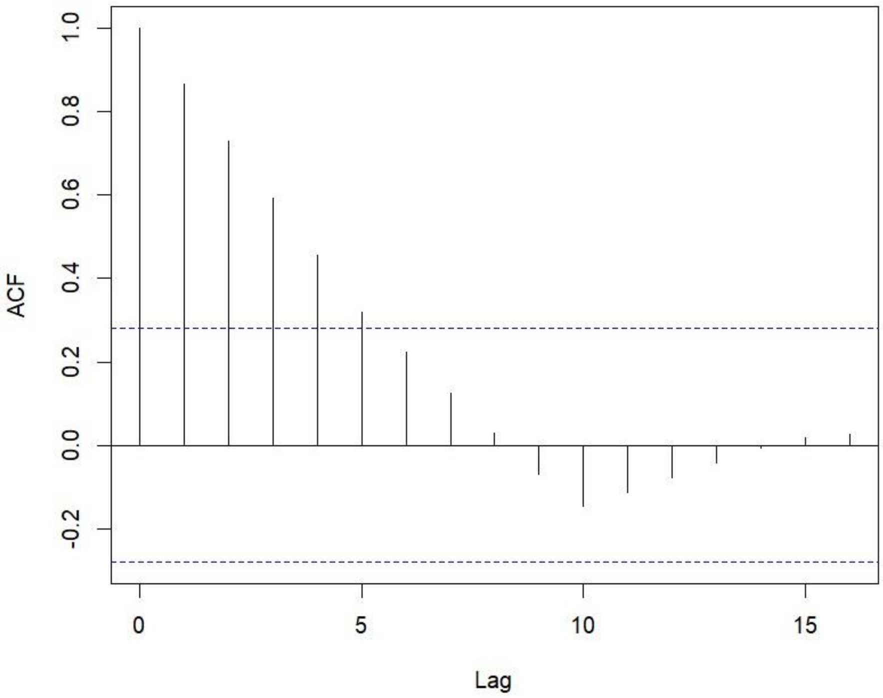

Figure B.1. Autocorrelation function of $\Delta \log HC$

Source: Graphic generated by RStudio

Appendix C

Table C.1. Formulas for unique and commonality components of variance for three independent variables

$$U(1) = R^2(123) - R^2(23)$$

$$U(2) = R^2(123) - R^2(13)$$

$$U(3) = R^2(123) - R^2(12)$$

$$\mathrm{C}(1,2) = R^2(13) + R^2(23) - R^2(3) - R^2(123)$$

$$\mathrm{C}(1,3) = R^2(12) + R^2(23) - R^2(2) - R^2(123)$$

$$\mathrm{C}(2,3) = R^2(12) + R^2(13) - R^2(1) - R^2(123)$$

$$\mathrm{C}(1,2,3) = R^2(1) + R^2(2) + R^2(3) - R^2(12) - R^2(13) - R^2(23) + R^2(123)$$

Source: Formula based on Seibold & McPhee (1979), with visual adaptation from Rowell (1991)

Table C.2. Correlation matrix after the appropriate transformations

	$\Delta\,\mathbf{log}\,\boldsymbol{GDP}$	$\Delta\,\mathbf{log}\,\boldsymbol{GFCF}$	$\Delta\,\mathbf{log}\,\boldsymbol{HC}$	$\Delta\,\mathbf{log}\,\boldsymbol{Trade}$
$\Delta \log GDP$	1			
$\Delta \log GFCF$	0.91186778	1		
$\Delta \log HC$	-0.02163631	-0.07501268	1	
$\Delta \log Trade$	0.28369269	0.25294067	0.37063824	1

Source: Author's output via RStudio

Table C.3. Variance Inflation Factor

$\Delta \log GFCF$	$\Delta \log HC$	$\Delta \log Trade$
1.107414	1.201634	1.276545

Source: Author's output via RStudio

Table C.4. Variance

$\Delta \log GFCF$	$\Delta \log HC$	$\Delta \log Trade$
0.002530218	7.893196e-06	0.004566679

Source: Author's output via RStudio