2006

Impact of the Investments in Information and Communication Technologies on Total Factor Productivity in the Context of the Economies in Transition

Sergey Samoilenko
Virginia Commonwealth University

IMPACT OF THE INVESTMENTS IN INFORMATION AND COMMUNICATION TECHNOLOGIES ON TOTAL FACTOR PRODUCTIVITY IN THE CONTEXT OF THE ECONOMIES IN TRANSITION

A Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Virginia Commonwealth University.

by

SERGEY SAMOILENKO
M.S. in Information Systems, Virginia Commonwealth University, 2002
B.S. in Industrial Engineering, Institute of Soviet Trade Technology, 1990

Directors:
Dr. Ojelanki Ngwenyama
Professor, Department of Information Systems
Dr. Kweku-Muata Osei-Bryson
Professor, Department of Information Systems

Virginia Commonwealth University
Richmond, Virginia
May, 2006

Acknowledgement

This section is turning out to be the most difficult part of my dissertation. It would be nice to run some sort of statistical analysis and show the significance of the obtained results, for I cannot adequately express my gratitude in words. I feel that even the best I can do shall trivialize the profound impact that my mentors and contributors made on my development. Nevertheless, following is an attempt to show my gratitude.

Dr. Ojelanki Ngwenyama is one of the major reasons for what I am now. He provided support and guidance, and continues to do so. He is a mentor, a friend, and an overall example to me of what a person should be. When I was in the dark, he was holding the light.

Dr. Kweku-Muata Osei-Bryson is another major reason that caused my transformation over the years into the person who was able to write this dissertation. He is a man of highest integrity, honor, but most of all, he is a man of kindness. Let the blind man see… It is because of him that I can do what I can do.

Dr. Ngwenyama and Dr. Osei-Bryson were co-chairs of my dissertation. They remain co-chairs of my development and future academic pursuits.

Their standards are high, and their demands are great. But so the rewards for those students who are willing and able to commit themselves to their leadership. This dissertation is the example of my reward, and let it serve as an example to the future students of Drs. Ngwenyama and Osei-Bryson. Looking back, all it took was to make myself available to what my mentors have to offer.

Dr. Richard Redmond made my student life feel like a vacation. He is the reason that I feel sorry for not being a student in the department of Information Systems anymore. When I was lost, he was always there to find me. When I needed an advice or help, I could always find it in his office. I should have made my appointments, though, instead of just barging in… Sorry about that, Rich…

Dr. Roland Weistroffer exhibited otherworldly tolerance while I was improving the quality of my dissertation drafts. He was supportive of my research and unforgiving of my presentation of it. Scores and scores of my "Yoda-style" sentences were exterminated with the vengeance by the red ink of his pen. If my dissertation is at all readable, it is because of Dr. Weistroffer. As usual, all the remaining errors are proudly mine…

I was very fortunate to have Dr. Jose Dulá on my dissertation committee. Not only did he put a human face on DEA for me, he also guided me through the process of learning and made sure that what I do makes sense. It is because of Dr. Dulá that I now share a special relationship with DEA.

Dr. Jacob Norbjerg has proven that one does not have to be near to make an impact. While being far away, he was able to transcend the distance by providing his generous appraisals, insightful comments and invaluable critique.

If my dissertation makes any sense from the point of view of economics, this is because Dr. Carol Lehr made sure of it. Dr. Larry Williams had his door always opened when I had any SEM-related questions. Invaluable help with PLS was provided by Dr. J. Hubona. The conversations about hypothetico-deductive logic with Dr. Allen Lee were

truly invaluable, and so the discussion about formative and reflective constructs with Dr. Peg Williams.

Finally, I would like to thank my wife Monica, who over many years supported me in all my endeavors, and who still loves me after all these years of me being a student. If this dissertation is an accomplishment, it is more Monica's accomplishment than it is mine.

Table of Contents

List of Tables

List of Figures

Abstract

IMPACT OF THE INVESTMENTS IN INFORMATION AND COMMUNICATION TECHNOLOGIES ON TOTAL FACTOR PRODUCTIVITY IN THE CONTEXT OF THE ECONOMIES IN TRANSITION

Sergey Samoilenko, MS

A Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Virginia Commonwealth University.

Virginia Commonwealth University, 2006

Major Directors:
Dr. Ojelanki Ngwenyama
Department of Information Systems,
Dr. Kweku-Muata Osei-Bryson
Department of Information Systems

The goal of this research is to establish a link between investments in *information and communication technology (ICT)* and economic growth in the context of countries that are currently classified by the international community as *transitional economies (TE)*. More specifically, in this study we focus on the relationship between ICT and one of the determinants of economic growth, *total factor productivity (TFP). Neoclassical growth*

accounting and the *theory of complementarity* provide the theoretical framework on which we build this research. By combining the data obtained from two sources, the World Bank Database and the IT Yearbook, we were able to construct a 10-year data set for 18 TEs spanning the period from 1993 to 2002.

Our inquiry is structured as a seven-step process that utilizes six data analytic methods. The first step in our investigation involves *Cluster analysis* (CA) with the purpose of determining whether or not the selected set of TEs is homogenous. Use of CA allowed us to identify two distinct groups of TEs in our sample, which suggests the heterogeneity of the sample.

In the second part of our inquiry, we employ *Decision Tree* (DT) analysis with the goal of investigating the differences between the clusters of TEs that were generated by the CA in the previous step. We were able to determine that one of the groups of TEs, the "leaders," appears to be wealthier than the other group, the "majority."

In the next step of our investigation, we perform *Data Envelopment Analysis* (DEA) to determine the efficiency of the TEs in our set. We were able to determine that the "leaders" are more efficient than the "majority" not only in terms of the production of the output, but also in terms of the utilization of the inputs.

The fourth part of our investigation takes advantage of the DT analysis with the purpose of obtaining the insights into the nature of the differences between the efficient and inefficient TEs. By incorporating the results of the CA into DT analysis we were able to construct the model that suggests, with the high degree of precision, some of the criteria according to which the efficient TEs differ from the inefficient ones.

The fifth stage of our investigation involves the use of the *Translog regression model* for the purpose of determining whether or not there exists a set of investments that are complementary to the investments in ICT. We have determined that there exists a statistically significant interaction effect between the investments in ICT and other variables, representing state of labor, as well as capital investments.

The sixth part of our investigation relies on using *Structural Equation Modeling* (SEM) implemented with *Partial Least Squares* (PLS) to test for the presence of the relationship between the investments in ICT and the unexplained part of the macroeconomic growth, TFP. We were able to establish the presence of the relationship between the two constructs of our conceptual model, "ICT Capitalization" and "TFP" for the "leaders" group of our sample. The construct "ICT Capitalization" was represented by the three ratio measures, all of which contain variable "Annual investment in telecom" in the denominator, while the Malmquist Index and its components, TC and EC., represented the construct "TFP." Thus, it allows us to state that we have established the presence of the relationship between the investments in ICT and TFP.

The last step of the data analysis involves using Classification DT and *Neural Network* (NN) analyses with the aim of investigating the reasons why some of the TEs exhibit statistically significant relationship between the investments in ICT and TFP, while other TEs do not. We were able to determine that one of the reasons why the "leaders" exhibit the statistically significant relationship between the investments in ICT and TFP is that they have higher level of inputs and more efficient processes of converting the inputs into the outputs than the "majority."

CHAPTER 1 Introduction

1.1 Connecting ICT and Economic Growth: Where It All Began and Why It Is Important

The goal of this research is to establish a link between the investments in Information and Communication Technology (ICT) and economic growth in the context of the set of the countries that are currently classified by the international community as 'Transitional Economies' or 'Economies in Transition'. The term "transition" in this context refers to transition from a command economy to a market economy (Piatkowski 2003).More specifically, in this study we shall focus on the relationship between ICT and one of the determinants of the economic growth, that of *Total Factor Productivity* (TFP).

According to *Neoclassical Growth accounting model* (also known as Solow's growth model), which provides theoretical framework for this study, overall economic growth could be broadly partitioned into the two components: first, explained growth and second, TFP, or unexplained growth. Explained growth comes from the growth in the inputs, such as increases in labor and the investments, while unexplained growth cannot be explained by the growth in inputs. Consequently, because TFP cannot be explained by the Neoclassical Growth model, it is often referred to as *Solow's residual*. Another term that is commonly used to refer to TFP is *Multi Factor Productivity*, or MFP.

In this light, this research aims to establish the relationship between the investments in ICT and the unexplained economic growth, namely, the subset of the economic growth that cannot be accounted for by any investments in ICT.

Broadly, the relevance and importance of the research question that we are trying to address in this paper could be perceived from three hierarchical perspectives. At the lowest level of granularity, we ask "How do investments in ICT affect TFP of the Transitional Economies (TE)?" At the second level, knowing, that TFP is one of the determinants of macro-economic growth, the question could take the form of "How do investments in ICT impact and contribute to the overall economic growth of TEs?"

We assume, together with the research and economic development community of the world, that the results of the overall economic growth in TEs would, at least partially, be directed at the reduction of poverty. Thus, finally, and we believe, most importantly, this research could be perceived as an important contribution to answering the following question: "How do investments in ICT, by affecting TFP, contribute to economic growth and, as a result, to poverty reduction in the Transitional Economies of the World?"

The original reason for linking investments in ICT and economic growth is the following. It is commonly agreed that there exists a set of factors that determine macro-economic growth, which could manifest itself in such a form as, let us take the most commonly used measure, an increase in Gross Domestic Product (GDP). A number of the developed countries (especially the US and some of the countries of the EU) have experienced a well-documented strong economic growth over an unusually long (early1990s-2000) period. Multiple studies, including research by the Organization for

Economic Development and Cooperation[1] (OECD 2001), attempted to identify the sources of the uncharacteristic 'growth spurt', as well as to uncover the reasons explaining why some countries have done better than others have. The modern economic theory helps in this inquiry, for it provides a few vantage points from which this research question could be approached and investigated. From the perspective of neoclassical growth accounting, for example, economic growth could come from two sources. First, the increase in GDP may be brought about by the increase in the available resources, i.e., labor and capital, as well as by the efficiency and effectiveness of labor and capital utilization. On the other hand, economic growth could be the outcome of the increase in TFP. The studies have concluded that none of the factors listed above (i.e., increase in TFP, or increase in availability and efficiency of labor and capital) appeared to be the most important. Instead, a new factor, associated with information and communication technologies, has been identified as driving the economic growth (Colecchia and Schreyer 2002).

[1] The OECD is a forum where the governments of 30 democracies work together to address the economic, social and environmental challenges of globalization.. The Organization provides a setting where governments can compare policy experiences, seek answers to common problems, identify good practice and work to co-ordinate domestic and international policies. The OECD member countries are: Australia, Austria, Belgium, Canada, the Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Japan, Korea, Luxembourg, Mexico, the Netherlands, New Zealand, Norway, Poland, Portugal, the Slovak Republic, Spain, Sweden, Switzerland, Turkey, the United Kingdom and the United States. The Commission of the European Communities takes part in the work of the OECD. (source: www.oecd.org)

1.2 ICT and Economic Growth in the Context of TE: Proposed Contribution

The research in the area of ICT-driven economic growth is not driven by any grand theory according to which the link between ICT and economic growth must exist. Quite contrary, the research is driven by the desire to build a theory that could explain what took place in some economies, or dismiss the observed relationship as a one-time fluke. All what is known at this point is that the relationship between investments in ICT and rapid economic growth has been observed over the period of time in one context. It is currently conjectured, that the relationship of the same sort could exist in a different setting, that of TEs. However, the research in this area is too scarce to answer this question conclusively. Thus, we intent to contribute to the existing body of knowledge in this area by analyzing the data that we obtained from public sources, and by determining, if in fact the relationship between ICT and economic growth, which have been observed in the setting of developed countries, can be observed in the setting of TEs. As we proceed, we attempt to identify a set of complimentary factors that may affect the strength of the relationship and offer the possible implications for policy-making.

There are no obvious reasons to suspect why ICT, investments in which can fuel the economic growth of the developed countries, may not produce the same result for the economies in transition. The plummeting cost of ICT, after all, makes the investments in ICT attractive for the policy makers and investors alike. However, in the absence of a sufficient number of empirical studies, researching the nature of the relationships between investments in ICT and economic growth in the context of TE, the two shall remain linked by not much else than the broad generalizations, some anecdotal evidence, and hope.

With this short preamble, we now offer a brief overview of, and an introduction to, the issues that we consider essential for understanding the problem this research is trying to address. In addition to its introductory role, the material provided below attempts to draw a broad picture, the purpose of which is to demonstrate how four different concepts, namely, the investments in ICT, TFP, economic growth, and poverty reduction, fit together.

1.3 Role of ICT in Economic Growth and Poverty Reduction: Current Perspectives

On September 8, 2000, the General Assembly of the United Nations (UN) adapted a UN Millennium Declaration containing Millennium Development Goals (MDG). One of the eight goals is directed at the eradication of extreme poverty and hunger in the world. The concept of poverty is commonly and understandably considered to be a multidimensional construct; thus, alleviation of poverty is not perceived by the global community as a problem with a single factor solution. Rather, poverty reduction calls for a multidimensional approach that incorporates a number of factors working together. Economic growth, commonly represented by the increase in Gross Domestic Product (GDP), is considered to be one of the key factors that could contribute to the reduction of poverty.

Granted, economic growth does not necessarily lead to 'pro-poor growth'[2], however, reductions in poverty are much more difficult to achieve without economic growth. Moreover, the World Bank's economists (Daly 2003) have documented strong correlation between national economic growth and poverty reduction. Studies by Adams (2002) in the context of 50 developing countries over the period from 1980s to 1990s, Ravallion (2001) in the context of 50 developing countries in 1990s , and Bruno et al. (1998) in the context of 44 developing countries over the period from 1981 to 1992, have found that regression analysis of poverty on growth yields a negative, statistically significant, relationship. This implies that lack of economic growth is associated with lack of reduction in poverty. Similarly, Christiansen, Demery and Paternostro (2002) used a sample of eight African countries to establish that economic growth is responsible for poverty reduction via growth in mean income. Overall, the current literature on economic growth and poverty suggests that there exists a consensus regarding the relationship between the two, namely, that economic growth actually reduces the incidence of poverty.

The Department for International Development (DFID)[3], in agreement with the consensus shared by the research' and economic development communities, identified four conditions that may contribute to the acceleration of the overall economic growth (OECD 2005b). One of these conditions is the creation of strong incentives for

[2] It is common to refer to the economic growth resulting in significant poverty reduction as 'pro-poor growth' (OECD 2005a)

[3] The Department for International Development (DFID) is the part of the UK Government that manages Britain's aid to poor countries and works to get rid of extreme poverty. (source: http://www.dfid.gov.uk/aboutdfid/)

investment. The role of investment is to promote growth and innovation, which, as a result, bring about an increase in productivity, thus contributing to overall economic growth (Foy 2003). The overall relationship between determining factors (including investment), economic growth, and poverty reduction is depicted in Figure 1(reproduced from OECD 2005b).

Figure 1 below illustrates that achievement of poverty reduction is dependent on the economic development of a country, and each of the determining factors must contribute to the overall economic growth expressed as an increase in GDP. The significance of this for investment is that in order for the investment incentives to exist, it must be demonstrated that investments do actually translate into an increase in GDP. As for the problem of poverty reduction, it is becoming a matter of distribution of the increasingly available resources that are brought about by economic growth.

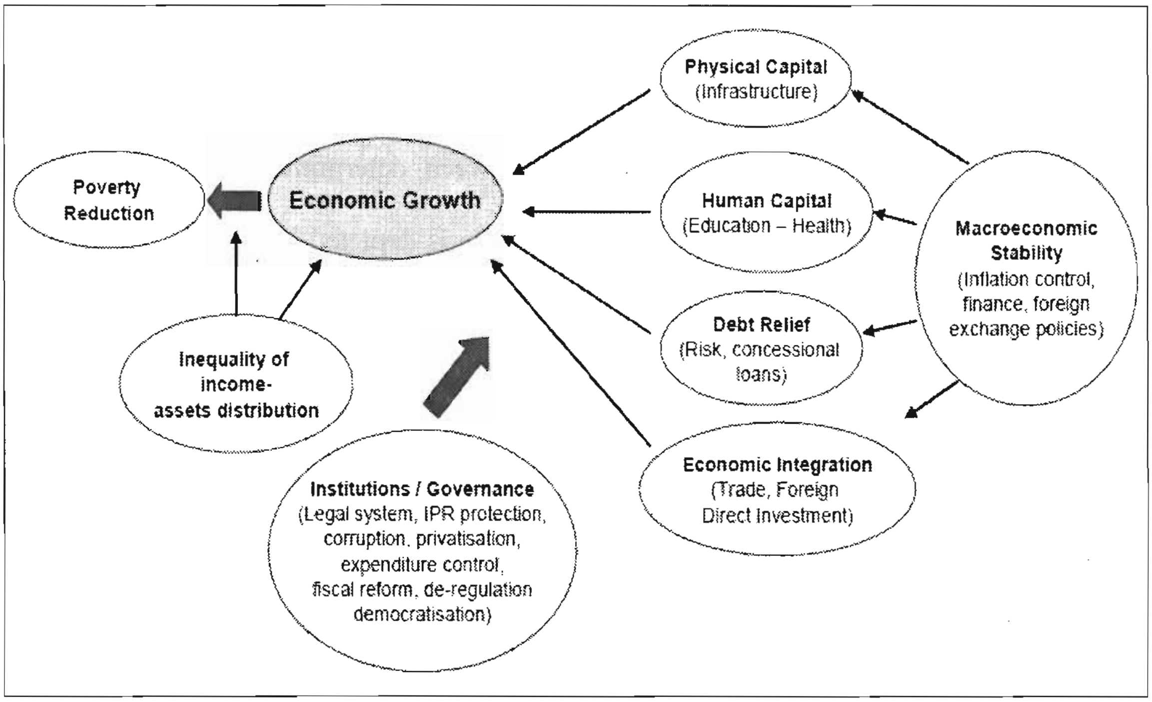


Figure 1 Economic growth and poverty reduction: determining factors

1.4 Linking ICT and Economic Development: Implications for the Poverty Reduction Policies

On November 20, 2001, the UN Secretary-General established, per request from the United Nations Economic and Social Council, an Information and Communication Technologies Task Force. The purpose of this initiative was to provide a global dimension to the efforts in bridging the 'digital divide[4]', to encourage digital opportunity,

[4] The digital divide is a social/political issue referring to the socio-economic gap between communities that have access to computers and the Internet and those who do not. The term also refers to gaps that exist between groups regarding their ability to use ICTs effectively, due to differing levels of literacy and technical skills, as well as the gap between those groups that have access to quality, useful digital content and those that do not (source: "http://en.wikipedia.org/wiki/Digital_divide")

and to place information and communication technologies (ICT) at the service of the development for all countries (Martinez-Frias 2003).

In the same year, the UN Commission on Science and Technology for Development (UNCTAD) selected "Technology development and capacity-building for competitiveness in digital society" as a theme for its 2001-2003 intercessional period. The particular attention within this theme was dedicated to the question of how the level of economic development of the developing countries could be increased by means of assimilation and application of ICTs.

The Development Assistance Committee (DAC)[5] of the Organization for Economic Co-operation and Development (OECD) suggests that in order to understand the potential contribution of ICT to pro-poor growth, it is necessary to understand the role of ICT in overall macro-economic growth, of which pro-poor growth is a subset. The DAC proposed the existence of the relationship between ICT and pro-poor economic growth, which is depicted by Figure 2 (reproduced from OECD 2005a).

If, indeed, as proposed, ICT does have a positive impact on the increase in GDP, than this contribution must be factored in in the general poverty reduction policies, and any good policy development depends on the understanding of the true nature of

[5] In order to achieve its aims the OECD has set up a number of specialized committees. One of these is the Development Assistance Committee, whose members have agreed to secure an expansion of aggregate volume of resources made available to developing countries and to improve their effectiveness. The members of the Development Assistance Committee are Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Japan, Luxembourg, the Netherlands, New Zealand, Norway, Portugal, Spain, Sweden, Switzerland, the United Kingdom, the United States and the Commission of the European Communities. (source: www.oecd.org).

economic impact of ICT, as well as the factors that might constrain or enhance it (OECD 2004).

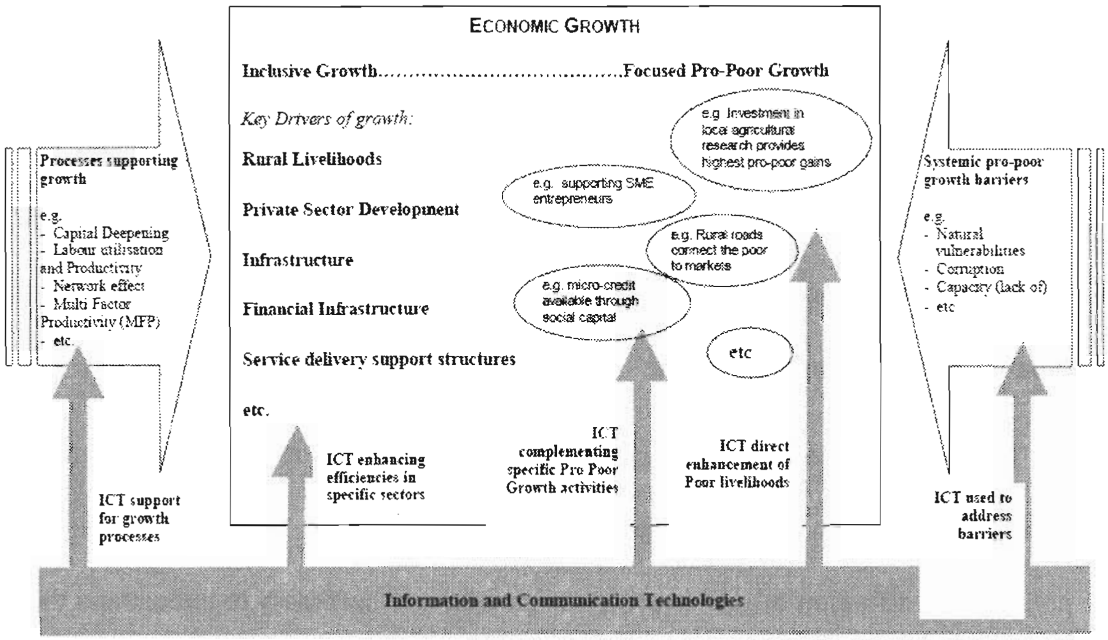


Figure 2 Contribution of ICTs to pro poor economic growth

1.5 Investments in ICT and Economic Growth: Is There Enough Evidence Linking the Two?

The need for the evidence of the relationships between ICT investments and economic growth, while evident in the majority of the economies of the world, is magnified in the developing countries, least developed countries and the economies in transition. This need for proof arises because many of such countries have not exhausted yet more traditional, straightforward growth reserves as, for example, privatization, institution building, elimination of loss-making state-owned enterprises, and others. As a result, investment in ICT might appear as a form of 'luxury', which has to compete with the more traditional, 'bread and butter' investments, such as those made in agriculture, infrastructure and industry. And unless ICT can be demonstrated to be a solid contributor to the country's economic well-being, it will be considered by many as unaffordable and impractical.

If, however, the link between ICT and economic growth is established, then ICT may be perceived as one of the tools for economic and overall development. And in this context, according to UN task force of ICT, it may provide support not just for the goal associated with pro-poverty development, but for all eight goals included in the UN Millennium Declaration (Martinez-Frias 2003). As a result, experiments and initiatives to use ICTs to combat poverty and promote economic growth and sustainable development in developing countries, economies in transition, and least developed countries, rooted in recognition that quickly accessed, properly adapted and broadly shared information and knowledge are key drivers of economic growth and social opportunity, has proliferated in recent years (OECD 2003c).

While still being short on the actual evidence, the latest research activities in this area provide some important insights into the possible relationship between macroeconomic growth and the investments in ICT. DAC suggests, for example, that in regard to the overall economic growth, contribution of ICT could come from the increases in efficiency and productivity, as well as from its positive impact on the sectors of economy that are acknowledged to be key drivers of overall growth (OECD 2005b).

In particular, because ICT products and services represent both inputs to the industries that use ICT products, and, at the same time, outputs from ICT-producing industries, Piatkowski (2004) suggests that ICT may contribute to the increase in GDP in the following four ways. First, production of ICT goods and services can contribute directly to the aggregate value added generated in an economy. Second, contribution to GDP could come from the increase in total factor productivity (TFP) of production in ICT sector, which contributes to aggregate TFP growth in an economy. Third, ICT capital could serve as an input in the production of non-ICT goods and services. Finally, increase in productivity in non-ICT producing sectors, induced by the production and use of ICT (so called 'spillover effects'), could contribute to economy-wide increase in TFP, and, as a result, to the increase in GDP.

In the context of OECD countries, investments in ICT could influence macroeconomic growth through the following four channels. First, the ICT investments could contribute to the technological innovation and high volumes of demand generated by an ICT production sector. Second, contribution could be made by means of "capital deepening," through the increased capital input per worker, which enables more efficient production that results, in turn, in increased labor productivity. Third, investments in ICT could contribute to TFP by means of increasing the efficiency in combining capital and labor. Finally, impact to the macroeconomic growth could be achieved through the increased ICT-based networking between firms, which reduces transaction costs and accelerates innovation (OECD 2004).

Investments in ICT can clearly facilitate macro-economic growth in the developed countries (OECD 2005a), and contribute to an acceleration in GDP and labor productivity rates (Piatkowski 2004), just as they did during the 1990s for the US (Jorgenson 2001; Jorgenson and Stiroh 2000; Oliner and Sichel 2002; Stiroh 2002) and number of countries of European Union (Colecchia and Schreyer 2001; van Ark *et al.* 2002; Daveri 2002; Jalava and Pohjola 2002).

Despite the ample evidence of the positive impact of ICT investments on the economies of the developed countries, the research concerning the effects of ICT on developing and transition economies is scarce. Among the few, IMF (2001) and Lee and Khatri (2003) reported on positive effect of ICT to economic growth in selected countries of South Asia, and Piatkowski (2003a, 2004) provided some estimates of ICT to growth and labor productivity in the Eastern European countries –former members of the Soviet block.

However, even if the research concerning the outcomes of ICT on economic development of developing countries produces evidence of a positive effect, the magnitude of the effect is incomparable with the effects of ICT on economic growth in developed countries. According to Pohjola (2001), the relative contribution of ICT to GDP growth in developing countries (China, India, Argentina, Chile, Brazil, Thailand and Venezuela) during the period between 1980 and 1995 was less than 2 percent, while for such developed countries as US, Finland, Canada, Sweden, and UK, contribution was more than 10 percent (Piatkowski 2003a).

It is important to note at this point, that when ICT-driven economic growth occurred in some of the developed countries, ICT deployment there was almost always accompanied by the complementary investments in, for example, organizational infrastructure (more on this topic and the relevance of it to this research later). The research and development community currently considers the absence or insufficient level of the complementary investments (not limited to the micro-economic level of the firm) to be one of the reasons why the impact of ICT investments does not manifest itself in the form of an improved productivity growth at the macroeconomic level (Pohjola 2002). Not much at this point is known about what type of complementary factors determine the successful translation of the ICT investments into productivity-driven growth in GDP.

1.6 Impact of ICT on Economic Growth: Mixed Evidence and Plenty of Suggestions

Thus, the actual impact of ICTs on macro-economic growth of developing, least developed countries and economies in transition is not yet clear. Some of the studies put in doubt the assumption that ICT investments in developing and least developed countries would result in an increase in the level of economic development (Avgerou 1998; Morales-Gomez and Melesse 1998). Kraemer and Dedrick (2001) state, quoting the result of a comprehensive cross-country empirical study on returns of IT investments in developed and developing countries by Dewan and Kraemer (2000), that the returns on the investments in ICT are "positive and significant for developed countries, but not statistically significant for developing countries" (p.262). Dewan and Kraemer (2000) suggest that this disparity may be the results of the differences in the capital stock and infrastructure, which, as the authors conclude, serve as prerequisites for ICT investments to be productive. So far, only the study by Pohjola (2001) was able to find that the ICT contributes to economic growth in developing countries. Kraemer and Dedrick (2001) suggest that one of the possible reasons for the lack of contribution of ICT investments to economic growth is the lack of complementary investments that are required in order for ICT to be properly utilized. Piatkowski (2002) concurs, in addition suggesting, in agreement with Murakami (1997), that size of ICT investments could be too small to bear upon any sizable economic growth.

There is no immediate reason why the economies in transition would not profit from the investments in ICT on the macroeconomic level. After all, to benefit from the growth impulses supplied by the use of ICT, as a capital input, a country does not have to have an ICT producing sector (Colecchia and Schreyer 2002). Nevertheless, at this point the conclusion whether or not ICT affects the overall economic growth in transition economies could not be made, for the existing empirical evidence is limited. Theoretically, however, the arguments linking ICT and economic growth have been presented (Madden and Savage 1998; Eggleston, Jensen and Zeckhauser 2002), and this theoretical support provides an additional incentive to conduct more studies in this area. But so far, as a result of the scarcity of the empirical evidence, the only option at this point for the development community is to assess, hypothetically, how the differences between the economies in transition and developed economies are likely to affect the impact of ICTs. And while the researchers in this area might disagree on whether or not ICT contributes to the increase in GDP, the consensus is that more research is needed because hardly any relevant studies have been done outside the OECD area (OECD 2004). Thus, "substantive research is urgently required if investment commitments are to be made – by the private sector or development agencies – with any real understanding of likely outcomes" (ibid p.4).

1.7 Impact of ICT on Economies in Transition: Building on the Existing Research

Any research dedicated to discovering the links (or lack thereof) between ICT and economic growth of TE could benefit, instead of starting from scratch, by taking into consideration the relevant findings that have been discovered by the OECD research in this area, while adjusting, however, for the intrinsic differences that exist between developed countries and the rest of the world. Thus, for example, OECD research demonstrated that a number of complementary factors in the business environment influence the extent of the impact of ICT on business performance. In the context of the developed countries, the following five factors are considered particularly important.

First factor refers to the nature of the business, i.e., whether or not the business can make more extensive use of ICTs to change processes and their relationships with customers and suppliers. The second factor refers to the extent of competition and the nature of the regulatory environment. In this regard, OECD research concluded that more competitive and less regulated business environments could take greater advantage of innovation that is provided by ICT, thus contributing to the macro-economic performance. The third factor reflects the relative costs of ICT deployment, which takes into consideration both direct and indirect costs. The fourth factor considers the amount and quality of human capital available, where a better skilled and equipped workforce is more likely to achieve higher rates of ICT-related innovation and increased productivity. Finally, the fifth factor considers the importance of the flexibility of the business environment, where capability for the restructuring and reorganization allows for better utilization of the opportunities offered by ICT (OECD 2004).

Granted, it is highly unlikely that the applicability of the complete set of complimentary factors can be directly transferred from OECD to the TE environment. However, the notion of complementary factors, which are required if we are to take advantage of what ICT has to offer to TEs, must be kept in mind with the purpose of discovering another set of factors, quite possibly unique ones, which may influence the impact of diffusion and use of ICT, affecting, therefore, the impact of ICT on business performance and overall economic growth.

There is, undoubtedly, a significant difference between the developed countries and the economies in transition. However, the presence of the differentiating factors does not indicate that the investments in ICT would not take place in the context of TEs. Rather, the differences suggest that the rate of the investments will probably be slower. Consequently, the effect of ICT on productivity and economic growth would likely manifest itself over a longer period of time. The economies in transition present a particularly interesting case for the research because they share economic characteristics with both developed and less developed economic regions (OECD 2004). As a result, transition economies may be perceived as a good vantage point from which the relationship between the investments in ICT and growth in TFP can be investigated.

On this somewhat positive note, we conclude the introduction and the general overview of the problem area to which this research is intended to contribute. Next two parts of the paper are dedicated to the overview of the theoretical framework used in this paper, as well as to a more detailed and rigorous definition of the research problem at hand.

CHAPTER 2 Overview of the Theoretical Framework

2.1 Growth Accounting

To approach our research problem we rely on a neoclassical framework of growth accounting. This framework originated from the work of Solow (1957) and since then has been used widely by other researchers (Oliner and Sichel 2002).

The objective of growth accounting is to decompose, using neoclassical production function, the rate of growth of the economy into the contributions from the different inputs. Neoclassical production function relates output and inputs in the following manner

$Y = f(A, K, L)$

Where

Y= output (most often in the form of GDP),

A= Total Factor Productivity (sometimes also referred to as the *level of technology*),

K= capital stock, and

L= quantity of labor, or size of labor force.

Based on the function shown above, growth accounting uses the Cobb-Douglas production function:

$Y = A * K^{\alpha} * L^{\beta}$

Where

α and β are constants determined by technology.

In the case of constant returns to scale[6], $\alpha + \beta = 1$, and thus $\beta = 1 - \alpha$, which gives the following formulation:

$Y = A * K^{\alpha} * L^{1-\alpha}$

It is important to note that this function does not necessarily represent the true relationships between the inputs and the output; rather, its purpose is simply to serve as a vehicle of the exploration and interpretation of the macroeconomic growth.

Out of the three inputs used by growth accounting, only capital, *K*, and labor, *L*, can be observed in the data, while TFP serves as a residual (often referred to as *Solow's residual*) term capturing that contribution to Y (GDP), which is left unexplained by the inputs of capital and labor. Thus, important difference between the three inputs used by the Cobb-Douglas production function is that two inputs (K and L)[7] are provided by the data, while the third input (A) is computationally derived. Therefore, the computed value of A would include not only all improvements that are derived from the utilization of the capital and labor, but any errors that have been made from estimation of the inputs of K and L as well.

---

[6] If $\alpha + \beta > 1$, returns are increasing to scale, and if $\alpha + \beta < 1$, returns are decreasing to scale

[7] In the context of this research K represents a combination of the capital investments in ICT that we denote K_{ICT}, and the other types of the investments that we designate as $K_{non\text{-}ICT}$. Similarly, L represents a total of ICT-related labor force, denoted by L_{ICT}, and the other types of the labor that we designate as $L_{non\text{-}ICT}$. Thus, Cobb-Douglas production function becomes $\boldsymbol{Y = A * K_{ICT}^{\alpha} * K_{non\text{-}ICT}^{\alpha} * L_{ICT}^{1-\alpha} * L_{non\text{-}ICT}^{1-\alpha}}$(more on that in the next section, where we formally define our problem).

If we are interested in changes in output at a period of time *t*, then formulation of the production function becomes

$$Y_t = A_t * K_t^{\alpha} * L_t^{1-\alpha}$$

Where

t = time period of interest. We could, having empirical data on Y_t, K_t, and L_t, express and compute changes in TFP as follows:

$$A_t = Y_t / (K_t^{\alpha} * L_t^{1-\alpha})$$

One of the appeals of using the neoclassical growth accounting framework lies in its simplicity; after all, only two factors, the TFP growth and rate of increase in inputs, are used to explain the growth rate of the output. As a result, this relationship reflects the fundamental assumptions of the framework, namely, presence of technological progress and the growth of labor. However, the flip side of the simplicity is somewhat limited explanatory capability of economic growth. For example, while assuming technological progress, the framework neither explains the sources of the progress or the factors that affect the progress, nor does it account for any possible interactions between the technological progress and capital growth. In reality, though, capital investments would be affected by the technological progress, for the progress in information technologies has fueled capital investments in the economies of the US and the other developed countries.

Finally, according to another assumption of the growth accounting framework, that the capital is subject to the law of diminishing returns, convergence of the poor and wealthy economies must take place. The reality, however, reflects that the gap between poor and rich countries of the world is widening.

Nevertheless, the use of the growth accounting framework for the purposes of researching contributions of ICT investments to macroeconomic growth of TEs appears to be warranted, for this analytical framework has been used widely to estimate contribution of ICT to economic growth in the context of developed and developing countries (Oliner and Sichel 2000; Schreyer 2000; Davery, 2000; Jorgenson and Stiroh 2000; Whelan 1999; Hernando and Nunez 2002).

Tuomi (2004) notes, however, that while the reliance on neoclassical framework of the most influential studies on ICT productivity and growth allows taking advantage of its theoretical assumptions[8], in the ICT-related and knowledge-intensive sectors of economy, such assumptions are not easy to justify empirically. Nevertheless, despite presenting the argument that in the strict sense the common interpretation of the rapid growth in TFP as being ICT-driven is not in complete agreement with the neoclassical framework's exogenous nature of TFP, Tuomi (2004) agrees that it would be natural to expect that ICTs would reveal their productivity impact on the overall economic efficiency and become visible in TFP. And, indeed, Van Ark et al. (2002) have demonstrated that the extensive investment in ICTs is associated with the improvements in total factor productivity.

[8] Such as constant returns to scale of production, perfect allocation of resources and investments, and competitive markets, where suppliers cannot influence prices (Tuomi, 2004)

Rarely, however, should any type of an investment made on the macroeconomic level be perceived and considered in isolation. Rather, it is more beneficial to search and identify a set of complementary factors that allow magnifying the potential benefits of the investment. Often, such complementary factor, or a set of factors, is represented by other types of investments, sometimes in entirely different areas of the economy. In this paper we use Theory of Complementarity as a theoretical framework that supports our search for the complementary to ICT investments. A brief overview of Theory of Complementarity is offered next.

2.2 Theory of Complementarity

Initially introduced in economics by Edgeworth (1881), the concept of Complementarity refers to the notion that the increase in one factor could result in the increased benefit received from its complementary factors. We apply theory of complementarity to our research problem in order to argue that if the benefits of the investments in ICT are to be reaped successfully at the macroeconomic level, then such investments could not be made in isolation from the investments in other areas. Thus, if the two investments, one in ICT and another one elsewhere, are more effective when taken jointly, rather than separately, we consider such investments complementary.

Complementarity of investments have been investigated in the context of the Research and Development (R&D) portfolios by Lambertini (2003), Lin and Saggi

(2002), Rosenkranz (2003) and in the context of process and product innovation by Athey and Schmutzler (1995). In a context more relevant to this research, Giuri, Torrisi, and Zinovyeva (2005) explored the complementarity between skills, organizational change, and investments in information and ICT; Bugamelli and Pagano (2004) studied the complementarity between investment in ICT and the related investment in human and organizational capital; Gera and Wulong (2004) examined complementarity of the investment in ICT and organizational changes and worker skills, and Loukis and Sapounas (2004) inquired into the complementarity between IS investment and the set of IS management factors.

However, even if the complementarity of the investments exists within a given production function, it could not be identified through the formulation offered by Cobb-Douglas production function. Complementarity of the investments could only be discerned if the formulation allows for the presence of the interaction term between the specified investments. Thus, we turn our attention to the transcendental logarithmic production function, a brief overview of which is offered next.

2.3 Transcendental Logarithmic (Translog) Production Function

The standard Cobb-Douglas production function, as it was mentioned before, has the form of

$Y = A * K^{\alpha} * L^{\beta}$

By taking the logarithm, the following formulation is obtained:

$\log Y = \log A + \alpha \log K + \beta \log L$

An extension to the given above formulation of the Cobb-Douglas production function, which has a more general and flexible functional form, was proposed by Christensen, Jorgenson and Lau (1973) and is called the *transcendental logarithmic (translog) production function*. The formulation of the translog is:

$\log Y = \alpha_0 + \alpha_k \log K + \alpha_L \log L + \frac{1}{2} \alpha_{k\,k} (\log K)^2 + \frac{1}{2} \alpha_{LL} (\log L)^2 + \alpha_{kL} \log L * \log K$

(Christensen, Jorgenson and Lau 1973)

Thus, the following general forms for the Cobb-Douglas production function and the translog production function may be offered, given $Y = f(A, K, L)$ and considering that A is a residual that can be expressed as an error term "*e*"

Cobb-Douglas Production Function:

$\log Y = \beta_0 + \beta_1 * \log K + \beta_2 * \log L + e$

Translog Production Function:

$\log Y = \beta_0 + \beta_1 * \log K + \beta_2 * \log L + \beta_3 * \log K^2 + \beta_4 * \log L^2 + \beta_5 * \log K * \log L + e$

It is easy to see that the Cobb-Douglas function is 'nested' in translog function, and test of the hypothesis that both functions describe production process equally well would entail testing the following null hypothesis:

H_0: $\beta_3 = \beta_4 = \beta_5 = 0$

Translog production function is more flexible than Cobb-Douglas function in the sense that it allows testing for the presence of the interactions between the variables. For example, let us say that we are interested in investigating the following production function:

$Y = f(A, K_{ICT}, K_{non\text{-}ICT}, L)$,

Where

Y= output (GDP),

A= the level of technology/ Total Factor Productivity,

K_{ICT} = investments in Information and Communication Technology

$K_{non\text{-}ICT}$ = investments in non-Information and Communication Technology

L= quantity of labor/size of labor force.

Then we could test for the presence of the statistically discernible interaction between the two types of the investments (ICT and non-ICT) using translog function as follows (again, keeping in mind that A is a residual expressed as error term "*e*"):

$$\log Y = \beta_0 + \beta_1 * \log K_{ICT} + \beta_{2*} K_{non\text{-}ICT} + \beta_{3*} \log L + \beta_{4*} \log K_{ICT}{}^2 +$$

$$+\beta_{5*} \log K_{non\text{-}ICT}{}^2 + \beta_{6*} \log L^2 + \beta_{7*} \log K_{ICT} * \log K_{non\text{-}ICT} + \beta_{8*} \log K_{ICT} * \log L + \beta_{9*} \log K_{non\text{-}ICT} * \log L + e$$

Then the test for the presence of the interaction would involve testing of the following hypothesis:

H0: β_7 is not statistically discernible from 0 at the given level of α

And, again, in the case of $\beta_4 = \beta_5 = \beta_6 = \beta_7 = \beta_8 = \beta_9 = 0$ translog production function would end up being formulated as corresponding Cobb-Douglas production function:

$\log Y = \beta_0 + \beta_1 * \log K_{ICT} + \beta_2 * K_{non\text{-}ICT} + \beta_3 * \log L + e$

Consequently, if the interaction term between investments in ICT and non-ICT is significant (i.e., we reject the null hypothesis of $\beta_7 = 0$), then we have a reason to assume that such investments are complementary.

CHAPTER 3 Formal Definition of the Research Problem

At this point, we formulate the following two broad research questions that we are addressing in this study:

The first objective of this research is to establish the presence of the relationship between investments in ICT (I_{ICT}) and Total Factor Productivity within the context of Transitional Economies (TE).

Let us consider the production function used in this study:

$Y = f(A, K_{ICT}, K_{non\text{-}ICT}, L)$

where

Y = output (GDP),

A = Total Factor Productivity(the level of technology),

K_{ICT} = investments in Information and Communication Technology

$K_{non\text{-}ICT}$ = investments in non-Information and Communication Technology

L = quantity of labor, or size of labor force.

Keeping in mind that the Cobb-Douglas function is nested in translog and equal to translog function under the condition of

$\beta_4 = \beta_5 = \beta_6 = \beta_7 = \beta_8 = 0$, translog function yields following formulation:

$$\log Y = \beta_0 + \beta_1 {*} \log K_{ICT} + \beta_2 {*}\, K_{non\text{-}ICT} + \beta_3 {*} \log L +$$

$$+ \beta_4 {*} \log K_{ICT}{}^2 + \beta_5 {*} \log K_{non\text{-}ICT}{}^2 + \beta_5 {*} \log L^2 +$$

$$+ \beta_6 {*} \log K_{ICT} * \log K_{non\text{-}ICT} + \beta_7 {*} \log K_{ICT} * \log L + \beta_8 {*} \log K_{non\text{-}ICT} * \log L + e$$

Thus, the null hypothesis corresponding to the first research question can be stated as follows:

$H1_0$: For a given sample of TEs, there exists no statistically discernible relationship between the capital investments in ICT (I_{ICT}) and Total Factor Productivity.

It is tempting to propose that in the case of the correlation we are going to test the following hypothesis:

*$H1.1_0$: correlation between the investments in ICT (**K_{ICT}**) and the Total Factor Productivity (**A**) is not statistically significant*

Which in the case of translog function amounts to testing whether or not error term e and K_{ICT} are correlated to a statistically discernible degree.

While in the case of the testing for the presence of the causal relationship we would test the relationship directly, as shown below.

$A = F(K_{ICT})$;

Which, in the case of using linear regression, takes the form of

$A = \gamma_0 + \gamma_1 * K_{ICT} + u$,

and involves testing whether or not $\gamma_1 = 0$.

However, there are two reasons why such approach of using direct correlation of regression cannot be used. First, from the perspective of economics, TFP is exogenous to Solow's production function, and, as such, it cannot be explained by the variables endogenous to the function such as labor or capital. Second, from the perspective of statistics, one of the fundamental assumptions of regression analysis is that an error term of the regression is not correlated with the independent variables. If we attempt to regress an independent variable against the error term, then we do so in violation of the fundamental assumption that deems our results pretty much useless.

Thus, we have chosen a different approach. We hypothesize that the relationship between the investments in ICT and TFP is indirect, mediated by some latent variables, let us call them *ICT Capitalization* and *TFP*. Then the relationship between the investments in ICT and TFP, mediated by ICT Capitalization can be depicted as shown in Figure 3 below.

According to the proposed model, investments in ICT and TFP serve as the indicators reflecting the latent constructs *ICT Capitalization* and *TFP*, respectively. We propose that there exists the actual direct relationship between *ICT Capitalization* and

TFP, which, in the case if the indicators indeed accurately represent the latent variables, implies indirect relationship between the indicators. In order to validate the proposed model, we would use structural equation model to test the hypothesized relationship.

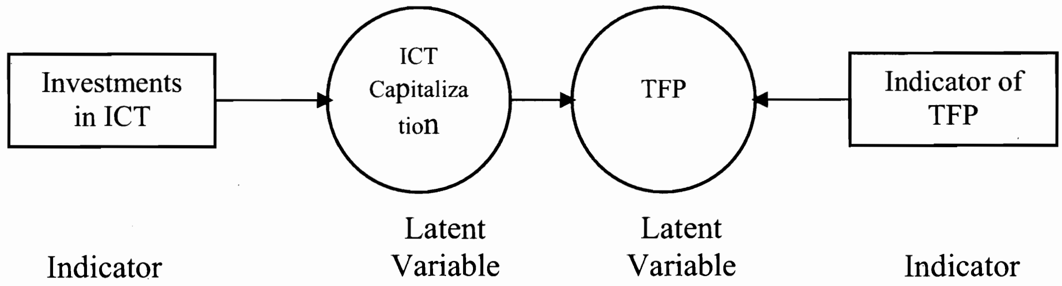


Figure 3 Mediated relationship between the investments in ICT and TFP

Based on the propose model, the hypothesis

$H1.2_0$: there is no statistically discernible relationship between the investments in ICT (expressed as $\boldsymbol{K_{ICT}}$ in Cobb-Douglas production function) and the Total Factor Productivity (expressed as **A** in Cobb-Douglas)

could be tested based on the results of the measurement and structural models of SEM.

The second objective of this research is to establish whether or not there exists a non-empty subset CCI of the set of total capital investments TCI, the presence of which affect the relationship between investments in ICT and Total Factor Productivity.

We denote a hypothetical set of complimentary to ICT capital investments as K_{comp}.

Thus, the null hypothesis corresponding to the second research question can be stated as follows:

$H2_0$: For a given sample of TEs, there exists no subset of complimentary investments K_{comp} presence of which produces a statistically discernible change in relationship between investments in ICT (I_{ICT}) and Total Factor Productivity

Following is the formulation of the Translog production function:

$Y = f(A, K_{ICT}, K_{other}, L)$,

Where

Y= output (GDP),

A= the level of technology/ Total Factor Productivity,

K_{ICT} = investments in Information and Communication Technology

K_{other} = investments in other areas then Information and Communication Technology

L= quantity of labor/size of labor force.

Then we can test for the presence of the statistically discernible interaction between the two types of the investments (ICT and non-ICT) using translog function as follows (again, keeping in mind that A is a residual expressed as the error term "*e*"):

$$\log Y = \beta_0 + \beta_1 * \log K_{ICT} + \beta_2 * K_{other} + \beta_3 * \log L + \beta_4 * \log K_{ICT}^{\ 2} + \beta_5 * \log K_{other}^{\ 2} + \beta_5 * \log L^2 + \beta_6 * \log K_{ICT} * \log K_{other} + \beta_7 * \log K_{ICT} * \log L + \beta_8 * \log K_{other} * \log L + e.$$

Then the test for the presence of the interaction would involve testing of the following hypothesis:

$H2.1_0$: β_6 is not statistically discernible from 0 at the given level of α

If the interaction term turns out to be statistically significant, allowing us to reject $H2.1_0$, we are able to state that we can establish the presence of complementary investments.

The purpose of the next part of the paper is to provide a review of the relevant research literature.

CHAPTER 4 Review of the Literature

The body of the research concerning the topic of economic impact of the investments in ICT could be presented as coming in two waves. The first wave of the studies, conducted in the period from mid 1980s to early 1990s, brought about very little empirical evidence of contribution of the investments in ICT to economic outcomes (Roach 1987, 1989, 1991; Strassman 1990; Yosri 1992; Loveman 1994; Oliner and Sichel 1994; Jorgenson and Stiroh, 1995; Hitt and Brynjolfsson 1996; Rai, Patnayakuni R. and Patnayakuni N. 1996, 1997; Strassman 1997). Despite the contradicting the theoretical arguments empirical results, the issue was not closed. Brynjolfsson (1993) warned that researchers "must be careful not to overinterpret these findings; a shortfall of evidence is not necessarily evidence of a shortfall" (p.67).

The desired evidence, in the form of empirical findings of positive and statistically significant relationship between investments in ICT and economic growth, has been provided by the second wave of the studies in this area, spanning the period from mid 1990s on (Lichtenberg 1995; Brynjolfsson and Hitt 1996; Gurbaxani , Melville and Kraemer 1998; Lehr and Lichtenberg 1999; Gilchrist, Gurbaxani and Towne 2001; Devaraj and Kohli 2003). The disparity between the results of the earlier and later studies has been attributed to various factors (Brynjolfsson 1993).

It would appear, according to common consensus, that there exist two factors that affect the relationship between investments in ICT and economic outcomes (Dewan and Kraemer 1998, 2000). The first factor is the size of the ICT investment and overall accumulated ICT capital, which should be above a certain threshold to manifest itself at the detectable level (Oliner and Sichel 2000; Jorgenson 2001; Jorgenson and Stiroh 2000; Council of Economic Advisors 2001). The second factor is the presence of the complimentary to ICT investments that are able to significantly increase the benefits provided by ICT (Black and Lynch 1997; Francalanci and Galal 1998; Tallon, Kraemer and Gurbaxani 2000; Devaraj and Kohli 2000; Brynjolfsson and Hitt 2000; Brynjolfsson, Hitt and Yang 2000; Ramirez 2003; Arvanitis 2003; OECD 2003).

Overall, research concerning the questions of the economic impact of the investments in ICT could be undertaken at different levels of analysis, such as firm, industry, sector, or country level. Because the unit of analysis in this study is country, we offer a review of only that subset of the 'second wave' of the available literature, which offers a country-level analysis of the impact of the investments in ICT on economic growth.

Kraemer and Dedrick (1994) examined the data from 11 Asia-Pacific countries over the period 1983-1990 and found a positive correlation between ICT investments and economic growth; the authors concluded that the results of the study could be interpreted as an evidence of ICT-driven growth.

Bryjolfsson and Yang (1996) conducted a literature review of more than 150 articles, published over the period from 1980s to 1995, which inquired into the relationship between information technology and productivity. The authors conclude, that the weight of evidence suggests that information technology capital generates its contribution for the US economy in terms of output growth and that without IT the economy would probably be in a worse situation (Brynjolfsson and Yang 1996).

Roller and Waverman (1996) have found statistically significant positive cause-and-effect type relationship between telecommunications infrastructure and economic growth for the sample of 21 developed and 35 developing countries over the period 1970-1990.

Murakami (1997) looked at the relationship between ICT and macroeconomic growth (among other things) in Japan and suggested ,based on the results of the analysis, that demand for ICT services, especially new services, may have a strong positive effect on the growth of the economy.

Motohashi (1997), on the other hand, found that the correlation between productivity and the IT intensity of the capital stock is weak in the context of five developed countries.

Study by Wong (1998) estimated the impact of ICT investments on the overall productivity of the Singapore economy for the period 1977 to 1997; results of the study indicated favorable productivity payoff from ICT investments, for the rate of generated return exceeded significantly the return to non-ICT capital.

Study by Canning and Pedroni (1999) found positive relationship between growth in telephones and paved roads per worker and economic growth; study used data over the period 1950-1992 and the sample of 150 countries.

Dewan and Kraemer (2000) used data pooled from 22 developed and 16 developing countries, over the period 1985-1993. The results of the analysis revealed that for developed countries returns from IT capital investments are positive and statistically significant. As for the developing countries of the sample, non-IT capital investments have shown also to be quite productive, while the evidence on IT investments was inconclusive.

Results of the study by Pohjola (2000), which analyzed the data from 39 developed and developing countries over the period 1980-1995, confirmed the conclusion of Dewan and Kraemer's (2000) that while ICT plays a significant role in macroeconomic growth of developed countries, similar contribution of ICT in the context of developing countries is yet to be made. The author suggests that it might simply reflect the fact that the developed countries lack IT-enhancing complimentary factors that enhance and amplify the effects of the ICT investments (Pohjola 2000).

Cette, Mairesse and Kocoglu (2001) attempted to quantify the contribution of the ICT to the economic growth in France. According to the authors' estimates total contribution of ICT to GDP growth amounted to approximately 0.2% per year over the period 1969-1999. Moreover, Cette, Mairesse and Kocoglu (2001) concluded that over the last years (1995-1999) the contribution of ICT to economic growth has gradually increased, reaching approximately 0.3% per year.

Research by Lee and Khatri (2001), among the only few studies that inquire into the impact of ICT on developing economies, reports the evidence of positive contribution of ICT to economic growth of the countries of South-East Asia. Similar results for Malaysia, Philippines, Thailand, South Korea, and Taiwan are reported by a study by the International Monetary Fund (IMF 2001).

Niininen (2001), in the macroeconomic study of Finland, first in this area of research, demonstrated that in Finland, ICT contributed to the growth in GDP at the average annual rate of 2.4 % in the period from 1983 to 1996.

Somewhat unexpectedly, the study of Pohjola (2002) does not find any significant correlation between ICT investment and economic growth in the period 1985 to 1999 for a sample of 42 countries; moreover, in contrast with some previous studies, the relationship is not statistically significant for the sub samples of industrial or high-income countries either. The author suggests three possible explanations, first, low level of investment in ICT, second, lack of complimentary infrastructure and, third, inability of neoclassical method to assess the benefits provided by investments in ICT.

Daveri (2002) inquired into the impact of ICT investments on economic growth of the countries of the EU, over the period 1992 to 2001. According to the results of the study growth contributions from ICT capital rose in six EU countries (UK, Denmark, Finland, Sweden, Ireland and Greece) over the period from 1996 to 2001. However, unlike in the United States, in the countries of the EU (exceptions being Ireland and Greece) this growth has not been associated with growth in the total factor productivity rates (Daveri 2002).

Kegels, van Overbeke and Van Zandweghe (2002) inquired into the contribution of the investments in ICT to macroeconomic growth in Belgium over the period from 1995 to 2000. Based on the results of the analysis of the preliminary data the authors conclude that while Belgium does not present the profile of the advanced new ICT-driven economy in Europe, the growth contribution of ICT capital accelerated in the late 1990s, thus putting Belgium slightly above the EU average (Kegels et al. 2002).

The paper by Baliamoune (2002) described a study examining the relationship between ICT diffusion and a set of macroeconomic policy variables for a sample of 47 developing countries. According to the results of the study, there is evidence that ICT enhances income and can provide an additional source of economic growth (Baliamoune 2002).

Colecchia and Schreyer (2002) have examined the contribution of ICT capital to economic growth in nine developed (members of OECD) countries, concluding that the US, while having benefited the most, has not been alone in taking advantage from ICT-driven economic growth. This study (Colecchia and Schreyer 2002) updates and extends the previous work by Colecchia (2001) and Schreyer (2000) on the contribution to ICT to output growth in G7 countries.

Van Ark et al. (2002), inquiring into the relationship between ICT investments and economic growth of the countries of the European Union over the period 1980-2000, found evidence of the contribution of the investments in ICT to economic growth and arrived to a similar conclusions as Colecchia and Schreyer (2002).

A study by Jalava and Pohjola (2002), which aimed to update and extend findings of Niininen (2001), confirmed that ICT has been the factor behind the improved economic performance of the US in the 1990s and estimated that in Finland the contribution of ICT to economic output has more than doubled over the period form early to late 1990s. The results of the research are in line with earlier similar findings reported by Oliner and Sichel (2000) and Jorgenson (2001), who have presented evidence of doubling of the ICT contribution to the economic growth of the US.

Research by Vijselaar and Albers (2002) inquired into the importance of ICT for the growth in productivity in the euro area (such as France, Germany, Italy, Netherlands, among others) over the period from the mid-1990s to 2000. Based on the available data the authors concluded that while there is evidence of an increased contribution of ICT to economic growth both in terms of production and in terms of investment, there is little evidence of significant positive spillover effects from the use of ICT to overall productivity growth (Vijselaar and Albers 2002).

A comprehensive analysis of ICT impact on productivity was performed by OECD (2003). According to the results of the study, which focused on developed countries, the strong evidence is present allowing identifying ICT as one of the key drivers of economic growth that took place in 1995-1999 in the USA, Canada, Finland, Netherlands, and other countries. The study identified the USA and Finland as the two countries that benefited the most from the ICT-driven growth.

Melka, Nayman, Zignago, and Mulder (2003) assessed the contribution of ICT to the macro economic growth in France over the period 1982-2000. The authors concluded that the contribution of ICT to the French growth is comparable to the contribution of ICT to British growth, but it is far behind that of the United States. Melka et al. (2003) suggest that one of the reasons why France lags behind the U.S. in terms of ICT contribution to growth is that the proportion of US ICT investment in total investments is more than twice as high as the French share.

Baliamoune-Lutz (2003) examined the relationships between ICT diffusion and a set of macroeconomic and policy variables in the context of 47 developing countries, using data over the period from 1998 to 2000; the author reports that the results of the study indicate that ICT dissemination fosters economic development.

Piatkowski (2003a) inquired into the contribution of investments in ICT to output growth and productivity in Poland (country of transition economy). According to the results of the study, contribution of ICT to economic growth in Poland is comparable to the countries of European Union and the US.

In the follow-up, Piatkowski (2003b) extended his study of Poland (Piatkowski 2003a) to eight transition economies of Europe (Bulgaria, Czech Republic, Hungary, Poland, Russia, Slovakia, and Slovenia). According to the results of the study, the contribution of investments in ICT between 1995 and 2000 in most of the featured in the study countries was “much higher than what might be expected on the basis of the level of their GDP per capita” (Piatkowski 2003b).

Dedrick, Gurbaxani and Kraemer (2003) critically reviewed more than 50 articles on computers and productivity that have appeared between 1985 and 2002. The conclusion of the review was that greater investment in IT is associated with greater productivity growth, for "recent studies have found that IT investments have had a major impact on labor productivity and economic growth at the country level."

Van Ark and Piatkowski (2004) investigated and compared the productivity performance of the 10 countries of Central and Eastern Europe (CEE-10) with the 15 countries of European Union (EU-15) during the 1990s. According to the results of the study, ICT capital was an important source of growth during the 1990s, and in the context of the CEE-10, it has contributed as much to labor productivity growth as for the EU-15.

In the conclusion of our review of the literature on the subject, we would like to offer a brief summary. It would appear that, indeed, the majority of the research power has been concentrated on studying the relationship between ICT and macroeconomic growth only in the context of the developed countries of the world. According to the results, there exists a relatively small group of developed countries (US, Finland, UK, and others) in which investments in ICTs have clearly paid off in terms of the economic growth due to improvements in TFP. The second, much larger group of the developed countries, while not achieving the same level of success, still did get the positive results reflected in small increase in TFP. Finally, the third group of the developed countries, representing the rest of the developed world, appears to be unable to show any ICT-related improvements in TFP.

The body of the research dedicated to the same topic, but in the context of developing countries and transition economies is much smaller. According to the scarce results, there is a relatively much smaller group of developing economies and economies in transition that demonstrated significant positive relationship between ICT investment and the growth in TFP, while the rest, literally, 'have nothing to show'. At this point in time, however, the overall empirical evidence is inconclusive for making any decisions regarding the impact of the investments in ICT. Thus, further research is required in order strengthen the foundation based on which impact of the investments in ICT on macroeconomic growth could be evaluated. And it is to this area that our research aims to contribute.

CHAPTER 5 Description of the Data

The data for this study were obtained from two sources. The first source was represented by the database of World Development Indicators, which is the World Bank's comprehensive database on development data. The second source of the data was represented by the *Yearbook of Statistics,* which is published yearly by International Telecommunication Union (ITU). In our choice of variables, we were greatly restricted by the availability of the data. For example, while the development data of the World Bank's database covers more than 600 indicators for 208 economies, data on many of the indicators relevant to our research were not available, or were available only for a few countries, or contained too few data points to be useful in statistical analysis. In terms of the length of the time series, we were restricted to the period from 1992 to 2002, data for which were provided by *Yearbook of Statistics* of ITU. After close examination of the available data, we have decided to construct two data sets, one covering the period from 1998 to 2002, and another one covering the period from 1993 to 2002.

There are several reasons why we have chosen such an approach. First, we decided to begin our analysis from year 1993 because it provided, in our opinion, a common starting point for the transitional economies. Our reasoning is that it took a year from the fall of the Berlin' Wall in 1991 for the process of transition to start, and year 1992 as a starting point could have favored "early starters." Second, we discovered that the 5-year period from 1998 to 2002 was reflected in both of our sources by richer data sets than any other period between 1993 and 2002. As a result, we ended up with a "lean" data set covering the years from 1993 to 2002, and a "rich" data set covering the years from 1998 to 2002. Finally, over the ten years of transition, the nature of the relationship between variables might change, and such change could be assessed by comparing whether the relationships of the late stage of the transition (1998-2002) hold for the whole period (1993-2002).

The following transitional economies have been selected for this study: Albania, Armenia, Azerbaijan, Belarus, Bulgaria, Czech Republic, Estonia, Hungary, Kazakhstan, Kyrgyz Republic, Latvia, Lithuania, Moldova, Poland, Romania, Slovak Republic, Slovenia, Ukraine, and Russian Federation. The data offered for Tajikistan, Turkmenistan, Uzbekistan, Georgia, Macedonia, and Croatia was insufficient to warrant the inclusion of these economies in this study.

The data from the two sources could not be merged by simply placing the data values in the same table; for the *Yearbook of Statistics* contains revenues and spending data expressed in local currency of each TE, while WDI data was expressed in US dollars. However, the "Average annual exchange rate per US$" was also provided by the Yearbook, thus, after the application of the exchange rate to each appropriate data point we could obtain the values expressed in current US dollars. The data values expressed in similar units, "Current US $," could also be found in the WDI data set, thus, after the application of the exchange rate to the local currency values in the *Yearbook of Statistics* , merging of the two data sets became relatively straightforward

Next, we describe the variables that were obtained from the two sources that were available to us.

5.1 WDI database of the World Bank

The names of all variables, as well as the descriptions of all variables, are reproduced here exactly, word for word, as they appear in the source.

Computer, communications and other services (% of commercial service exports)-

Computer, communications and other services (% of commercial service exports) include such activities as international telecommunications, and postal and courier services; computer data; news-related service transactions between residents and nonresidents; construction services; royalties and license fees; miscellaneous business, professional, and technical services; and personal, cultural, and recreational services.

High-technology exports (% of manufactured exports)-

Computer, communications and other services (% of commercial service imports) include such activities as international telecommunications, and postal and courier services; computer data; news-related service transactions between residents and nonresidents; construction services; royalties and license fees; miscellaneous business, professional, and technical services; and personal, cultural, and recreational services.

High-technology exports (current US$)-

High-technology exports are products with high R&D intensity, such as in aerospace, computers, pharmaceuticals, scientific instruments, and electrical machinery. Data are in current U.S. dollars.

Computer, communications and other services (% of commercial service imports)-

Computer, communications and other services (% of commercial service imports) include such activities as international telecommunications, and postal and courier services; computer data; news-related service transactions between residents and nonresidents; construction services; royalties and license fees; miscellaneous business, professional, and technical services; and personal, cultural, and recreational services.

Military expenditure (% of GDP)-

Military expenditures are based on the NATO definition, which includes all current and capital expenditures on the armed forces, including peacekeeping forces; defense ministries and other government agencies engaged in defense projects; paramilitary forces, if these are judged to be trained and equipped for military operations; and military space activities. Such expenditures include military and civil personnel, including retirement pensions of military personnel and social services for personnel; operation and maintenance; procurement; military research and development; and military aid (in the military expenditures of the donor country). Excluded are civil defense and current expenditures for previous military activities, such as for veterans' benefits, demobilization, conversion, and destruction of weapons. This definition cannot be applied for all countries, however, since that would require much more detailed information than is available about what is included in military budgets and off-budget military expenditure items. (For example, military budgets might or might not cover civil defense, reserves and auxiliary forces, police and paramilitary forces, dual-purpose forces such as military and civilian police, military grants in kind, pensions for military personnel, and social security contributions paid by one part of government to another.)

Military personnel (% of total labor force)-

Armed forces personnel refer to active duty military personnel, including paramilitary forces if those forces resemble regular units in their organization, equipment, training, or mission. Labor force comprises all people who meet the International Labor Organization's definition of the economically active population.

Fixed line and mobile phone-

Fixed lines are telephone mainlines connecting a customer's equipment to the public switched telephone network. Mobile phone subscribers refer to users of portable telephones subscribing to an automatic public mobile telephone service using cellular technology that provides access to the public switched telephone network.

International telecom, outgoing traffic (minutes per subscriber)-

International telecommunications outgoing traffic refers to the telephone traffic, measured in minutes per subscriber, that originated in the country with a destination outside the country.

Internet users (per 1,000 people)-

Internet users are people with access to the worldwide network.

Mobile phones (per 1,000 people)-

Mobile phones refer to users of portable telephones subscribing to an automatic public mobile telephone service using cellular technology that provides access to the public switched telephone network, per 1,000 people.

Telephone mainlines (per 1,000 people)-

Telephone mainlines are telephone lines connecting a customer's equipment to the public switched telephone network. Data are presented per 1,000 people for the entire country.

Telephone mainlines per employee-

Telephone mainlines per employee are calculated by dividing the number of mainlines by the number of telecommunications staff (with part-time staff converted to full-time equivalents) employed by telecommunications enterprises providing public telecommunications services.

International tourism, expenditures (current US$)-

International tourism expenditures are expenditures of international outbound visitors in other countries, including payments to foreign carriers for international transport. These may include expenditures by residents traveling abroad as same-day visitors, except in cases where these are so important as to justify a separate classification. Data are in current U.S. dollars.

Physicians (per 1,000 people)-

Physicians are defined as graduates of any facility or school of medicine who are working in the country in any medical field (practice, teaching, research).

Health expenditure per capita (current US$)-

Total health expenditure is the sum of public and private health expenditures as a ratio of total population. It covers the provision of health services (preventive and curative), family planning activities, nutrition activities, and emergency aid designated for health but does not include provision of water and sanitation. Data are in current U.S. dollars.

Health expenditure, private (% of GDP)-

Private health expenditure includes direct household (out-of-pocket) spending, private insurance, charitable donations, and direct service payments by private corporations.

Health expenditure, public (% of GDP)-

Public health expenditure consists of recurrent and capital spending from government (central and local) budgets, external borrowings and grants (including donations from international agencies and nongovernmental organizations), and social (or compulsory) health insurance funds.

Health expenditure, total (% of GDP)-

Total health expenditure is the sum of public and private health expenditure. It covers the provision of health services (preventive and curative), family planning activities, nutrition activities, and emergency aid designated for health but does not include provision of water and sanitation.

Pupil-teacher ratio, primary-

Primary school pupil-teacher ratio is the number of pupils enrolled in primary school divided by the number of primary school teachers (regardless of their teaching assignment).

School enrollment, secondary (% gross)-

Gross enrollment ratio is the ratio of total enrollment, regardless of age, to the population of the age group that officially corresponds to the level of education shown. Secondary education completes the provision of basic education that began at the primary level, and aims at laying the foundations for lifelong learning and human development, by offering more subject- or skill-oriented instruction using more specialized teachers.

School enrollment, tertiary (% gross)-

Gross enrollment ratio is the ratio of total enrollment, regardless of age, to the population of the age group that officially corresponds to the level of education shown. Tertiary education, whether or not to an advanced research qualification, normally requires, as a minimum condition of admission, the successful completion of education at the secondary level.

Research and development expenditure (% of GDP)-

Expenditures for research and development are current and capital expenditures (both public and private) on creative, systematic activity that increases the stock of knowledge. Included are fundamental and applied research and experimental development work leading to new devices, products, or processes.

Researchers in R&D (per million people)-

Researchers in R&D are people trained to work in any field of science who are engaged in professional R&D activity. Most such jobs require completion of tertiary education

Technicians in R&D (per million people)-

Technicians in R&D are people engaged in professional R&D activity who have received vocational or technical training in any branch of knowledge or technology. Most such jobs require three years beyond the first stage of secondary education.

Scientific and technical journal articles-

Scientific and technical journal articles refer to the number of scientific and engineering articles published in the following fields: physics, biology, chemistry, mathematics, clinical medicine, biomedical research, engineering and technology, and earth and space sciences.

Labor force, total-

Total labor force comprises people who meet the International Labour Organization definition of the economically active population: all people who supply labor for the production of goods and services during a specified period. It includes both the employed and the unemployed. While national practices vary in the treatment of such groups as the armed forces and seasonal or part-time workers, in general the labor force includes the armed forces, the unemployed, and first-time job-seekers, but excludes homemakers and other unpaid caregivers and workers in the informal sector.

Unemployment, female (% of female labor force)-

Unemployment refers to the share of the labor force that is without work but available for and seeking employment. Definitions of labor force and unemployment differ by country.

Unemployment, male (% of male labor force)-

Unemployment refers to the share of the labor force that is without work but available for and seeking employment. Definitions of labor force and unemployment differ by country.

Unemployment, total (% of total labor force)-

Unemployment refers to the share of the labor force that is without work but available for and seeking employment. Definitions of labor force and unemployment differ by country.

Roads, paved (% of total roads)-

Paved roads are those surfaced with crushed stone (macadam) and hydrocarbon binder or bituminized agents, with concrete, or with cobblestones, as a percentage of all the country's roads, measured in length.

Roads, total network (km)-

Total road network includes motorways, highways, and main or national roads, secondary or regional roads, and all other roads in a country

Vehicles (per 1,000 people)-

Motor vehicles include cars, buses, and freight vehicles but do not include two-wheelers. Population refers to midyear population in the year for which data are available.

Passenger cars (per 1,000 people)-

Passenger cars refer to road motor vehicles, other than two-wheelers, intended for the carriage of passengers and designed to seat no more than nine people (including the driver).

Foreign direct investment, net inflows (% of GDP)-

Foreign direct investment is net inflows of investment to acquire a lasting management interest (10 percent or more of voting stock) in an enterprise operating in an economy other than that of the investor. It is the sum of equity capital, reinvestment of earnings, other long-term capital, and short-term capital as shown in the balance of payments. This series shows net inflows in the reporting economy.

GDP (current US$)-

GDP at purchaser's prices is the sum of gross value added by all resident producers in the economy plus any product taxes and minus any subsidies not included in the value of the products. It is calculated without making deductions for depreciation of fabricated assets or for depletion and degradation of natural resources. Data are in current U.S. dollars. Dollar figures for GDP are converted from domestic currencies using single year official exchange rates. For a few countries where the official exchange rate does not reflect the rate effectively applied to actual foreign exchange transactions, an alternative conversion factor is used.

Population, total-

Total population is based on the de facto definition of population, which counts all residents regardless of legal status or citizenship—except for refugees not permanently settled in the country of asylum, who are generally considered part of the population of their country of origin.

Private investment in telecoms (current US$)-

Private investment in telecoms covers infrastructure projects in telecommunications that have reached financial closure and directly or indirectly serve the public. Movable assets are excluded. The types of projects included are operations and management contracts, operations and management contracts with major capital expenditure, greenfield projects (in which a private entity or a public-private joint venture builds and operates a new facility), and divestiture. Data are in current U.S. dollars.

Public spending on education, total (% of GDP)-

Public expenditure on education consists of public spending on public education plus subsidies to private education at the primary, secondary, and tertiary levels.

Exports of goods and services (% of GDP)-

Exports of goods and services represent the value of all goods and other market services provided to the rest of the world. They include the value of merchandise, freight, insurance, transport, travel, royalties, license fees, and other services, such as communication, construction, financial, information, business, personal, and government services. They exclude labor and property income (formerly called factor services) as well as transfer payments.

Trade (% of GDP)-

Trade is the sum of exports and imports of goods and services measured as a share of gross domestic product.

5.2 Yearbook of Statistics of International Telecommunication Union

The names of all variables, as well as the descriptions of all variables, are reproduced here exactly, word for word, as they appear in the source.

Average annual exchange rate per US$-

Refers to the yearly average of the number of national currency units per one United States dollar. Yearly averages are generally calculated as arithmetic averages of monthly rates. Note that there can be differences between official, market and unofficial rates. In some countries for which average rates cannot be obtained, an end of period rate is used.

Telecom equipment exports-

Telecommunication equipment exports and imports are shown in national currency. The data come from national sources.

Telecom equipment imports-

Telecommunication equipment exports and imports are shown in national currency. The data come from national sources.

Main telephone lines in operation-

The number of telephone lines connecting the subscriber's terminal equipment to the public switched network and which have a dedicated port in the telephone exchange equipment. This term is synonymous with the term main station or Direct Exchange Line (DEL) which are commonly used in telecommunication documents. It may not be the same as an access line or a subscriber. The definition of access line used by some countries varies. In some cases, it refers to the total installed capacity (rather than lines in service). In other cases it refers to all network access points including mobile cellular subscribers. Telephone subscribers would not generally include public telephones which are included in main lines.

Main telephone lines per 100 inhabitants-

Calculated by dividing the number of main lines by the population and multiplying by 100.

Public payphones-

The total number of all types of public telephones, including coin and card operated and public telephones in call offices. Public phones installed in private places should also be included as should mobile public telephones. All public telephones regardless of capability (e.g., local calls or national only) should be counted

Cellular mobile telephone subscribers-

Refers to users of portable telephones subscribing to an automatic public mobile telephone service which provides access to the Public Switched Telephone Network

(PSTN) using cellular technology. This can include analogue and digital cellular systems (including Microcellular systems such as DCS-1800, Personal Handyphone System (PHS) and others) but should not include non-cellular systems. Subscribers to fixed wireless (e.g., Wireless Local Loop (WLL)), public mobile data services, or radio paging services are not included.

Line capacity of local public switching exchanges-

The total capacity of public switching exchanges corresponds to the maximum number of main lines that can be connected. This number includes, therefore, main lines already connected and main lines available for future connection, including those used for the technical operation of the exchange (test numbers).

Cellular subscribers per 100 inhabitants-

Calculated by dividing the number of cellular mobile subscribers by the population and multiplying by 100.

Residential telephone connection charge-

The one time charge involved in applying for residential telephone service. Where there are different charges for different exchange areas, the charge is generally for the largest urban unless otherwise noted. Also, when the business and residential fixed telephone tariffs are the same, then business tariff has been omitted in some cases.

Business telephone connection charge-

Same as above but applied to business users.

Residential telephone monthly subscription-

Refers to the recurring fixed charge for a residential subscriber to the PSTN. The charge covers the rental of the line but not the rental of the terminal (e.g., telephone set). In some cases, the rental charge includes an allowance for free or reduced rate call units. If there are different charges for different exchange areas, the largest urban area is used. Also, when the business and residential fixed telephone tariffs are the same, then business tariff has been omitted in some cases.

Business telephone monthly subscription-

Same as above but applied to business subscribers.

Cost of a 3-minute local call (peak rate)-

Local call refers to the cost of a peak rate 3-minute call within the same exchange area using the subscriber's own terminal (i.e., not from a public telephone).

Cost of a 3-minute local call (off-peak rate)-

Local call refers to the cost of an off-peak rate 3-minute call within the same exchange area using the subscriber's own terminal (i.e., not from a public telephone).

Total telecommunication services revenue-

Refers to earnings from the direct provision of facilities for providing telecommunication services to the public (i.e., not including revenues of resellers). This includes revenue from fixed telephone, mobile communications, text (telex, telegraph and facsimile), leased circuits and data communications services. Some countries include telecommunication-related revenue such as directory advertising and equipment rental or sales. Others include value-added telecommunication services such as the provision of electronic mail or on-line services.

Annual telecommunication investment-

Refers to expenditure associated with acquiring the ownership of telecommunication equipment infrastructure (including supporting land and buildings and intellectual and non-tangible property such as computer software). These include expenditure on initial installations and on additions to existing installations.

Personal computers-

The number of personal computers (i.e., designed to be operated by a single user at a time) in use in the country. Primarily ITU estimates based on a number of national and international sources.

Estimated Internet users-

The number of Internet users.

Internet users per 100 inhabitants-

Calculated by dividing the number of Internet users by the population and multiplying by 100.

International outgoing telephone traffic (minutes)-

This covers the effective (completed) traffic originating in a given country to destinations outside that country. Many countries have now shifted to reporting international traffic volumes based on point of billing. This means that the data refers to traffic billed in the country.

International incoming telephone traffic (minutes)-

Effective (completed) traffic originating outside the country with a destination inside the country.

5.3 Data Issues

While compiling our data sets we had to primarily deal with two issues. The first issue was how to find a scale that minimizes the bias arising from comparing and contrasting the countries that are different in size, population and level of wealth. The second issue was how to deal with the missing data.

In dealing with the issue of bias minimization, we decided to use, where it was possible, percentages rather than actual number values. Thus, for example, while the variable "full-time telecommunication staff" was represented in the *Yearbook of Statistics* as a number value, we have transformed it into the percentage of the variable *labor force, total* provided by WDI data set by using following formula:

Full-time telecommunication staff, % of total labor force =

= (*Full-time telecommunication staff* / *Labor force, total*) * 100%

In similar fashion, we have transformed numerical dollar-figures of revenues and investments into the percentages of GDP by using the same approach. For example, to transform the variable *Annual telecom investment* into the percentage of GDP we used following formula:

Annual telecom investment, % of GDP =

= (*Annual telecom investment (in current US \$) / GDP(in current US \$)*) * 100%

The reasoning behind this approach was that the percentages would maximize the objectivity of representing a specific structure of the economy and the labor force, while minimizing the subjectivity associated with the size and level of economic development of a given transitional economy.

In dealing with the issue of missing data, we have adopted two approaches, interpolation and extrapolation. As an example of the first approach, in some cases we have substituted the missing value by the value of the previous or the next year. For example, in the case of Ukraine the variable *Physicians, per 1000 people* is missing values for the years 2002 and 2003. However, the values for the 2001 and 2004 are present and are both equal to "2.97387." In this situation, we have assumed that the missing value for this variable for 2002 is highly likely not going to be very different from the years for which the data were present and, therefore, interpolated the missing value with the value of "2.97387."

In other cases, however, a different approach was taken, for a simple substitution of the missing value by the value for the previous year, or year after, was judged to be inadequate. For example, in the case of Kyrgyz Republic variable *Full-time telecommunication staff* has a missing value for the year 2002. Upon a close examination of the data, we have noticed that the values for this variable were changing significantly over the years leaving us no reason to believe that the value for the year 2002 would replicate the value for the year 2001. Moreover, it was noticed that the value of the variable *Annual telecom investment*, which potentially could influence the value of the variable *Full-time telecommunication staff*, has increased significantly for the year 2002 as compared to the year 2001. In order to impute the missing value in this case, as well as in all similar cases, we have chosen to plot the data and add a trend line to the resulting graph. In the imputation of the substitute variable we decided to choose a value that would minimize the existing trend, for we rather err on the side of the underestimation than obtain inflated by our own device results.

Figure 4, provided below, demonstrates the original data set, with a missing value for the year 2002, and a data set with the extrapolated value for 2002. The data point representing the extrapolated value of the variable *Full-time telecommunication staff* is shown as a black oval on the bottom diagram, with the trend lines added to demonstrate the effect of the substitution.

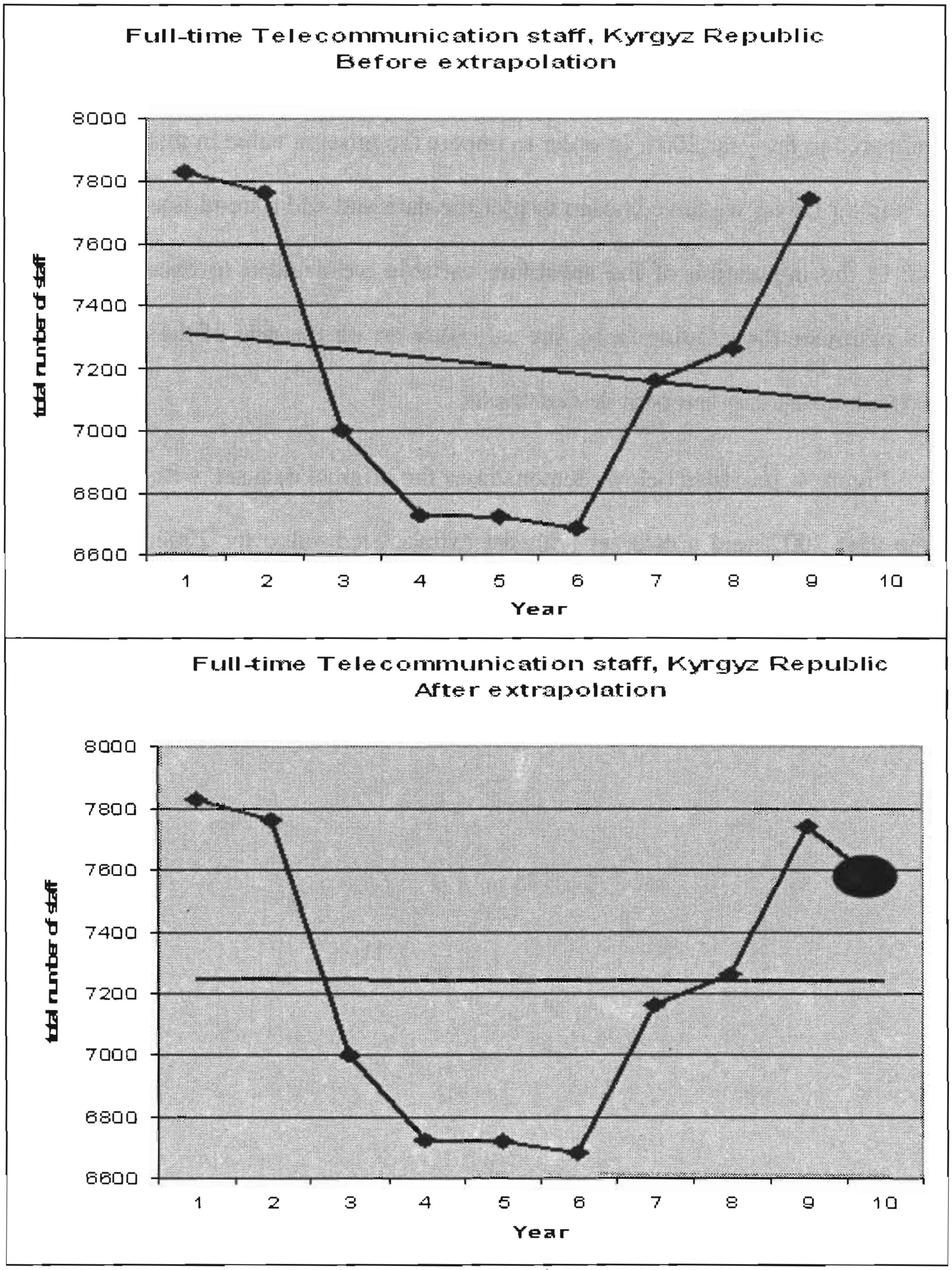


Figure 4 Example of Extrapolation

CHAPTER 6 Overview of the research methodology

6.1 Introduction

The data analysis methods and statistical tools could be generally separated into two categories, the ones that allow us to generalize and could be used to build and test a theory, and the other ones that are usually used to inquire into the narrower realm of the data at hand. In this research, we are going to employ both types of tools, with the intent of building a theory based on the uncovered relationship within the data at hand. Clustering, Data Envelopment Analysis (DEA), and Classification (Decision) Trees are the methods that are commonly used to inquire into and analyze the existing data set at hand, without making any generalizations or assumptions about the rest. We use these tools in this research to gain some insights into the nature of the relationships between the variables that comprise our real-life data set. Then, upon gaining the "localized" knowledge, we employ translog production function and structural equation modeling to generalize the discovered relationships and construct the possible theoretical explanations. Let us see how we could place our methodology within the framework of one of the better tools that the positivist science could offer, namely, the hypothetico-deductive logic. The common structure is depicted in the Figure 5 below. According to the proposed methodology, we begin at the very bottom, by attempting to find the link between the investments in ICT and TFP. This corresponds to "Conclusion" part of the hierarchy below. Once the connection is established, we would attempt to find out the set of conditions, which predicate the existence of the link. Because we are restricted to the data at hand, we still operate at the local, yet higher level of the "Minor premise" within the hierarchy. Finally, we shall attempt to produce a generalized statement or rule that would allow us to enter the "Major premise" level of the hierarchy.

Once the major premise is generated, our newly minted hypothesis could be tested in another environment, such as that of developed, or developing, or least developed countries. In this sense our setting of the economies in transition is a fortunate one, for it is distinctive enough to be warranted its own appellation within a taxonomy, and yet containing the general characteristic of other types of countries. This gives us at least some reasons to assume that conclusion of this study could be applied beyond its setting.

Thus, our methodology proceeds "bottom-up," spanning all the way from a conclusion, to the minor premise and the major premise. It is not difficult to map the tools and methods that we use to each of the boxes in the Figure 5.

For example, starting in the "Conclusion" part of the framework of the hypothetico-deductive logic, we first are going to determine the existence of the indirect relationship between TFP and investments in ICT. We are going to use Cluster Analysis, Decision Trees analysis, Data Envelopment Analysis, and Correlation as the data analytic tools at this stage.

Once the presence of the indirect relationship is established, we are going to determine the conditions required for such relationship to exist. This stage utilizes Decision Trees and Multiple Regression as the data analytic tools and corresponds to the "Minor Premise" part of the framework.

Next, in the "Major Premise" of the framework, we are going to formulate our research model relating *investments in ICT* to *TFP* in the context of TEs. The hypothesis is going to take form of the structural equation model. Structural Equation Modeling implemented with Partial Least Squares is the tool that we are going to use at this stage.

Finally, we are going to test our formulated hypothesis, expressed as our research model, by running Structural Equation Model. Based on the results of the corresponding to our research model measurement and the structural models, we are going to reject or accept our hypothesis. This would bring us to the "Conclusion" part of the framework.

Now we present our course of action in more details.

