WHAT ARE THE POTENTIAL COVARIATES OF TOTAL FACTOR PRODUCTIVITY?

(AN APPLICATION OF ENCOMPASSING AND LASSO TECHNIQUE)

Pakistan Institute of Development Economics

By

IJAZ AHMAD
PROVPIDE2019FMPHILETS13

Supervisor

Dr. Hafsa Hina

MPhil Econometrics

PIDE School Economics

Pakistan Institute of Development Economics, Islamabad

2023

Pakistan Institute of Development Economics

Pakistan Institute of Development Economics, Islamabad

PIDE School of Economics

CERTIFICATE

This is to certify that this thesis entitled **"What are the potential Covariates of Total Factor Productivity? (An Application of Encompassing and LASSO Technique)"** submitted by **Mr. Ijaz Ahmad** is accepted in its present form by the School of Economics, Pakistan Institute of Development Economics (PIDE), Islamabad as satisfying the requirements for partial fulfillment of the degree in Master of Philosophy in Econometrics.

Supervisor:	Dr. Hafsa Hina	Signature: ____________
External Examiner:	Dr. Tanveer ul Islam	Signature: ____________
Head, PIDE School of Economics:	Dr. Shujaat Farooq	Signature: ____________

Acknowledgements

In the name of ALLAH, the most compassionate, without Whose blessing I would never been able to do this work.

First and foremost, I would like to give my cordial gratitude to my venerated supervisor

Dr. Hafsa Hina for her kind guidance, inspiring criticism, and valuable suggestions. Without her supervision I would never been able to do this research work. He provided me new insight intoresearch and widened my horizon.

There is a contribution of many people in completing my studies and research work, without their encouragement and support, this would have been a difficult task for me.

Very special thanks to **Wagma Naseer** for their support, care, patience, guidance, love and who blessed me with unforgettable memories and contributed a lot in my career building.

My last words of appreciation for my Parents, without them I would never be what I am today. My **father** always encourage and supports me. No word can express my thanks to my **loving parents, my brothers, my sisters** who's love, affection, prayers, care, and support help me not only during my studies but throughout my life

Ijaz Ahmad

ABSTRACT

This study pursues to identify the potential covariates of total factor productivity from the existing literature models of total factor productivity by using encompassing and LASSO techniques. The annual time series data of Pakistan are collected for the period 1982-2020. In this study we have considered the covariates of total factor productivity provided in literature in case of Pakistan. The final model is found by encompassing and LASSO techniques, which are $TFPM_{\textbf{Encompassing}}$ and $TFPM_{\textbf{LASSO}}$. Then we chose the final model for total factor productivity by ranking both models according to their standard errors. It has been observed that $TFPM_{\textbf{Encompassing}}$ has the minimum standard error and encompasses the $TFPM_{\textbf{LASSO}}$. Therefore, $TFPM_{\textbf{Encompassing}}$ is considered as parsimonious model for the covariates of total factor productivity in Pakistan. The results of parsimonious model suggest that the major covariates of total factor productivity are expenditure on education, foreign direct investment, unemployment rate, inflation, imported machinery, residential patents, and nonresidential patents of Pakistan. This is considered superior because it contains all those regressors, which are not collectively present in one model in the existing literature. ARDL bound test has been applied to find the long run cointegrating relationships among the covariates of total factor productivity. In the long run the expenditure on education, Inflation, unemployment rate and residential patent of Pakistan have negative and significant, whereas the imported machineries and nonresidential patents have positive and significant impact on total factor productivity of Pakistan. While in the short run, imported machineries, unemployment, foreign direct investment, residential and nonresidential patents have positive and significant effects on total factor productivity of Pakistan. Whereas, Inflation have negative and significant effect on total factor productivity both short and long run. Therefore, it is recommended that the Government should need to improve the technique of production by importing machinery, more investment and that will lead high productivity in every sector.