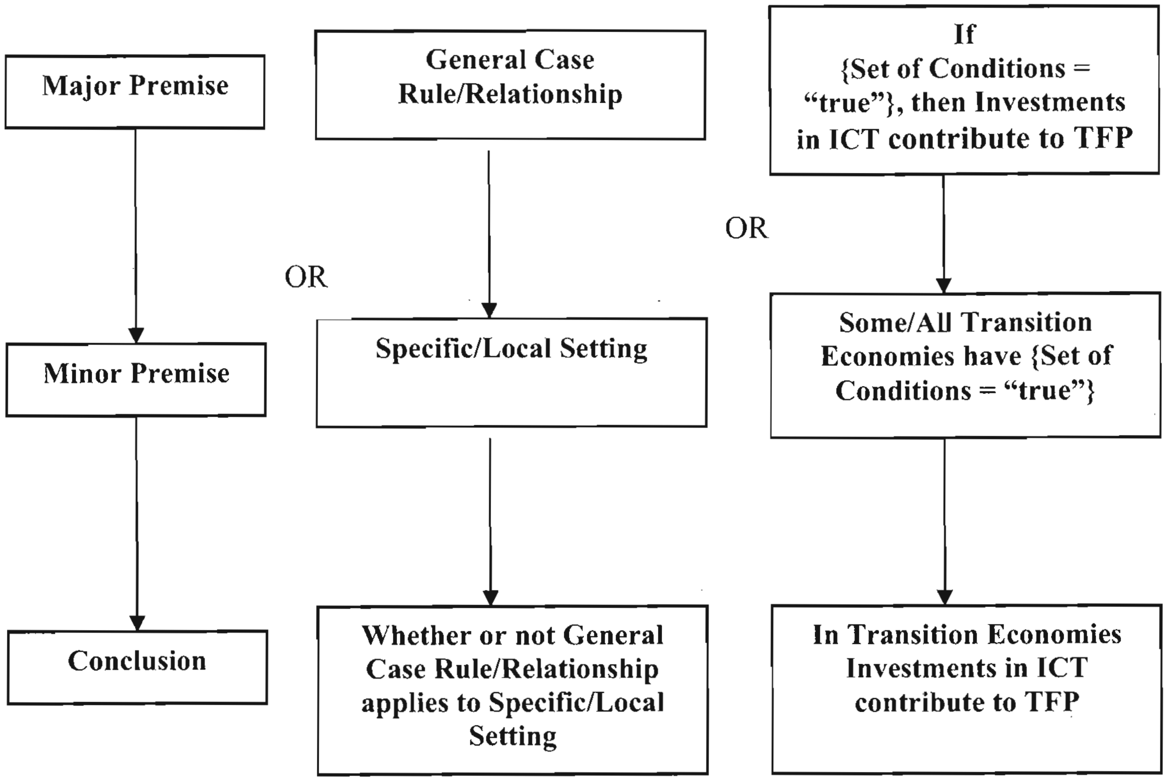


Figure 5
The research methodology within the framework of the hypothetico-deductive logic

6.2 Research methodology

The first step in our investigation involves Cluster analysis (CA). The purpose of the clustering is to determine whether or not the selected set of TEs is homogenous. Specifically, we would like to obtain some hints on whether or not there are clear and specific distinctions between some of the TEs.

In the second part of our inquiry, we employ DT analysis. At this stage, the goal is to investigate the differences between the clusters of TEs that were generated by the

CA in the previous step. Taking a step-by-step approach we attempt to uncover set of dimensions along which TEs in our sample vary the most.

In the next step of our investigation, we perform Data Envelopment Analysis (DEA). This method allows us to inquire into the efficiency of the TEs of our set. By assigning an efficiency score to each TE, DEA allows us to determine the contents of the two subsets of our sample: efficient and inefficient TEs.

The fourth part of our investigation would take advantage of the Decision Tree (DT) analysis. By creating a binary dummy variable, we are able to incorporate the results of DEA and create a Classification DT. As a result, we are able to obtain the insights into the nature of the differences between the efficient and inefficient TEs.

The fifth stage of our investigation involves using of the Translog regression model. The purpose of this part of our analysis is to determine whether or not there exists a set of investments that are complementary to the investments in ICT. The use of Translog allows us to incorporate the interaction terms in the production function and to test the significance of the interaction terms. Based on the results of this stage we are able to find out some of the types of the investments affecting the investments in ICT.

The sixth part of our investigation relies on using Structural Equation Modeling (SEM). The goal of this stage is to determine the presence of the relationship between the investments in ICT and the unexplained part of the macroeconomic growth, TFP. We apply SEM to the whole sample of our TEs, as well as to the subsets of the sample that were determined by the CA in the first step. As a result, we are able to identify what subset of our sample exhibits statistically discernible relationship, and which subset does not.

The last step of the data analysis involves using Classification DT and Neural Network (NN) analyses. During the final stage of our research, we aim to investigate the reasons why some of the TEs exhibit statistically significant relationship between the investments in ICT and TFP, while other TEs do not. By using DT, we are able to inquire into the differences regarding the ICT Capitalization, while use on NN allows us to obtain some hints regarding the differences in effectiveness of the process by which the investments in ICT contribute to TFP.

We offer an overview of the tools and techniques used in this study next.

6.3 Overview of the Methods

In this part of the paper, we provide brief introductions and overviews of the methods that are used in this study. Our seven-step research methodology takes advantage of the following six data analytic methods: cluster analysis, decision tree analysis, data envelopment analysis, multivariate regression, structural equation modeling, and neural networks.

The statistical tools used in the order as listed. However, we are going to offer an overview of data envelopment analysis (DEA) prior to cluster analysis (CA), for the use of CA is warranted by the specifics of DEA.

6.3.1 Data Envelopment Analysis

Data Envelopment Analysis (DEA) is a nonparametric method of measuring the efficiency of a decision-making unit (DMU). Any collection of entities that operates with the same set of the inputs and produces the same outputs, be it a firm or a country, could be designated as DMUs. In this DEA study, DMU would represent a country from the set of TE. This method, which is nonparametric in the sense that DEA is entirely based on the observed input-output data, was originated as a collection of techniques for measuring the relative efficiency of a set of DMUs with unknown or unavailable price values for data inputs and outputs (Sengupta 1995). While it is beyond the scope of this paper to provide an overview of the theory behind the computations in DEA, we would like to direct the interested reader to the comprehensive presentation of the theoretical underpinnings of the DEA by Dula (2002b).

One of the benefits of DEA lies in its flexibility, for a researcher could take advantage of several models and orientations that this method has to offer. Thus, for example, the choice of a given DEA model would depend on the underlying economic assumptions about the returns to scale of the process that transforms the inputs into the outputs (Dula 2002a). Consequently, the different sets of assumptions would yield the different models. As a result, instead of forcing a single perspective, DEA offers multiple vantage points in the form of the several models and orientations available to the researcher.

The three commonly mentioned orientations of DEA model are the following: input-oriented, output oriented, and base-oriented (Charnes, Cooper, Lewin and Seiford 1994). An input-oriented model is concerned with the minimization of the use of the inputs for achieving a given level of the output. Thus, it deals with the efficiency of the input utilization and inputs within the model are controllable. In our research, we employ the input-oriented model in order to approach the issue of efficiency from the perspective of the policy maker, who would probably be interested in how efficiently the allocated investment resources are utilized by a given TE.

Output-oriented DEA model, on the other hand, is concerned with the maximization of the level of the outputs per given level of the inputs. Thus, it deals with the efficiency of the output production where outputs are controllable. In this study, we use output-oriented model, for it appears reasonable to expect that the macroeconomic goal of any economy is maximization of the output as represented by GDP.

Base-oriented model, unlike the first two, has dual orientation and is concerned with the optimal combination of the inputs and outputs. Therefore, this type of DEA model deals with the efficiency of the input utilization and efficiency of the output production, having control over both inputs and outputs within the model.

As a result, definition of efficiency depends on the orientation of the model. In the case of an input-oriented model, no DMU would be considered efficient if it is possible to decrease any of its inputs without affecting any other inputs and without decreasing the level of the outputs. While in the case of an output-oriented model, a DMU could not be considered efficient if it is possible to increase the level of any of its outputs without affecting other outputs or increasing level of any of its inputs (Charnes, Cooper and Rhodes 1978).

The empirical foundation of DEA eliminates the need for some of the assumptions and limitations of traditional efficiency measurement approaches (Bowlin 1998). As a result, DEA can be used in cases where the relationships between the multiple inputs and multiple outputs involved in DMUs are complex or unknown (Cooper, Seiford, and Zhu 2004).

The original DEA model was introduced in 1978 by Charnes, Cooper and Rhodes and it is commonly called CCR (an abbreviation consisting of first letters of the authors' names). This model allows representing multiple inputs and outputs of each DMU as a single abstract "meta input" and "meta output." Consequently, the efficiency of each DMU can be represented as a ratio of the abstract input to the abstract output, and the resulting efficiency value can then be used for comparison with other DMUs in the set.

Mathematically, this ratio can be expressed as the following objective function:

$\max h_0(u,v_i) = \Sigma_r u_r y_{ro} / \Sigma_i v_i x_{io}$

where,

the u_r is the variable representing the output value

the v_i is the variable representing the observed input value

the y_{ro} is the observed amount y of output r produced by DMU_0 from the input x_{io}

x_{io} is the observed amount of input i consumed in order to produce y amount of output r by the DMU_0 (the DMU to be evaluated).

Adding the normalizing constraint, according to which ratio of virtual input to virtual output for each DMU must be less than one, the following linear programming (LP) problem can be formulated:

$\max h_0(u,v) = \Sigma_r u_r y_{ro} / \Sigma_i v_i x_{io}$

subject to

$\Sigma_r u_r y_{rj} / \Sigma_i v_i x_{ij} \leq 1$ for $j = 1,\ldots, n$

$u_r, v_i \geq 0$ for all i and r,

where in the case of fully rigorous development $u_r, v_i \geq 0$ would be replaced with the constraint including a *non-Archimedian element* ε such as

$$\frac{u_r}{\sum_{i=1}^{m} v_i x_{i0}}, \frac{v_r}{\sum_{i=1}^{m} v_i x_{i0}} \geq \varepsilon > 0$$, and ε is smaller than any positive real number.

(adapted from Cooper et al. (2004).

However, the given above formulation yields infinite number of solutions, thus, transformation of the formulation has been developed by Charnes and Cooper(1962), which changes (u, v) variables to (μ, v) and yields the equivalent linear programming problem to the one above. Following is a Charnes-Cooper transformation

$$\max z = \sum_{r=1}^{s} \mu_r y_{r0}$$

Subject to

$$\sum_{r=1}^{s} \mu_r y_{r0} - \sum_{i=1}^{m} v_i x_{ij} \leq o$$

$$\sum_{i=1}^{m} v_i x_{ij} = 1$$

$$\mu_r, v_i \geq 0$$

The LP Dual for this linear problem, which sometimes referred to as the "Farrell model," can be formulated in its "relaxed" form as

$$\Theta^* = \min \Theta$$

Subject to

$$\sum_{J=1}^{n} x_{ij}\lambda_j \leq \Theta x_{i0} \qquad i = 1,2,\ldots,m;$$

$$\sum_{J=1}^{n} y_{rj}\lambda_j \geq y_{r0} \qquad r = 1,2,\ldots,s;$$

$$\lambda_j \geq 0 \qquad j = 1,2,\ldots,n;$$

The provided above model operates under assumption of *strong disposal* by ignoring the presence of non-zero slacks, allowing, consequently, for the solutions with *weakly efficient* DMUs (more on this below).

The optimal solution to this problem, Θ^*, yields the efficiency score for a particular DMU, which implies that

$$\Theta^* \leq 1$$

As a result, all DMUs for which

$$\Theta^* < 1$$

are inefficient, while those for which

$$\Theta^* = 1$$

are considered to be 100 percent efficient.

A relative efficiency of a DMU can be characterized as being *strong* or *weak*. For example, a DMU is considered to be *strongly* (fully) efficient if $\Theta^* = 1$ and all slacks are equal to zeroes. On another hand, a DMU could be *weakly* efficient if it obtained the same score of $\Theta^* = 1$, but some slacks are not equal to zero. As it was shown above, the type of relative efficiency of a given DMU is determined by the constraint of the LP problem: if the constraint utilizes a *non-Archimedian element* ε, the solution would allow for *fully efficient* DMUs, and if not, then all relatively efficient DMUs must be qualified as *being weakly* efficient.

As we stated previously, in this research we are going to utilize input-oriented DEA model as well as output-oriented model. The type of relative efficiency of a DMU would not be affected by the orientation of the model, for it is dependent on whether the slacks are equal to zero or not. Next, we provide a brief overview and formulation of the output-oriented DEA model.

In the case of the output-orientation, the objective function would be based on the ratio of virtual output to input, which would yield similar, but "re-oriented" objective function

$$\min h_0(u,v) = \Sigma_i v_i x_{io} / \Sigma_r u_r y_{ro}$$

subject to

$$\Sigma_i v_i x_{ij} / \Sigma_r u_r y_{rj} \leq 1 \text{ for } j = 1,\ldots, n$$

$$u_r, v_i \geq 0 \text{ for all } i \text{ and } r$$

or, again, in the case of the fully rigorous development

$$u_r, v_r \geq \varepsilon > 0 \text{ for all } i \text{ and } r$$

Charnes-Cooper transformation of which would take the form as follows

$$\min q = \sum_{i=1}^{m} v_i x_{i0}$$

Subject to

$$\sum_{i=1}^{m} v_i x_{ij} - \sum_{r=1}^{s} \mu_r y_{rj} \geq o$$

$$\sum_{r=1}^{s} \mu_r y_{r0} = 1$$

$$\mu_{r,} \, v_i \geq \varepsilon$$

Consequently, the corresponding output-oriented LP dual could be represented as the following output-based Farrell model

$$\Theta^* = \max \Theta$$

Subject to

$$\sum_{j=1}^{J} z_j x_{jn} \geq \Theta u_{jm} \qquad m = 1,2,\ldots,M;$$

$$\sum_{j=1}^{J} z_j x_{jn} \leq x_{jn} \qquad n = 1,2,\ldots,N;$$

$$z_j \geq 0 \qquad j = 1,2,\ldots,J;$$

Where

Θ = output technical efficiency measure,

u_{jm} = quantity of output m produced by DMU j,

xj_n = quantity of input n produced by DMU j, and

z_j = intensity variable for DMU j (adapted from Cooper et al. (2004).

To briefly recap, in the case of "non-relaxed" LP, a score less than one means that some other unit(s) from the sample could produce the given level of outputs using less inputs (in output-oriented model), or, could utilize the given level of the inputs more efficiently by produce higher level of the outputs(in the case of input-oriented model). In the case of the "relaxed" LP, however, a DMU receiving a score of less than 1 could still be considered "weakly" efficient. This 'weak' efficiency could take place in the case if the optimal solution with positive slacks exists and there no other DMU which is better in every input or output.

Thus, by using the techniques of LP, this comparison results in efficiency ranking of each DMU in the given set, where the highest ranking DMU is considered to be 100% efficient (Sengupta 1996). Because multiple DMUs could receive the same score, there could be multiple 100% efficient DMUs in the given set. As a result, DEA 'envelops' the data set with the boundary points represented by the 100% efficient DMUs.

According to Charnes, Cooper, Lewin, and Seiford (1994), the concepts and methodologies of DEA are now incorporated into four models: CCR, BCC, Multiplicative, and Additive. The authors summarize the interpretive possibilities of each of the listed above models as follows:

1. The CCR (***C***harnes, ***C***ooper and ***R***hodes) ratio model yields an objective evaluation of overall efficiency, identifies the sources, and estimates the amounts of thus identified inefficiencies.
2. The BCC (***B***anker, ***C***harnes, and ***C***ooper) model distinguishes between technical and scale inefficiencies by estimating pure technical efficiency at the given scale

of operation and identifying whether increasing, decreasing, or constant returns to scale possibilities are present for further exploitation.

3. The Multiplicative model provides a log-linear envelopment or a piecewise Cobb-Douglas interpretation of the production process.
4. The Additive model and the extended Additive model that relate DEA to the Charnes-Cooper inefficiency analysis and in the process relate the efficiency results to the economic concept of Pareto optimality (Charnes et al. 1994).

The DEA models with the different assumptions regarding their return-to-scale would, naturally, may produce the different interpretations of the data. As a result, the models with the different return-to-scale' assumptions would produce the different envelopment surfaces (commonly called 'efficiency frontier') consisting of the set of the 100% efficient DMUs. In this regard, based on the created efficiency frontier, the results of DEA could provide descriptive, as well as prescriptive information to each DMU.

Descriptively, results of the analysis could yield the relative efficiency of each DMU in comparison to the rest of the set and in comparison to the efficiency frontier. While prescriptively, DEA could provide some suggestion regarding improvement of the inefficient DMUs. Specifically, how much a given DMU must it alter its outputs and inputs in order to become more efficient (Soteriou and Zenios, in Markoulides 1998). We describe in more details the model used in this study next.

DEA, BCC model

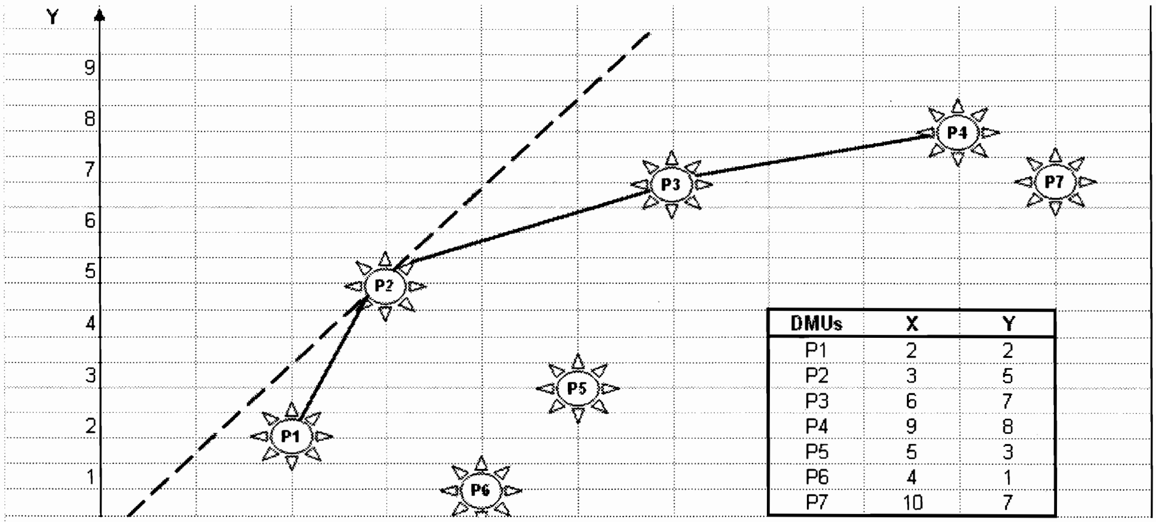
A widely used DEA model was introduced in 1984 by Banker, Charnes, and Cooper and is commonly called BCC model. The main difference between CCR model and the model of Banker, Charnes, and Cooper lies in how the returns to scale are handled. While CCR model assumes constant return to scale, BCC model is more flexible in that it allows for the variable returns to scale. Hence, a given DMU is considered to be efficient by CCR model only if is both scale[1] and technically[2] efficient, while for the same DMU to be considered efficient by BCC model it must only be technically efficient (Bowlin 1998). Thus, if a DMU is considered to be efficient by CCR model, it will also be considered as such by BCC model, while reverse not necessarily being true.

As well as CCR model, BCC makes the efficiency of a given DMU to be contingent on two conditions: the value of proportional reduction being equal to one and all slacks being equal to zero. This, of course, is equivalent to zero distance from the efficiency frontier. Thus, all the sources on inefficiency stem from nonzero slacks and value of proportional reduction being less than one.

Similarly to CCR model, for a given DMU to be qualified as efficient under the input oriented model of BCC, it must simultaneously qualify as an efficient under the output oriented model and vice versa. Figure below illustrates the differences in the

[1] Scale efficiency is a measure of the deviation from the constant return to scale
[2] Technical efficiency is represented and measured as the greatest proportion of the maximum potential output to the actual or observed output in the case of output-oriented model, or the greatest proportion of

enveloping surfaces produced by CCR and BCC models for the set of 7 DMUs with a single input and a single output (adapted from Charnes et al. 1994).



DMUs	X	Y
P1	2	2
P2	3	5
P3	6	7
P4	9	8
P5	5	3
P6	4	1
P7	10	7

Figure 6 DEA: CCR (Dashed line) and BCC (Solid line) Models

In the Figure 6 above, dashed line represents the frontier of CCR model, with constant returns to scale, while the solid line depicts the enveloping surface produced by BCC model with the variable returns to scale. As we could see, CCR model produces efficiency frontier defined by the single DMU (P2), while the efficiency frontier of the BCC model is produced by the four DMUs (P1, P2, P3 and P4). The DMUs that do not belong to this envelopment surface (or efficient frontier) are operating inefficiently.

In both cases, whether we use CCR or BCC model, the use of DEA allows for producing of the efficiency frontier consisting of the efficient countries. The efficiency of

---

the minimal feasible input usage to the actual or observed input usage for input-oriented model (from OnFront Reference guide 1998-2000) the countries of TEs is determined by how well the set of inputs, one of which is the investment in ICT, is transformed in the output, represented by GDP. Next, for the purposes of our study we need to isolate yearly changes in TFP, as values of GDP change from year-to-year. This is accomplished by means of using Malmquist index, covered next.

DEA, Malmquist Index

It is commonly assumed, that economic growth could be determined by two factors. The first factor, resource accumulation, could lead to high rates of growth, albeit, due to the law of diminishing return, only for a limited period of time. Thus, it is the second factor, growth in productivity that is assumed to allow for attaining of the sustained economic growth. The productivity is commonly referred to as Total Factor Productivity (TFP) and its growth is measured by the use of Malmquist index.

Based on the idea of productivity index, originally suggested by Malmquist (1953), Caves, Christensen and Diewert (1982) defined the Malmquist index of TFP growth. Later, Färe, Grosskopf, Norris, and Zhang (1994) have shown that the Malmquist index could be constructed based on the results of DEA. Essentially, the approach is based on performing DEA analysis in two points in time; let us say T1 and T2. Let us again consider running example based on the set of 7 DMUs, each with single input and a single output. Figure 7 below shows constructed efficiency frontier at the time T1, and, designated by a dashed line, inefficiency of the DMU P5 relative to the frontier.

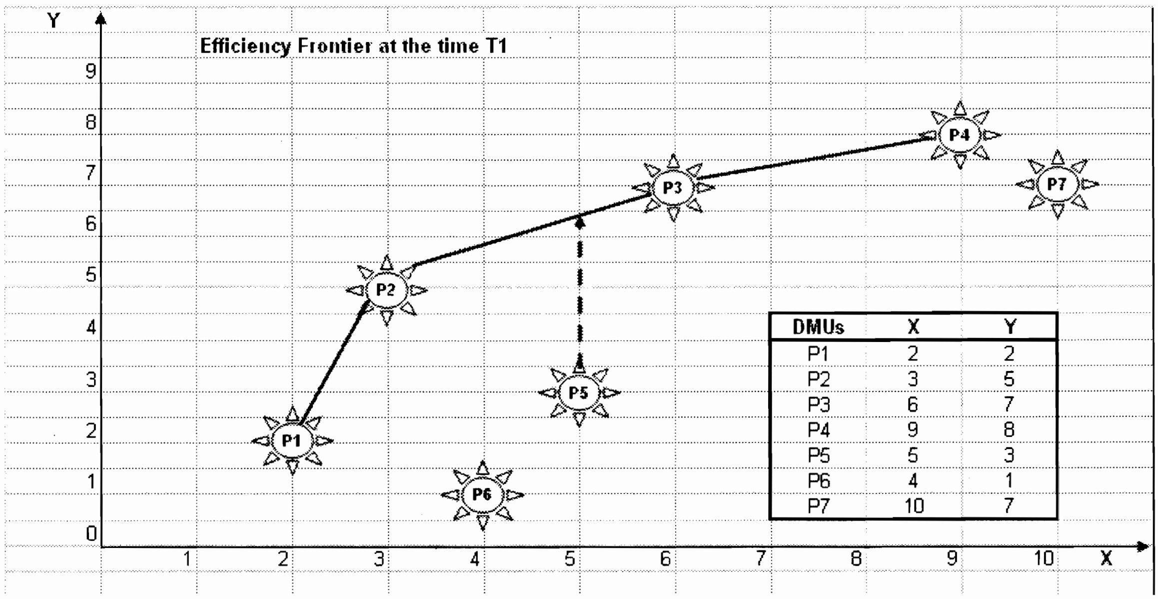
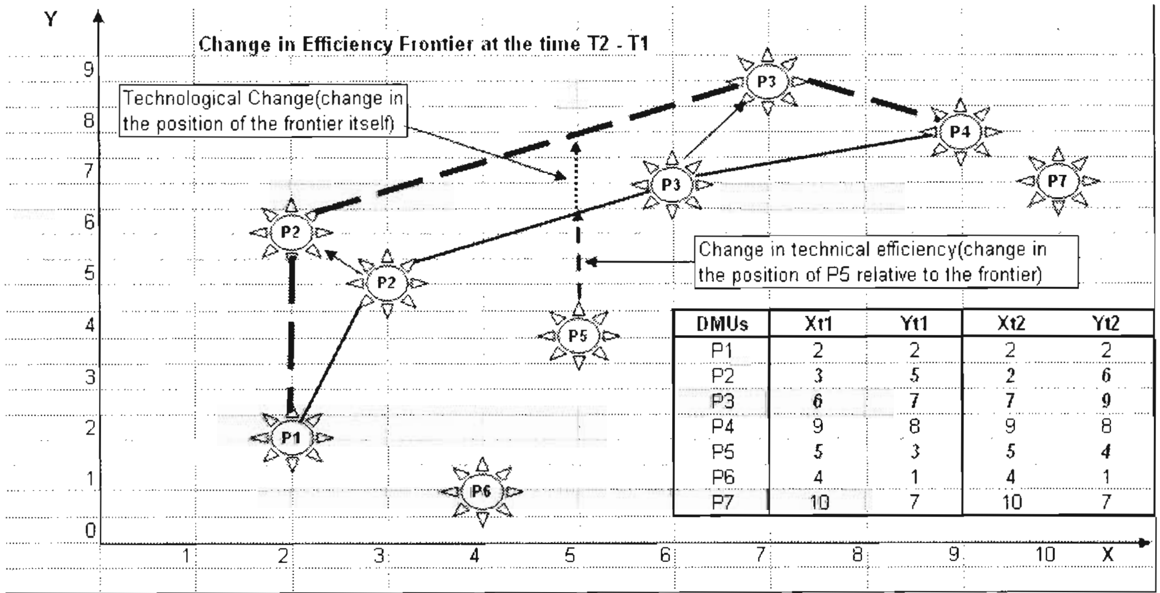


Figure 7 Efficiency frontier, efficient and inefficient DMUs

Then, for a given DMU, the period of time (t2-t1) could be represented by the distance between the data point at the time t1 and the data point at the time t2. For each DMU the distance between these data points would be reflective of the change in this DMU's TFP, which is, of course, is represented by the Malmquist index.

In the case of economic growth, we would expect that the efficiency frontier for a given set of DMUs would change its position over a period of time. Let us suppose that three DMUs, P2, P3 and P5, have changed their position over the period of time T2 –T1 (as depicted in the figure below). Such change is reflected by the new positions of these DMUs, as well as a new position of the efficiency frontier (represented by dashed line). As a result, for a given period in time, change in the position of each DMU could be perceived as consisting of the two components. The first component is the change in distance between a given DMU and the efficient frontier, which reflects the changes in technical efficiency, and the second is the change in position of the efficient frontier itself, reflective of the technological change that took place over (t2-t1) period of time. Figure 8 below depicts such changes in the frontier, as well as the change in the position of the DMU P5 relative to the frontier. In this case, change in the position of DMU P5 relative to the frontier, less the change in the position of the frontier itself, would be represented by a particular value of the Malmquist index for DMU P5.



DMUs	Xt1	Yt1	Xt2	Yt2
P1	2	2	2	2
P2	3	5	2	6
P3	6	7	7	9
P4	9	8	9	8
P5	5	3	5	4
P6	4	1	4	1
P7	10	7	10	7

Figure 8 Components of Malmquist Index: TC and EC

Formally, the Malmquist (output-oriented) TFP change index between period 0 (the base period) and period 1 could be defined (following is adapted from (Milana and Zeli 2002))by using the concept of distance functions $\boldsymbol{d^t(x,y)}$.

In the case of output-orientation distance function is defined as $d^t(x,y) = \min\{d : (y/d) \in A^{-t}(x)\}$, where $A^{-t}(\mathbf{x})$ is the set of all possible levels of the output **y** for a given technology **t** and the input level **x** (Milana and Zeli 2002).

Consequently, the Malmquist (output-oriented) TFP change index between period 0 and period 1 is given by

$$TFP_{\mathrm{m}}(y_0,x_0,y_1,x_1) \equiv \left[\frac{d^0(y_1,x_1)}{d^0(y_0,x_0)} * \frac{d^1(y_1,x_1)}{u\ (y_0,x_0)} \right]^{1/2}$$

Which is equivalent to:

$$TFP_{\mathrm{m}}(y_0,x_0,y_1,x_1) = EC*TC = \frac{d^1(y_1,x_1)}{d^0(y_0,x_0)} \left[\frac{d^0(y_1,x_1)}{d^1(y_1,x_1)} * \frac{d^0(y_0,x_0)}{u\ (y_0,x_0)} \right]^{1/2}$$

Where EC is an index of efficiency change between periods 0 and 1

$$EC \equiv \frac{d^1(y_1,x_1)}{d^0(y_0,x_0)}$$

And TC is an index of technological change between periods 0 and

$$TC \equiv \left[\frac{d^0(y_1,x_1)}{d^1(y_1,x_1)} * \frac{d^0(y_0,x_0)}{u\ (y_0,x_0)} \right]^{1/2}$$

(Milana and Zeli 2002)

DEA, Summary

We have described how by using DEA we are able to obtain values of changes in TFP for each of the DMUs used in the analysis. The choice of DEA for our purposes appears to be warranted, for this technique has been used widely in the research investigating the impact of IT on productivity (references to be added).

There are multiple inherent advantages and disadvantages to using DEA, as well as multiple endemic strengths and weaknesses of the technique itself. Thus, for example, it would appear that DEA would require homogeneity of the DMU within a given sample, for this type of analysis makes sense only if one determines relative efficiency of the similar DMUs. It is also quite obvious that the presence of any outliers within a set of DMUs would be detrimental to analysis due to inevitable distortion of the enveloping surface. In these two cases the problems lie not within the technique of DEA itself, but are inherent to the data set that is the subject to analysis.

Thus, to maximize the benefits offered by DEA we are going to analyze the data set that we use with the purpose of determining whether or not it contains the obvious outliers and whether the countries of TEs appear to be homogenous. In order to do so we are going to use Cluster Analysis, overview of which is presented next.

6.3.2 Cluster Analysis

Cluster analysis(CA) (Tryon 1939) could be described as an analytic tool purpose of which is to organize, or partition, a set of data into the number of subsets in such a way, that the similarity within the observations within a subset is high, while the similarity between the observations from the different subsets is low. Commonly, variance is used as main tool for identification and classification of the data points into the clusters, such as variance within each cluster is minimized and the variance between the clusters is maximized. In our case, the original data set is represented by the group of countries of transition economies and the purpose of clustering would be to determine if the original group could be partitioned in a number of subgroups, so there are substantial differences between the subgroups, but the individual countries within a single group are similar.

There are three general approaches to clustering. These are hierarchical clustering, k-means clustering, and two-step clustering. We provide the brief overview of each approach next.

Hierarchical clustering could form clusters by one of the two methods, agglomerative or divisive. Agglomerative method assumes that each data point is its own cluster, and with each step of the clustering process, these clusters are combined to form larger clusters, which are eventually combined to form a single cluster. Divisive method of the hierarchical clustering, on the other hand, starts with the single cluster encompassing all data points within the sample and proceeds to divide it into the smaller dissimilar clusters. Agglomerative method of hierarchical clustering appears to be the most appropriate type of clustering for the purposes of our research. We expect that all the transition economies within our data set are relatively similar. Thus, as agglomeration of data sets together proceeds, we expect to see the reduction of the number of clusters with the proportional increase of the membership within each cluster. If, on the other hand, we end up at the last stages with two clusters, one containing a single data point, and another containing the rest of the data set, it may reasonably suggest that the outlier is present. Hierarchical clustering could be depicted as a two dimensional diagram, commonly called *dendrogram*, which shows the agglomerations or divisions made at each successive stage of the cluster analysis. An example of a dendrogram is presented in the Figure 9 below.

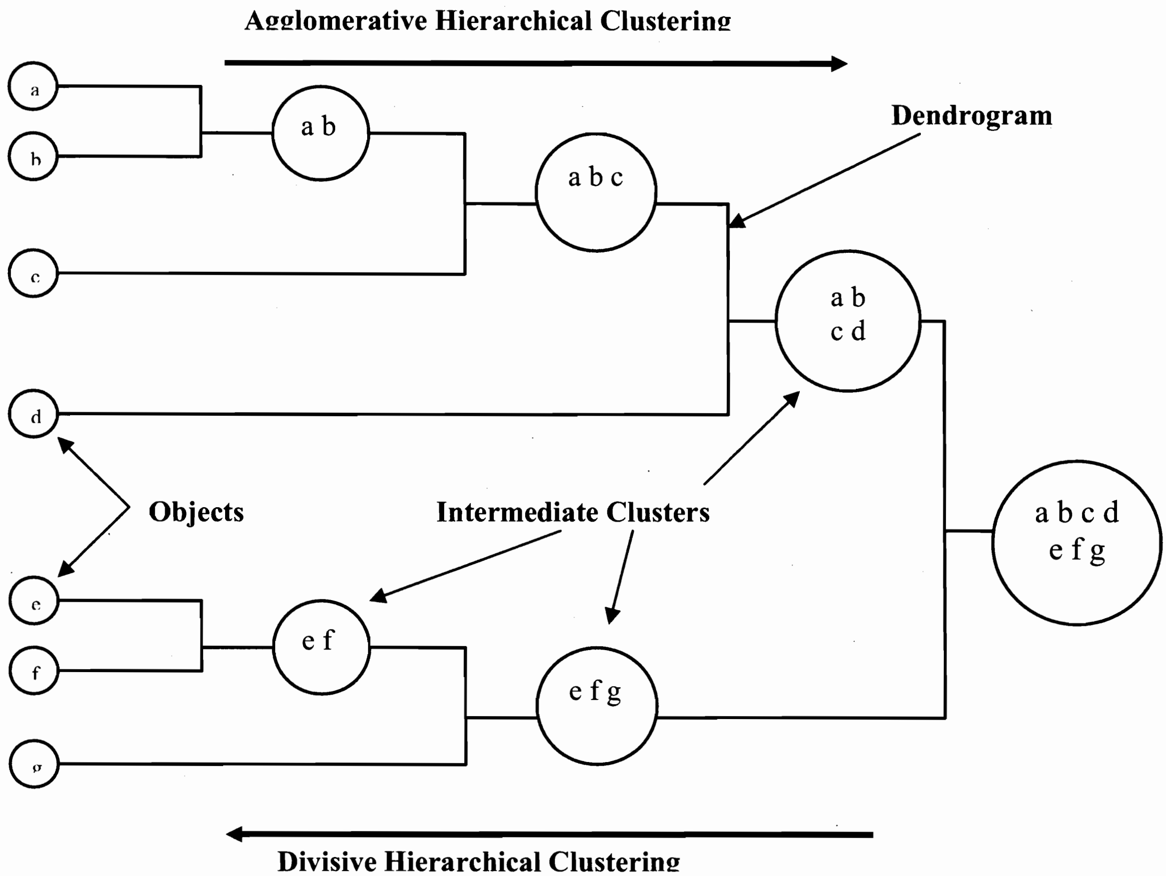


Figure 9 Hierarchical clustering

Unlike hierarchical clustering, K-means clustering requires the number of resulting cluster, k, to be specified prior to analysis. Thus, k-means clustering will produce *k* different clusters of greatest possible distinction. This method of clustering could be useful in the case if our data set included countries of the different types, i.e., developed, developing, transition economies, and least developed countries. Consequently, it would have been reasonable to pre-specify number of clusters according to the establish classification types. In our case, however, we do not have any apparent reasons to specify

k as being anything but 1. An example of K-means clustering, with K=2 and K=3, is provided below in the Figure 10.

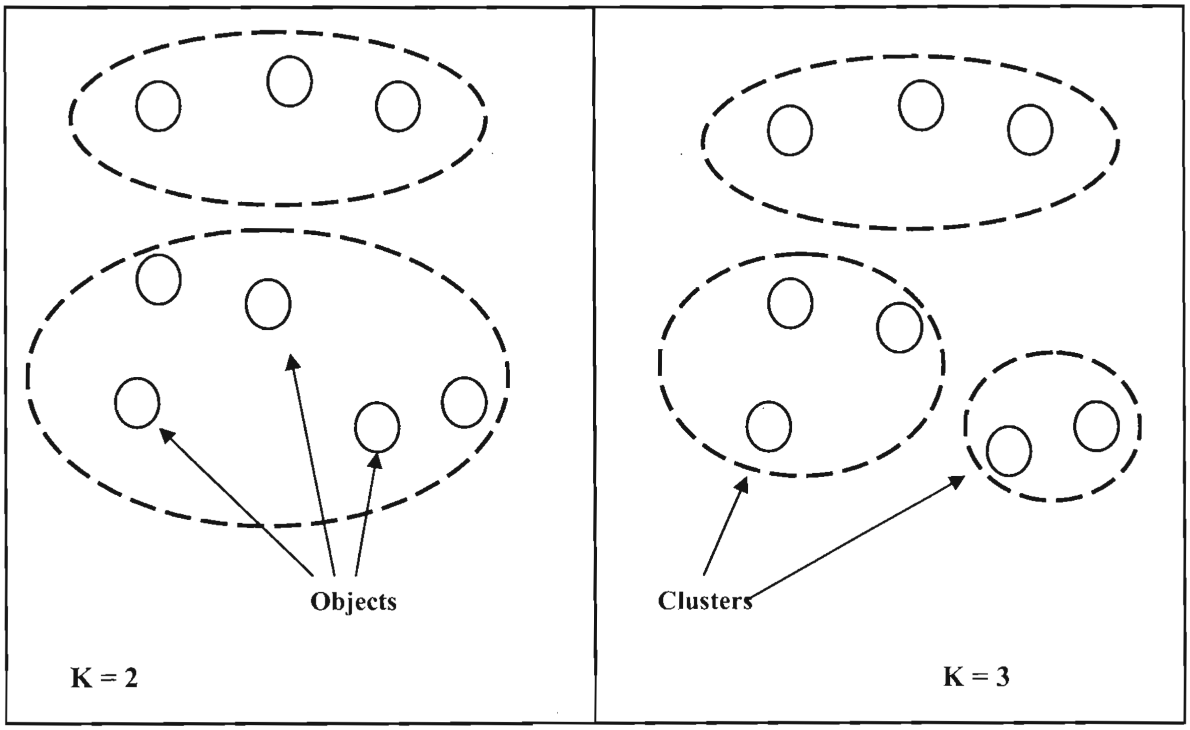


Figure 10 K-means clustering

Two-way (or two-step) cluster analysis based on the combining of the multiple data points into the single clusters, which are then treated as a single data point and subjected to the hierarchical clustering. This type of clustering could have been useful in our research if, aside of having the data set containing multiple types of countries, we were concerned with identifying and grouping different types of countries together based on a specific characteristic. For example, two-way clustering could have been useful if we were interested in combining in a single cluster ICT-producing countries. Such pre-clustering would have resulted in members of the developed countries, such as USA, being grouped together with the developing countries of South-East Asia.

Prior to cluster analysis it is usually required to specify measure of distance that is going to be used in formation of the clusters. While there are multiple distance measures could be used, we are going to utilize Euclidian distance, the most commonly chosen type.

However, despite our intention to identify a particular, advantageous for our specific purpose type of clustering, it does not guarantee that we shall end up using it. The simple reason is that it is not clear what makes a particular clustering approach to be effective. Depending on the size of the sample and the number of variables used in Cluster Analysis, a specific clustering method can perform extremely well on one data set, while very poorly on another.

6.3.3 Multiple Regression

Multiple Regression: General Overview

The purpose of the *Multiple Regression* (MR)(Pearson 1908) is to model the relationship between a single dependent variable and the multiple independent variables. Generally, MR procedure estimates a linear relationship within the following model:

$$Y = a + b_1 * X_1 + b_2 * X_2 + b_3 * X_3 + \ldots + b_n * X_n + e$$

Where,

Y= dependent variable

a = intercept

b_k = slope coefficient

X_k = independent variable

e = error term

Unlike the *General Linear Model* (GLM), purpose of which is to test the relationship between the dependent and a single independent variable of the model

$$Y = a + a + b_1 {*} X_1 + e$$

the interpretation of the MP is not as straightforward. While GLM models dependent variable as a function of the intercept and the product of a slope and the value of the independent variable, MR relates each independent variable in the partial fashion to a dependent variable. Meaning, a coefficient of each independent variable represents partial contribution to the dependent variable, while controlling for the rest of the independent variables in the MP equation.

As a result, unlike the GLM that directly correlates the independent and dependent variables, MR correlates the dependent and independent variables in a partial fashion, which is often referred to as a *partial correlation* (Yule 1907).

Multiple Regression: Interaction Effects

In some cases while using MR, we are interested in the interaction effect of the independent variables on the dependent variable. In this case, the general model of MR takes form of:

$$Y = a + b_1 {*} X_1 + b_2 {*} X_2 + b_3 {*} X_1 X_2 + \ldots + b_n {*} X_n + b_k {*} X_k + b_m {*} X_n {*} X_k + e$$

And the test for interaction amounts to testing the null hypothesis

$$H0: b_3 = 0;$$

And in the case of $b_3 \neq 0$ we are able to reject the null hypothesis of no interaction between X_1 and X_2.

The interpretation of the interaction term in MR, however, is not as straightforward as the interpretation of the slope coefficient of an independent variable. For example, b_3 in the equation above reflects the relationship between Y and X_1 and X_2 when X_1 and X_2 increase *jointly*. Furthermore, b_3 in the equation above reflects conditional relationship between Y and X_1 and X_2, for the impact of X_1 on Y would depend on the level of X_2 and vice versa.

For the purposes of this research, we are interested only in testing the null hypothesis of no interaction between the investments in ICT and other types of investments. As a result, we are not going to inquire into such potentially affecting the interaction term issues as the presence of thresholds, level-dependent dynamic of the interacting variables, and so on.

The brief overview that we have provided in this section is not intended to reflect the true complexity of this topic. Consequently, any reader who is interested in the subject of interpreting the interactions in MR, would be well-served by referring to such sources as Jaccard et al. (1990), Aiken and West (1991), and Braumoeller (2004).

6.3.4 Decision Trees

DT: General Overview

Few of the data mining tools are as widely used as Decision Trees (DT) for the purposes of classification and prediction. DT got its name because the visual representation of a classificatory or predictive DT model resembles an upside down tree. The process of the creation of the DT model, or "growing of the tree," is called *DT induction*, and is based on the algorithmic partitioning of the data set into the multiple subsets. There are number of the algorithms that are commonly used in DT induction, the most popular being CART (Classification and Regression Trees) (Breiman et al 1984), CHAID (Chi-squared Automatic Interaction Detection), ID3 (Quinlan 1986), C4.5(Quinlan 1993), and C5.0(Quinlan 1998). When a given algorithm is applied to a data set, the result is represented in the form of the tree that depicts a path along which the partitioning took place.

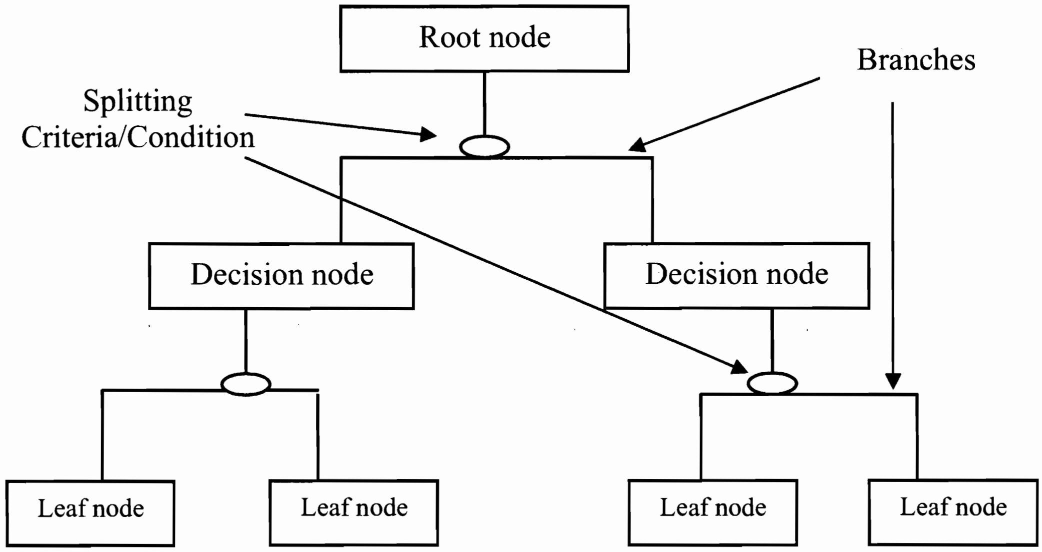


Figure 11 Basic structure of a Decision Tree

Structurally, the resultant DT model could be perceived as consisting of the four types of the components, which are *root node, leaf nodes, decision nodes* and the *branches* that connect the nodes to form a tree-like model. The Figure 11 above depicts the general structure of a DT.

At the top of the DT is a *root node*, which represents the starting point (in the form of the complete data set) from which the building of the DT model begins. Once an algorithm applied, the data set becomes gradually partitioned more and more according to the specific *splitting rules*. The point where each partitioning takes place called a decision node, for it is a point where the "decision" to partition data is made. Most often this splitting of the data into the subsets is based on the value of a single independent variable from the data set (therefore, data sets with many independent variables generally tend to yield larger trees), however, multivariate splits are possible as well.

Despite that, there are no set in stone rules regarding the number of branches that each split could produce, most of the algorithms yield *binary trees*, where the each split made in the decision node produces two branches. While this splitting restriction does affect representation of the model (binary trees tend to be taller and non-binary DTs tend to be more compact), it produces no effect on the modeled by the chosen algorithm structure of the data set.

This process of DT induction takes place until the stopping criteria prevent further splitting of the data set. Most of the algorithms use several *stopping rules*, plus, some additional stopping rules could be added by the model builder. The nodes where stopping criteria prevents further partitioning of the data set are called *leaf nodes* or *terminal nodes*.

Thus, because the process of DT induction basically proceeds in accordance with the splitting and stopping criteria, the use of the different criteria would result in the different DT models, even if they are applied to the same data set. Once the decision tree is constructed, it is not only presented in an easily understandable visual form (see Figure 12), but could be as well easily converted into the equivalent, yet more readable set of the *decision rules*.(depicted in Figure 13).

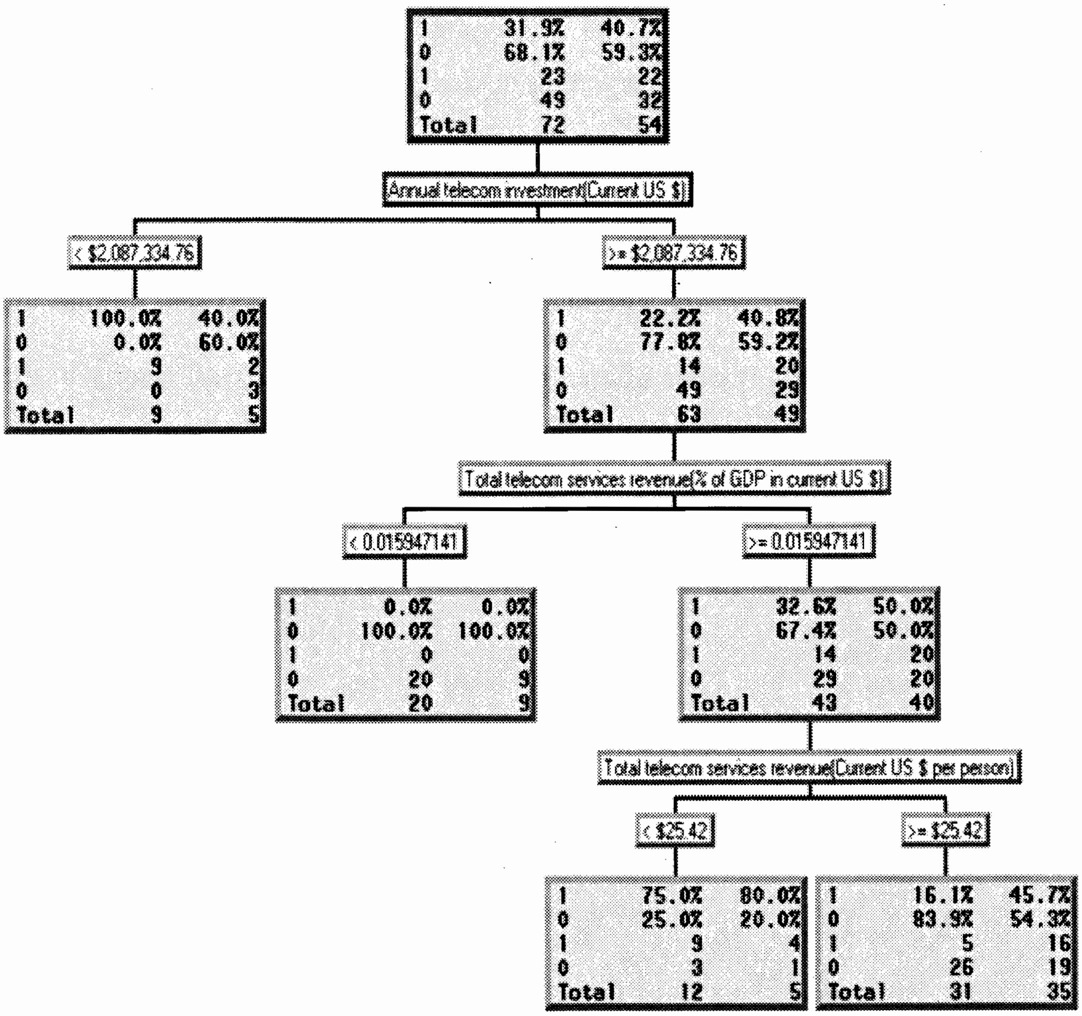


Figure 12 Representation of the Decision Tree

This conversion takes place through the tracing of the every partitioning path from the root node to each of the leaf nodes. As a result, every DT model would yield, at most, the same number of the decision rules as there are leaf nodes. In the case of the large trees with many leaf nodes, it is often possible to combine some rules, reducing, therefore, the total number of the decision rules produced by the model.

Nevertheless, the reduction of the total number of the rules does not solve the endemic problem of the large DTs, namely, their tendency to overfit the data and produce the models that have low generalizability beyond the data set at hand. Thus, development of the algorithms that would allow for smaller DTs is one of the challenges of the decision tree' induction. Specifically, Breiman et al. (1984) suggested that despite the importance of the good splitting rules offered by an algorithm, the overall quality of the tree is more dependent on the quality of the stopping rules.

```
IF  Annual telecom investment(Current US $) <         $2,087,335
THEN
 NODE  :    2
 N     :    9
 1     : 100.0%
 0     :   0.0%

IF  Total telecom services revenue(% of GDP in current US $) <  0.015947141
AND         $2,087,335 <= Annual telecom investment(Current US $)
THEN
 NODE  :    4
 N     :    20
 1     :   0.0%
 0     : 100.0%

IF  Total telecom services revenue(Current US $ per person)
   <           $25
AND  0.015947141 <= Total telecom services revenue(% of GDP in current US $)
AND         $2,087,335 <= Annual telecom investment(Current US $)
THEN
 NODE  :    6
 N     :    12
 1     :  75.0%
 0     :  25.0%

IF             $25 <= Total telecom services revenue(Current US $ per
     person)
AND  0.015947141 <= Total telecom services revenue(% of GDP in current US $)
AND         $2,087,335 <= Annual telecom investment(Current US $)
THEN
 NODE  :    7
 N     :    31
 1     :  16.1%
 0     :  83.9%
```

Figure 13
Set of the Decision Rules corresponding to the Decision Tree in Figure 12

There are three general approaches that have been used to achieve smaller decision tree models. The first approach attempts to tackle the problem by suggesting different criteria to partition the data set at each decision node. The second approach is based on pruning of a decision tree during the process of DT induction, or after the process is complete. Finally, the *look-ahead* methods attempt to establish a split at a decision node by analyzing the classifiability of the examples resulting from the split (Kothari and Dong 2000).

Generally, the approaches that aim to control the size of a DT during the induction phases are often referred to as *pre-pruning*, in contrast to *post-pruning* activities that take place during the pruning phase. Despite taking the different paths to reduce the size of DT, most of the pre-pruning and post-pruning methods are guided by the single common criteria, the best generalization accuracy of the tree. Apart from pre- and post-pruning approaches, the most common method utilized for boosting of the generalization accuracy of the DT is through the partitioning of the data. Under this approach, the original data set is divided into the three subsets, *training*, *validation*, and *test*. Training set is then used for preliminary DT induction, or, *training of the model.* Once the preliminary model is created, it is then tuned up using the validation data set. Finally, the resultant model is then assessed and tested using the holdout subset of the test data.

However, there are situations when other performance measures such as stability, interpretability and simplicity of a DT could become important and must be considered along with the generalization accuracy as well. For such cases, the multi-criteria approaches for evaluating DTs (Osei-Bryson 2004) would provide the greater benefit and should be preferred over the single criterion approach.

Use of DT for Classification and Prediction

Decision Tree modeling could be used in cases when the dependent variable of the data set is categorical (i.e., "1" or "0", as in "efficient" or "inefficient") or continuous (i.e, any value between "0" and "1"). Classification Tree' models are constructed in the cases when the dependent variable of the data set is categorical, while regression trees are used in the cases of the continuous dependent variable. If used for the purposes of classification, DT allows predicting the membership of a particular case to a group, while in the case of regression DT predicts a value.

In the case of this research, we could potentially use both versions of the DT, for the dependent variable of the data set is the efficiency of a given DMU (i.e., transitional economy). In the case of the using direct scores provided by DEA, we would use Regression Tree to create a predictive model. However, we as well could transform the assigned scores to create a classificatory model. Let us recall that the assignment of the scores by DEA is essentially dichotomous, i.e., a DMU is either found to be efficient, or not. In this case we could assign the inefficient score of "0" to all transitional economies that received any score but "1".

The decision to use regression or classification model would affect the conditions of the splitting criteria. For example, in the case of the continuous dependent variable (i.e., the values of the scores as assigned by DEA) each condition would split the data set using “≤”, “≥”, “>”,or “<” operators, while in the case of the categorical “efficient”/“inefficient” dependent variable split would occur based on the condition of equality (i.e., “=”).

One of the criteria that are often used to evaluate a DT model is *goodness of fit.* And while both, classification and regression trees utilize this criterion, they utilize the different measures to express it. In the case of the classificatory model *misclassification rate* is used as a measure of goodness of fit, while the *residual sum of squares* often chosen as a measure to evaluate the quality of the regression trees. Nevertheless, despite having different goals (i.e., classification vs. prediction) and using different measures of goodness of fit, both models, given their use of the same algorithm, would employ the same methods to construct the tree, thus, both models would have the same form of the tree.

6.3.5 Structural Equation Modeling implemented with Partial Least Squares method

SEM: General overview

Structural Equation Modeling (SEM) is a methodology that could be referred to as representing the second generation of multivariate analysis (Fornell 1987). Unlike the statistical tools of the first generation, exemplified by such techniques as cluster analysis, multiple regression, principal component analysis and others, SEM, according to Gefen et al. (2000), "enables researchers to answer a set of interrelated research questions in a single, systematic, and comprehensive analysis." Use of SEM allows researcher to posit the presence of the relationships between the multiple unobserved, or latent, variables, where every latent variable is associated with multiple observed variables, often called indicators or measures. Because the "ICT Investment Success" model of this study attempts to test the relationship between the multiple theoretical constructs, SEM is considered to be the appropriate methodology to use.

After the creation of the research model by an investigator, the process of SEM could be perceived as consisting of the two parts. The first part involves testing of the measurement model and primarily deals with the validation of the latent constructs included the model. Once validity of the measurement model has been established, testing of the structural model, the second part of SEM, takes place. Assessment of the structural model involves testing of the hypothesized relationships between the latent constructs of the research model. The results of the assessment are based on the significance of the structural paths, which could be estimated by using different methods, such as General Least Squares (GLS), Ordinary Least Squares (OLS), Maximum Likelihood Estimation (MSL), Partial Least Squares (PLS), and others. Basic structure of a SEM is depicted in the Figure 14 below.

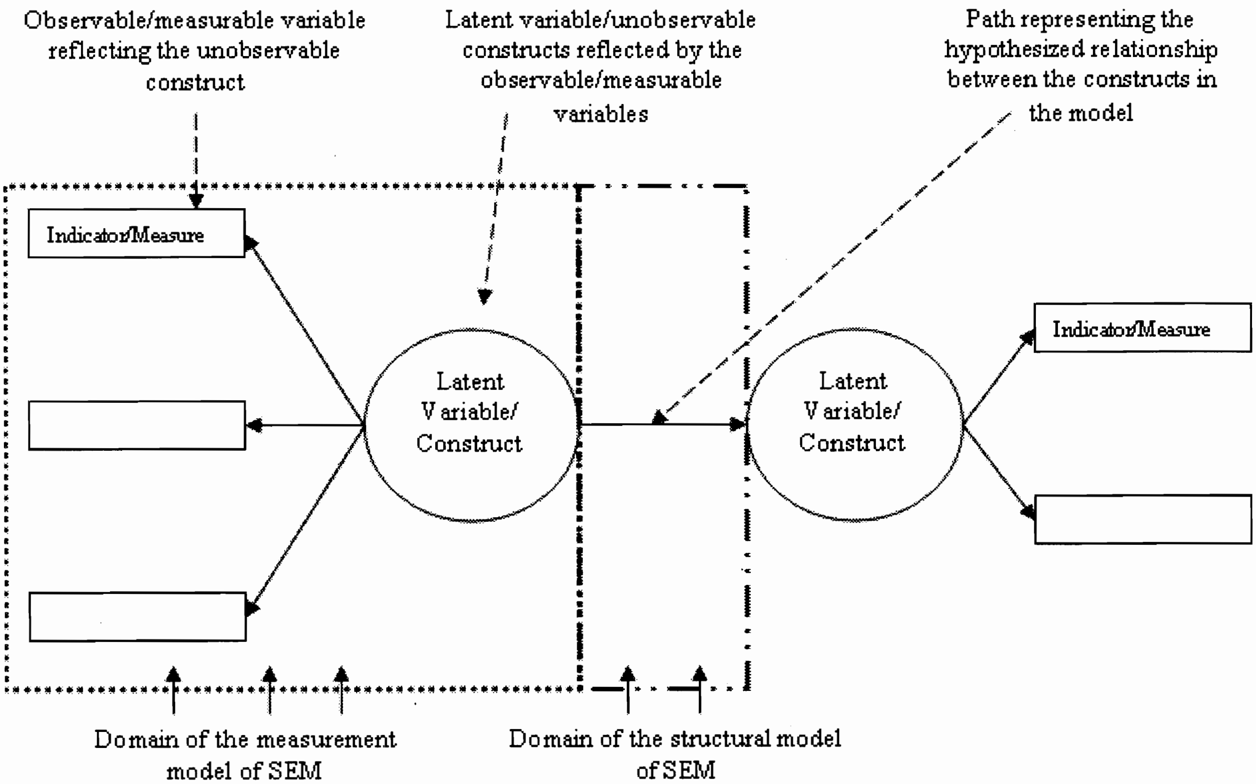


Figure 14 Basic Structure and Components of SEM

SEM: Two Common Approaches

There are two common approaches to SEM, covariance-based and variance-based. The covariance-based approach is based on the objective of minimization of the difference between the covariance matrix of the sample and the covariance matrix of the model. Thus, this approach is also commonly called factor-based, for the goal is to maximize the fit of the model by means of minimizing the unique variance. It is because of this goal of optimization of the fit the covariance-based approach is suitable for the situation "where prior theory is strong and further testing and development is the goal" (Chin 2000).

In contrast to covariance-based approach that attempts to optimize fit, variance-based approach attempts to optimize predictive capability of the research model relative to the sample. The optimization of the prediction is achieved by estimating, as close as possible, the parameters of the model by means of the minimization of the residual variances of the variables in the model (Chin 1998a) Thus, this method is commonly referred to as component-based, for "under this approach it is assumed that all the measured variance is useful variance to be explained" (Chin 1997). One of the methods that could be used to estimate parameters in covariance-based SEM model is Partial Least Squares (PLS), overview of which is presented next.

PLS: Brief Overview of Benefits

Partial Least Squares method , originally introduced by Wald (1966), is one of the least restrictive extensions of the multiple linear regression models that has the minimal demands on measurement scales, sample size, and residual distribution (Chin 1997). While covariance-based methods are more appropriate when research model is supported by the strong theory and well developed measures, PLS is recommended "for predictive research models where the emphasis may be more on theory development" (Barclay et al. 1995).

Two commonly mentioned benefits of PLS method refer to the size and distribution normality of the sample. Regarding the sample size, Chin (1995) explains that because "PLS is a limited information procedure, an appropriate sample size tends to be much smaller than that needed for a full information procedure," such as a covariance-based approach. As for the sample data distribution, Barclay et al. (1995) note that PLS does not require normally distributed data sample.

Another benefit of PLS refers to, albeit indirectly, its convenience to an investigator. This estimation method is now available to a researcher as a GUI-based PLS-Graph software package (Chin 1998b). Besides its intuitive interface, which makes PLS-G easy to learn and use, this package offers bootstrapping option for significance testing, adequate reporting capabilities, while being computationally efficient.

PLS: Why It Is Appropriate for This Study

We have decided to use PLS methodology to investigate our research problem for following reasons. First, the measures that we use to represent our constructs are not well developed. Second, our research model is at the early stage of the development. Thus, we operate under the condition of the low theoretical knowledge, which is not uncommon for the type of study that we undertake (Barclay et al. 1995). In regard to these points Chin (1995) notes that

> *"When using PLS, low theoretical knowledge does not necessarily imply a researcher's inability to define construct nor the nomological network in which*

these constructs reside. Instead, it likely depicts an exploratory stage where a researcher is testing and ad hoc model with newly developed items."

This is in agreement with Joreskog and Wold (1982) statement that PLS is primarily intended for the analysis in the situations of low theoretical knowledge.

Third, the data is our sample may not have a normal distribution. While this would violate one of the fundamental assumptions in other methods (i.e., Maximum Likelihood used in LISREL) and thus preclude us from employing them, PLS does not rely on the assumption of the normally distributed data (Barclay et al. 1995).

Finally, another reason why we have chosen PLS pertains to the sample size. In our case, where we are dealing with 18 TEs over 10 year period of time, our sample size is expected to be small (less than 180 cases). Moreover, the sample size would decrease even further when dealing with the 5-year period. Even by itself, this issue of the small sample size would completely preclude us from using any other SEM method (i.e., LISREL).

6.3.6 Neural Networks

Neural Networks (NN) is a rather diverse group of techniques that are primarily used in the modeling of the unknown complex relationships in the data. The most common architecture of a NN, depicted in the Figure 15 below, could be represented as consisting of the three layers: input, hidden, and output layers, where each layer consists of one or more nodes. This type of architecture, originally introduced by Rumelhart and McClelland (1986), is referred to as *Multilayer Peceptron* (MLP).

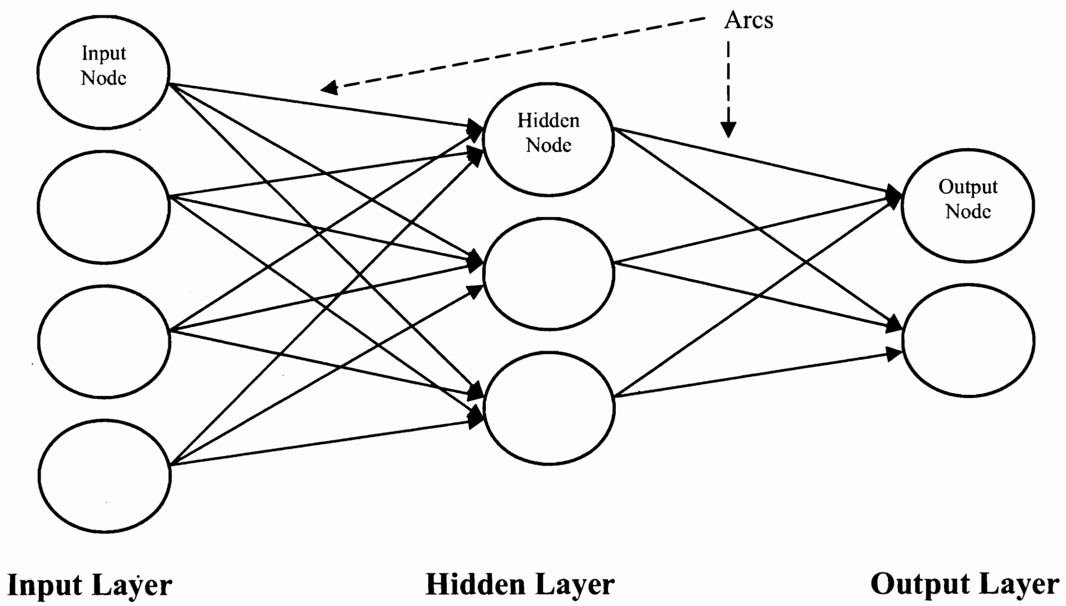


Figure 15 Basic structure and components of the Neural Network

An input layer of NN consists of the input units or *input nodes*, representing the input of the data into the NN. The hidden layer consists of the hidden units or *hidden nodes*, purpose of which is to determine and assign weights to the connections between the input nodes and hidden nodes. Finally, the output layer consists of one or more output units or *output nodes*, which represent the output of the NN. A common taxonomy used to describe NN is based on the three criteria.

The first criterion is the architecture of a NN, which refers to the type of connection between the nodes of the network. The type of connection that allows for the signals between the nodes to travel in only one way, forward, is called feed-forward connection. The NNs using this type of connection are called *feed-forward networks*. Another type of connection allows for the presence of the loops between the nodes and is called feedback connection. The NNs using this type of connection are called *feedback networks*.

The second criterion refers to the learning process of NN. In general, learning takes place in NN by means of changing the weights that the nodes of the hidden layer assign to the inputs. If the weights are fixed in advance and cannot be changed, then the NN is called *fixed network*. If the weights could be changed, the NN is called *adaptive network*. The adaptive networks are capable of changing the weights based on the two major types of learning: supervised and unsupervised. The *unsupervised learning* relies only on the local information in the form of the available data set and does not require the presence of an investigator to adjust the process of learning. The process of *supervised learning*, on the other hand, requires presence of the investigator, who affects the process of learning by supplying the global information in the form of the output target to the NN.

The third criterion refers to the *transfer function* that NN uses to convert inputs into the outputs. There are three major categories of transfer function: linear, threshold, and sigmoid. In the case of *linear function,* the output is in the linear relationship to the weighted inputs. The *threshold function* uses some threshold value applied to the input to determine the binary output. Thus, the output value would be dependent on whether the input is greater or less than the preset threshold value. Finally, the *sigmoid function* uses non-linear relationship between the input and output to determine the output of the NN.

Thus, not entirely unlike the Decision Trees, Neural Networks could be used, based on the chosen transfer function, for the purposes of classification or regression. In our case, we are interested in the use of the NN for the purposes of performing the regression task. Specifically, given two subsets of the data, our aim is to predict the output for the one subset using NN that was trained on the other subset. For the purposes of this research, we would use supervised mode of learning because the data set that we are going to use contains not only the inputs (indicators representing the construct "ICT Capitalization"), but also the outputs(indicators representing the construct "TFP").

This brief overview of such complex subject as NN cannot do justice to the topic. Thus, we direct the interested reader to Bishop (1995) for a comprehensive treatment of the subject.

CHAPTER 7 Results of the Data Analysis

7.1 Clustering

7.1.1 Clustering: Data

In this section, we describe the results of the cluster analysis that we have performed on the two data sets that were compiled from the available to us sources, Yearbook of Statistics and WDI database. The both sets include the data on the same 18 countries, with only difference being a length of the time series used for each set. While the first data set includes the data on the 5-year-period from 1998 to 2002, the second data set includes the data on the 10-year period from 1993 to 2002. The variables that we use for the cluster analysis listed and described below.

Country&Year- Represents the name of the country and a year, such that the "Albania1993" represents the data for Albania in the year 1993. Thus, Albania during a 5-year period from 1993- 2002 would be represented by the set of the data points {Albania1993,….., Albania2002}

Total telecom services revenue(Current US $)- Represents a converted in US $ total telecom service revenue for a given transitional economy, for a given year

Total telecom services revenue per person (Current US $) - A converted in US $ total telecom service revenue for a given transitional economy divided by the total population of a given country.

Total telecom services revenue per worker (Current US $) - A converted in US $ total telecom service revenue for a given transitional economy divided by the total labor force of a given country.

Annual telecom investment (Current US $) - Represents a converted in US $ total annual telecom investment for a given transitional economy, for a given year

Annual telecom investment per person (Current US $) - A converted in US $ total annual telecom investment for a given transitional economy divided by the total population of a given country.

Annual telecom investment (% of GDP) - Represents a percentage, relative to GDP, of the converted in US $ total annual telecom investment for a given transitional economy, for a given year. Obtained by dividing a converted in US $ total annual telecom investment for a given transitional economy, for a given year, by the country's GDP, in US $.

Annual telecom investment per worker (Current US $) - A converted in US $ total annual telecom investment for a given transitional economy divided by the total labor force of a given country.

Annual telecom investment per telecom worker (Current US $) - A converted in US $ total annual telecom investment for a given transitional economy divided by the number of full-time telecom employees of a given country.

Total telecom services revenue per telecom worker (Current US $) - A converted in US $ total telecom service revenue for a given transitional economy divided by the number of the full-time telecom employees of a given country.

Productivity ratio per telecom worker (revenue/investment) - Ratio of Annual telecom investment (Current US $ per telecom worker) to Total telecom services revenue(Current US $ per telecom worker)

The purpose that we pursue with cluster analysis is to inquire whether or not all 18 economies in transition that we study are, indeed, similar in regard to their investments in and revenues from ICT. If the results of the cluster analysis demonstrate that the 18 countries are not the same, we would like to further inquire whether the countries could shift their position from cluster to cluster with time, or, whether the countries are permanently bound to the cluster that they have been assigned to at the beginning.

Thus, two hypotheses could be generated prior to the cluster analysis regarding the outcome:

H01: 18 transitional economies are homogenous regarding their level of investments in and revenues from ICT.

We would reject H01 if it would be possible to achieve such result of cluster analysis, where there is more than one cluster and, given a set of data points representing a transitional economy over 5 and 10-year period of time, every cluster would contain a complete set of the data points representing a given economy.

H02: Relative to each other, 18 transitional economies maintain static relative level of investments in and revenues from ICT.

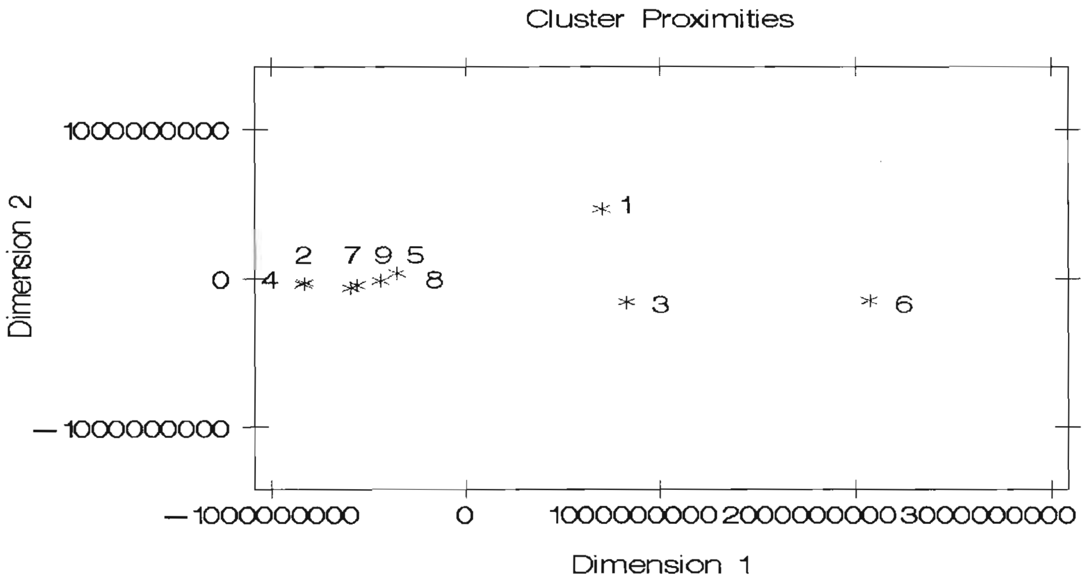
We would reject H02 if it would be possible to achieve such result of cluster analysis, so that all the data points that represent any given transitional economy from the sample would be assigned to the different clusters

We used SAS Enterprise Miner(EM) to perform cluster analysis of the two data sets. The variables that we decided to use are not measured on the same scale, so, prior to cluster analysis, according to the common practice, we have transformed the data by standardizing the variables. Then we have performed three sets of analysis, one for the first 5-year period (from 1993 to 1997), another one for the second 5-year period (1998-2002), and, finally, the third set for 10-year period (from 1993 to 2002). The results are presented below.

7.1.2 Clustering: Results

7.1.2.1 First 5-years data set (1993-1997)

First 5-years data set (1993-1997), Automatic setting



Automatic setting yielded 9-cluster solution, with the clusters 1, 3, and 6 being clearly removed from the conglomeration of the clusters 2, 4, 5, 7, 8, and 9.

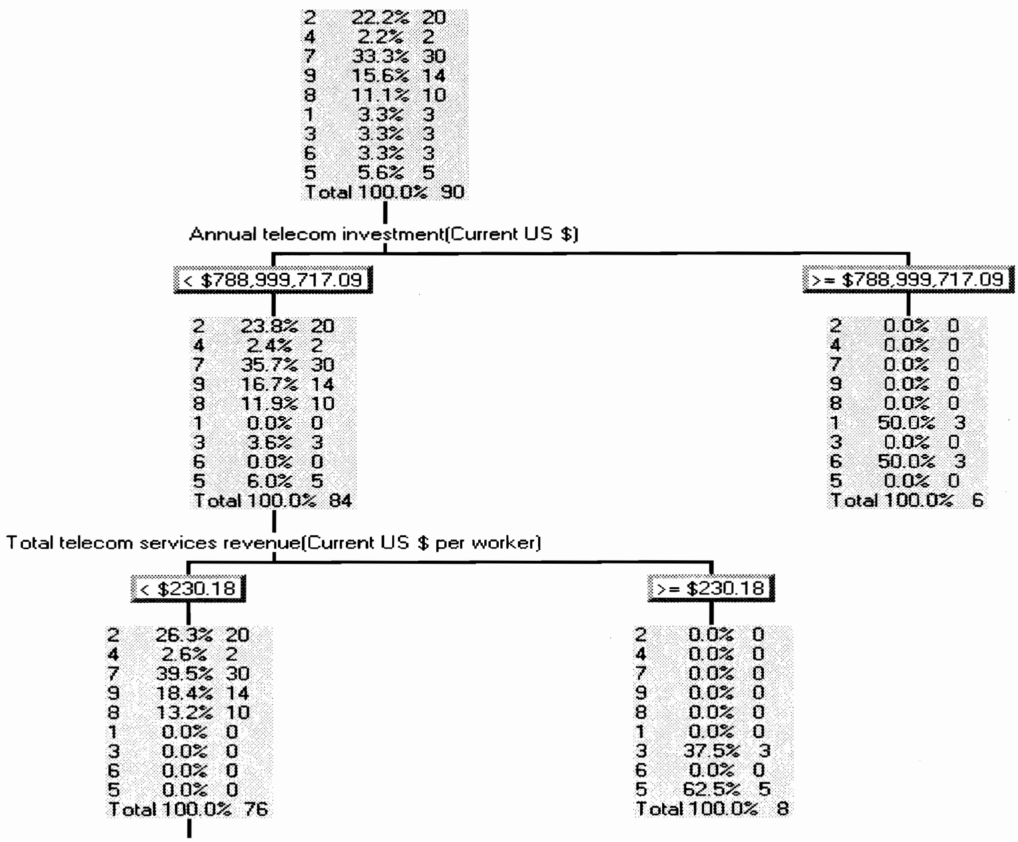
CLUSTER	Frequency of Cluster	Root-Mean-Square Standard Deviation	Maximum Distance from Cluster Seed	Nearest Cluster	Distance to Nearest Cluster	Total telecom services revenue(Current U
1	3	0.6998425936	2.9131015752	3	4.0448010779	1215146543.3
2	20	0.3222487868	2.2505782408	7	2.1699959716	35026556.072
3	3	0.3606273971	1.261267894	5	3.1715021858	1590473236.3
4	2	0.7259759632	1.7025645496	2	6.0794513147	36867867.021
5	5	0.3476510782	1.4179977778	3	3.1715021858	284330838.94
6	3	0.4507162016	1.5621759723	3	4.6578391346	2715412171.7
7	30	0.2717098631	1.4756153759	9	2.0782970203	263403667.19
8	10	0.531926549	2.4662418631	9	2.2251475824	439863615.49
9	14	0.5321557482	2.7590672214	7	2.0782970203	384039671.98

Results in the table above demonstrate the memberships of each cluster.

Name	Importance	Measurement	Type	Label
TOTAL_TELECOM_SERVICES_REVENUE_C	0	interval	num	Total telecom services revenue(Current US $)
TOTAL_TELECOM_SERVICES_REVENUE_0	0	interval	num	Total telecom services revenue(Current US $ per per
TOTAL_TELECOM_SERVICES_REVENUE_1	0.543509935	interval	num	Total telecom services revenue(Current US $ per wor
ANNUAL_TELECOM_INVESTMENT_CURREN	0.4548246341	interval	num	Annual telecom investment(Current US $)
ANNUAL_TELECOM_INVESTMENT_CURRE0	0	interval	num	Annual telecom investment(Current US $ per person)
ANNUAL_TELECOM_INVESTMENT___OF_	0	interval	num	Annual telecom investment(% of GDP)
ANNUAL_TELECOM_INVESTMENT_CURRE1	0.8223503269	interval	num	Annual telecom investment(Current US $ per worker)
ANNUAL_TELECOM_INVESTMENT_CURRE2	0.738151599	interval	num	Annual telecom investment(Current US $ per telecom
TOTAL_TELECOM_SERVICES_REVENUE_2	0	interval	num	Total telecom services revenue(Current US $ per te
PRODUCTIVITY_RATIO_PER_TELECOM_W	0	interval	num	Productivity ratio per telecom worker (revenue/inve
FULL_TIME_TELECOMMUNICATION_STAF	1	interval	num	Full-time telecommunication staff(% of total labor

In clustering of the data, three variables were considered to be important: Full-time telecom staff, Annual telecom investment per worker, and Annual telecom investment per telecom worker.

Decision Tree below depicts the partitioning of the data that was used to obtain 9-cluster solution.



Decision Tree for 9-cluster solution , continued from above.

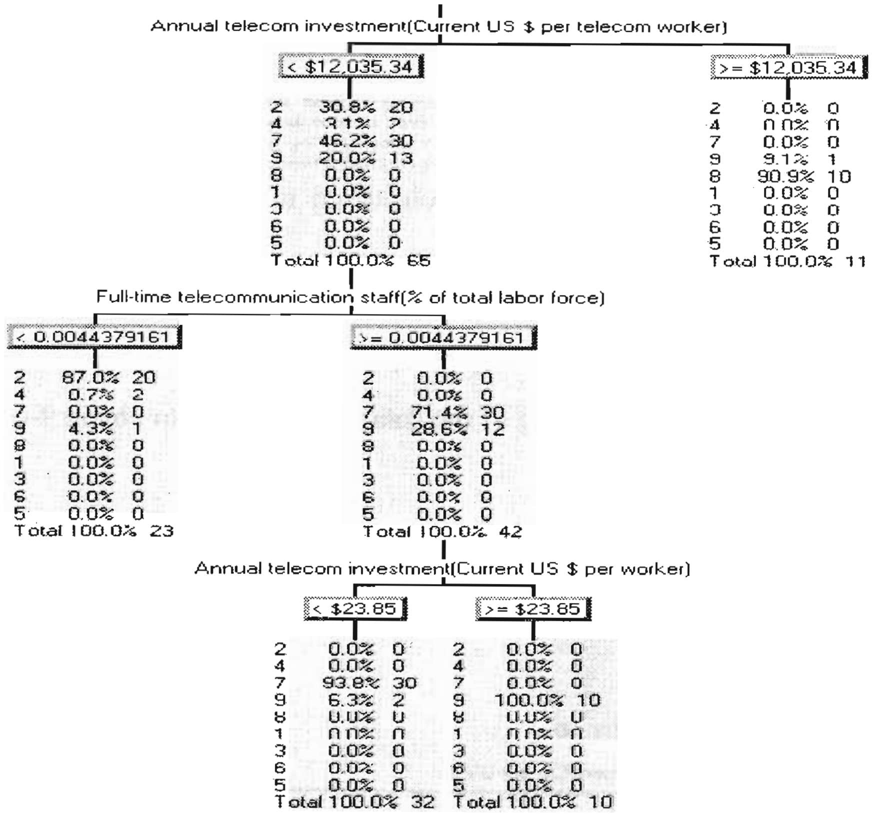


Table below contains the decision rules that were created based on the Decision Tree above.

IF $788,999,717 <= Annual telecom investment(Current US $)
THEN
NODE : 3
N : 6
2 : 0.0%
4 : 0.0%
7 : 0.0%
9 : 0.0%
8 : 0.0%
1 : 50.0%
3 : 0.0%
6 : 50.0%
5 : 0.0%

IF $230 <= Total telecom services revenue(Current US $ per

worker)
AND Annual telecom investment(Current US $) < $788,999,717
THEN
NODE : 5
N : 8
2 : 0.0%
4 : 0.0%
7 : 0.0%
9 : 0.0%
8 : 0.0%
1 : 0.0%
3 : 37.5%
6 : 0.0%
5 : 62.5%

IF $12,035 <= Annual telecom investment(Current US $ per telecom worker)
AND Total telecom services revenue(Current US $ per worker) < $230
AND Annual telecom investment(Current US $) < $788,999,717
THEN
NODE : 7
N : 11
2 : 0.0%
4 : 0.0%
7 : 0.0%
9 : 9.1%
8 : 90.9%
1 : 0.0%
3 : 0.0%
6 : 0.0%
5 : 0.0%

IF Full-time telecommunication staff(% of total labor force) < 0.0044379161
AND Annual telecom investment(Current US $ per telecom worker) < $12,035
AND Total telecom services revenue(Current US $ per worker) < $230
AND Annual telecom investment(Current US $) < $788,999,717
THEN
NODE : 8
N : 23
2 : 87.0%
4 : 8.7%

7 : 0.0%
9 : 4.3%
8 : 0.0%
1 : 0.0%
3 : 0.0%
6 : 0.0%
5 : 0.0%

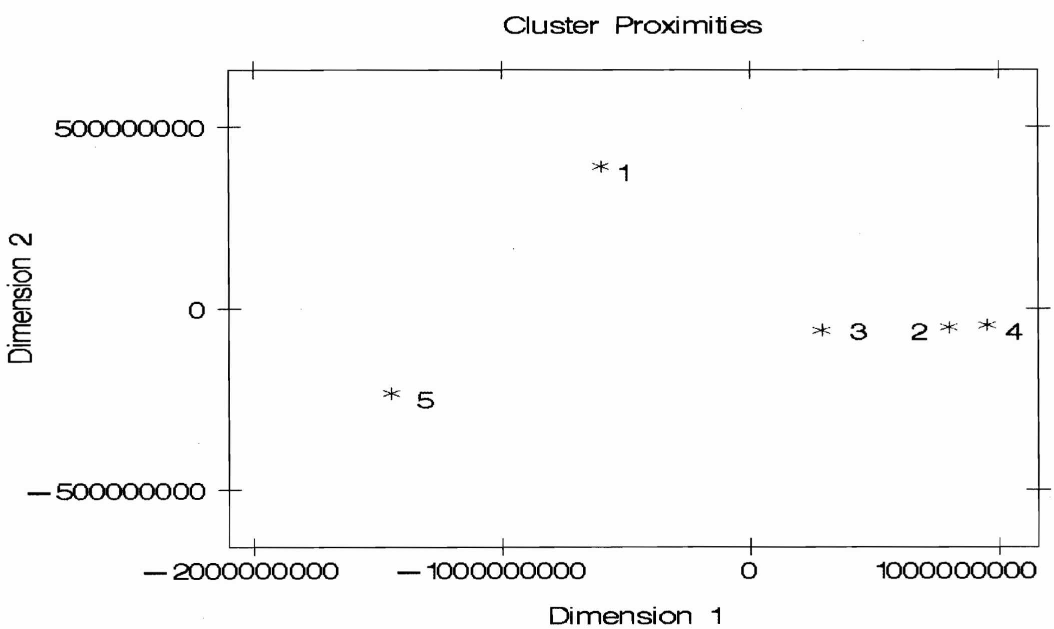
IF Annual telecom investment(Current US $ per worker) < $24
AND 0.0044379161 <= Full-time telecommunication staff(% of total labor force)
AND Annual telecom investment(Current US $ per telecom worker)
< $12,035
AND Total telecom services revenue(Current US $ per worker)
< $230
AND Annual telecom investment(Current US $) < $788,999,717
THEN
NODE : 10
N : 32
2 : 0.0%
4 : 0.0%
7 : 93.8%
9 : 6.3%
8 : 0.0%
1 : 0.0%
3 : 0.0%
6 : 0.0%
5 : 0.0%

IF $24 <= Annual telecom investment(Current US $ per
worker)
AND 0.0044379161 <= Full-time telecommunication staff(% of total labor force)
AND Annual telecom investment(Current US $ per telecom worker)
< $12,035
AND Total telecom services revenue(Current US $ per worker)
< $230
AND Annual telecom investment(Current US $) < $788,999,717
THEN
NODE : 11
N : 10
2 : 0.0%
4 : 0.0%
7 : 0.0%
9 : 100.0%
8 : 0.0%

1	:	0.0%
3	:	0.0%
6	:	0.0%
5	:	0.0%

Clearly, 9-cluster solution is not very conclusive, thus, we have decided to gradually decrease the number of clusters. From the automatic setting, we have proceeded to the 5-cluster solution, below.

First 5-year data set (1993-1997), 5 clusters



It would appear that the 5-cluster solution is somewhat similar to the 9-cluster solution, for it produces clusters (in this case 5 and 1) that are far removed from the majority of the data points.

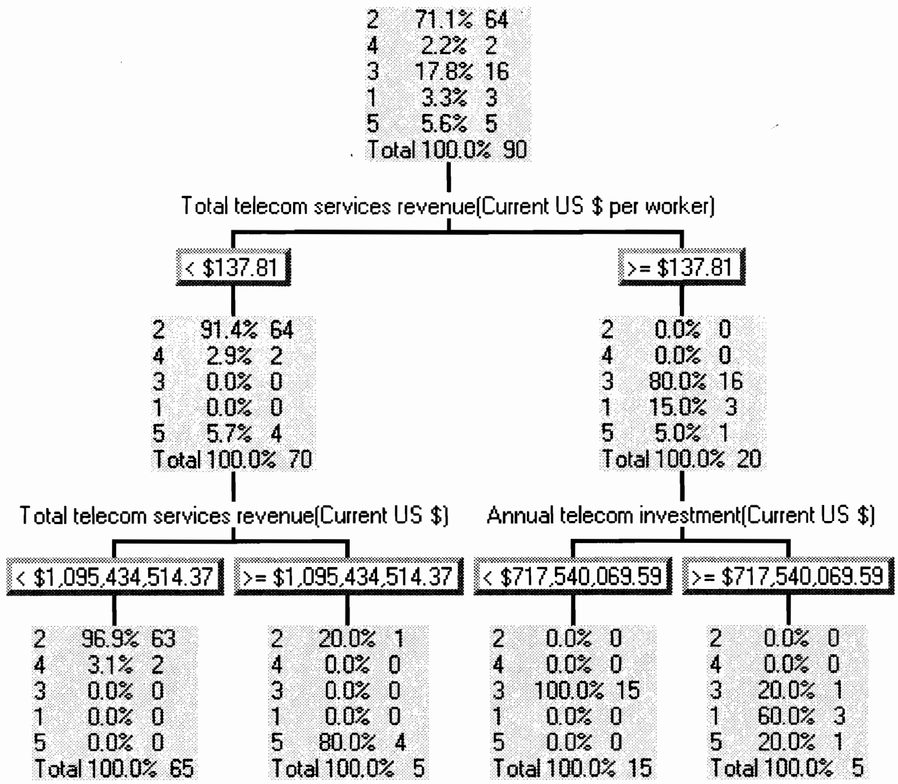
Name	Importance	Measurement	Type	Label
TOTAL_TELECOM_SERVICES_REVENUE_C	0.4953701428	interval	num	Total telecom services revenue(Current US $)
TOTAL_TELECOM_SERVICES_REVENUE_0	0	interval	num	Total telecom services revenue(Current US $ per pers
TOTAL_TELECOM_SERVICES_REVENUE_1	1	interval	num	Total telecom services revenue(Current US $ per worl
ANNUAL_TELECOM_INVESTMENT_CURREN	0.4089284618	interval	num	Annual telecom investment(Current US $)
ANNUAL_TELECOM_INVESTMENT_CURRE0	0	interval	num	Annual telecom investment(Current US $ per person)
ANNUAL_TELECOM_INVESTMENT____OF_	0	interval	num	Annual telecom investment(% of GDP)
ANNUAL_TELECOM_INVESTMENT_CURRE1	0	interval	num	Annual telecom investment(Current US $ per worker)
ANNUAL_TELECOM_INVESTMENT_CURRE2	0	interval	num	Annual telecom investment(Current US $ per telecom
TOTAL_TELECOM_SERVICES_REVENUE_2	0	interval	num	Total telecom services revenue(Current US $ per te
PRODUCTIVITY_RATIO_PER_TELECOM_W	0	interval	num	Productivity ratio per telecom worker (revenue/inves
FULL_TIME_TELECOMMUNICATION_STAF	0	interval	num	Full-time telecommunication staff(% of total labor

Partitioning of the data set was made based on the three variables (see table above), the most important of which was Total telecom service revenue per worker.

CLUSTER	Frequency of Cluster	Root-Mean-Square Standard Deviation	Maximum Distance from Cluster Seed	Nearest Cluster	Distance to Nearest Cluster	Total telecom services revenue(Current U	T
1	3	0.6998425936	2.5239631464	3	4.7058451621	1215146543.3	**
2	64	0.5295749423	3.7314510481	5	4.8487874807	176480665.98	**
3	16	0.7373524727	2.8809512242	5	4.4861687205	634554919.63	**
4	2	0.7259759632	1.7025645496	2	6.5997317192	36867867.021	**
5	5	0.5570569269	2.7625459645	3	4.4861687205	2253900356.3	**

After examining the information provided by Reporter node we have determined that Cluster 5 contains data points corresponding to Poland (1993-1997), Hungary (1994), and Czech Republic (1996, 1997).

The Decision Tree, reproduced below, demonstrates how the partitioning took place. Decision rules, based on the tree, provided below as well.



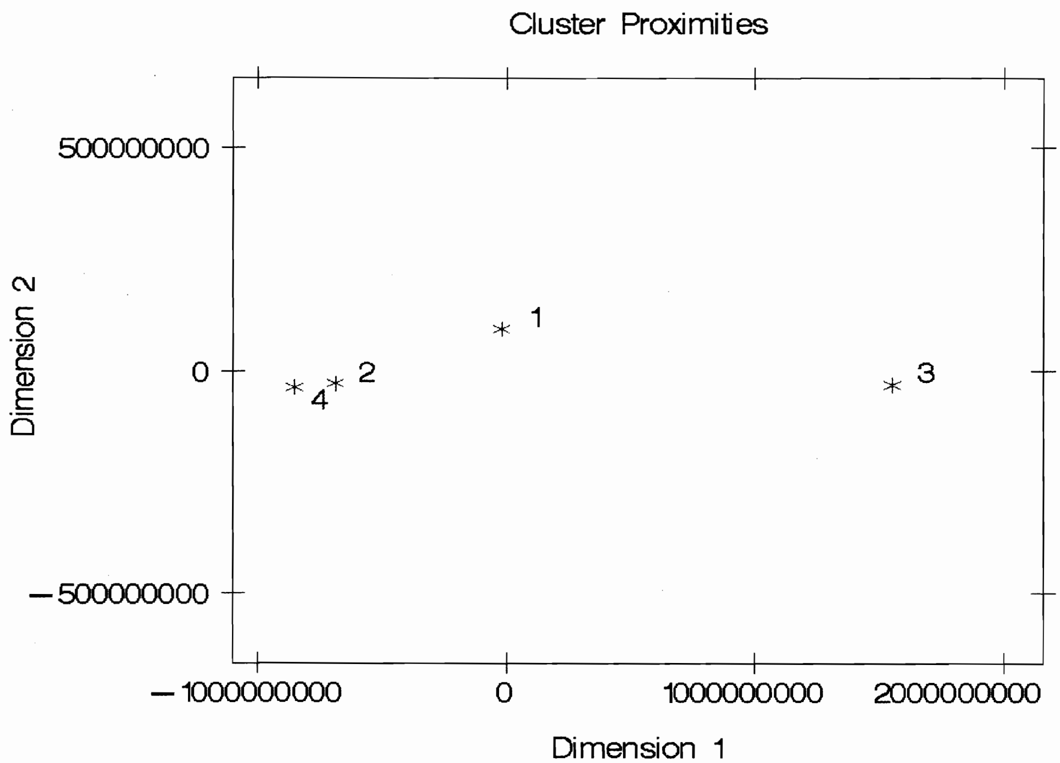
IF Total telecom services revenue(Current US $) < $1,095,434,514
AND Total telecom services revenue(Current US $ per worker)
< $138
THEN
NODE : 4
N : 65
2 : 96.9%
4 : 3.1%
3 : 0.0%
1 : 0.0%
5 : 0.0%

IF $1,095,434,514 <= Total telecom services revenue(Current US $)
AND Total telecom services revenue(Current US $ per worker)
< $138
THEN
NODE : 5
N : 5
2 : 20.0%
4 : 0.0%
3 : 0.0%

```
 1      :   0.0%
 5      :  80.0%

IF  Annual telecom investment(Current US $) <        $717,540,070
AND                $138 <= Total telecom services revenue(Current US $ per
     worker)
THEN
 NODE   :     6
 N      :     15
 2      :   0.0%
 4      :   0.0%
 3      : 100.0%
 1      :   0.0%
 5      :   0.0%

IF        $717,540,070 <= Annual telecom investment(Current US $)
AND                $138 <= Total telecom services revenue(Current US $ per
     worker)
THEN
 NODE   :     7
 N      :     5
 2      :   0.0%
 4      :   0.0%
 3      :  20.0%
 1      :  60.0%
 5      :  20.0%
```

First 5-year data set (1993-1997), 4 clusters

In this case we have obtained a 4-cluster solution, with somewhat consistent to the previous cases outcome. We could see that the two, removed from the majority of the data points, clusters move in the opposite directions. Cluster 1 is converging with the clusters 2 and 4, while cluster 3 is becoming farther removed from the rest.



Name	Importance	Measurement	Type	Label
TOTAL_TELECOM_SERVICES_REVENUE_C	0.4796603265	interval	num	Total telecom services revenue(Current US $)
TOTAL_TELECOM_SERVICES_REVENUE_0	0	interval	num	Total telecom services revenue(Current US $ per per
TOTAL_TELECOM_SERVICES_REVENUE_1	1	interval	num	Total telecom services revenue(Current US $ per wor
ANNUAL_TELECOM_INVESTMENT_CURREN	0	interval	num	Annual telecom investment(Current US $)
ANNUAL_TELECOM_INVESTMENT_CURRE0	0	interval	num	Annual telecom investment(Current US $ per person)
ANNUAL_TELECOM_INVESTMENT____OF_	0	interval	num	Annual telecom investment(% of GDP)
ANNUAL_TELECOM_INVESTMENT_CURRE1	0	interval	num	Annual telecom investment(Current US $ per worker)
ANNUAL_TELECOM_INVESTMENT_CURRE2	0	interval	num	Annual telecom investment(Current US $ per telecom
TOTAL_TELECOM_SERVICES_REVENUE_2	0	interval	num	Total telecom services revenue(Current US $ per te
PRODUCTIVITY_RATIO_PER_TELECOM_W	0	interval	num	Productivity ratio per telecom worker (revenue/inve
FULL_TIME_TELECOMMUNICATION_STAF	0	interval	num	Full-time telecommunication staff(% of total labor

As we could see in the table above, Total telecom service revenue per worker is again considered to be the most important variable in partitioning of the data set

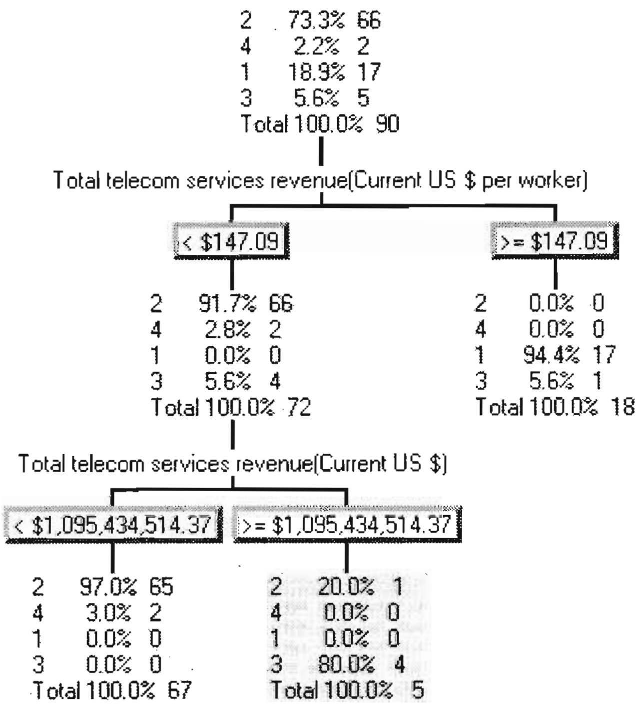
CLUSTER	Frequency of Cluster	Root-Mean-Square Standard Deviation	Maximum Distance from Cluster Seed	Nearest Cluster	Distance to Nearest Cluster	Total telecom services revenue(Current U	T
1	17	0.8893477386	5.9360703067	3	4.6378603964	754507281.76	*
2	66	0.5454882724	3.6935660532	3	4.7953452319	185855411.77	*
3	5	0.5570569269	2.7625459645	1	4.6378603964	2253900356.3	*
4	2	0.7259759632	1.7025645496	2	6.6137271274	36867867.021	*

By examining the report produced by Reporter node, we were able to determine the membership of the clusters 3 and 1.

Cluster 3 – Poland (1993-1997).

Cluster 1 – Czech Republic (1995-7), Estonia (1996-7), Hungary (1993-7), Slovak Republic (1996-7), Slovenia (1993-1997).

Corresponding to the 4-cluster solution Decision Tree is provided below.



Upon the examination of the tree, we could see that clusters 1 and 3 consist of the data points with the higher level of Telecom revenue than the clusters 2 and 4.

Set of the decision rules, corresponding to the DT, is provided below.

```
IF              $147 <= Total telecom services revenue(Current US $ per
      worker)
THEN
 NODE   :    3
 N     :    18
 2     :   0.0%
 4     :   0.0%
 1     :  94.4%
 3     :   5.6%
```

```
IF  Total telecom services revenue(Current US $) <      $1,095,434,514
AND Total telecom services revenue(Current US $ per worker)
   <              $147
THEN
 NODE  :     4
 N     :     67
 2     :  97.0%
 4     :   3.0%
 1     :   0.0%
 3     :   0.0%

IF       $1,095,434,514 <= Total telecom services revenue(Current US $)
AND Total telecom services revenue(Current US $ per worker)
   <              $147
THEN
 NODE  :     5
 N     :      5
 2     :  20.0%
 4     :   0.0%
 1     :   0.0%
 3     :  80.0%
```

First 5-year data set (1993-1997), 3 clusters

We continue gradually minimize the number of clusters. In this case, we have chosen 3-cluster solution. The diagram of distances is no longer available (must have more than 3 clusters to get a pictorial representation of the solution). However, we still could obtain the table containing the distances between the clusters(provide below).

CLUSTER	Cluster 1	Cluster 2	Cluster 3
1	0	285364799.37	1215365433
2	285364799.37	0	930669312.
3	1215365433.8	930669312.95	

Somewhat surprisingly, the most important variable in partitioning of the data becomes Annual telecom investment, not revenue, as in the solutions above.

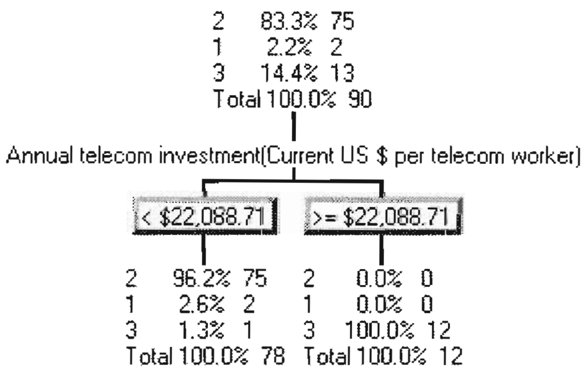
Name	Importance	Measurement	Type	Label
TOTAL_TELECOM_SERVICES_REVENUE_C	0	interval	num	Total telecom services revenue(Current US $)
TOTAL_TELECOM_SERVICES_REVENUE_0	0	interval	num	Total telecom services revenue(Current US $ per perso
TOTAL_TELECOM_SERVICES_REVENUE_1	0	interval	num	Total telecom services revenue(Current US $ per worke
ANNUAL_TELECOM_INVESTMENT_CURREN	0	interval	num	Annual telecom investment(Current US $)
ANNUAL_TELECOM_INVESTMENT_CURRE0	0	interval	num	Annual telecom investment(Current US $ per person)
ANNUAL_TELECOM_INVESTMENT____OF_	0	interval	num	Annual telecom investment(% of GDP)
ANNUAL_TELECOM_INVESTMENT_CURRE1	0	interval	num	Annual telecom investment(Current US $ per worker)
ANNUAL_TELECOM_INVESTMENT_CURRE2	1	interval	num	Annual telecom investment(Current US $ per telecom wo
TOTAL_TELECOM_SERVICES_REVENUE_2	0	interval	num	Total telecom services revenue(Current US $ per tele
PRODUCTIVITY_RATIO_PER_TELECOM_W	0	interval	num	Productivity ratio per telecom worker (revenue/invest
FULL_TIME_TELECOMMUNICATION_STAF	0	interval	num	Full-time telecommunication staff(% of total labor fo

Table below provides the membership for each cluster. We could see that cluster 1 contains only 2 data points (out of 90).

CLUSTER	Frequency of Cluster	Root-Mean-Square Standard Deviation	Maximum Distance from Cluster Seed	Nearest Cluster	Distance to Nearest Cluster	Total telecom services revenue(Current U
1	2	0.7259759632	1.7025645496	2	6.6876085173	36867667.021
2	75	0.6584285885	5.080063669	3	6.0155715706	294236653.68
3	13	0.9764468829	5.9319266929	2	6.0155715706	1099602594

Examination of the report yields the information that Cluster 1 contains the data points corresponding to Albania (1996, 1997) and Cluster 3 consists of Czech republic (1995-1997), Hungary (1994-7), Slovenia (1993-1997), Poland (1997).

Decision tree and set of rules for 3-cluster solution are provided below.



IF Annual telecom investment(Current US $ per telecom worker)
< $22,089
THEN
NODE : 2
N : 78
2 : 96.2%
1 : 2.6%
3 : 1.3%

IF $22,089 <= Annual telecom investment(Current US $ per telecom worker)
THEN
NODE : 3
N : 12
2 : 0.0%
1 : 0.0%
3 : 100.0%

First 5-year data set (1993-1997), 2 clusters

Distance between the two clusters is provided in the table below.

CLUSTER	Cluster 1	Cluster 2
1	0	813376548.2
2	813376548.23	

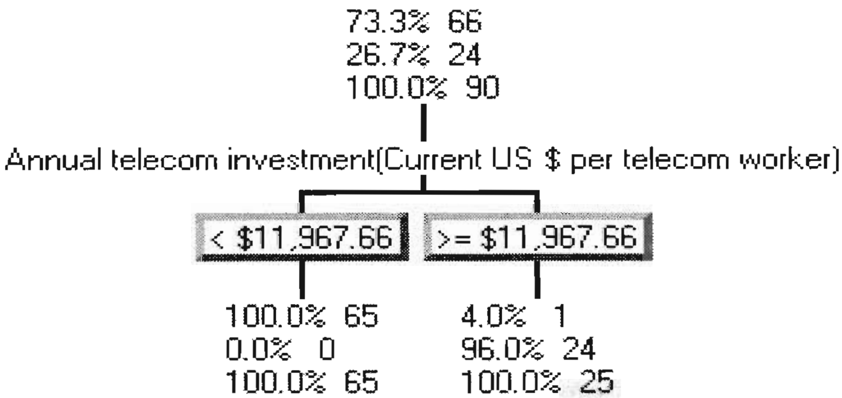
Having 2-cluster solution we could see that again, Annual telecom investment (this time per telecom worker) becomes the most important variable in partitioning of the data.

Name	Importance	Measurement	Type	Label
TOTAL_TELECOM_SERVICES_REVENUE_C	0	interval	num	Total telecom services revenue(Current US $)
TOTAL_TELECOM_SERVICES_REVENUE_0	0	interval	num	Total telecom services revenue(Current US $ per persor
TOTAL_TELECOM_SERVICES_REVENUE_1	0	interval	num	Total telecom services revenue(Current US $ per worker
ANNUAL_TELECOM_INVESTMENT_CURREN	0	interval	num	Annual telecom investment(Current US $)
ANNUAL_TELECOM_INVESTMENT_CURRE0	0	interval	num	Annual telecom investment(Current US $ per person)
ANNUAL_TELECOM_INVESTMENT____OF_	0	interval	num	Annual telecom investment(% of GDP)
ANNUAL_TELECOM_INVESTMENT_CURRE1	0	interval	num	Annual telecom investment(Current US $ per worker)
ANNUAL_TELECOM_INVESTMENT_CURRE2	1	interval	num	Annual telecom investment(Current US $ per telecom wor
TOTAL_TELECOM_SERVICES_REVENUE_2	0	interval	num	Total telecom services revenue(Current US $ per telec
PRODUCTIVITY_RATIO_PER_TELECOM_W	0	interval	num	Productivity ratio per telecom worker (revenue/investm
FULL_TIME_TELECOMMUNICATION_STAF	0	interval	num	Full-time telecommunication staff(% of total labor for

The membership of each cluster is provided in the table below.

CLUSTER	Frequency of Cluster	Root Mean Square Standard Deviation	Maximum Distance from Cluster Seed	Nearest Cluster	Distance to Nearest Cluster	Total telecom services revenue(Current U	T
1	24	0.9960394767	6.5635801015	2	5.0665605681	923409967.74	*
2	66	0.6215336579	7.9951632373	1	5.0665605681	216279988.74	*

Examination of the results provided by Reporter node yield the following regarding the membership of the Cluster 1 – Czech Republic (1994-1997), Estonia (1996-1997), Hungary (1993-1997), Latvia (1994-5, 1997), Poland (1995-1997), Slovak Republic (1996-1997), Slovenia (1993-7).



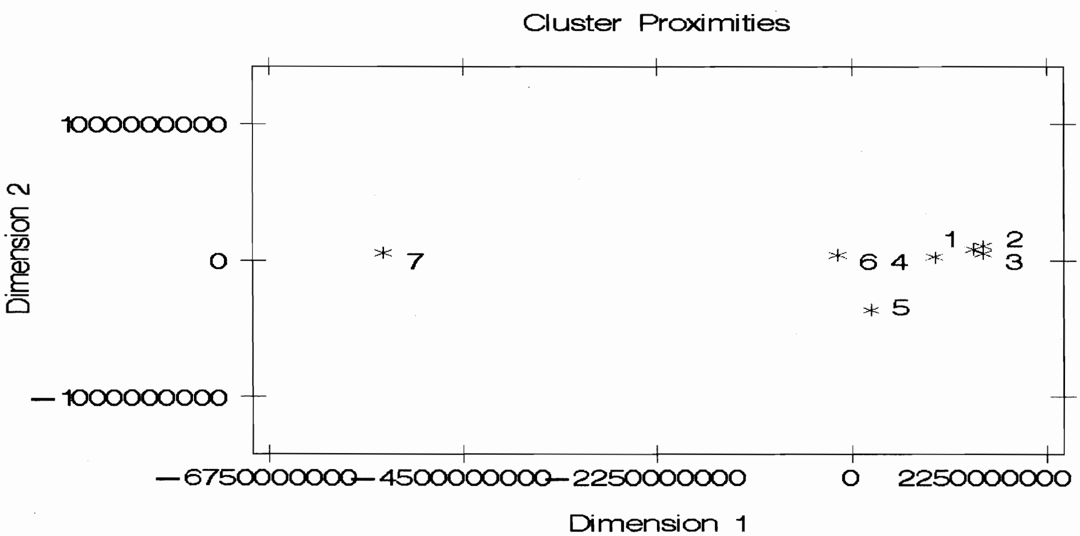
IF Annual telecom investment(Current US $ per telecom worker)
< $11,968
THEN
NODE : 2
N : 65
2 : 100.0%
1 : 0.0%

IF $11,968 <= Annual telecom investment(Current US $ per telecom worker)
THEN
NODE : 3
N : 25
2 : 4.0%
1 : 96.0%

7.1.2.2 Second 5-year data set (1998 – 2002)

Second 5-year data set (1998 – 2002), Automatic setting

First, we have chosen the "Automatic" setting of the Clustering node. This setting does not require input regarding the desired number of clusters. We have decided to find out what is the number of "naturally", from point of view of EM, occurring clusters. Set of the data covering the period from 1998 to 2002 has been partitioned into the seven clusters.



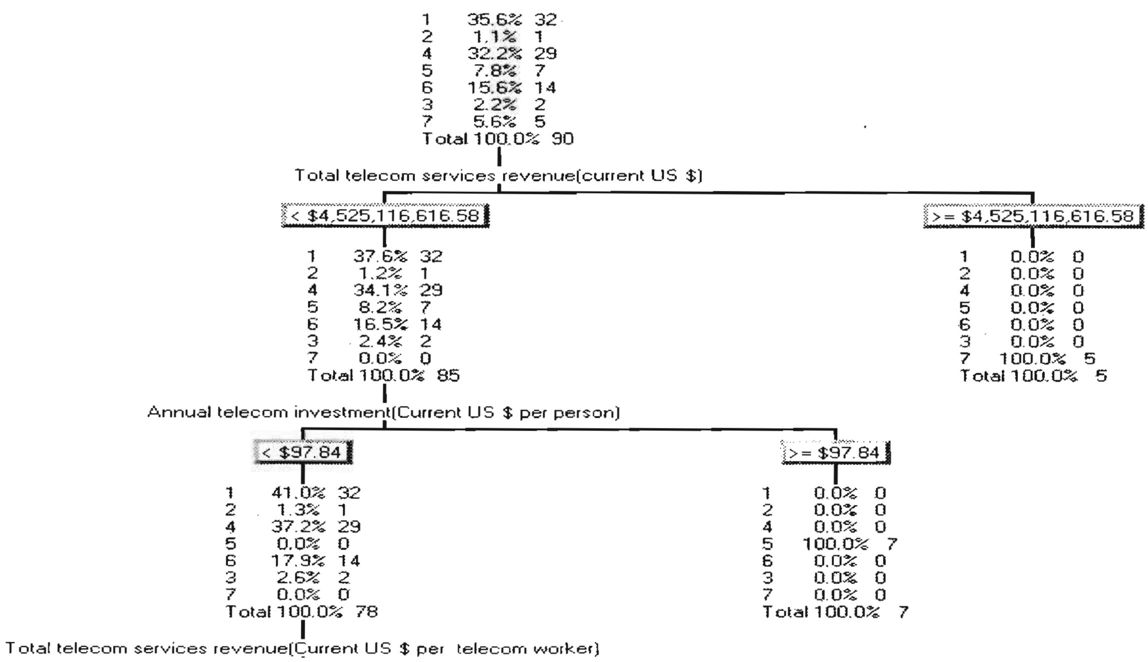
It looks like the cluster number 7 is really different from other clusters, while the clusters 1, 2 and 3 are almost indistinguishable one from another and the clusters 4, 5, and 6 being close by.

Name	Importance	Measurement	Type	Label
TOTAL_TELECOM_SERVICES_REVENUE_C	0.548563068	interval	num	Total telecom services revenue(current US $)
TOTAL_TELECOM_SERVICES_REVENUE_0	0	interval	num	Total telecom services revenue(Current US $ per per
TOTAL_TELECOM_SERVICES_REVENUE_1	0	interval	num	Total telecom services revenue(Current US $ per wor
ANNUAL_TELECOM_INVESTMENT_CURREN	0	interval	num	Annual telecom investment(current US $)
ANNUAL_TELECOM_INVESTMENT_CURRE0	0.6512832409	interval	num	Annual telecom investment(Current US $ per person)
ANNUAL_TELECOM_INVESTMENT____OF_	0	interval	num	Annual telecom investment(% of GDP)
ANNUAL_TELECOM_INVESTMENT_CURRE1	0	interval	num	Annual telecom investment(Current US $ per worker)
ANNUAL_TELECOM_INVESTMENT_CURRE2	0	interval	num	Annual telecom investment(Current US $ per telecom
TOTAL_TELECOM_SERVICES_REVENUE_2	0.8692634383	interval	num	Total telecom services revenue(Current US $ per te
PRODUCTIVITY_RATIO_PER_TELECOM_W	0.539092322	interval	num	Productivity ratio per telecom worker (revenue/inve
FULL_TIME_TELECOMMUNICATION_STAF	1	interval	num	Full-time telecommunication staff(% of total labor

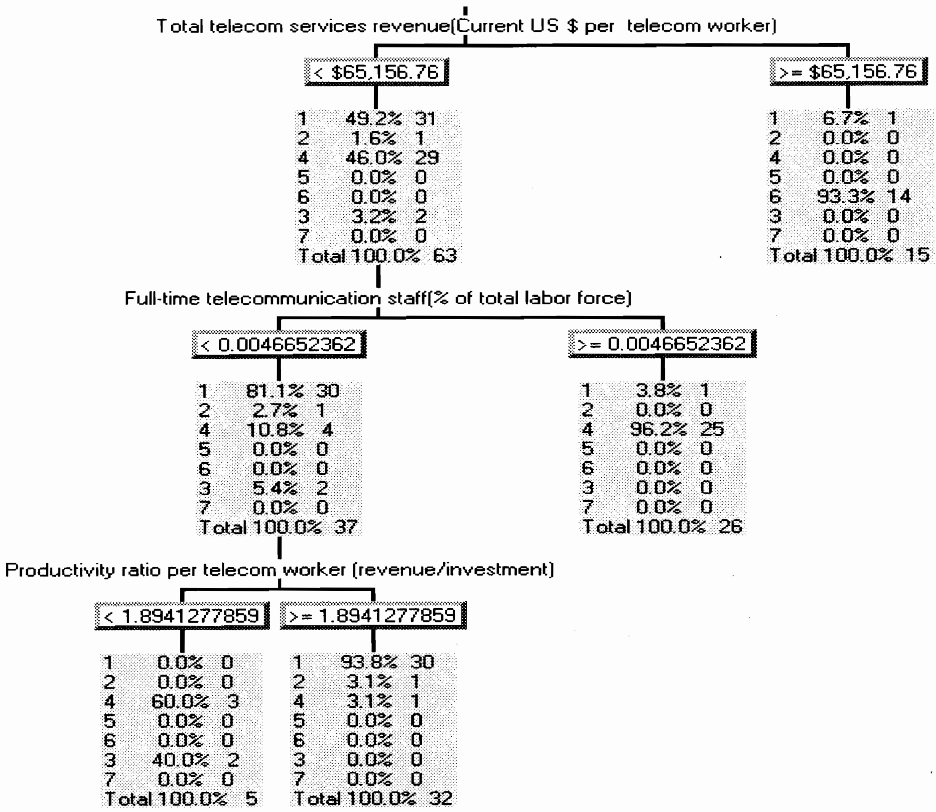
As the table above demonstrates, in partitioning the data set three variables were considered to be relevant, in diminishing order of importance: Full-time telecom staff, Total telecom service revenue per telecom worker, and Annual telecom investment per capita. Table below contains the information regarding the membership of each cluster.

CLUSTER	Frequency of Cluster	Root-Mean-Square Standard Deviation	Maximum Distance from Cluster Seed	Nearest Cluster	Distance to Nearest Cluster	Total telecom services revenue(current U	T
1	32	0.4185646439	3.2406333118	4	1.9342286928	184751824.31	
2	1		0	1	8.5307650418	83738833.612	
3	2	0.5637707038	1.322159497	4	4.3697658153	73187951.274	
4	29	0.4472444393	3.3005618407	1	1.9342286928	615972725.77	
5	7	0.6941647281	2.3928696332	6	4.2847540184	1266938559.4	
6	14	0.6249276297	3.0733437536	4	4.2590931508	1714339274.4	
7	5	0.3535707375	1.7904973734	6	4.2929036722	6880321221.3	

Decision Tree for 7-cluster solution.



Decision Tree for 7-cluster solution , continued from above.



Decision tree corresponding to the 7-cluster solution is depicted above, while the set of the decision rule offered below.

```
IF              $151 <= Annual telecom investment(Current US $ per
     worker)
THEN
 NODE   :    3
 N      :    8
 1      :   0.0%
 2      :   0.0%
 5      :   0.0%
 4      : 100.0%
 3      :   0.0%

IF 0.0151074521 <= Annual telecom investment( % of GDP)
AND Annual telecom investment(Current US $ per worker) <          $151
THEN
 NODE   :    5
```

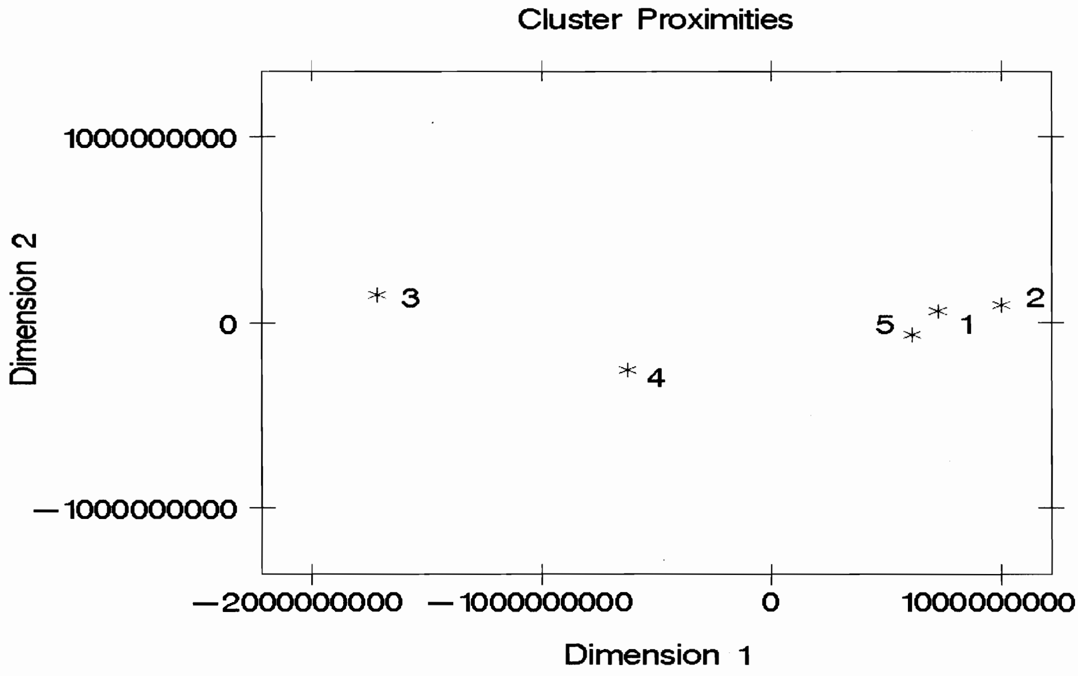
N : 5
1 : 0.0%
2 : 0.0%
5 : 100.0%
4 : 0.0%
3 : 0.0%

IF Total telecom services revenue(Current US $ per worker)
< $163
AND Annual telecom investment(% of GDP) < 0.0151074521
AND Annual telecom investment(Current US $ per worker) < $151
THEN
NODE : 6
N : 55
1 : 98.2%
2 : 1.8%
5 : 0.0%
4 : 0.0%
3 : 0.0%

IF $163 <= Total telecom services revenue(Current US $ per worker)
AND Annual telecom investment(% of GDP) < 0.0151074521
AND Annual telecom investment(Current US $ per worker) < $151
THEN
NODE : 7
N : 22
1 : 4.5%
2 : 0.0%
5 : 0.0%
4 : 0.0%
3 : 95.5%

Second 5-year data set (1998 – 2002), 5 clusters

5-cluster solution produces somewhat familiar picture of two clusters being removed from the rest. Clusters 5, 1, and 2 together contain 61 data points out of 90, cluster 3 contains 21, and cluster 4 contains another 8 data points.



According to the table below the most important variable in partitioning of the data was Total telecom service revenue per worker, followed by Annual telecom investment per worker.

Name	Importance	Measurement	Type	Label
TOTAL_TELECOM_SERVICES_REVENUE_C	0	interval	num	Total telecom services revenue(current US $)
TOTAL_TELECOM_SERVICES_REVENUE_0	0	interval	num	Total telecom services revenue(Current US $ per pe
TOTAL_TELECOM_SERVICES_REVENUE_1	1	interval	num	Total telecom services revenue(Current US $ per wo
ANNUAL_TELECOM_INVESTMENT_CURREN	0	interval	num	Annual telecom investment(current US $)
ANNUAL_TELECOM_INVESTMENT_CURRE0	0	interval	num	Annual telecom investment(Current US $ per person)
ANNUAL_TELECOM_INVESTMENT____OF_	0.514565923	interval	num	Annual telecom investment(% of GDP)
ANNUAL_TELECOM_INVESTMENT_CURRE1	0.6277606063	interval	num	Annual telecom investment(Current US $ per worker)
ANNUAL_TELECOM_INVESTMENT_CURRE2	0	interval	num	Annual telecom investment(Current US $ per telecom
TOTAL_TELECOM_SERVICES_REVENUE_2	0	interval	num	Total telecom services revenue(Current US $ per t
PRODUCTIVITY_RATIO_PER_TELECOM_W	0	interval	num	Productivity ratio per telecom worker (revenue/inv
FULL_TIME_TELECOMMUNICATION_STAF	0	interval	num	Full-time telecommunication staff(% of total labor

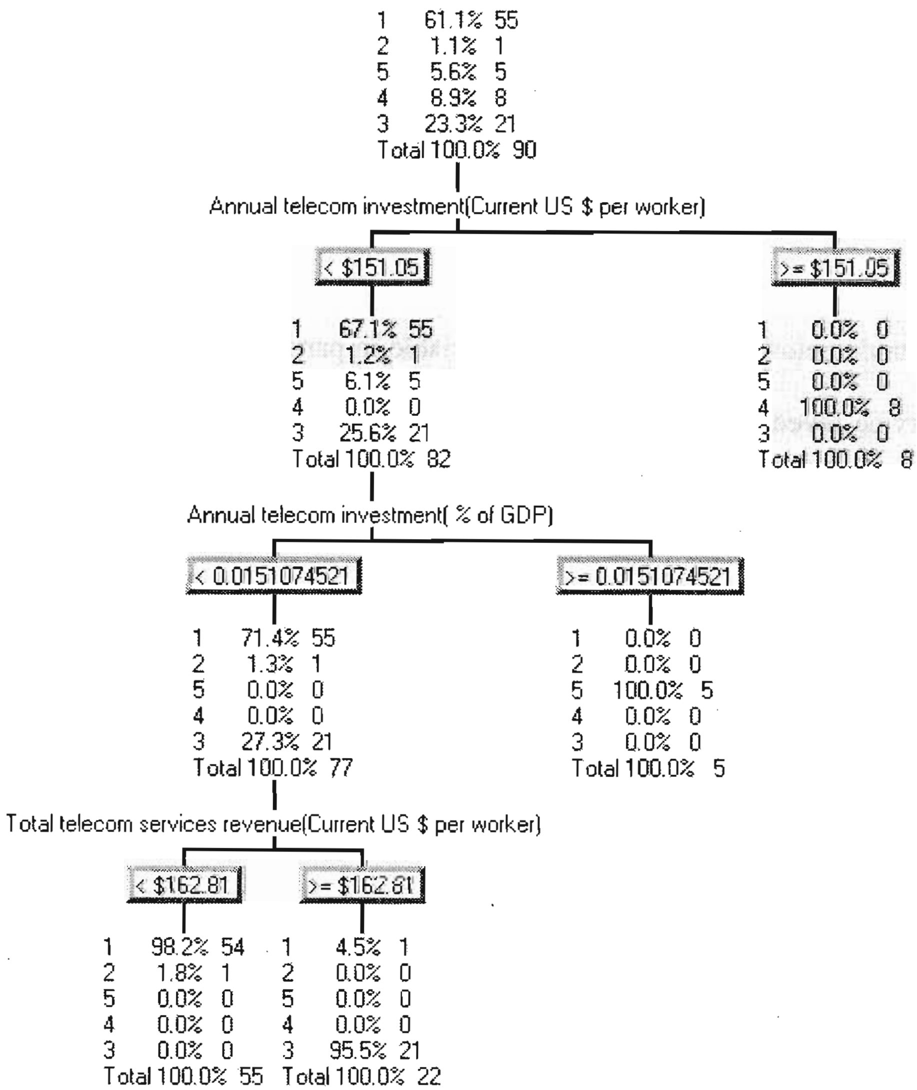
The membership of each cluster is provided in the table below.

CLUSTER	Frequency of Cluster	Root-Mean-Square Standard Deviation	Maximum Distance from Cluster Seed	Nearest Cluster	Distance to Nearest Cluster	Total telecom services revenue(current U	Tot
1	55	0.4799484028	3.4061702401	5	3.1325635398	344028764.97	****
2	1	.	0	1	8.7455694628	83738833.612	****
3	21	0.8094228691	4.8613569275	1	3.9517923548	2737648310.3	****
4	8	0.8063792151	4.2540643243	3	4.2831127649	1582846083.3	****
5	5	0.7578771748	2.7697292441	1	3.1325635398	423520787.38	****

According to the report, Cluster 3 consist of the data points corresponding to Czech Republic (1999, 2002), Estonia (1998-2002), Hungary (1998-2000), 2002), Lithuania (2002), Poland (1998-2002), Slovak Republic (2001-2), Slovenia (2002).

As for Cluster 4, it has following members - Czech Republic (1998, 2000-1), Hungary (2001), Slovenia (1998-2001)

Decision tree of the 5-cluster solution is provided below.



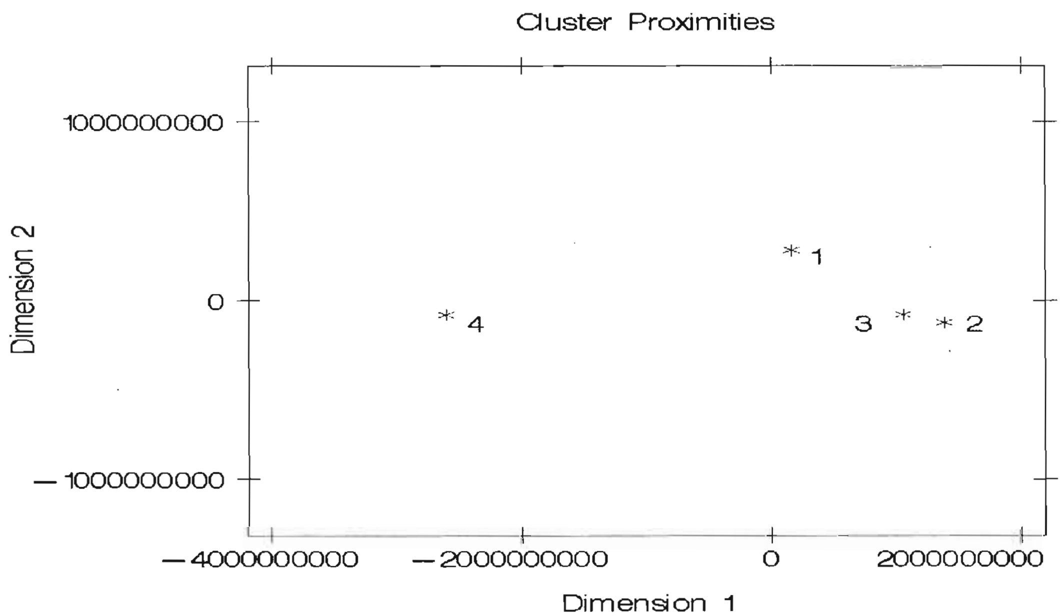
And following is a set of decision rules.

```
IF              $151 <= Annual telecom investment(Current US $ per
     worker)
THEN
 NODE  :     3
 N     :     8
 1     :   0.0%
 2     :   0.0%
 5     :   0.0%
 4     : 100.0%
 3     :   0.0%

IF 0.0151074521 <= Annual telecom investment( % of GDP)
AND Annual telecom investment(Current US $ per worker) <           $151
THEN
 NODE  :     5
 N     :     5
 1     :   0.0%
 2     :   0.0%
 5     : 100.0%
 4     :   0.0%
 3     :   0.0%

IF  Total telecom services revenue(Current US $ per worker)
   <            $163
AND Annual telecom investment( % of GDP) < 0.0151074521
AND Annual telecom investment(Current US $ per worker) <           $151
THEN
 NODE  :     6
 N     :    55
 1     :  98.2%
 2     :   1.8%
 5     :   0.0%
 4     :   0.0%
 3     :   0.0%

IF              $163 <= Total telecom services revenue(Current US $ per
     worker)
AND Annual telecom investment( % of GDP) < 0.0151074521
AND Annual telecom investment(Current US $ per worker) <           $151
THEN
 NODE  :     7
 N     :    22
```

1	:	4.5%
2	:	0.0%
5	:	0.0%
4	:	0.0%
3	:	95.5%

Second 5-year data set (1998 – 2002) , 4 Clusters

Again, we could see that the trend of having obvious heterogeneity holds for 4-cluster solution. Clearly, cluster 4 is far removed from the majority of the data points represented by the clusters 2 and 3 (68 out of 90 data points).



Total telecom service revenue, followed by the Annual telecom investment per capita, became the variables that were used in partitioning of the data set.

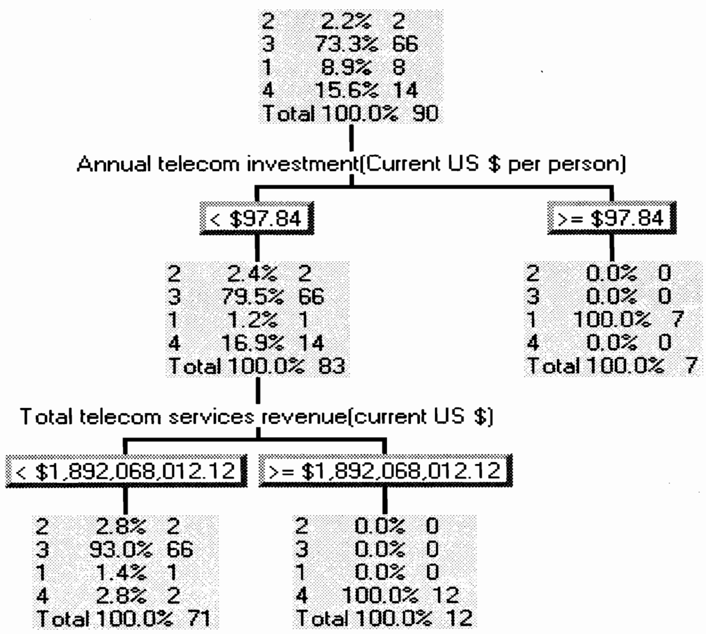
Name	Importance	Measurement	Type	Label
TOTAL_TELECOM_SERVICES_REVENUE_C	1	interval	num	Total telecom services revenue(current US $)
TOTAL_TELECOM_SERVICES_REVENUE_0	0	interval	num	Total telecom services revenue(Current US $ per per
TOTAL_TELECOM_SERVICES_REVENUE_1	0	interval	num	Total telecom services revenue(Current US $ per wor
ANNUAL_TELECOM_INVESTMENT_CURREN	0	interval	num	Annual telecom investment(current US $)
ANNUAL_TELECOM_INVESTMENT_CURRE0	0.7543551029	interval	num	Annual telecom investment(Current US $ per person)
ANNUAL_TELECOM_INVESTMENT____OF_	0	interval	num	Annual telecom investment(% of GDP)
ANNUAL_TELECOM_INVESTMENT_CURRE1	0	interval	num	Annual telecom investment(Current US $ per worker)
ANNUAL_TELECOM_INVESTMENT_CURRE2	0	interval	num	Annual telecom investment(Current US $ per telecom
TOTAL_TELECOM_SERVICES_REVENUE_2	0	interval	num	Total telecom services revenue(Current US $ per te
PRODUCTIVITY_RATIO_PER_TELECOM_W	0	interval	num	Productivity ratio per telecom worker (revenue/inve
FULL_TIME_TELECOMMUNICATION_STAF	0	interval	num	Full-time telecommunication staff(% of total labor

Membership of each cluster is provided in the table below.

CLUSTER	Frequency of Cluster	Root-Mean-Square Standard Deviation	Maximum Distance from Cluster Seed	Nearest Cluster	Distance to Nearest Cluster	Total telecom services revenue(current U	T
1	8	0.7991970092	3.2145259023	4	4.1482993838	1193259585	*
2	2	1.1577678035	2.7152061759	3	6.2477868882	71099981.766	*
3	66	0.5709900839	5.4464164955	4	4.8077841629	391797069.99	*
4	14	0.7633443188	3.3819963816	1	4.1482993838	3980673197.4	*

According to the results provided by Reporter node, Cluster 4 consists of the data points corresponding to Czech republic (1999, 2002), Estonia (2001-2), Hungary (1998-2002), Poland (1998-2002), while Cluster 1 contains Czech republic (1998, 2000-1), and Slovenia (1998-2002).

Below is the Decision Tree generated for the 4-cluster' solution.



Corresponding decision rules are provided below.

IF $98 <= Annual telecom investment(Current US $ per person)
THEN
NODE : 3
N : 7
2 : 0.0%
3 : 0.0%
1 : 100.0%
4 : 0.0%

IF Total telecom services revenue(current US $) < $1,892,068,012
AND Annual telecom investment(Current US $ per person) < $98
THEN
NODE : 4
N : 71
2 : 2.8%
3 : 93.0%
1 : 1.4%
4 : 2.8%

IF $1,892,068,012 <= Total telecom services revenue(current US $)
AND Annual telecom investment(Current US $ per person) < $98
THEN
NODE : 5
N : 12
2 : 0.0%
3 : 0.0%
1 : 0.0%
4 : 100.0%

Second 5-year data set (1998 – 2002) , 3 clusters

Below is the table of distances between the clusters.

CLUSTER	Cluster 1	Cluster 2	Cluster 3
1	0	339216343.56	2657856689.
2	339216343.56	0	2996061545.
3	2657856689.7	2996061545.9	

In partitioning of the data set Total telecom services revenue per telecom worker was considered to be the most important variable (see table below).

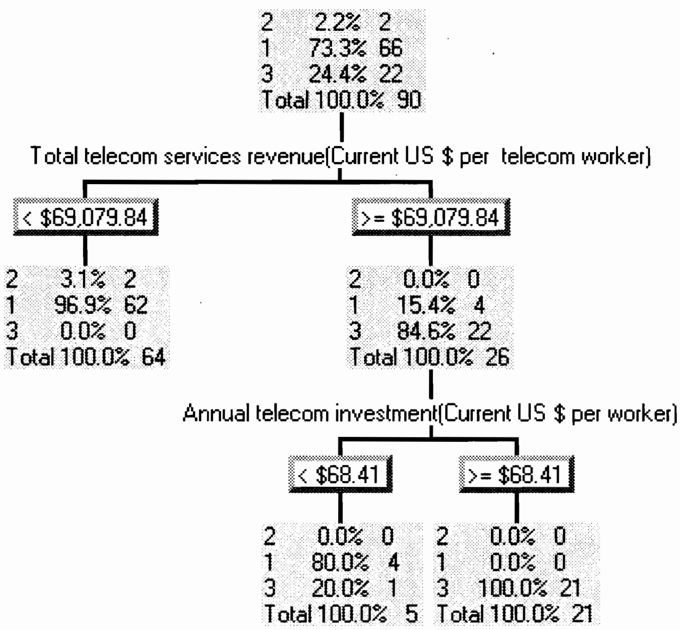
Name	Importance	Measurement	Type	Label
TOTAL_TELECOM_SERVICES_REVENUE_C	0	interval	num	Total telecom services revenue(current US $)
TOTAL_TELECOM_SERVICES_REVENUE_0	0	interval	num	Total telecom services revenue(Current US $ per pers
TOTAL_TELECOM_SERVICES_REVENUE_1	0	interval	num	Total telecom services revenue(Current US $ per work
ANNUAL_TELECOM_INVESTMENT_CURREN	0	interval	num	Annual telecom investment(current US $)
ANNUAL_TELECOM_INVESTMENT_CURRE0	0	interval	num	Annual telecom investment(Current US $ per person)
ANNUAL_TELECOM_INVESTMENT____OF_	0	interval	num	Annual telecom investment(% of GDP)
ANNUAL_TELECOM_INVESTMENT_CURRE1	0.449942882	interval	num	Annual telecom investment(Current US $ per worker)
ANNUAL_TELECOM_INVESTMENT_CURRE2	0	interval	num	Annual telecom investment(Current US $ per telecom w
TOTAL_TELECOM_SERVICES_REVENUE_2	1	interval	num	Total telecom services revenue(Current US $ per tel
PRODUCTIVITY_RATIO_PER_TELECOM_W	0	interval	num	Productivity ratio per telecom worker (revenue/inves
FULL_TIME_TELECOMMUNICATION_STAF	0	interval	num	Full-time telecommunication staff(% of total labor f

Upon the examination, we found that cluster 2 consists of only 2 data points, which is quite unusual, for it represents less than 2.5% of the data points in the set.

CLUSTER	Frequency of Cluster	Root-Mean-Square Standard Deviation	Maximum Distance from Cluster Seed	Nearest Cluster	Distance to Nearest Cluster	Total telecom services revenue(current U
1	66	0.5709900839	5.4850476145	3	5.0820437309	391797069.99
2	2	1.1577678035	2.7152061759	1	6.2477868882	71099981.766
3	22	0.9761501806	4.6588911923	1	5.0820437309	2967068247.4

According to the generated report Cluster 2 contains 2 data point corresponding to Albania (1998-9), while Cluster 3 contains Czech Republic (1998-2002), Estonia (2001-2), Hungary (1998-2002), Poland (1998-2002), Slovenia (1998-2002).

Decision tree for 3-cluster solution is provided below.



Below is the set of the decision rules corresponding to DT for this solution.

IF Total telecom services revenue(Current US $ per telecom worker)
< $69,080
THEN
NODE : 2
N : 64
2 : 3.1%
1 : 96.9%
3 : 0.0%

IF Annual telecom investment(Current US $ per worker) < $68
AND $69,080 <= Total telecom services revenue(Current US $ per telecom worker)
THEN
NODE : 6
N : 5
2 : 0.0%
1 : 80.0%
3 : 20.0%

IF $68 <= Annual telecom investment(Current US $ per worker)
AND $69,080 <= Total telecom services revenue(Current US $ per telecom worker)
THEN
NODE : 7
N : 21
2 : 0.0%
1 : 0.0%
3 : 100.0%

Second 5-year data set (1998 – 2002), 2 Clusters

Table of distances for 2-cluster solution is shown below.

CLUSTER	Cluster 1	Cluster 2
1	0	1568266276
2	1568266276.7	

Annual telecom investment per capita becomes the most important in partitioning of the data.

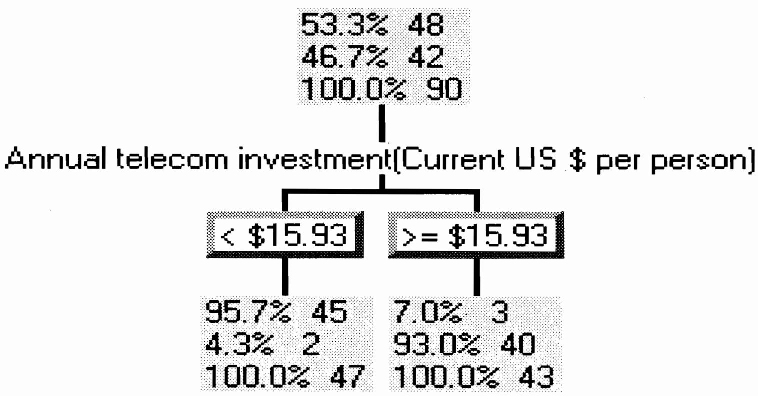
Name	Importance	Measurement	Type	Label
TOTAL_TELECOM_SERVICES_REVENUE_C	0	interval	num	Total telecom services revenue(current US $)
TOTAL_TELECOM_SERVICES_REVENUE_0	0	interval	num	Total telecom services revenue(Current US $ per perso
TOTAL_TELECOM_SERVICES_REVENUE_1	0	interval	num	Total telecom services revenue(Current US $ per worke
ANNUAL_TELECOM_INVESTMENT_CURREN	0	interval	num	Annual telecom investment(current US $)
ANNUAL_TELECOM_INVESTMENT_CURRE0	1	interval	num	Annual telecom investment(Current US $ per person)
ANNUAL_TELECOM_INVESTMENT____OF_	0	interval	num	Annual telecom investment(% of GDP)
ANNUAL_TELECOM_INVESTMENT_CURRE1	0	interval	num	Annual telecom investment(Current US $ per worker)
ANNUAL_TELECOM_INVESTMENT_CURRE2	0	interval	num	Annual telecom investment(Current US $ per telecom wo
TOTAL_TELECOM_SERVICES_REVENUE_2	0	interval	num	Total telecom services revenue(Current US $ per tele
PRODUCTIVITY_RATIO_PER_TELECOM_W	0	interval	num	Productivity ratio per telecom worker (revenue/invest
FULL_TIME_TELECOMMUNICATION_STAF	0	interval	num	Full-time telecommunication staff(% of total labor fo

And, we could see, that clusters become almost equal in size.

CLUSTER	Frequency of Cluster	Root-Mean-Square Standard Deviation	Maximum Distance from Cluster Seed	Nearest Cluster	Distance to Nearest Cluster	Total telecom services revenue(current U	T
1	42	1.0109905398	6.5860503159	2	3.8348341767	1822278211.1	**
2	48	0.6040141432	8.1872284584	1	3.8348341767	307096315.8	**

According to the results obtained from the report, Cluster 1 consists of the data points which represent Czech Republic(1998-2002), Estonia (1998-2002), Hungary (1998-2002), Latvia (1998-2002), Poland (1998-2002) , Slovak Republic (1998,2000-2002) Slovenia (1998-2002), Ukraine (2002), Romania (2000, 2002), Moldova (2002), Lithuania (1999, 2000, 2002), Bulgaria (2002).

Decision tree is depicted below.



We provide the corresponding set of decision rules as well.

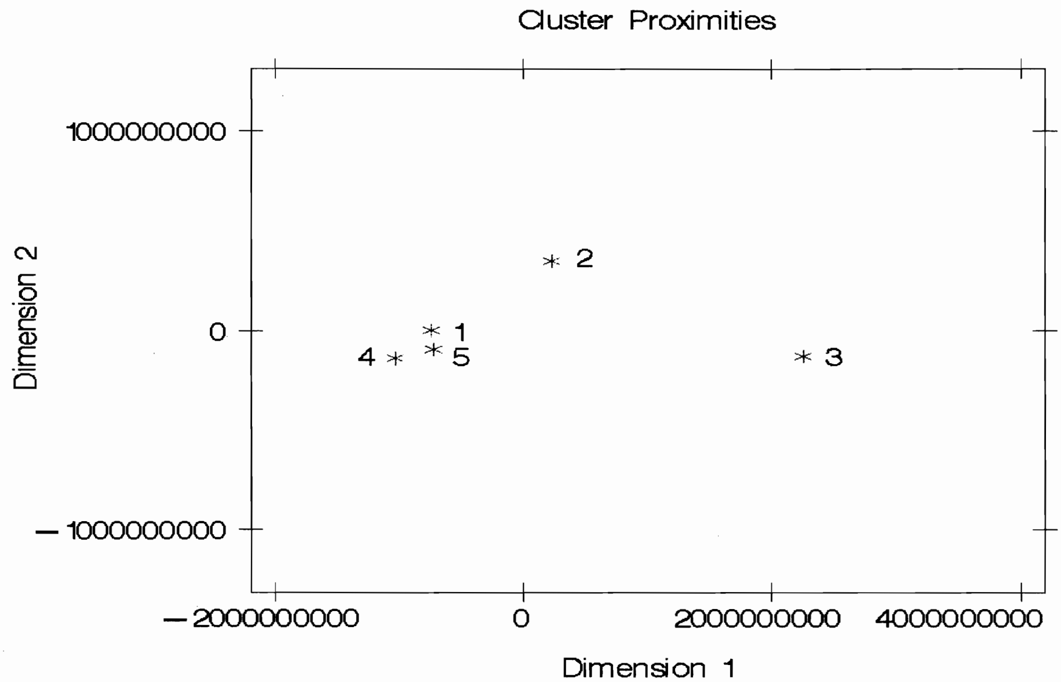
IF Annual telecom investment(Current US $ per person) < $16
THEN
NODE : 2
N : 47
2 : 95.7%

```
 1     :  4.3%

IF              $16 <= Annual telecom investment(Current US $ per
     person)
THEN
 NODE  :    3
 N     :    43
 2     :  7.0%
 1     :  93.0%
```

7.1.2.3 10-year data set (1993-2002)

10-year data set (1993-2002), 5 clusters

The "Automatic" setting has produced somewhat similar to the previous 5-Year cases results in partitioning the 10-year dataset. We could see that "Automatic" setting generated 5 clusters, 3 of which, clusters number 1, 4, and 5, are close together, while clusters 2 and 3 situated farther away.



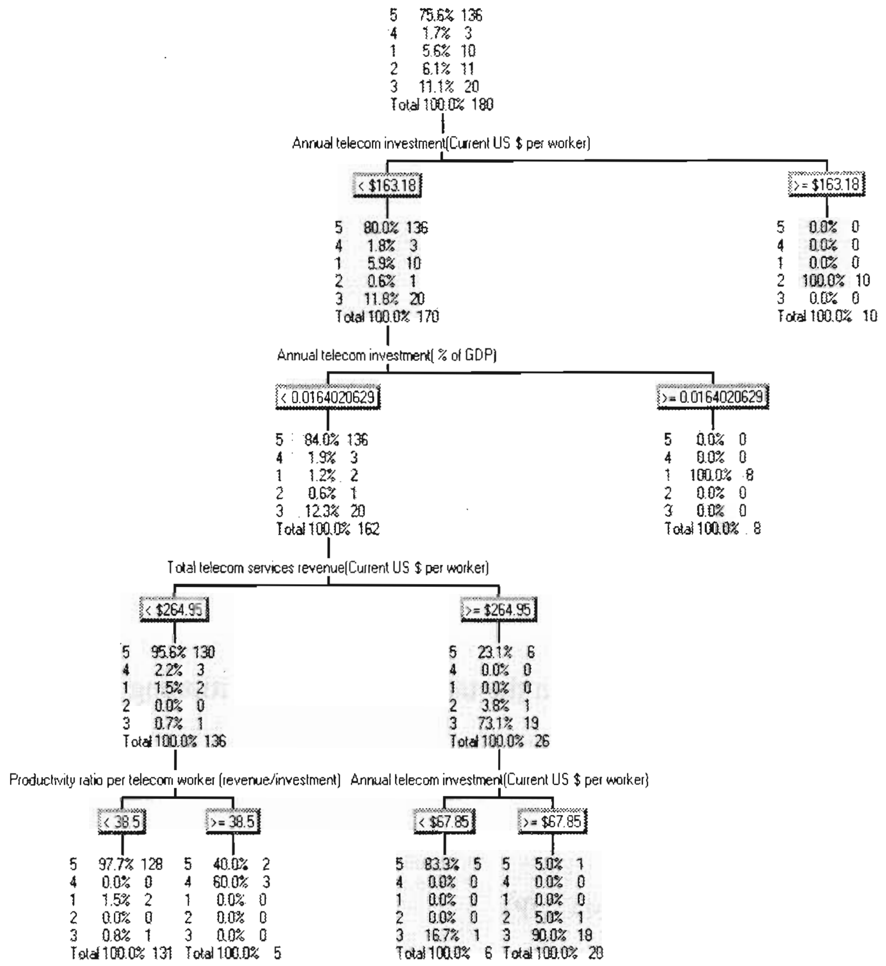
Based on the information provided in the table below, in partitioning of the data set three variables were considered to be the most relevant, in diminishing order of importance: Total telecom service revenue per worker, Annual telecom investment per worker, and Annual telecom investment (% of GDP) .

Name	Importance	Measurement	Type	Label
TOTAL_TELECOM_SERVICES_REVENUE_C	0	interval	num	Total telecom services revenue(Current US $)
TOTAL_TELECOM_SERVICES_REVENUE_0	0	interval	num	Total telecom services revenue(Current US $ per per
TOTAL_TELECOM_SERVICES_REVENUE_1	1	interval	num	Total telecom services revenue(Current US $ per wor
ANNUAL_TELECOM_INVESTMENT_CURREN	0	interval	num	Annual telecom investment(Current US $)
ANNUAL_TELECOM_INVESTMENT_CURRE0	0	interval	num	Annual telecom investment(Current US $ per person)
ANNUAL_TELECOM_INVESTMENT____OF_	0.7506022753	interval	num	Annual telecom investment(% of GDP)
ANNUAL_TELECOM_INVESTMENT_CURRE1	0.95367343	interval	num	Annual telecom investment(Current US $ per worker)
ANNUAL_TELECOM_INVESTMENT_CURRE2	0	interval	num	Annual telecom investment(Current US $ per telecom
TOTAL_TELECOM_SERVICES_REVENUE_2	0	interval	num	Total telecom services revenue(Current US $ per te
PRODUCTIVITY_RATIO_PER_TELECOM_W	0.3814699143	interval	num	Productivity ratio per telecom worker (revenue/inve
FULL_TIME_TELECOMMUNICATION_STAF	0	interval	num	Full-time telecommunication staff(% of total labor

Table describing the membership for each cluster is provided below.

CLUSTER	Frequency of Cluster	Root-Mean-Square Standard Deviation	Maximum Distance from Cluster Seed	Nearest Cluster	Distance to Nearest Cluster	Total telecom services revenue(Current U
1	10	0.6910952169	3.3838196355	5	2.8685729343	301079379.57
2	11	0.9125016132	3.6211143581	3	4.3069207936	1167300217.7
3	20	0.9272675431	4.5743045211	2	4.3069207936	3247370982
4	3	0.79872865	3.039460512	5	7.3359108746	52491522.551
5	136	0.5521570041	3.5951652596	1	2.8685729343	343798794.92

Below are the Decision tree and a set of corresponding decision rules.



Set of the Decision Rules, based on the model above.

IF $163 <= Annual telecom investment(Current US $ per worker)
THEN
NODE : 3
N : 10
5 : 0.0%
4 : 0.0%
1 : 0.0%
2 : 100.0%
3 : 0.0%

IF 0.0164020629 <= Annual telecom investment(% of GDP)
AND Annual telecom investment(Current US $ per worker) < $163
THEN
NODE : 5
N : 8
5 : 0.0%
4 : 0.0%
1 : 100.0%
2 : 0.0%
3 : 0.0%

IF Productivity ratio per telecom worker (revenue/investment) < 38.5
AND Total telecom services revenue(Current US $ per worker)
< $265
AND Annual telecom investment(% of GDP) < 0.0164020629
AND Annual telecom investment(Current US $ per worker) < $163
THEN
NODE : 8
N : 131
5 : 97.7%
4 : 0.0%
1 : 1.5%
2 : 0.0%
3 : 0.8%

IF 38.5 <= Productivity ratio per telecom worker (revenue/investment)
AND Total telecom services revenue(Current US $ per worker)
< $265
AND Annual telecom investment(% of GDP) < 0.0164020629
AND Annual telecom investment(Current US $ per worker) < $163
THEN
NODE : 9
N : 5
5 : 40.0%
4 : 60.0%
1 : 0.0%
2 : 0.0%
3 : 0.0%

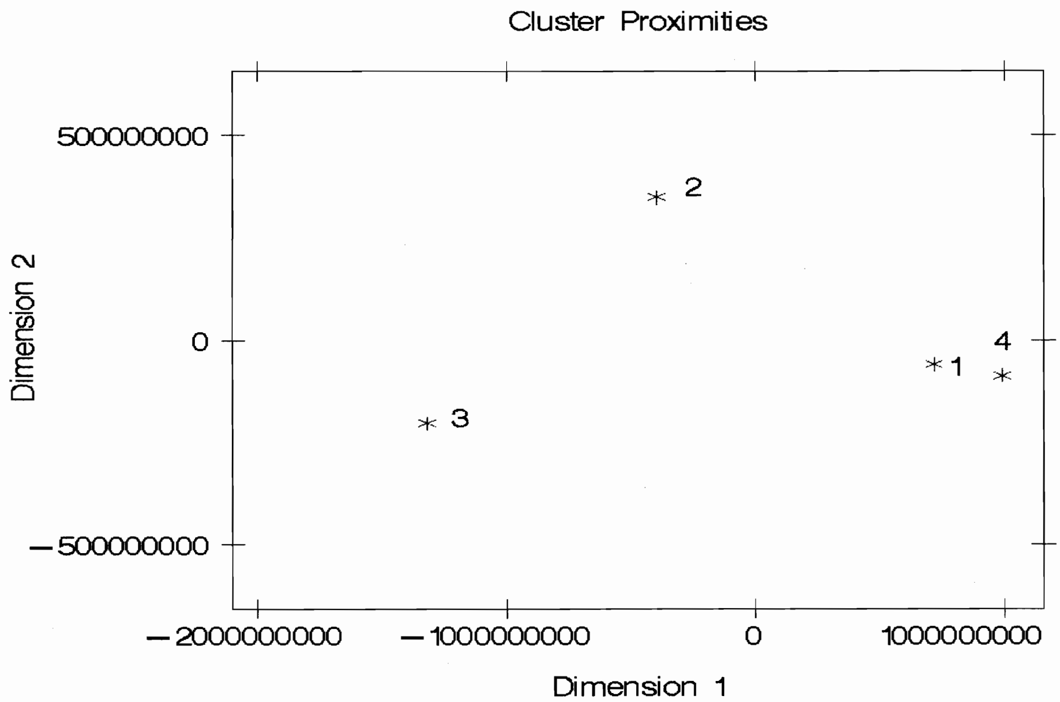
IF Annual telecom investment(Current US $ per worker) < $68
AND $265 <= Total telecom services revenue(Current US $ per worker)
AND Annual telecom investment(% of GDP) < 0.0164020629

```
THEN
 NODE  :    10
 N     :     6
 5     :  83.3%
 4     :   0.0%
 1     :   0.0%
 2     :   0.0%
 3     :  16.7%

IF              $68 <= Annual telecom investment(Current US $ per
     worker) <              $163
AND              $265 <= Total telecom services revenue(Current US $ per
     worker)
AND Annual telecom investment( % of GDP) < 0.0164020629
THEN
 NODE  :    11
 N     :    20
 5     :   5.0%
 4     :   0.0%
 1     :   0.0%
 2     :   5.0%
 3     :  90.0%
```

Next, we perform the cluster analysis on the set of 5-year data using preset number of clusters, 4.

10-year data set (1993-200)2, 4 clusters

Somewhat similar result was obtained through 4-cluster solution, Clusters 1 and 4 are close together (accounting for 138 data points out of 180) with the clusters 3 and 2 being fairly removed.



Total telecom service revenue per worker was considered to be the most important variable in partitioning of the data.

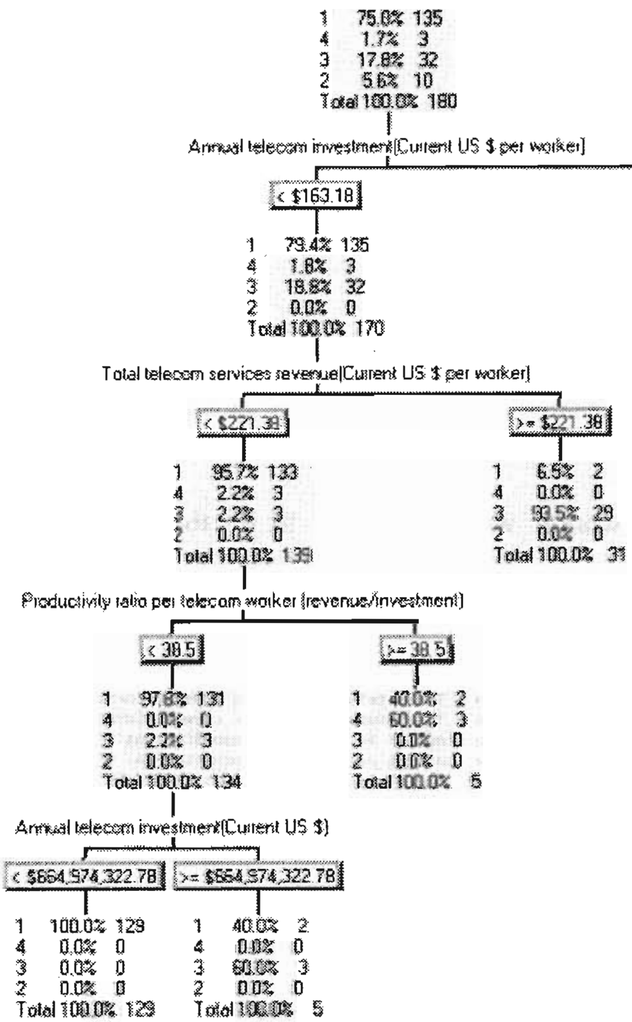
Name	Importance	Measurement	Type	Label
TOTAL_TELECOM_SERVICES_REVENUE_C	0	interval	num	Total telecom services revenue(Current US $)
TOTAL_TELECOM_SERVICES_REVENUE_0	0	interval	num	Total telecom services revenue(Current US $ per per
TOTAL_TELECOM_SERVICES_REVENUE_1	1	interval	num	Total telecom services revenue(Current US $ per wor
ANNUAL_TELECOM_INVESTMENT_CURREN	0.2894552211	interval	num	Annual telecom investment(Current US $)
ANNUAL_TELECOM_INVESTMENT_CURRE0	0	interval	num	Annual telecom investment(Current US $ per person)
ANNUAL_TELECOM_INVESTMENT____OF_	0	interval	num	Annual telecom investment(% of GDP)
ANNUAL_TELECOM_INVESTMENT_CURRE1	0.6168238947	interval	num	Annual telecom investment(Current US $ per worker)
ANNUAL_TELECOM_INVESTMENT_CURRE2	0	interval	num	Annual telecom investment(Current US $ per telecom
TOTAL_TELECOM_SERVICES_REVENUE_2	0	interval	num	Total telecom services revenue(Current US $ per te
PRODUCTIVITY_RATIO_PER_TELECOM_W	0.2844069536	interval	num	Productivity ratio per telecom worker (revenue/inve
FULL_TIME_TELECOMMUNICATION_STAF	0	interval	num	Full-time telecommunication staff(% of total labor

Each cluster membership is provided in the table below.

CLUSTER	Frequency of Cluster	Root-Mean-Square Standard Deviation	Maximum Distance from Cluster Seed	Nearest Cluster	Distance to Nearest Cluster	Total telecom services revenue(Current U
1	135	0.549179016	5.8872861968	3	4.510702041	308195775.29
2	10	0.8792442624	3.5227988604	3	4.5254221857	1251400954.6
3	32	0.9412023199	5.5697550356	1	4.510702041	2294834772.9
4	3	0.79872865	3.039460512	1	7.3562791919	52491522.551

The size of the cluster 4, which is 3 data points, suggests that we should probably go ahead and minimize the number of clusters further. As for the cluster 2, it consists of Slovenia (1998-2001) and Czech Republic (1996-1998, 2000-2001)

Decision tree for the 4-cluster solution is provided below.



And so the corresponding set of the decision rules, based on the Decision Tree above.

IF $163 <= Annual telecom investment(Current US $ per worker) THEN NODE : 3 N : 10

```
 1     :   0.0%
 4     :   0.0%
 3     :   0.0%
 2     : 100.0%

IF              $221 <= Total telecom services revenue(Current US $ per
     worker)
AND Annual telecom investment(Current US $ per worker) <         $163
THEN
 NODE   :     5
 N      :     31
 1     :   6.5%
 4     :   0.0%
 3     :  93.5%
 2     :   0.0%

IF        38.5 <= Productivity ratio per telecom worker (revenue/investment)
AND Total telecom services revenue(Current US $ per worker)
  <           $221
AND Annual telecom investment(Current US $ per worker) <         $163
THEN
 NODE   :     7
 N      :     5
 1     :  40.0%
 4     :  60.0%
 3     :   0.0%
 2     :   0.0%

IF  Annual telecom investment(Current US $) <       $664,974,323
AND Productivity ratio per telecom worker (revenue/investment) <      38.5
AND Total telecom services revenue(Current US $ per worker)
  <           $221
AND Annual telecom investment(Current US $ per worker) <         $163
THEN
 NODE   :     8
 N      :    129
 1     : 100.0%
 4     :   0.0%
 3     :   0.0%
 2     :   0.0%

IF        $664,974,323 <= Annual telecom investment(Current US $)
AND Productivity ratio per telecom worker (revenue/investment) <      38.5
AND Total telecom services revenue(Current US $ per worker)
```

< $221
AND Annual telecom investment(Current US $ per worker) < $163
THEN
NODE : 9
N : 5
1 : 40.0%
4 : 0.0%
3 : 60.0%
2 : 0.0%

10-year data set(1993-2002), 3 clusters

We cannot get a visual representation of the clusters; thus, we provide the table below, which provides the distances between the clusters.

CLUSTER	Cluster 1	Cluster 2	Cluster 3
1	0	2536016348.8	335212194.:
2	2536016348.8	0	22026195:
3	335212194.34	2202619539	

As in the previous solution, Total telecom services revenue per worker is considered to be the most important variable to distinguish the subsets of the data.

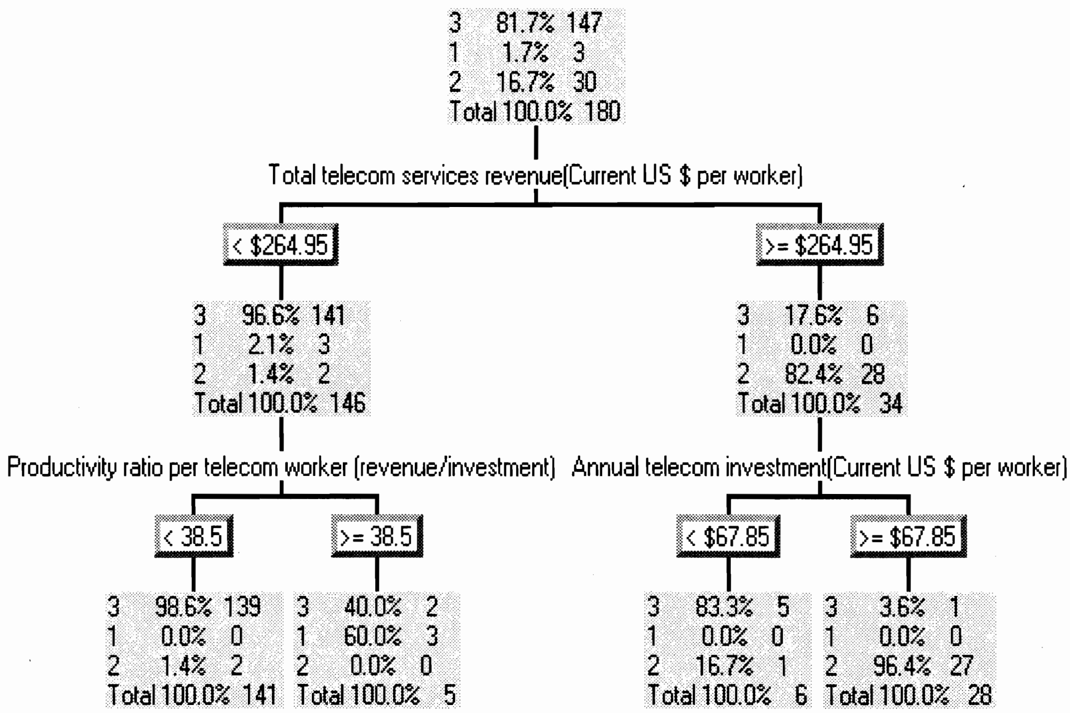
Name	Importance	Measurement	Type	Label
TOTAL_TELECOM_SERVICES_REVENUE_C	0	interval	num	Total telecom services revenue(Current US $)
TOTAL_TELECOM_SERVICES_REVENUE_0	0	interval	num	Total telecom services revenue(Current US $ per per
TOTAL_TELECOM_SERVICES_REVENUE_1	1	interval	num	Total telecom services revenue(Current US $ per wor
ANNUAL_TELECOM_INVESTMENT_CURREN	0	interval	num	Annual telecom investment(Current US $)
ANNUAL_TELECOM_INVESTMENT_CURRE0	0	interval	num	Annual telecom investment(Current US $ per person)
ANNUAL_TELECOM_INVESTMENT____OF_	0	interval	num	Annual telecom investment(% of GDP)
ANNUAL_TELECOM_INVESTMENT_CURRE1	0.4221572041	interval	num	Annual telecom investment(Current US $ per worker)
ANNUAL_TELECOM_INVESTMENT_CURRE2	0	interval	num	Annual telecom investment(Current US $ per telecom
TOTAL_TELECOM_SERVICES_REVENUE_2	0	interval	num	Total telecom services revenue(Current US $ per te
PRODUCTIVITY_RATIO_PER_TELECOM_W	0.3102858647	interval	num	Productivity ratio per telecom worker (revenue/inve
FULL_TIME_TELECOMMUNICATION_STAF	0	interval	num	Full-time telecommunication staff(% of total labor

Cluster 1, consisting of 3 data points, suggests that we could probably benefit from even further reduction of the number of clusters.

CLUSTER	Frequency of Cluster	Root-Mean-Square Standard Deviation	Maximum Distance from Cluster Seed	Nearest Cluster	Distance to Nearest Cluster	Total telecom services revenue(Current U
1	3	0.79872865	3.039460512	3	7.3910226812	52491522.551
2	30	1.1012423434	6.0830364668	3	5.6707912186	2477730978.4
3	147	0.6092375767	5.871283852	2	5.6707912186	362062738.7

It turns out to be that the Cluster 1 is represented by Albania (1996, 1997, 1999).

Decision tree, as well as a set of corresponding rules, is provided below.



IF Productivity ratio per telecom worker (revenue/investment) < 38.5
AND Total telecom services revenue(Current US $ per worker)
< $265
THEN
NODE : 4
N : 141
3 : 98.6%
1 : 0.0%
2 : 1.4%

IF 38.5 <= Productivity ratio per telecom worker (revenue/investment)
AND Total telecom services revenue(Current US $ per worker)
< $265
THEN
NODE : 5
N : 5
3 : 40.0%
1 : 60.0%
2 : 0.0%

IF Annual telecom investment(Current US $ per worker) < $68
AND $265 <= Total telecom services revenue(Current US $ per worker)
THEN
NODE : 6
N : 6
3 : 83.3%
1 : 0.0%
2 : 16.7%

IF $68 <= Annual telecom investment(Current US $ per worker)
AND $265 <= Total telecom services revenue(Current US $ per worker)
THEN
NODE : 7
N : 28
3 : 3.6%
1 : 0.0%
2 : 96.4%

10-year data set(1993-2002), 2 clusters

The last solution for the 10-year data set consists of 2 clusters, distance between which is given in the table below.

CLUSTER	Cluster 1	Cluster 2
1	0	123231662
2	1232316620.9	

Partitioning of the data set is made based on the Annual telecom investment per telecom worker(see table below).

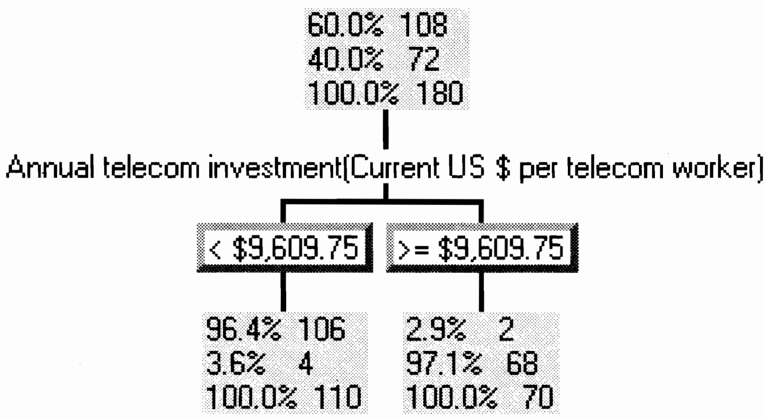
Name	Importance	Measurement	Type	Label
TOTAL_TELECOM_SERVICES_REVENUE_C	0	interval	num	Total telecom services revenue(Current US $)
TOTAL_TELECOM_SERVICES_REVENUE_0	0	interval	num	Total telecom services revenue(Current US $ per perso
TOTAL_TELECOM_SERVICES_REVENUE_1	0	interval	num	Total telecom services revenue(Current US $ per worke
ANNUAL_TELECOM_INVESTMENT_CURREN	0	interval	num	Annual telecom investment(Current US $)
ANNUAL_TELECOM_INVESTMENT_CURRE0	0	interval	num	Annual telecom investment(Current US $ per person)
ANNUAL_TELECOM_INVESTMENT____OF_	0	interval	num	Annual telecom investment(% of GDP)
ANNUAL_TELECOM_INVESTMENT_CURRE1	0	interval	num	Annual telecom investment(Current US $ per worker)
ANNUAL_TELECOM_INVESTMENT_CURRE2	1	interval	num	Annual telecom investment(Current US $ per telecom wo
TOTAL_TELECOM_SERVICES_REVENUE_2	0	interval	num	Total telecom services revenue(Current US $ per tele
PRODUCTIVITY_RATIO_PER_TELECOM_W	0	interval	num	Productivity ratio per telecom worker (revenue/invest
FULL_TIME_TELECOMMUNICATION_STAF	0	interval	num	Full-time telecommunication staff(% of total labor fo

Cluster's membership is shown in the table below.

CLUSTER	Frequency of Cluster	Root-Mean-Square Standard Deviation	Maximum Distance from Cluster Seed	Nearest Cluster	Distance to Nearest Cluster	Total telecom services revenue(Current U
1	72	1.046794019	8.0433260935	2	4.0488755304	1410038116.9
2	108	0.5865984355	9.8373259377	1	4.0488755304	242498908.25

According to the generated report, contents of the cluster 1 are as follows(72 data points):Czech republic (1993-2002), Estonia (1994-2002), Hungary (1993-2002), Bulgaria (2002), Latvia (1994, 1995, 1997-2002), Lithuania (1999-2002), Slovenia (1993-2002), Poland (1993-2002), Slovak Republic (1995-8, 2000-2).

Decision tree, as well as set of decision rules, is provided below.



IF Annual telecom investment(Current US $ per telecom worker)
< $9,610
THEN
NODE : 2
N : 110
2 : 96.4%
1 : 3.6%

IF $9,610 <= Annual telecom investment(Current US $ per telecom worker)
THEN
NODE : 3
N : 70
2 : 2.9%
1 : 97.1%

7.1.3 Clustering: Summary of the Analysis

Table 1 Clustering: Contents of the 2 clusters

Data Set	**Contents of the first cluster**	**Contents of the second cluster**
First 5-year data set (1993 – 1997), 2 Clusters	**Albania (1993-1997)** **Armenia(1993-1997)** **Azerbaijan(1993-1997)** **Belarus(1993-1997)** **Bulgaria(1993-1997)** Czech Republic (1993) Estonia (1993-1995) **Kazakhstan(1993-1997)** **Kyrgyz Rep (1993-1997)** **Lithuania(1993-1997)** Latvia (1996) **Moldova(1993-1997)** **Romania(1993-1997)** Slovak Rep(1993-1995) **Ukraine(1993-1997)** Poland (1993-1994)	Czech Rep(1994-1997) Estonia (1996-1997) **Hungary (1993-1997)** Latvia (1994-5, 1997) Poland (1995-1997) Slovak Rep (1996-1997) **Slovenia (1993-1997)**.
Second 5-year data set (1998 – 2002), 2 Clusters	**Albania (1998-2002)** **Armenia(1998-2002)** **Azerbaijan(1998-2002)** **Belarus(1998-2002)** Bulgaria(1998-2001) Slovak Rep(1999) **Kazakhstan(1998-2002)**	**Czech Rep (1998-2002)** **Estonia (1998-2002)** **Hungary (1998-2002)** **Latvia (1998-2002)** **Poland (1998-2002)** Slovak Rep(1998,2000-2002)

	Kyrgyz Rep (1998-2002) Lithuania(1998, 2001) Moldova(1998, 1999,2001) **Romania(1998-2002)** **Ukraine (1998-2001)**	**Slovenia (1998-2002)** Ukraine (2002) Romania (2000, 2002) Moldova (2002) Lithuania(1999, 2000,2002) Bulgaria (2002)
<u>*10-year data set(1993-2002), 2 clusters*</u>	**Albania (1993-2002)** **Armenia(1993-2002)** **Azerbaijan(1993-2002)** **Belarus(1993-2002)** Bulgaria(1993-2001) Slovak Rep(1993, 1994, 1999) **Kazakhstan(1993-2002)** **Kyrgyz Rep (1993-2002)** Latvia (1993, 1996) Lithuania (1993-1998) Moldova **(1993-2002)** **Romania(1993-2002)** **Ukraine (1993-2001)**	**Czech rep (1993-2002)** Estonia (1994-2002) **Hungary (1993-2002)** Bulgaria (2002) Latvia (1994, 1995, 1997-2002) Lithuania (1999-2002) **Slovenia (1993-2002)** **Poland (1993-2002)** Slovak Rep(1995-1998, 2000-2002)

Below is the table listing the clustering solutions that were obtained, as well as the variables that were used for splitting of the data set into the clusters.