Key Words: Total factor productivity, Encompassing, Cox test, LASSO, Unit Root, ARDL Bound Test

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVATIONS

ADF	Augmented Dickey Fuller
ARDL	Autoregressive Distributive Lag
CPI	Consumer price index
DE	Development Expenditure
ECM	Error correction model
Edu	Education expenditure
F_ Depth	Financial depth
FD	Financial Development
FDI	Foreign direct investment
GUM	General Unrestricted Model
HC	Human Capital
ICT	Information and communication technology
IMM	import of machinery
IR	Interest rate
LASSO	Least absolute shrinkage and selection operator
PTN	Number of nonresidential patents
PTR	Number of residential patents
R&D	Research and Development
SEC	Secondary education
SG	Size of the government
TFP	Total factor productivity
TFPM	Total factor productivity Model
TO	Trade Openness
UMP	Unemployment Rate
URN	Urbanization
WDI	World Development Indicator

CHAPTER 1

INTRODUCTION

1.1 Background of the study

Total factor productivity is defined as, such increase in output which is not defined by growth in inputs(Solow, 1957). The total factor of productivity is used as an indicator to measure the usage of all the factors of production by the economy of a country to generate valuable outputs. It is also known as multi factor productivity and used to measure the output economic growth of a country. For every economy, sustainability of economic growth is the main objective of economic policies. Developing economies are trying to implement policies through which they could boost their economic growth and will attain the growth trajectory of developed countries. Therefore, assessment of productivity growth and its sources plays a significant role for policy formulation and resource allocation.

Solow (1956) and Swan (1956) are considered that the production function of neoclassical, where labor and capital are used as factors of production. Later 1960, when other theories were introduced (e.g., consumption, education, and life expectancy) as a factor of economic growth and this way the theory of economic growth move toward a new direction. As like a physical capital, human capital were also becoming an essential factor of economic growth literature. As a proxy for human capital, indicators like expenditures on education, primary and secondary schooling enrollment ratio, (R&D) Research and Development, life expectancy, expenditures on health were used by different researcher. Becker (2009) investigated that investment on education, training and health improves human capital accumulation (skilled, semi-skilled & healthy labour force). Educated labor force can affect economic growth through factor accumulation as well as through increase in Total Factor Productivity. TFP growth, the best overall measure of competitiveness reflects technological change, increases the welfare of a nation. TFP growth is contributed by technical change, efficiency change, and scale change (Iliyasu et al., 2015).

Over a long period of time the economic growth of Pakistan is declining due to the low rate of TFP. During 1972-2019 the TFP of Pakistan grew at a rate of 1.62 percent. However, the long-term trend of TFP are showing decline as like the trend of GDP (PIDE Reform Agenda 2021). O. Siddique (2020) provides the decade wise estimates of TFP of Pakistan. That is 1.71 percent in 1970s, 2.77 in 1980s, 0.33 in 1990s, 1.31 in 2000s and 2.07 in 2010s. The highest rate of TFP in the history of Pakistan were 2.77 in 1980s due to high investment to GDP ratio and the lowest were 0.33 in 1990s, which are due to instability of macroeconomic variables and the failure of policymakers to implement and sustain reforms (López-Cálix et al., 2012). According to Hussain (2010), the low TFP rate in 1990s were due to flawed by political instability and poor macroeconomic management. In late 2000s, TFP rate grew to 1.31% and GDP growth rate was 5.06%, according to Muslehuddin (2007) argued that the improvement in growth were taken due to betterment in macroeconomic fundamentals, structural reforms/changes, institutions, governance, and the private sector dynamism of economy. There are many reasons for the low growth of the TFP, some of the important causes is the lack of latest up-to-date technology, unskilled labor, poor infrastructure, and political instability in the country and decline in the industrial revolution (O. Siddique, 2020). The aim of this research is to identify the significant covariates of TFP from all the relevant determinants of TFP from literature by using encompassing and LASSO techniques and adopting encompassing and LASSO in this study because no literature has used these techniques for finding covariates of TFP from all the relevant determinants, to determine the significant covariates of TFP. This may guide the policy makers to formulate the policies based on few determinants that will increase the TFP and consequently the economic growth of Pakistan.

Leamer (1978) found that the regression model is valid only when all the relevant regressors that are the determinants of the dependent variable should be included in the model. If the relevant variables are excluded from the regression model, then the model will be mis-specified and the result which are drawn from the regression model can be completely misleading. We will use Encompassing and LASSO approach to estimate significant determinants. Encompassing approach is proposed by (Hendry & Richard, 1987) and (Mizon & Richard, 1986), which is model selection strategy method (Zaman, 2017). LASSO is a machine learning technique that was developed in 1989 and presented by Robert Tibshirani in 1996. LASSO regression is a powerful technique which performs two main responsibilities such as the regularization and the feature selection. . This technique is used when there are more independent variables and high multicollinearity. Finally, this method will make a parsimonious model in the presence of a large number of variables.

In order to obtain the true regressors of the dependent variable, one should start with the general model which includes all the potential determents. General unrestricted model (GUM) is developed by considering all the determinants of dependent variable from the existing literature. Whenever the number of variables is greater than the number of observation (k>n) then we are unable to estimate the GUM. To solve this problem, we choose the general model either by encompassing technique or by applying Least Absolute Shrinkage and Selection Operator (LASSO) technique for feature selection. Both encompassing and LASSO have advantages over other techniques such as OLS, Ridge regression etc. because encompassing and LASSO provide unique solutions in case the number of variables is greater than the number of observations. They both include the best subset of predictors in our final model instead of including all the predictors. On the other hand, different researchers have built various models in the earlier studies to examine determinants of TFP. They also concluded that if we omit any of these variables then these will cause omitted variables bias. While if we consider all the variables which are used in the earlier studies instantaneously then our model will be too large, and our result will be insignificant. Therefore, the aim of this study is to find the best suitable model of TFP in case of Pakistan by using encompassing and LASSO methodology.

1.2 Research Problem

According to PIDE growth agenda, the TFP grew at a rate of 1.62 percent and GDP showing decline during 1972 to 2019. For higher GDP growth it is needed to nudge TFP, to achieve the higher TFP growth rate. First, we need a specific model which have more power of explanation to evaluate determinants of TFP, and which may be close to the true model and identify the main covariates which have impact on TFP.

1.3 Objective of the study

The key objective of the study is:

- To identify the potential Covariates of TFP by using encompassing methodology.
- To identify the potential Covariates of TFP by using LASSO methodology.
- To investigate the symmetric impacts of potential Covariates on TFP.

1.4 Significance of the study

The most important task of the study is to find the most suitable model of TFP and its covariates among different existing models by using the encompassing and LASSO methodology. It is studied from the previous literature as, Adnan et al. (2020), Tufail and Ahmed (2015), Saleem et al. (2019), Akinlo and Adejumo (2016), Kolasa and Żółkiewski (2004), Gehringer et al. (2016), M. Ajide (2004), that different models have found the effect of the covariates on TFP. But it is very difficult to say which model is the best model among the large number of models. Thus, this study will contribute to choose only one model which may be close to the true model.

CHAPTER 2

LITERATURE REVIEW

In this chapter we will explain and discuss briefly the previous work interrelated to our study. The covariates of total factor productivity have been explored by various researchers with various conclusions. Some of them suggested a positive relationship whereas other have drawn a negative relationship of TFP with determinants of macroeconomic variables.

2.1 Endogenous and Exogenous economic growth theory

Solow-Swan and neoclassical growth theories of economics refer to those of exogenous economic growth theories. Model development by (Solow, 1956) and (Swan, 1956) laid the foundation of literature in exogenous growth model theories that took the production function of neoclassical where labor and capital are used as factors of production. The model explains the idea that the long run growth rate is determined exogenously. For instance, technological advancement is a significant factor of growth rate that is determined outside from the model. Moreover, other factors such as population growth rate that too affect economic growth rate is yet another factor which is determined exogenously. Introducing consumer optimization by Cass and Koopmans (1965) extended the neoclassical growth model; however, the economic growth rate in long run remained dependent on as an exogenous factor, i.e., advancement in technological factors.

Following the empirical and innovative work of Romer (1986, 1990), Lucas Jr (1988), Agion and Hewitt (1992) the groundwork of new economic growth theories as endogenous growth theories were set on; nonetheless, the emergence of eight endogenous growth theories were but the extending work of neoclassical growth theories. The theories exert the importance of policy makers' effect on economic growth in the economy. Therefore, explaining endogenous nature of the growth model which is determined inside the model. Endogenous growth model introduced knowledge, human capital, R&Ds in the growth model, which were determined endogenously inside the model. The valuable contribution of Jones and Manuelli (1990), Rebelo (1991) and others to the development of endogenous-growth model and eventually with introduction of various growth theories. So were introduced several large sample international datasets.

2.2 Total Factor Productivity and Economic Growth

Total factor productivity is defined as, increase in output which is not defined by growth in inputs (Solow, 1957).It is now considered as an important factor of growth. The first attempt was probably made by Abramovitz (1989) at measuring TFP , when the study tries to identify the main source of U.S productivity growth. Solow (1956, 1957) also recognizes that only a fraction of growth in output is explained by physical inputs, the exogenous residual capture TFP. (Prescott, 1998) give evidence for need for theory of TFP, as new classical growth theory cannot explain the differences in income level and economic development level of different countries unless the variation in TFP in different point of time and in different countries are considered. The difference in physical capital, intangible capital, saving rate, technical knowledge all are important, but it is total factor productivity that matter.