Table 2 Clustering: Split variables used in separating the data set into 2 clusters

Data Set	**The Variables that were used for Splitting**
First 5-years data set (1993-	• Full-time telecom staff,

1997), Automatic setting, 9 clusters	• Annual telecom investment per worker • Annual telecom investment per telecom worker
First 5-year data set (1993-1997), 5 clusters	• Total telecom service revenue per worker
First 5-year data set (1993-1997), 4 clusters	• Total telecom service revenue per worker
First 5-year data set (1993-1997), 3 clusters	• Annual telecom investment per telecom worker
First 5-year data set (1993-1997), 2 clusters	• Annual telecom investment per telecom worker
Second 5-year data set (1998 – 2002), Automatic setting,7 clusters	• Full-time telecom staff, • Total telecom service revenue per telecom worker • Annual telecom investment per capita
Second 5-year data set (1998 – 2002), 5 clusters	• Total telecom service revenue per worker • Annual telecom investment per worker
Second 5-year data set (1998 – 2002) , 4 Clusters	• Total telecom service revenue • Annual telecom investment per capita
Second 5-year data set (1998 – 2002) , 3 clusters	• Total telecom services revenue per telecom worker
Second 5-year data set (1998 – 2002), 2 Clusters	• Annual telecom investment per capita
10-year data set (1993-2002), 5 clusters	• Total telecom service revenue per worker • Annual telecom investment per worker • Annual telecom investment (% of GDP
10-year data set (1993-200)2, 4 clusters	• Total telecom service revenue per worker • Annual telecom investment per worker
10-year data set(1993-2002), 3 clusters	• Total telecom services revenue per worker

10-year data set(1993-2002), 2 clusters	• Annual telecom investment per telecom worker

The most frequently used for the partitioning of the data variables are summarized in the table below. For each of those variables we have as well calculated the number of times that each variable was used.

Table 3 Clustering: Frequency of use of the split variables

Variable used for splitting of the Data	**Frequency of use**
Total telecom service revenue per worker	**6**
Annual telecom investment per worker	**4**
Annual telecom investment per telecom worker	**4**
Annual telecom investment per capita	**3**
Total telecom service revenue per telecom worker	**2**

7.1.4 Clustering: conclusion and the analysis of the results

By using cluster analysis, we were able to come up with a solution that partitions both available to us data sets into two clusters. For both data sets, one of the clusters contained the data points completely representing Poland, Czech Republic, Hungary and Slovenia for 5- and 10-year periods. Thus, these results suggest that we are able to reject H01. Moreover, as it is illustrated by the examples of Slovenia and Estonia, transitional economies are not static in their regards to the relative position to the other countries of the sample. Therefore, these results suggest that we are able to reject H02 as well.

Finally, we should ask ourselves the following question: What is the significance of the separation of 18 transitional economies into the two clusters? One of the possible answers is provided in the research by Piatkowski (2003b), who concluded that in the period "between 1995 and 2000 ICT capital has most potently contributed to output growth in the Czech Republic, Hungary, Poland, and Slovenia." Thus, it could be suggested that we were able to separate 18 transitional economies into the two groups, one group of transitional economies that benefits the most from the investments in telecom, and another group where the benefits are less pronounced.

7.2 Decision Trees: Characteristics of the "Leaders"

In this part of our analysis, we aim to identify the characteristics of those TEs, which are able to benefit the most from the investments in ICT. " We use Decision Trees as a tool to determine the characteristics of the "leaders" and proceeded as follows.

7.2.1 DT: Data

First, we identified the subset of all available to us data with the purpose of balancing out two goals. First, the subset must be large enough to be representative of 18 countries that we use in our analysis. Second, it must contain as much variables as possible. We ended up using only 14 countries out of our sample of 18 because the 'richest' subset of 5-year data from 1998 to 2002 was available only for these 14 countries. In our analysis we were able to use 34 variables, which are listed below.

Variables used in the Analysis, 5-year data set, 14Transitional Economies:

1. Exports of Computer, communications and other services (% of commercial service exports)
2. High-technology exports (% of manufactured exports)
3. Imports of Computer, communications and other services (% of commercial service imports)
4. Military expenditure (% of GDP)
5. Military personnel (% of total labor force)
6. Fixed line and mobile phone subscribers (per 1,000 people)
7. International telecom, outgoing traffic (minutes per subscriber)
8. Internet users (per 1,000 people)
9. Mobile phones (per 1,000 people)
10. Telephone mainlines (per 1,000 people)
11. Telephone mainlines per employee
12. Health expenditure per capita (current US$)
13. Health expenditure, private (% of GDP)
14. Health expenditure, public (% of GDP)
15. Health expenditure, total (% of GDP)

16. Immunization, DPT (% of children ages 12-23 months)

17. Immunization, measles (% of children ages 12-23 months)

18. Pupil-teacher ratio, primary

19. School enrollment, secondary (% gross)

20. School enrollment, tertiary (% gross)

21. Research and development expenditure (% of GDP)

22. Researchers in R&D (% of total labor force)

23. Technicians in R&D (% of total labor force)

24. Roads, paved (% of total roads)

25. Roads, total network (km)

26. Full-time telecommunication staff(% of total labor force)

27. Annual telecom investment(% of GDP in current US $)

28. Urban population (% of total)

29. Urban population growth (annual %)

30. Population growth (annual %)

31. Foreign direct investment, net inflows (% of GDP)

32. GDP growth (annual %)

33. GDP per capita (constant 2000 US$)

34. GDP per capita growth (annual %)

As for the 10-year set of data, we ended up using 16 countries: Albania, Armenia, Azerbaijan, Belarus, Bulgaria, Estonia, Hungary, Kazakhstan, Kyrgyz Republic, Latvia, Lithuania, Moldova, Poland, Romania, Slovenia, and Ukraine. We were able to put together set of 12 variables, which are listed below.

Variables used in the Analysis, 10-year data set, 16 Transitional Economies:

1. Fixed line and mobile phone subscribers (per 1,000 people)
2. International telecom, outgoing traffic (minutes per subscriber)
3. Military expenditure (% of GDP)
4. Military personnel (% of total labor force)
5. Mobile phones (per 1,000 people)
6. Physicians (per 1,000 people)
7. Telephone mainlines (per 1,000 people)
8. Telephone mainlines per employee
9. International tourism, expenditures (current US$)
10. GDP growth (annual %)
11. GDP per capita (constant 2000 US$)
12. Annual telecom investment(% of GDP in Current US $)

7.2.2 DT: Description of the Method

Second, we created a binary dummy variable, which was set as a "target" of the DT analysis. In our 5-year data set, we assigned the values of '1' of the target variable to the following countries: Czech Republic, Hungary, Latvia, Poland, Slovak Republic, and Slovenia. The values of '0' were assigned to Ukraine, Romania, Lithuania, Kyrgyz

Republic, Bulgaria, Belarus, Armenia, and Azerbaijan. As for the 10-year data set, we have assigned values of '1' to Estonia (1994-2002), Hungary (1993-2002), Latvia (1994, 1995, 1997-2002), Lithuania (1999-2002), Slovenia (1993-2002), and Poland (1993-2002).

Finally, we run DT analysis to generate a model. Once the model was generated, we proceeded to identify the variable that was used for the first split and set its status to "don't use" in the Decision Tree node. Then the process was repeated.

In our evaluation of the resulting models, we were looking for those variables, splits along which resulted in the cleanest possible separation of the data set according to the value of the target variable. Moreover, we have calculated the average value of the split variable, as well as the (approximate) percentile within which value of the split falls.

Next, we provide the results of our analysis in the form of the resulting for each model DT, accompanied by the relevant information about the split variables. First, we introduce the results of the analysis of the 5-year data set, followed by the 10-year data set.

7.2.3 DT: Results of the Analysis

<u>5-year data set, 14 TEs</u>

<u>Step 1: Complete data set.</u>

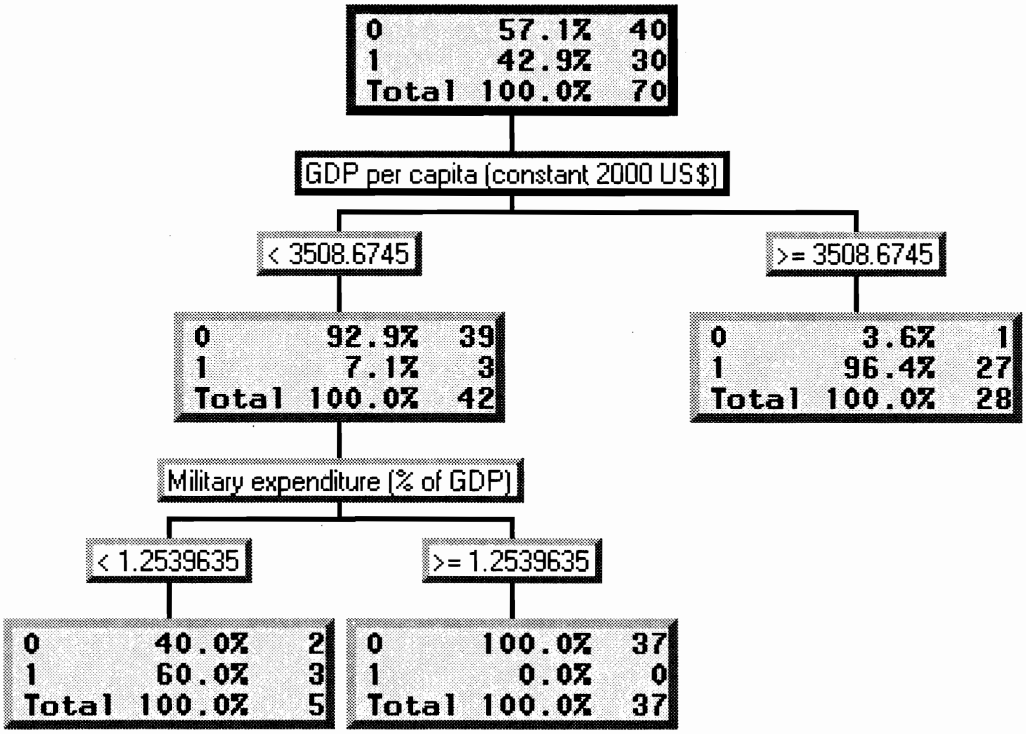


Table of the Split Variables, Average Values of the Split Variables, and Approximate Percentile at which Split was made

GDP per capita, Average:	$2920.674
GDP per capita, 60th percentile:	$3502.359
Military expenditure, Average:	1.980002%
Military expenditure, Bottom 10%:	1.260230%

According to this model over 90% of the "leaders" have GDP per capita above 60th percentile, which implies that "leaders" are wealthier, in general, group of TEs than "majority."

Step 2: "GDP per capita" variable removed from the data set

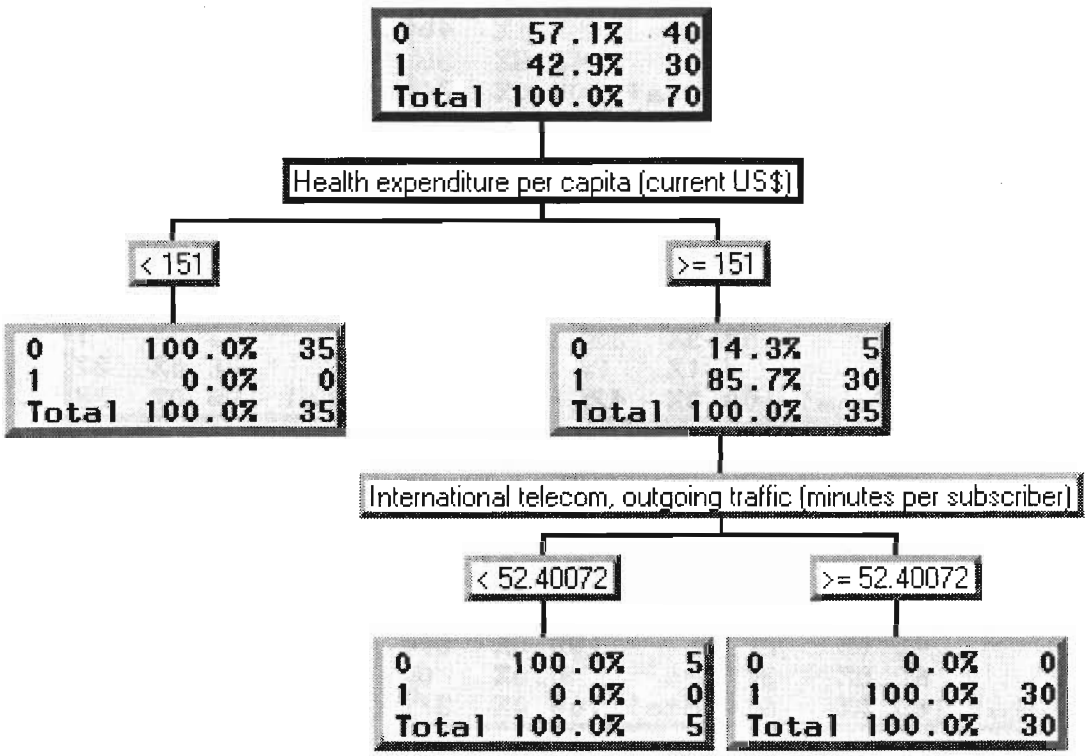


Table of the Split Variables, Average Values of the Split Variables, and Approximate Percentile at which Split was made

Health expenditure per capita, Average:	$208.2857
Health expenditure per capita, 50th percentile:	$151
International telecom, outgoing traffic, Average:	114.1463
International telecom, outgoing traffic, 35th percentile:	51.5195

This model shows that "leaders" have higher level of health expenditure than the most of the "majority." But even in the cases where the "majority" and the "leaders" have the same level of health expenditure per capita, "majority" has a lower level of international telecom traffic (i.e., "majority" places lower number of the international phone calls).

Step 3: "Health expenditure per capita" variable removed from the data set

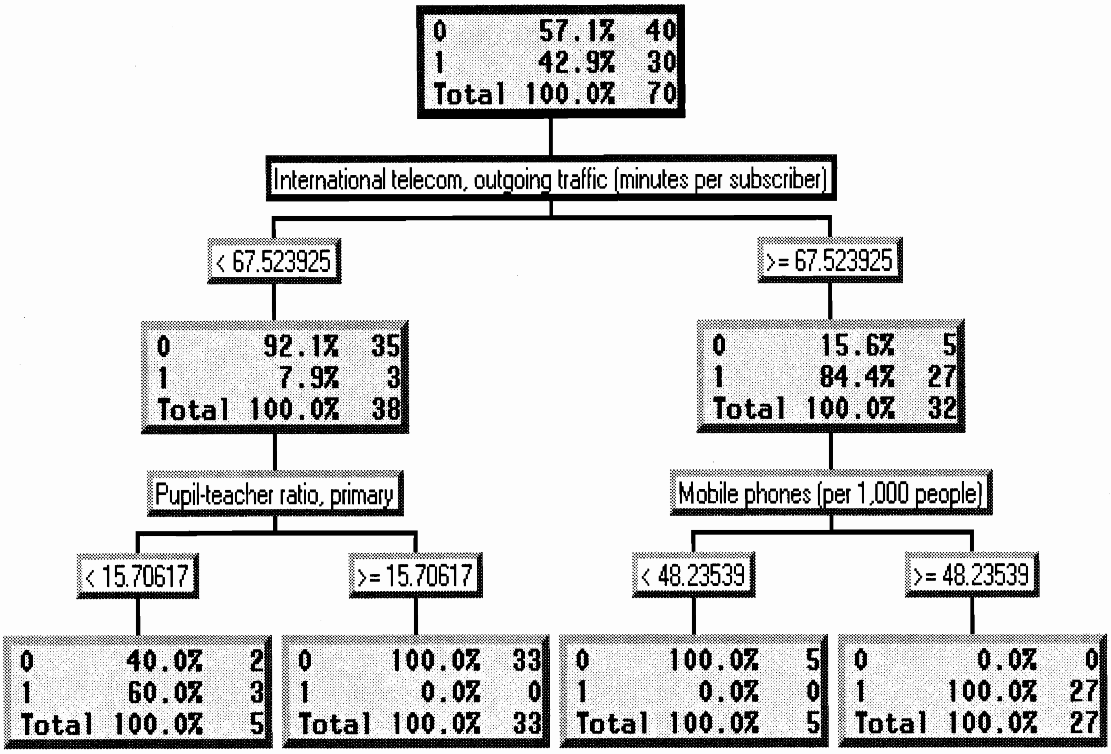


Table of the Split Variables, Average Values of the Split Variables, and Approximate Percentile at which Split was made

International telecom, outgoing traffic, Average:	114.1463
International telecom, outgoing traffic, 55th percentile:	68.27844
Mobile phones, Average:	183.2279
Mobile phones,30th percentile:	43.64353
Mobile phones,35th percentile:	50.36778

Pupil-teacher ratio, Average:	16.7074
Pupil-teacher ratio, Top 30%:	15.50373

According to this model 90% of the “leaders” have a higher level of international telecom traffic and have higher number of mobile phones, while most of the “majority” have higher pupil-teacher ratio.

Step 4: “International telecom, outgoing traffic” variable removed from the data set

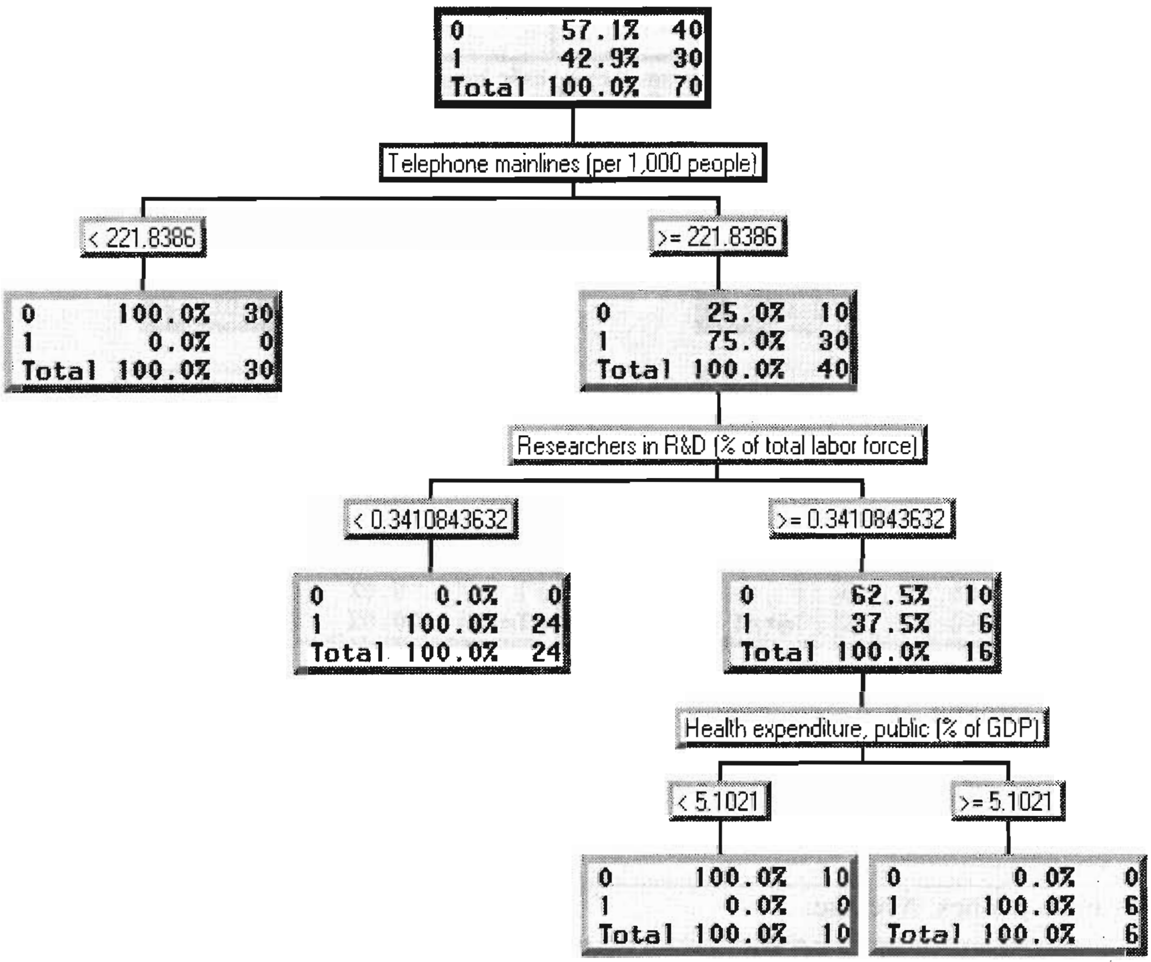


Table of the Split Variables, Average Values of the Split Variables, and Approximate Percentile at which Split was made

Telephone mainlines, Average	243.0411
Telephone mainlines, 40th percentile	209.8646
Telephone mainlines, 45th percentile	248.6087
Researchers in R&D, average	0.293753721%
Researchers in R&D,70th percentile	0.339521721%
Health expenditure, public, Average	3.96483%
Health expenditure, public, 75th percentile	5.1069%

Based on this model "majority" has a less number of telephone mainlines than "majority," while most of the "majority" has less percentage of the researchers among their labor force involved. This is surprising, for we would have expected to see more "leaders" than the "majority" in the top 30% of the percentage of the researchers in R&D among the labor force.

Step 5: "Telephone mainlines" variable removed from the data set

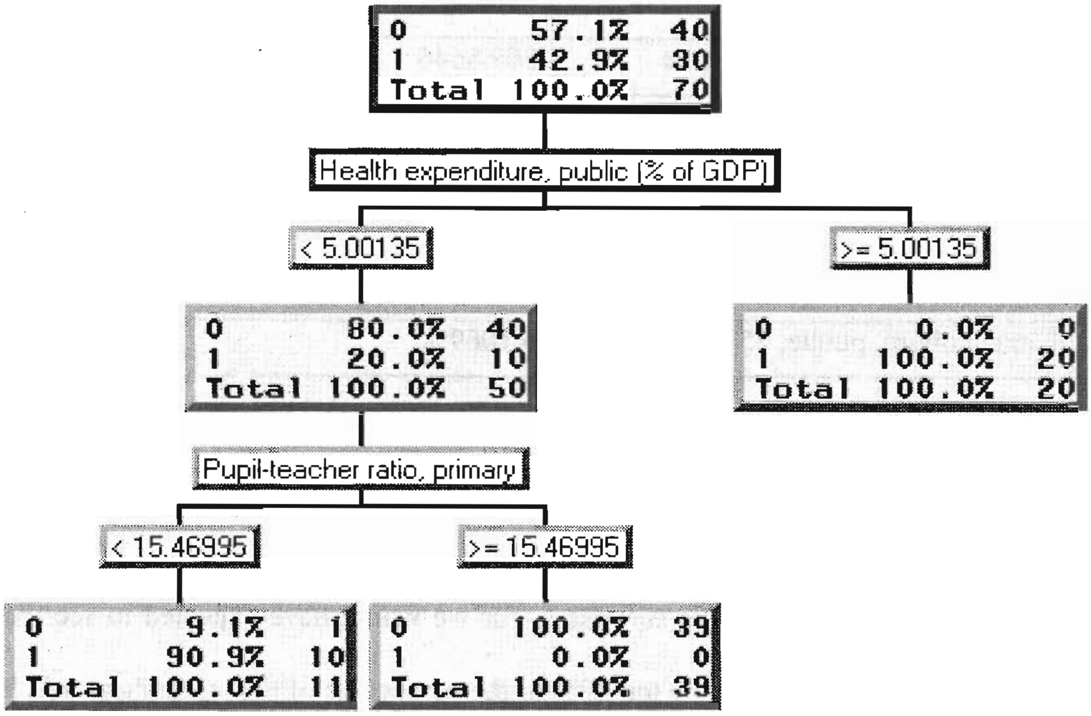


Table of the Split Variables, Average Values of the Split Variables, and Approximate Percentile at which Split was made

Health expenditure, public, Average	3.96483%
Health expenditure, public, 75th percentile	5.1069%
Pupil-teacher ratio, Average:	16.7074
Pupil-teacher ratio, Top 30%:	15.50373

This model tells us that the most of the "leaders" spend more money than the "majority" on Health care, while having more teachers per pupil in primary education.

Step 6: "Health expenditure, public" variable removed from the data set

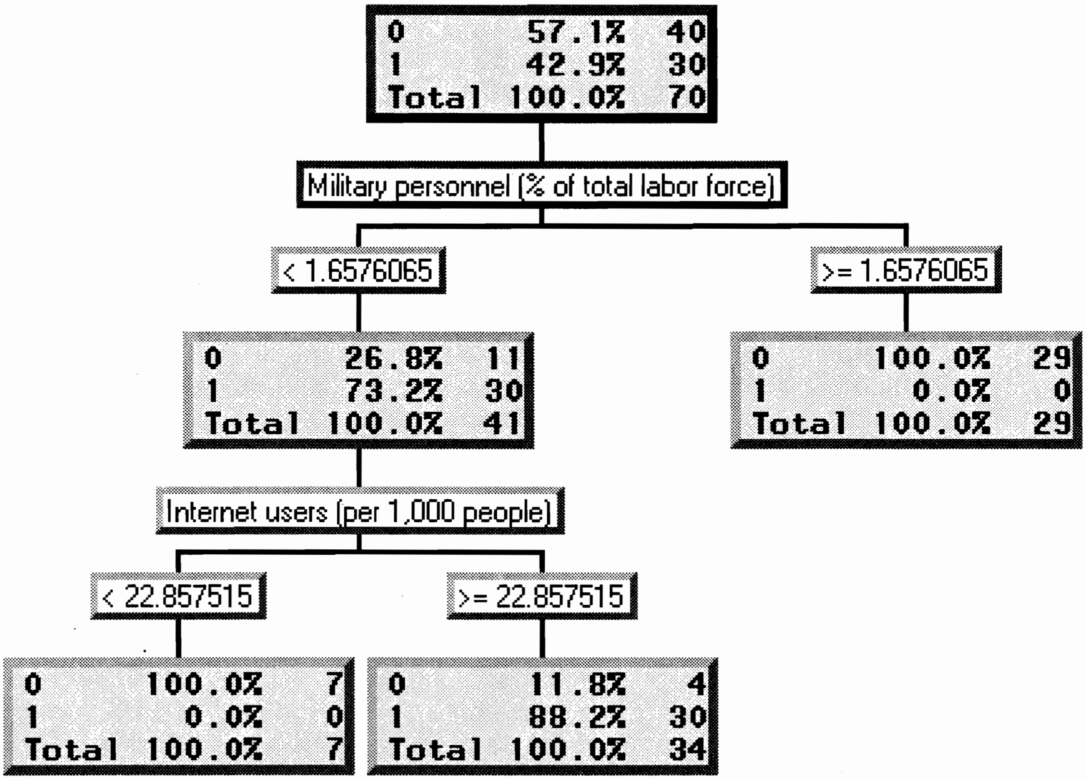


Table of the Split Variables, Average Values of the Split Variables, and Approximate Percentile at which Split was made

Military personnel, Average	1.655823%
Military personnel,, 60th percentile	1.665328%
Internet users, Average:	64.93297
Internet users, 35th percentile	22.92829

According to this model most of the "majority" have larger percentage of their labor force being military personnel, while the "leaders" have higher percentage of the internet users among the population.

Step 6: "Military personnel" variable removed from the data set

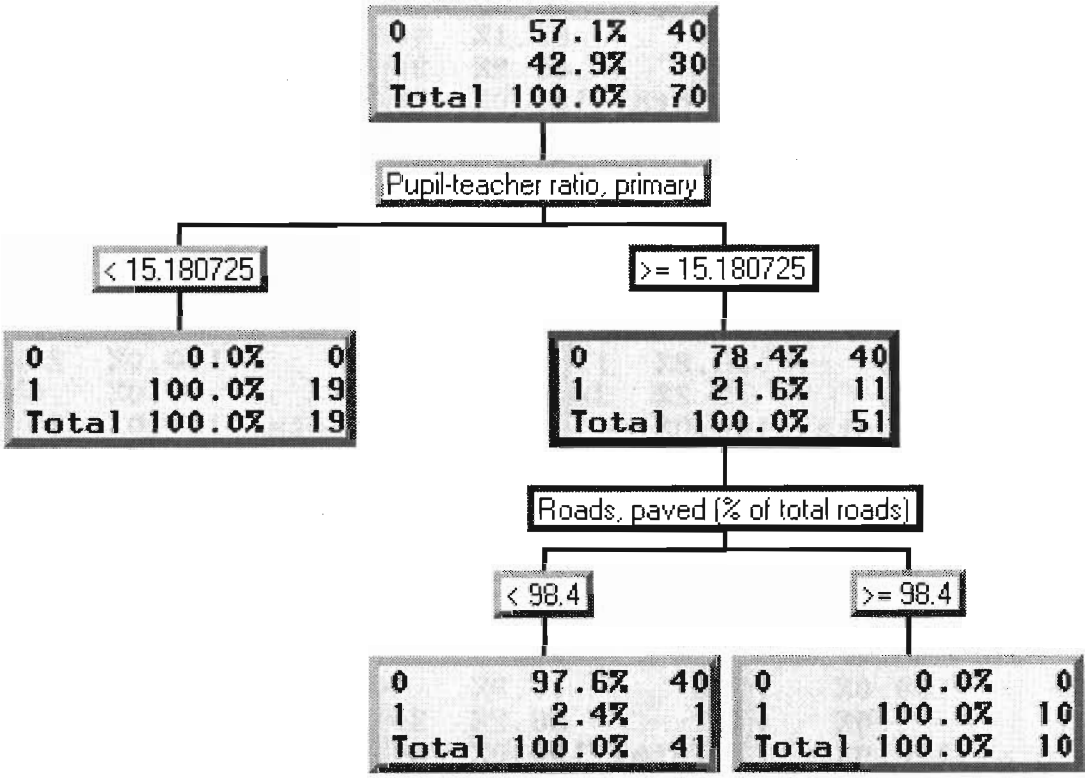


Table of the Split Variables, Average Values of the Split Variables, and Approximate Percentile at which Split was made

Pupil-teacher ratio, Average	16.7074
Pupil-teacher ratio, Top 30%	15.50373
Roads paved, Average	95.17429%
Roads paved, 79th percentile	98.38100%

This model shows us that in general the "leaders" have more teachers per pupil in the system of primary education, and in those cases where 'majority" and the "leaders" are comparable in this area, "majority" has a lower percentage of the paved roads than "leaders" do.

Step 6: "Pupil-teacher ratio, primary" variable removed from the data set

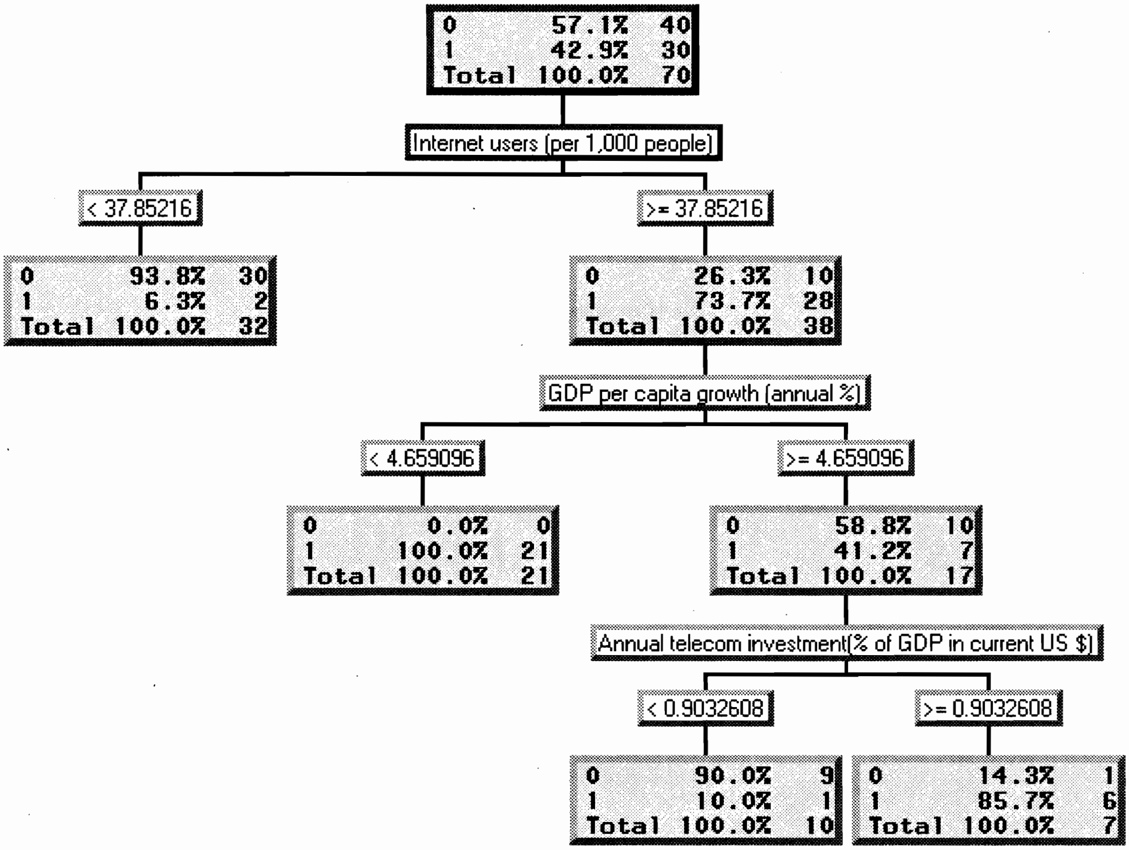


Table of the Split Variables, Average Values of the Split Variables, and Approximate Percentile at which Split was made

Internet users, Average	64.93297
Internet users, 35th percentile	22.92829
Internet users, 45th percentile	36.95068
GDP per capita growth, Average	4.678119%
GDP per capita growth, 50th percentile	4.586358%
Annual telecom investment, Average	0.780932959%
Annual telecom investment, 65th percentile	0.887426121%

Based on this model we could see that three forth of the countries in the "majority" group have lower percentage on Internet users. We could also see that in those cases where the percentage of Internet users is compatible between the two groups, "leaders" have lower level of GDP per capita growth (possible indication of the wealthier economies), while spending more on telecoms.

Step 7: "Internet users per 1000 people" variable removed from the data set

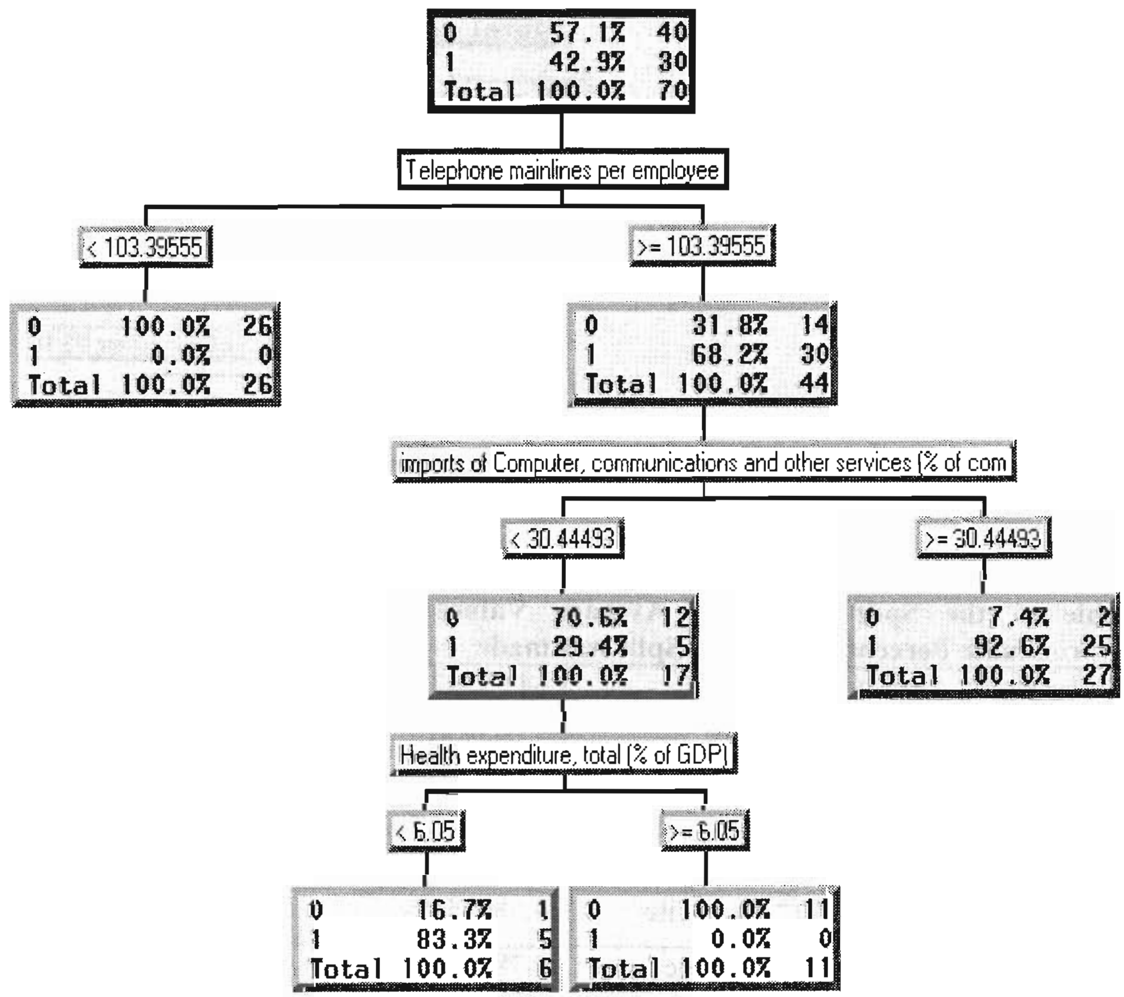


Table of the Split Variables, Average Values of the Split Variables, and Approximate Percentile at which Split was made

Telephone mainlines per employee, Average	142.5712
Telephone mainlines per employee, 37^{th} percentile	103.412

Imports of Computer, comm. and other services, Average	37.65674%
Imports of Computer, comm. and other services, 31st percentile	29.57723%
Imports of Computer, comm. and other services, 32nd percentile	31.60094%

By looking at this model, we could determine that only 35% of the "majority" has the comparable to "leaders" infrastructure of the Telephone mainlines per employee. Moreover, even for those TEs with the comparable infrastructure, "leaders" import more ICT-related services than the "majority."

Step 8:"Telephone mainlines per employee" variable removed from the data set

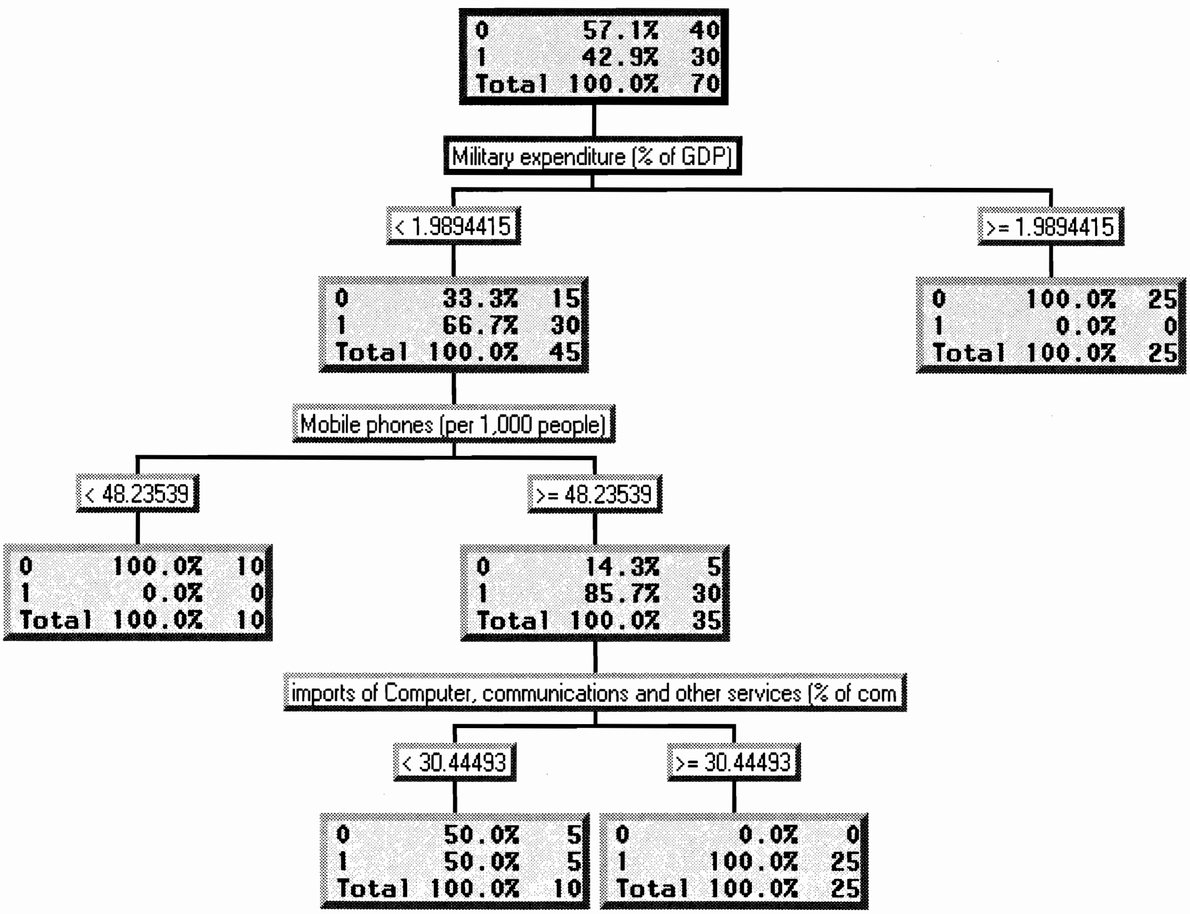


Table of the Split Variables, Average Values of the Split Variables, and Approximate Percentile at which Split was made

Military expenditure, Average	1.980002%
Military expenditure, 64th percentile	1.985057%
Military expenditure, 65th percentile	1.993955%
Mobile phones, Average	183.2279
Mobile phones, 34th percentile	48.86391
Imports of Computer, comm. and other services, Average	37.65674%
Imports of Computer, comm. and other services, 31st percentile	29.57723%
Imports of Computer, comm. and other services, 32nd percentile	31.60094%

Sixty percent of the "majority" have spend more on Military, and where the spending is comparable, only 12 % of the "majority" have the same number of the mobile phones per 1000 people as the "leaders."

Step 9:"Military expenditure" variable removed from the data set

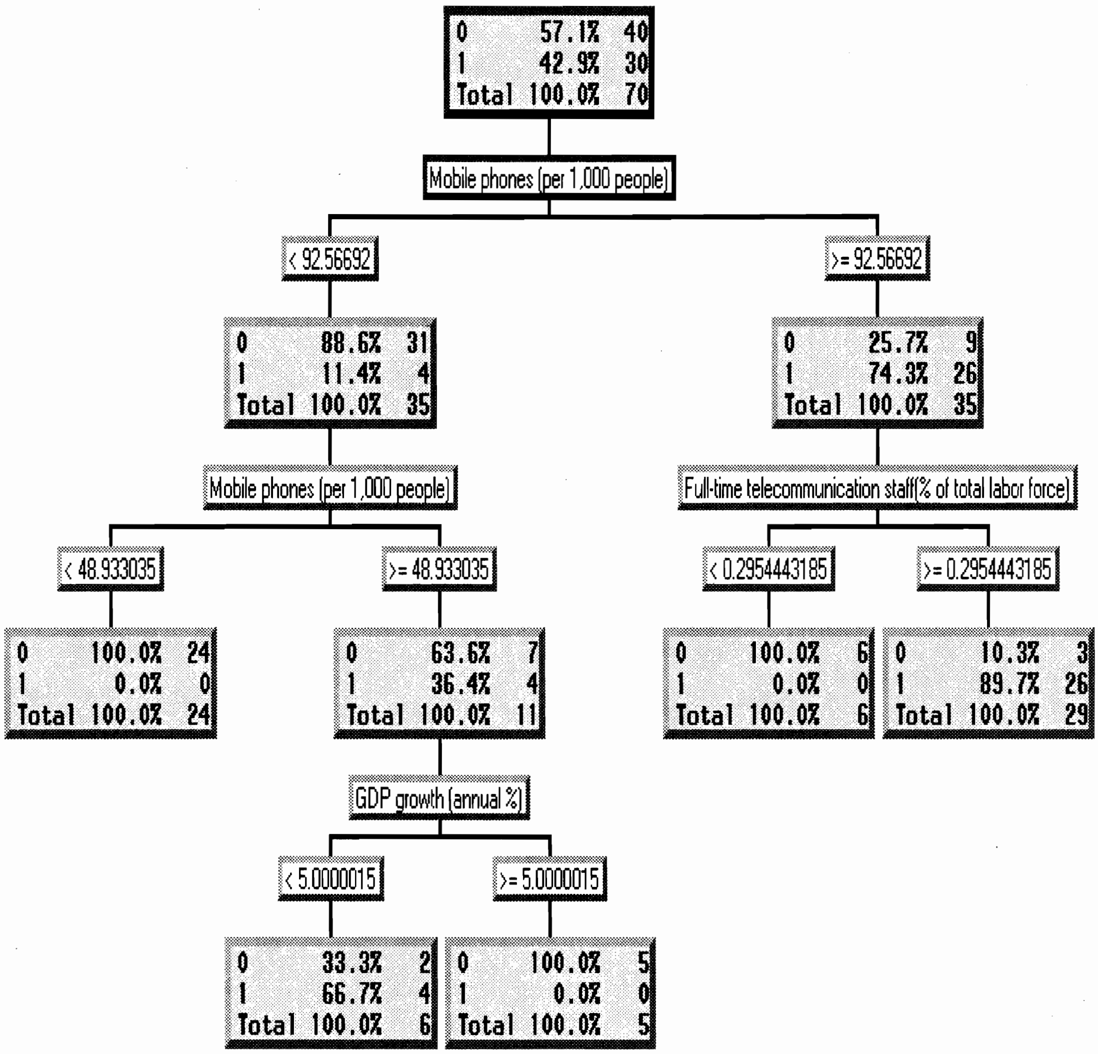


Table of the Split Variables, Average Values of the Split Variables, and Approximate Percentile at which Split was made

Mobile phones, Average	183.2279
Mobile phones, 50th percentile	92.56692
Mobile phones, 34th percentile	48.86391
Full-time telecommunication staff, Average	0.394519203%
Full-time telecommunication staff, 15th percentile	0.303569884%

Almost 90% of the "leaders" have larger number of mobile phones per 1000 people, and have a higher percentage of Full-time telecom staff as a percentage of the labor force, while 60% of the "majority" located in the bottom third of the sample in terms of the number of mobile phones per 1000 people.

Step 10:"Mobile phones per 1000 people" variable removed from the data set

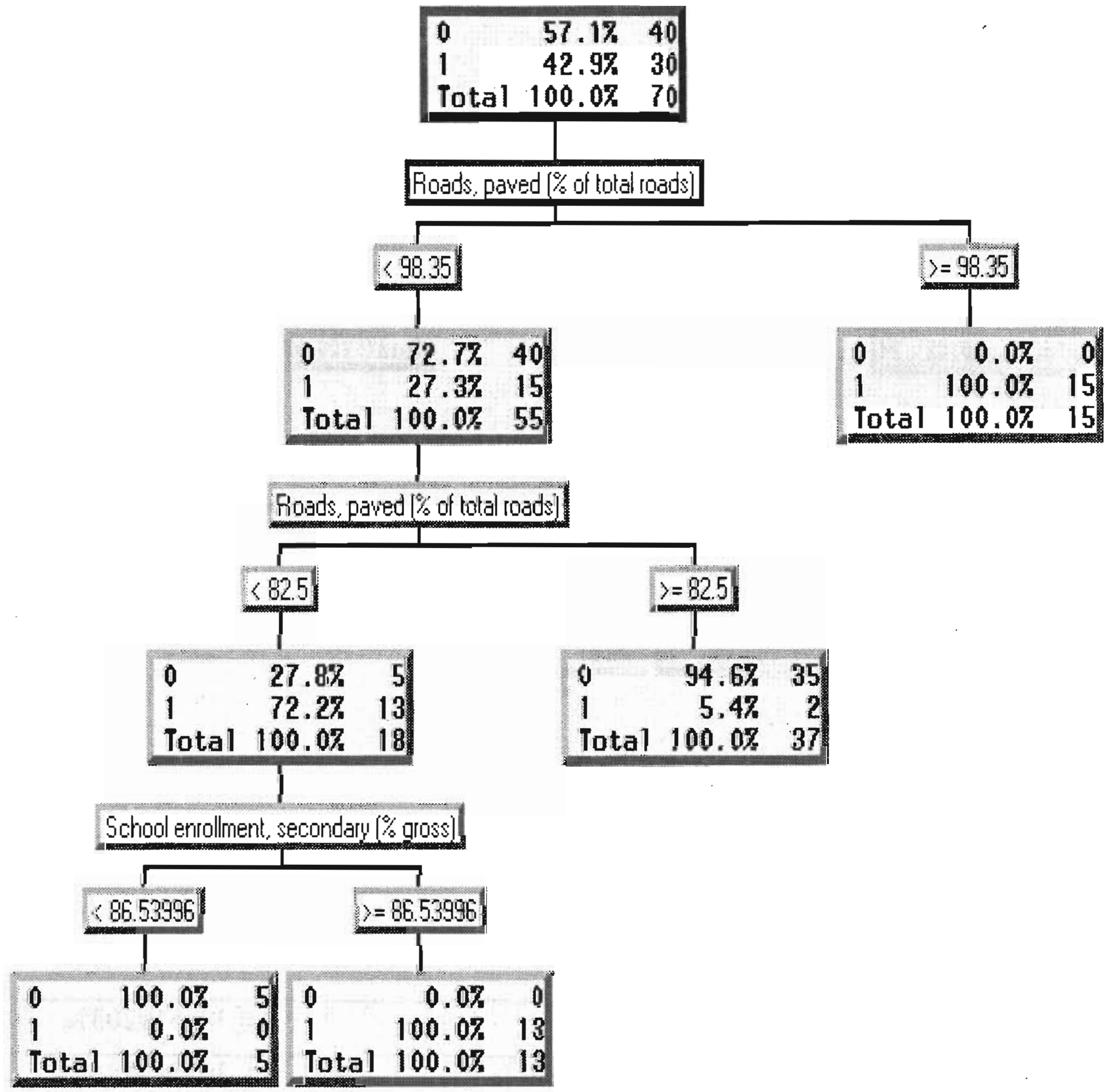


Table of the Split Variables, Average Values of the Split Variables, and Approximate Percentile at which Split was made

Roads, paved, Average	95.17429%
Roads, paved, 79th percentile	98.381%
Roads, paved, 26th percentile	86.196%
Roads, paved, 25th percentile	80.4%
School enrollment, secondary, Average	92.40741%
School enrollment, secondary, 30th percentile	86.58104%

One half of the TEs of the "leaders" group have a higher percentage of the paved roads, and the 40 % that have lower percentage than the "majority" exhibit higher secondary school enrollment rate. This lower percentage of the paved roads could be indicative of the mountainous, or tourist-destination countries.

Step 11:"Roads, paved" variable removed from the data set

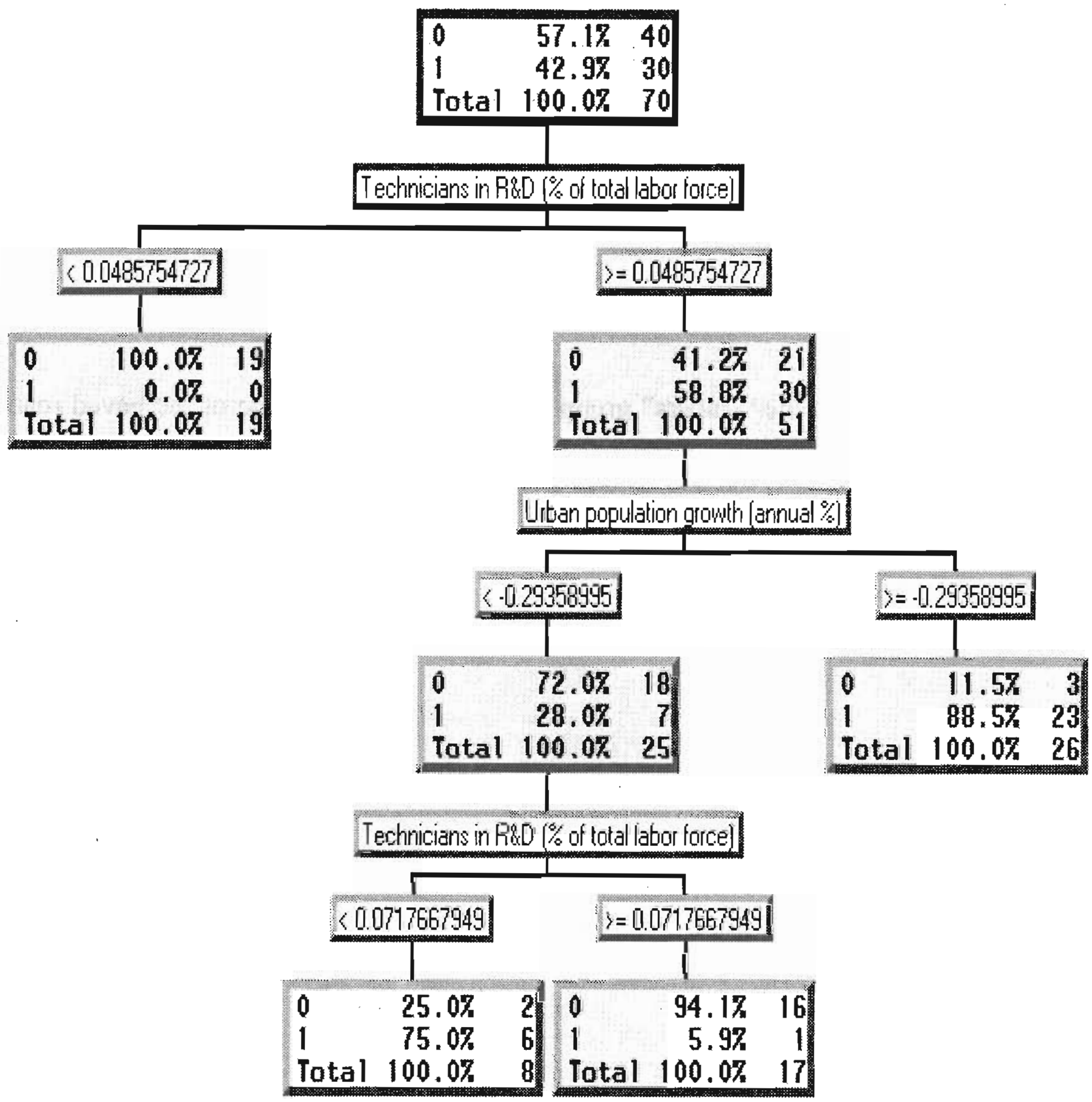


Table of the Split Variables, Average Values of the Split Variables, and Approximate Percentile at which Split was made

Technicians in R&D, Average	0.085047059%
Technicians in R&D, 44th percentile	0.071529012%
Technicians in R&D, 27th percentile	0.048610444%
Urban population growth, Average	-0.31052%
Urban population growth, 43rd percentile	-0.28804%

Urban population growth, 42nd percentile	-0.31009%

The "leaders" positioned somewhere in the middle in terms of the number of Technicians in R&D as the percentage of the total labor force. However, the most of the 'leaders' have higher growth rate of the urban population than "majority."

<u>Step 12:"Technicians in R&D" variable removed from the data set</u>

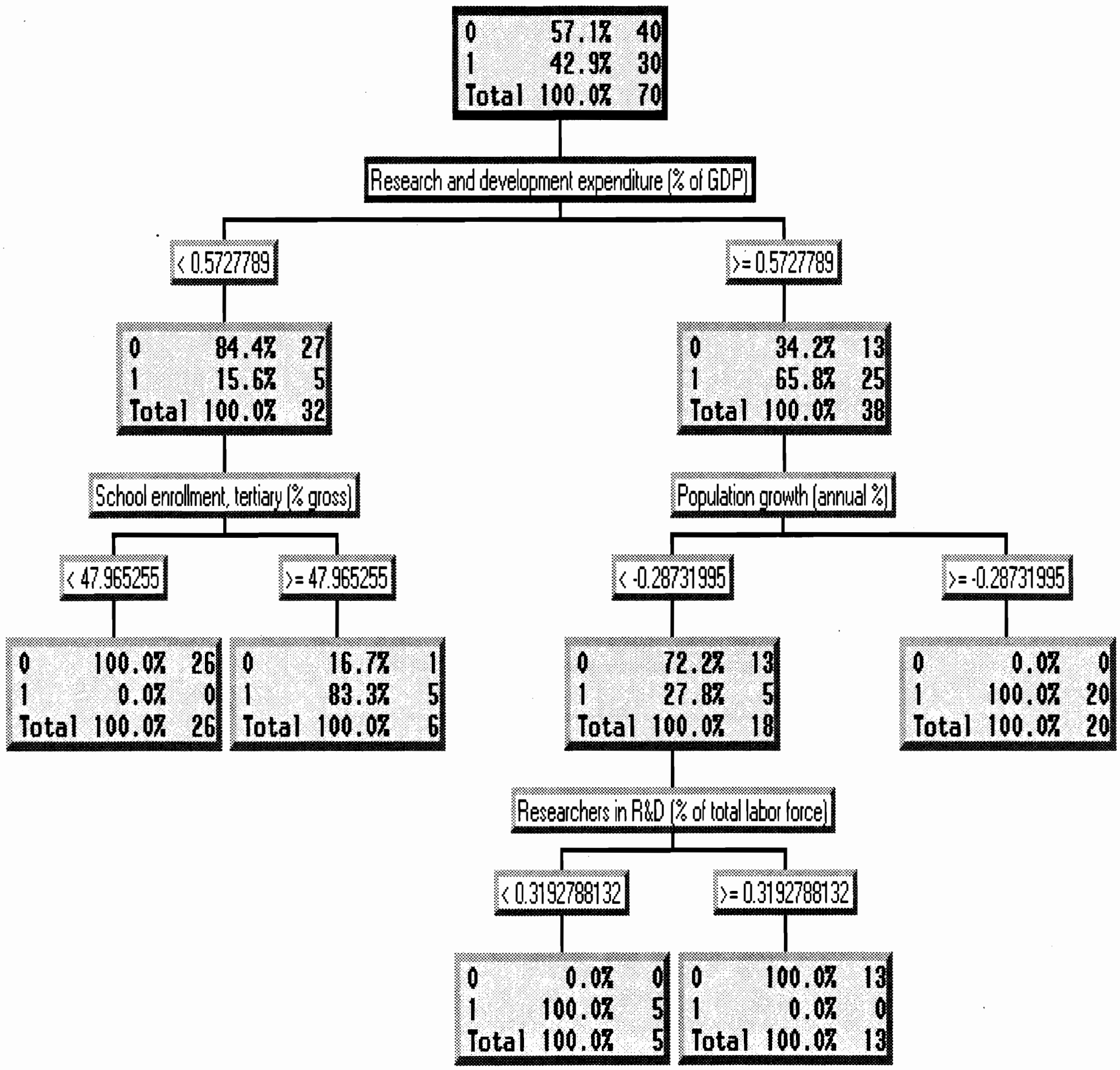


Table of the Split Variables, Average Values of the Split Variables, and Approximate Percentile at which Split was made

R&D expenditure, Average	0.672765%
R&D expenditure, 45th percentile	0.569465%
Population Growth, Average	-0.27597%
Population Growth, 53rd percentile	-0.28563%
School enrollment, tertiary, Average	42.69178%
School enrollment, tertiary, 70th percentile	47.72809%
Researchers in R&D, Average	0.293753721%
Researchers in R&D, 65th percentile	0.318635354%

On average, higher percent of the "leaders" spends more on R&D research, while having a higher rate of population growth.

10-year data set, 16TEs

Step 1: Complete data set

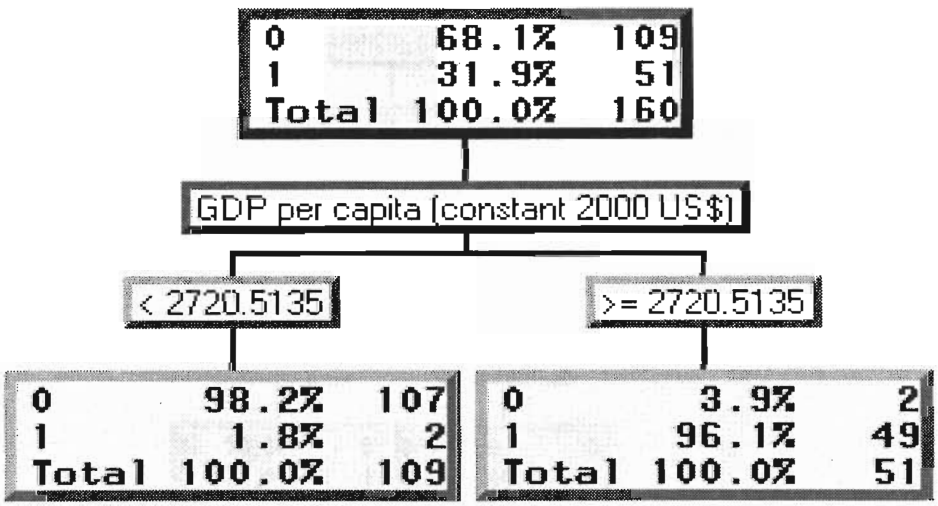


Table of the Split Variables, Average Values of the Split Variables, and Approximate Percentile at which Split was made

GDP per capita, Average	$2,182.51
GDP per capita, 50th percentile	$1404.279
GDP per capita, 68th percentile	$2715.5564

The first split yields the simple characteristic that separates the "leaders" from the "majority" in our set: "leaders" have higher GDP per capita.

<u>Step 2:"GDP per capita" variable removed from the data set</u>

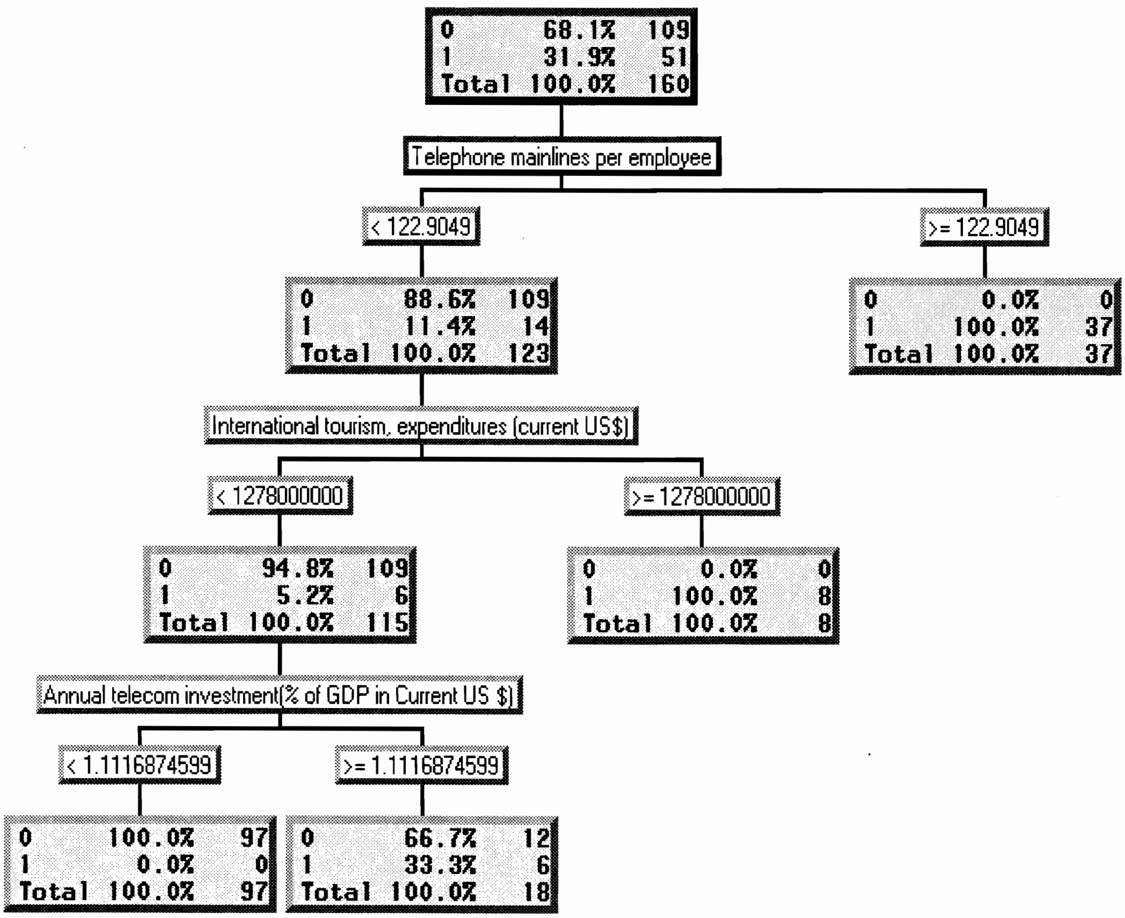


Table of the Split Variables, Average Values of the Split Variables, and Approximate Percentile at which Split was made

Telephone mainlines per employee, Average	97.78
Telephone mainlines per employee, 50^{th} percentile	85.1817
Telephone mainlines per employee, 77^{th} percentile	122.640748
International Tourism, expenditures, Average	$643,200,000.00
International Tourism, expenditures, 50^{th} percentile	$311,000,000.00
International Tourism, expenditures, 30^{th} percentile	$121,700,000.00
International Tourism, expenditures, 31^{st} percentile	$130,870,000.00
Annual Telecom investment, Average	%0.76
Annual Telecom investment, 77^{th} percentile	%1.104108227

Based on this model we could learn that 72% of the "leaders" occupy upper quartile in the variable "Telephone mainlines per employee," while the entire "majority" resides below in terms of this variable. In addition, another characteristic of any TEs within a "majority" group is that none of them have International tourism expenditures above 31^{st} percentile.

Step 3:"Telephone mainlines per employee" variable removed from the data set

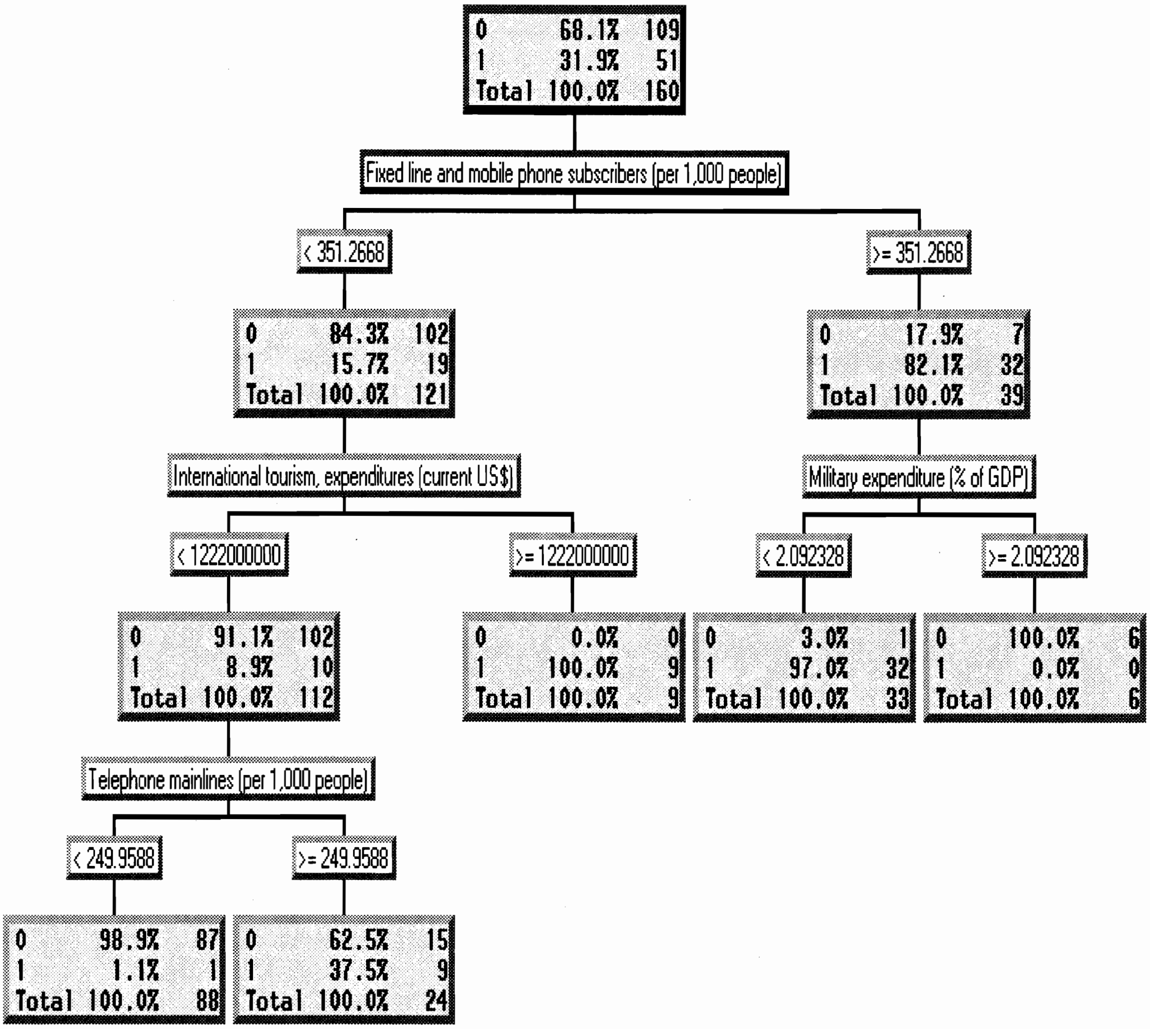


Table of the Split Variables, Average Values of the Split Variables, and Approximate Percentile at which Split was made

Fixed line and mobile phone subscribers, Average	281.24
Fixed line and mobile phone subscribers, 50th percentile	212.8928
Fixed line and mobile phone subscribers, 75th percentile	346.4978
International Tourism, expenditures, Average	$643,200,000.00
International Tourism, expenditures, 50th percentile	$311,000,000.00
International Tourism, expenditures, 30th percentile	$121,700,000.00
International Tourism, expenditures, 31st percentile	$130,870,000.00

Military expenditure, Average	%1.82
Military expenditure, 50th percentile	%1.656081
Military expenditure, 67th percentile	%2.090049
Telephone mainlines per 1000 people, Average	203.87
Telephone mainlines per 1000 people, 50th percentile	184.21165
Telephone mainlines per 1000 people, 62nd percentile	252.891742

Based on this model we determine that roughly 63% of the "leaders" and only 6% of the "majority" occupy upper quartile within the values for "Fixed line and mobile phone subscribers." However, even where some "leaders" are comparable to "majority," none of the "leaders" has Military expenditures higher than 67 percentile, while 85% of the "majority" has Military expenditures in 67 percentile.

On another hand, if "majority" and "leaders" have positioned in terms of "Fixed line and mobile phone subscribers" at the bottom three quartiles, the spending of the "majority" on International tourism is not above 30th percentile, while only 19% of the "leaders" have exhibited such low level of expenditures. Moreover, even where expenditures on tourism are comparable, 80% of the "majority" has lower number of Telephone mainlines per 1000 people than 17% of the "leaders."

Step 4:"Fixed line and mobile phone subscribers" variable removed from the data set

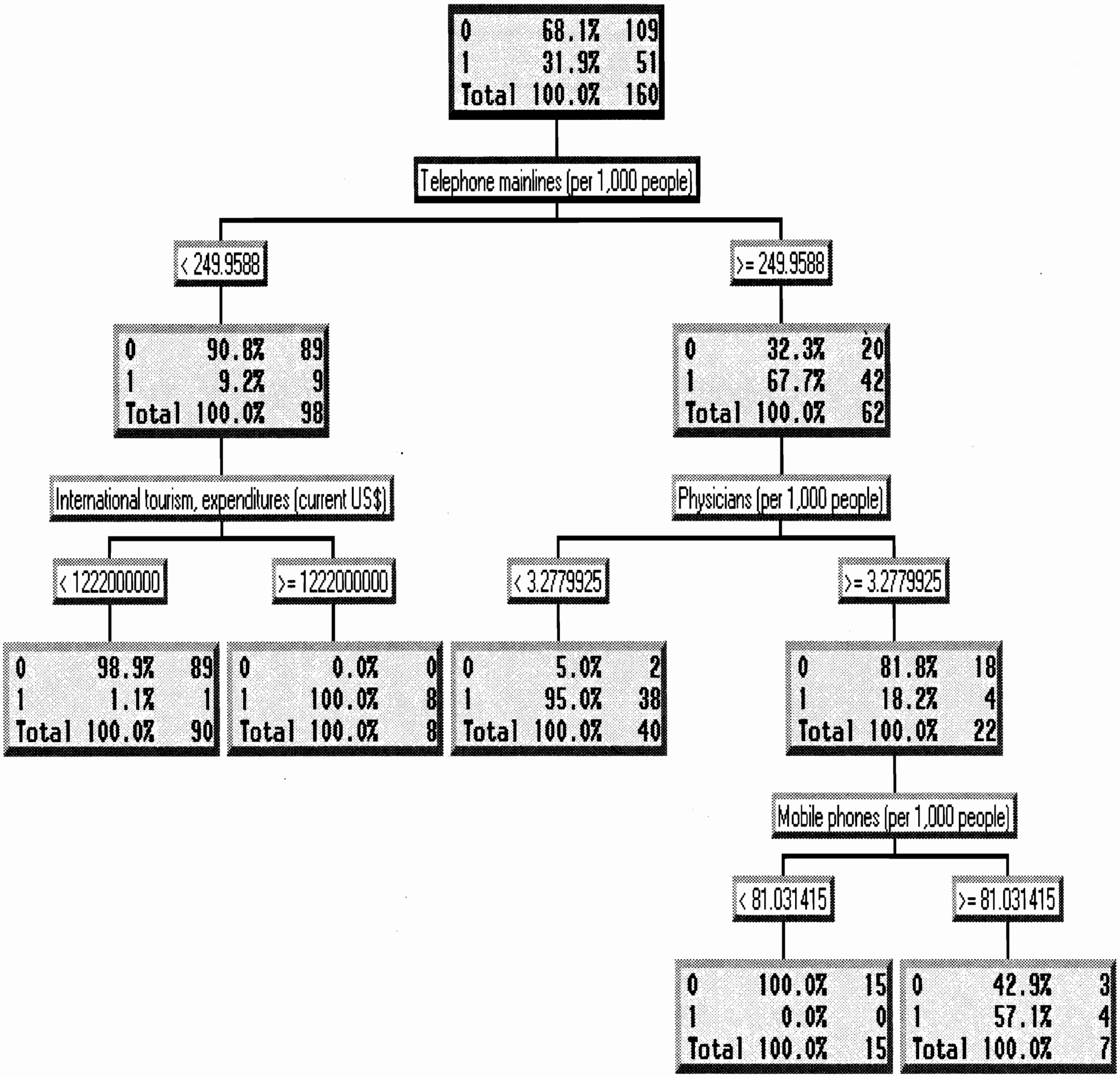


Table of the Split Variables, Average Values of the Split Variables, and Approximate Percentile at which Split was made

Telephone mainlines, Average	203.87
Telephone mainlines per 1000 people, 50^{th} percentile	184.21165
Telephone mainlines per 1000 people, 62^{nd} percentile	252.891742
International Tourism, expenditures, Average	$643,200,000.00
International Tourism, expenditures, 50^{th} percentile	$311,000,000.00

International Tourism, expenditures, 30th percentile	$121,700,000.00
International Tourism, expenditures, 31st percentile	$130,870,000.00
Physicians, Average	3.06
Physicians, 60th percentile	3.2352286
Mobile phones, Average	77.95
Mobile phones, 50th percentile	7.460091
Mobile phones, 76th percentile	80.4120116

This model is somewhat similar to the previous one. However, one of the new insights provided by this DT is that where "leaders" are comparable to the "majority" in terms of the telephone mainlines, 90% of the "majority" has higher number of physicians per 1000 people than 90% of the "leaders."

Step 5:"Telephone mainlines" variable removed from the data set

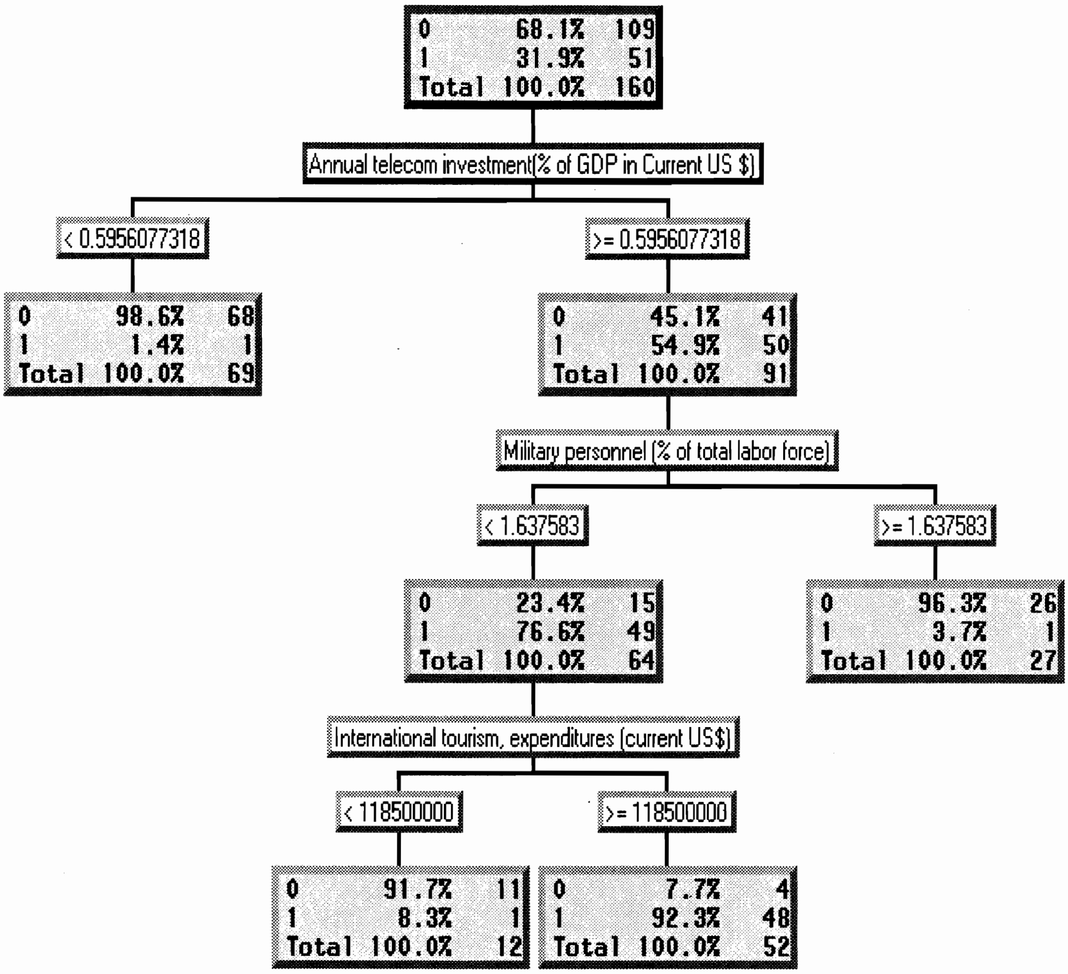


Table of the Split Variables, Average Values of the Split Variables, and Approximate Percentile at which Split was made

Annual Telecom investment, Average	%0.76
Annual Telecom investment, 50th percentile	%0.698127187
Annual Telecom investment, 43rd percentile	%0.594465742
Military personnel, Average	%1.68
Military personnel, 50th percentile	%1.35825
Military personnel, 60th percentile	%1.635823
International Tourism, expenditures, Average	$643,200,000.00

International Tourism, expenditures, 50th percentile	$311,000,000.00
International Tourism, expenditures, 30th percentile	$121,700,000.00
International Tourism, expenditures, 28th percentile	$117,560,000.00

Based on this diagram we could say that 94% of the leaders have higher levels of telecom investments, lower percentage of the labor force involved in the military, while spending more on the international tourism.

<u>Step 6:"Annual telecom investment" variable removed from the data set</u>

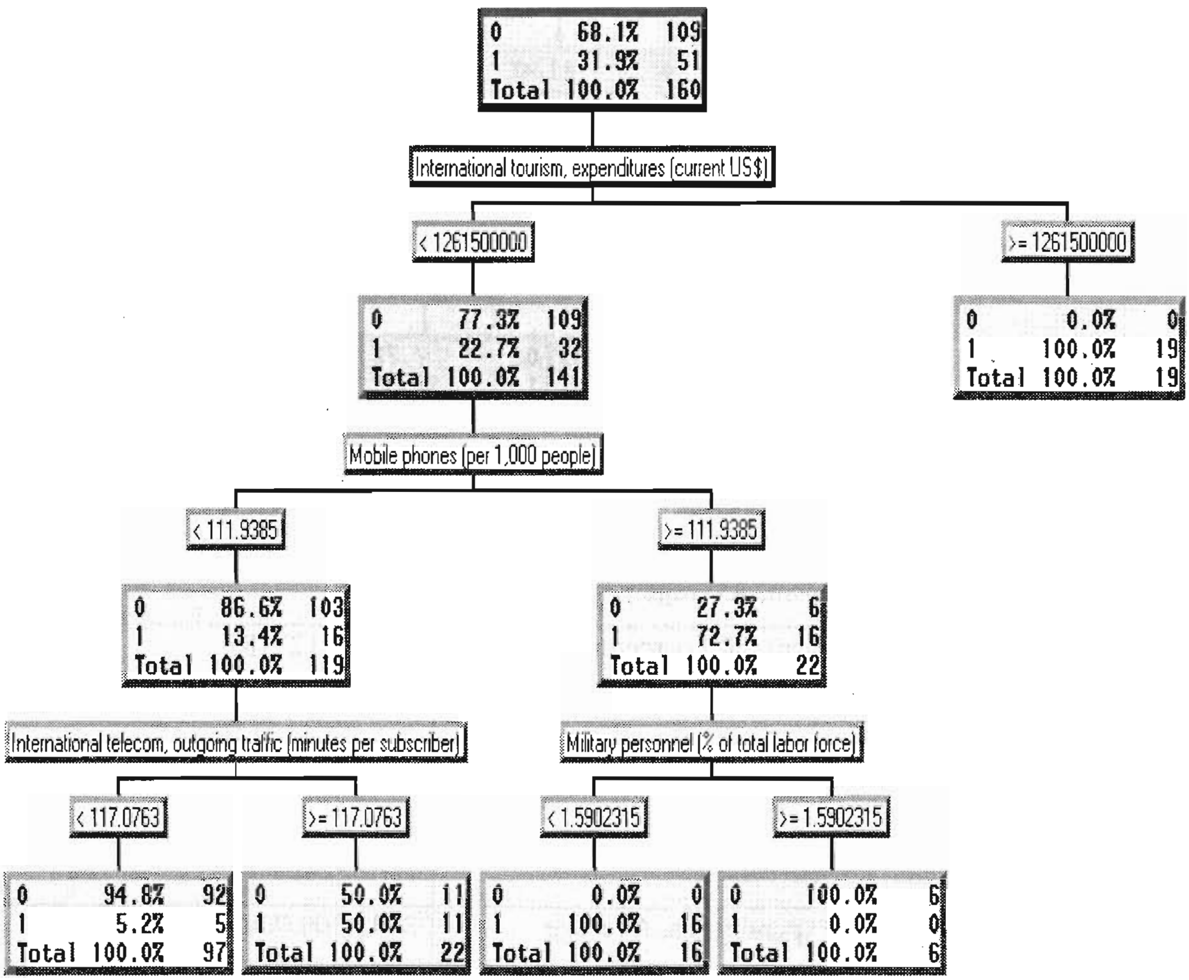


Table of the Split Variables, Average Values of the Split Variables, and Approximate Percentile at which Split was made

International Tourism, expenditures, Average	$643,200,000.00
International Tourism, expenditures, 50^{th} percentile	$311,000,000.00
International Tourism, expenditures, 30^{th} percentile	$121,700,000.00
International Tourism, expenditures, 31^{st} percentile	$130,870,000.00
Mobile phones, Average	77.95
Mobile phones, 50^{th} percentile	7.460091
Mobile phones, 82^{nd} percentile	111.812452
International telecom, outgoing traffic, Average	96.32
International telecom, outgoing traffic,50^{th} percentile	65.194105
International telecom, outgoing traffic,80^{th} percentile	117.60494
Military personnel, Average	%1.68
Military personnel, 50^{th} percentile	%1.35825
Military personnel, 60^{th} percentile	%1.635823

According to this model, 84% of the "majority" is similar to less than 10% of the leaders in terms of spending on International tourism, number of mobile phones, and outgoing traffic of international telecom. As well as no member of the "majority" have a level of expenditures on International tourism above 31^{st} percentile.

Step 7:"International tourism, expenditures" variable removed from the data set

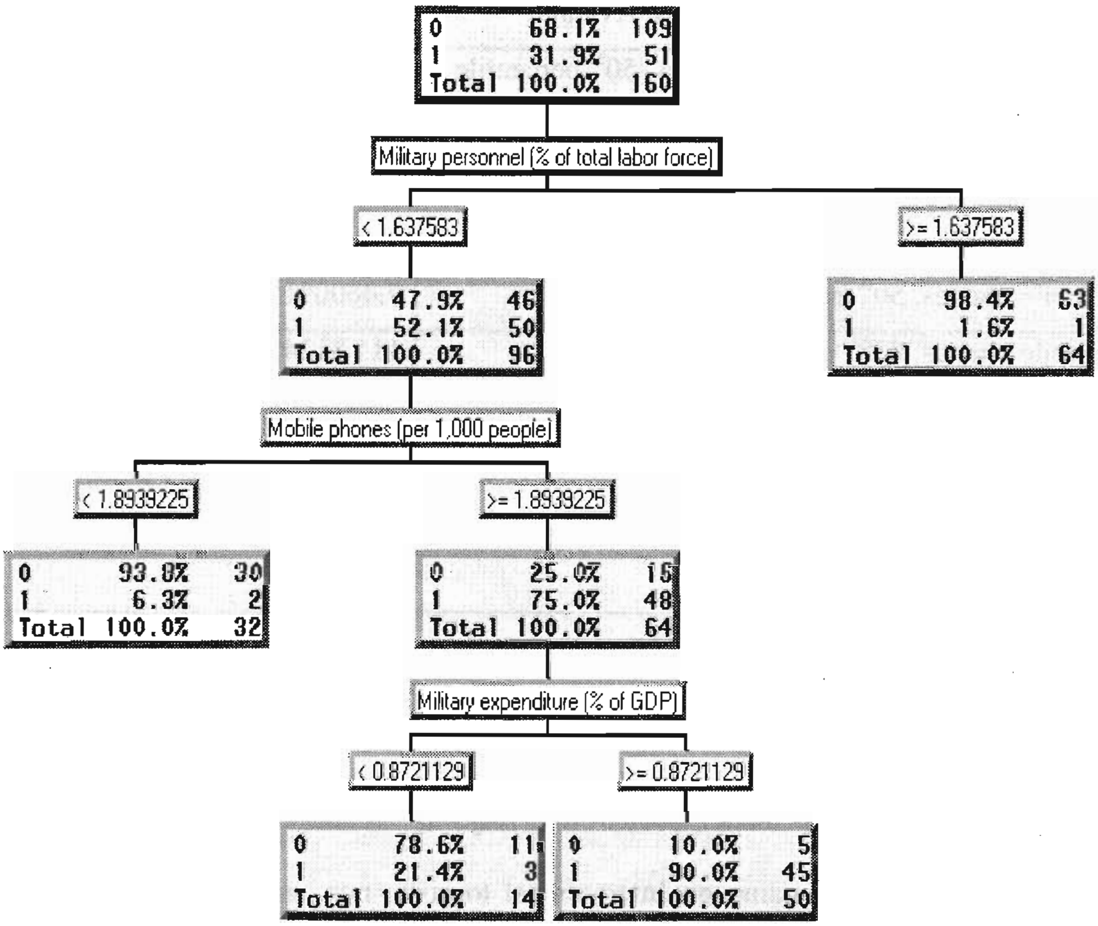


Table of the Split Variables, Average Values of the Split Variables, and Approximate Percentile at which Split was made

Military personnel, Average	%1.68
Military personnel, 50th percentile	%1.35825
Military personnel, 60th percentile	%1.635823
Mobile phones, Average	77.95
Mobile phones, 50th percentile	7.460091
Mobile phones, 87th percentile	180.071403
Military expenditure, Average	%1.82
Military expenditure, 50th percentile	%1.656081

Military expenditure, 14th percentile	%0.880652

This model tells us that 58% of the "majority" have larger percentage of the labor force in the military. Moreover, 94% of the "leaders" are comparable to only less than 15% of the "majority" in terms of similarly small percentage of the labor force in military and the high number of mobile phones per 1000 people.