The economic growth experiments are incomplete with no resonation with TFP. Nonetheless, one of the major issues with the Slow model is the residual value, TFP is of a great importance which is mostly the computation of the residuals; therefore, TFP model is identified as Solow residuals. Growth Accounting Model (GAM) is used for calculating TFP that splits economic growth to its analogous component. Also, that the economic growth literature is very rich with the TFP for both aspects as developed and for developing economies.

2.3 Empirical Literature

The following studies are exploring the influence of different determinants of total factor productivity on national level of Pakistan and as well as international level.

As far as determinants of TFP are concern, Isaksson and Ng (2006) consider two modes of analysis for 15 countries, it shows that human capital, physical capital, infrastructure, financial development, technology transfer through trade have significant effect on TFP .Then Danquah et al. (2011) investigate factors effecting TFP growth. It measure TFP using non-parametric frontier technique to compute Malmquist productivity index. This technique results in decomposing TFP into two components, technical efficiency, and technological changes. To check the robustness of TFP growth and its components. The study has used Bayesian model averaging technique. Using the data for 67 countries (20 OECD, 47 non-OECD) from 1960-2000, the research found that initial GDP and trade openness were significant. In OECD countries investment price, consumption share, labour force was significant and in non-OECD countries population density also have effect on TFP growth.

Pasha et al. (2002)also strongly support the idea that TFP plays very important role in effecting the economic growth rate in Pakistan. High growth during 80‘s and low economic growth in 90‘s was mainly due to variation in Pakistan‘s TFP level .The paper used simple OLS regressions for determinants of TFP in Pakistan. The factors which contribute to growth in TFP were human capital, vintage capital, development expenditures, manufactured exports, cotton production and workers. Sabir et al. (2003) consider TFP in Pakistan from different point. They establish the link between TFP and Structural Adjustment Program (SAP) in Pakistan. The study considers the pre-reform period 1972-73 and reform period 1987-88 to 2001-02.SAP results in low growth in factor inputs which results in lowering growth rate in economy , and this was mainly due to decrease in TFP level in country.

Khan (2005) also estimates and then determine the macro determinants of TFP by using OLS technique for the data from 1960-2003 for Pakistan. And found that inflation, FDI, financial sector, budget deficit, population growth, employment and government consumption plays important roles in determining TFP level. Then Qazi and Hyder (2007) found that cotton production, export of manufacturing goods, human capital real development expenditure, and real credit to private sector plays important role in effecting TFP level sector wise and overall, for the whole economy.

Due to Pakistan's unfavorable political and socioeconomic environment, foreign direct investment inflows have declined to a low level during the past decade. Adnan et al. (2020) studied the dynamics of FDI, human capital, and openness to trade in relation to TFP for Pakistan between 1970 and 2018. They used the ARDL bound approach to analyses the long-run and short-run relationships between the variables and the Granger causality test to determine the causal relationships between TFP and explanatory variables. According to their findings, the TFP has a long-standing link between FDI and human capital. Several studies also showed a correlation between innovation and production. Saleem et al. (2019) Studied that the Innovation, economic growth, and TFP in Pakistan, and used annual time series data for 1972–2016 from World development Indicator (WDI) by applying 2SLS model. The overall results conclude that all variables of their study are statistically significant. Besides, innovation are positively significant and contributory factor to economic growth in Pakistan. Pakistan is basically an agricultural economy, provides round about 38.5 % employment to the labor force and employing more than about 65-70% of the population are depends on agriculture for its livelihood according to Pakistan Economic Survey (2020–2021).Tufail and Ahmed (2015) used ARDL and ECM and explained that the magnitude of elasticity estimate for the agriculture sector is pretty impressive for all factors. The agriculture sector found that trade openness, development expenditures, human capital, and imports of intermediate goods and machinery have a positive impact on growth and TFP. However, FDI contribution is positive but has low elasticity with respect to TFP, whereas the coefficient of financial development has a negative sign due to the availability of credit to the agricultural sector and the fact that the agricultural sector faced numerous obstacles in gaining access to agricultural credit. Using secondary data for the period 1960-2003 and estimating the macroeconomic determinants of TFP in Pakistan. Khan (2005) concluded that TFP played a crucial role in economic growth. The primary contributors to TFP in Pakistan between 1965 and 2005 were macroeconomic stability, Foreign Direct Investment, and banking sector expansion. The intriguing finding of the study was that the effect of schooling on TFP was negligible. Abbas and Nasir (2001) employed primary schooling, secondary schooling, and higher education as proxy for human capital to determine its relationship with economic growth in Pakistan and Sri Lanka from 1970 to 1994. The empirical results based on the growth accounting method showed that the effects of primary school attendance on economic growth is negative for both economies. When human capital is proxied by secondary school enrollment, the effect becomes positive for both countries. The results indicate that human capital contributes positively and significantly to Pakistan's economic growth. According to the report, Pakistan's primary education system is in a horrible condition, and raising the level of basic education in Pakistan and Sri Lanka is essential. They recommend that investments in secondary and higher education should be increased because these levels of education have a positive impact on the economic growth of both economies. Research and development (R&D) is a crucial determinant of total factor productivity (TFP) via technological progress and spillover effects. Spending on Research & Development in the high-tech sector can be more effective and productive than in other areas (Nadiri, 1993). Ajide (2021) analyzed the asymmetric impacts of TFP on Nigeria's crime rate by employing linear and nonlinear ARDL modelling techniques and the Granger causality test for time series data. They observed that, in the near term, the crime rate has a significant positive effect on the TFP of the positive components, whereas the negative components had negative effects on TFP. However, both positive and negative shocks have negative effects on TFP in the long run.

2.4 Literature Review on Application of Encompassing Technique

The encompassing technique relates to a model's capacity to account for the characteristics of other models. The previous researchers such as, Mizon and Richard (1986), Hendry and Richard (1987), and Lu and Mizon (1996), focused on variance and parameter encompassing. The Cox test of the non-nested hypothesis is a variance-encompassing test, according to Mizon and Richard (1986) who's focused on the wide range of encompassing tests. In 1990 another researcher developed a test of conditional mean encompassing and compared that test with Cox and Richard tests (Wooldridge, 1990).

The application of the encompassing technique has been found in the existing literature of Pakistan. Such as Nazir (2017) applied the encompassing technique on the three energy growth models. Those three models were proposed by Kraft and Kraft (1978) and Dantama et al (2011). She built the third model by using the determinants of two existing models. She has tested these three models with the help of nested and non-nested encompassing by using F and Cox test. She also 12 found that the independent variables in the first two models defined the GDP growth very well. Finally, the third model encompassed the first two modes.

M. A. Siddique et al. (2016) explored the internal and external contributing factor of Islamic banking growth of Pakistan. They collected the quarterly data from the period 2004-2012. The researchers used the encompassing approach to find the parsimonious model. Firstly, they used the encompassing technique to find the GUM from the existing model. Then they used the Wald restrictions test on the GUM to find the parsimonious model. Finally, the researchers have found the Islamic banking will be in progress if there is efficient management.

2.5 Literature Review on Application of LASSO Technique

LASSO is a powerful technique which performs two main responsibilities such as the regularization and the feature selection. This method makes a parsimonious model in the presence of a large number of variables. The previous researchers i.e., Epprecht et al. (2021) Compared two approaches for the purpose of model selection for the linear regression models such as Auto-metrics (automatically selection from general to specific) and LASSO (the regularization and feature selection method) and ada-LASSO (adaptive LASSO). Their result concluded that all the techniques will improve their performances as increasing the sample size and decreasing the number of relevant and candidate variables. Ferraro determined the LASSO technique, which is a statistical tool that obtain sparse solutions for regression problems. He also found that LASSO technique has so many applications, from biology to economics. But he suggested the application of social economics, especially, the investigation of poverty rate determinants. The aim of this study was to identify the explanatory variables that have higher impact on poverty of Latin American countries.

Fonti and Belitser (2017) examined the use of LASSO technique to describe the feature selection task. While using different setups, they tested this technique. They mostly focused on the two types of statistical models such as linear model and generalized linear model. They concluded that the LASSO technique has benefits to select a model that have the most relevant features.