Step 8:"Military personnel" variable removed from the data set

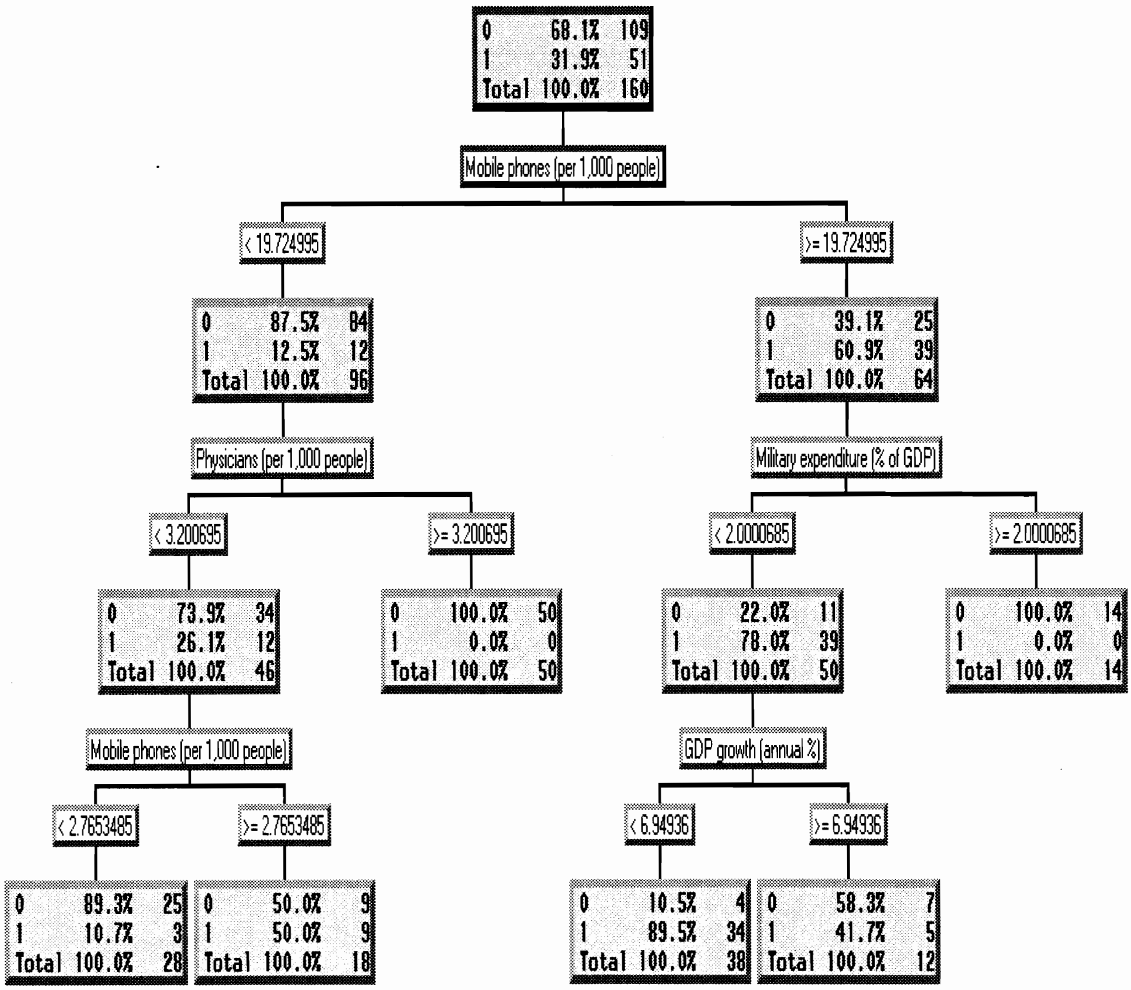


Table of the Split Variables, Average Values of the Split Variables, and Approximate Percentile at which Split was made

Mobile phones, Average	77.95
Mobile phones, 50th percentile	7.460091
Mobile phones, 60th percentile	19.56579
Mobile phones, 17th percentile	0.27131788
Military expenditure, Average	%1.82
Military expenditure, 50th percentile	%1.656081
Military expenditure, 64th percentile	%2.000479
GDP growth annual, Average	%1.77
GDP growth annual, 50th percentile	%3.904678
GDP growth annual, 78th percentile	%6.891061

This diagram demonstrates that 23% of the "majority" represent lower 17 % of the values for the number of the mobile phones, while only 10% of the "majority" is comparable to 76% of the "leaders" in terms of the number of mobile phones and military expenditure.

Step 9:"Mobile phones" variable removed from the data set

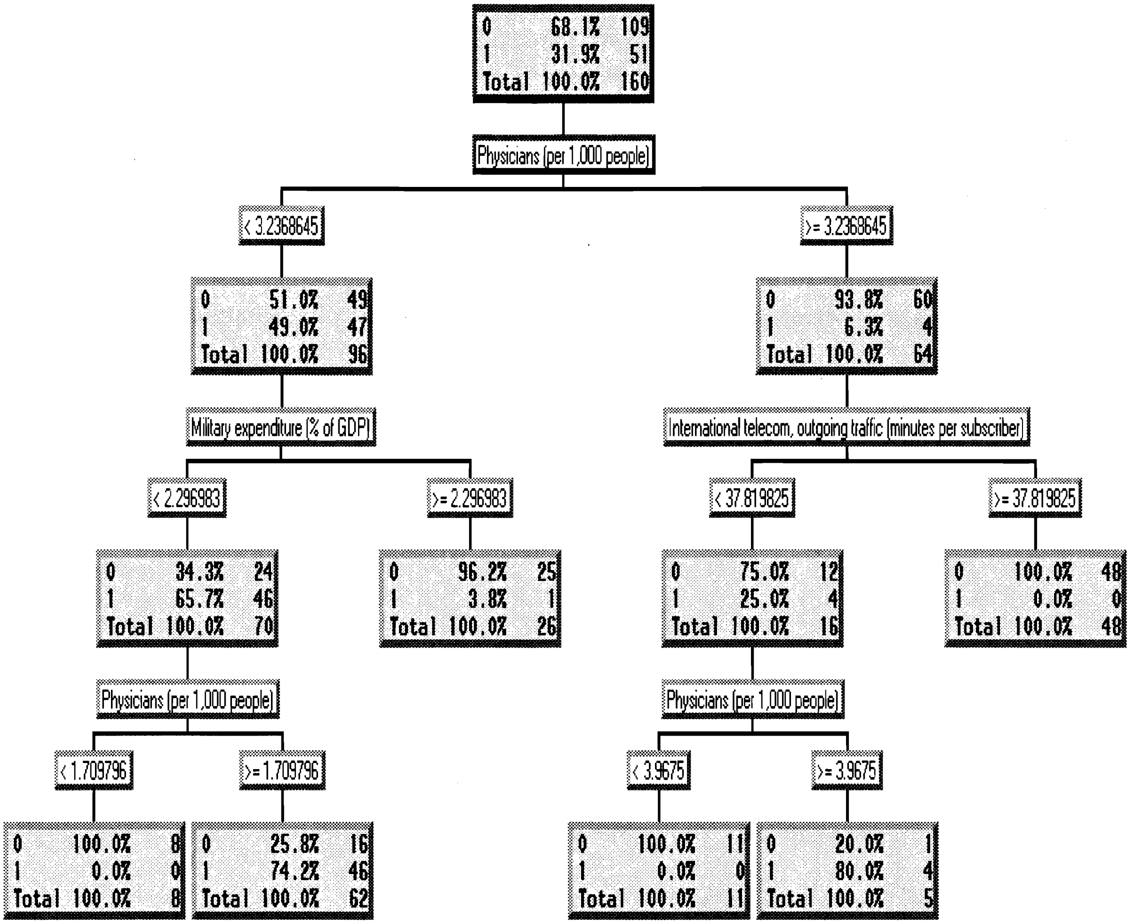


Table of the Split Variables, Average Values of the Split Variables, and Approximate Percentile at which Split was made

Physicians, Average	3.06
Physicians, 50th percentile	3.1523
Physicians, 60th percentile	3.2352286
Physicians, 10th percentile	1.8819187
Physicians, 86th percentile	3.964626
Military expenditure, Average	%1.82
Military expenditure, 50th percentile	%1.656081
Military expenditure, 72.5th percentile	%2.291784

International telecom, outgoing traffic, Average	96.32
International telecom, outgoing traffic,50^{th} percentile	65.194105
International telecom, outgoing traffic,16^{th} percentile	37.7779192

Based on this diagram, we could determine that 55 % of the "majority" has a larger number of physicians than 92 % of the "leaders." Moreover, the 50% of the "majority" that has a comparable to the "leaders" number of physicians spends too much on military.

7.2.4 DT: Brief summary of the analysis

We have performed the DT analysis of the two data sets, one spanning 5-year period from 1998 to 2002, and the second one spanning 10-year period from 1993 to 2002. The goal of the analysis was identifying the characteristics of the two groups of TEs, which we named the "leaders" and "majority."

The results of the conducted DT analysis suggest that on average "leaders" have higher:

- GDP per capita
- level of health expenditure than the most of the "majority"
- level of international telecom traffic
- number of mobile phones
- number of telephone mainlines than "majority"
- level of spending on Health care

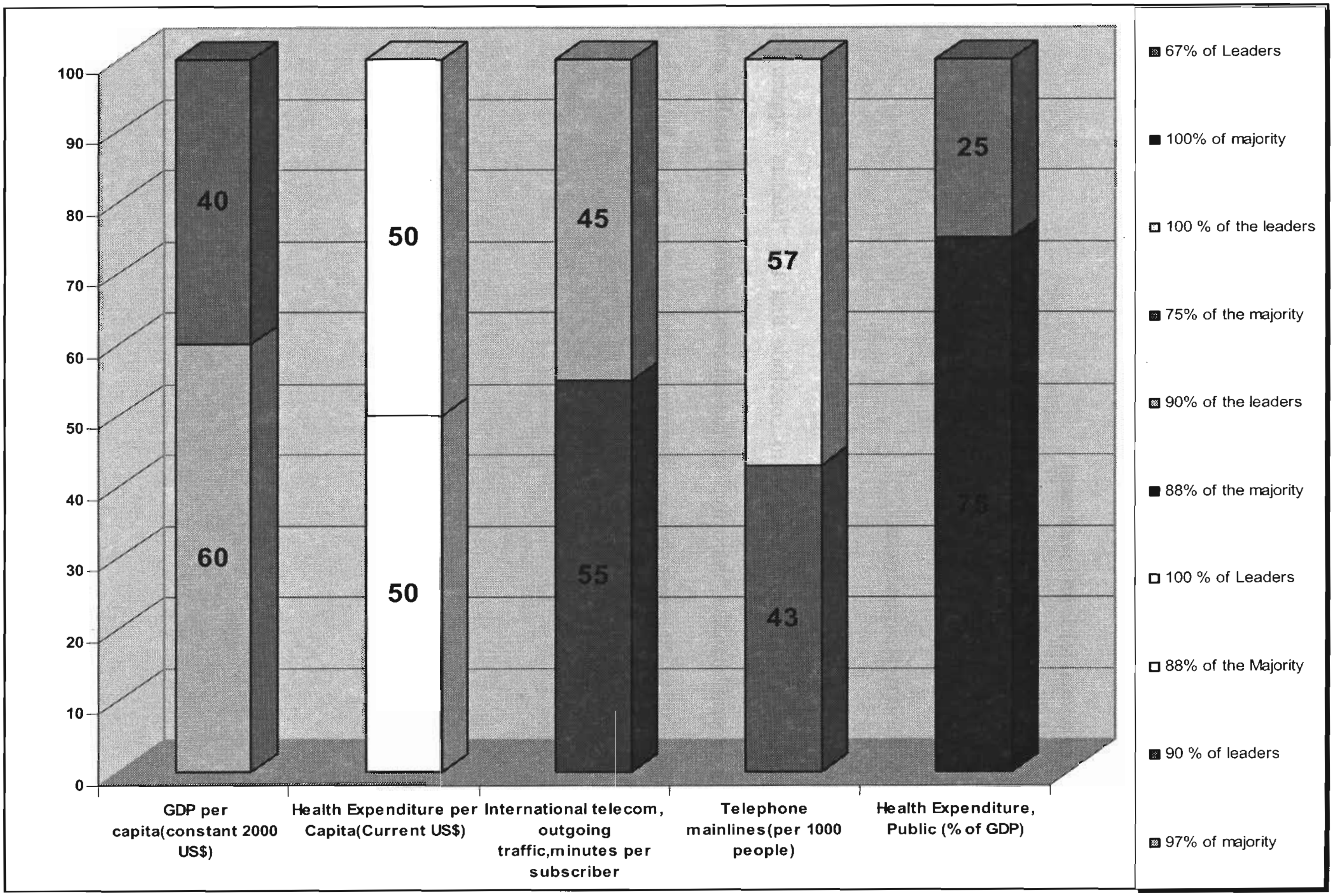
- percentage of the internet users among the population
- number of teachers per pupil in the system of primary education
- percentage of the total labor force employed as R&D technicians
- level of annual telecom investment
- level of expenditures on international tourism.

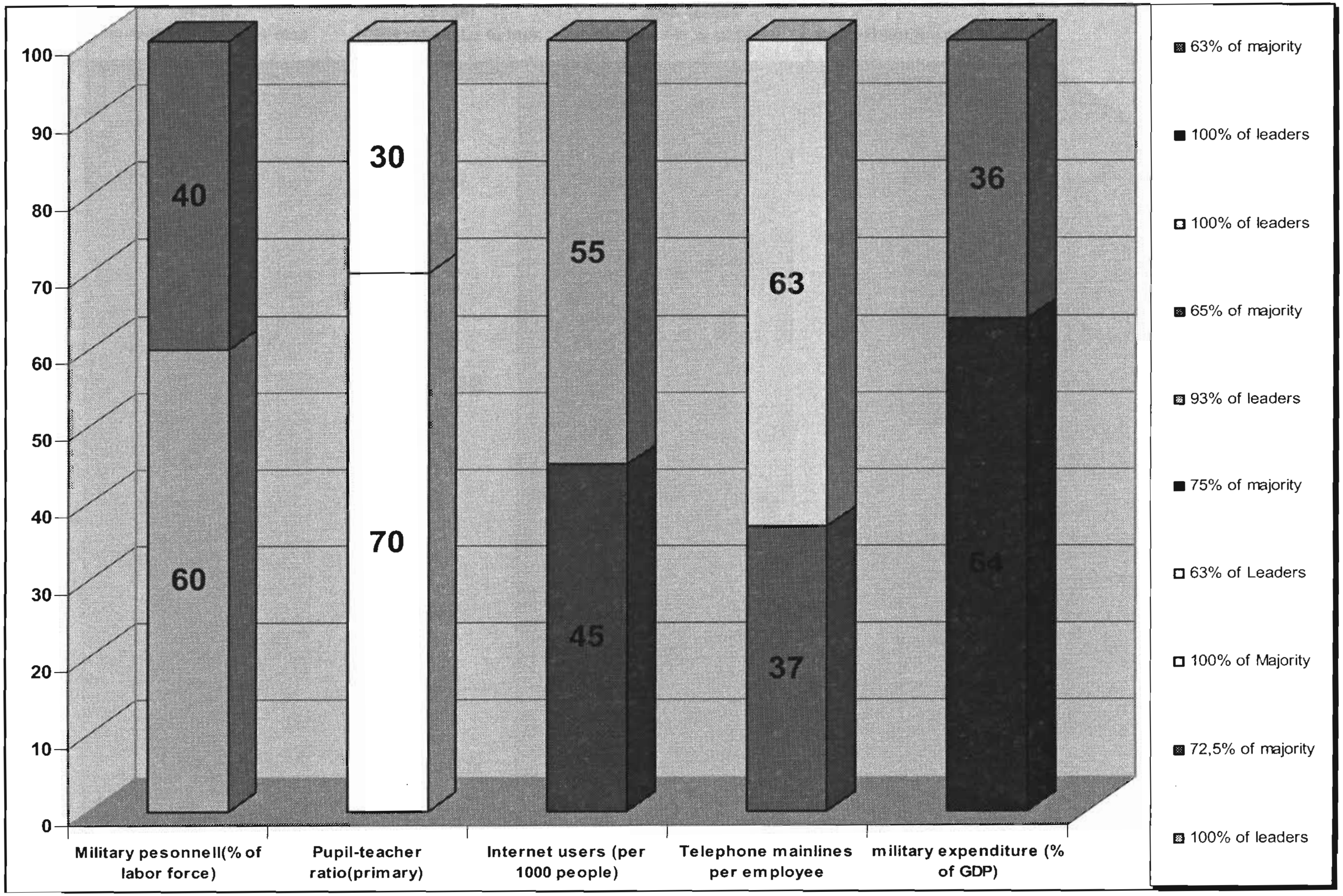
In the same time, “leaders” have lower than the “majority”:

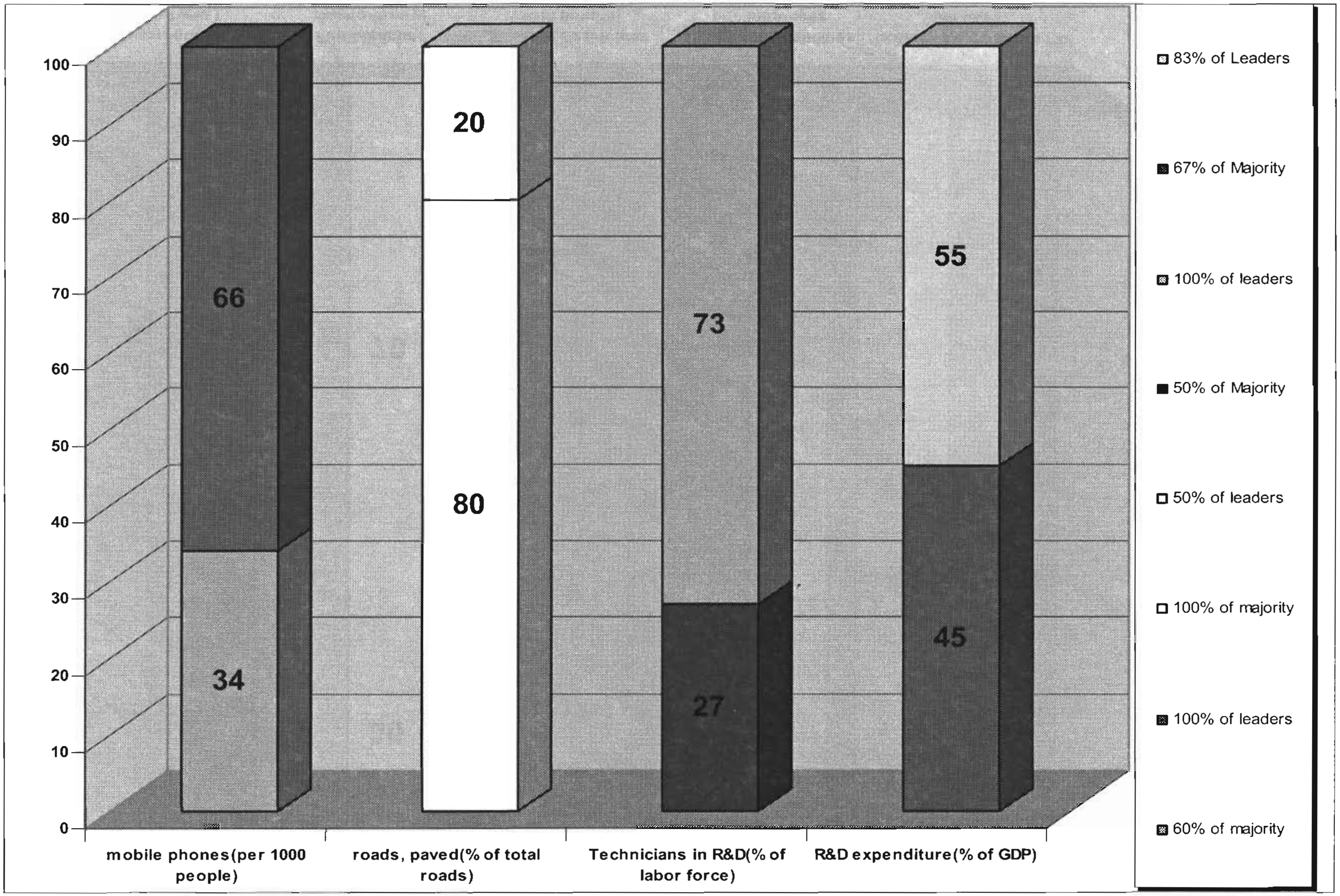
- level of military expenditure
- percentage of the labor force serving in military.

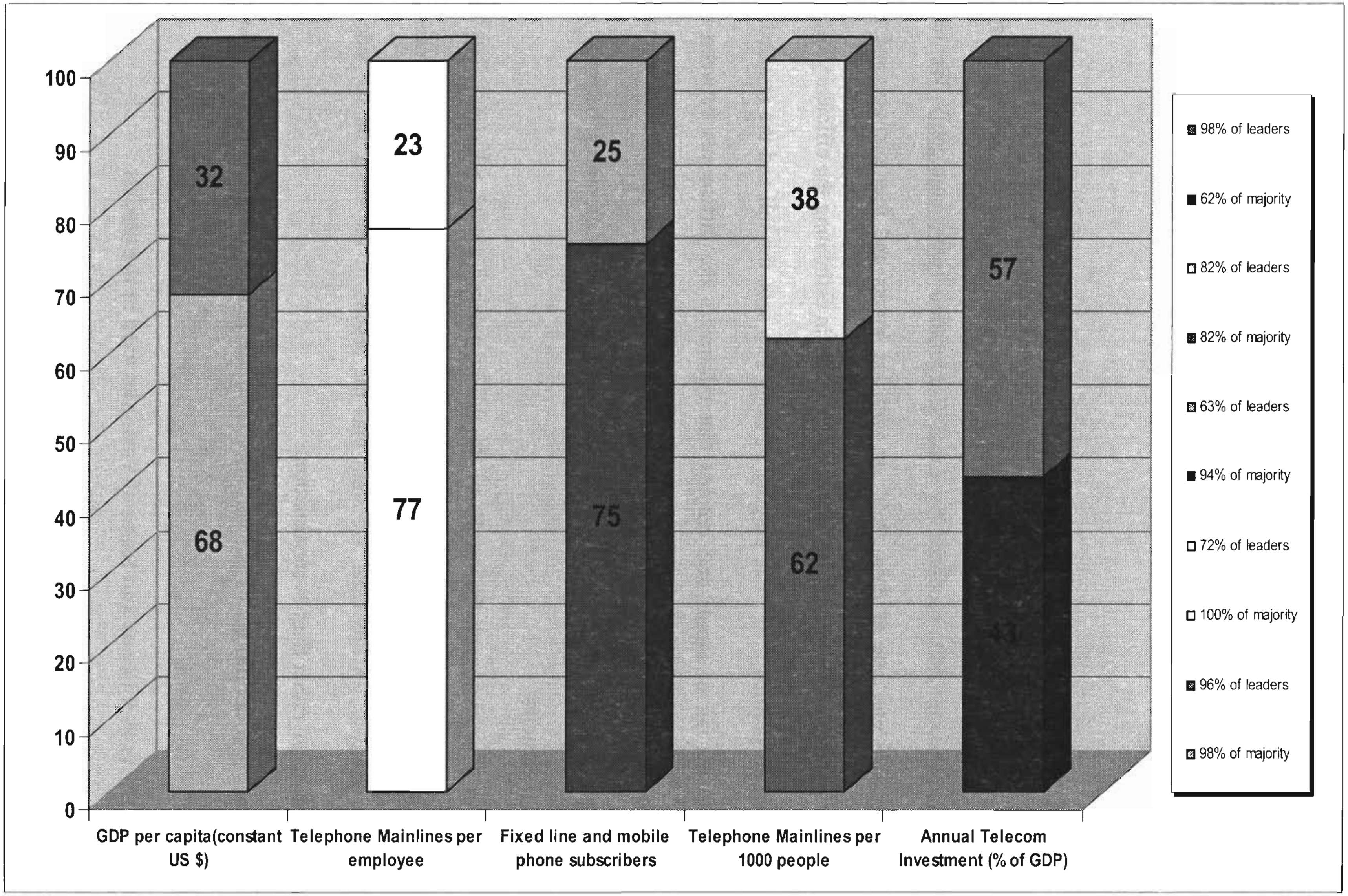
All the compiled information allows us to conclude that the “leaders” appear to be wealthier, in general, than the “majority,” having better infrastructure and smaller armies.

Below we provide the diagrams showing summarized information that we have obtained from analysis of each of the DT model.


100
90
80
70
60
50
40
30
20
10
0
40
60
50
50
45
55
57
43
25
75
GDP per capita(constant 2000 US$)
Health Expenditure per Capita(Current US$)
International telecom, outgoing traffic,minutes per subscriber
Telephone mainlines(per 1000 people)
Health Expenditure, Public (% of GDP)
67% of Leaders
100% of majority
100 % of the leaders
75% of the majority
90% of the leaders
88% of the majority
100 % of Leaders
88% of the Majority
90 % of leaders
97% of majority


100
90
80
70
60
50
40
30
20
10
0
40
60
30
70
55
45
63
37
36
64
Military pesonnell(% of labor force)
Pupil-teacher ratio(primary)
Internet users (per 1000 people)
Telephone mainlines per employee
military expenditure (% of GDP)
63% of majority
100% of leaders
100% of leaders
65% of majority
93% of leaders
75% of majority
63% of Leaders
100% of Majority
72,5% of majority
100% of leaders


100
90
80
70
60
50
40
30
20
10
0
66
34
20
80
73
27
55
45
mobile phones(per 1000 people)
roads, paved(% of total roads)
Technicians in R&D(% of labor force)
R&D expenditure(% of GDP)
83% of Leaders
67% of Majority
100% of leaders
50% of Majority
50% of leaders
100% of majority
100% of leaders
60% of majority


100
90
80
70
60
50
40
30
20
10
0
32
68
23
77
25
75
38
62
57
43
GDP per capita(constant US $)
Telephone Mainlines per employee
Fixed line and mobile phone subscribers
Telephone Mainlines per 1000 people
Annual Telecom Investment (% of GDP)
98% of leaders
62% of majority
82% of leaders
82% of majority
63% of leaders
94% of majority
72% of leaders
100% of majority
96% of leaders
98% of majority

7.3 Data Envelopment Analysis

7.3.1 DEA: Data

In this section, we describe the data that we used to perform DEA, as well as the preliminary results that were obtained. To perform DEA we used the software application "OnFront," version 2.02, produced by Lund Corporation (www.emq.com). We have decided to use two sets of data, 'full' set with six inputs and four outputs, and the 'reduced' one with four input and two output variables in estimating the efficiency of the transitional economies. The purpose of using two sets, the full and the reduced, is to identify the set of inputs and outputs that minimizes the differences between the "leaders" and the "majority." We shall run the analysis and then calculate the average efficiency for each cluster, for every DEA model, and the differences between the average efficiencies.

Then we shall pick the results of the full or reduced model, whichever has the smaller differences between the "majority" and the "leaders," and use the chosen results in further analysis. In this case, we shall bias the results of the study on the side of underestimation, rather than overestimation.

The names and definitions of the variables from the full set are provided in the table below.

Table 4 DEA: Definitions and Roles of Variables used in the models

Variable	Role	Definition
GDP per capita (in current US $)	DEA Input, Full set DEA Input, Reduced	A converted in current US $ total GDP of a

	set	transitional economy divided by the number representing a total population of the country
Full-time telecommunication staff(% of total labor force)	DEA Input ,Full set DEA Input, Reduced set	A number representing the percentage of the full-time telecom workers in the total labor force of the country. Obtained by dividing the number of the telecom employees by the total labor force.
Annual telecom investment per telecom worker	DEA Input, Full set DEA Input, Reduced set	A converted in US $ total annual telecom investment for a given transitional economy divided by the number of full-time telecom employees of a given country.
Annual telecom investment(% of GDP in current US $)	DEA Input, Full set DEA Input, Reduced set	A converted in US $ total annual telecom investment for a given transitional economy divided by the converted in current US $ total GDP of a transitional economy
Annual telecom investment per capita	DEA Input, Full set only	A converted in US $ total annual telecom investment for a given transitional economy divided by the total population of a given country.
Annual telecom investment per worker	DEA Input, Full set only	A converted in US $ total annual telecom investment for a given transitional economy divided by the total labor force of a given country

Total telecom services revenue per telecom worker	DEA Output, Full set DEA Output, Reduced set	A converted in US $ total telecom service revenue for a given transitional economy divided by the number of the full-time telecom employees of a given country.
Total telecom services revenue(% of GDP in current US $)	DEA Output, Full set DEA Output, Reduced set	Represents a percentage, relative to GDP, of the converted in US $ total telecom revenue for a given transitional economy, for a given year. Obtained by dividing a converted in US $ total telecom revenue for a given transitional economy, for a given year, by the country's GDP, in US $.
Total telecom services revenue per worker	DEA Output, Full set only	A converted in US $ total telecom service revenue for a given transitional economy divided by the number of the total labor force of a given country
Total telecom services revenue per capita	DEA Output, Full set only	A converted in US $ total telecom service revenue for a given transitional economy divided by the total population of a given country.

The main goal that we were pursuing in performing DEA was to find out how efficient the 18 transitional economies were in purely converting investment inputs into the revenue outputs. Therefore, we did not include any other types of inputs or outputs such as those related to infrastructure, capabilities, utilization, etc. The reason for using two sets of data to perform DEA was based on our assumption that the full set would give us an opportunity to evaluate the efficiency economically more broadly,(i.e., relative to the whole population, labor force of a country and the telecom industry), while reduced set allowed only for narrow evaluation of efficiency (i.e., relative to the telecom worker).

7.3.2 DEA: Results

In using "OnFront" to obtain the efficiency scores we have chosen to use, first, Farrel Input-Saving Measure of Efficiency as a direct efficiency measure for the three types of models: CRS (constant return to scale), VRS (variable return to scale) and NIRS (non-increasing return to scale). Then we performed the same type of analysis using Farrel Output-Oriented Measure of Technical efficiency.

Results for the two data sets, for each model, are listed below.

Input-Saving Measure of Efficiency

Reduced set (four DEA inputs and two outputs)

Farrel Input-Saving Measure of Efficiency, CRS

Country	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002
Albania	1	1	1	1	1	1	1	0.68	1	1
Armenia	0.79	1	1	1	0.96	0.69	0.71	0.73	0.75	0.80
Azerbaijan	0.40	0.57	1	1	1	0.87	0.51	0.71	0.45	0.46

Belarus	0.42	0.17	0.55	0.60	0.50	0.37	0.46	0.42	0.35	0.44
Bulgaria	0.77	0.85	0.76	0.71	0.56	0.52	0.58	0.81	0.49	0.55
Czech Rep	0.63	0.77	0.59	0.46	0.68	0.83	0.82	0.74	0.55	0.78
Estonia	0.50	0.62	0.94	0.89	0.94	0.96	1	1	1	0.75
Hungary	0.57	0.90	1	1	1	1	0.92	1	1	1
Kazakhstan	0.52	0.22	0.44	0.50	0.41	0.41	0.44	0.52	0.47	0.88
Kyrgyz Rep	1	1	1	0.66	0.56	1	1	1	1	1
Latvia	0.27	0.87	0.94	0.85	1	0.85	0.69	0.60	0.52	0.51
Lithuania	0.57	0.47	0.55	0.56	0.39	0.42	0.49	0.44	0.52	0.93
Moldova	0.22	1	1	1	1	1	1	1	1	1
Romania	0.35	0.43	0.51	0.49	0.44	0.43	0.52	0.52	0.50	0.73
Slovak Rep	0.45	0.60	0.60	0.64	0.50	0.46	0.48	1	1	0.89
Slovenia	1	1	1	1	1	1	1	0.71	0.74	1
Ukraine	0.88	0.48	0.61	1	0.71	0.83	0.90	1	0.67	0.75
Poland	0.77	0.92	0.85	0.71	0.75	1	0.84	0.94	1	0.99

Farrel Input-Saving Measure of Efficiency, VRS

Country	**1993**	**1994**	**1995**	**1996**	**1997**	**1998**	**1999**	**2000**	**2001**	**2002**
Albania	1	1	1	1	1	1	1	0.95	1	1
Armenia	1	1	1	1	1	0.73	0.73	0.75	0.76	0.80
Azerbaijan	0.82	0.92	1	1	1	0.94	1	1	1	1
Belarus	0.68	0.59	0.62	0.69	0.61	0.58	0.62	0.64	0.62	0.77
Bulgaria	1	1	1	0.71	0.74	0.55	0.66	1	0.51	1
Czech Rep	0.65	0.87	0.65	0.69	0.79	0.90	0.86	0.88	0.66	0.81
Estonia	0.53	0.66	1	0.94	1	1	1	1	1	1
Hungary	0.65	1	1	1	1	1	1	1	1	1
Kazakhstan	1	0.48	0.48	0.56	0.52	0.59	0.70	0.76	0.70	1

Kyrgyz Rep	1	1	1	1	1	1	1	1	1	1
Latvia	0.57	1	1	0.85	1	0.93	0.90	0.94	0.80	0.70
Lithuania	0.65	0.56	0.55	0.58	0.57	0.59	0.79	0.91	0.84	1
Moldova	1	1	1	1	1	1	1	1	1	1
Romania	0.56	0.57	0.57	0.60	0.61	0.63	0.69	0.75	0.61	0.81
Slovak Rep	0.50	0.62	0.61	0.66	0.62	0.61	0.63	1	1	0.93
Slovenia	1	1	1	1	1	1	1	0.95	0.76	1
Ukraine	0.88	0.63	0.68	1	1	1	1	1	0.69	1
Poland	0.81	0.99	0.86	0.84	0.92	1	0.96	1	1	1

Farrel Input-Saving Measure of Efficiency, NIRS

Country	**1993**	**1994**	**1995**	**1996**	**1997**	**1998**	**1999**	**2000**	**2001**	**2002**
Albania	1	1	1	1	1	1	1	0.68	1	1
Armenia	0.79	1	1	1	1	0.73	0.73	0.73	0.76	0.80
Azerbaijan	0.40	0.57	1	1	1	0.87	0.51	0.71	0.45	0.46
Belarus	0.42	0.17	0.55	0.60	0.50	0.37	0.46	0.42	0.35	0.44
Bulgaria	1	1	1	0.71	0.74	0.55	0.66	1	0.49	1
Czech Rep	0.65	0.87	0.59	0.46	0.68	0.83	0.82	0.74	0.55	0.78
Estonia	0.50	0.66	1	0.94	1	1	1	1	1	1
Hungary	0.65	1	1	1	1	1	1	1	1	1
Kazakhstan	0.52	0.22	0.44	0.50	0.41	0.41	0.44	0.52	0.47	0.88
Kyrgyz Rep	1	1	1	0.66	0.56	1	1	1	1	1
Latvia	0.27	1	1	0.85	1	0.85	0.69	0.60	0.52	0.51
Lithuania	0.57	0.47	0.55	0.56	0.39	0.42	0.49	0.44	0.52	0.93
Moldova	0.22	1	1	1	1	1	1	1	1	1
Romania	0.35	0.43	0.51	0.49	0.44	0.43	0.52	0.52	0.50	0.73
Slovak Rep	0.45	0.62	0.61	0.64	0.50	0.46	0.48	1	1	0.89

Slovenia	1	1	1	1	1	1	1	0.71	0.74	1
Ukraine	0.88	0.48	0.61	1	1	1	1	1	0.69	1
Poland	0.81	0.99	0.86	0.71	0.75	1	0.84	0.94	1	0.99

Full set (six DEA inputs and four outputs)

Farrel Input-Saving Measure of Efficiency, CRS

Country	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002
Albania	1	1	1	1	1	1	1	0.80	1	1
Armenia	0.79	1	1	1	1	0.89	0.81	0.73	0.86	0.80
Azerbaijan	0.49	0.60	1	1	1	0.88	0.51	0.84	0.49	0.48
Belarus	0.56	0.31	0.68	0.64	0.56	0.47	0.57	0.44	0.39	0.52
Bulgaria	1	1	1	0.88	0.86	0.83	0.92	1	0.72	1
Czech Rep	1	1	0.78	0.60	0.81	0.97	0.98	0.81	0.68	0.90
Estonia	0.81	0.79	1	1	1	1	1	1	1	1
Hungary	1	1	1	1	1	1	1	1	1	1
Kazakhstan	0.80	0.40	0.60	0.67	0.75	0.65	0.51	0.54	0.51	1
Kyrgyz Rep	1	1	1	0.72	0.56	1	1	1	1	1
Latvia	0.41	0.99	0.94	0.94	1	0.85	0.69	0.67	0.55	0.52
Lithuania	0.82	0.82	0.74	0.85	0.45	0.46	0.50	0.46	0.55	0.96
Moldova	0.22	1	1	1	1	1	1	1	1	1
Romania	0.55	0.52	0.62	0.59	0.48	0.46	0.55	0.52	0.63	0.90
Slovak Rep	0.82	0.80	0.79	0.87	0.55	0.51	0.71	1	1	1
Slovenia	1	1	1	1	1	1	1	0.71	0.81	1
Ukraine	1	0.70	0.70	1	0.92	1	1	1	0.85	0.88
Poland	0.93	1	0.90	0.76	0.85	1	0.84	1	1	1

Farrel Input-Saving Measure of Efficiency, VRS

Country	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002
Albania	1	1	1	1	1	1	1	0.95	1	1
Armenia	1	1	1	1	1	0.89	0.83	0.75	0.87	0.80
Azerbaijan	0.82	0.92	1	1	1	0.95	1	1	1	1
Belarus	0.75	0.59	0.77	0.72	0.64	0.60	0.64	0.67	0.66	0.77
Bulgaria	1	1	1	0.93	0.95	0.83	1	1	0.72	1
Czech Rep	1	1	0.80	0.75	0.87	0.97	1	0.95	0.74	0.94
Estonia	0.86	0.83	1	1	1	1	1	1	1	1
Hungary	1	1	1	1	1	1	1	1	1	1
Kazakhstan	1	0.48	0.62	0.70	1	0.65	0.70	0.78	0.75	1
Kyrgyz Rep	1	1	1	1	1	1	1	1	1	1
Latvia	0.57	1	1	0.95	1	0.93	0.90	0.94	0.83	0.76
Lithuania	0.87	0.93	0.82	0.87	0.60	0.63	0.81	0.91	0.86	1
Moldova	1	1	1	1	1	1	1	1	1	1
Romania	0.61	0.63	0.67	0.66	0.63	0.65	0.73	0.75	0.72	0.91
Slovak Rep	0.84	0.82	0.80	0.87	0.68	0.66	0.74	1	1	1
Slovenia	1	1	1	1	1	1	1	0.96	0.87	1
Ukraine	1	0.70	0.75	1	1	1	1	1	0.86	1
Poland	0.93	1	0.91	0.85	0.93	1	0.96	1	1	1

Farrel Input-Saving Measure of Efficiency, NIRS

Country	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002
Albania	1	1	1	1	1	1	1	0.80	1	1
Armenia	0.79	1	1	1	1	0.89	0.81	0.73	0.86	0.80

Azerbaijan	0.49	0.60	1	1	1	0.88	0.51	0.84	0.49	0.48
Belarus	0.56	0.31	0.68	0.64	0.56	0.47	0.57	0.44	0.39	0.52
Bulgaria	1	1	1	0.93	0.95	0.83	1	1	0.72	1
Czech Rep	1	1	0.78	0.60	0.81	0.97	1	0.81	0.68	0.90
Estonia	0.81	0.79	1	1	1	1	1	1	1	1
Hungary	1	1	1	1	1	1	1	1	1	1
Kazakhstan	0.80	0.40	0.60	0.67	1	0.65	0.51	0.54	0.51	1
Kyrgyz R	1	1	1	0.72	0.56	1	1	1	1	1
Latvia	0.41	1	1	0.94	1	0.85	0.69	0.67	0.55	0.52
Lithuania	0.82	0.93	0.74	0.87	0.45	0.46	0.50	0.46	0.55	0.96
Moldova	0.22	1	1	1	1	1	1	1	1	1
Romania	0.55	0.52	0.62	0.59	0.48	0.46	0.55	0.52	0.63	0.90
Slovak Rep	0.82	0.80	0.79	0.87	0.55	0.51	0.74	1	1	1
Slovenia	1	1	1	1	1	1	1	0.71	0.81	1
Ukraine	1	0.70	0.70	1	1	1	1	1	0.85	1
Poland	0.93	1	0.90	0.76	0.85	1	0.84	1	1	1

We have obtained the scale efficiency scores for each data set as well. The results are presented below

Scale efficiency, Full data set

Country	**1993**	**1994**	**1995**	**1996**	**1997**	**1998**	**1999**	**2000**	**2001**	**2002**
Albania	1	1	1	1	1	1	1	0.84	1	1
Armenia	0.79	1	1	1	1	1	0.98	0.98	0.99	1
Azerbaijan	0.60	0.65	1	1	1	0.93	0.51	0.84	0.49	0.48
Belarus	0.75	0.53	0.88	0.89	0.86	0.79	0.89	0.65	0.59	0.68
Bulgaria	1	1	1	0.94	0.90	1	0.92	1	1	1
Czech Rep	1	1	0.98	0.80	0.94	0.99	0.98	0.86	0.91	0.95

Estonia	0.94	0.96	1	1	1	1	1	1	1	1
Hungary	1	1	1	1	1	1	1	1	1	1
Kazakhstan	0.80	0.84	0.96	0.96	0.75	1	0.72	0.69	0.68	1
Kyrgyz Rep	1	1	1	0.72	0.56	1	1	1	1	1
Latvia	0.71	0.99	0.94	1	1	0.91	0.77	0.71	0.66	0.68
Lithuania	0.95	0.87	0.90	0.97	0.74	0.74	0.62	0.51	0.64	0.96
Moldova	0.22	1	1	1	1	1	1	1	1	1
Romania	0.91	0.82	0.92	0.90	0.75	0.71	0.75	0.69	0.88	1
Slovak Rep	0.98	0.97	0.98	0.99	0.81	0.77	0.96	1	1	1
Slovenia	1	1	1	1	1	1	1	0.74	0.93	1
Ukraine	1	1	0.93	1	0.92	1	1	1	0.99	0.88
Poland	1	1	0.99	0.90	0.91	1	0.88	1	1	1

Scale efficiency, Reduced data set

Country	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002
Albania	1	1	1	1	1	1	1	0.71	1	1
Armenia	0.79	1	1	1	0.96	0.95	0.98	0.98	0.98	1
Azerbaijan	0.49	0.62	1	1	1	0.93	0.51	0.71	0.45	0.46
Belarus	0.62	0.30	0.88	0.87	0.81	0.63	0.75	0.67	0.56	0.58
Bulgaria	0.77	0.85	0.76	1	0.75	0.94	0.88	0.81	0.96	0.55
Czech Rep	0.96	0.88	0.92	0.67	0.86	0.93	0.96	0.84	0.84	0.96
Estonia	0.94	0.94	0.94	0.94	0.94	0.96	1	1	1	0.75
Hungary	0.87	0.90	1	1	1	1	0.92	1	1	1
Kazakhstan	0.52	0.47	0.91	0.89	0.78	0.71	0.62	0.68	0.67	0.89
Kyrgyz Rep	1	1	1	0.66	0.56	1	1	1	1	1
Latvia	0.48	0.87	0.94	1	1	0.91	0.78	0.64	0.65	0.72
Lithuania	0.87	0.84	1	0.96	0.69	0.70	0.62	0.49	0.61	0.93

Moldova	0.22	1	1	1	1	1	1	1	1	1
Romania	0.62	0.75	0.89	0.82	0.72	0.67	0.75	0.69	0.82	0.90
Slovak Rep	0.91	0.96	0.99	0.97	0.80	0.75	0.77	1	1	0.96
Slovenia	1	1	1	1	1	1	1	0.75	0.97	1
Ukraine	1	0.76	0.89	1	0.71	0.83	0.90	1	0.98	0.75
Poland	0.94	0.93	0.99	0.84	0.81	1	0.87	0.94	1	0.99

Output-Oriented Measure of Efficiency

Reduced set (four DEA inputs and two outputs)

Farrel Output-Oriented Measure of Efficiency, CRS

Country	**1993**	**1994**	**1995**	**1996**	**1997**	**1998**	**1999**	**2000**	**2001**	**2002**
Albania	1	1	1	1	1	1	1	1.48	1	1
Armenia	1.27	1	1	1	1.04	1.44	1.40	1.37	1.34	1.25
Azerbaijan	2.49	1.77	1	1	1	1.15	1.98	1.42	2.22	2.18
Belarus	2.38	5.72	1.83	1.66	2.01	2.72	2.16	2.36	2.88	2.27
Bulgaria	1.29	1.17	1.32	1.40	1.80	1.93	1.72	1.23	2.04	1.82
Czech Rep	1.59	1.30	1.68	2.17	1.47	1.20	1.22	1.35	1.81	1.28
Estonia	1.99	1.61	1.07	1.13	1.06	1.04	1	1	1	1.33
Hungary	1.76	1.11	1	1	1	1	1.09	1	1	1
Kazakhstan	1.93	4.47	2.29	2.02	2.47	2.42	2.29	1.94	2.13	1.13
Kyrgyz Rep	1	1	1	1.51	1.77	1	1	1	1	1
Latvia	3.68	1.15	1.06	1.18	1	1.18	1.44	1.67	1.92	1.97
Lithuania	1.77	2.13	1.82	1.79	2.55	2.41	2.04	2.25	1.94	1.07
Moldova	4.61	1	1	1	1	1	1	1	1	1

Romania	2.87	2.34	1.97	2.03	2.28	2.35	1.93	1.93	2	1.36
Slovak Rep	2.21	1.68	1.65	1.56	2.01	2.19	2.08	1	1	1.12
Slovenia	1	1	1	1	1	1	1	1.41	1.35	1
Ukraine	1.14	2.1	1.65	1	1.41	1.20	1.11	1	1.49	1.33
Poland	1.30	1.09	1.17	1.41	1.33	1	1.20	1.06	1	1.01

Farrel Output-Oriented Measure of Efficiency, VRS

Country	**1993**	**1994**	**1995**	**1996**	**1997**	**1998**	**1999**	**2000**	**2001**	**2002**
Albania	1	1	1	1	1	1	1	1.17	1	1
Armenia	1	1	1	1	1	1.09	1.17	1.32	1.17	1.1
Azerbaijan	2.36	1.69	1	1	1	1.14	1	1	1	1
Belarus	2.18	3.64	1.52	1.58	1.57	1.96	1.72	2.07	2.43	2.18
Bulgaria	1	1	1	1.17	1.06	1.17	1.11	1	1.51	1
Czech Rep	1.14	1.11	1.56	1.90	1.38	1.18	1.12	1.30	1.69	1.28
Estonia	1.77	1.35	1	1.04	1	1	1	1	1	1
Hungary	1.12	1	1	1	1	1	1	1	1	1
Kazakhstan	1.86	2.66	1.48	1.63	1.32	1.55	1.93	1.93	2	1.01
Kyrgyz Rep	1	1	1	1	1	1	1	1	1	1
Latvia	2.47	1	1	1.11	1	1.12	1.41	1.23	1.91	1.94
Lithuania	1.64	1.40	1.48	1.34	2.12	2.10	1.96	1.56	1.89	1
Moldova	1	1	1	1	1	1	1	1	1	1
Romania	1.90	2.04	1.56	1.78	1.83	2.01	1.78	1.92	1.60	1.35
Slovak Rep	1.35	1.41	1.43	1.35	1.74	1.92	1.67	1	1	1.1
Slovenia	1	1	1	1	1	1	1	1.19	1.27	1
Ukraine	1.08	1.48	1.35	1	1	1	1	1	1.26	1
Poland	1.06	1.01	1.13	1.40	1.20	1	1.20	1	1	1

Farrel Output-Oriented Measure of Efficiency, NIRS

Country	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002
Albania	1	1	1	1	1	1	1	1.48	1	1
Armenia	1.27	1	1	1	1	1.09	1.17	1.32	1.17	1.10
Azerbaijan	2.36	1.69	1	1	1	1.14	1.98	1.42	2.22	2.18
Belarus	2.18	3.64	1.52	1.58	1.57	1.96	1.72	2.07	2.43	2.18
Bulgaria	1	1	1	1.17	1.06	1.17	1.11	1	1.51	1
Czech Rep	1.14	1.11	1.56	1.90	1.38	1.18	1.12	1.35	1.69	1.28
Estonia	1.77	1.35	1	1.04	1	1	1	1	1	1
Hungary	1.12	1	1	1	1	1	1	1	1	1
Kazakhstan	1.93	2.66	1.48	1.63	1.32	1.55	1.93	1.94	2	1.13
Kyrgyz Rep	1	1	1	1.51	1.77	1	1	1	1	1
Latvia	2.47	1	1	1.11	1	1.18	1.44	1.67	1.91	1.97
Lithuania	1.64	1.40	1.48	1.34	2.12	2.10	1.96	2.25	1.94	1.07
Moldova	4.61	1	1	1	1	1	1	1	1	1
Romania	1.9	2.04	1.56	1.78	1.83	2.01	1.78	1.93	1.60	1.35
Slovak Rep	1.35	1.41	1.43	1.35	1.74	1.92	1.67	1	1	1.12
Slovenia	1	1	1	1	1	1	1	1.41	1.27	1
Ukraine	1.08	1.48	1.35	1	1	1	1	1	1.26	1
Poland	1.06	1.01	1.13	1.40	1.33	1	1.20	1.06	1	1.01

<u>Full set (six DEA inputs and four outputs)</u>

Farrel Output-Oriented Measure of Efficiency, CRS

Country	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002
Albania	1	1	1	1	1	1	1	1.25	1	1
Armenia	1.27	1	1	1	1	1.13	1.24	1.37	1.16	1.25

Azerbaijan	2.02	1.68	1	1	1	1.14	1.98	1.19	2.04	2.07
Belarus	1.77	3.20	1.47	1.56	1.80	2.14	1.75	2.29	2.54	1.90
Bulgaria	1	1	1	1.13	1.17	1.21	1.08	1	1.39	1
Czech Rep	1	1	1.28	1.67	1.23	1.03	1.02	1.23	1.48	1.12
Estonia	1.24	1.26	1	1	1	1	1	1	1	1
Hungary	1	1	1	1	1	1	1	1	1	1
Kazakhstan	1.25	2.50	1.67	1.49	1.33	1.53	1.97	1.86	1.96	1
Kyrgyz Rep	1	1	1	1.4	1.77	1	1	1	1	1
Latvia	2.47	1.01	1.06	1.06	1	1.18	1.44	1.50	1.83	1.94
Lithuania	1.22	1.22	1.36	1.18	2.24	2.17	2.01	2.15	1.83	1.04
Moldova	4.61	1	1	1	1	1	1	1	1	1
Romania	1.81	1.92	1.62	1.69	2.10	2.16	1.81	1.93	1.59	1.11
Slovak Rep	1.22	1.26	1.27	1.15	1.81	1.95	1.42	1	1	1
Slovenia	1	1	1	1	1	1	1	1.41	1.23	1
Ukraine	1	1.44	1.43	1	1.09	1	1	1	1.17	1.14
Poland	1.08	1	1.11	1.32	1.18	1	1.19	1	1	1

Farrel Output-Oriented Measure of Efficiency, VRS

Country	**1993**	**1994**	**1995**	**1996**	**1997**	**1998**	**1999**	**2000**	**2001**	**2002**
Albania	1	1	1	1	1	1	1	1	1.13	1
Armenia	1	1	1	1	1	1.09	1.17	1.1	1.32	1.16
Azerbaijan	1.89	1.60	1	1	1	1.14	1	1	1	1
Belarus	1.72	2.36	1.46	1.55	1.57	1.93	1.70	1.86	2.02	2.33
Bulgaria	1	1	1	1.04	1.02	1.11	1	1	1	1.37
Czech Rep	1	1	1.24	1.43	1.14	1.03	1	1.1	1.15	1.44
Estonia	1.22	1.26	1	1	1	1	1	1	1	1
Hungary	1	1	1	1	1	1	1	1	1	1

Kazakhstan	1.16	1.96	1.48	1.34	1	1.23	1.93	1	1.73	1.93
Kyrgyz Rep	1	1	1	1	1	1	1	1	1	1
Latvia	2.47	1	1	1.06	1	1.12	1.41	1.93	1.23	1.82
Lithuania	1.21	1.04	1.31	1.13	2.10	2.10	1.96	1	1.56	1.83
Moldova	1	1	1	1	1	1	1	1	1	1
Romania	1.80	1.89	1.54	1.60	1.80	1.96	1.78	1.10	1.92	1.58
Slovak Rep	1.18	1.26	1.25	1.12	1.74	1.92	1.25	1	1	1
Slovenia	1	1	1	1	1	1	1	1	1.17	1.22
Ukraine	1	1.17	1.35	1	1	1	1	1	1	1.17
Poland	1.06	1	1.10	1.30	1.17	1	1.18	1	1	1

Farrel Output-Oriented Measure of Efficiency, NIRS

Country	**1993**	**1994**	**1995**	**1996**	**1997**	**1998**	**1999**	**2000**	**2001**	**2002**
Albania	1	1	1	1	1	1	1	1.25	1	1
Armenia	1.27	1	1	1	1	1.09	1.17	1.32	1.16	1.10
Azerbaijan	1.89	1.60	1	1	1	1.14	1.98	1.19	2.04	2.07
Belarus	1.72	2.36	1.46	1.55	1.57	1.93	1.70	2.02	2.33	1.86
Bulgaria	1	1	1	1.04	1.02	1.11	1	1	1.37	1
Czech Rep	1	1	1.24	1.43	1.14	1.03	1	1.23	1.44	1.12
Estonia	1.24	1.26	1	1	1	1	1	1	1	1
Hungary	1	1	1	1	1	1	1	1	1	1
Kazakhstan	1.25	1.96	1.48	1.34	1	1.23	1.93	1.73	1.93	1
Kyrgyz Rep	1	1	1	1.40	1.77	1	1	1	1	1
Latvia	2.47	1	1	1.06	1	1.18	1.44	1.50	1.82	1.94
Lithuania	1.21	1.04	1.36	1.13	2.10	2.10	1.96	2.15	1.83	1.04
Moldova	4.61	1	1	1	1	1	1	1	1	1
Romania	1.81	1.92	1.54	1.60	1.80	1.96	1.78	1.93	1.58	1.10

Slovak Rep	1.18	1.26	1.25	1.12	1.74	1.92	1.25	1	1	1
Slovenia	1	1	1	1	1	1	1	1.41	1.23	1
Ukraine	1	1.17	1.35	1	1	1	1	1	1.17	1
Poland	1.06	1	1.10	1.30	1.17	1	1.19	1	1	1

We have obtained the scale efficiency scores for each data set as well. The results are presented below

Scale efficiency, Full data set

Country	**1993**	**1994**	**1995**	**1996**	**1997**	**1998**	**1999**	**2000**	**2001**	**2002**
Albania	1	1	1	1	1	1	1	1.11	1	1
Armenia	1.27	1	1	1	1	1.04	1.06	1.04	1	1.13
Azerbaijan	1.07	1.05	1	1	1	1	1.98	1.19	2.04	2.07
Belarus	1.03	1.35	1	1.01	1.14	1.11	1.03	1.13	1.09	1.02
Bulgaria	1	1	1	1.09	1.15	1.09	1.08	1	1.01	1
Czech Rep	1	1	1.04	1.17	1.09	1.01	1.02	1.07	1.02	1.01
Estonia	1.01	1	1	1	1	1	1	1	1	1
Hungary	1	1	1	1	1	1	1	1	1	1
Kazakhstan	1.07	1.28	1.13	1.11	1.33	1.25	1.02	1.07	1.02	1
Kyrgyz Rep	1	1	1	1.40	1.77	1	1	1	1	1
Latvia	1	1.01	1.06	1	1	1.05	1.02	1.22	1.01	1
Lithuania	1.01	1.18	1.03	1.05	1.07	1.03	1.03	1.38	1	1.04
Moldova	4.61	1	1	1	1	1	1	1	1	1
Romania	1.01	1.02	1.05	1.05	1.17	1.10	1.02	1	1.01	1
Slovak Rep	1.04	1	1.02	1.03	1.04	1.02	1.14	1	1	1
Slovenia	1	1	1	1	1	1	1	1.20	1.01	1
Ukraine	1	1.22	1.06	1	1.09	1	1	1	1	1.14
Poland	1.02	1	1.01	1.01	1	1	1	1	1	1

Scale efficiency, Reduced data set

Country	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002
Albania	1	1	1	1	1	1	1	1.26	1	1
Armenia	1.27	1	1	1	1.04	1.32	1.20	1.04	1.14	1.13
Azerbaijan	1.05	1.04	1	1	1	1.01	1.98	1.42	2.22	2.18
Belarus	1.09	1.57	1.21	1.05	1.28	1.39	1.25	1.14	1.19	1.04
Bulgaria	1.29	1.17	1.32	1.20	1.69	1.65	1.55	1.23	1.35	1.82
Czech Rep	1.39	1.18	1.08	1.14	1.07	1.02	1.08	1.04	1.07	1
Estonia	1.13	1.19	1.07	1.08	1.06	1.04	1	1	1	1.33
Hungary	1.57	1.11	1	1	1	1	1.09	1	1	1
Kazakhstan	1.04	1.68	1.55	1.23	1.86	1.56	1.19	1	1.06	1.12
Kyrgyz Rep	1	1	1	1.51	1.77	1	1	1	1	1
Latvia	1.49	1.15	1.06	1.06	1	1.05	1.02	1.36	1	1.02
Lithuania	1.07	1.53	1.23	1.34	1.20	1.15	1.04	1.44	1.03	1.07
Moldova	4.61	1	1	1	1	1	1	1	1	1
Romania	1.51	1.15	1.26	1.14	1.24	1.17	1.09	1	1.25	1.01
Slovak Rep	1.64	1.19	1.16	1.16	1.16	1.14	1.24	1	1	1.02
Slovenia	1	1	1	1	1	1	1	1.18	1.06	1
Ukraine	1.06	1.41	1.22	1	1.41	1.20	1.11	1	1.19	1.33
Poland	1.23	1.08	1.04	1.01	1.11	1	1	1.06	1	1.01

7.3.3 DEA: Conclusion and the analysis of the results

The results of DEA demonstrate that a number of transitional economies per given year have obtained a rating of being hundred percent relatively efficient. It does not, however, mean that all of the countries that were deemed relatively efficient are in fact efficient. This, sometimes overly generous assignment of efficiency scores, is a common characteristic of the majority of DEA models (Lins et al. 2003). Another common characteristic of DEA models is that they tend to evaluate as efficient those DMUs that have the smallest input values, or, the DMUs with the largest outputs (Ali 1994). There are multiple ways by means of which the number of the efficient DMUs in the outcome of the analysis could be reduced (Adler et al. 2002), and we might consider using them, if necessary, later in our analysis.

Meanwhile, however, we are going to use the efficiency scores as a target variable to perform Decision Tree analysis in our attempt to find out the characteristics of the efficient DMUs.

7.3.4 Comparison of the results of DEA based on the segmentation provided by Clustering

Let us recall that as a result of Cluster analysis we were able to obtain two clusters. One of the clusters, it is conjectured, contains the set of TEs that are able to take a greater advantage from the investments in ICT. The data points comprising the other cluster represent the countries (or the countries in a certain time period) which receive lesser benefits from the investments in ICT. We named the first cluster "leaders" and the second cluster "majority."

Table 5 DEA: Choosing the "leaders"

Data Set	**TEs – members of the "leaders" cluster**
First 5-year data set (1993 – 1998), 2 Clusters	Czech Republic (1994-1997), Estonia (1996-1997), Hungary (1993-1997), Latvia (1994-5, 1997), Poland (1995-1997), Slovak Republic (1996-1997), Slovenia (1993-1997).
Second 5-year data set (1998 – 2002), 2 Clusters	Czech Republic(1998-2002), Estonia (1998-2002), Hungary (1998-2002), Latvia (1998-2002), Poland (1998-2002) , Slovak Republic (1998,2000-2002), Slovenia (1998-2002), Ukraine (2002), Romania (2000, 2002), Moldova (2002), Lithuania (1999, 2000, 2002), Bulgaria (2002)
10-year data set(1993-2002), 2 clusters	Czech republic (1993-2002), Estonia (1994-2002), Hungary (1993-2002), Bulgaria (2002), Latvia (1994, 1995, 1997-2002), Lithuania (1999-2002), Slovenia (1993-2002),

	Poland (1993-2002), Slovak Republic (1995-8, 2000-2).

The data points representing "leaders" are provided in the table above. We have decided to compare the "leaders" and "majority" in terms of the criteria that were obtained through the DEA. These criteria are: scores of efficiency for CRS, VRS, and NIRS models, as well as the scores for Scale efficiency. Moreover, we have also compared the two clusters in terms of the values of Malmquist index and components of the Malmquist index, Technical change and Efficiency change. The obtained results are displayed in the tables below.

Input-Oriented Measure of Efficiency

First 5-years (1993-1997) data set

Table 6 DEA: Comparison of the clusters based on the input-oriented model, Reduced set (four DEA inputs and two outputs)

Criterion for comparison	"Leaders" cluster	"Majority" cluster	Difference	Difference %
Average efficiency score, CRS	**0.84**	**0.70**	**0.14**	**19.60%**
Average efficiency score, VRS	**0.90**	**0.81**	**0.09**	**10.78%**
Average efficiency score, NIRS	**0.86**	**0.72**	**0.14**	**19.13%**

Average efficiency score, SE	**0.93**	**0.85**	**0.07**	**8.40%**
Malmquist Index	**1.18**	**1.22**	**-0.04**	**-3.31%**
Malmquist Index, TC	**1.08**	**1.13**	**-0.05**	**-4.71%**
Malmquist Index, EC	**1.14**	**1.13**	**0.01**	**0.72%**

Table 7 DEA: Comparison of the clusters based on the input-oriented model, Full set (six DEA inputs and four outputs)

Criterion for comparison	"Leaders" cluster	"Majority" cluster	Difference	Difference %
Average efficiency score, CRS	**0.92**	**0.81**	**0.11**	**13.77%**
Average efficiency score, VRS	**0.94**	**0.88**	**0.06**	**6.93%**
Average efficiency score, NIRS	**0.92**	**0.82**	**0.10**	**12.69%**
Average efficiency score, SE	**0.97**	**0.91**	**0.06**	**6.85%**
Malmquist Index	**1.18**	**1.21**	**-0.03**	**-2.49%**
Malmquist Index, TC	**1.14**	**1.13**	**0.01**	**0.89%**
Malmquist Index, EC	**1.04**	**1.08**	**-0.04**	**-3.70%**

Second 5-years (1998-2002) data set

Table 8 DEA: Comparison of the clusters based on the input-oriented model, Reduced set (four DEA inputs and two outputs)

Criterion for comparison	"Leaders" cluster	"Majority" cluster	Difference	Difference %
Average efficiency score, CRS	**0.82**	**0.71**	**0.11**	**15.45%**
Average efficiency score, VRS	**0.93**	**0.85**	**0.08**	**9.45%**
Average efficiency score, NIRS	**0.85**	**0.72**	**0.12**	**16.93%**
Average efficiency score, SE	**0.88**	**0.83**	**0.05**	**5.78%**
Malmquist Index	**1.159**	**1.159**	**0.000**	**-0.03%**
Malmquist Index, TC	**1.106**	**1.103**	**0.003**	**0.29%**
Malmquist Index, EC	**1.058**	**1.056**	**0.001**	**0.14%**

Table 9 DEA: Comparison of the clusters based on the input-oriented model, Full set (six DEA inputs and four outputs)

Criterion for comparison	"Leaders" cluster	"Majority" cluster	Difference	Difference %
Average efficiency score, CRS	**0.87**	**0.78**	**0.09**	**11.79%**
Average efficiency score, VRS	**0.95**	**0.87**	**0.08**	**9.01%**

Average efficiency score, NIRS	**0.88**	**0.78**	**0.09**	**11.93%**
Average efficiency score, SE	**0.91**	**0.88**	**0.03**	**3.90%**
Malmquist Index	**1.155**	**1.154**	**0.001**	**0.09%**
Malmquist Index, TC	**1.111**	**1.102**	**0.009**	**0.80%**
Malmquist Index, EC	**1.042**	**1.052**	**-0.010**	**-0.93%**

Complete 10-years (1993-2002) data set

Table 10 DEA: Comparison of the clusters based on the input-oriented model, Reduced set (four DEA inputs and two outputs)

Criterion for comparison	"Leaders" cluster	"Majority" cluster	Difference	Difference %
Average efficiency score, CRS	**0.82**	**0.70**	**0.12**	**16.61%**
Average efficiency score, VRS	**0.90**	**0.84**	**0.07**	**7.90%**
Average efficiency score, NIRS	**0.84**	**0.72**	**0.12**	**16.58%**
Average efficiency score, SE	**0.90**	**0.84**	**0.06**	**7.19%**
Malmquist Index	**1.17**	**1.20**	**-0.027**	**2.32%**
Malmquist Index, TC	**1.10**	**1.12**	**-0.025**	**2.27%**

Malmquist Index, EC	**1.09**	**1.09**	**0.004**	**-0.32%**

Table 11 DEA: Comparison of the clusters based on the input-oriented model, Full set (six DEA inputs and four outputs)

Criterion for comparison	"Leaders" cluster	"Majority" cluster	Difference	Difference %
Average efficiency score, CRS	**0.89**	**0.79**	**0.10**	**12.54%**
Average efficiency score, VRS	**0.95**	**0.88**	**0.07**	**7.48%**
Average efficiency score, NIRS	**0.89**	**0.80**	**0.09**	**11.63%**
Average efficiency score, SE	**0.94**	**0.89**	**0.04**	**4.96%**
Malmquist Index	**1.18**	**1.18**	**0.00**	**-0.13%**
Malmquist Index, TC	**1.13**	**1.11**	**0.02**	**1.52%**
Malmquist Index, EC	**1.04**	**1.07**	**-0.03**	**-2.11%**

Output-Oriented Measure of Efficiency

First 5-years (1993-1997) data set

Table 12 DEA: Comparison of the clusters based on the output-oriented model, Reduced set (four DEA inputs and two outputs)

Criterion for comparison	"Leaders" cluster	"Majority" cluster	Difference	Difference %
Average efficiency score, CRS	**1.27**	**1.73**	**-0.46**	**-26.73%**
Average efficiency score, VRS	**1.16**	**1.35**	**-0.19**	**-14.11%**
Average efficiency score, NIRS	**1.17**	**1.42**	**-0.25**	**-17.68%**
Average efficiency score, SE	**1.08**	**1.26**	**-0.18**	**-13.99%**
Malmquist Index	**1.18**	**1.21**	**-0.03**	**-2.39%**
Malmquist Index, TC	**1.08**	**1.11**	**-0.03**	**-2.57%**
Malmquist Index, EC	**1.14**	**1.15**	**0.00**	**-0.41%**

Table 13 DEA: Comparison of the clusters based on the output-oriented model, Full set (six DEA inputs and four outputs)

Criterion for comparison	"Leaders" cluster	"Majority" cluster	Difference	Difference %
Average efficiency score, CRS	**1.12**	**1.36**	**-0.24**	**-17.62%**

Average efficiency score, VRS	**1.09**	**1.24**	**-0.14**	**-11.66%**
Average efficiency score, NIRS	**1.09**	**1.30**	**-0.21**	**-16.12%**
Average efficiency score, SE	**1.02**	**1.10**	**-0.08**	**-7.38%**
Malmquist Index	**1.18**	**1.21**	**-0.03**	**-2.39%**
Malmquist Index, TC	**1.08**	**1.11**	**-0.03**	**-2.57%**
Malmquist Index, EC	**1.14**	**1.15**	**0.00**	**-0.41%**

<u>Second 5-years (1998-2002) data set</u>

Table 14 DEA: Comparison of the clusters based on the output-oriented model, Reduced set (four DEA inputs and two outputs)

Criterion for comparison	"Leaders" cluster	"Majority" cluster	Difference	Difference %
Average efficiency score, CRS	**1.30**	**1.68**	**-0.38**	**-22.73%**
Average efficiency score, VRS	**1.21**	**1.39**	**-0.19**	**-13.41%**
Average efficiency score, NIRS	**1.25**	**1.47**	**-0.22**	**-14.96%**
Average efficiency score, SE	**1.08**	**1.203167**	**-0.12**	**-10.36%**
Malmquist	**1.159**	**1.198**	**-0.040**	**-3.30%**

Index				
Malmquist Index, TC	**1.106**	**1.153**	**-0.047**	**-4.08%**
Malmquist Index, EC	**1.063**	**1.043**	**0.020**	**1.90%**

Table 15 DEA: Comparison of the clusters based on the output-oriented model, Full set (six DEA inputs and four outputs)

Criterion for comparison	"Leaders" cluster	"Majority" cluster	Difference	Difference %
Average efficiency score, CRS	**1.21**	**1.44**	**-0.22**	**-15.58%**
Average efficiency score, VRS	**1.18**	**1.30**	**-0.12**	**-8.88%**
Average efficiency score, NIRS	**1.21**	**1.38**	**-0.18**	**-12.78%**
Average efficiency score, SE	**1.03**	**1.093611**	**-0.06**	**-5.84%**
Malmquist Index	**1.159**	**1.159**	**0.000**	**-0.03%**
Malmquist Index, TC	**1.106**	**1.103**	**0.003**	**0.29%**
Malmquist Index, EC	**1.058**	**1.056**	**0.001**	**0.14%**

Complete 10-years (1993-2002) data set

Table 16 DEA: Comparison of the clusters based on the output-oriented model, Reduced set (four DEA inputs and two outputs)

Criterion for comparison	"Leaders" cluster	"Majority" cluster	Difference	Difference %
Average efficiency score, CRS	**1.30**	**1.77**	**-0.47**	**-26.54%**
Average efficiency score, VRS	**1.19**	**1.42**	**-0.23**	**-16.06%**
Average efficiency score, NIRS	**1.22**	**1.49**	**-0.27**	**-18.07%**
Average efficiency score, SE	**1.09**	**1.24**	**-0.15**	**-12.15%**
Malmquist Index	**1.17**	**1.20**	**-0.02**	**-1.79%**
Malmquist Index, TC	**1.10**	**1.14**	**-0.04**	**-3.83%**
Malmquist Index, EC	**1.09**	**1.08**	**0.02**	**1.45%**

Table 17 DEA: Comparison of the clusters based on the output-oriented model, Full set (six DEA inputs and four outputs)

Criterion for comparison	"Leaders" cluster	"Majority" cluster	Difference	Difference %
Average efficiency score, CRS	**1.17**	**1.41**	**-0.24**	**-16.71%**
Average efficiency score, VRS	**1.15**	**1.29**	**-0.14**	**-11.00%**

Average efficiency score, NIRS	**1.16**	**1.36**	**-0.19**	**-14.29%**
Average efficiency score, SE	**1.02**	**1.10**	**-0.07**	**-6.81%**
Malmquist Index	**1.17**	**1.20**	**-0.02**	**-1.79%**
Malmquist Index, TC	**1.10**	**1.14**	**-0.04**	**-3.83%**
Malmquist Index, EC	**1.09**	**1.08**	**0.02**	**1.45%**

Summary

Two of the many purposes of the data analysis performed in this part of the paper were to determine relative efficiencies of the "majority" and the "leaders," and to compare the relative efficiencies of these two groups of TEs produced by the DEA based on the full and the reduced data sets. We present the summarized results for the input-oriented and output-oriented model in the form of the tables and provide some brief comments after each table.

Input-Oriented Measure of Efficiency

This part of the summary demonstrates the differences in the relative efficiency between the "leaders" and the "majority" in terms of the utilization of the inputs. The input-orientation does not concern itself with the maximization of the outputs, but rather with maximization of the utilization of the inputs. Thus, it is probably reflective of the perspective of the policy maker, especially in the case when the available resources are limited.

Table 18 DEA input-oriented model: the differences between the clusters, in %, reduced data set

Criterion for comparison	1993-1998	1998-2002	1993-2002
Average efficiency score, CRS	**19.60%**	**15.45%**	**16.61%**
Average efficiency score, VRS	**10.78%**	**9.45%**	**7.90%**
Average efficiency score, NIRS	**19.13%**	**16.93%**	**16.58%**
Average efficiency score, SE	**8.40%**	**5.78%**	**7.19%**
Malmquist Index	**-3.31%**	**-0.03%**	**2.32%**
Malmquist Index, TC	**-4.71%**	**0.29%**	**2.27%**
Malmquist Index, EC	**0.72%**	**0.14%**	**-0.32%**

The table above demonstrates that the "leaders" were much more efficient than the "majority," especially during the period from 1993 to 1998. However, at that time "majority" had a higher growth in efficiency, albeit due to the change in technology. The picture changed slightly during the period fro 1998 to 2002, where the "leaders," while being still more efficient than the "majority," have exhibited practically the same growth in efficiency as the "majority."

Table 19 DEA input-oriented model: the differences between the clusters, in %, full data set

Criterion for comparison	**1993-1998**	**1998-2002**	**1993-2002**
Average efficiency score, CRS	**13.77%**	**11.79%**	**12.54%**
Average efficiency score, VRS	**6.93%**	**9.01%**	**7.48%**
Average efficiency score, NIRS	**12.69%**	**11.93%**	**11.63%**
Average efficiency score, SE	**6.85%**	**3.90%**	**4.96%**
Malmquist Index	**-2.49%**	**0.09%**	**-0.13%**
Malmquist Index, TC	**0.89%**	**0.80%**	**1.52%**
Malmquist Index, EC	**-3.70%**	**-0.93%**	**-2.11%**

This table presents the results, which were produced by DEA of the full data set. The summarized information presents very similar results in terms of the differences in relative efficiency between the "leaders" and the "majority." Similarly, the "majority" had a higher growth in efficiency during the period from 1993 to 1998; however, almost all of it came from the growth in efficiency.

Overall, comparing the results of the DEA performed using the full and the reduced data sets, we could see, that the full data set produces the smaller differences in the relative efficiencies between the "majority" and the "leaders." However, even the smaller differences obtained by the analysis of the full data set are still significant. This allows us to suggest that the "leaders" are more efficient in utilizing the inputs, and the difference in efficiencies is not likely to lessen, for the "majority" does not appear to have significantly higher growth than the "leaders."

Output-Oriented Measure of Efficiency

This part of the summary demonstrates the differences in the relative efficiency between the "leaders" and the "majority" in terms of the maximization of the outputs. Unlike in the case of the input-oriented model, the output-orientation does not concern itself with the efficient utilization of the inputs, but rather with the maximization of the outputs. Thus, it is probably reflective of the perspective of the investor, especially in the case when the resources are abundant and the primary goal is to obtain the maximum revenue.

Table 20 DEA output-oriented model: the differences between the clusters, in %, reduced data set

Criterion for comparison	**1993-1998**	**1998-2002**	**1993-2002**
Average efficiency score, CRS	**26.73%**	**22.73%**	**26.54%**
Average efficiency score, VRS	**14.11%**	**13.41%**	**16.06%**
Average efficiency score, NIRS	**17.68%**	**14.96%**	**18.07%**
Average efficiency score, SE	**13.99%**	**10.36%**	**12.15%**

Malmquist Index	**2.39%**	**3.30%**	**1.79%**
Malmquist Index, TC	**2.57%**	**4.08%**	**3.83%**
Malmquist Index, EC	**0.41%**	**-1.90%**	**-1.45%**

The table above demonstrates that the "leaders" were much more efficient than the "majority" during the period from 1993 to 1998. However, unlike in the case of the input-oriented model, where the "majority" had a higher growth in efficiency, this time the "leaders" exhibit a significantly higher growth in efficiency than the majority. The picture remains the same during the period from 1998 to 2002, where the "leaders," while being still more efficient than the "majority," have exhibited even higher growth in efficiency.

Table 21 DEA output-oriented model: the differences between the clusters, in %, full data set

Criterion for comparison	**1993-1998**	**1998-2002**	**1993-2002**
Average efficiency score, CRS	**17.62%**	**15.58%**	**16.71%**
Average efficiency score, VRS	**11.66%**	**8.88%**	**11.00%**
Average efficiency score, NIRS	**16.12%**	**12.78%**	**14.29%**
Average efficiency score, SE	**7.38%**	**5.84%**	**6.81%**
Malmquist Index	**2.39%**	**0.03%**	**1.79%**
Malmquist Index, TC	**2.57%**	**-0.29%**	**3.83%**

Malmquist Index, EC	**0.41%**	**-0.14%**	**-1.45%**

This table presents the results of DEA based on the full data set. The results are similar to the ones obtained by DEA of the reduced set in terms of the "leaders" being more efficient than the "majority." However, the magnitudes of differences are smaller. In terms of the changes in efficiency, we could see that the "leaders" were increasing their efficiency faster than the "majority" during the period from 1993 to 1998. During the period from 1998 to 2002, however, the "majority" start increasing their efficiency faster than the "leaders."

Overall, just like in the case of the input-oriented model, we could see that the full data set produces the smaller differences in the relative efficiencies between the "majority" and the "leaders." However, even the smaller differences obtained by the analysis of the full data set are still significant. This allows us to suggest that the "leaders" are more efficient in converting the inputs into the outputs, and the difference in efficiencies is not likely to lessen, for the "majority" does not appear to have a higher growth in efficiency than the "leaders."

Based on the comparison of the results produced by DEA of the reduced and the full data sets, we have concluded that the DEA of the full data set yields the smaller magnitudes of the differences in the efficiencies between the "leaders" and the "majority." Thus, consistent with our intent to bias the results on the side of underestimation, we would use the results of the DEA based on the full data set in further analyses.

7.3.5 Correlation between changes in Productivity and Malmquist Index

In this part of the paper, we provide the data regarding the correlational relationship between the changes in productivity and values of the Malmquist index. As it was mentioned previously, the variable "productivity" was created by obtaining the ratio of the variable "Telecom revenue per telecom worker" to the "Annual telecom investment per telecom worker." However, because the total number of the telecom workers was considered to be unchanging in the period of one year, the variable "productivity" represents, in fact, a ratio of the "Total Telecom revenue" to "Annual telecom investment" for each year.

Despite the intent, however, we could not correlate "productivity" and Malmquist index, for "productivity" is static and Malmquist index represents a change that took place in the period of one year. Thus, we have created a new variable, "change in productivity," the values for which were obtained by subtracting, let us say, a value of productivity for year 1994 from the value of productivity for the previous year, 1993. As a result, we were able to construct the variable with the values representing, conceptually, the magnitude of change that took place over the period of one year.

Once the values for the variable "change in productivity" were calculated, we have correlated those variables with the values provided by Malmquist index. Again, we have used three data sets representing the first 5-Year period (1993-1998), the second 5-Year period(1998-2002), and the 10-Year period from 1993 to 2002. The obtained values are provided in the table below.

Table 22 Correlation coefficients for the relationships between Malmquist index and Change in Productivity

	1st 5 years (1993-1998)	2nd 5 years (1998-2002)	10 years (1993-2002)
Albania	**0.94 ***	**0.88***[11]	**0.91*****
Armenia	-0.29	**0.96*****[12]	0.15
Azerbaijan	**0.92***	***0.82*****[13]	**0.78*****
Belarus	0.04	0.64	0.11
Bulgaria	-0.19	**0.88***	***0.56*****
Czech Republic	0.77	0.64	***0.60*****
Estonia	0.47	0.21	0.36
Hungary	0.48	0.78	**0.63***
Kazakhstan	0.35	0.50	0.39
Kyrgyz Republic	0.75	**0.88***	**0.72***
Latvia	0.25	**0.87***	0.20
Lithuania	**0.96*****	0.03	**0.68***
Moldova	0.26	**0.95*****	0.24
Poland	**0.92***	**0.99*****	**0.91*****
Romania	-0.23	**0.89***	**0.78*****

[11] *- statistically significant at $\alpha = 0.1$ level;

[12] ***- statistically significant at $\alpha = 0.01$ level;

[13] **- statistically significant at $\alpha = 0.05$ level;

Slovak Republic	0.67	**0.99***	**0.94***
Slovenia	***0.86***	**0.89***	**0.87***
Ukraine	**0.88***	0.08	**0.74***

One of the ways of testing the null hypothesis of non-zero correlation happening by chance is by using Student's t-test. We have obtained the values corresponding to the 1%, 5% and 10% significance levels, which are provided in the table below.

Data set/Significance level	**10% significance level**	**5% significance level**	**1% significance level**
5-year series(N = 5)	0.805	0.878	0.959
10-year series(N = 10)	0.549	0.632	0.765

Most of the values were found to be statistically significant at 1% and 5% level. Let us elaborate the significance of the obtained results: we were able to establish the presence of the indirect relationship between the Malmquist index and the investments in telecom for 67% of the TEs in our sample. In this case, the variable "productivity" serves as a proxy that mediates the relationship between the investments in telecom and TFP. Let us recall, that we suggested the way around the "endogeneity problem" that was based on the mediated by the latent variable relationship between the investments in ICT and TFP. The presented in the table of correlation coefficients results are suggestive that our proposed approach could be successful.

In the same time, we were investigating the presence of the direct correlation between the changes in the investments in ICT and TFP. Following is the reason for this inquiry: we would like to propose and test our theoretical framework based on indirect relationship between the investments in ICT and TFP, but only while being firmly rooted in the Neoclassical Growth Accounting.

While the presented above results of indirect correlation between the investments in ICT and TFP are encouraging, we would like to make sure that it is in the same time consistent with the Neoclassical Growth Accounting framework, which does not allow for the direct relationship between the investments in ICT and TFP. Thus, in this case the goal is to reject the null hypothesis of the presence of the direct relationship between the TFP and the investments in ICT. The obtained results are presented in the table below.

Table 23 Correlation coefficients for the relationships between Malmquist index and Change in ICT investments

	1st 5 years(1993-1998)	**2nd 5 years(1998-2002)**	**10 years (1993-2002)**
Albania	**-0.88**	-0.72	**-0.74**
Armenia	0.59	**-0.96**	-0.03
Azerbaijan	**-0.94**	-0.38	**-0.68**
Belarus	0.57	0.55	0.56
Bulgaria	0.58	-0.17	0.04
Czech Republic	-0.10	-0.53	-0.25
Estonia	-0.40	0.09	-0.20
Hungary	-0.49	0.27	-0.20
Kazakhstan	0.07	-0.62	-0.02
Kyrgyz Republic	**-0.96**	**-0.98**	**-0.93**
Latvia	0.35	-0.84	0.36
Lithuania	**-0.92**	0.64	-0.13

Moldova	0.28	**-0.97**	-0.21
Poland	-0.28	-0.85	-0.39
Romania	0.60	-0.74	-0.50
Slovak Republic	0.41	-0.28	-0.28
Slovenia	-0.61	-0.84	**-0.74**
Ukraine	-0.06	0.59	0.02

The presented above results suggest that we are able to reject the null hypothesis of the presence of the direct relationship between the TFP and the investments in ICT for the majority of TEs in our sample. It could not be overlooked, though, that we are not able to reject the null hypothesis in the case of Slovenia, Kyrgyz Republic, Azerbaijan, and Albania, which taken together represent 22% of our sample. Overall, however, the required by the Neoclassical Growth Accounting framework absence of the direct relationship between the TFP and the investments in ICT hold for the majority of TEs in our sample.

Next, we decided to obtain the average values of change in productivity for each TE in our sample, and rank the economies in terms of these values. The purpose of this ranking is to inquire whether or not we would be able to observe similar to the produced by the Cluster analysis results. Specifically, we would like to find out whether or not it would be possible to produce the groupings of TEs based on the change in Productivity, and whether or not the resultant groups would bear some similarity to the clusters produced by the Cluster analysis. The results are presented in the table below.

Table 24 Rankings of the countries of TEs in terms of the Average Change in Productivity

Rank	Percent	Country	Change in Productivity, Average
1	100.00%	Slovak Republic	0.51
2	94.10%	Romania	0.48
3	88.20%	Hungary	0.43
4	82.30%	Poland	0.37
5	76.40%	Estonia	0.28
6	70.50%	Latvia	0.23
7	64.70%	Czech rep	0.22
8	58.80%	Slovenia	0.16
9	52.90%	Lithuania	-0.13
10	47.00%	Belarus	-0.17
11	41.10%	Moldova	-0.22
12	35.20%	Kazakhstan	-0.39
13	29.40%	Bulgaria	-0.51
14	23.50%	Azerbaijan	-0.51
15	17.60%	Armenia	-0.58
16	11.70%	Ukraine	-0.94
17	5.80%	Albania	-0.99
18	0.00%	Kyrgyz rep	-2.80

To make the presentation of the results more vivid, we plotted the values of the average change in productivity for each economy side-by-side in the Figure 16.