2.6 Summary

The previous literature showed that different models are used for different sets of covariates to explain the phenomena of TFP. These all studies concluded that, in the above studies all models are different from each other. When different models are used for the same phenomena, then it make sense that all the models are incorrect or there will be only one model that might be adjacent to the true model. If we omit any of these variables, then these will cause omitted variables bias. While if we consider all the variables which are used in the earlier studies instantaneously then our model will be too large, and our result will be insignificant. Thus, it is important to find out the best suitable model of TFP in case of Pakistan by using encompassing and LASSO methodology. These methodologies will give us a parsimonious model which may be close to the true model and identify the main covariates which have impact on TFP.

CHAPTER 3

DATA AND METHODOLOGY

3.1 Theoretical Framework

In literature of economics, there is an essential role of productivity in accelerating the momentum of economic growth. According to the framework of neo-classical growth accounting, output growth is a summation of productivity/efficiency growth and growth of inputs accumulation. The Solow residual is the portion of an economy's output growth that cannot be attributable to the accumulation of capital and labor, the production factors. The Solow residual represents growth in output that is beyond the basic growth of inputs. As a result, the Solow residual is frequently regarded as a measure of technical innovation-driven productivity growth. The Solow residual is also known as the TFP. Hence, for a given combination of input factors such as land, capital, and labor, improvements in productivity or efficiency generate a shift in production frontier. The technological progress or surge in efficiency is treated as an exogenous process in neo-classical framework, Solow Growth Model can be considered as an example. However, endogenous growth theorists have challenged these models and argued that technological advancement is an endogenous process that can be measured as TFP. The technical process (i.e., endogenous) enables government actions to influence the technological process, which is reflected in TFP and ultimately growth. Consequently, it implies that output is controlled by multiple variables, although indirectly via labor and capital efficiency.

Total factor productivity growth is taken as dependent variable and will examine the effect of different determinants on TFP growth of the economy. If conditions in country improves different determinants of TFP, the growth of TFP will increases.

Neoclassical production function can be used to estimate TFP

$$Y = F(A, K, L) \qquad \ldots\ldots. \ (3.1)$$

In the equation above, Y represents output, K represents capital stock, L represents employed labor force, and A represents the residual term, which is TFP. We can write the above equation in the form of growth as follow,

$$gY = gL + g\text{k} - gTFP \qquad \ldots\ldots. \ (3.2)$$

Here, gY is the growth rate of total output, gL is the growth rate of labor, gk is the growth rate of capital, and gTFP is the growth rate of total factor productivity. So, we can also express the preceding equation as

$$g = \alpha gL + (1 - \alpha) - gTFP \qquad \ldots\ldots. \ (3.3)$$

α represents the proportion of labour to output, while (1- α) represents the proportion of capital to output. According to equation 3.3, the growth rate of production is a weighted average of increase in the employed labour force, capital stock, and technical development, given by the growth of TFP, where the weights are labour and capital factor shares.

Assuming output and inputs are observable, the TFP can be calculated using the following equation:

$$gTF = gY - \alpha gL - (1 - \alpha) \qquad \ldots\ldots. \ (3.4)$$

Therefore, the growth in TFP is seen essentially as a residual.

3.3 Model Selection by Encompassing Method

As the objective of our study, we are to choose the most suitable model of TFP among various models. To fulfill this objective, we will use encompassing approach which is proposed by (Hendry & Richard, 1987) and (Mizon & Richard, 1986). (Leamer & Leamer, 1978) found that the regression model is valid only when all the relevant regressors that are the determinants of the dependent variable should be included in the model. If the relevant variables are excluded from the regression model, then the model will be mis-specified and the result which are drawn from the regression model can be completely misleading. Two types of model speciation error; under fitted model and over fitted model. If our model are under fitted, our result will be biased, and we cannot take policy recommendation. For correctly specified our model, we will use Encompassing approach, which is model selection strategy method (Zaman, 2017). Different researchers have constructed various models in the earlier studies to study determinants of TFP. So, we will use encompassing method to choose one model among the various models, which may be close to the true model.

3.3.1 Nested Model

If variables of one model are present in another model, we call it nested model. For nested model we use parametric encompassing technique, to develop a general model that explain the earlier model, and which contain all the regressor of earlier models with lags, we called it GUM. When we estimate GUM, we will make sure that the number of observations must be greater than number of variables. In two-model compression by parametric encompassing techniques, M1 & M2 as.

$$M1: \quad TFP = \beta_1 + \beta_2 X + \beta_3 Y + \mu$$

$$M2: \quad TFP = b1+b2X + \mu$$

In the above models, M1 is nested in model M2 and simply we say that M1 is our GUM. To estimate a general model either M1 is correct or M2, for that we will impose restriction on additional variables. Because the correlation between M1 and M2 are dependent on additional variables such as, coefficient of b3.

We will estimate M1 general model and apply F-test for additional variables that whether they are significant or not? If they are significant, then we will conclude that they provide additional information to explain our dependent variable. We will conclude that by using parametric encompassing method M1 encompassing M2.

3.3.2 Non-Nested Model

Parametric encompassing is possible when we estimate GUM, and we estimate GUM when $n > k$.

Otherwise, we will go toward Non-Nested Model encompassing. For Non-Nested Model encompassing we will use variance encompassing method or forecasting encompassing.

3.3.2.1 Variance encompassing

In 2-model compression by variance encompassing method such as M1 encompassing M2 in term of variance encompassing. It mean that maximum variation of the dependent variable is explain by M1 and nothing left for M2.

Steps involved in J-test for variance encompassing.

- Estimate M1 and get estimated dependent variable from M1.
- Augment M2 by estimated dependent variable as a regressor.
- Estimate the augmented model and test the hypothesis

If null hypothesis, coefficient of augment estimated variable are equal to zero, which mean that it insignificant. It's concluded that augment estimated variables haven't effects on dependent variable. So M1 does not encompass M2. If they are significant, then M1 encompasses M2. So, we will conclude that they have extra power of explanation of dependent variables.

3.3.3 Encompassing of multiple models

In common literature, non-nested encompassing is applied on two-model comparison only but Zaman study on the multiple models encompassing. Multiple models encompassing one phenomenon, the dependent variable of all models will be same. First, we will estimate all models and calculate the **standard error of residual** and choose a model which have minimum standard error. Then making 2-2 sets of all models with minimum S.E model and then we will apply some tests to perform the null hypothesis of encompassing. If null hypothesis of models are accepted, we will ignore those models because they have no extra explanatory power. Other models, which null hypothesis are rejected, we will add it and get a general model. The general model will be simplified further by utilizing general to specific methodology. To get specific models from GUM we will use restrictions on all variables. Those variables which are highly insignificant then we will drop them from the model. In this way finally we get a specific model.

3.4 Model Selection by LASSO

LASSO is a machine learning technique that was developed in 1989 and presented by Robert Tibshirani in 1996. LASSO is a shrinkage-based linear regression. It is a powerful technique which performs two main responsibilities such as the regularization and the feature selection. This technique is used when there are more independent variables and high multicollinearity in the model. It is an alternative method to the least squares estimate. Moreover, in this model when the variables are insignificant or do not have relationship with the response variable then the lasso makes their coefficient approximately equal to zero and finally drop them from the model. In this way the over fitting is also reduced.

3.4.1 Feature Selection

The main purpose of feature selection process is to omit those variables which are redundant, to make the model easier to interpret and to reduce the over fitting. The feature selection is a very important task because here the number of variables are very high and sometime the number of variables is larger than the number of observations. In this case it is not easy to say which one of the variables is relevant and which one is irrelevant. Therefore, the feature selection process has a great importance (Fonti & Belitser, 2017).

3.4.2 Methodology of LASSO

A commonly used procedure to find a linear relationship among variables is the linear regression model which involve the minimization of RSS.

$$\boldsymbol{RSS} = \sum_{i=1}^{n} (\boldsymbol{y_i} - \boldsymbol{\beta_0} - \sum_{j=1}^{p} \boldsymbol{\beta_j x_{0ij}})^2 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \mathbf{3.5}$$

But this linear regression model has the problem of variability in the least square fit as the number of variables exceeds the number of observations. The solution is suggested in the form of ridge regression. The ridge regression has advantage over OLS because as the penalty λ increases the variance decreases substantially at the expenses of very small increase in bias. Secondly the OLS does not provide a unique solution in case of the number of variables is larger than the number of observations. While in this case the ridge regression works well which is given in the bellow equation.

$$\sum_{i=1}^{n} (\boldsymbol{y_i} - \boldsymbol{\beta_0} - \sum_{j=1}^{p} \boldsymbol{\beta_j x_{0ij}})^2 + \boldsymbol{\lambda} \sum_{j=1}^{p} \boldsymbol{\beta^2}_{\boldsymbol{j}} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \mathbf{3.6}$$