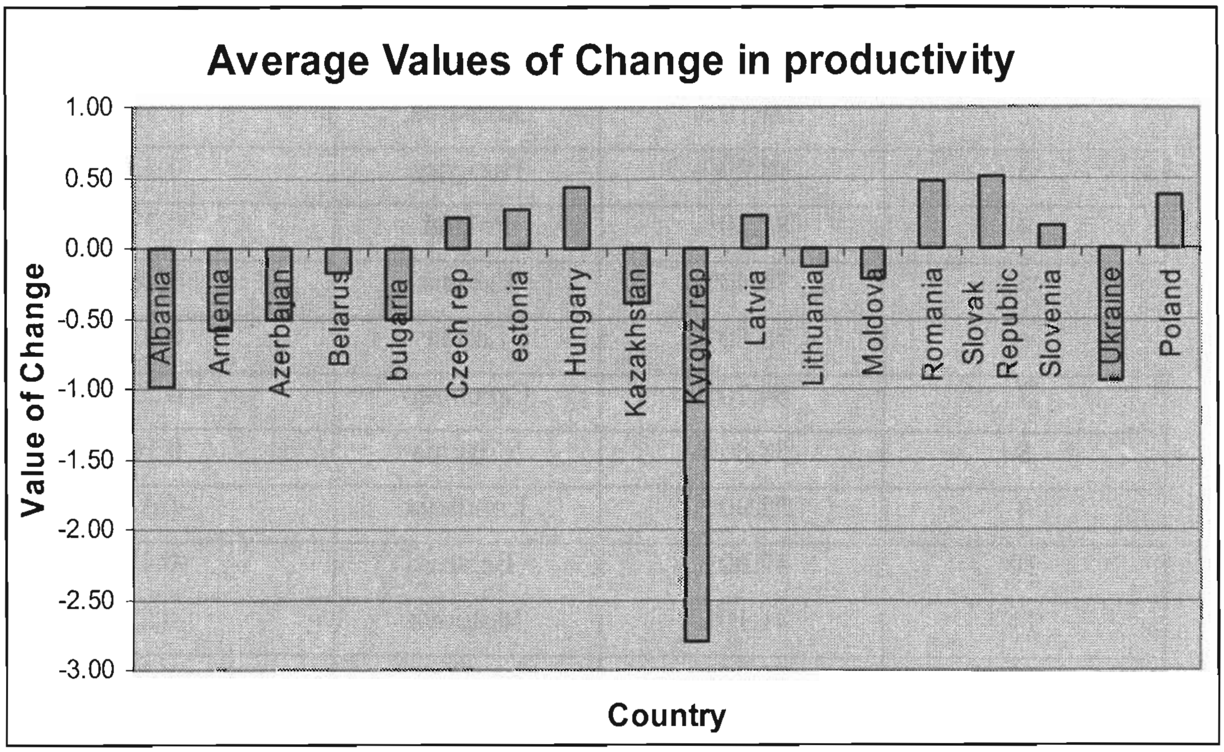


Figure 16 Plotted values of the average change in productivity for each economy

As we expect, the highest averages in the set belong to the group of countries that were isolated into the cluster "leaders." Unexpectedly, Romania made its way into the group of "leaders." We have decided to investigate the reasons for such 'anomaly'. One of the possible explanation that we found is that in the period from 2000 to 2002 Romania has tripled its Total telecom revenue, all while reducing on 20% the investments in ICT. We have calculated the average change in productivity for Romania during the period from 1993 to 2000, and it turned out to be 0.010. Clearly, the unusual data values for the last two years should be considered as outliers, rather than norm; for Romania did not exhibit any indications of the normality of this pattern.

7.4 Decision Tree Analysis

7.4.1 DT: Data

In this section, we briefly present preliminary results obtained after the application of the Decision Tree analysis. The purpose of the analysis was to find out what the specific characteristics of the efficient DMUs are. We have created a new binary variable "Efficiency" which could take a value of "1" or "0." The value of "1" was assigned in the case if DEA analysis yielded such value for a given DMU in a given year, while the value of "0," which was assigned to any inefficient DMU for a given year, regardless of whether it was "0.99" or "0.22." Then the 2 data sets used for Cluster analysis and DEA were combined together and the duplicate columns were removed. After that, "Efficiency" variable was added and the resulting data set was used as a data source for Classification Tree analysis. Because we have 3 DEA models, CRS, VRS, and NIRS, we have ended up with 3 data sets, thus, 3 results are presented. During the analysis, following nodes were used: "Data Source," "Variable Selection," "Decision Tree" and "Reporter" for classification analysis, and "Data Partition" node was added in order to create the predictive models. We have chosen to use the default settings of the "Decision Tree" node during the Classification Tree analysis.

Following is the list of variables used in DT analysis:

1. Total telecom services revenue(Current US $)
2. Total telecom services revenue(Current US $ per person)

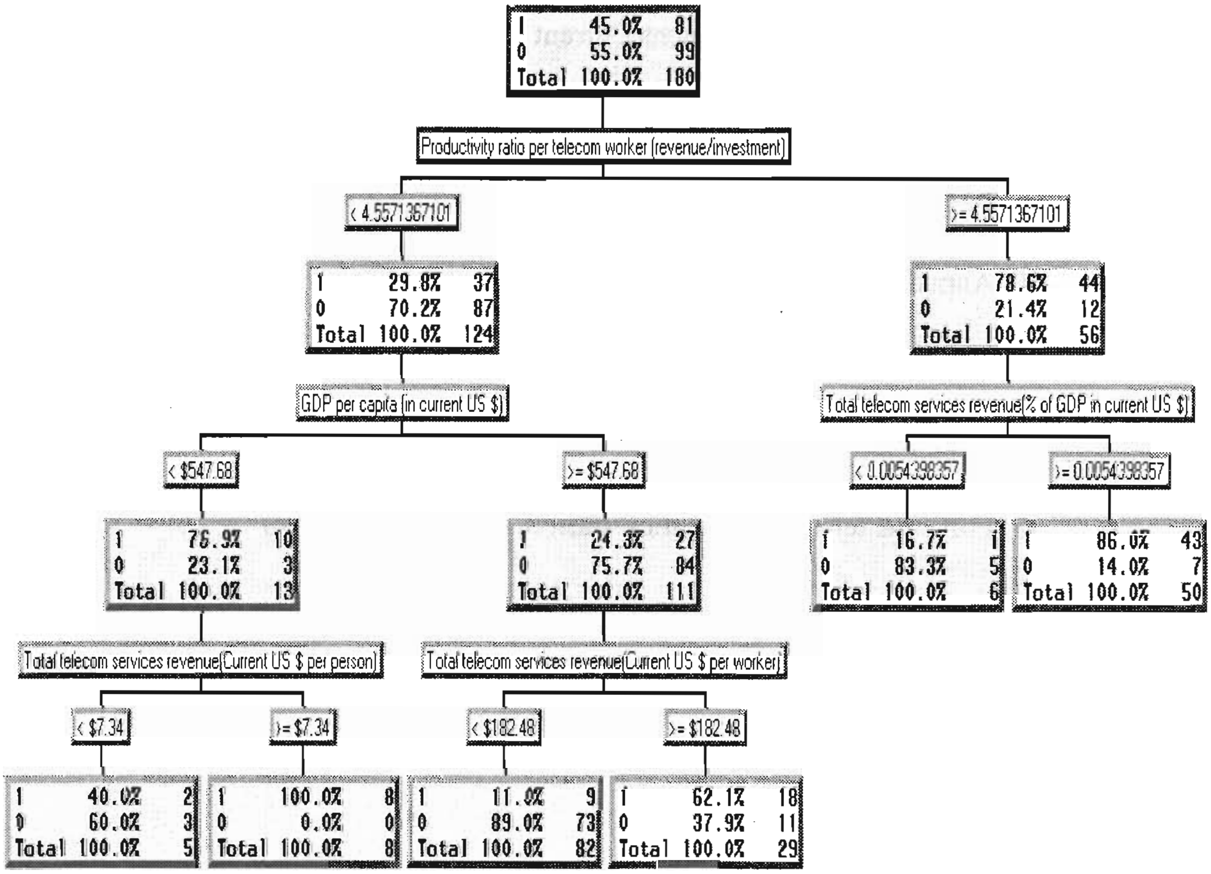
3. Total telecom services revenue(Current US $ per worker)
4. Annual telecom investment(Current US $)
5. Efficiency
6. Annual telecom investment(Current US $ per worker)
7. Annual telecom investment(Current US $ per telecom worker)
8. Total telecom services revenue(Current US $ per telecom worker)
9. Productivity ratio per telecom worker (revenue/investment)
10. Annual telecom investment(% of GDP)
11. GDP per capita (in current US $)
12. Total telecom services revenue(% of GDP in current US $)
13. Full-time telecommunication staff(% of total labor force)

7.4.2 DT: Results

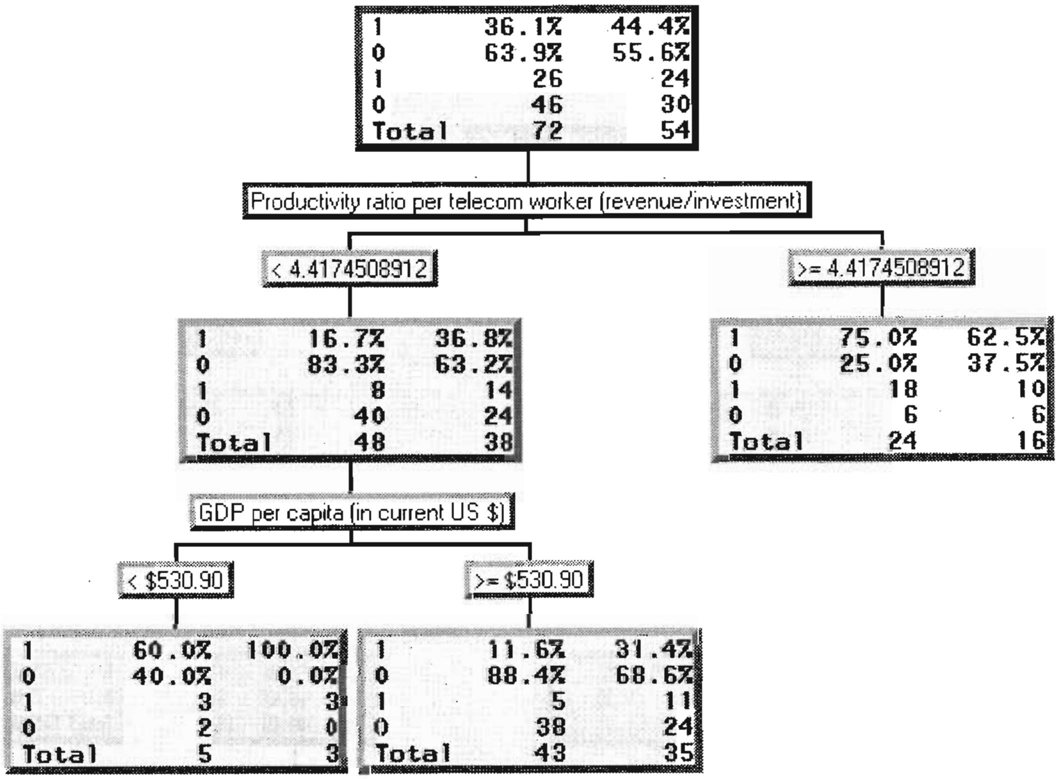
Part 1

Full data set, CRS setting of DEA

Classification Tree



Predictive Model



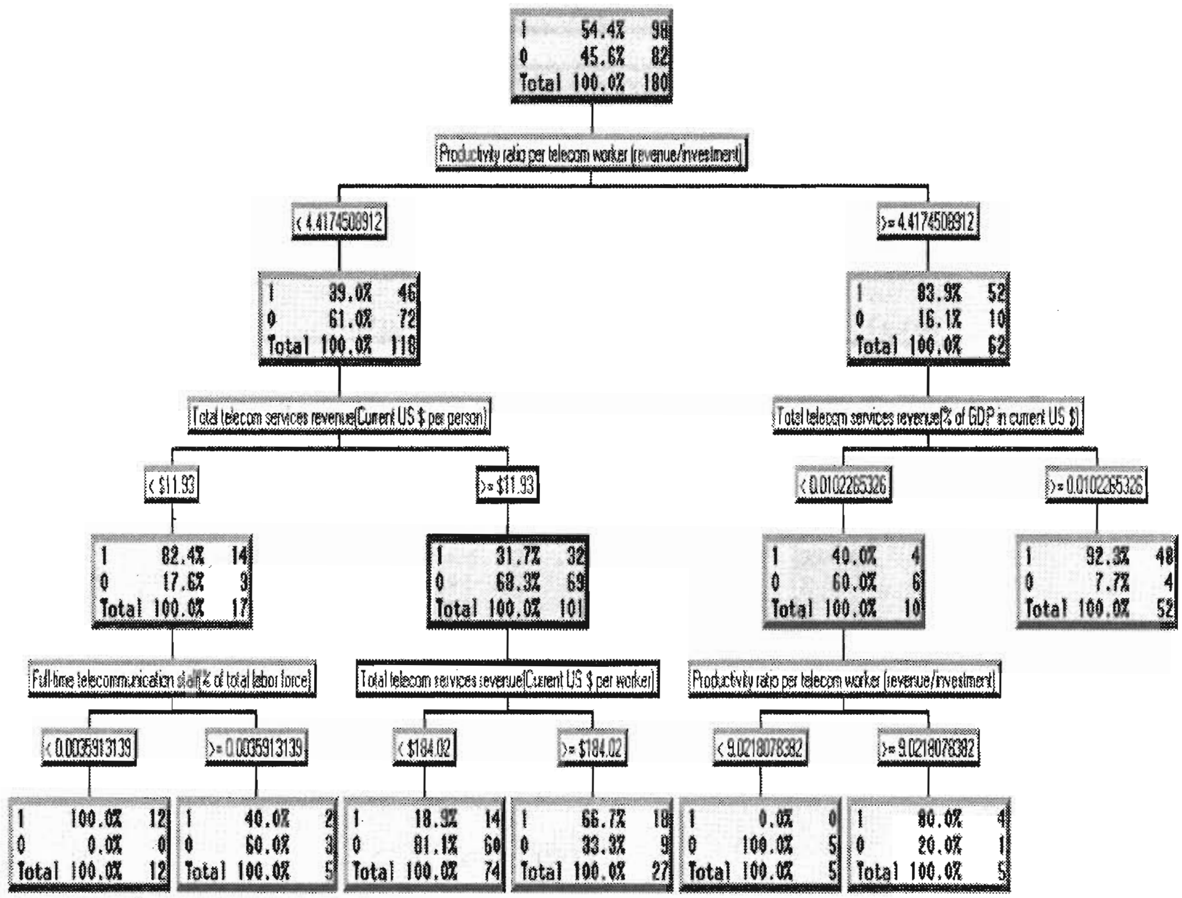
Confusion Matrix

target output

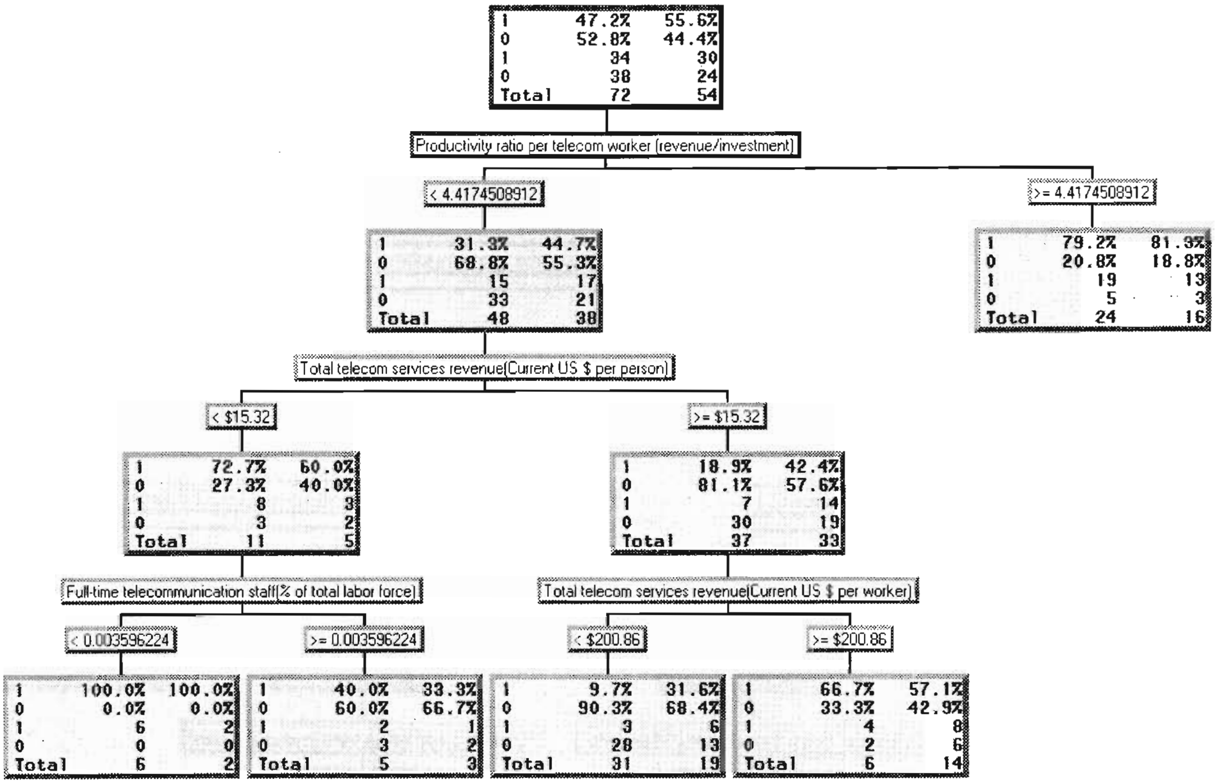
Frequency Percent Row Pct Col Pct	0	1	Total
0	38 52.78 82.61 88.37	8 11.11 17.39 27.59	46 63.89
1	5 6.94 19.23 11.63	21 29.17 80.77 72.41	26 36.11
Total	43 59.72	29 40.28	72 100.00

Full data set, VRS setting of DEA

Classification Tree



Predictive Model

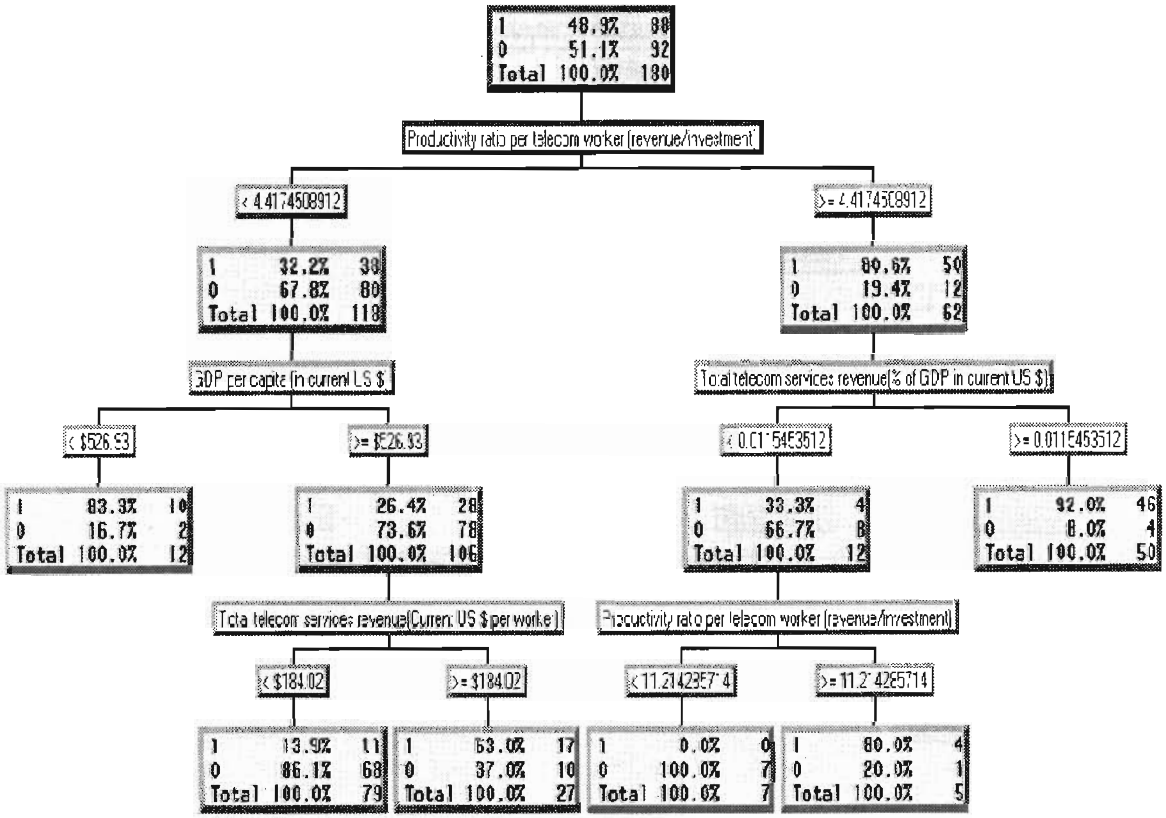


Confusion matrix

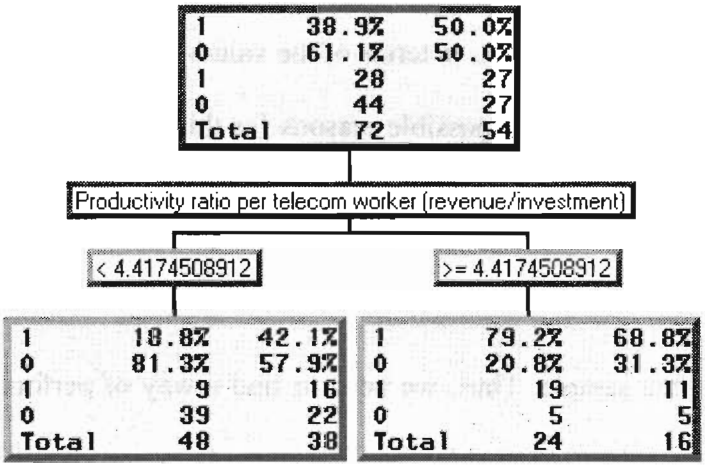
```
          target     output

Frequency|
Percent  |
Row Pct  |
Col Pct  |0       |1       |  Total
---------+--------+--------+
0        |     31 |      7 |     38
         |  43.06 |   9.72 |  52.78
         |  81.58 |  18.42 |
         |  86.11 |  19.44 |
---------+--------+--------+
1        |      5 |     29 |     34
         |   6.94 |  40.28 |  47.22
         |  14.71 |  85.29 |
         |  13.89 |  80.56 |
---------+--------+--------+
Total          36       36       72
            50.00    50.00   100.00
```

Full data set, NIRS setting of DEA

<u>Classification Tree</u>



Predictive Model



Confusion Matrix

Table of target by output

target	output		
Frequency Percent Row Pct Col Pct	0	1	Total
0	39 54.17 88.64 81.25	5 6.94 11.36 20.83	44 61.11
1	9 12.50 32.14 18.75	19 26.39 67.86 79.17	28 38.89
Total	48 66.67	24 33.33	72 100.00

Summary of the Part 1

Based on the results of this part of the DT analysis we cannot conclude that the "efficient" TEs are clearly different, in terms of the values of variables that we used, from the "inefficient" TEs. One of the possible reasons for this could be the fact that 18 TEs that comprise our sample are not, as it was demonstrated by CA, homogenous.

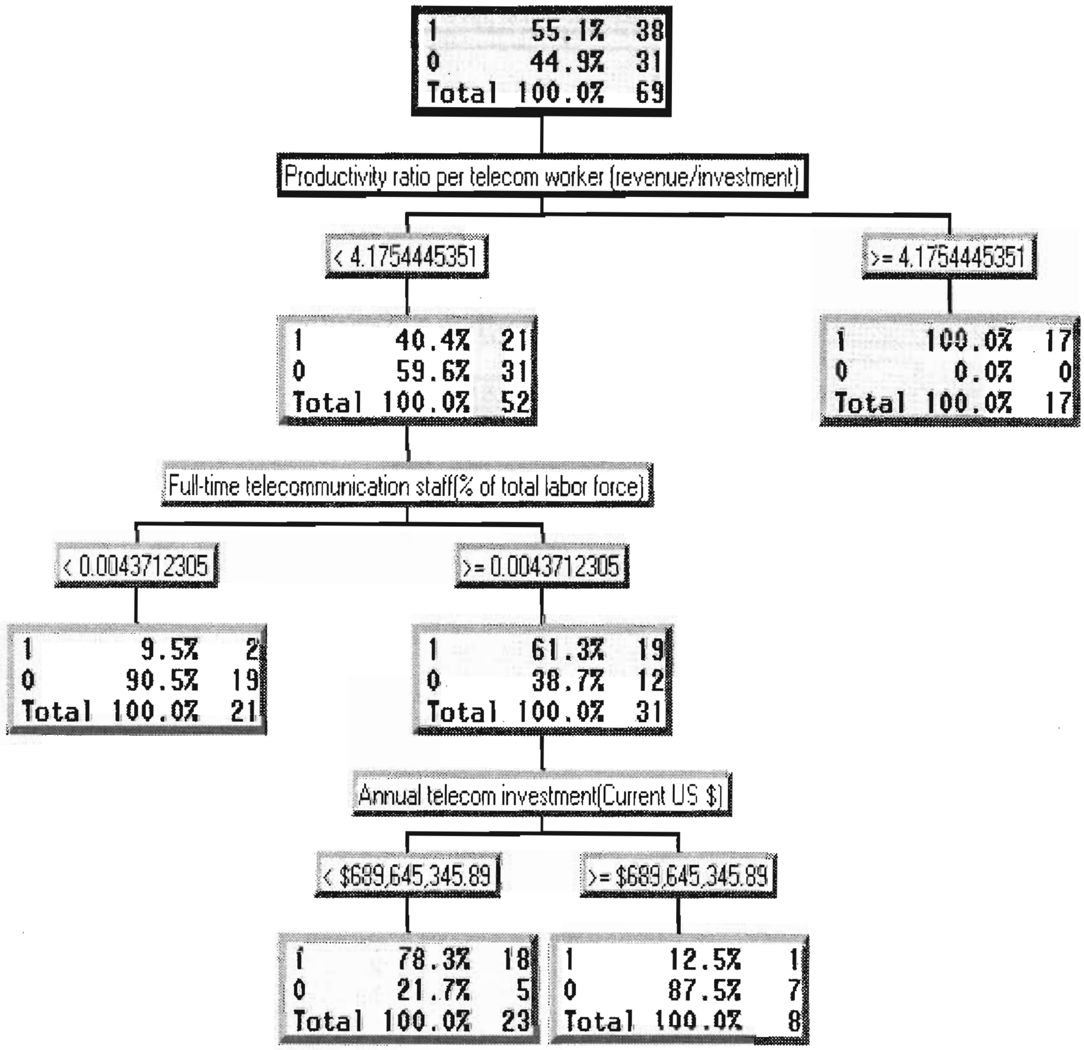
One of the fundamental assumptions of DEA, however, is functional similarity of DMUs comprising the sample. Thus, we need to find a way of performing DT analysis that would account for the identified by CA differences between TEs of our sample.

We have decided to separate the provided by DEA results into two subsets: first, the results for the "leaders," and, second, the results for the "majority." Consequently, we ended up with the two subsets of the data, and each of the subsets, corresponding to the "leaders" and "majority," was comprised of "efficient" and "inefficient" DMUs. Having done so, we performed the DT analysis again. We present the results of the second part of the DT analysis next.

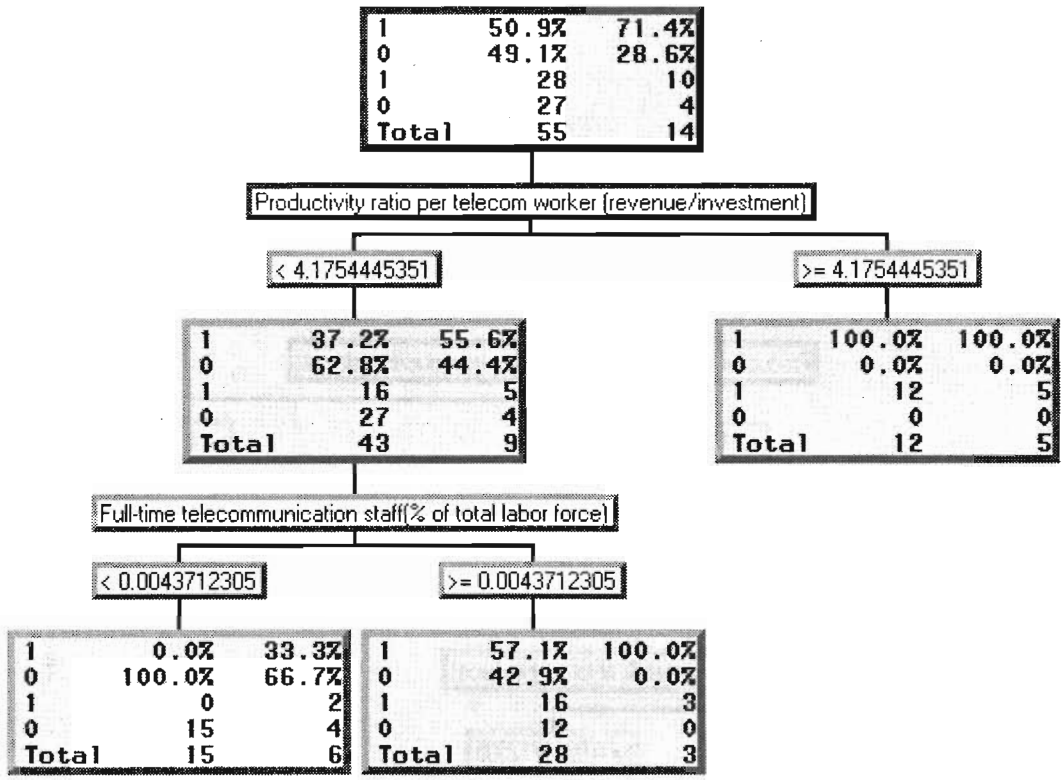
Part 2

Full data set, CRS setting of DEA, "Leaders" cluster

Classification Tree



Predictive Model



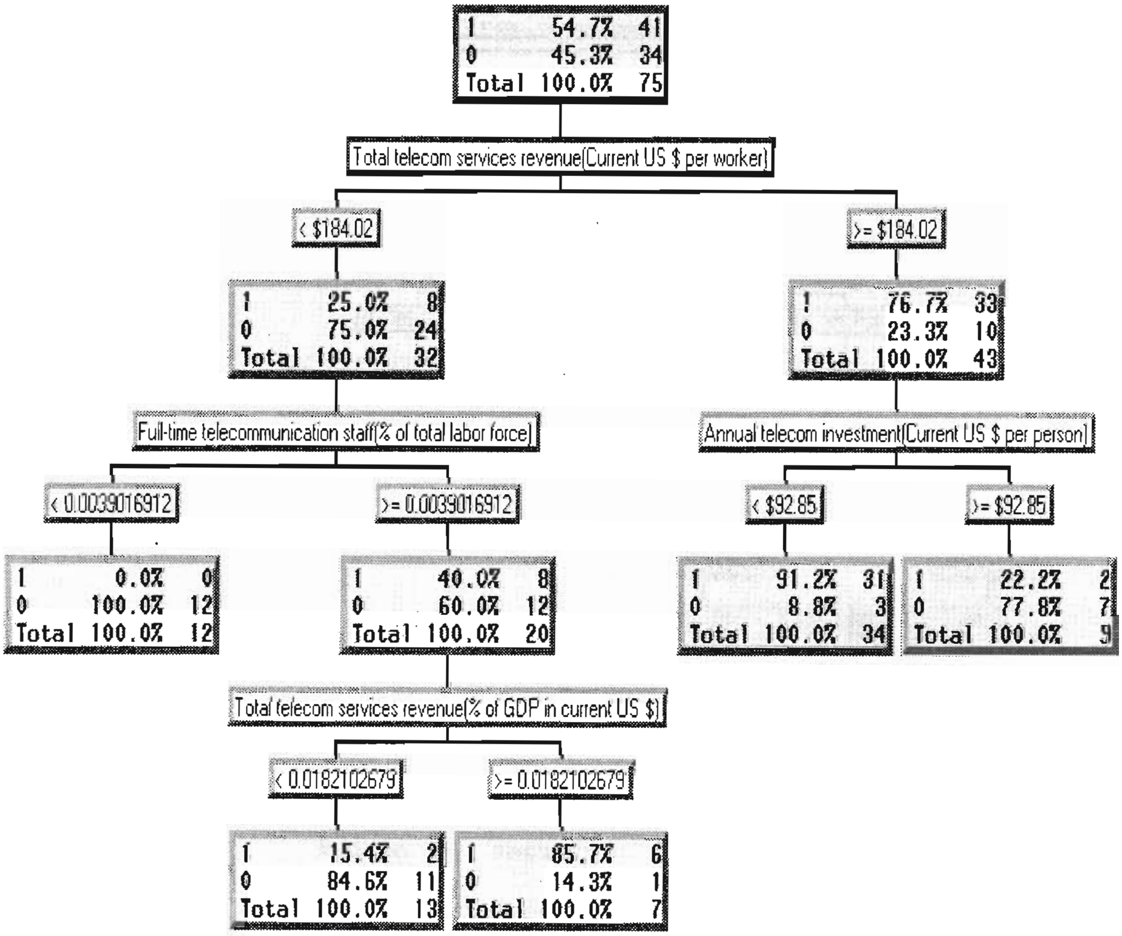
Confusion matrix

target output

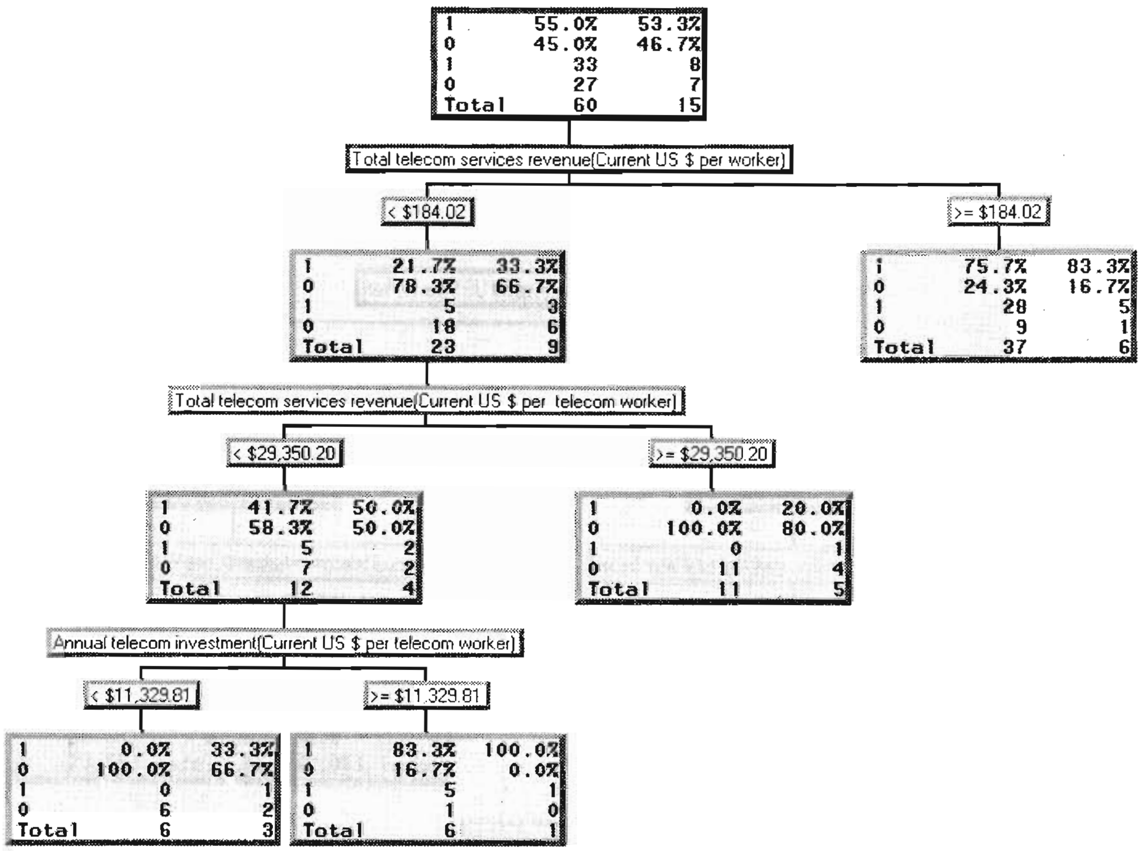
Frequency / Percent / Row Pct / Col Pct	0	1	Total
0	15 27.27 55.56 100.00	12 21.82 44.44 30.00	27 49.09
1	0 0.00 0.00 0.00	28 50.91 100.00 70.00	28 50.91
Total	15 27.27	40 72.73	55 100.00

Full data set, NIRS setting of DEA, "Leaders" cluster

Classification Tree



Predictive Model



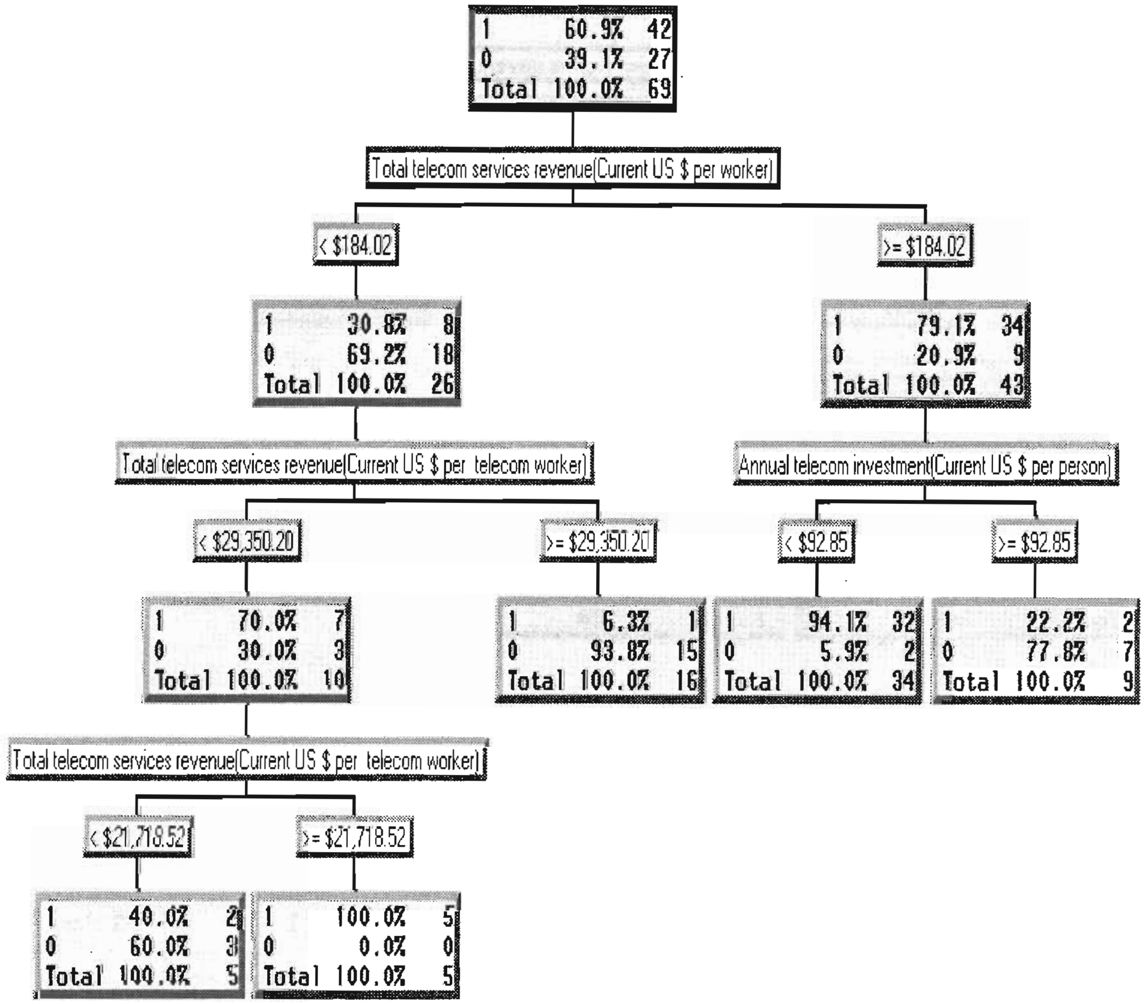
Confusion matrix

target output

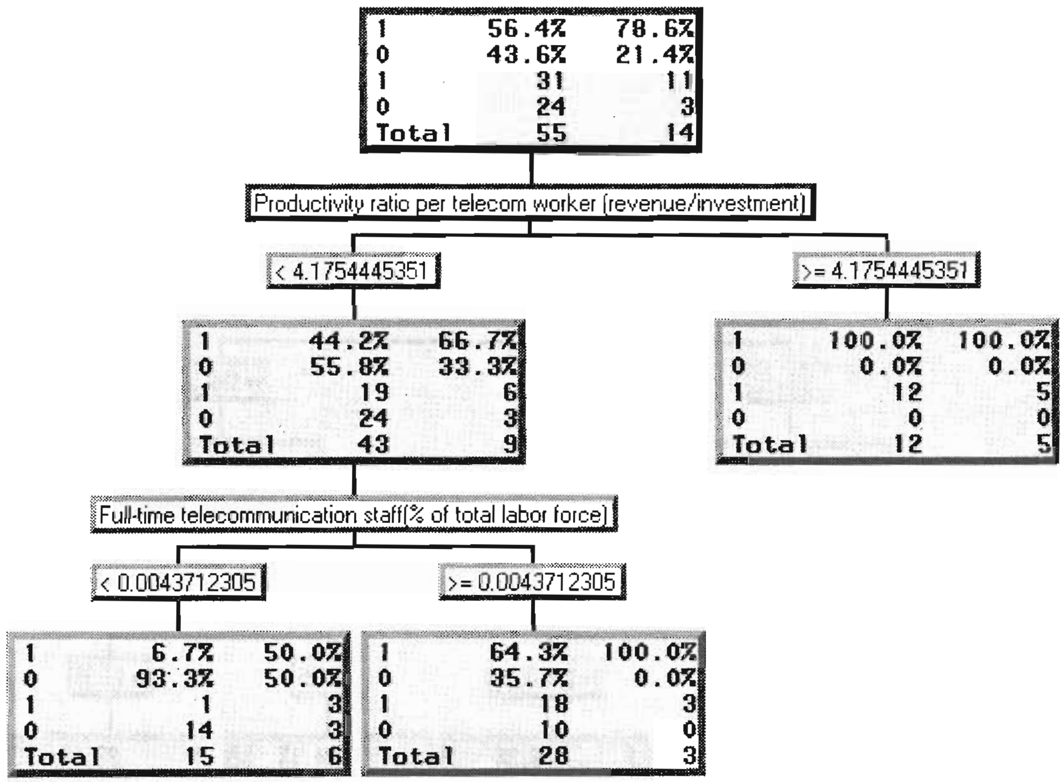
Frequency Percent Row Pct Col Pct	0	1	Total
0	17 28.33 62.96 100.00	10 16.67 37.04 23.26	27 45.00
1	0 0.00 0.00 0.00	33 55.00 100.00 76.74	33 55.00
Total	17 28.33	43 71.67	60 100.00

Full data set, VRS setting of DEA, "Leaders" cluster

Classification Tree



Predictive Model

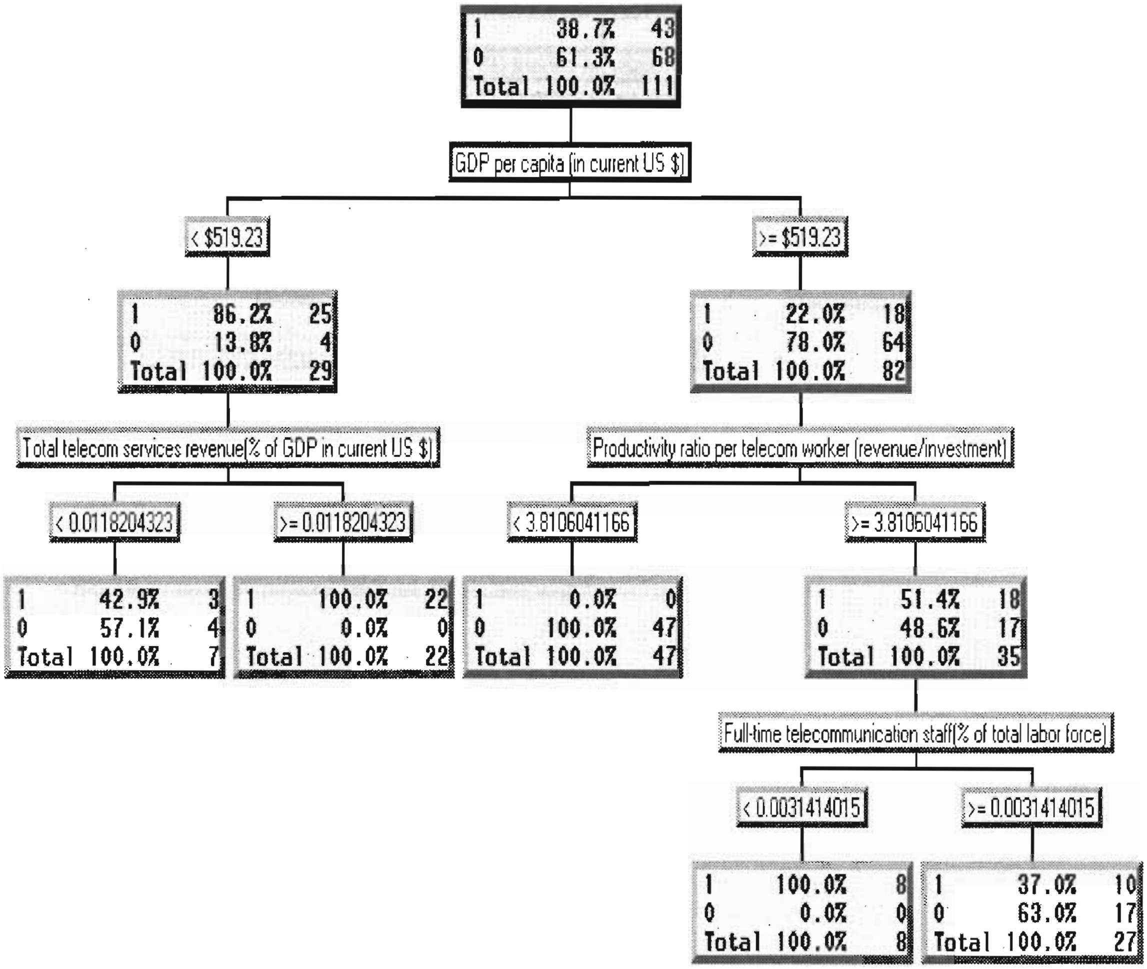


Confusion matrix

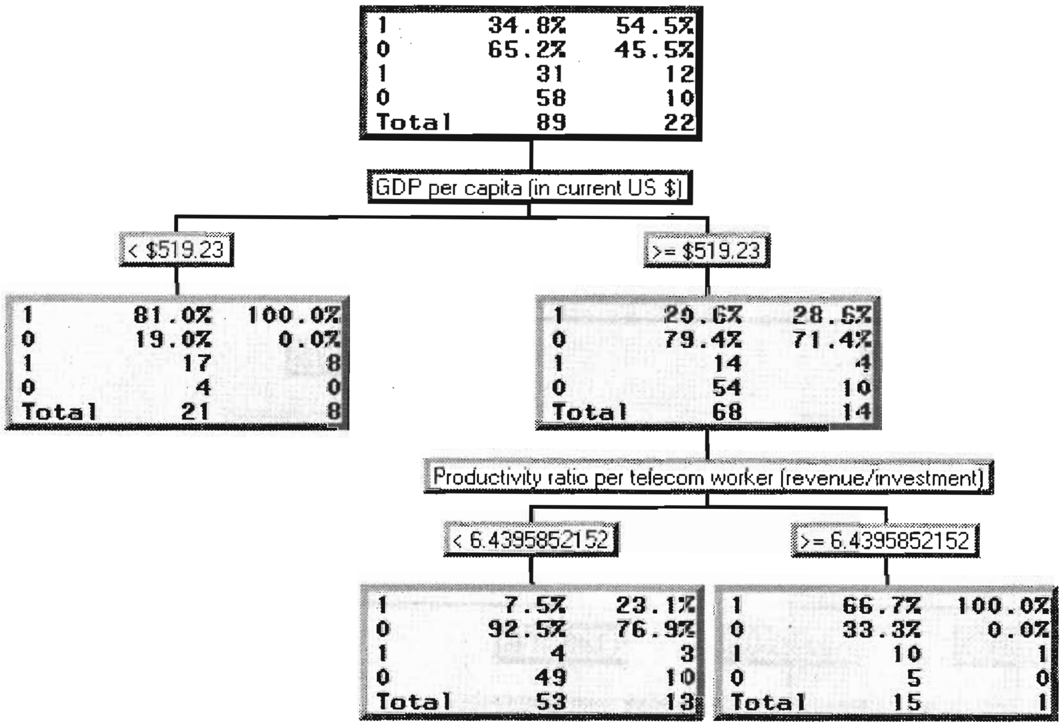
```
    target      output

Frequency|
Percent  |
Row Pct  |
Col Pct  |0        |1        |  Total
---------+---------+---------+
0        |      14 |      10 |     24
         |   25.45 |   18.18 |  43.64
         |   58.33 |   41.67 |
         |   93.33 |   25.00 |
---------+---------+---------+
1        |       1 |      30 |     31
         |    1.82 |   54.55 |  56.36
         |    3.23 |   96.77 |
         |    6.67 |   75.00 |
---------+---------+---------+
Total         15        40        55
           27.27     72.73    100.00
```

Full data set, CRS setting of DEA, "Majority" cluster

Classification Tree



Predictive Model

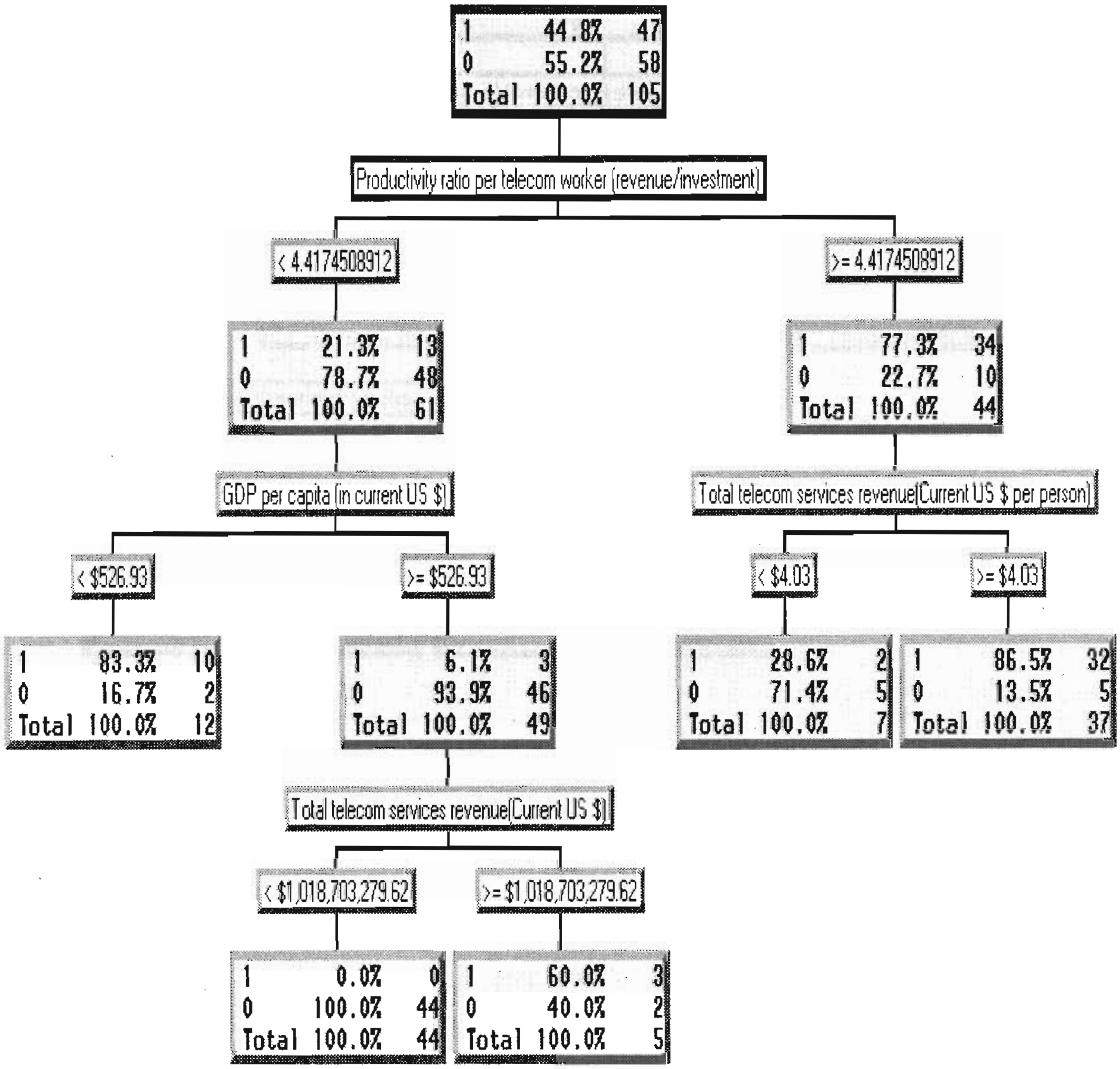


Confusion matrix

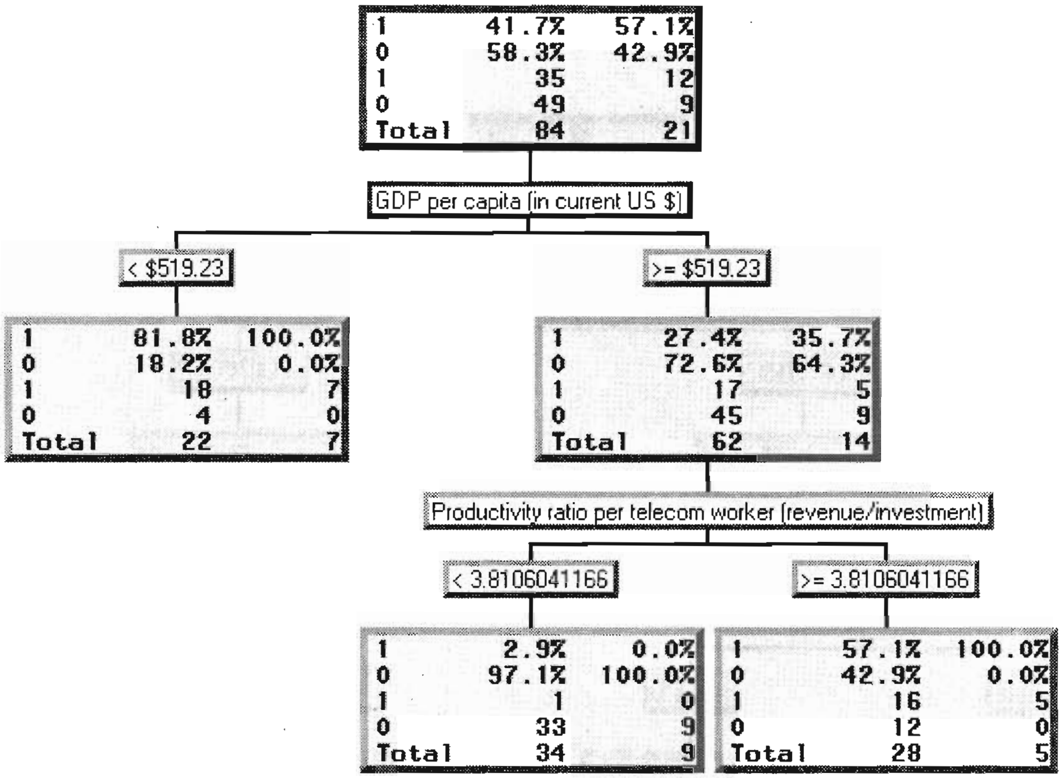
```
          target      output

Frequency|
Percent  |
Row Pct  |
Col Pct  |0       |1       |  Total
---------+--------+--------+
0        |     49 |      9 |     58
         |  55.06 |  10.11 |  65.17
         |  84.48 |  15.52 |
         |  92.45 |  25.00 |
---------+--------+--------+
1        |      4 |     27 |     31
         |   4.49 |  30.34 |  34.83
         |  12.90 |  87.10 |
         |   7.55 |  75.00 |
---------+--------+--------+
Total          53       36       89
            59.55    40.45   100.00
```

Full data set, NIRS setting of DEA, "Majority" cluster

Classification Tree



Predictive Model

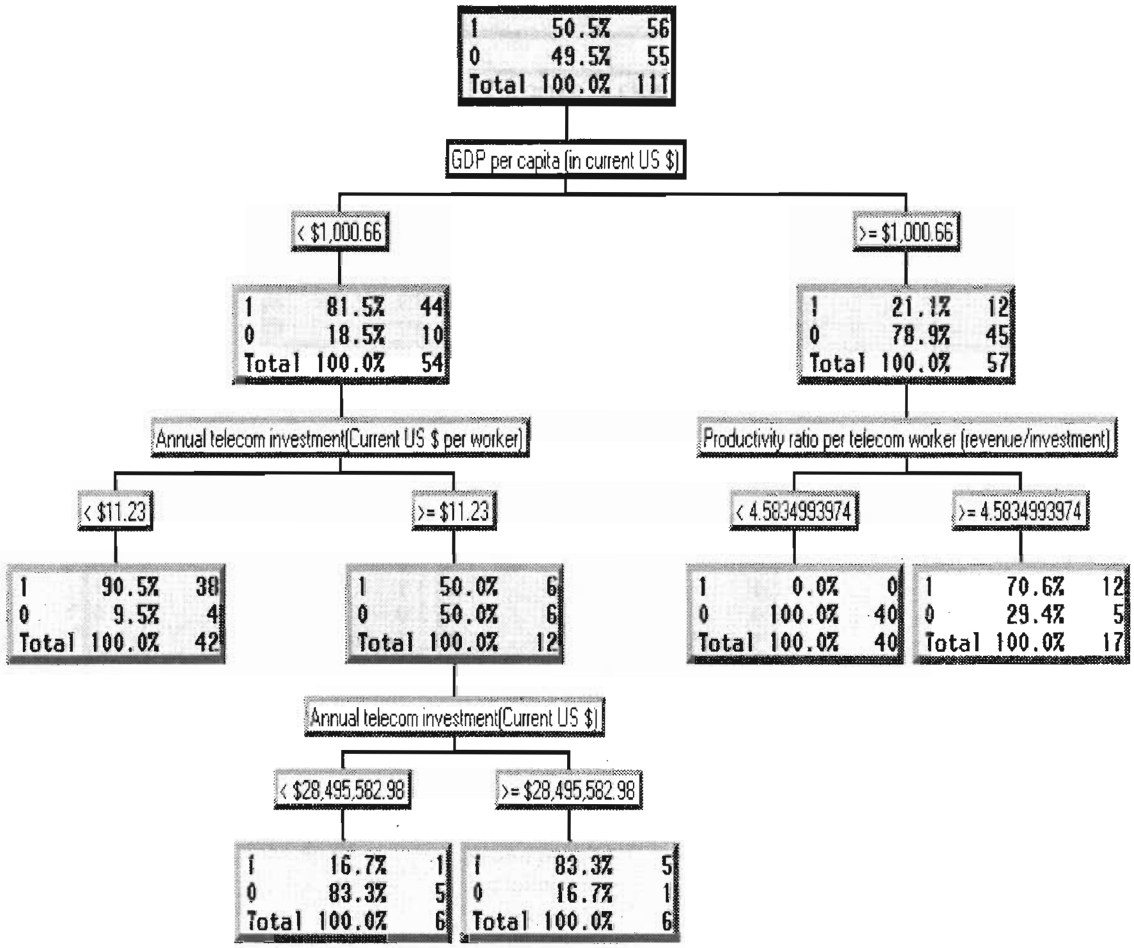


Confusion matrix

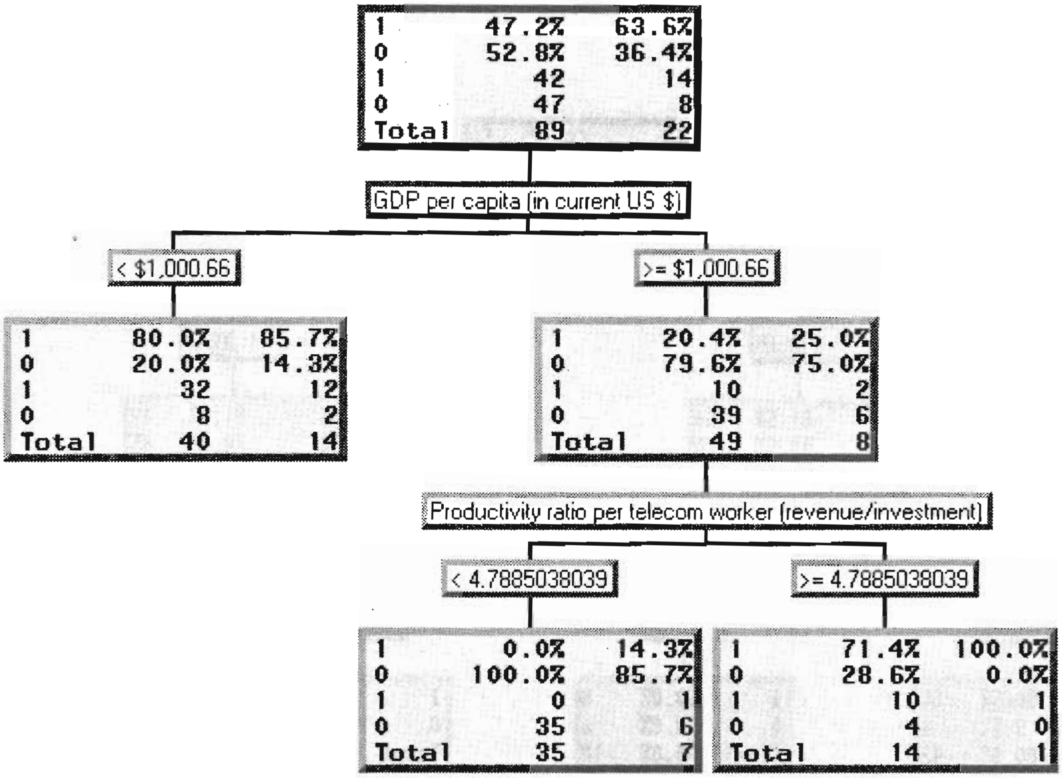
```
          target     output

Frequency|
Percent  |
Row Pct  |
Col Pct  |0       |1       |  Total
---------+--------+--------+
0        |     33 |     16 |     49
         |  39.29 |  19.05 |  58.33
         |  67.35 |  32.65 |
         |  97.06 |  32.00 |
---------+--------+--------+
1        |      1 |     34 |     35
         |   1.19 |  40.48 |  41.67
         |   2.86 |  97.14 |
         |   2.94 |  68.00 |
---------+--------+--------+
Total          34       50       84
            40.48    59.52   100.00
```

Full data set, VRS setting of DEA, "Majority" cluster

<u>Classification Tree</u>



Predictive Model



Confusion matrix

target output

Frequency / Percent / Row Pct / Col Pct	0	1	Total
0	35 39.33 74.47 100.00	12 13.48 25.53 22.22	47 52.81
1	0 0.00 0.00 0.00	42 47.19 100.00 77.78	42 47.19
Total	35 39.33	54 60.67	89 100.00

Summary of the Part 2

The new DT models provide much better results than the ones presented in the Part 1. However, these models still provide only limited benefit primarily because they force the researcher to consider two different groups of TEs, the "leaders" and "majority," in isolation.

Based on the results of CA we have established that while some of the countries are permanently located in one of the cluster, there is a fairly large group of TEs that tend to "migrate" from one cluster to another. As a matter of fact, almost 28% of the TEs in our data set are the "migrants" (see table 25 below). Consequently, in order to deal with such countries we would need to have both DT models, one for the periods when such countries are considered to be among the "leaders" and another model for the periods when these countries are among the "majority."

Table 25 DT: Permanent leaders, permanent majority, and migrants

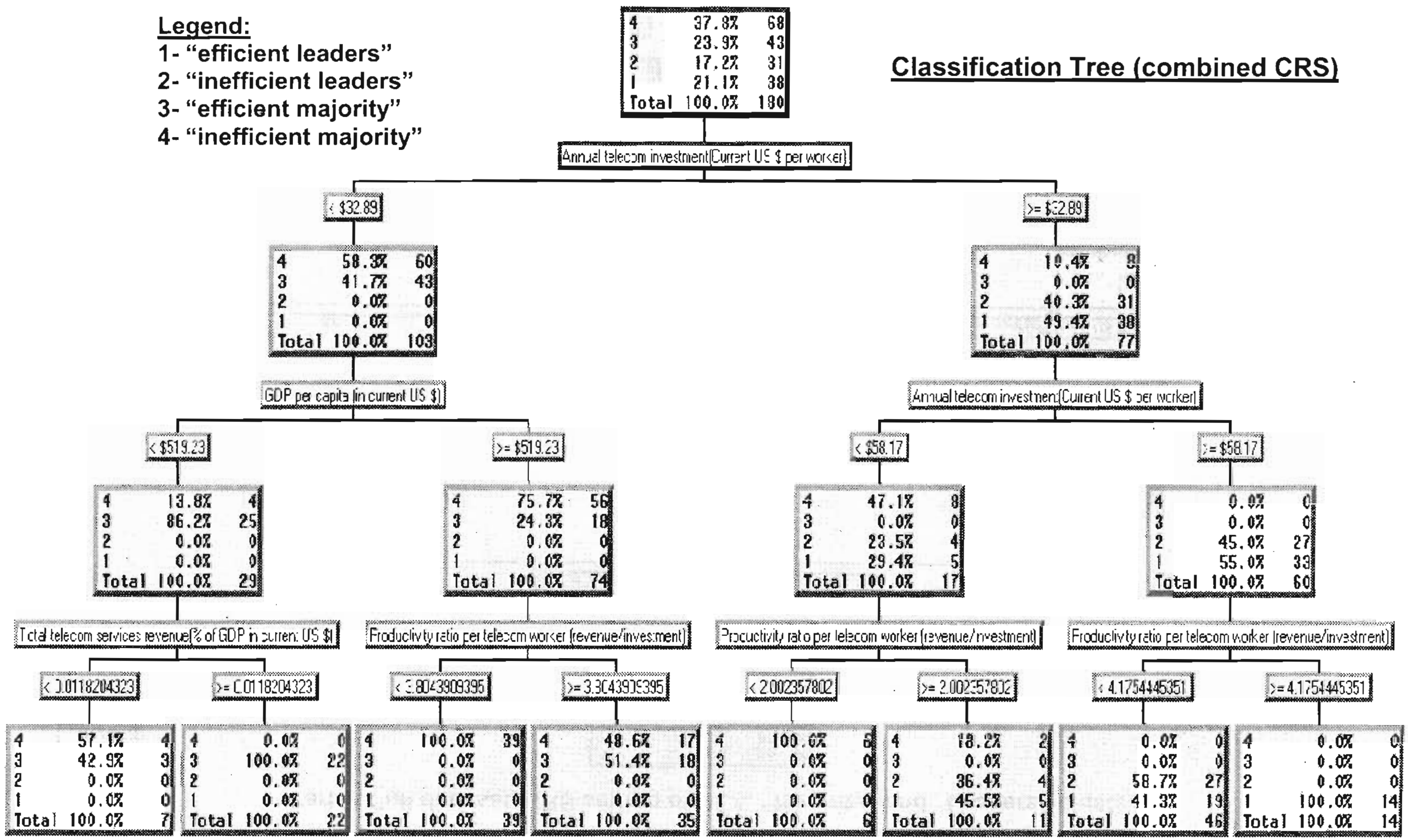
Permanent leaders	Permanent majority	Migrants (years as leaders)
Czech rep (1993-2002) Hungary (1993-2002) Slovenia (1993-2002) Poland (1993-2002)	Albania (1993-2002) Armenia(1993-2002) Azerbaijan(1993-2002) Belarus(1993-2002) Kazakhstan(1993-2002) Kyrgyz Rep (1993-2002) Moldova (1993-2002) Romania(1993-2002) Ukraine (1993-2001)	Estonia (1994-2002) Bulgaria (2002) Latvia (1994, 1995, 1997-2002) Lithuania (1999-2002) Slovak Rep(1995-1998, 2000-2002)

We have decided to modify our DT models further, by incorporating both, the information about the TEs membership in a given cluster, and the efficiency score, into the single DT model. We have realized that we could categorize our TEs into the four groups:

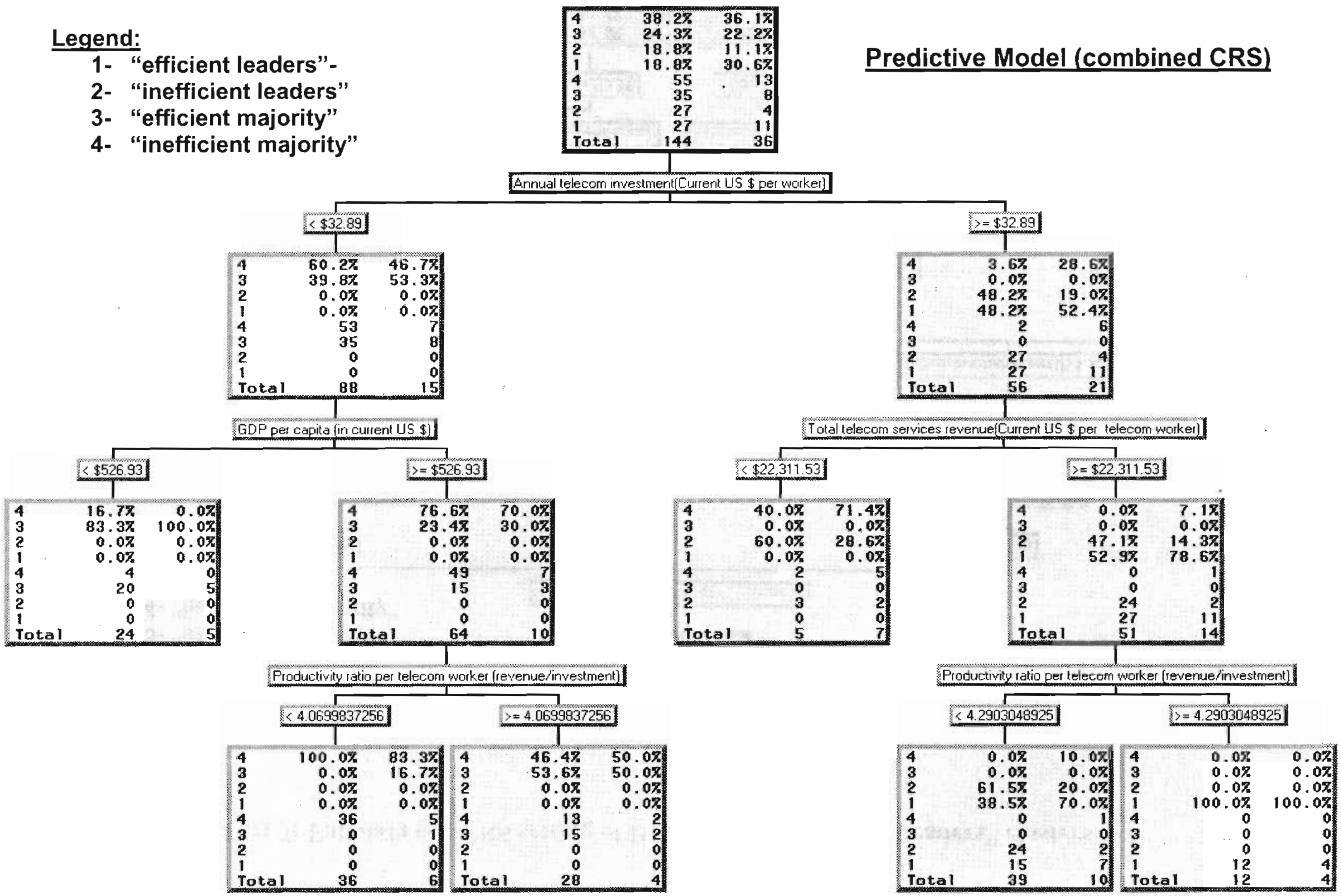
- "efficient leaders", those TEs that belong to the "leaders" cluster and were assigned the score of "1" by DEA
- "inefficient leaders" , those TEs that belong to the "leaders" cluster and were assigned the score of less than "1" by DEA
- "efficient majority", those TEs that belong to the "majority" cluster and were assigned the score of "1" by DEA
- "inefficient majority", those TEs that belong to the "majority" cluster and were assigned the score of less than "1" by DEA.

Next, we have created a new variable that we named "ID" and assigned the values of "1" to the "efficient leaders," "2" to the "inefficient leaders," "3" to the "efficient majority" and "4" to the "inefficient majority." After that we have re-ran the DT analysis, results of which are presented next.

Part 3: Full data set, CRS setting of DEA, “Majority” and “Leaders” clusters



Part 3: Full data set, CRS setting of DEA, "Majority" and "Leaders" clusters



Part 3: Confusion matrix, Combined CRS model

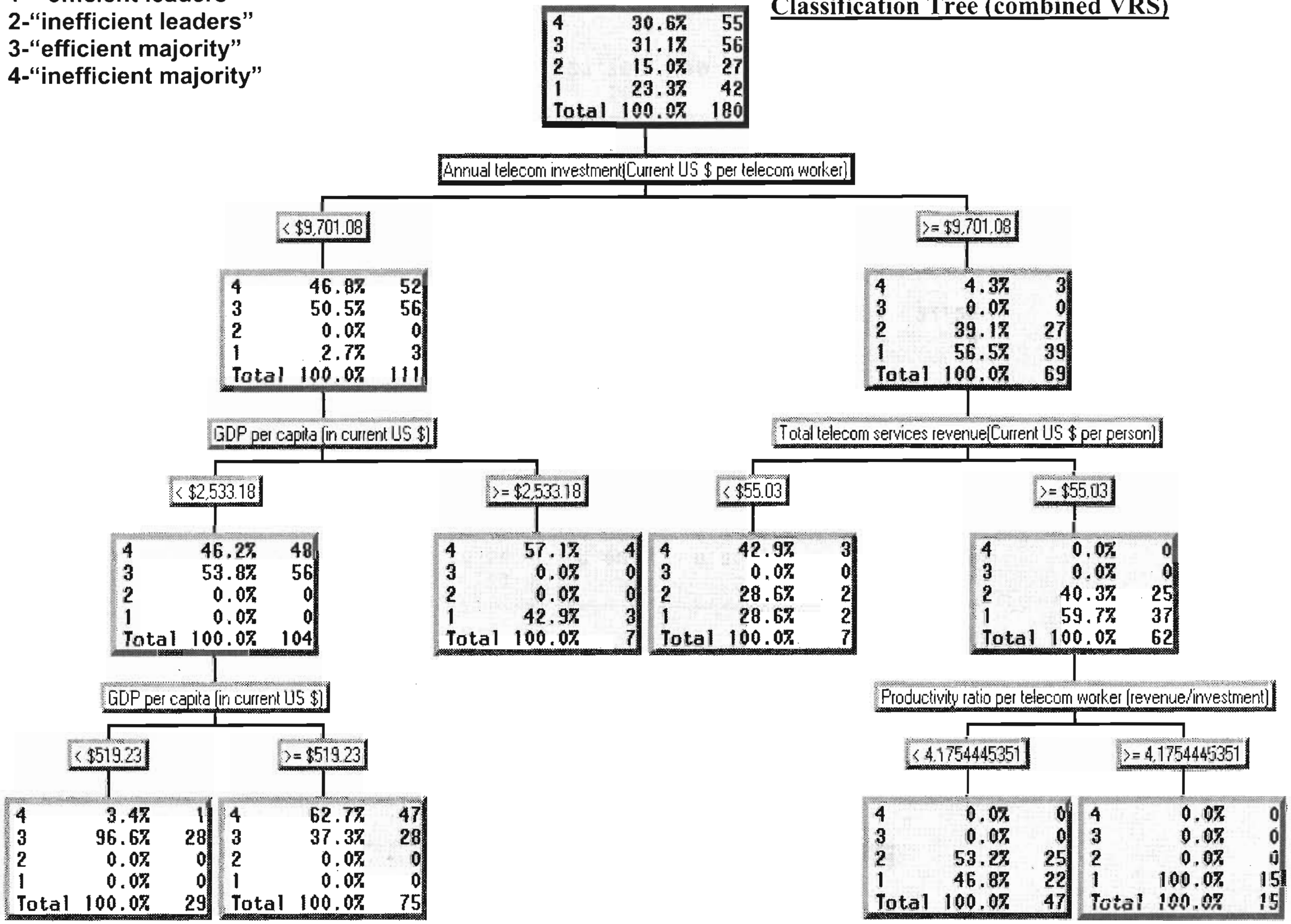
target output

Frequency / Percent / Row Pct / Col Pct	1	2	3	4	Total
1	11	0	0	0	11
	30.56	0.00	0.00	0.00	30.56
	100.00	0.00	0.00	0.00	
	100.00	0.00	0.00	0.00	
2	0	4	0	0	4
	0.00	11.11	0.00	0.00	11.11
	0.00	100.00	0.00	0.00	
	0.00	40.00	0.00	0.00	
3	0	0	7	1	8
	0.00	0.00	19.44	2.78	22.22
	0.00	0.00	87.50	12.50	
	0.00	0.00	100.00	12.50	
4	0	6	0	7	13
	0.00	16.67	0.00	19.44	36.11
	0.00	46.15	0.00	53.85	
	0.00	60.00	0.00	87.50	
Total	11	10	7	8	36
	30.56	27.78	19.44	22.22	100.00

Legend:
1- "efficient leaders"
2-"inefficient leaders"
3-"efficient majority"
4-"inefficient majority"

Part 3 Full data set, VRS setting of DEA, "Majority" and "Leaders" clusters

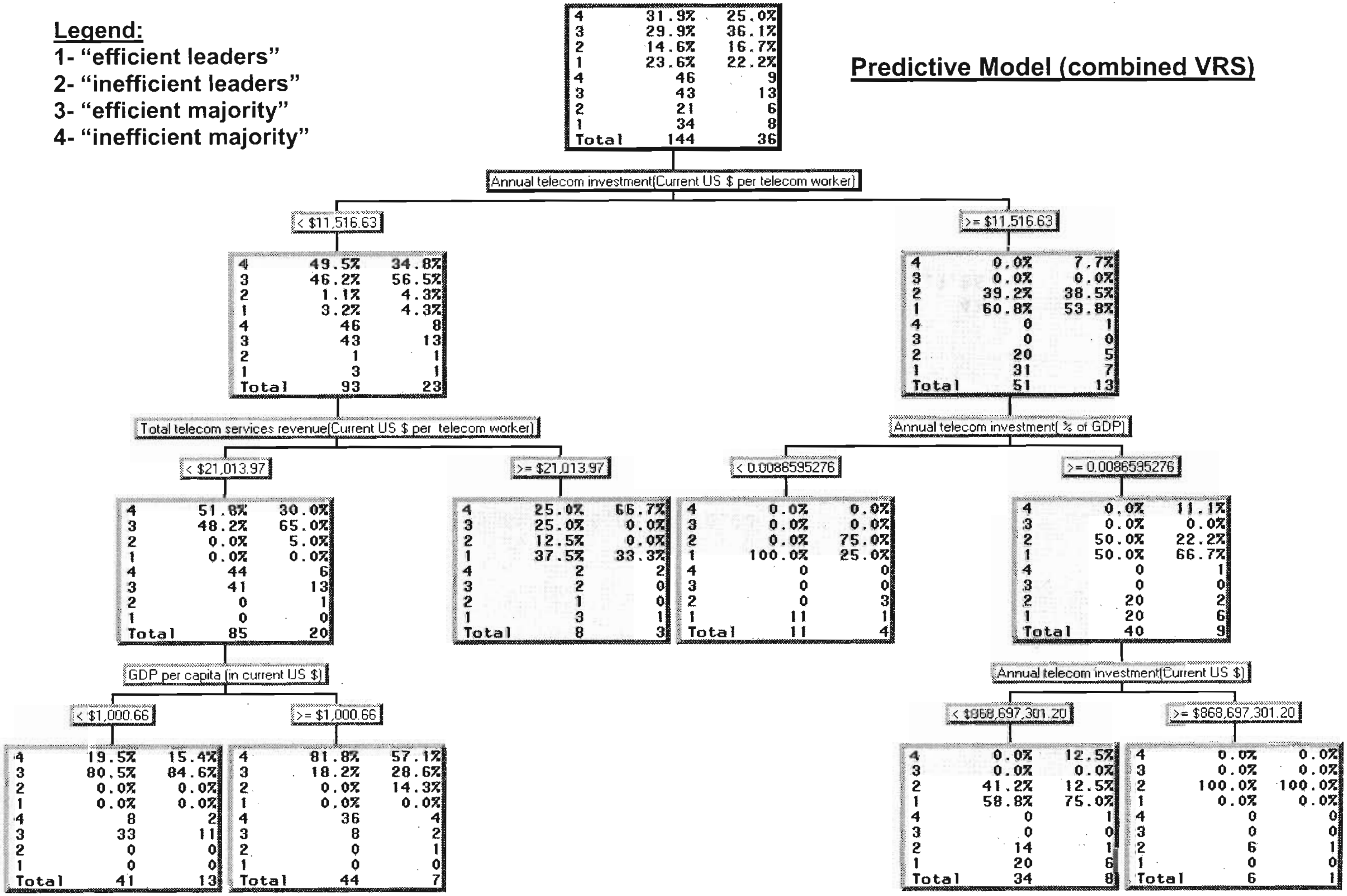
Classification Tree (combined VRS)



Part 3: Full data set, VRS setting of DEA, "Majority" and "Leaders" clusters

Legend:
1- "efficient leaders"
2- "inefficient leaders"
3- "efficient majority"
4- "inefficient majority"

Predictive Model (combined VRS)



Part 3: Confusion matrix, Combined VRS model

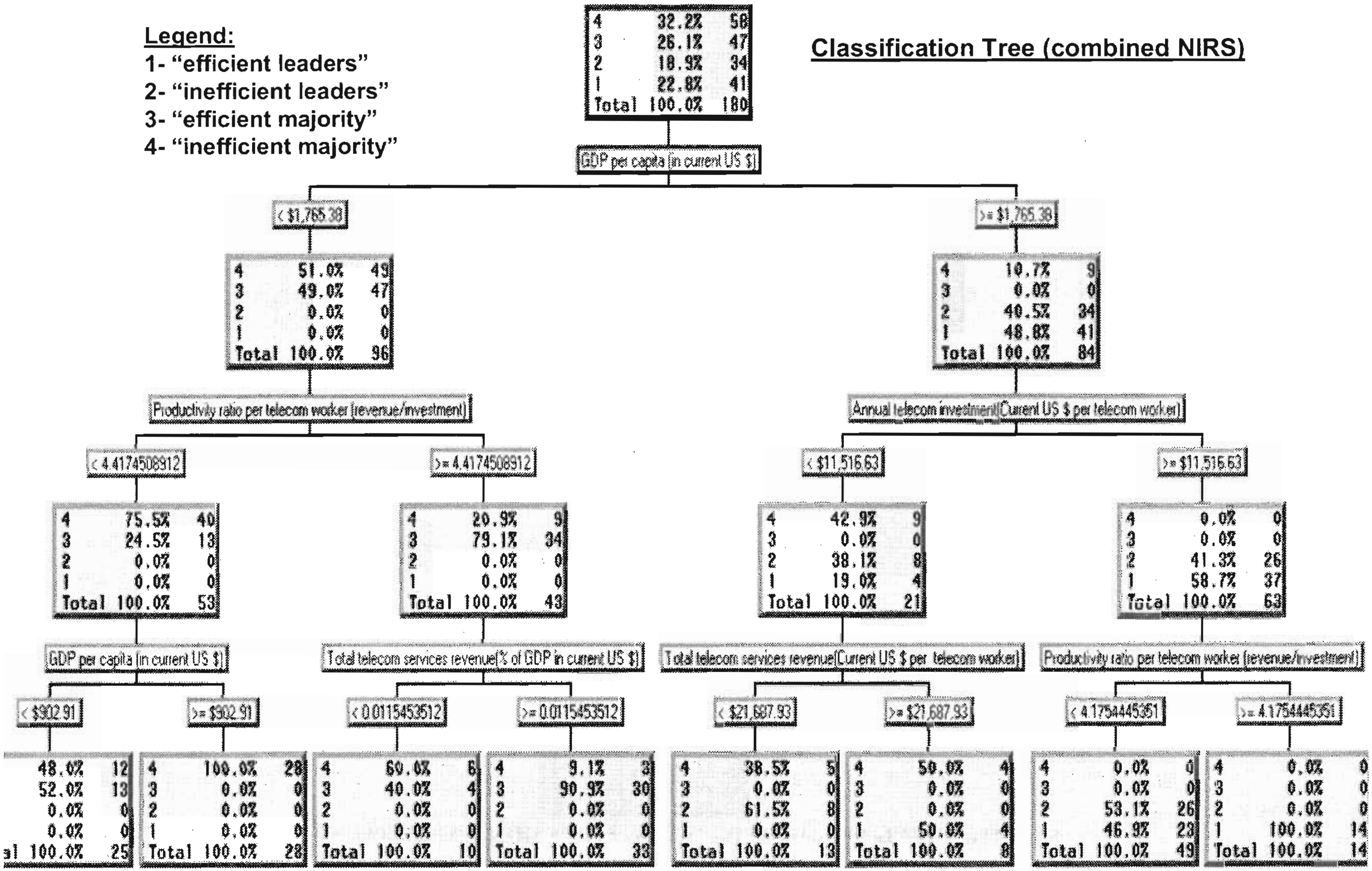
Table of target by output

target	Frequency / Percent / Row Pct / Col Pct	output 1	2	3	4	Total
1	Frequency	34	0	0	0	34
	Percent	23.61	0.00	0.00	0.00	23.61
	Row Pct	100.00	0.00	0.00	0.00	
	Col Pct	64.15	0.00	0.00	0.00	
2	Frequency	15	6	0	0	21
	Percent	10.42	4.17	0.00	0.00	14.58
	Row Pct	71.43	28.57	0.00	0.00	
	Col Pct	28.30	100.00	0.00	0.00	
3	Frequency	2	0	41	0	43
	Percent	1.39	0.00	28.47	0.00	29.86
	Row Pct	4.65	0.00	95.35	0.00	
	Col Pct	3.77	0.00	78.85	0.00	
4	Frequency	2	0	11	33	46
	Percent	1.39	0.00	7.64	22.92	31.94
	Row Pct	4.35	0.00	23.91	71.74	
	Col Pct	3.77	0.00	21.15	100.00	
Total	Frequency	53	6	52	33	144
	Percent	36.81	4.17	36.11	22.92	100.00

Part 3: Full data set, NIRS setting of DEA, "Majority" and "Leaders" clusters

<u>Legend:</u>
1- "efficient leaders"
2- "inefficient leaders"
3- "efficient majority"
4- "inefficient majority"

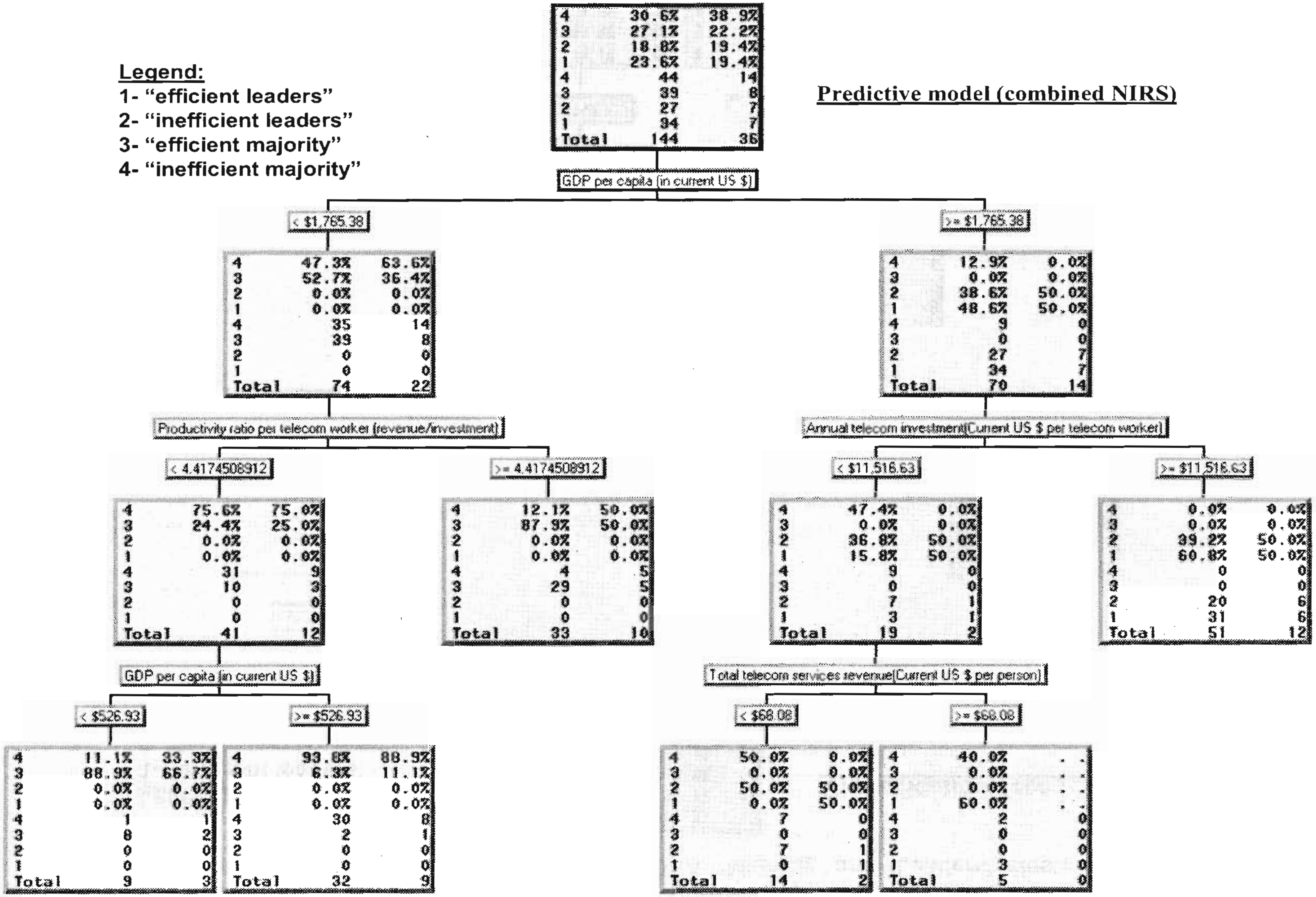
<u>Classification Tree (combined NIRS)</u>



Part 3: Full data set, NIRS setting of DEA, "Majority" and "Leaders" clusters

<u>**Legend:**</u>
1- "efficient leaders"
2- "inefficient leaders"
3- "efficient majority"
4- "inefficient majority"

<u>**Predictive model (combined NIRS)**</u>



Part 3: Confusion matrix, Combined NIRS model

Table of target by output

target / output Frequency Percent Row Pct Col Pct	1	3	4	Total
1	34 23.61 100.00 48.57	0 0.00 0.00 0.00	0 0.00 0.00 0.00	34 23.61
2	27 18.75 100.00 38.57	0 0.00 0.00 0.00	0 0.00 0.00 0.00	27 18.75
3	0 0.00 0.00 0.00	37 25.69 94.87 88.10	2 1.39 5.13 6.25	39 27.08
4	9 6.25 20.45 12.86	5 3.47 11.36 11.90	30 20.83 68.18 93.75	44 30.56
Total	70	42	32	144

Summary of the Part 3

We consider the results provide by the last DT model, developed in the Part 3, to be the most satisfactory out of the three that we have used to conduct the analysis of our data in this section. This model offers a much more clear insight into the specific aspects characterizing the each of the four groups that we have identified. Following is some important excerpts from the "English rules" describing each DT, which we obtained from the "Reporter" node.

CRS model:

- "inefficient majority"

IF Annual telecom investment (Current US $) < $19,070,765
AND 0.0031414015 <= Full-time telecommunication staff (% of total labor force)
AND 3.8043909395 <= Productivity ratio per telecom worker (revenue/investment)
AND $519 <= GDP per capita (in current US $)
AND Annual telecom investment (Current US $ per worker) < $33
THEN
NODE : 22
N : 7
4 : 100.0% (inefficient majority)
3 : 0.0%
2 : 0.0%
1 : 0.0%

IF Productivity ratio per telecom worker (revenue/investment) < 3.8043909395
AND $519 <= GDP per capita (in current US $)

AND Annual telecom investment (Current US $ per worker) < $33

THEN

NODE : 10

N : 39

4 : 100.0% (inefficient majority)

3 : 0.0%

2 : 0.0%

1 : 0.0%

IF Productivity ratio per telecom worker (revenue/investment) < 2.002357802

AND $33 <= Annual telecom investment (Current US $ per worker) < $58

THEN

NODE : 12

N : 6

4 : 100.0% (inefficient majority)

3 : 0.0%

2 : 0.0%

1 : 0.0%

- "efficient majority"

IF Full-time telecommunication staff (% of total labor force) < 0.0031414015

AND 3.8043909395 <= Productivity ratio per telecom worker (revenue/investment)

AND $519 <= GDP per capita (in current US $)

AND Annual telecom investment (Current US $ per worker) < $33

THEN

NODE : 16

N : 8

4 : 0.0%

3 : 100.0% (efficient majority)

2 : 0.0%

1 : 0.0%

IF 0.0118204323 <= Total telecom services revenue(% of GDP in current US$)

AND GDP per capita (in current US $) < $519

AND Annual telecom investment (Current US $ per worker) < $33

THEN

NODE : 9

N : 22

4 : 0.0%

3 : 100.0% (efficient majority)

2 : 0.0%

1 : 0.0%

- "inefficient leaders"

IF $836,899,003 <= Annual telecom investment (Current US $)

AND 0.0039016912 <= Full-time telecommunication staff (% of total labor force)

AND Productivity ratio per telecom worker (revenue/investment) < 4.1754445351

AND $58 <= Annual telecom investment (Current US $ per worker)

THEN

NODE : 25

N : 9

4 : 0.0%

3 : 0.0%

2 : 100.0% (inefficient leaders)

1 : 0.0%

IF Full-time telecommunication staff (% of total labor force) < 0.0039016912

AND Productivity ratio per telecom worker (revenue/investment) < 4.1754445351

AND $58 <= Annual telecom investment(Current US $ per worker)
THEN
NODE : 20
N : 11
4 : 0.0%
3 : 0.0%
2 : 100.0% (inefficient leaders)
1 : 0.0%

- "efficient leaders"

IF 1.5674014075 <= Productivity ratio per telecom worker (revenue/investment) < 4.1754445351
AND Annual telecom investment (Current US $) < $836,899,003
AND 0.0039016912 <= Full-time telecommunication staff (% of total labor force)
AND $58 <= Annual telecom investment (Current US $ per worker)
THEN
NODE : 29
N : 17
4 : 0.0%
3 : 0.0%
2 : 5.9%
1 : 94.1% (efficient leaders)

IF 4.1754445351 <= Productivity ratio per telecom worker (revenue/investment)
AND $58 <= Annual telecom investment (Current US $ per worker)
THEN
NODE : 15
N : 14
4 : 0.0%

3 : 0.0%

2 : 0.0%

1 : 100.0 %(efficient leaders)

VRS model:

- “inefficient majority”

IF Productivity ratio per telecom worker (revenue/investment) < 4.5834993974

AND $1,001 <= GDP per capita (in current US $)

AND Annual telecom investment (% of GDP) < 0.0206545843

AND Total telecom services revenue (Current US $ per person) < $58

THEN

NODE : 14

N : 40

4 : 97.5% (inefficient majority)

3 : 0.0%

2 : 2.5%

1 : 0.0%

- “efficient majority”

IF Full-time telecommunication staff (% of total labor force) < 0.0032180894

AND $423 <= GDP per capita (in current US $)< $1,001

AND Annual telecom investment (% of GDP) < 0.0206545843

AND Total telecom services revenue (Current US $ per person) < $58

THEN

NODE : 18

N : 10

4 : 0.0%

3 : 100.0% (efficient majority)

2 : 0.0%

1 : 0.0%

IF $24 <= Total telecom services revenue (Current US$ per person) < $58

AND 4.5834993974 <= Productivity ratio per telecom worker (revenue/investment)

AND $1,001 <= GDP per capita (in current US $)

AND Annual telecom investment (% of GDP) < 0.0206545843

THEN

NODE : 23

N : 8

4 : 0.0%

3 : 100.0% (efficient majority)

2 : 0.0%

1 : 0.0%

- "inefficient leaders"

IF Full-time telecommunication staff (% of total labor force) < 0.0039016912

AND Annual telecom investment (Current US $) < $956,450,275

AND Productivity ratio per telecom worker (revenue/investment) < 4.1754445351

AND $25 <= Annual telecom investment (Current US $ per person)

AND $58 <= Total telecom services revenue (Current US $ per person)

THEN

NODE : 24

N : 13

4 : 0.0%

3 : 0.0%

2 : 92.3% (inefficient leaders)

1 : 7.7%

IF $956,450,275 <= Annual telecom investment(Current US $)

AND Productivity ratio per telecom worker (revenue/investment) < 4.1754445351

AND $25 <= Annual telecom investment(Current US $ per person)

AND $58 <= Total telecom services revenue(Current US $ per person)

THEN

NODE : 17

N : 8

4 : 0.0%

3 : 0.0%

2 : 100.0 %(inefficient leaders)

1 : 0.0%

- <u>"efficient leaders"</u>

IF 4.1754445351 <= Productivity ratio per telecom worker (revenue/investment)

AND $25 <= Annual telecom investment (Current US $ per person)

AND $58 <= Total telecom services revenue (Current US $ per person)

THEN

NODE : 11

N : 15

4 : 0.0%

3 : 0.0%

2 : 0.0%

1 : 100.0% (efficient leaders)

NIRS Model:

- "inefficient majority"

IF \$903 <= GDP per capita (in current US \$)< \$1,734

AND Productivity ratio per telecom worker (revenue/investment) < 4.4174508912

THEN

NODE : 9

N : 28

4 : 100.0% (inefficient majority)

3 : 0.0%

2 : 0.0%

1 : 0.0%

IF Full-time telecommunication staff (% of total labor force) < 0.003363351

AND GDP per capita (in current US \$) < \$903

AND Productivity ratio per telecom worker (revenue/investment) < 4.4174508912

THEN

NODE : 16

N : 6

4 : 100.0% (inefficient majority)

3 : 0.0%

2 : 0.0%

1 : 0.0%

IF 4.4174508912 <= Productivity ratio per telecom worker (revenue/investment) < 11.214285714

AND Total telecom services revenue (% of GDP in current US \$) < 0.0115453512

AND GDP per capita (in current US \$) < \$1,734

THEN

NODE : 18

N : 5

4 : 100.0% (inefficient majority)

3 : 0.0%

2 : 0.0%

1 : 0.0%

- "efficient majority"

IF 0.0115453512 <= Total telecom services revenue(% of GDP in current US$)

AND 4.4174508912 <= Productivity ratio per telecom worker (revenue/investment)

AND GDP per capita (in current US $) < $1,734

THEN

NODE : 11

N : 33

4 : 9.1%

3 : 90.9 %(efficient majority)

2 : 0.0%

1 : 0.0%

IF GDP per capita (in current US $) < $548

AND 0.003363351 <= Full-time telecommunication staff (% of total labor force)

AND Productivity ratio per telecom worker (revenue/investment) < 4.4174508912

THEN

NODE : 26

N : 11

4 : 9.1%

3 : 90.9 %(efficient majority)

2 : 0.0%

1 : 0.0%

- "inefficient leaders"

IF Full-time telecommunication staff (% of total labor force) < 0.0039016912
AND Productivity ratio per telecom worker (revenue/investment) < 4.1754445351
AND $11,517 <= Annual telecom investment (Current US $ per telecom worker)
AND $1,734 <= GDP per capita (in current US $)
THEN
NODE : 24
N : 13
4 : 0.0%
3 : 0.0%
2 : 100.0 %(inefficient leaders)
1 : 0.0%

IF 0.0051205112 <= Full-time telecommunication staff (% of total labor force)
AND Total telecom services revenue (Current US $ per telecom worker)
< $21,688
AND Annual telecom investment (Current US $ per telecom worker) < $11,517
AND $1,734 <= GDP per capita (in current US $)
THEN
NODE : 23
N : 7
4 : 14.3%
3 : 0.0%
2 : 85.7% (inefficient leaders)
1 : 0.0%

- "efficient leaders"

IF 4.1754445351 <= Productivity ratio per telecom worker (revenue/investment)
AND \$11,517 <= Annual telecom investment (Current US \$ per telecom worker)
AND \$1,734 <= GDP per capita (in current US \$)
THEN
NODE : 15
N : 14
4 : 0.0%
3 : 0.0%
2 : 0.0%
1 : 100.0 %(efficient leaders)

IF Annual telecom investment (Current US \$) < \$868,697,301
AND 0.0039016912 <= Full-time telecommunication staff (% of total labor force)
AND Productivity ratio per telecom worker (revenue/investment) < 4.1754445351
AND \$11,517 <= Annual telecom investment (Current US \$ per telecom worker)
AND \$1,734 <= GDP per capita (in current US \$)
THEN
NODE : 30
N : 27
4 : 0.0%
3 : 0.0%
2 : 14.8%
1 : 85.2 %(efficient leaders)

7.4.3 DT: Summary of the Analysis

In this part of the paper, we described the process of obtaining the DT models that allow us to inquire into the differences between the efficient and the inefficient TEs. The examples provided in the last section suggest that we were able to obtain such models.

The complete and detailed analysis of the differences between the efficient and inefficient TEs is beyond the scope of this study. Instead, in this section we were pursuing a goal of demonstrating, first, that the DT model could be applied to our data in order to describe the differences between efficient and inefficient TEs, and, second, that such DT model could do so with enough precision to be useful.

One of the major assumptions of the DEA is that all DMUs in the data set are functionally similar. However, in reality, by obvious reasons, it is very difficult condition to obtain. The approach that we have taken resulted in the DT models that allow for classification and prediction of the relative efficiency of the functionally dissimilar DMUs, thus allowing to compare, side-by-side, the TEs that otherwise should not be compared to each other.

Consequently, the resultant DT models are able to provide a policy maker with a suggestion regarding the differences between the efficient and inefficient TEs of the same type (i.e., the "majority"). Moreover, these models could also provide the insights regarding how TEs of one group could differ from the TEs of another group (i.e., "ineffective" majority from "ineffective" leaders). Furthermore, it is also possible to use these DT models in order to gain some insights in the case of the 'migrant' TEs. For example, it is possible to find out how TE of the "effective majority" group could transition into the group of the "ineffective leaders."

The purity of the nodes of the resultant DT models suggests that we were able to accomplish the outlined goal. Consequently, the provided by the DT model English rules, describing properties of each of the four groups of the TEs, are descriptive and precise enough to provide some utility to a policy maker.

7.5 Translog: Testing for Interaction

In this section we describe our inquiry into the presence of the complementary relationships between the investments in ICT and other endogenous (i.e., labor and the other types of investments) variables. To test for the presence of the interactions we have constructed four data sets, three for the 5-year period from 1998 to 2002, and one for the 10-year period from 1993 to 2002.

We ended up constructing three 5-year data sets because the data for some countries were not available for the variables of interests. Thus, we ended up testing for the presence of interactions by analyzing various subsets of the original 18-countries set of the data.

We describe the four data sets in more details next.

7.5.1 Translog: Description of the data sets

Our first 5-year data set consisted of following thirteen TEs: Armenia, Azerbaijan, Belarus, Bulgaria, Hungary, Kyrgyz Republic, Latvia, Lithuania, Poland, Romania, Slovak Republic, Slovenia, and Ukraine.

This data set was constructed using the following variables: Military personnel (% of total labor force), Researchers in R&D (% of total labor force), Technicians in R&D (% of total labor force) , Full-time telecommunication staff(% of total labor force), Military expenditure (% of GDP in current US $), Health expenditure, private (% of GDP in current US $), Health expenditure, public (% of GDP in current US $), Research and development expenditure (% of GDP in current US $), International tourism expenditures (% of GDP in current US $), Annual telecom investment(% of GDP in current US $), Foreign direct investment, net inflows (% of GDP), and GDP (current US$).

The second 5-year data set consisted of the following 14 TEs: Armenia, Azerbaijan, Belarus, Bulgaria, Czech Republic, Hungary, Kyrgyz Republic, Latvia, Lithuania, Poland, Romania, Slovak Republic, Slovenia, and Ukraine.

The second data set was constructed using the following variables: Military personnel (% of total labor force), Researchers in R&D (% of total labor force), Technicians in R&D (% of total labor force) , Full-time telecommunication staff(% of total labor force), Military expenditure (% of GDP in current US $), Health expenditure, private (% of GDP in current US $), Health expenditure, public (% of GDP in current US $), Research and development expenditure (% of GDP in current US $), and GDP (current US$). This data set was constructed by dropping off the variable 'International tourism, expenditures (% of GDP in current US $)', which allowed us to include Czech Republic into the data set.

The third 5-year data set consisted of the following 15 TEs: Armenia, Azerbaijan, Belarus, Bulgaria, Estonia, Hungary, Kyrgyz Republic, Latvia, Lithuania, Moldova, Poland, Romania, Slovak Republic, Slovenia, and Ukraine.

We constructed this data set using the following variables: Military personnel (% of total labor force), Researchers in R&D (% of total labor force), Technicians in R&D (% of total labor force) , Full-time telecommunication staff(% of total labor force), Military expenditure (% of GDP in current US $), Health expenditure, private (% of GDP in current US $), Health expenditure, public (% of GDP in current US $), International tourism expenditures (% of GDP in current US $), Foreign direct investment, net inflows (% of GDP), and GDP (current US$). This data set was obtained by dropping off the variable 'Research and development expenditure (% of GDP in current US $)', which allowed us to include Moldova into the data set.

Finally, the 10-year data set consisted of following 16 TEs: Albania, Armenia, Azerbaijan, Belarus, Bulgaria, Estonia, Hungary, Kazakhstan, Latvia, Lithuania, Moldova, Poland, Romania, Slovak Republic, Slovenia, and Ukraine.

This data set was constructed using following variables: Military personnel (% of total labor force), Physicians (% of total labor force), Unemployment, total (% of total labor force), Full-time telecommunication staff(% of total labor force), Military expenditure (% of GDP in Current US $), International tourism, expenditures (% of GDP in Current US $), Annual telecom investment(% of GDP in Current US $), Foreign direct investment, net inflows (% of GDP), and GDP (current US$).

In the next part of this section of the paper, we describe the method and the tool that we used to perform the Translog regression analysis.

7.5.2 Translog: Description of the method

We used SAS Enterprise Miner to perform our data analysis. The general approach that we have followed is depicted in the diagram below. In our case we had four data sets, thus, four different diagrams were created. Next, we are going to describe in details all the relevant steps that have been taken at each node of the diagram.

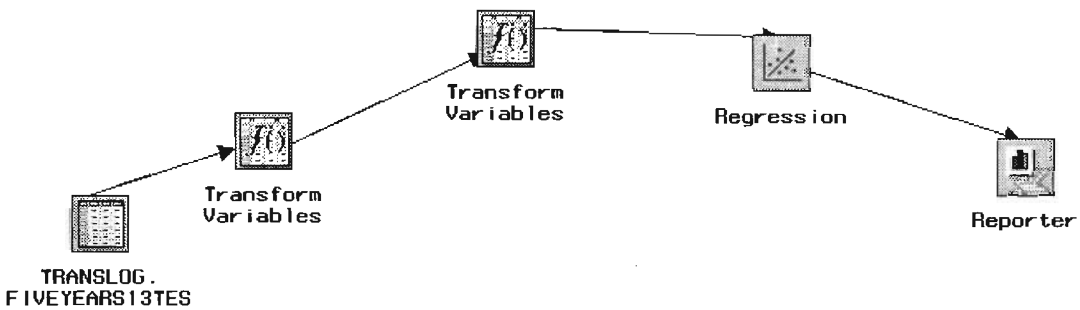


Figure 17 Translog: implementation in SAS

In the first step of our analysis, we used the Data Source node primarily for two purposes, first, we have selected the data set that we were going to use for each case, and second, we defined "GDP" variable as a 'target' of the data analysis. However, the original values contained in the data set could not be used in Translog function, since Translog requires logs, squared logs, and log interaction terms. The transformation of the data was performed in the next step of our analysis.

Second step involved the use of two Transform Variables nodes. In the first Transform Variables node, we have used Log-transformation to produce logs of the original data values in the data set defined in the Data Source node. Thus, at the first Transform Variables node we ended up with the transformed original variables. However, we still needed squared logs and interaction terms.

In the second Transform Variables node, we used the option "Create New Variable" to create all of the required squared log terms. After it was done, we ended up with the logs of the original variables, plus, with the squared logs of the original variables. At this point, we still needed to define the interaction terms prior the data analysis.

The third step involved the use of the Regression node, where the interaction terms were added to the model and the actual model was actually run. To add the interaction term we used "Interaction Builder" to create all the required for our purposes interaction terms. Later in this part, we list all the interaction terms that were created for each of the four models. Thus, at this point we have all the required by Translog function variables. Finally, we have chosen "General Linear Model" as an input coding option of the regression model and then run the data analysis.

The Reporter node was used to obtain the results of the data analysis in a better formatted way than it was offered by the output of the Regression node. After four diagrams were constructed, we run the data analysis, the preliminary results of which are presented next.

7.5.3 Translog: Results of the Interaction Analysis

In this section, we present some of the relevant results that were obtained from the output provided by the Reporter node. The results are presented in the form of the four tables, one for each data set that was used to run a Translog regression analysis.

We present the interaction terms that were found to be significant, in the form as they appear in the each model (column "Interaction Term in the model"). We as well provide a description of the each term (column "Description of the Interaction Term").

Let us recall, that in this part of the data analysis we test the null hypothesis that coefficients β of each interaction term is equal to 0. The column "Estimate" provides the values of the parameter estimates for the β of each interaction terms. Thus, in the table below the interaction term represented by the variable "Log (Annual Telecom Investment)*Log (Researchers in R&D)" has value of $\beta = 5.4950$.

Finally, the last column (labeled "Pr > |t| at 95%") provides a 2-tailed P-value used in testing of our null hypothesis that $\beta = 0$. We test the hypothesis at 95% confidence level, or, as it is commonly expressed, at a level of $\alpha = 0.05$. As a result, all the coefficients having a p-value of 0.05 or less would be considered statistically significant, which would allow us to reject the null hypothesis of $\beta = 0$ at the level of $\alpha =$ 0.05.

We provide the summary of the relevant findings below, while the complete reports of the Translog regression analysis, obtained from the output of the Reporter node, are provided at the end of this section

Table 26 Translog: Interaction terms in 5-Year 13 TEs Data Set

Interaction Term in the model	Description of the Interaction Term	Estimate (value of β)	Pr > |t| at 95%
ANNU_RZY*RESE_MA1	Log (Annual Telecom Investment)*Log(Researchers in R&D)	5.4950	0.0116
ANNU_RZY*TECH_KDM	Log (Annual Telecom Investment)*Log(Technicians in R&D)	-3.8542	0.0302
ANNU_RZY*FULL_RZM	Log (Annual Telecom Investment)*Log(Full-time Telecom Staff)	-2.5693	0.0124
ANNU_RZY*HEAL_QHZ	Log (Annual Telecom Investment)*Log(Health Expenditure, Private)	-3.8643	0.0012
ANNU_RZY*INTE_IFE	Log (Annual Telecom Investment)*Log(International Tourism, Expenditures)	3.4912	0.0086

Table 27 Translog: Interaction terms in 5-Year 14 TEs Data Set

Interaction Term in the model	Description of the Interaction Term	Estimate (value of β)	Pr > |t| at 95%
ANNU_74V*MILI_BN1	Log (Annual Telecom Investment)*Log(Military	1.6918	0.0341

	Personnel)		
ANNU_74V*TECH_04C	Log (Annual Telecom Investment)*Log(Technicians in R&D)	-2.5606	0.0306
ANNU_74V*HEAL_VKR	Log (Annual Telecom Investment)*Log(Health Expenditure, Private)	-1.6527	0.0206

Table 28 Translog: Interaction terms in 5-Year 15 TEs Data Set

Interaction Term in the model	**Description of the Interaction Term**	**Estimate (value of β)**	**Pr > |t| at 95%**
ANNU_KW1*INTE_8V1	Log (Annual Telecom Investment)*Log(International Tourism, Expenditures)	-1.7474	0.0566
ANNU_KW1*TECH_07Y	Log (Annual Telecom Investment)*Log(Technicians in R&D)	2.1306	0.0395

Table 29 Translog: Interaction terms in 10-Year 16 TEs Data Set

Interaction Term in the model	**Description of the Interaction Term**	**Estimate (value of β)**	**Pr > |t| at 95%**
ANNU_EL1*PHYS_RNY	Log (Annual Telecom Investment)*Log(Physicians, % of total labor force)	-5.2154	0.0009
ANNU_EL1*FULL_2KW	Log (Annual Telecom Investment)*Log(Full-time Telecom Staff)	7.8231	<.0001

ANNU_EL1*INTE_1AH	Log (Annual Telecom Investment)*Log(International Tourism, Expenditures)	0.6283	0.0412
ANNU_EL1*FORE_X0E	Log (Annual Telecom Investment)*Log(Foreign Direct Investment)	0.8495	0.0122

7.5.4 Summary and the Interpretation of the Results

As a summary of our preliminary findings, we would like to state, that based on the result of the data analysis we have established the presence of the statistically significant interaction effect between the variable "Annual telecom investment (% of GDP in current US $)" and the following variables:

- Labor

- Researchers in R&D (% of total labor force),
- Technicians in R&D (% of total labor force) ,
- Full-time telecommunication staff(% of total labor force),
- Military personnel (% of total labor force),
- Physicians (% of total labor force)

- Investment:

- International tourism, expenditures (% of GDP in Current US $),
- Foreign direct investment, net inflows (% of GDP), and GDP (current US$)

- Health expenditure, private (% of GDP in current US $)

Thus, the null hypothesis of no interaction between the investments in ICT and other types of the investments (i.e., labor and capital) could be rejected. We provide some possible interpretations of the interactions in the table below. Of especial interest are two interaction terms: first, between Annual Telecom Investment and Technicians in R&D, and, second, between Annual Telecom Investment and Full-time Telecom Staff, because the estimate values of β for this terms change the signs (the first one from negative to positive in three 5-year data sets, and the second one from negative, in 5-year data set, to positive in 10-year data set). This change suggests that the different levels of these variables could produce different types of the effects.

Table 30 Translog: Interpretation of the interaction terms

Interactions	**Interpretation**
Log (Annual Telecom Investment) *Log(Researchers in R&D)	One of the possible interpretations of the effect of this interaction term is that the larger the number of the researchers in a given TE, the greater the impact of the Annual Telecom Investment on macroeconomic growth. Similarly, it could be interpreted that the increase in Annual Telecom Investment would result in the greater contribution of the researchers to the increase in GDP. **Implication**: Wealthier countries spend more on

	science and R&D, and have a higher level of investments in Telecom. The possible conclusion, in line with the existing research, is that certain level of capital accumulation (already achieved by certain TEs) is required for the investments in ICT to contribute noticeably to the macroeconomic growth. Most of the TEs have not obtained yet the required capital depth in the area of ICT.
Log (Annual Telecom Investment) *Log(Technicians in R&D)	The greater the Annual Telecom investment is, the greater the contribution of the Technicians in R&D would be to the economic growth. Similarly, the greater the number of Technicians in a given TE, the greater contribution of the Telecom investments to GDP. On the other hand, low level of investments in telecoms prevents Technicians in R&D from contributing highly to the economic growth. **Implication:** Number of technicians in R&D could possibly reflect the state of "physical and intellectual infrastructure" of an economy. Up to the certain level of wealth TEs could simply grow by having "more" of it, while after the certain point

	any growth would only occur through the increase in quality of the infrastructure- by having “better” rather than “more.”
Log (Annual Telecom Investment) *Log(Full-time Telecom Staff)	It is conceivable to interpret this effect in the following way: there is a certain ‘optimal percentage’ of Full-time Telecom staff that is required for Annual Telecom Investment to contribute maximally to the increase in GDP. If this is the case, then there exists a ‘golden middle’ of the investment-to-staff ratio. For example, it is possible that up to the certain level of the investment in Telecoms the number of the staff must grow rapidly, while after that point only increase in productivity of the Full-time Telecom staff could provide increased contribution of the investments in Telecoms to the macroeconomic growth. **Implication:** Same as above
Log (Annual Telecom Investment) *Log(Military Personnel)	This interaction term could be interpreted in the following way: the larger the number of the military personnel in a given TE, the greater the impact of the Annual Telecom Investment on

	macroeconomic growth (i.e., larger army requires greater levels of the investments in ICT). Similarly, it could be interpreted that the increase in Annual Telecom Investment would result in the greater contribution of the military personnel to the increase in GDP(possibly, by allowing a smaller, more efficient armed forces). **Implication:** Larger the army, more workforce is lost for macroeconomic growth, but larger armies could also be the larger consumers of ICT/Telecom, for such armies require larger ICT/Telecom infrastructure to support communication.
Log (Annual Telecom Investment) *Log(Physicians, % of total labor force)	The argument could be made that the less number of Physicians, as a % of total labor force, in a given TE, the less Annual Telecom investment contributes to the increase in GDP. Similarly, in those TEs that have lower levels of the investment in Telecom, Physicians contribute less to the macroeconomic growth of the given TE. **Implication**: Same situation as with "R&D Researchers" where both, # of Researchers and # of Physicians is representative of the level of wealth

	of a given TE. Wealthier countries spend more on science, R&D, and medicine and also have a higher level of investments in Telecom.
Log (Annual Telecom Investment) *Log(International Tourism, Expenditures)	We could interpret the effect of this interaction term in the following way: the greater the Expenditures on the International Tourism in a given TE, to the grater degree would be the impact of the Annual Telecom Investment on macroeconomic growth. Similarly, it could be interpreted that the increase in Annual Telecom Investment would result in the greater contribution of the Expenditures on the International Tourism to the increase in GDP. **Implication**: Same as above, Expenditures on International Tourism reflect the level of wealth of a country, plus, reflect the degree of "openness" of a given TE. In this case, the level of expenditures on International Tourism could be reflective of the level of integration of a country into the international community, as well as into the market economy of the world. Another possible implication is that such 'integrated' countries are

	capable of taking advantage of lower market prices on ICT/Telecom equipment.
Log (Annual Telecom Investment) *Log(Foreign Direct Investment)	It is possible, that the greater the level of FDI in a given TE, the more significant would be the contribution of the Annual Telecom Investment to the increase in GDP. The case could be made that TE which has a higher level of Annual Telecom investment would attract additional FDI, thus resulting in the increase in GDP. Alternatively, given the same level of FDI, contribution to the macroeconomic growth could be greater due to the more effective and efficient management of the investments. **Implication**: Same as above, both factors reflect the level of wealth of TE.
Log (Annual Telecom Investment) *Log(Health Expenditure, Private)	One of the possible interpretations of this interaction term is that lower the level of Private Health Expenditure in a given TE, the less Annual Telecom investment contributes to the increase in GDP. Similarly, in those TEs that have lower levels of the investment in Telecom, Private Health Expenditure contributes

	less (i.e., smaller) to the macroeconomic growth of the given TE. **Implication**: Same as above, both factors reflect the level of wealth of TE, where in wealthier TEs population has a higher level of disposable income that could be spend on private health care, as well as on the products and services of ICT/Telecom. This, of course, results in the increased contribution of Telecom investments to the macroeconomic growth of the country.

In the conclusion of this part of the paper, we would like to acknowledge, that despite the offered above interpretations of the interaction terms, such interpretations should not be taken as final and definitive. Jaccard, Turrisi and Wan (1990), Aiken and West (1991), and Braumoeller (2004) note, that there exists a number of difficulties in interpreting interaction effects between continuous variables. Thus, our interpretations of the interaction terms should be taken with a grain of salt, and considered to be suggestive, rather than indicative, in nature.

7.6 SEM with PLS: Testing the relationship between the Investments in ICT and TFP

In this part of the paper, we test the hypothesis of presence of the relationship between TFP and the investments in ICT, where the appropriate null hypothesis is formulated as follows:

H0: there exists no statistically significant relationship between TFP and the investments in ICT.

To test the null hypothesis we use Partial Least Squares (PLS) method (Wold 1989), implemented in PLS-Graph (Chin 1994) technique. But before we proceed with the description of the conceptual model, variables that we use and the results of the analysis, we would like to briefly outline the scope of this section.

7.6.1 SEM with PLS: Scope of the Inquiry

We propose our research model not with the purpose of describing, as close as possible, the process of how the investments in ICT contribute to the unexplained growth in the form of TFP. Rather, the purpose of our model is to serve as a vehicle of understanding whether or not the mentioned above process indeed takes place. Therefore, our research model is not intended to mimic, in the simplified form, the actual process, but rather its intent is to provide the insights on the basis of which a better model could be constructed.

According to Ackoff (1999) " the effectiveness of any model used to describe and understand behavior of a particular system as a whole ultimately depends on the degree to which that model accurately represents that system" (p.34). We contend that before a model that 'accurately represents' the actual system could be build there is a required step of validating that the actual system, in fact, exists. In our case, that system in need of validation would be the contribution of the investments in ICT to TFP.

Thus, this part of the paper intends to offer some proof that the actual system within which investments in ICT contribute to TFP does, in fact, exists and therefore, could be modeled. As a result, the conceptual research model that we use in this part of the paper would be better served to be called a "prototype of the model." Nevertheless, our prototype, just as any model, would still need to be validated.

The process of model validation could be perceived as a following the scientific method cyclic process that consists of six phases (Ackoff, 1999). These phases are: first, formulation of the problem, second, construction of the model, third, testing of the model, fourth, derivation of the solution from the tested model, fifth, testing and controlling the solution, and, finally, implementation of the solution.

In this part of the paper we intent to accomplish the first three phases of the validation process. Namely, we formulate the problem by stating the null hypothesis, and then we put forward our conceptual research model by means of which we intend to test the hypothesis. Finally, we would conclude this section with the evaluation of the results of the testing, by means of accepting or rejecting the null hypothesis. The other three stages of the validation process are beyond the scope of this paper.

Despite putting forward our research model to test hypothesized relationship, we are, fundamentally, at this point interested in not much else but testing only the presence of the relationship between the investments in ICT and TFP. Meaning, within the scope of this paper we are not interested in answering the questions of why, what are the requisite, or pre-requisite conditions that are necessary for the relationship to exist. Having acknowledged that, we now proceed to introduce our conceptual research model intended to test the null hypothesis stated in the beginning of this part of the paper.

7.6.2 SEM with PLS: Conceptual model

The conceptual model that we have chosen to test the hypothesized relationship consists of the six unidimensional constructs. The first construct is "TFP," which represents the unexplained economic growth. The second construct is "ICT Capitalization," role of which to reflect the general state of the investments in ICT in a given TE. The third construct is "ICT Utilization"; this construct is thought to represent the effectiveness and efficiency of the utilization of the available to a given TE ICT-related infrastructure. We named the construct reflecting the state of ICT infrastructure "ICT Diffusion." The purpose of the fifth construct in the model, "Militarization," is to reflect the state of the military, from the perspective of the military spending and military personnel, in a given TE. Finally, the purpose of the sixth construct, "Health of Economy," is to represent the economic environment of TE.

We offer following definitions of the relevant to this study constructs. First, we define a construct "ICT Capitalization" as a "measurable/quantifiable fiscal attitude of a TE towards the labor- and capital investments in ICT." Second, we offer a definition of a construct "ICT Utilization" as an "extent of the ability and capability of the users to take advantage of the available to them ICT-related resources." Finally, "ICT Diffusion" is defined as a "level of pervasiveness/ubiquity of the socio-technical ICT-related infrastructure in a given TE."

The overall relationships between the listed above constructs are depicted in the figure below.

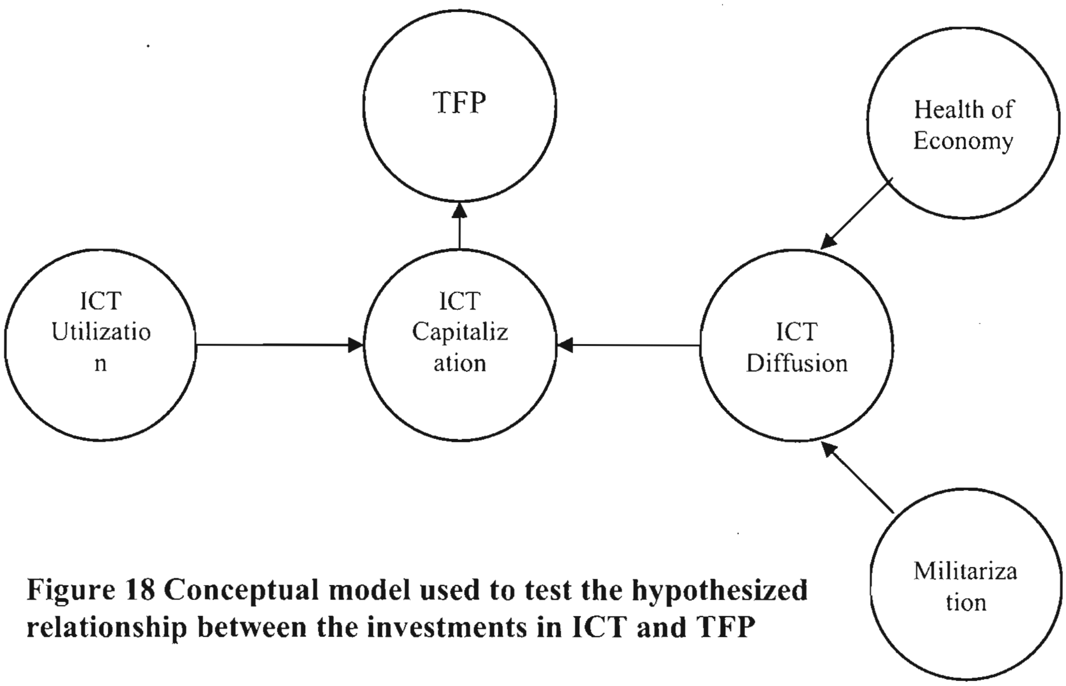


Figure 18 Conceptual model used to test the hypothesized relationship between the investments in ICT and TFP

According to the model depicted above, we hypothesize that there exist direct relationships between "ICT Capitalization" and "ICT Utilization," and between "ICT Capitalization" and "ICT Diffusion." We also assume the existence of the direct relationships between "ICT Diffusion" and "Health of Economy" and "Militarization." But mostly, of course, we are interested in the relationship between "TFP" and "ICT Capitalization," for testing the statistical significance of this relationship is central to our research. We describe the variables representing the each of the constructs next.

7.6.3 SEM with PLS: Measures of the Constructs

The purpose of the six constructs that we use is to reflect the general state of the ICT-related structure of the economy of TEs. Therefore, we needed to represent all 18 TEs in such way that differences in geographical size, population, economic wealth and so on would be countered. Thus, we avoided using such measures as "Annual Telecom Investment", "Full-time Telecom Staff" and the like because these types of variables are relative in the sense that they only could be used to compare similar in size, wealth, and population countries.

Therefore, we ended up using only those indicators from the "Yearbook of IT" and WDI Database, which were represented as ratios or percentages. In addition, we have created a set of what could be called as "meta-indicators," which are the measures that were obtained using original variables and represented as ratios.

We describe the variables that we use as the measures to represent each of the constructs next.

TFP – this construct is represented by the values obtained from the results of the Data Envelopment Analysis that we have performed earlier. There are altogether three available measures that we could use to represent this construct. The first is "MI," or Malmquist Index, which represents the change in TFP from one time period to another. Let us recall, that the MI is depicted in the following way: "1" for no change, less than "1" for decrease, and greater than "1" for increase in productivity over the time period. The DEA offers decomposition of MI into "TC" and "EC," where TC stands for change in technology and EC for the change in efficiency. Thus, the second and third measure that could be used to represent TFP is TC and EC.

Health of Economy- we use two measures that were provided by WDI database, "Exports of goods and services (% of GDP)" and "Trade (% of GDP)". The source of the data offers following definitions of these measures:

> Exports of goods and services (% of GDP): Exports of goods and services represent the value of all goods and other market services provided to the rest of the world. They include the value of merchandise, freight, insurance, transport, travel, royalties, license fees, and other services, such as communication, construction, financial, information, business, personal, and government services. They exclude labor and property income (formerly called factor services) as well as transfer payments.
>
> Trade (% of GDP): Trade is the sum of exports and imports of goods and services measured as a share of gross domestic product.

Militarization- we have created two ratio variables to represent this construct. The first variable is "ratioMilExp/MilPers", which represent the military expenditure of the TE (in current US %) per "military person". This ratio was obtained by dividing total military expenditure of a country, represented in current US $, by the total size of its military force. The second ratio is "MilExpPerCapita", which represent the military expenditure of the TE (in current US %) per capita. This ratio was obtained by dividing total military expenditure of a country, represented in current US $, by the total population of the country.

ICT Diffusion – this construct is represented by the three indicators obtained from WDI Database. These indicators are: "TelephoneMainlines" (Telephone Mainlines per 1000 people), "MobilePhones"(Mobile Phones per 1000 people), and "FixedMobilePhoneSubscr" (Fixed Line and Mobile Phone Subscribers per 1000 people).

ICT Utilization – this construct is represented by the following two measures. The first one is "InterTelecomOutgoing" obtained from WDI Database where it was labeled as "International telecom, outgoing traffic in minutes per subscriber". The second measure is "revenuePerSubscriber", which represents the ratio of the total telecom revenue to the total number of fixed and mobile lines subscribers. In our view the first measure depicts utilization of the available infrastructure, while the second one the extent of utilization in dollar term.

ICT Capitalization – in our model this construct is represented by three measures. The first one is "ratioGDPInvestPerCapita", which is the ratio of GDP per capita to the annual telecom investment per capita. The values for "GDP per capita" provided by WDI Database, and values for the Annual Investment in Telecoms provided by the Yearbook of IT. The second measure is "ProductivityRatio" which is a ratio of Total telecom revenue to annual investments in telecom per year. Both nominator and denominator of the ratio were obtained from the Yearbook of IT(both values are represented as percentages of GDP). Finally, the third measure is "ratioFtTW/ICTinvest", which is the ratio of Full-time telecom stuff to the Annual investment in telecoms (in current US$).

The purpose of these three measures is to provide a general picture depicting the investments in ICT in relation to GDP and total population(first measure), revenues from ICT (second measure) and the full-time telecom staff. All the measures that we use to represent the construct "ICT Capitalization" are ratios, and they all have Annual Investment in telecom in denominator of the ratio, in one form or another (per capita, in current US$, or as a % of GDP). Therefore, we assume that if we would be able to demonstrate the presence of statistically significant relationship between "ICT Capitalization" and "TFP", we would be able to claim the presence of the indirect relationship between the investments in ICT and TFP. However, this could only be done if we would demonstrate that the measures that we use to represent our constructs are valid and reliable. First, we conducted exploratory Principal Component Analysis (PCA) in order to see that the measures that we have chosen to represent our six constructs would demonstrate a specific pattern of loadings. The results of PCA analysis are presented next.

7.6.4 SEM with PLS: PCA Analysis

The purpose of the conducting the PCA analysis was to see whether the measures that were intended to represent a construct would align in the same direction and load together on the same principal component. At this point we already knew the number of the constructs that we were going to use in our research model, thus, we requested five components to be extracted.

We also requested the results of the Kaiser-Meyer-Olkin (KMO) test of sampling adequacy and Bartlett's test of sphericity to be included in the output, for these two measures are commonly used to determine whether or not a data set could be successfully analyzed using factor analysis. Part of the SPSS output, containing these two measures, is presented the table below.

Table of the results of the KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.661
Bartlett's Test of Sphericity	Approx. Chi-Square	3116.174
	df	66
	Sig.	.000

In order for our data set to pass these two tests, KMO value must be above 0.5, and Bartletts' test value must be less than 0.05. Based on the provided above output we could conclude that our data set passed the two tests and is suitable to be analyzed by PCA.

Next, we performed PCA itself, specifying 5 components, according to the number of latent variables that we have in our research model. We also have requested the most common rotation option, varimax, in order to obtain an easy to interpret solution, where each of our measures would be maximally associated with a single construct. Part of the SPSS output, displaying rotated solution, is presented below.

Rotated Component Matrix(a)

	Component				
	1	2	3	4	5
Fixed&MobilePhoneSubscr				**.823**	
InterTelecomOutgoing					**.941**
MobilePhones				**.932**	
TelephoneMainlines		**.562**	**.538**	**.414**	
ProductivityRatio	**.909**				
ratioGDPInvestPerCapita	**.972**				
ratioFtTW/ICTinvest	**.922**				
ratioMilExp/MilPers		**.837**			
MilExpPerCapita		**.890**			
Exports%ofGDP			**.944**		
Trade%ofGDP			**.964**		
revenuePerSubscriber		**.695**			**.624**

Extraction Method: Principal Component Analysis.

Rotation Method: Varimax with Kaiser Normalization.

Based on the results of the output we could determine that our measures appear to be suitable for representing their constructs. Two measures, "revenuePerSubscriber" and "TelephoneMainlines," appear to be less than perfect but we decided to retain them anyway. Our reasoning was that because the focus of our investigation is on the relationship between the "TFP" and "ICT Capitalization," the results should not be affected by the two imperfect measures representing other constructs in the model. At his point, we could continue our inquiry and perform PLS analysis, results of which are presented in the next section.

7.6.5 PLS Analysis: Steps, Procedures, and Results

Assessment of a research model using PLS analysis could be perceived as consisting of two distinct steps. The first step includes the assessment of the measurement model that primarily deals with the evaluation of the characteristics of the latent variables and measurement items that represent them. The second step involves the assessment of the structural model; this step involves evaluation of the specified by the research model relationships between the latent variables. We present more detailed description of the assessment process below.

7.6.5.1 PLS Analysis: Assessment of the Measurement Model

We assess the adequacy of our measurement model by evaluating the following three criteria: first, the reliability of the individual items and their constructs, second, the convergent validity of the measures representing each construct, and, finally, discriminant validity of the measures (Hulland 1999).

A commonly accepted test of the reliability of the individual items consists of the assessment of the loadings of the measures on their construct, where the loadings of 0.7 and higher are considered acceptable. Assessment of the reliability of the constructs is commonly performed by evaluating the composite reliability of the constructs, with the acceptable values being above 0.7.

Convergent validity of the measures is assessed through the evaluation of the measure of internal consistency (Fornell and Larcker 1981), with values above 0.7 being acceptable (Nunnaly 1978). The process of evaluation involves assessment of the loadings of the measures on their own constructs as well. It is expected that the measures representing a construct would exhibit high loadings on that construct (high convergent validity), and low loadings on the all other constructs in the model (discriminant validity). We would also look at the magnitude and significance of the t-values for the loadings of each of the individual items.

One of the suggested (e.g., Fornell and Larcker 1981) ways of determining discriminant validity in PLS-G is by assessing the average variance that is shared by a construct and the construct's measures. This measure is provided by PLS-G output as AVE. The commonly accepted practice is to substitute diagonal elements of the correlation matrix that includes the correlations between the model's constructs with the squared root of AVE. The adequacy of the discriminant validity is demonstrated if the diagonal elements of the matrix, represented by squared roots of AVEs, are greater then the off-diagonal elements (Hulland 1999). We present the results of the assessment of the measurement model next.

7.6.5.2 PLS Analysis: Results of the assessment of the measurement model

To test our research model we have created three base models, first one where "TFP" was represented by the measure "MI" (Malmquist Index), the second one where "TFP" was represented by the measures "TC" (Technology change component of the Malmquist Index) and "MI", and the third model where "TFP" was represented by the measures "EC"(Efficiency change component of the Malmquist Index) and "MI". We decided that three-model approach would allow us not only to test the presence of the relationship between "TFP" and "ICT Capitalization," but also to test whether or not the relationships exist between "ICT Capitalization" and the each of the two components of Malmquist Index, TC and EC. While such representation would not allow for 'pure' representation of the TC or EC components, it would heavily bias "TFP" along TC and EC components of MI.

Thus, these three different models allow for representing construct "TFP" in the three following ways. First, unexplained growth would be reflected by the measure MI, which while allowing us to determine whether or not relationship between "TFP" and "ICT Capitalization" exists, would not allow discerning the relevant component due to which unexplained growth took place.

Second, by presenting "TFP" as being reflected by the measures MI and TC, we would be able to represent that part of the unexplained growth, which was due to the changes in technology. Thus, by using this model to test the significance of the relationship between "TFP" and "ICT Capitalization" we would be able to test whether or not there is a relationship between "technology" component of TFP and the investments in ICT.

Third, by substituting TC component of MI for EC component, just as it was described above, we would be able to test the relationship between "efficiency" component of TFP and "ICT Capitalization." We begin presentation of the results of the assessment by providing screen shots of the three models, depicted by the Figures 19, 20, and 21 below.

Base Model 1, MI as a measure of "TFP"

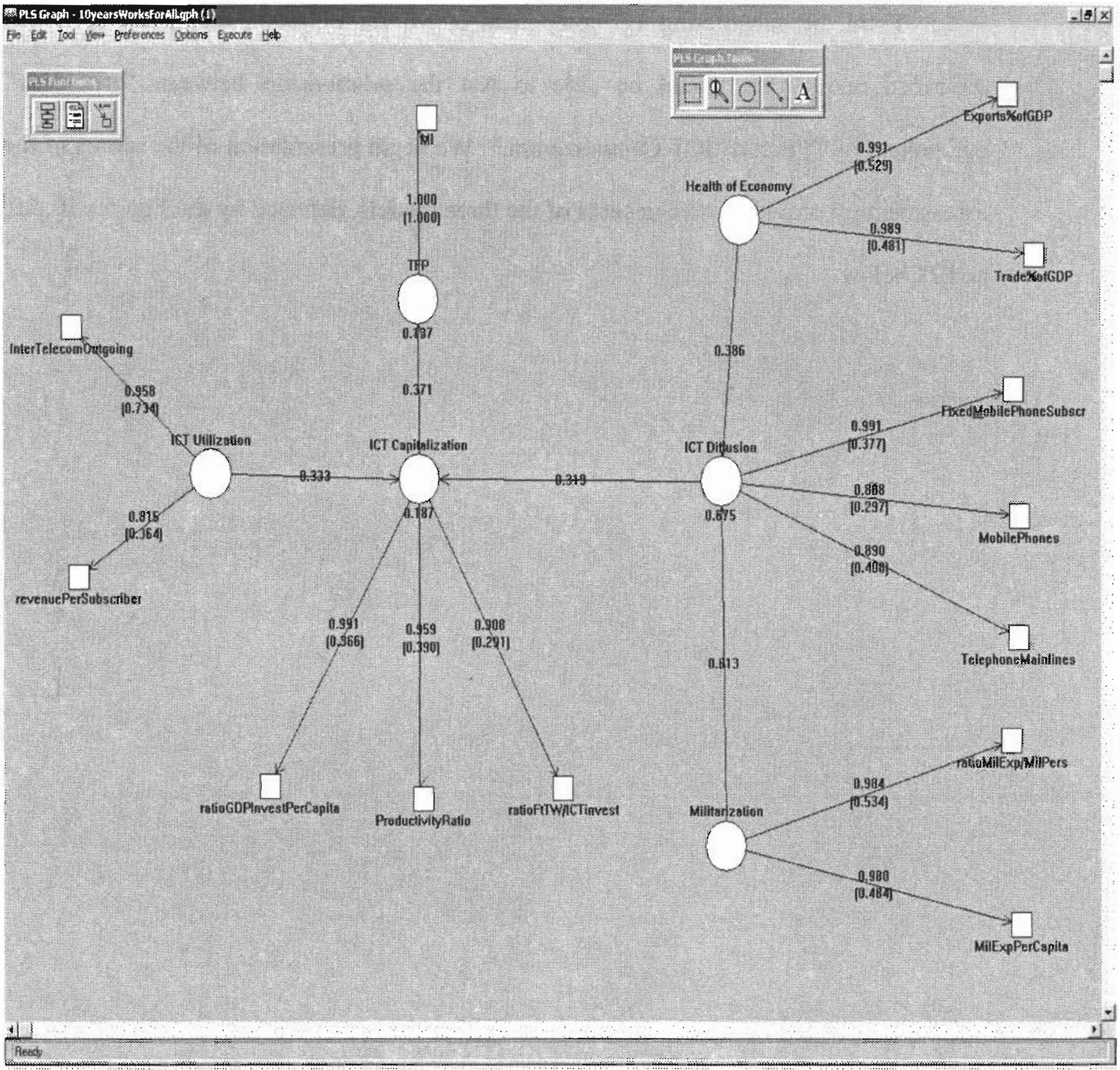


Figure 19 Base model 1

Base Model 2, MI and TC as the measures of "TFP"

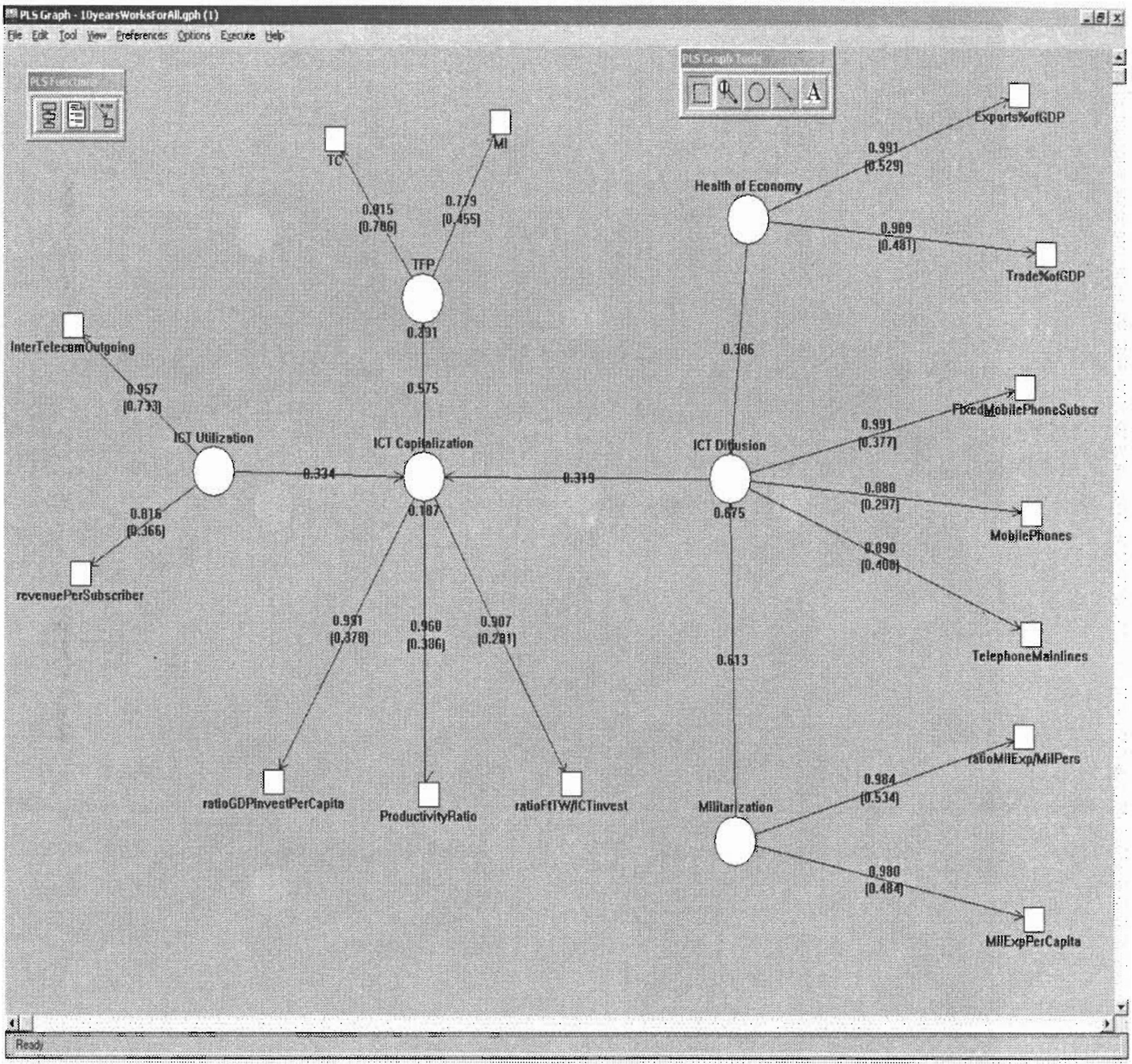


Figure 20 Base model 2

<u>Base Model 3, MI and EC as the measures of "TFP"</u>

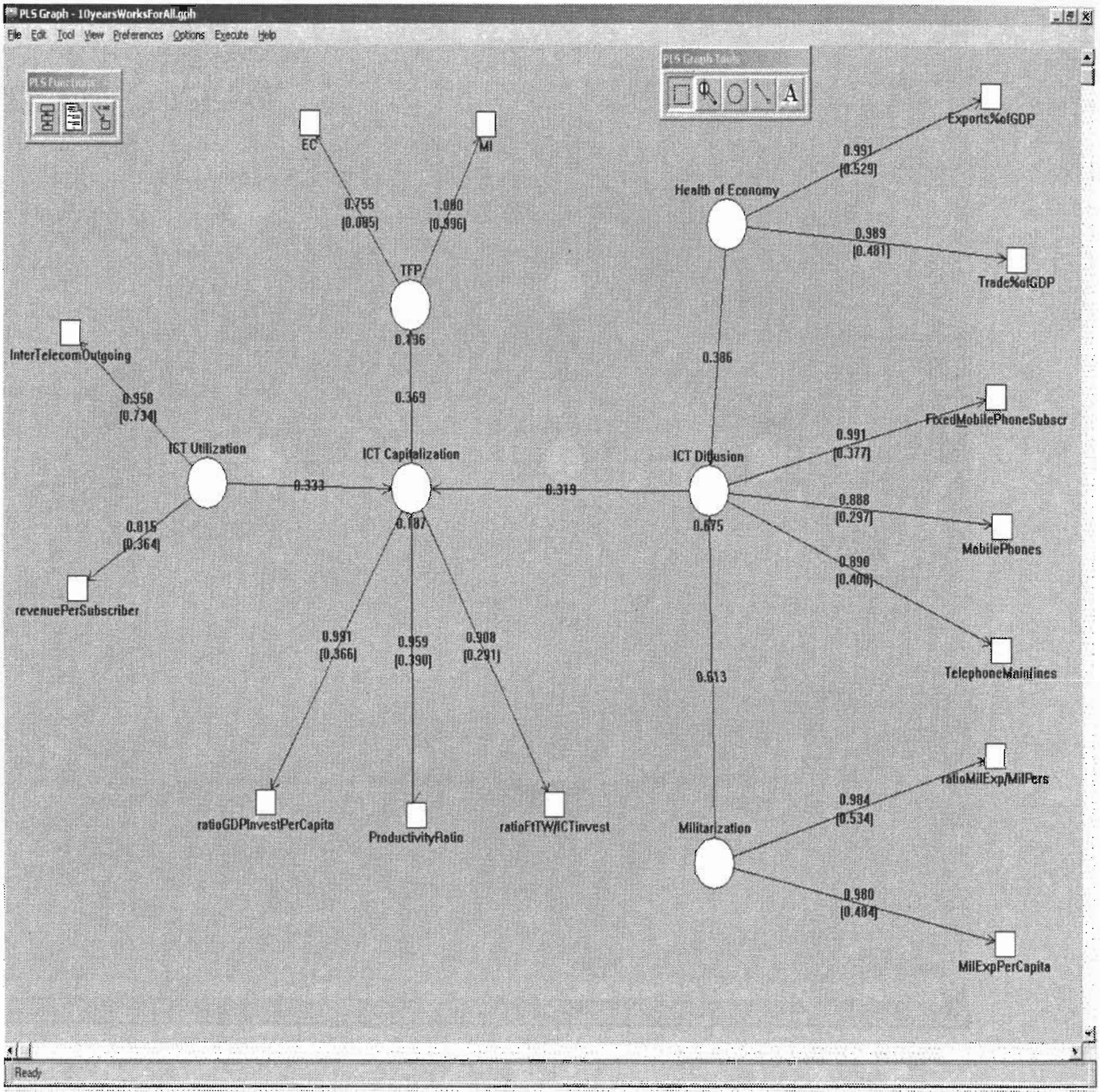


Figure 21 Base model 3

1. <u>Reliability assessment</u>

1.1 Reliability of the Constructs

Composite Reliability (Criteria: should be greater than 0.7)
AVE (Average Variance extracted, Criteria: should be greater than 0.5)

Base Model 1, MI as a measure of "TFP"

Construct/Measure	Composite Reliability	AVE	Squared Root of AVE
TFP	1.000	1.000	1.000
Militarization	0.982	0.964	0.9818
Health of Economy	0.990	0.979	0.9894
ICT Diffusion	0.946	0.855	0.9246
ICT Capitalization	0.968	0.909	0.9534
ICT Utilization	0.883	0.791	0.8893

Base Model 2, MI and TC as the measures of "TFP"

Construct/Measure	Composite Reliability	AVE	Squared Root of AVE
TFP	0.838	0.722	0.8497
Militarization	0.982	0.964	0.9818
Health of Economy	0.990	0.979	0.9894
ICT Diffusion	0.946	0.855	0.9246
ICT Capitalization	0.968	0.909	0.9534
ICT Utilization	0.883	0.791	0.8893

Base Model 3, MI and EC as the measures of "TFP"

Construct/Measure	Composite Reliability	AVE	Squared Root of AVE
TFP	0.878	0.785	0.8860
Militarization	0.982	0.964	0.9818
Health of Economy	0.990	0.979	0.9894
ICT Diffusion	0.946	0.855	0.9246
ICT Capitalization	0.968	0.909	0.9534
ICT Utilization	0.883	0.791	0.8893

Based on the information presented in the table above, we could see that our research model passed the first test of the constructs' reliability assessment. For each one of the three models the measures of the internal consistency (Composite reliability column) are higher than suggested by Nunnaly (1978) baseline of 0.7, and the shared by each construct and its measures variance(AVE column) is significantly higher than suggested by Rivard and Huff (1988) value of 0.5. We examine reliability of the individual measures next.

1.2 Reliability of the individual items

Loadings (Criteria: Must be above 0.5; higher the better)

Base Model 1, MI as a measure of "TFP"

Variable/ Measure	Loading	Communality
MI	1.0000	1.0000
ratioMilExp/MilPers	0.9838	0.9678

MilExpPerCapita	0.9802	0.9607
Exports%ofFDP	0.9906	0.9814
Trade%ofGDP	0.9886	0.9774
FixedMobilePhoneSubscr	0.9912	0.9825
MobilePhone	0.8884	0.7893
TelephoneMainlines	0.8905	0.7930
ratioGDPInvestPerCapita	0.9910	0.9822
ProductivityRatio	0.9591	0.9199
ratioFtTW/ICTInvest	0.9082	0.8249
InterTelecomOutgoing	0.9578	0.9173
revenuePerSubscriber	0.8151	0.6644

Base Model 2, MI and TC as the measures of "TFP"

Variable/ Measure	Loading	Communality
MI	0.7793	0.6073
TC	0.9148	0.8368
ratioMilExp/MilPers	0.9838	0.9678
MilExpPerCapita	0.9802	0.9607
Exports%ofFDP	0.9906	0.9814
Trade%ofGDP	0.9886	0.9774
FixedMobilePhoneSubscr	0.9912	0.9825
MobilePhone	0.8884	0.7893
TelephoneMainlines	0.8905	0.7930
ratioGDPInvestPerCapita	0.9910	0.9822
ProductivityRatio	0.9591	0.9199
ratioFtTW/ICTInvest	0.9082	0.8249
InterTelecomOutgoing	0.9578	0.9173
revenuePerSubscriber	0.8151	0.6644

Base Model 3, MI and EC as the measures of "TFP"

Variable/ Measure	Loading	Communality
MI	1.0000	1.0000
TC	0.7551	0.5702
ratioMilExp/MilPers	0.9838	0.9678
MilExpPerCapita	0.9802	0.9607
Exports%ofFDP	0.9906	0.9814
Trade%ofGDP	0.9886	0.9774
FixedMobilePhoneSubscr	0.9912	0.9825
MobilePhone	0.8884	0.7893
TelephoneMainlines	0.8905	0.7930
ratioGDPInvestPerCapita	0.9910	0.9822
ProductivityRatio	0.9591	0.9199
ratioFtTW/ICTInvest	0.9082	0.8249
InterTelecomOutgoing	0.9578	0.9173
revenuePerSubscriber	0.8151	0.6644

After examining the loadings of the measures provided in the three tables above, we could conclude that our research model also passed the second test of the individual items reliability assessment. Individual loadings of the all items in each of the three models are higher than 0.8, which indicates that the measures and the construct share the significant amount of variance. Next, we are going to examine convergent and discriminant validity of our model.

2. Convergent and Discriminant Validity

2.1 Convergent Validity

Convergent validity (Criteria: T-value, must be significant)

Base Model 1, MI as a measure of "TFP"

Measure	T-value
ratioMilExp/MilPers	318.0371
MilExpPerCapita	241.0288
Exports%ofFDP	799.7814
Trade%ofGDP	555.8673
FixedMobilePhoneSubscr	482.3323
MobilePhone	49.9610
TelephoneMainlines	84.5551
ratioGDPInvestPerCapita	206.0844
ProductivityRatio	31.5843
ratioFtTW/ICTInvest	21.9873
InterTelecomOutgoing	6.5862
revenuePerSubscriber	5.4178

Base Model 2, MI and TC as the measures of "TFP"

Measure	T-value
MI	3.3060
TC	4.8584
ratioMilExp/MilPers	318.0371
MilExpPerCapita	241.0288
Exports%ofFDP	799.7814
Trade%ofGDP	555.8673
FixedMobilePhoneSubscr	482.3323

MobilePhone	49.9610
TelephoneMainlines	84.5551
ratioGDPInvestPerCapita	206.0844
ProductivityRatio	31.5843
ratioFtTW/ICTInvest	21.9873
InterTelecomOutgoing	6.5862
revenuePerSubscriber	5.4178

Base Model 3, MI and EC as the measures of "TFP"

Measure	**T-value**
MI	7.2233
EC	4.5371
ratioMilExp/MilPers	318.0371
MilExpPerCapita	241.0288
Exports%ofFDP	799.7814
Trade%ofGDP	555.8673
FixedMobilePhoneSubscr	482.3323
MobilePhone	49.9610
TelephoneMainlines	84.5551
ratioGDPInvestPerCapita	206.0844
ProductivityRatio	31.5843
ratioFtTW/ICTInvest	21.9873
InterTelecomOutgoing	6.5862
revenuePerSubscriber	5.4178

All t-values are significant, which indicates that our research model passed the first test of the convergent validity. Next, we are going to assess how high the individual measures load on their own constructs, in comparison to the other constructs present in the model.

2.2 Convergent and Discriminant validity

Convergent and Discriminant validity Loadings (Criteria: loadings of the measures on their own construct; measure must load high on its own construct and low on all other constructs)

Base Model 1, MI as a measure of "TFP"

Measure	**TFP**	**Militarization**	**Health of Eco**	**ICTDiffusion**	**ICT Capital**	**ICT_Utilizat**
Fixed&MobilePhoneSubscr	-0.05	0.72	0.54	**0.99**	-0.24	0.14
InterTelecomOutgoing	0.04	0.20	-0.11	0.05	0.32	**0.96**
MobilePhones	-0.02	0.62	0.38	**0.89**	-0.13	0.20
TelephoneMainlines	-0.09	0.69	0.65	**0.89**	-0.36	0.02
ProductivityRatio	0.37	-0.18	-0.30	-0.24	**0.96**	0.39
MI	**1.00**	-0.09	-0.01	-0.06	0.37	0.04
ratioGDPInvestPerCapita	0.38	-0.19	-0.26	-0.26	**0.99**	0.29
ratioFtTW/ICTinvest	0.30	-0.25	-0.23	-0.31	**0.91**	0.13
ratioMilExp/MilPers	-0.08	**0.98**	0.35	0.76	-0.21	0.33
MilExpPerCapita	-0.10	**0.98**	0.27	0.69	-0.20	0.36
Exports%ofGDP	-0.01	0.36	**0.99**	0.60	-0.30	-0.12

Trade%ofGDP	-0.01	0.27	**0.99**	0.55	-0.25	-0.10
revenuePerSubscriber	0.03	0.57	-0.09	0.23	0.16	**0.82**

Base Model 2, MI and TC as the measures of "TFP"

Measure	**TFP**	**Militar ization**	**Health of Eco**	**ICTDif fusion**	**ICT Capital**	**ICT_U tilizat**
Fixed&MobilePhoneSubscr	-0.03	0.72	0.54	**0.99**	-0.24	0.15
InterTelecomOutgoing	0.08	0.20	-0.11	0.05	0.32	**0.96**
MobilePhones	-0.01	0.62	0.38	**0.89**	-0.13	0.20
TelephoneMainlines	-0.07	0.69	0.65	**0.89**	-0.36	0.02
ProductivityRatio	0.58	-0.18	-0.30	-0.24	**0.96**	0.39
MI	**0.78**	-0.09	-0.01	-0.06	0.37	0.04
TC	**0.91**	-0.03	-0.04	-0.02	0.58	0.10
ratioGDPInvestPerCapita	0.61	-0.19	-0.26	-0.26	**0.99**	0.29
ratioFtTW/ICTinvest	0.43	-0.25	-0.23	-0.31	**0.91**	0.13
ratioMilExp/MilPers	-0.06	**0.98**	0.35	0.76	-0.21	0.33
MilExpPerCapita	-0.07	**0.98**	0.27	0.69	-0.20	0.36
Exports%ofGDP	-0.04	0.36	**0.99**	0.60	-0.30	-0.12
Trade%ofGDP	-0.02	0.27	**0.99**	0.55	-0.25	-0.10
revenuePerSubscriber	0.10	0.57	-0.09	0.23	0.16	**0.82**

Base Model 3, MI and EC as the measures of "TFP"

Measure	TFP	Militarization	Health of Eco	ICTDiffusion	ICT Capital	ICT_Utilizat
Fixed&MobilePhoneSubscr	-0.05	0.72	0.54	**0.99**	-0.24	0.14
InterTelecomOutgoing	0.04	0.20	-0.11	0.05	0.32	**0.96**
MobilePhones	-0.02	0.62	0.38	**0.89**	-0.13	0.20
TelephoneMainlines	-0.09	0.69	0.65	**0.89**	-0.36	0.02
ProductivityRatio	0.36	-0.18	-0.30	-0.24	**0.96**	0.39
MI	**1.00**	-0.09	-0.01	-0.06	0.37	0.04
EC	**0.76**	-0.08	0.02	-0.06	0.00	-0.06
ratioGDPInvestPerCapita	0.38	-0.19	-0.26	-0.26	**0.99**	0.29
ratioFtTW/ICTinvest	0.30	-0.25	-0.23	-0.31	**0.91**	0.13
ratioMilExp/MilPers	-0.08	**0.98**	0.35	0.76	-0.21	0.33
MilExpPerCapita	-0.10	**0.98**	0.27	0.69	-0.20	0.36
Exports%ofGDP	-0.01	0.36	**0.99**	0.60	-0.30	-0.12
Trade%ofGDP	-0.01	0.27	**0.99**	0.55	-0.25	-0.10
revenuePerSubscriber	0.03	0.57	-0.09	0.23	0.16	**0.82**

In order to obtain the required loadings we have followed the method outlined by Chin and Hubona (forthcoming 2006), which is as accurate, while being more efficient, as the one offered by Gefen and Straub (2005).

Table above demonstrates that all measures present in our research model share a lot of variance with their own construct, which indicates high convergent validity. In the same time, we could notice that no measure loads highly on any other construct, but its own, thus demonstrating high discriminant validity. Next, we undertake the final test of the convergent and discriminant validity of our research model.

Convergent and Discriminant validity (Criteria: smallest on-diagonal value must be greater than largest off-diagonal value)

Base Model 1, MI as a measure of "TFP"

Construct	**Variance (original on-diagonal values replaced with squared root of AVE)**					
TFP	**1.000**					
Militarization	-0.087	**0.9818**				
Health of Economy	-0.012	0.318	**0.9894**			
ICT Diffusion	-0.059	0.736	0.581	**0.9246**		
ICT Capitalization	0.371	-0.211	-0.281	-0.279	**0.9534**	
ICT Utilization	0.044	0.354	-0.111	0.120	0.295	**0.8893**

Base Model 2, MI and TC as the measures of "TFP"

Construct	**Variance (original on-diagonal values replaced with squared root of**

	AVE)					
TFP	**0.8497**					
Militarization	-0.064	**0.9818**				
Health of Economy	-0.032	0.318	**0.9894**			
ICT Diffusion	-0.044	0.736	0.581	**0.9246**		
ICT Capitalization	0.575	-0.211	-0.281	-0.278	**0.9534**	
ICT Utilization	0.092	0.355	-0.111	0.121	0.295	**0.8893**

Base Model 3, MI and EC as the measures of "TFP"

Construct	**Variance (original on-diagonal values replaced with squared root of AVE)**					
TFP	**0.8860**					
Militarization	-0.087	**0.9818**				
Health of Economy	-0.012	0.318	**0.9894**			
ICT Diffusion	-0.059	0.736	0.581	**0.9246**		
ICT Capitalization	0.369	-0.212	-0.281	-0.279	**0.9534**	
ICT Utilization	0.044	0.354	-0.111	0.120	0.295	**0.8893**

After the comparison of the on-diagonal values of squared root of AVEs with the off-diagonal values, we could easily determine that the lowest on-diagonal value is at least an order of magnitude (Chin 1995) greater than the largest off-diagonal value. Thus, it would appear that our research model successfully passed the last test of convergent and discriminant validity, allowing us to use our measures and proceed with the assessment of the structural model.

7.6.5.3 PLS Analysis: Assessment of the Structural Model

The process of assessment of the structural model involves testing of the significance of the hypothesized relationships between specified in the research model constructs. By running PLS-G analysis, we could observe the path coefficients between the constructs in the model. The significance of the path coefficients then could be evaluated by running bootstrapping procedure, which yields T-values for each path.

Structural models generated for the study

We begin this section by describing the different subsets of the data that we ended up using for testing of the structural model Let us recall that we have generated three base models, where construct "TFP" was represented by MI, MI and TC, and MI and EC. Let us recall as well that the values provided by MI represent change that took place over the period of time, in our case, 1-year period. Thus, we ended up using the data set covering a 9-year period from 1994 to 2002, instead of the original 10-year period from 1993 to 2002. We applied each of the base models to the full data set, covering a 9-year period, and to two reduced data set, first, covering a 5-year period from 1994 to 1998, and, second, covering a 5-year period from 1998 to 2002. Next, based on the results of the Cluster analysis, we generated two subsets for each of the three periods. The two subsets are the "leaders" and the "majority." Thus, we ended up testing multiple structural models, for each of the three base models were applied to the 5- and 9-year periods, and for each period we tested complete data set for the period (we call it "all"), as well as two subsets representing the "leaders" and "majority".

We present the results of the assessment of the structural model next.

7.6.5.4 PLS Analysis: Results of the Assessment of the Structural Model

Altogether, we have generated and tested 27 structural models. All of these models have the same basic structural relationships between their constructs, as it was presented earlier in the section. Despite the same structure, results of the assessment vary from model to model. This variability of the results is due to the differences between the representations of the construct "TFP," as well as the differences in samples that were used to test the models.

We have evaluated fit of our structural models using the R-squared for the endogenous constructs ("ICT Capitalization," "ICT Diffusion," and "TFP"). To evaluate structural paths of the models, we used bootstrapping procedure to obtain t-values, significance level of which was then established using 2-tailed t-distribution table.

Results are presented below.

Strength of the Structural Path from "ICT Capitalization" to "TFP," t-values

Base Model 1, (MI as a measure of "TFP")

Group	1994-1998	1998-2002	1994-2002
All	**1.5192**	**1.1576**	**1.8633**
Majority	**1.5204**	**1.2120**	**1.8218**
Leaders	**0.8258**	**2.1233**	**2.2881**

Base Model 2, (TC and MI as the measures of "TFP")

Group	1994-1998	1998-2002	1994-2002
All	**1.8301**	**0.9866**	**2.2180**
Majority	**1.8935**	**0.9626**	**2.1697**
Leaders	**0.9938**	**1.6786**	**2.1445**

Base Model 3, (EC and MI as the measures of "TFP")

Group	1994-1998	1998-2002	1994-2002
All	**1.5601**	**1.0776**	**1.8489**
Majority	**1.5728**	**1.2026**	**1.8021**
Leaders	**0.3920**	**2.1440**	**2.4328**

Estimated path effects, at $p < 0.05$ level of significance

Table 31 PLS: Significance of the structural path from "ICT Capitalization" to "TFP"

Group, Measures of TFP	1994-1998	1998-2002	1994-2002
All, MI	**Not significant**	**Not significant**	**Not significant**
All, MI and TC	**Not significant**	**Not significant**	**Significant**
All, MI and EC	**Not significant**	**Not significant**	**Not significant**
Majority, MI	**Not significant**	**Not significant**	**Not significant**
Majority, MI and TC	**Not significant**	**Not significant**	**Significant**
Majority, MI and EC	**Not significant**	**Not significant**	**Not significant**
Leaders, MI	**Not significant**	**Significant**	**Significant**
Leaders, MI and TC	**Not significant**	**Not significant**	**Significant**
Leaders, MI and EC	**Not significant**	**Significant**	**Significant**

Structural Path from "ICT Capitalization" to "ICT Diffusion," t-values

Base Model 1, (MI as a measure of "TFP")

Group	1994-1998	1998-2002	1994-2002
All	**6.8921**	**6.2202**	**7.7878**
Majority	**3.4862**	**1.9030**	**3.5353**
Leaders	**2.2375**	**0.4770**	**1.1696**

Base Model 2, (TC and MI as the measures of "TFP")

Group	1994-1998	1998-2002	1994-2002

All	**8.0460**	**6.7620**	**8.1474**
Majority	**3.5995**	**1.8986**	**3.9015**
Leaders	**2.8165**	**0.6449**	**1.2156**

Base Model 3, (EC and MI as the measures of "TFP")

Group	1994-1998	1998-2002	1994-2002
All	**8.0293**	**6.9154**	**7.0644**
Majority	**3.7608**	**1.7279**	**3.6958**
Leaders	**2.5932**	**0.4503**	**1.3350**

Table 31 PLS: Significance of the structural path from "ICT Capitalization" to "ICT Diffusion"

Group, Measures of TFP	1994-1998	1998-2002	1994-2002
All, MI	**Significant**	**Significant**	**Significant**
All, MI and TC	**Significant**	**Significant**	**Significant**
All, MI and EC	**Significant**	**Significant**	**Significant**
Majority, MI	**Significant**	**Not significant**	**Significant**
Majority, MI and TC	**Significant**	**Not significant**	**Significant**
Majority, MI and EC	**Significant**	**Not significant**	**Significant**
Leaders, MI	**Significant**	**Not significant**	**Not significant**
Leaders, MI and TC	**Significant**	**Not significant**	**Not significant**
Leaders, MI and EC	**Significant**	**Not significant**	**Not significant**

<u>Structural Path from "ICT Capitalization" to "ICT Utilization," t-values</u>

Base Model 1, (MI as a measure of "TFP")

Group	1994-1998	1998-2002	1994-2002
All	**2.3323**	**2.3110**	**2.7958**
Majority	**1.8207**	**2.4978**	**2.8019**
Leaders	**2.8694**	**0.8171**	**1.7954**

Base Model 2, (TC and MI as the measures of "TFP")

Group	1994-1998	1998-2002	1994-2002
All	**2.1243**	**2.4560**	**3.6197**
Majority	**2.0245**	**2.3577**	**2.8175**
Leaders	**3.0852**	**0.8831**	**1.5100**

Base Model 3, (EC and MI as the measures of "TFP")

Group	1994-1998	1998-2002	1994-2002
All	**1.9664**	**2.4163**	**2.3787**
Majority	**2.0904**	**2.4867**	**2.7202**
Leaders	**3.7546**	**0.8606**	**1.7955**

Table 32 PLS: Significance of the structural path from "ICT Capitalization" to "ICT Utilization"

Group, Measures of TFP	1994-1998	1998-2002	1994-2002
All, MI	**Significant**	**Significant**	**Significant**
All, MI and TC	**Significant**	**Significant**	**Significant**
All, MI and EC	**Not significant**	**Significant**	**Significant**
Majority, MI	**Not significant**	**Significant**	**Significant**

Majority, MI and TC	**Significant**	**Significant**	**Significant**
Majority, MI and EC	**Significant**	**Significant**	**Significant**
Leaders, MI	**Significant**	**Not significant**	**Not significant**
Leaders, MI and TC	**Significant**	**Not significant**	**Not significant**
Leaders, MI and EC	**Significant**	**Not significant**	**Not significant**

Structural Path from "Health of Economy" to "ICT Diffusion," t-values

Base Model 1, (MI as a measure of "TFP")

Group	1994-1998	1998-2002	1994-2002
All	**7.7468**	**7.5468**	**8.0871**
Majority	**5.3933**	**6.7070**	**7.1869**
Leaders	**14.6129**	**3.7424**	**6.4556**

Base Model 2, (TC and MI as the measures of "TFP")

Group	1994-1998	1998-2002	1994-2002
All	**7.8042**	**6.9607**	**8.4675**
Majority	**5.6178**	**6.0506**	**6.3335**
Leaders	**14.7947**	**4.0979**	**6.1240**

Base Model 3, (EC and MI as the measures of "TFP")

Group	1994-1998	1998-2002	1994-2002

All	**7.2710**	**7.5536**	**7.9041**
Majority	**5.5202**	**6.1593**	**6.2246**
Leaders	**12.9959**	**4.0536**	**6.7515**

Table 33 PLS: Significance of the structural path from "Health of Economy" to "ICT Diffusion"

Group, Measures of TFP	1994-1998	1998-2002	1994-2002
All, MI	**Significant**	**Significant**	**Significant**
All, MI and TC	**Significant**	**Significant**	**Significant**
All, MI and EC	**Significant**	**Significant**	**Significant**
Majority, MI	**Significant**	**Significant**	**Significant**
Majority, MI and TC	**Significant**	**Significant**	**Significant**
Majority, MI and EC	**Significant**	**Significant**	**Significant**
Leaders, MI	**Significant**	**Significant**	**Significant**
Leaders, MI and TC	**Significant**	**Significant**	**Significant**
Leaders, MI and EC	**Significant**	**Significant**	**Significant**

<u>Structural Path from "Militarization" to "ICT Diffusion," t-values</u>

Base Model 1, (MI as a measure of "TFP")

Group	1994-1998	1998-2002	1994-2002
All	**6.5338**	**16.8321**	**11.4635**
Majority	**2.9722**	**10.8319**	**7.2093**
Leaders	**1.3863**	**6.4294**	**5.8360**

Base Model 2, (TC and MI as the measures of "TFP")

Group	1994-1998	1998-2002	1994-2002
All	**6.7023**	**15.4541**	**11.4060**
Majority	**3.0635**	**9.8833**	**6.4577**
Leaders	**1.6243**	**6.7051**	**5.5956**

Base Model 3, (EC and MI as the measures of "TFP")

Group	1994-1998	1998-2002	1994-2002
All	**6.7046**	**17.2956**	**12.0016**
Majority	**3.2787**	**9.3495**	**6.9558**
Leaders	**1.5698**	**7.4594**	**5.5637**

Table 34 PLS: Significance of the structural path from "Militarization" to "ICT Diffusion"

Group, Measures of TFP	1994-1998	1998-2002	1994-2002
All, MI	**Significant**	**Significant**	**Significant**
All, MI and TC	**Significant**	**Significant**	**Significant**
All, MI and EC	**Significant**	**Significant**	**Significant**
Majority, MI	**Significant**	**Significant**	**Significant**

Majority, MI and TC	**Significant**	**Significant**	**Significant**
Majority, MI and EC	**Significant**	**Significant**	**Significant**
Leaders, MI	**Not significant**	**Significant**	**Significant**
Leaders, MI and TC	**Not significant**	**Significant**	**Significant**
Leaders, MI and EC	**Not significant**	**Significant**	**Significant**

Fit of the models, R-squared

R-squared, "TFP"

Group, Measures of TFP	1994-1998	1998-2002	1994-2002
All, MI	**0.150**	**0.058**	**0.137**
All, MI and TC	**0.405**	**0.076**	**0.331**
All, MI and EC	**0.155**	**0.045**	**0.136**
Majority, MI	**0.167**	**0.084**	**0.164**
Majority, MI and TC	**0.440**	**0.101**	**0.374**
Majority, MI and EC	**0.176**	**0.067**	**0.165**
Leaders, MI	**0.015**	**0.173**	**0.640**
Leaders, MI and TC	**0.090**	**0.154**	**0.650**
Leaders, MI and EC	**0.088**	**0.154**	**0.626**

R-squared, "ICT Diffusion"

Group, Measures of TFP	1994-1998	1998-2002	1994-2002
All, MI	**0.586**	**0.856**	**0.675**
All, MI and TC	**0.586**	**0.856**	**0.675**
All, MI and EC	**0.586**	**0.856**	**0.675**
Majority, MI	**0.417**	**0.707**	**0.551**
Majority, MI and TC	**0.417**	**0.707**	**0.551**
Majority, MI and EC	**0.417**	**0.707**	**0.551**
Leaders, MI	**0.624**	**0.710**	**0.471**
Leaders, MI and TC	**0.624**	**0.710**	**0.471**
Leaders, MI and EC	**0.624**	**0.710**	**0.471**

R-squared, "ICT Capitalization"

Group, Measures of TFP	1994-1998	1998-2002	1994-2002
All, MI	**0.237**	**0.351**	**0.187**
All, MI and TC	**0.237**	**0.351**	**0.187**
All, MI and EC	**0.237**	**0.351**	**0.187**
Majority, MI	**0.254**	**0.470**	**0.261**
Majority, MI and TC	**0.254**	**0.470**	**0.261**
Majority, MI and EC	**0.254**	**0.470**	**0.261**
Leaders, MI	**0.458**	**0.186**	**0.174**
Leaders, MI and TC	**0.458**	**0.186**	**0.174**

Leaders, MI and EC	**0.458**	**0.186**	**0.174**

7.6.6 PLS Analysis: Discussion of the results

Based on the results of the testing of our research model we can conclude, that there exist statistically significant relationship between:

1. "ICT Capitalization" and "Technological Change" component of TFP for all 18 of the TEs over 9-year period of time between 1994 and 2002
2. "ICT Capitalization" and "Efficiency Change" component of TFP for the "leaders" group of the TEs over 9-year period of time between 1994 and 2002
3. "ICT Capitalization" and "Technological Change" component of TFP for the "leaders" group of the TEs over 5-year period of time between 1998 and 2002
4. "ICT Capitalization" and "Efficiency Change" component of TFP for the "leaders" group of the TEs over 9-year period of time between 1998 and 2002

Thus, it would appear that obtained evidence is sufficient to allow us to reject the null hypothesis of no relationship between the constructs "ICT Capitalization" and "TFP." Some of the obtained results are, however, surprising. For example, it is not clear why the relationship between "ICT Capitalization" and "TFP" for "leaders" group for the first 5-year period(from 1994 to 1998) turned out to be not significant for both components of the unexplained growth, TC and EC. We would have expected that because the relationship between "ICT Capitalization" and "Efficiency Change" component of "TFP" turned out to be statistically significant for both 10-year period and for the second 5-year period (1998-2002), such relationship should as well hold for the first 5-year period (1994-1998).

One of the possible explanations could be the sample size of the "leaders" subset of the data for the both 5-year periods. In the case of the first 5-year period (1994-1998), the sample size is 34 and for the second 5-year period, the sample size is 42. These sample sizes were considered to be sufficient under "10 times" heuristic, according to which the minimal sample size in PLS should be 10 times larger than the number of the indicators associated with the largest(in terms of the associated with the latent variable indicators) construct. In the case of our research model, the largest construct is "ICT Capitalization," which is associated with 3 measures. Thus, according to "10 times" heuristic sample sizes of 34 and 42 should be large enough in terms of giving us sufficient statistical power to test for the presence of the hypothesized relationship.

Research by Goodhue et al. (2006) investigated, among other things, the effect of the small sample size on the statistical power when using PLS analysis. According to the results of the study (which used four sample sizes, smallest being 40), the "10 times" rule should not be used when employing PLS "for anything but a strong effect size with high reliability" (p. 9). Consequently, regarding the cases where small sample sizes, based of "10 times" rule, were used and no statistically significant results were found Goodhue et al. (2006) suggest that "it would be incorrect to assume that the relationships tested do not exist. Because power was likely too low, these low power non-statistically significant results do not convey any scientific knowledge" (p.9).

Thus, the interpretation of the results of this part of our study should be probably modified according to the suggestion of Goodhue et al. (2006) that "significant results are probably there; non-significant results convey no scientific knowledge" (p. 9). As a result, we should be suspect that the samples sizes that we used to test the significance of the tested path between "ICT Capitalization" and "TFP" for the "leaders" were too small to provide acceptable level of power to detect anything but presence of the strong relationship. Consequently, we feel justified to change the interpretation of the level of significance of the tested path between "ICT Capitalization" and "TFP" from "Not Significant" to "Could Not Be Determined" (CNBD). Accordingly, the updated table is provided below.

Table 35 PLS: Significance of the structural path from "ICT Capitalization" to "TFP," modified

Group, Measures of TFP	1994-1998	1998-2002	1994-2002
All, MI	**Not significant**	**Not significant**	**Not significant**
All, MI and TC	**Not significant**	**Not significant**	**Significant**
All, MI and EC	**Not significant**	**Not significant**	**Not significant**
Majority, MI	**Not significant**	**Not significant**	**Not significant**
Majority, MI and TC	**Not significant**	**Not significant**	**Significant**
Majority, MI and EC	**Not significant**	**Not significant**	**Not significant**
Leaders, MI	**CNBD**	**Significant**	**Significant**

Leaders, MI and TC	**CNBD**	**CNBD**	**Significant**
Leaders, MI and EC	**CNBD**	**Significant**	**Significant**

7.7 Neural Networks and Decision Tree analysis

The purpose of this part of our study is to inquire into the nature of some of the differences between the "leaders" and the "majority" that affects the relationship between their constructs "ICT Capitalization" and "TFP." For this purpose, we constructed two tables (presented below) containing the values and the differences between the values in the percents, of the measures representing the two constructs for the "leaders" and the "majority."

Table 36: NN and DT: Difference between the "leaders" and the "majority" in terms of the values of the indicators of "TFP" and "ICT Capitalization," 10-year period (1993-2002)

	majority	leaders	difference in %
MI	**1.20**	**1.23**	**2.70%**
EC	**1.09**	**1.13**	**3.41%**
TC	**1.13**	**1.11**	**-1.70%**
ProductivityRatio	**10.70**	**3.04**	**-71.61%**
ratioGDPInvestPerCapita	**732.86**	**100.41**	**-86.30%**
ratioFtTW/ICTinvest	**1.94**	**0.06**	**-97.07%**

Table 37: NN and DT: Difference between the "leaders" and the "majority" in terms of Difference between the values of the indicators of "TFP" and "ICT Capitalization," 5-year period (1998-2002)

	majority	leaders	difference in %
MI	**1.16**	**1.17**	**1.06%**
EC	**1.04**	**1.04**	**0.51%**
TC	**1.12**	**1.13**	**0.25%**
ProductivityRatio	**8.35**	**3.62**	**-56.68%**
ratioGDPInvestPerCapita	**370.56**	**96.78**	**-73.88%**
ratioFtTW/ICTinvest	**0.89**	**0.04**	**-95.75%**

We decided to inquire into what are some of the reasons why the relationship between "ICT Capitalization" and "TFP" turned out to be significant for the "leaders" but not for the "majority." These reasons, we propose, could stem from the three following conditions.

First, the relationship between the two construct could be affected by the level of the inputs. In this case, we should be able to determine whether or not the "leaders" differ from the "majority" in terms of the values of the measures that we use to represent the construct "ICT Capitalization." Second, the relationship could be affected by the process of converting the "inputs" (measures of "ICT Capitalization") into the "outputs" (measures of "TFP"). If this is the case, we should be able to determine whether or not the relationship between "TFP" and "ICT Capitalization" for the "majority" would become significant if the "majority" would utilize the process of conversion of inputs into the outputs of the "leaders", while keeping the same level of inputs.

Finally, we propose that the relationship between the construct "ICT Capitalization" and "TFP" could be affected simultaneously by the both factors, namely, insufficient level of the inputs and the ineffective process of the conversion of the inputs into the outputs. If this is the case, then the "majority", while keeping the same level of inputs but utilizing the process of conversion of inputs into the outputs of the "leaders", would exhibit no change in significance of the relationship between the constructs "ICT Capitalization" and "TFP" .

Our inquiry proceeds in two steps: first, we conduct the DT analysis to determine whether or not the levels of the inputs, in the form of the measures of the construct "ICT Capitalization", differ significantly between the "majority" and the "leaders". Second, we use Neural Networks to model the process of the conversion of the inputs into the outputs of the "leaders." We then use the inputs of the "majority" to generate the predicted by NN model outputs in the form of the predicted values of the measures of "TFP." The predicted values then substituted for the original values of the MI, EC, and TC of the "majority," after the path coefficients of the structural model of SEM is tested again and the significance is determined based on the new t-values.

In the following two parts of this section, we describe the DT and NN analysis that we have conducted using SAS Enterprise Miner.

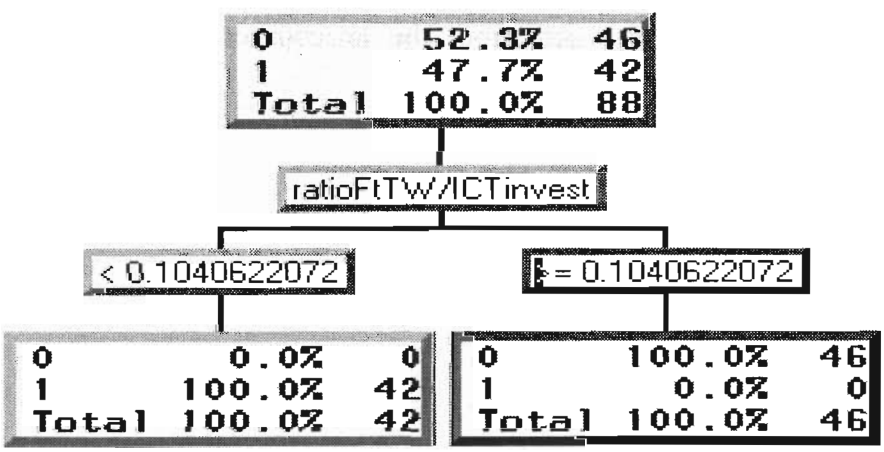
7.7.1 DT: Determining the Difference in the Inputs

We have decided to find out the difference between the "leaders" and the "majority" in terms of the values of the indicators representing the "ICT Capitalization" construct. By creating a binary dummy variable ("0" representing the "majority and "1" representing the "leaders") we were able to use SAS to perform DT analysis to inquire into this matter. The two data sets that we used, one for 5-year period and the second one for 10-year period, contained 4 variables, 1 dummy variable and 3 variables representing the indicators of the "ICT Capitalization."

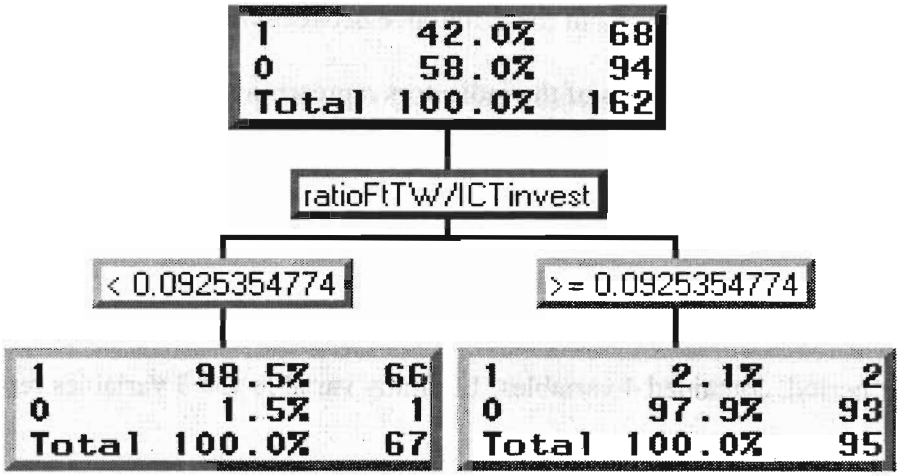
Once the first DT model was generated, we reset the split variable to "don't use" and re-run the analysis. Altogether 2 single-split models were generated, giving us some insights regarding the differences between the "majority" and the "leaders."
The results of the analysis presented below.

First Split

5-year period



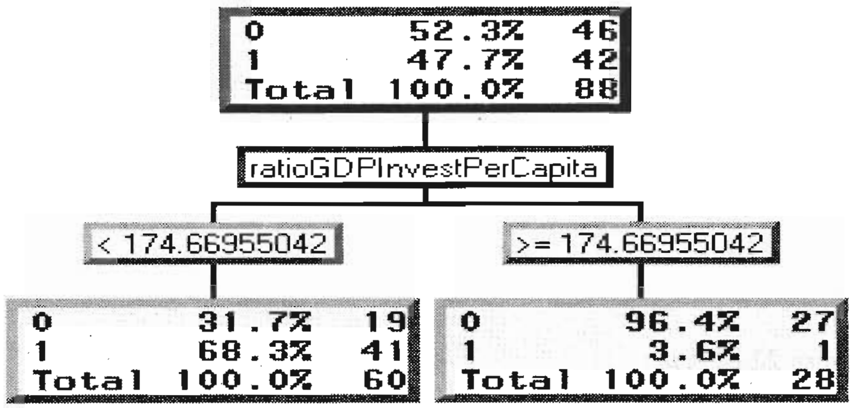
10 year period



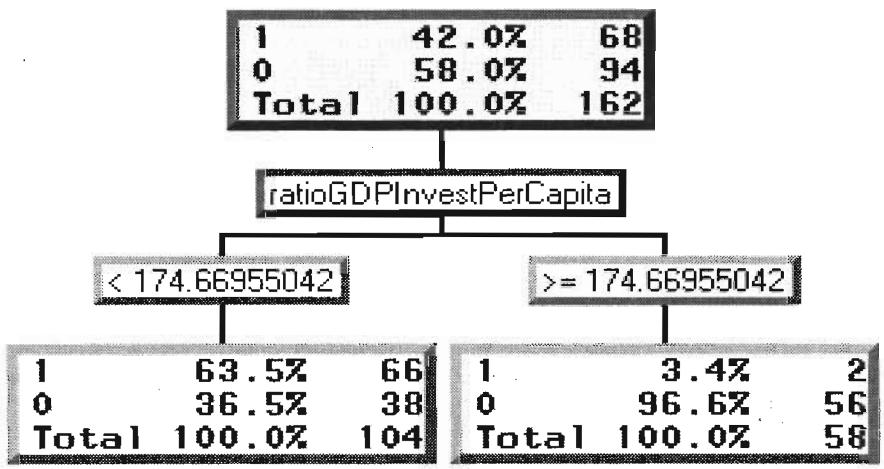
According to the first model "leaders" clearly differ from the "majority" in terms of the values of the variable "ratioFtW/ICTinvest," which is the ratio of the number of Full-time telecom employees to the annual telecom investment, in 1000 US $. This ratio is the inverse of the "investment per telecom employee, in 1000 US $" and the purpose of its creation was to represent, from the perspective of the policy-makers, the 'attention' or 'supervision' that the investment gets from the telecom employees. The simplest interpretation of this split is that given the same level of the investment, "leaders" have less telecom employees than the "majority," or, conversely, that the "leaders" invest in ICT more, per telecom worker, then the "majority."

Second Split

5-year period



10 year period



The second model creates a split based on the value of the variable "ratioGDPInvestPerCapita," which is the ratio of the "GDP per capita" and "Annual Investment in telecom, per capita." Similarly to the results of the first model, we could see that over 60% of the "majority" invest less in telecoms per capita than the "leaders." Thus, based on the conducted DT analysis we could conclude that the "leaders" do differ from the "majority" in terms of the inputs.

7.7.2 NN: Determining the Difference in the Process of Converting the Inputs into the Outputs

We present the way that we conducted NN analysis by using the diagram depicted by the Figure 22 below.

The diagram uses two Input Nodes, one containing the data of the "leaders," and another one containing the data of the "majority."

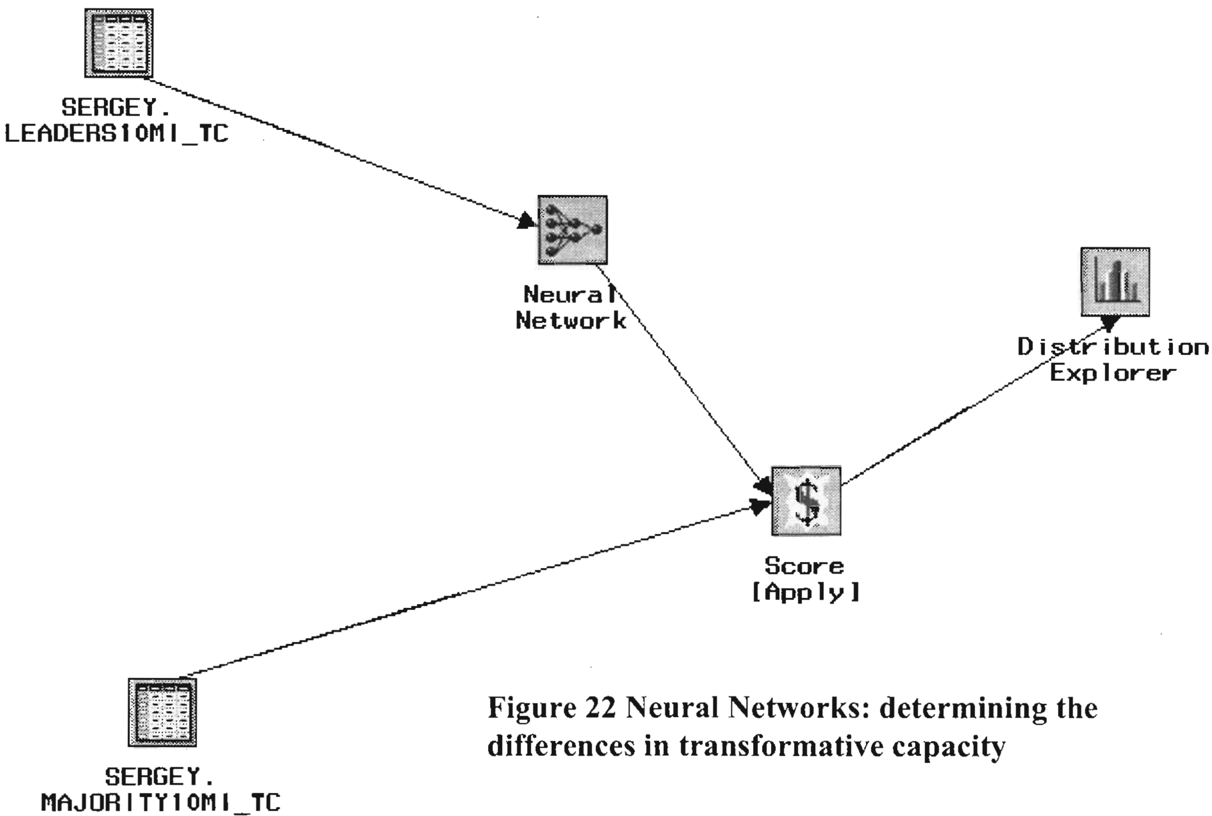


Figure 22 Neural Networks: determining the differences in transformative capacity

Below we provide the screen shot of the contents of the "leaders" Input Node. As one can see, input is provided in the form of the measures that we use to represent the construct "ICT Capitalization," while output (designated as "target") is provided in the form of the measures that we use to represent the construct "TFP."

Name	Model Role	Measurement	Type	Format	Informat	Variable Label
PRODUCTIVITYRATIO	input	interval	num	BEST12.	12.	ProductivityRatio
MI	target	interval	num	BEST12.	12.	MI
TC	target	interval	num	BEST12.	12.	TC
RATIOGDPINVESTPERCAPITA	input	interval	num	BEST12.	12.	ratioGDPInvestPerCapita
RATIOFTTW_ICTINVEST	input	interval	num	BEST12.	12.	ratioFtTW/ICTinvest

Unlike the Input Node of the "leaders," the Input Node of the "majority" contains only the input measures, with no output ("target") measures available.

Name	Model Role	Measurement	Type	Format	Informat	Variable Label
PRODUCTIVITYRATIO	input	interval	num	BEST12.	12.	ProductivityRatio
MI	rejected	unary	char	$1.	$1.	MI
TC	rejected	unary	char	$1.	$1.	TC
RATIOGDPINVESTPERCAPITA	input	interval	num	BEST12.	12.	ratioGDPInvestPerCapita
RATIOFTTW_ICTINVEST	input	interval	num	BEST12.	12.	ratioFtTW/ICTinvest

The purpose of the Neural Network node is to create a model according to which inputs of the "leaders" are converted into the outputs. The created model then is stored in the "Score!" node, and the predicted values of the outputs for "majority," created based on the NN's model, then could be retrieved from the "Distribution Explorer" node. Screen shot depicting part of the predicted output, obtained from the "Distribution Explorer," is provided below.

	ProductivityRatio	MI	TC	ratioGDPInvestPerCapita	ratioFtTW/ICTinvest	Hidden: H1=1	Predicted: MI	Predicted: TC
1	27			2041.1066433	4.7042595652	-0.998764792	11.704774368	1.0333460364
2	6.0966386555			471.69538109	0.8808447479	0.4961341433	3.3832941018	1.0951744649
3	211.61111111			17493.224528	26.868111111	-1	11.711650256	1.0332949487
4	123.4			7159.5284378	15.180931333	-1	11.711650256	1.0332949487
5	45.391752577			2125.3093263	3.4994299485	-0.992908568	11.672175205	1.0335882475
6	114.15841584			4701.756324	6.0474538614	-0.999996334	11.711629847	1.0332951003
7	3.5227272727			201.10309254	0.2455045833	0.8414268017	1.4611935675	1.1094556327
8	7.5899529042			184.83232979	0.1666916797	0.8377457431	1.4816844889	1.1093033855
9	7.7893569845			139.91824731	0.1044207539	0.8528531588	1.3975877928	1.1099282223
10	11.232217573			794.18545052	6.5459539749	-0.999968538	11.711475121	1.03329625
11	11.014157014			767.05862737	4.8583822394	-0.998866012	11.705337819	1.03334185
12	1.6057888941			52.00637445	0.3093739185	0.8253585683	1.5506387359	1.1087910569
13	1.7863359443			70.038698712	0.3766600914	0.8009469539	1.6865280342	1.1077814022
14	2.0288028115			73.0521913	0.3248091069	0.8182056371	1.5904561276	1.1084952145
15	2.5833472245			82.300683357	0.3543133606	0.8037463319	1.6709450621	1.1078971833
16	2.1609900571			72.72732	0.2940573563	0.8277501237	1.5373259423	1.1088899707
17	4.5502482622			116.77450024	0.3946194359	0.7730645254	1.8417379102	1.106628196
18	3.6081754736			104.49203714	0.2598741084	0.8306563754	1.5211480488	1.1090101721
19	18.314720812			3773.266106	15.884487672	-1	11.711650256	1.0332949487
20	23.167189305			2223.499786	10.197979865	-0.999999987	11.711650181	1.0332949493

Altogether, we have created six models, three for the 5-year period and three for the 10-year period. For each period we have created the models with the predicted output based on the, first, MI, second, MI and TC, and third, MI and TC measures of the construct "TFP." After that, the predicted values have been imported into Excel spreadsheet and substituted into the original data sets for the "majority."

After that, we have re-run the SEM analysis for each of the six datasets and recorded the path coefficients and the t-values for the path from the construct "ICT Capitalization" to the construct "TFP." We then summarized the results in the form of a table, which is presented below.

Table 38 NN and DT: Path Coefficients and t-values of the relationship between "ICT Capitalization" and "TFP" for "majority," based on the predicted measures of "TFP"

	MI	MI and TC	MI and EC
5 years			
Path	0.141	0.113	0.250
t-value	1.6363	1.2619	**4.7536**
10 years			
Path	0.615	0.615	0.607
t-value	**8.6325**	**7.8834**	**8.1683**

Based on the results of the NN analysis we can conclude, that the "majority" is capable, even with the unchanged levels of the input, of displaying statistically significant relationship between the construct "ICT Capitalization" and "TFP." The prerequisite for this, however, would be the improvement of the process of converting the inputs, in the form of the measures of the construct "ICT Capitalization," into the outputs, as represented by the measures of the construct "TFP."

7.7.3 NN and DT Analyses: Summary

Based on the results of the analysis we could conclude that there exists a difference between the "majority" and the "leaders" in terms of the inputs (measures of the construct "ICT Capitalization"). Moreover, we also can conclude there exists a difference between the how "majority" and the "leaders" convert the inputs into the outputs (measures of the construct "TFP").

However, it would appear that the reason why the “majority” does not exhibit statistically significant relationship between the constructs “ICT Capitalization” and “TFP” lies not in the insufficiently low levels of the inputs, but rather in the inefficiency of the process by which the inputs are converted into the outputs. Thus, the “majority” would likely benefit more not from the increasing the levels of the inputs, but from the increasing the efficiency and effectiveness of the utilization of the current level of the inputs.

CHAPTER 8 Summary and Conclusion

In the last chapter of this paper, we discuss the theoretical and methodological contributions of our study. We as well comment on the key findings of our research, and the practical implications of the main findings. In conclusion, we discuss the limitations of this study and provide some directions for the future research.

8.1 Contribution of the study

8.1.1 Theoretical Contribution

The Neoclassical Growth Accounting framework and Theory of Complementarity provided a theoretical foundation for this study. By applying these two theories to the context of the economies in transition, we were able to test them in the underresearched setting and corroborate their validity. Thus, the first theoretical contribution of this study refers to the strengthening of the Neoclassical Growth Accounting framework and Theory of Complementarity by means of corroborating their validity in the setting of the Economies in Transition.

In this paper, we also proposed and tested a prototype of the theoretical framework that we called "ICT Investment Success." Despite the incompleteness of the developed in this study model, we believe that it could serve as a platform for the development of a more comprehensive theoretical model relating the investments in ICT and TFP. As a result, the second theoretical contribution of this study refers to the proposing and testing of the theoretical model; albeit in the preliminary stage of its development.

Additionally, we have suggested an approach, implemented as our seven-step methodology, to the data analysis in the absence of theory. It is often the case that in the situations of high complexity, especially when a researcher inquires into an interdisciplinary problem, no comprehensive theory is available to guide the analysis of the data. Our research proposes, albeit implicitly, and demonstrates the sketch of the theoretical framework of the data analysis in such circumstance. Let us recall that the ultimate goal of our research was to outline a theoretical framework linking the investments in ICT to the TFP. However, to our knowledge there is no theory purpose of which is to guide a theory development. This research, by demonstrating rigorous approach to the theory development, suggests the answer to the question 'What is the theory of theory building?' Consequently, we consider the suggested by this study framework of 'theory building in the absence of guiding theory' to be the third theoretical contribution of our study.

8.1.2 Methodological Contribution

It is our belief that in addition to the theoretical contributions, this study also makes some important methodological contributions.

We consider used in this study what could be called as a 'multi-method methodology of data analysis' to be the most important methodological contribution of our research. The outcome of this contribution is an approach that is structured, rigorous, and consistent. Our multi-method approach allows for the reliable data analysis where each finding is cross-validated by multiple methods. Moreover, the analysis of the data performed in accordance with our multi-method methodology is self-sufficient; for while corroborating the results of the other studies it does not depend on them. Furthermore, the use of such multi-method methodology increases internal reliability and external validity of the findings, allowing for a higher level of confidence in the results of the data analysis. Next, in addition to what we consider to be our primary contribution in this area, we would like to discuss some other methodological contributions of this research.

First, our data analysis followed a seven-step approach, where the output of each step, in the form of the results, was used as an input to a next step. As a result, we were able to propose and test what we call a 'sequential methodology,' which is systematic, reliable and based on the widely accepted data analytic methods. Thus, the other methodological contribution of this study is the proposed methodology that utilizes a sequential use of methods; this methodology was successfully applied and tested in the context of the real-world data.

Second, by using fairly large variety of methods we are able to provide the multiple "entry points" into our study, in the form of the findings and insights obtained by each method. Consequently, this research could be accessed from the perspective of any of the six methods that we used in the study, which allows for easier incorporation of the results of our study into the nomological network of the existing body of knowledge.

Third, in our research we were able to incorporate and use side-by-side parametric and non-parametric methods, which allowed us to consistently relate the available to us data to the theory. Use of non-parametric methods allowed us to minimize the assumptions about our sample data, while use of parametric methods allowed us to place the findings obtained by non-parametric methods in the theoretical context.

Despite that usually use of non-parametric methods in a study does not allow the researchers to test the theory, by incorporating parametric and non-parametric methods we were able to do so. We suggest that such combination of methods could allow the researchers to lessen their reliance on the assumptions about the data, while increasing their confidence about the actual relationships that exist within the sample of the study. Such side-by-side incorporation of the parametric and non-parametric methods allowed us almost simultaneously test the theory against the data sample.

Fourth, we have proposed and tested the approach allowing for relating the investments in ICT to TFP. As a result, we have demonstrated one possible way of dealing with the "endogeneity problem," which could be applied in similar situations.

Fifth, we have demonstrated a possibility of creating a structural equation model where each of the measures that represent each of the model's construct is represented objectively. Meaning, to represent each measure we use a commonly accepted indicators obtained from the public sources, which have been used extensively across multiple fields of scientific inquiry and practice. Thus, such representation of our constructs makes it easier for fitting of our model into the nomological network and relating the results of the study across the different fields of study, such as economics, public policy, finance, etc.

Moreover, we have demonstrated that the use of Cluster Analysis prior to DEA allows for testing of the one of the important assumptions of the DEA method: functional similarity of DMUs. The results of the CA in our study suggest that such assumption should not be taken for granted, but rather tested for.

Another contribution of this research is that we suggested a way by which the DT analysis could be used to inquire into the differences between the efficient and inefficient DMUs in the case of the heterogeneity of the sample. As a result, a nature of the efficiency of a DMU could be investigated with higher degree of precision by allowing considering the important differences between the DMUs in the sample.

Finally, in our study we have employed a multiple confirmation of the results, where each finding is corroborated by multiple methods. For example, lower effectiveness of the "majority" is confirmed by DEA and NN analysis, a finding regarding a higher economic development of the "leaders" supported by both Cluster and

DT analyses. Thus, such approach allows us to have a higher level of confidence in the validity and reliability of the results of the study.

8.1.3 Key Findings

In this section of the conclusion, we describe the major findings of our research. We present the results in order that they were obtained, following the steps of our methodology.

Step 1: *Cluster Analysis*

The first finding of our study involves use of CA, which allowed us to identify two distinct groups of TEs in our sample. One of the clusters, which we named "leaders," contained such countries as Poland, Czech Republic, Hungary and Slovenia, while another cluster, "majority," was represented by Albania, Armenia, Kyrgyz Republic, Kazakhstan, Moldova and others. Another interesting findings provided by CA was that unlike the countries listed above, which 'reside' within the identified clusters permanently, there is a group of the TEs which 'migrate' from cluster to cluster. The membership of 'migrants' depends on a year, for example, for a 10-year period from 1993 to 2002 following countries have changed memberships:

- Bulgaria resides in "leaders" for 2002, while being in "majority" the rest of the time

- Latvia occupies the "majority" cluster for 1993 and 1996,while being among the "leaders" the rest of the time
- Lithuania has a changing point in year 1999, during which this TE migrates from the "majority" cluster, where it resided for the period from 1993 to 1998, to the "leaders," where it stays from 1999 to 2002
- Slovak Republic despite having the "leaders" membership for the most of the years, has the 'regressive' points in 1993, 1994, and 1999, where this TE shares the cluster with the "majority."

This allows us to suggest that despite being labeled by the same term, Transition Economies, some of the countries differ quite significantly from each other, while for others differences are less pronounced.

In terms of the major differences between the "majority" and the "leaders," splits produced by CA suggest that the Annual Telecom investment and Total Telecom Services Revenue are two of the criteria along which the two clusters differ most.

Step 2: *Decision Tree Analysis*

The second finding of our study was obtained by using DT analysis; this method allowed us to inquire into some of the differences between the two clusters identified by CA. The goal of the analysis was identifying some of the characteristics of the "leaders" and the "majority."

Based on the results of the analysis we can conclude that on average "leaders" differ from the "majority" in the following areas:

A. economic wealth, where the "leaders have a higher GDP per capita than the "majority"

B. health care, where the "leaders" have higher level of health expenditure than the most of the "majority"

C. ICT-related infrastructure and its utilization, where the "leaders" have:

   a. larger number of mobile phones,

   b. greater number of telephone mainlines than "majority" and

   c. greater percentage of the internet users among the population.

D. investments in ICT, where the "leaders" have consistently higher level of annual telecom investment

E. education, where the "leaders" have greater number of teachers per pupil in the system of primary education

F. overall education level of the labor force, where the "leaders" have greater percentage of the total labor force employed as R&D technicians

G. communication with the international community, where the "leaders" have higher level of expenditures on international tourism, as well as higher level of international telecom traffic, than the "majority."

H. militarization, where the "leaders" have lower than the "majority":

   A. level of military expenditure

   B. percentage of the labor force serving in military.

Undoubtedly, based on the results of DT analysis we can conclude that the "leaders" appears to be wealthier than the "majority." Consequently, while the case for the 'migrant' countries that change the cluster memberships is not clear cut, we should be able to confidently state that for 'permanent residents' of the clusters the results should hold true. Therefore, based on the results of the DT analysis, we can conclude that such "leaders" as Poland, Czech Republic, Hungary and Slovenia appear to be wealthier in general, having better infrastructure and smaller armies than representing "majority" Albania, Armenia, Kyrgyz Republic, Kazakhstan, Moldova and Ukraine.

Step 3: *Data Envelopment Analysis*

The third finding of our study was obtained by means of using DEA. We were able to determine that the "leaders" are more efficient than the "majority" not only in terms of the production of the output, but also in terms of the utilization of the inputs. For example, according to the results provided by the input-oriented model (utilizing Farrell Input-Saving Measure of efficiency) the "leaders" are 7.5% to 12.5% more efficient than the "majority" over the 10-year period from 1993 to 2002.

As for the output-oriented model (utilizing Farrell Output-Oriented Measure of efficiency), the results indicate that the gap between the two groups of TES is even larger, for the "leaders" are 11% to 16.5% more efficient than the "majority" over the 10-year period from 1993 to 2002. The interpretation of the findings is straightforward: the "leaders" are not only more efficient than the "majority" in term of utilization of the inputs (i.e., investments, available infrastructure, etc.), but are also much more effective in terms of the transformation of the inputs into the outputs.

Moreover, similar comparison of the values of the Malmquist index demonstrate that the gap in performance, expressed as a change in MI, widens on average at the rate of 1.8% over 10-year period of time for output-orientation. In terms of the input-orientation, the gap would appear to be diminishing, albeit at such low rate of 0.13% that the "majority" would not be able to close on the "leaders" any time soon.

We could summarize the findings of DEA as follows: while the "leaders" are already more efficient and even more effective than the "majority," additionally, there are strong indications of the divergence between the two groups of TEs. Consequently, if the countries like Albania, Armenia, Kyrgyz Republic, Kazakhstan, Moldova and Ukraine want to get close in terms of the performance to the countries like Poland, Czech Republic, Hungary and Slovenia, they need to increase their efficiency and effectiveness at a higher rate than they do now.

Step 4: *Decision Tree Analysis*

The fourth finding of our research was obtained by DT analysis, when we decided to find out what are the characteristics that determine the efficiency of the "leaders" and the "majority." We have discovered that in the case of the "leaders" efficiency could be obtained by means of increasing the magnitude of the ratio of the revenues from telecoms to the investments in telecoms. In terms of the "majority," however, situation is slightly different, for increase in efficiency comes from the increase in revenues from the telecom investments. Thus, the increase in the efficiency of such countries as Poland, Czech Republic, Hungary and Slovenia comes from the balanced increase in revenues and investments, while TEs as Albania, Armenia, Kyrgyz Republic, Kazakhstan, Moldova and Ukraine could increase their efficiency simply by maximizing the revenues provided by the telecoms.

Step 5: *Translog Function*

The fifth finding of our study involved the use of the translog function. We have determined that there exists a statistically significant interaction effect between the investments in ICT and other variables, representing a state of labor, as well as representing a state of the investments. As a result, we were able to establish a following two sets of the investments that are complementary to the investments in ICT:

A. Labor

- Researchers in R&D (% of total labor force),
- Technicians in R&D (% of total labor force) ,
- Full-time telecommunication staff(% of total labor force),
- Military personnel (% of total labor force),
- Physicians (% of total labor force)

B. Investment:

- International tourism, expenditures (% of GDP in Current US $),
- Foreign direct investment, net inflows (% of GDP), and GDP (current US$)
- Health expenditure, private (% of GDP in current US $)

We would like to present some of the possible interpretations of the findings in this part.

For example, the number of the Researchers and Technicians involved in R&D activities are not only representative of overall level of education of the labor force in a given TE, it also could reflect a level of innovativeness present in the economy, as well as be reflective of the overall economic wealth of the country.

It would appear that the number of physicians and the level of private health expenditure are also suggestive of the overall wealth of the economy, for it would appear reasonable that citizens of the wealthier countries would have higher level of private spending on health, which also could possibly sustain the larger number of physicians. Additionally, it is clearly suggestive of the level of health care in a given TE, as well as the level of education that is ultimately responsible for the number of physicians.

The estimate values of β for the interaction term between Annual Telecom Investment and Full-time Telecom Staff change the signs (the first one from negative to positive in three 5-year data sets, and the second one from negative, in 5-year data set, to positive in 10-year data set) depending on the data set used in analysis. This change in direction of the effect suggests that the different levels of these variables could produce different types of the effects, thus suggesting the presence of the threshold effect. First and foremost this suggests that some form of the balance must be found between the number of the Full-time telecommunication workers and the level of their productivity.

Finally, expenditures on International tourism and the inflow of FDI are indicative of, first, international openness of the country, second, attractiveness of the economy to the foreign investors, and, third, the level of wealth of the economy, which translates into the levels of the disposable income of the individual citizens.

Overall, the results of the Translog analysis suggest that there are three 'grand' factors at play that manifest themselves in the form of complementary to ICT investments. The first factor is the overall wealth and health of the economy. The second factor is the wealth and health of the social capital, manifesting itself as the level of health care and education. Finally, socio-cultural and economic openness of a given economy could be proposed as a third 'grand' factor affecting the investments in ICT in the context of TEs.

Step 6: *Structural Equation Modeling with Partial Least Squares*

The sixth finding of our study was obtained through the use of SEM implemented with PLS.

Let us recall that our research model consists of five constructs, TFP, ICT Capitalization, ICT Diffusion, ICT Utilization, Health of Economy, and Militarization. In our model, ICT Capitalization is affected by ICT Diffusion and ICT Utilization, and ICT Diffusion is affected by Health of Economy, and Militarization. We tested the model using the data corresponding to three time periods: first 5 years (1994-1998), second 5 years (1998-2002) and complete period from 1994 to 2002. Because MI represents the change that took place, in the case of this research, within a year, first data point, 1993, was not used.

For the purposes of this research the most important path that was tested in the one between "ICT Capitalization" and "TFP," because this path represents the relationship between the investments in ICT and unexplained growth, a.k.a, TFP, Solow's residual, Multifactor productivity. We have represented the construct "TFP" in three different ways.

First, we tested the model when "TFP" was reflected by a single measure, MI. We were able to establish the presence of the relationship between the two constructs of our conceptual model, "ICT Capitalization" and "TFP" for the "leaders" group of our sample. The construct "ICT Capitalization" was represented by the three ratio measures, all of which contain variable "Annual investment in telecom" in the denominator, while the construct "TFP" was represented by the Malmquist Index. Thus, it allows us to state that we have established the presence of the relationship between the investments in ICT and TFP for such members of the "leaders" as Poland, Czech Republic, Hungary and Slovenia, and, of course, for all of the 'migrants' for those years that they enjoyed membership in the "leaders" cluster.

This model, however useful for establishing relationship between "ICT Capitalization" and "TFP," does not really help us to determine whether this relationship attributable to the TC or EC component of MI. This is extremely important, for TC represents the movement of the efficiency frontier, while EC represent the movement relative to the efficiency frontier.

Consequently, we have tested the second model where the construct "TFP" was represented by the two measures, MI and TC. Use of these measures allowed us to bias the representation of the "TFP" in the direction of the TC. Let us recall, that TC component of MI represents the change attributable to the change in technology. We were able to determine the presence of the statistically significant relationship between the two constructs for the both groups of TEs. The interpretation of this finding is quite simple, what it means is that every country out of 18 in our sample by investing in ICT could benefit from the associated with this investment technological change. However, this model still does not allow us to find out whether or not investments in ICT are associated with the EC component of MI.

Thus, we tested the third model, where the construct "TFP" was represented by the two measures, MI and EC. Similarly to the second model, by using these measures we were able to bias the representation of the "TFP," in this case in the direction of the EC. After we tested the third model, we were able to establish the presence of the relationship between the "ICT Capitalization" and "TFP" for the "leaders" group of our sample, but not for the "majority." Again, interpretation of this finding is relatively straightforward, for it means that such TEs as, for example, Albania, Armenia, Kyrgyz Republic, Kazakhstan, Moldova and Ukraine, do not translate their investments in ICT into the increased efficiency. Meaning, all that the members of the "majority" get in return for investing in ICT is a better technology, which, sadly, does not really affect their efficiency of utilizing the obtained technology.

Moreover, we were also able to establish the presence of the relationship between "ICT Capitalization" and "ICT Diffusion" for the "majority." One of the possible interpretations of this finding is that the investments in ICT for the "majority" are affected by the state of the ICT-related infrastructure, where the expansion of the infrastructure fuels the investments. While this relationship is significant for the "majority" for all three periods, for the "leaders" this relationship turned out to be significant only for the first 5-year period from 1994 to 1998. What it possibly suggests, is that the "leaders" used to invest in ICT in order to build up the infrastructure, but now this period is over and some other factor fuels the investments in ICT for such countries as Poland, Czech Republic, Hungary and Slovenia.

The very similar picture was obtained when we tested the strength of the path between "ICT Capitalization" and "ICT Utilization." It turned out that for the "majority" the relationship is significant for all of the three periods, while for the "leaders" this relationship turned out to be significant only for the fist 5-year period from 1994 to1998. One of the possible interpretations of this finding is that the investments in ICT could be fueled by the demand in the utilization of the ICT-related infrastructure. While for the "majority" investments in ICT are still driven by the expansion in infrastructure and high demand on the utilization of the infrastructure, this is no longer so for the "leaders." It could be suggested that starting from year 1998 and on the "leaders" acquired enough of the infrastructure and a sufficient enough level of the utilization of that infrastructure that these two factors no longer drive the investments in ICT.

Finally, the relationships between the "ICT Diffusion," "Health of Economy," and "Militarization" turned out to be statistically significant for both groups of the TEs. However, this finding is of special importance to the "majority," for their investments in ICT are still driven by the expanding ICT infrastructure.

Step 7: *Decision Tree Analysis and Neural Networks*

The seventh finding of our study was obtained by using DT an NN analyses. We were interested in gaining the insights into the nature of the differences due to which the "leaders" exhibit the presence of the statistically significant relationship between the investments in ICT and TFP, while the "majority" does not.

In order to do so, we conducted our inquiry in two phases. First, we used DT to find out whether or not the "leaders" differ from the "majority" in terms of the inputs, which are in our case the indicators reflecting the construct "ICT Capitalization.".

By using DT analysis, we were able to determine that one of the reasons why the "leaders" exhibit the statistically significant relationship between the investments in ICT and TFP is that given the same level of the investment, "leaders" have less telecom employees than the "majority." Alternatively, conversely, we can interpret it as that the "leaders" invest in ICT more, per telecom worker, then the "majority." Furthermore, we were able to establish that over 60% of the "majority" invest less in telecoms per capita than the "leaders." Consequently, based on the results of DT analysis we were able to conclude that the "leaders" do differ from the "majority" in terms of the inputs.

The second phase of our inquiry was directed at the determining whether or not the process by which the "leaders" convert the inputs into the outputs is significantly different from the process used by the "majority." We designated two types of the outputs representing the construct "TFP," first, MI and TC, and second, MI and EC. Then we used NN analysis to create a model of the process of the transformation that was based on the "leaders." Next, we have used the process of the "leaders" to predict the outputs given the inputs of the "majority." The predicted outputs of the "majority" were then substituted instead of the actual outputs and then used in SEM with PLS analysis.

The results have demonstrated that the "majority" exhibit statistically significant relationship between the constructs "ICT Capitalization" and "TFP," when "TFP" is represented by the predicted measures. Thus, we were able to conclude that the "majority" is capable, even with the unchanged levels of the inputs, of displaying statistically significant relationship between the constructs "ICT Capitalization" and "TFP." The prerequisite for this, however, would be the improvement of the process of converting the inputs into the outputs. We propose calling the process of conversion of the inputs into the outputs a "Transformative Capacity" of the TE.

At this point, we can only conjecture what some of the determinants of the transformative capacity are. However, based on the conducted research we could suggest some of the factors that possibly affect it. For example, one of the factors that we propose could be affecting the transformative capacity is the level of socio-cultural and economic openness of the country. This factor provides a necessary level of exposure and familiarity with what the "latest and greatest" in ICT has to offer. This factor, of course, must be complemented by the actual availability of the "latest and greatest" that manifests itself in the form of the, in the case of our research model, construct "ICT Diffusion."

Another factor that we suggest could be playing a role in the transformative capacity of the TEs is the level of socio-cultural and economic innovativeness. This factor, we conjecture, is required for the economy to try and eventually to adopt the "latest and greatest" that ICT has to offer. We would like to suggest that this factor must be complemented by the actual utilization of the "latest and greatest" in the form of the factor "ICT Utilization."

Overall, however, we conjecture that the actual process of transformation of inputs into the outputs follows the structure of the Technology Acceptance Model (TAM), where in order to become adopted technology must be perceived by the users as being easy to use and useful. In our case the perception of "ease of use" takes form of the ubiquity of the technology, which is affected by the suggested above level of socio-cultural and economic innovativeness, while usability of the technology, or at least the perception of the usability, is influenced by the proposed level of socio-cultural and economic innovativeness.

It would appear reasonable to think that despite the purpose of our model is to depict the relationship between the investments in ICT and TFP on the macroeconomic level, the actual process of transformation of the inputs takes place on the microeconomic level, namely, on the level of the individual users. Consequently, despite the possibility of depicting the transformative capacity on the global level, we suggest using the 'nested' approach, where our macroeconomic model incorporates the microeconomic model in the form of TAM, or TAM-like model.

We believe that such incorporation of the widely known and used model could provide the multiple benefits that we would be happy to discuss in our future research. As for now, our updated research model, with the incorporated model of the transformative capacity, is depicted in the Figure 23 below.

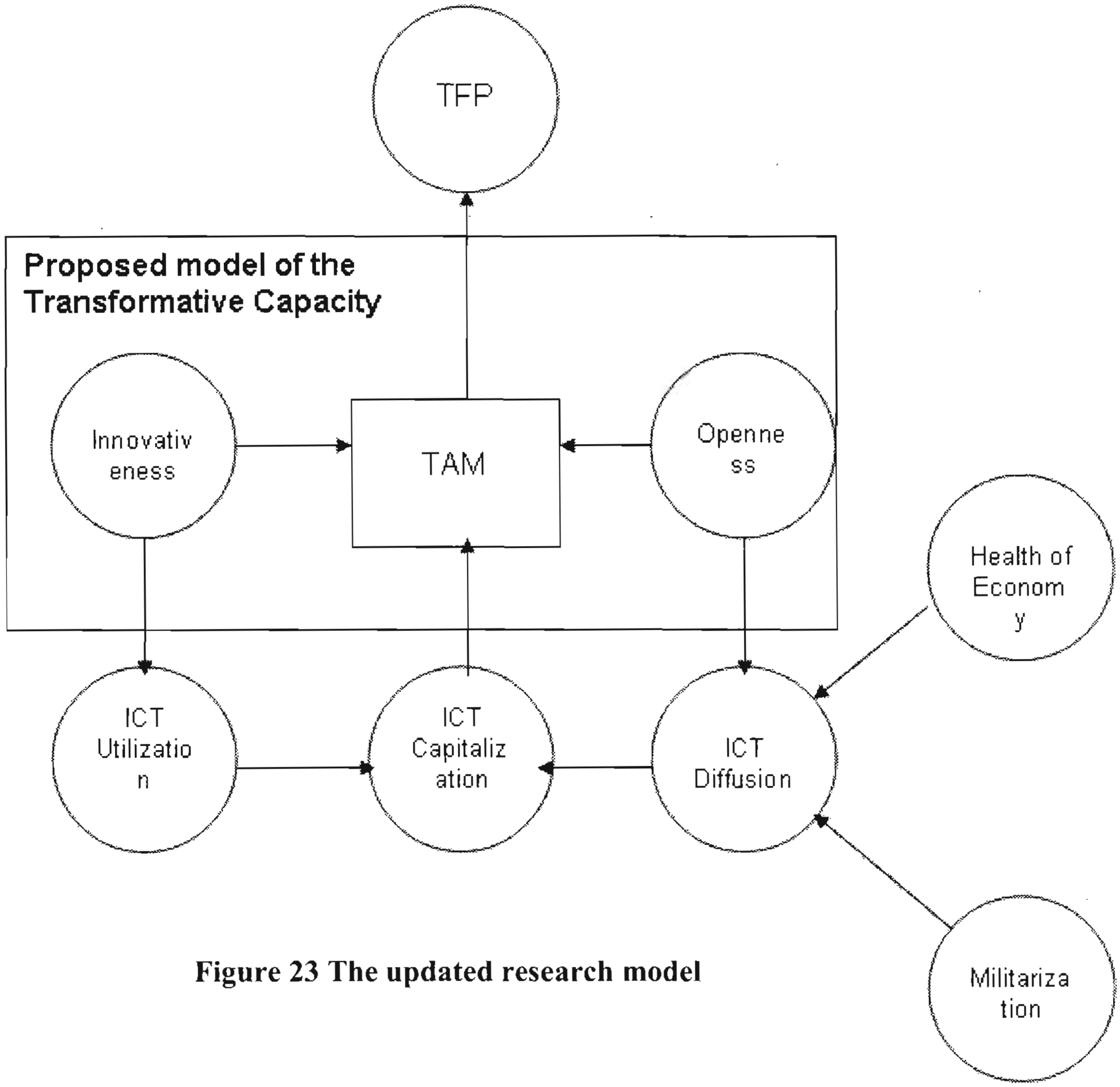


Figure 23 The updated research model

At this point, we would like to discuss some of the implications of the key findings for the policy makers.

8.1.4 Implications for Policy Makers

In this section of the conclusion, we discuss how the findings of our study might be of use to the policy makers.

First, the results of our study have demonstrated that the TEs should not be perceived as a homogenous group of economies. Consequently, a policy maker working in the area of the investments in ICT should be well served by identifying a subset of the TEs within which a given economy falls. It would be unwise to gauge the performance of the economy in comparison to the dissimilar group of TEs.

Second, we have established that the economies in transition are not static in terms of their performance. Meaning, some of the TEs change their membership from being part of the "majority" cluster to being part of the "leaders." Obviously, the reverse process is possible as well. Consequently, the policy makers for the "majority" should be pursuing the goal of not becoming the "best of the majority," but rather attempting to become a part of the "leaders."

Third, the results of the DEA suggest that performance of the TEs depends not only on the ability to effectively produce outputs, but as well on the ability to efficiently utilize the inputs. As a result, a policy maker would be well served by making sure that increase in performance of a TE does not come solely from the increase in the level of inputs.

Fourth, a policy maker should be aware of the presence of the complementary to ICT investments. Consequently, any policy maker in the area of ICT must take a broader perspective on the state of the investments in a given TE. From this vantage point the investments, revenues, and the level of performance of ICT should not be evaluated apart from the state of the complementary investments in other areas of the economy.

Fifth, when investing in ICT a policy maker should aware of the various drivers that affect the investments. For example, we have discussed that at the early stages the investments in ICT could be driven by the demand in infrastructure, then by the demand in utilization of the infrastructure, and so on. A policy maker would benefit from considering and identifying a specific driver, or drivers, that fuel the investments in ICT at the given time.

Sixth, the results of our study suggest the presence of the threshold number of the full-time telecom workers per given level of the investments. Consequently, while investing in ICT a policy maker should consider whether to correspondingly increase the number of the full-time telecom staff, or to increase the level of productivity of the current number. It would appear that the pattern is cyclical, for at some point increase in the investment in ICT must correspond with the increase in the number of telecom employees, and then with the increase in productivity of the employees until the certain threshold value is reached. After that point, the investment in ICT must be again accompanied by the increase in the number of the full-time telecom workers, and so on.

Seventh, the policy makers could benefit from identifying and explicitly stating whether the goal of the investments is maximization of the outputs, or the efficient use of the inputs. The results of DEA suggest that the expected outcome of the investments in ICT should be gauged against the appropriate goal, for the results could vary considerably depending on the orientation of the investments.

Finally, the results of NN analysis suggest that the policy makers in the less advanced countries could benefit from benchmarking of the best practices from their more successful counterparts, while, of course, taking into consideration the important differences of their respective settings.

8.2 Limitations of the study

Despite the theoretical and methodological contributions offered by this study, our research is not without its limitations. Among the commonly listed limitations of any study are those that are associated with the methods of data analysis and the data used in the study.

Fortunately, unlike a study that uses one or two data analytic methods, our multi-method approach is less reliant on a single statistical method, thus diminishing the limitations of our study associated with a given method or data analytic technique. While we are not suggesting that our research has no minor limitations in this area, we are confident in our thinking that no finding in this study could be invalidated due to our over reliance on a single assumption of any method that we used. The rationale for such reasoning is that not one of the six methods used in this study serves as a cornerstone of our multi-method approach.

Due to large number of methods used in our study, we present a single exemplar, demonstrating the limitations of one method, PLS, and its possible effect on our findings. Despite the multiple advantages, PLS methodology has some limitations associated with the estimation of a research model. For example, PLS estimation procedure could overestimate by up to 10 percent strengths of the measurement paths (the loadings of the measures on the constructs), while underestimating the strengths of the structural paths (the paths between the constructs) (Dijkstra 1983). While this is an inherent limitation of the method, it does not affect our results, because even considering 10 percent decrease in the values of the measurement paths the loadings of our items would end up being at or above 0.8, which is still more than acceptable. In terms of the structural paths, however, our study could only benefit from the underestimation provided by PLS; for what it means that we were able to establish the presence of the relationship between the constructs "TFP" and "ICT Capitalization" despite the underestimation, which only makes our case for reliability of the findings stronger.

Unfortunately, despite all the benefits that our multi-method approach could provide, we are defenseless against the problems associated with the quality of the available data. We could identify two limitations associated with the data that precluded us from obtaining better results.

The first limitation is the availability of the indicators in WDI Database and the "Yearbook of IT." Our original optimism, fueled by WDI Database' listing of 630 available indicators, has waned rather quickly when we realized that it does not necessarily implies presence of the data for the advertised indicators. For example, some of the most important indicators that we could have used for our study were not available for the TEs. Then some of the other indicators were available for some of the TEs, but not for the others. The same sentiment applies to the "Yearbook of IT," where scores of the important variables were not available for all of the countries in our sample. At this point, it was becoming clear what some of the reasons are responsible for scarcity of the research in the area associated with TEs. As a result, we were facing a dilemma: to decrease the size of our sample and have more variables, or to keep the sample as large as we could while being limited in the number of the available variables. We believe that in the end we were able to find a balance allowing us to have large enough sample and enough data to perform meaningful statistical analysis without sacrificing the power. However, our research could have been greatly improved in terms of the findings if more data were available to us.

The second limitation is the length of the time series that we use for this research. It could be argued that 10-year period of time is not sufficiently long for the true patterns of macroeconomic dynamics to play out and manifest themselves. We agree with this contention and might even revisit the topic of this study in 10 years, when 20-year period data becomes available. As for now, we must accept while getting the best out of what is available, admitting, nevertheless, that our research could have been greatly improved in terms of the findings if a longer time series data were available to us.

8.3 Directions for Future Research

Despite of the plethora of the results obtained during this study we consider that the most interesting findings are still lie ahead. Thus, our research should be perceived not as an end, but rather as the beginning of much wider and deeper inquiry into the matter of the relationship between the investments in ICT and TFP.

The main thrust of the future research could be identified as the one directed at the further development of the "ICT Investment Success Model," a preliminary version of which we developed and tested in this study. Earlier in this paper, we suggested the possibility that the investments ICT could be driven by different factors during the different phases of the economic development of the TEs. It would appear that at the initial stages, the investments are driven by the expanding infrastructure and the increasing demand associated with the utilization of the infrastructure by the users. Our study suggested that this was the case for the "majority." However, regarding the "leaders" the case still remains open; for it is not clear what drives the investments in ICT in their case. It would be interesting to take a closer look at this problem and to propose some form of the "Drivers of the ICT Investment" model.

Next, we suggest that the proposed and tested in this study model should be further tested in the different setting; it would be interesting to test whether the discovered relationship between the investments in ICT and TFP would hold in the context of the developing or the least developed countries.

Furthermore, we think that the proposed in this study model could be expanded to include what could be called a "Quality of Life" construct. It would allow us to inquire into how the level of personal wealth or, conversely, level of poverty, of the population affects and is affected by the investments in ICT.

Another topic worth further researching pertains to the level of the impact of the complementary investments on ICT. Interpretation of the interaction terms is not straightforward, for the magnitude and the direction of the effect are dependent of the levels of the variables in the interaction term. As a result, it is possible that the values of the complementary investments would have some sort of threshold values. Thus, it is important to determine at what point the benefits provided by the complementary investments start to diminish.

Finally, we suggest that the proposed "Transformative Capacity" model is worth further investigating, for it is important to inquire into the process and the mechanism by which the investments in ICT are actually being transformed in the components of TFP.

8.4 Final Conclusion

Ever decreasing prices make the investments in ICT look more and more attractive and increasingly affordable for the investors and policy makers alike. The outcome of the investing in ICT, however, was not clear cut, for some economies were able to benefit from the investments, while other economies were not. The goal of our study was to investigate whether or not the investments in ICT relate to the changes in

TFP in the context of the Economies in Transition. We were able to determine that such relationship do exist, but only for the group of the countries that we named "leaders."

The second objective of this study was to investigate whether or not the investments in ICT have complementary effect with the other investments, capital- or labor-related. We were able to demonstrate the existence of the subset of the investments that appear to be complementary to the investments in ICT. Thus, it would appear that we were able successfully accomplish the outlined goals.

Moreover, during the course of our inquiry we have obtained a large number of findings and insights regarding the matter of the investments in ICT and the relationship between the investments and macroeconomic growth. Despite the plethora of the results provided by this study, we still consider the most interesting and important findings to lie ahead.

We hope, therefore, that the results of our research would serve as a springboard for the new studies to be conducted in this area, for the new vantage points to be considered and new insight to be gained.

Literature Cited

TFP in the context of the Economies in Transition. We were able to determine that such relationship do exist, but only for the group of the countries that we named "leaders."

The second objective of this study was to investigate whether or not the investments in ICT have complementary effect with the other investments, capital- or labor-related. We were able to demonstrate the existence of the subset of the investments that appear to be complementary to the investments in ICT. Thus, it would appear that we were able successfully accomplish the outlined goals.

Moreover, during the course of our inquiry we have obtained a large number of findings and insights regarding the matter of the investments in ICT and the relationship between the investments and macroeconomic growth. Despite the plethora of the results provided by this study, we still consider the most interesting and important findings to lie ahead.

We hope, therefore, that the results of our research would serve as a springboard for the new studies to be conducted in this area, for the new vantage points to be considered and new insight to be gained.

Literature Cited

1. Ackoff, R. L. (1999). "Ackoff's Best. His Classic Writings on Management".New York: John Wiley & Sons.

2. Adams, Richard H., Jr. (2002). "Economic Growth, Inequality, and Poverty: Findings from a New Dataset", *World Bank Policy Research Working Paper No. 2972*, Washington, D.C.: World Bank, February.

3. Adler, N., Friedman, L., Sinuany-Stern, Z., (2002). "Review of ranking methods in data envelopment analysis context", *European Journal of Operational Research*, vol.140, No.2, pp. 249–265.

4. Aiken, L. and West, S., (1991), "Multiple regression: Testing and interpreting interactions". Newbury Park, CA: Sage Publications.

5. Ali, A.I., (1994). "Computational aspects of DEA." In: Charnes, A., Cooper, W.W., Lewin, A., Seiford, L.M. (Eds.), "Data Envelopment Analysis: Theory, Methodology and Applications". Kluwer Academic Publishers, Boston, pp. 63–88.

6. Arvantis, S. (2003), "Information Technology, Workplace Organization, Human Capital and Firm Productivity: Evidence for the Swiss Economy", *KOF-Arbeitspapiere/ Working Paper No. 70*, February, Zurich.

7. Athey, S. and Schmutzler, A. (1995). "Product and Process Flexibility in an Innovative Environment", *RAND Journal of Economics*, The RAND Corporation, vol.26, No.4, pp. 557-574.

8. Avgerou, C. (1998). "How Can IT Enable Economic Growth in Developing Countries?" *Information Technology for Development*, vol.8, pp. 15-28.

9. Baliamoune, Mina N. (2002). "The New Economy and Developing Countries: Assessing the Role of ICT Diffusion", United Nations University World Institute for Development Economics Research, Finland: Helsinki [A revised version of this paper was published in *Information Technology for Development*, vol.10, No.3 (2003), pp.151-169].

10. Baliamoune-Lutz, M. (2003). "An Analysis of the Determinants and Effects of ICT Diffusion in Developing Countries", *Information Technology for Development*, vol.10, pp.151-169.

11. Barclay, D., Thompson, R. and Higgins, C. (1995) "The Partial Least Squares (PLS) Approach to Causal Modeling: Personal Computer Adoption and Use an Illustration," *Technology Studies* , vol.2, No.2, pp. 285-309.

12. Bishop, C.M. (1995), "Neural Networks for Pattern Recognition", Oxford: Oxford University Press.

13. Black, S. and Lynch, L. (1997). "How to compete: The impact of workplace practices and information technology on productivity", *Working paper National Bureau of Economic Research*, Cambridge, MA.

14. Bowlin, W. F. (1998). "Measuring Performance: An Introduction to Data Envelopment Analysis (DEA)", *Journal of Cost Analysis*, Fall 1998, pp. 3-27.

15. Braumoeller, B., (2004), "Hypothesis Testing and Multiplicative Interaction Terms," *International Organization,* vol. 58, no. 4, pp. 807-820.

16. Breiman, L., Friedman, J., Olshen, R., and Stone, C., (1984) "Classification and Regression Trees" Wadsworth International Group.

17. Bruno, M., Ravallion, M. and Squire, L. (1998), "*Equity and Growth in Developing Countries: Old and New Perspectives on the Policy Issues*", Vito Tani and Ke-Young Chu, editors, "Income Distribution and High Growth", Cambridge, MA: MIT Press.

18. Brynjolfsson, E. and Hitt, L. (1996). "Paradox lost? Firm-level Evidence on the Returns to Information Systems Spending ", *Management Science*, vol. 42, no.4, pp. 541-558.

19. Brynjolfsson, E. and Hitt, L. (1997). "Computing Productivity: Are Computers Pulling Their Weight?", mimeo MIT and Wharton, available at *http://ccs.mit.edu/erik/cpg/*

20. Brynjolfsson, E. and Hitt, L. (2000). "Beyond Computation: Information Technology, Organizational Transformation and Business Performance", *Journal of Economic Perspectives,* vol.14, no. 4, pp. 23-48.

21. Brynjolfsson, E. and Yang, S. (1996). "Information Technology and Productivity: A Review of the Literature", *Advances in Computing,* vol. 43, pp. 179–214.

22. Brynjolfsson, E.(1993). "The Productivity Paradox of Information Technology: Review and Assessment", Communications *of ACM*, December, vol.36, no.12, pp. 67-77.

23. Bugamelli, M. and Pagano, P. (2004). "Barriers to investment in ICT", *Applied Economics, Taylor and Francis Journals*, vol. 36, no.20, pp. 2275-2286.

24. Canning, D. and Pedroni, P. (1999). "Infrastructure and Long Run Economic Growth", Consulting Assistance on Economic Reform II Discussion Paper No.57 Harvard Institute for International Development. Available at http://scholar.google.com/url?sa=U&q=http://www.arts.cornell.edu/econ/cae/infrastructure-7-99.pdf

25. Caves, D.W., Christensen, L.R. and Diewert, W.E. (1982)."The Economic Theory of Index Numbers and the Measurement of Input, Output, and Productivity" ,*Econometrica*, vol.50,pp. 1393-1414.

26. Cette, G., Mairesse, J. and Kocoglu, Y. (2001). "The contribution of information and communication technology to French economic growth", Bank of

France and Université de la Méditerranée paper. Available at http://www.bis.org/publ/cgfs19bdf2.pdf

27. Cette, G., Kokoglu, Y. and Mairesse, J. (2000). "The Diffusion of Information and Communication Technologies in France. Measurement and Contribution to Growth and Productivity", *Economie et Statistique*, no. 339-340.

28. Charnes, A. and Cooper, WW. (1962). "Programming with linear fractional functionals", *Naval Research Logistics Quarterly*, vol.9, pp.181–186.

29. Charnes, A., Cooper, W.W. and Rhodes, E. (1978). "Measuring the efficiency of decision making units", *European Journal of Operational Research.*, vol.2, pp.429-444.

30. Charnes, A., Cooper, W.W., Lewin, A.Y. and Seiford, L.M. (1994), "Data Envelopment Analysis: Theory, Methodology and Applications", Kluwer Academic Publishers, Norwell, MA.

31. Chay, K. Y., Powell, J. L., (2001). "Semiparametric Censored Regression Models", *Journal of Economic Perspectives*, vol.15,pp. 29–42.

32. Chin, W. W. (1995). "Partial Least Squares is to LISREL as principal components analysis is to common factor analysis" in *Technology Studies*, 2, pp. 315-319.

33. Chin, W.W. (1998a) The Partial Least Squares Approach to Structural Equation Modeling. In G. A. Marcoulides(Ed.), "Modern Methods for Business Research", Mahwah, NJ: Lawrence Erlbaum Associates, pp. 295-336.

34. Chin, W. W. (1998b) PLS-Graph (Version 03.00 build 1126).

35. Chin, W. W., (2000). "Frequently Asked Questions – Partial Least Squares & PLS-Graph. Home Page"[On-line]. Available: http://disc-nt.cba.uh.edu/chin/plsfaq.htm

36. Chin, W. and Hubona, G.S.(2006). "Structural Equation Modeling (SEM) Using PLS-Graph Software: A Tool For Quantitative Researchers Analyzing Path-Based Models," (forthcoming) tutorial for AMCIS conference, August 2006.

37. Christensen, R., Jorgenson, D. and Lau, L. (1973). "Transcendental Logarithmic Production Frontiers", *The Review of Economics and Statistics*, MIT Press, vol. 55, no.1, pp. 28-45.

38. Christiaensen, L., Demery, L. and Paternostro, S. (2002). "Growth, Distribution, and Poverty in Africa", World Bank: Washington, DC.

39. Colecchia, A. and Schreyer, P. (2001). "The Impact of Information Communications Technology on Output Growth", STI Working Paper 2001/7, OECD, Paris.

40. Colecchia, A., and Schreyer, P. (2002). "ICT Investment and Economic Growth in the 1990s: Is the United States a Unique Case? A Comparative Study of Nine OECD Countries" *Review of Economic Dynamics,* vol. 5, pp. 408-442.

41. Cooper, W.W., Seiford, L.M. and Zhu, Joe. (2004). "Data envelopment analysis: History, Models and Interpretations", in "Handbook on Data Envelopment Analysis", eds W.W. Cooper, L.M. Seiford and J. Zhu, Chapter 1, pp.1-39, Kluwer Academic Publishers, Boston. 2004.

42. Council of Economic Advisors. (2001)." The annual report of the council of economic advisors". In *The Economics of the President.* U.S. Government Printing Office, Washington, D.C.

43. Daly, J.(2003)." *ICT, Economic Growth and Poverty Reduction"* [online] *Development Gateway* , available at http://topics.developmentgateway.org/ict/sdm/previewDocument.do~activeDocumentId=495495.

44. Daveri, F. (2000). "Is Growth an Information Technology Story in Europe too?" Working Paper, University of Parma, Parma, Italy.

45. Daveri, F. (2002), "The New Economy in Europe, 1992-2001", *Oxford Review of Economic Policy*, vol. 18, no.3, pp. 345-362.

46. Dedrick, J. and Kraemer, K.. (2001). "The Productivity Paradox: Is it Resolved? Is There a New One? What Does It All Mean for Managers", Center for Research on Information Technology and Organizations, Irvine, February 2001. available at http://www.crito.uci.edu/itr/publications/pdf/it_prod_paradox.pdf

47. Dedrick, J., Gurbaxani, V., and Kraemer, K. (2003). "Information technology and economic performance: A critical review of the empirical evidence". *ACM Computing Survey,* vol.35, no,1 (March 2003), pp.1-28.

48. Demery, L. Christiansen, L., and Paternostro, S. (2002). "Growth, Distribution and Poverty in Africa – Messages from the 1990s", World Bank Policy Research Working Paper No. 2810, World Bank, Washington DC. Available at http://ssrn.com/abstract=636093

49. Devaraj, S. and Kohli, R. (2000). "Information technology payoff in the health care industry: A longitudinal study", *Journal of Management Information Systems*, vol.16, no.4, pp. 41-67.

50. Devaraj, S. and Kohli, R. (2003). "Performance Impacts of Information Technology: Is Actual Usage the Missing Link?", *Management Science*, vol.49, no.3, pp.273-289.

51. Dewan, S. and Kraemer, K. (1998). „International dimensions of the productivity paradox", *Communications of the ACM,* vol. 41, no.8 (Aug.), pp.56–62.

52. Dewan, S. and Kraemer, K. (2000). "Information Technology and Productivity: Evidence from Country Level Data", Management *Science*, special issue on the Information Industries, vol. 46, no.4 (April), pp.548–562.

63. Gera, S. and Wulong, Gu. (2004)."The Effect of Organizational Innovation and Information and Communications Technology on Firm Performance", *International Productivity Monitor*, Centre for the Study of Living Standards, vol. 9, pp. 37-51.

64. Gilchrist, S., Gurbaxani, V. and Towne R. (2001). "Productivity and the PC revolution", UC Irvine Working Paper.

65. Giuri, P., Torrisi S. and Zinovyeva, N. (2005). "ICT, Skills and Organizational Change: Evidence from a Panel of Italian Manufacturing Firms" ,LEM Papers Series 2005/11, Laboratory of Economics and Management (LEM), Sant'Anna School of Advanced Studies, Pisa, Italy.

66. Gurbaxani V., Melville N. and Kraemer K. (1998). "Disaggregating the Return on Investment to IT Capital", In *Proceedings of the international Conference on information Systems* (Helsinki, Finland). International Conference on Information Systems. Association for Information Systems, Atlanta, GA, 376-380.

67. Hernando, I. and Nunez, S., (2002). "The Contribution of ICT to Economic Activity: A Growth Accounting Exercise with Spanish Firm-Level Data" , Banco de España — Servicio de Estudios Documento de Trabajo No. 0203, available at http://www.bde.es/informes/be/docs/dt0203e.pdf

68. Hitt L. and Brynjolfsson, E. (1996). "Productivity, business profitability, and consumer surplus: Three different measures of information technology value", *MIS Quarterly,* vol.20, no.2, pp.121–142.

69. IMF (2001). International Financial Statistics. Washington, DC: IMF.

70. Jaccard, J., Turrisi, R. and Wan, C., (1990), "Interaction effects in multiple regression". Thousand Oaks, CA: Sage Publications. Series: Quantitative Applications in the Social Sciences,No.72

53. Dijkstra, T. (1983). "Some Comments on Maximum Likelihood and Partial Least Squares Methods." In *Journal of Econometrics*, vol. 22, pp. 67-90.

54. Dula, J.H., (2002a),"Data Envelopment Analysis (DEA)," in "Handbook of Applied Optimization", P.M. Pardalos and M.G.C. Resende, (eds.), pp. 531--543, Oxford University Press, New York, 2002. ISBN 0-19-512594-0.

55. Dula, J.H., (2002b), "Computations in DEA", *Pesquisa Operacional*, vol. 22, no.2, pp.165-182, Dec. 2002, ISSN 0101-7438.

56. Edgeworth, F.Y. (1881). "Mathematical Psychics", London: Kegan Paul.

57. Eggleston, K., Jensen, R. and Zeckhauser, R. (2002). "Information and Communication Technologies, Markets and Economic Development," Discussion Papers Series, Department of Economics, Tufts University 0203, Department of Economics, Tufts University.

58. Fare, R., Grosskopf, S., Norris, M., and Z. Zhang. (1994). "Productivity Growth, Technological Progress, and Efficiency in Industrialized Countries", *American Economic Review*, vol.84, pp. 374-380.

59. Fornell, C. (1987)."A Second Generation of Multivariate Analysis: Classification of Methods and Implications for Marketing Research" in Houston, M.J. (Ed). "Review of Marketing", Chicago, Illinois: American Marketing Association, pp. 407-450.

60. Foy, T. (2003), "Agriculture and Pro-Poor Growth" , Paper presented to OECD DAC POVNET Agriculture and Pro-Poor Growth, Florence Workshop, 7-9 October 2003, OECD:Paris.

61. Francalanci, C. and Galal, H. (1998). "Information technology and worker composition: Determinants of productivity in the life insurance industry", *MIS Quarterly,* Vol. 22, No. 2, pp.227–241.

62. Gefen, D., Straub, D. and Boudreau, M. (2000) "Structural Equation Modeling Techniques and Regression: Guidelines for Research Practice,"in *Communications of the Association for Information Systems* , vol.7, no.7, pp. 1-78.

71. Jalava, J., and Pohjola, M. (2001). "Economic Growth in the New Economy", WIDER Discussion Paper 2001/5. Helsinki: UNU/WIDER

72. Jalava, J., and Pohjola, M. (2002). "Economic growth in the New Economy: evidence from advanced economies," *Information Economics and Policy*, Elsevier, vol. 14, no.2, pp. 189-210.

73. Jorgenson, D.W. (2001). "Information Technology and the US economy," *American Economic Review*, vol.91, March, pp.1-32.

74. Jorgenson, D.W. and Stiroh, K.J. (2000). "Raising the Speed Limit: US Economic Growth in the Information Age", *Brookings Papers on Economic Activity*, vol.2.no.1, pp. 125-211.

75. Jorgenson, D.W. and Stiroh, K.J. (1995), "Computers and Growth", *Economics of Innovation and New Technology*, vol. 3, pp.295-316.

76. Kegels, C., van Overbeke, M. and Van Zandweghe, W. (2002). "ICT contribution to economic performance in Belgium: preliminary evidence – Revision of WP 7-02," Brussels: Federal Planning Bureau Working Paper 8-02.

77. Kothari, R. and Dong, M. (2001), "Decision Trees for Classification: A review and some new results," in *Lecture Notes in Pattern Recognition*, S.R. Pal and N.R. Pal, (Eds.), Singapore, 2001, World Scientific Publishing Company.

78. Kraemer, K. L., and Dedrick, J. (2001). 'Information Technology and Economic Development: Results and Policy Implications of Cross-Country Studies', in M. Pohjola (ed.) "Information Technology, Productivity, and Economic Growth", Oxford: Oxford University Press.

79. Kraemer, K. L. and J. Dedrick. (1994). "Payoffs from investment in information technology: Lessons from Asia-Pacific region", *World Development,* vol.22, no.12, pp.19–21.

80. Lambertini, L. (2003). 'The monopolist's optimal R&D portfolio', *Oxford Economic Papers*, vol.55, pp.561–78.

81. Lee, Il-Houng and Khatri, Y. (2001). "The Role of the New Economy in East Asia," (unpublished; Washington: International Monetary Fund, Asia and Pacific Department). Available at http://www.er.cna.it/biblioteca/files/2005060616255620050514010351O3WEO_2001_09_11__01_09_2610.pdf

82. Lehr, B. and Lichtenberg, F. (1999). „Information Technology and its Impact on Productivity: Firm- Level Evidence from Government and Private Data Sources, 1977-1993", *Canadian Journal of Economics*, vol.32, no.2, pp.335-362.

83. Lichtenberg, F. (1995). "The Output Contributions of Computer Equipment and Personal: A Firm-Level Analysis", *Economics of Innovation and New Technology*, vol. 3, pp. 201-217.

84. Lin, P. and Saggi, K. (2002). "Product Differentiation, Process R&D, and the Nature of Market Competition", *European Economic Review*, vol.46, pp.201-211.

85. Lins, MPE, Gomes, EG, de Mello, JCCBS, de Mello, AJRS, (2003), "Olympic ranking based on a zero sum gains DEA model", *European Journal Of Operational Research*, vol. 148, no.2, pp 312 – 322.

86. Loukis,E and Sapounas, I . (2004). "The impact of information systems investment and management on business performance in Greece", ECIS 2005 Conference paper.

87. Loveman, G.(1994). "An Assessment of the Productivity Impact of Information Technologies," in Allen, Thomas J. and Scott Morton, Michael S. (Ed.), "Information Technology and the Corporation of the 1990s: Research Studies", Oxford University Press, pp. 84-110.

88. Madden, G. and Savage, S. (1998). "CEE Telecommunications Investment and Economic Growth", *Information Economics and Policy*, no. 10, pp. 173–195.

89. Malmquist,S.(1953). "Index numbers and indifference surfaces", *Trabajos de Estatistica*, vol.4, pp.209–242.

90. Martinez-Frias, J. (2003). "The Importance of ICTs for Developing Countries", *Interdisciplinary Science Reviews*, 2003, vol. 28, no. 1.

91. Melka, J., , Mulder, N., Nayman, L. and Zignago, S.(2003). "Skills, Technology and Growth is ICT the Key to Success? An Analysis of ICT Impact on French Growth", Working Papers 2003-04, CEPII research center.

92. Milana C. and Zeli, A. (2002). "The Contribution of ICT to Production Efficiency in Italy - Firm-Level Evidence Using Data Envelopment Analysis and Econometric Estimations", STI Working Paper 2002/13, OECD, Directorate for Science, Technology and Industry, DSTI/DOC(2002)13, Paris, France, 2002.

93. Morales-Gomez, D. and Melesse, M. (1998). "Utilizing Information and Communication Technologies for Development: The Social Dimensions", *Information Technology for Development*, vol.8, pp. 3-13.

94. Motohashi, K. (1997). "ICT Diffusion and its Economic Impacts on OECD countries", *OECD Science, Technology and Industry Review,* no. 20, OECD, Paris.

95. Murakami, T. (1997). "*The Impact of ICT on Economic Growth and the Productivity Paradox*" available at http://www.tcf.or.jp/data/19971011_Takeshi_Murakami_2.pdf

96. Niininen, P. (2001). "Computers and Economic Growth in Finland", in M. Pohjola (ed.), "Information Technology, Productivity, and Economic Development", UNU/WIDER Studies in Development Economics. Oxford: Oxford University Press.

97. OECD (2001). The New Economy: Beyond the Hype, Paris: OECD.

98. OECD (2003). *The Sources of Economic Growth in OECD Countries*, Paris: OECD.

99. OECD (2003b). "ICT and Economic Growth: Evidence from OECD Countries, Industries and Firms", Paris: OECD.

100. OECD (2003c). "Integrating Information and Communication Technologies in Development Programmes", Policy Brief, Paris:OECD

101. OECD (2004). "DAC Network on Poverty Reduction: ICTs and economic growth in developing countries", Paris: OECD. Available at http://www.oecd.org/dataoecd/15/54/34663175.pdf

102. OECD (2005a). "Good Practice Paper on ICTs for Economic Growth and Poverty Reduction", *The DAC Journal 2005*, vol. 6, no. 3.

103. OECD (2005b). "Background Paper: The Contribution of ICTs to Pro-Poor Growth: No. 384", OECD Papers, vol. 5, no. 2, July 2005, pp.15-52.

104. OECD (2005c). "The Contribution of ICTs to Pro-Poor Growth :No. 379", OECD Papers, vol. 5, no. 1, March 2005, pp. 59-72.

105. Oliner, S.D. and Sichel, D.E. (2000). "The Resurgence of Growth in the late 1990s: Is Information Technology Story?", *Journal of Economic Perspectives*, vol.14, no.4, pp.3-22.

106. Oliner, S. and Sichel, D. (1994). "Computers and Output Growth Revisited: How Big is the Puzzle?" *Brookings Papers on Economic Activity*, 1994, vol.2, pp. 273-334.

107. Oliner, S. and Sichel, D.(2002). "Information Technology and Productivity: Where are We Now and Where are We Going? ", *Economic Review*, vol.3, no.3, pp. 15-41.

108. Osei-Bryson, K.-M. (2004) "Evaluation of Decision trees: A Multi-Criteria Approach" in *Computers & Operations Research*, vol. 31, no. 11, pp. 1933-1945.

109. Piatkowski, M. (2002). "The 'New Economy' and Economic Growth in Transition Economies", WIDER Discussion Paper No. 2002/63, Helsinki:WIDER.

110. Piatkowski, M. (2003a). "The Contribution of ICT Investment to Economic Growth and Labor Productivity in Poland 1995-2000" *TIGER Working Paper Series* No. 43, Warsaw, July. Available at www.tiger.edu.pl

111. Piatkowski, M. (2003b). "Does ICT Investment Matter for Output Growth and Labor Productivity in Transition Economies?" *TIGER Working Paper Series*, No. 47. December. Warsaw. Available at www.tiger.edu.pl

112. Piatkowski, M. (2004). "The Impact of ICT on Growth in Transition Economies", *TIGER Working Paper Series, No. 59*. July. Warsaw. From: http://www.tiger.edu.pl/publikacje/TWPNo59.pdf

113. Pearson, K. (1908). "On the generalized probable error in multiple normal correlation" in *Biometrika*, vol.6, pp. 59-68.

114. Pohjola, M. (2000). "Information Technology and Economic Growth: A Cross-Country Analysis", United Nations University World Institute for Development Economics Research (UNU/WIDER), Finland:Helsinki.

115. Pohjola, M. (2002)."New Economy in Growth and Development", WIDER Discussion Paper 2002/67, United Nations University World Institute for Development Economics Research (UNU/WIDER). Finland:Helsinki.

116. Pohjola, M. (ed.) (2001). "Information Technology, Productivity and Economic Growth: International Evidence and Implications for Economic Development", WIDER Studies in Development Economics, Oxford University Press, United Kingdom:Oxford.

117. Quinlan, R. J. (1986), "Induction of Decision Trees" in *Machine Learning*, vol.1, pp 88-106.

118. Quinlan, R. J. (1993), "C4.5: Programs for Machine Learning", Morgan Kaufmann Publishers, San Mateo, CA, 1993.

119. Quinlan, R.J. (1998). "*C5.0: An Informal Tutorial*", available on line at http://www.rulequest.com/see5-unix.html.

120. Rai, A, Patnayakuni, R and Patnayakuni, N. (1996) "Refocusing where and how IT value is realized", *Omega: The International Journal of Management Science*, vol.24, no.4, pp.399-412.

121. Rai, A, Patnayakuni, R and Patnayakuni, N. (1997). "Technology investments and business performance," *Communications of the ACM*, vol. 40, no.7, pp.89-97.

122. Ramirez, R. (2003). "The Influence of Information Technology and Organizational Improvement Efforts on the Performance of Firms . Doctoral Dissertation, University of California.

123. Ravallion, M. (2001). "Growth, Inequality, and Poverty: Looking beyond Averages", Policy Research Working Paper Series 2558, The World Bank.

124. Roach, S. (1987). "America's Technology Dilemma: A Profile of the Information Economy," Morgan Stanley Special Economic Study, April.

125. Roach, S. (1989). "America's White-Collar Productivity Dilemma", *Manufacturing Engineering*, August , pp. 104.

126. Roach, S. (1991). "Services under Siege: the Restructuring Imperative," *Harvard Business Review*, vol.39, no.2, pp. 82-92.

127. Röller, L.-H, and Waverman, L. (1994). "The Impact of Telecommunications Infrastructure on Economic Growth and Development", OCED ICCP Draft Working Paper, France:Paris.

128. Rosenkranz, S. (2003). "Simultaneous choice of process and product innovation when consumers have a preference for product variety", *Journal of Economic Behavior & Organization*, vol.50, no. 2, pp.183-201.

129. Rumelhart, D. E., & McClelland, J. L. (Eds.). (1986). "Parallel distributed processing: Explorations in the microstructure of cognition. Volume 1: Foundations". Cambridge, MA: MIT.

130. Schreyer, P. (2000). "The Contribution of Information and Communication Technology to Output Growth: a Study of the G7 Countries", *STI Working Papers* 2000/2, OECD, Paris.

131. Sengupta JK. (1996). "Data Envelopment analysis: a new tool for improving managerial efficiency", *International Journal of Systems Science*, vol.27, no.12, pp.1205–1210.

132. Shao, B. and Lin, W. (2002). "Technical efficiency analysis of information technology investments: a two-stage empirical investigation", *Information and Management*, vol. 39, pp. 391-401.

133. Solow R. (1960). " Investment and Technological Progress ", in K. Arrow, S. Karlin and P.Suppes eds., "Mathematical Methods in the Social Sciences 1959", Stanford, CA: StanfordUniversity Press, 1960.

134. Solow, R. (1957). 'Technical Change and the Aggregate Production Function', *Review of Economics and Statistics*, vol. 39, no. 3, pp. 312–20, August.

135. Soteriou, A. C. and Zenios, S. A. (1998). "An Introduction to Data Envelopment Analysis and an Application to Assessing the Performance of Bank

Branches" ,in G. Marcoulides (eds.), "Modern Methods for Business Research", Mahwah, NJ: Lawrence Erlbaum Associates, Inc. Publishers, pp. 125-142.

136. Stiroh, K. (2001). "Investing in Information Technology: Productivity Payoffs for US Industries ", *Current Issues in Economics and Finance*, FRBNY, vol. 7, no. 6, June.

137. Stiroh, K. (2002). "Information Technology and the U.S. Productivity Revival: What Do the Industry Data Say?" American *Economic Review*, American Economic Association, vol. 92, no.5, pp. 1559-1576.

138. Strassman, P. (1997). "The Squandered Computer". The Information Economic Press, New Canaan, Connecticut.

139. Strassmann, P. (1990). "The Business Value of Computers: An Executive's Guide." New Canaan, CT, Information Economics Press.

140. Tallon, P., Kraemer, K. and Gurbaxani, V. (2000). "Executives perspectives on the business value of information technology", *Journal of Management Information Systems,* vol. 16, no,4, pp.145–173.

141. Tobin, J. (1958). "Liquidity Preference as Behavior Towards Risk", *The Review of Economic Studies*, No. 67, February, pp. 65-86. Reprinted as Cowles Foundation Paper 118.

142. Tryon, R.C. (1939). "Cluster analysis" ,Edwards Brothers, Ann Arbor, MI.

143. Tuomi, I. (2004). "Knowledge Society and the New Productivity Paradigm: A Critical Review of Productivity Theory and the Impacts of ICT", IPTS Working Paper, February 2004.

144. Van Ark, B. and Piatkowski, M. (2004). "Productivity, Innovation and ICT in Old and New Europe", *International Economics and Economic Policy*, Springer-Verlag, vol. 1, Special Issue no. 2+3, pp. 215-246.

145. Van Ark, B., Melka, J., Mulder, N., Timmer, M. and Ypma, G. (2002). "ICT Investments and Growth Accounts for the European Union, 1980-2000", Research Memorandum GD-56, Groningen Growth and Development Centre, Groningen, *www.eco.rug.nl/ggdc/homeggdc.html*

146. Vijselaar, F. and Albers, R. (2002). "New Technologies and Productivity Growth in the Euro Area", European Central Bank Working Paper N°122. available at http://www.ecb.int/pub/wp/ecbwp122.pdf

147. Whelan, K. (1999). "Tax incentives, material inputs, and the supply curve for capital equipment", Federal Reserve Board. Finance and Economics Discussion Series Paper 1999-21. April. Available at www.federalreserve.gov/pubs/feds/1999/199921/199921pap.pdf

148. Whelan, K. (2000). "Computers, obsolescence, and productivity", Federal Reserve Board. Finance and Economics Discussion Series Paper 2000-6. February. Available at www.federalreserve.gov/pubs/feds/2000/200006/200006pap.pdf

149. Wold, H. (1966). "Estimation of principal components and related models by iterative least squares", in "Multivariate Analysis", P.R. Krishnaiah, ed. Academic Press, New York, 391-420.

150. Wold, H. (1982). "Systems under indirect observation using PLS" In C. Fornell (Ed.), "A Second Generation of Multivariate Analysis", Volume I: Methods (pp. 325-347). New York: Praeger.

151. Wong, P.K. (1998). "Leveraging the Global Information Revolution for Economic Development: Singapore's Evolving E-information Industry Strategy", *Information Systems Research*, vol. 9, no. 4, pp.323-341.

152. Yosri, A. (1992). “The relation between information technology expenditures and revenue contributing factors in large corporations”, Doctoral Dissertation, Walden University.

153. Yule, G. U. (1907), “On the theory of correlation for any number of variables, treated by a new system of notations”, in *Proceedings of Royal Society*, vol. 79A, pp. 182-93.

VITA

Sergey V. Samoilenko

Education

PhD Information Systems, Minor in Decision Sciences (Defended May 5, 2006)
Department of Information Systems, School of Business,
Virginia Commonwealth University, Richmond, VA

M.S. in Information Systems (2002)
Department of Information Systems, School of Business
Virginia Commonwealth University, Richmond, VA

B.S. in Industrial Engineering (1990)
Institute of Soviet Trade Technology, St. Petersburg, Russia

Research Interest

Information Systems Economics
Decision Models for IS Management
Information Technology and Development
Information Technology and Transitional Economies

Teaching Interest

Systems Analysis and Design
Database
Business Process Analysis and Redesign
Research Methods for Information Systems

Brief Biographical Information

Sergey Samoilenko was born in Russia (former U.S.S.R.), city of St. Petersburg (former Leningrad) in 1966. Arrived in the U.S.A. in 1991 and became a citizen of the U.S.A. in 2001. Current place of residence Richmond, Virginia.