Despite having a lot of advantages, the ridge procedure is not free of problem. The problem with Ridge regression is that it tell us to include all the predictors in our final model instead of the best subset which leads to shrink all the coefficient toward zero but not exactly equal to zero. Secondly, for selecting a good value of penalty λ, the ridge regression produces a different set of coefficients for each of λ. To get rid of this problem, a new procedure, LASSO which was introduced by Robert Tibshirani that is given below.

$$\sum_{i=1}^{n}(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{0ij})^2 + \lambda\sum_{j=1}^{p}|\beta_j| \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 3.7$$

$$RSS + \underbrace{\lambda}_{Tuning\ Parameter}\underbrace{\sum_{j=1}^{p}|\beta_j|}_{penalty} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 3.8$$

We have seen in the above equation (3.6) and (3.7) that Ridge and LASSO regression have same construction, but there's only difference that the β term in the ridge regression penalty equation (3.6) has been replaced by β in the LASSO penalty equation (3.7). In the LASSO equation (3.7) where RSS is residual sum of squares, λ is tuning parameter and ∑ is sum of absolute value of coefficients is the LASSO penalty.

3.4.3 Choosing the Value of the Tuning Parameter

The penalty's strength is controlled by the tuning parameter. The parameter coefficient (β) correlates with the tuning parameter value. Whenever λ =0 then the penalty term has no effect, and we will get the same coefficients as simple linear regression. When λ= ∞ then all the coefficients are zero. When λ is in between the two extremes (0< λ<∞) then we are balancing the two ideas. Such as fitting a linear model of Y on X and shrinking the coefficients. The range of tuning parameters is between zero to infinity and it is a crucial value for the identification of the true model. Whenever an intercept is included in the model, then it is left unchanged and, in the equation (3.8 the shrinkage penalty is applied to $\beta_1 \ldots . \beta_p$, but not to the intercept β_0. Furthermore, LASSO has a significant benefit over ridge regression in that it creates models that are easier to understand and include a smaller number of predictor variables. It is a regularization method that creates parsimonious model in the presence of large number of features. In this way the over fitting is reduced (James et al., 2013).

3.5 Estimation of Specific Model

When there are more than one no stationary time series, then there is the possibility of cointegration among them. ARDL co-integration by Pesaran et al. (2001) is commonly used for the identification of co-integration among the variables and its details are given bellow.

3.5.1 Descriptive statistics:

The GUM is reduced in size by performing a number of restrictions and then the resultant model is known as specific model. First, we will summarize descriptive statistics of the variables to check mean, median, dispersion, standard deviation, and some normality tests. After descriptive statistics we will move forward for unit root test.

3.5.2 Unit Root Test

The main issue of time series data analysis are autocorrelation problem. To start our estimation first I will check whether autocorrelation problems are present are not. For autocorrelation problem I will perform some unit root test such as DF test, ADF and Phillips perron test. So, these tests will confirm the nature of variables, which are important for econometrics techniques to be applied. If all variables are stationary at a level, estimation will do through OLS. If all variables are stationary at first order, then we will go toward Engle-Granger or Johnson cointegration. And if variables are stationary at mix of both level and first or second order, then will use auto-regressive distributed lag model (ARDL).

But here, we will get GUM model through encompassing techniques, and we define GUM model as “contain all the regressors of earlier models with lags”. So, we may call GUM as ARDL model.

This ARDL (1960) is different from ARDL proposed by (Pesaran et al., 2001). In model when more than one non-stationary variables, then possibility are more of cointegration among variables and ARDL by Pesaran et al. (2001) is commonly used for identification of cointegration among variables. The details of linear ARDL is follows.

3.5.3 Symmetric ARDL

In the previous literature all researchers assumed that all explanatory variables tend to have symmetric impact on the dependent variable. The meaning of symmetric assumption is that if a decrease an independent variable will bring increases in the dependent, or increases an independent variable will bring decreases in the dependent**.**

When all variables are integrated in different order then we use Autoregressive Distributed Lag. For example, some variables are integrated of I(0) and some are I(1). In this study ARDL approach will be used to study the co-integration relationship between TFP and its determinants because we are expecting both I(0) and I(1) regressors. This approach captures both short run and long run relationships. It states that the dependent variable must be stationary at level or stationary at first difference, and that the other explanatory variables may be stationary at either levels or first differences. However, the model's variables should not be stationary at I (2).

3.5.4 ARDL Model Specification

We have constructed a general model by using TFP_t as a dependent variable and X_{1t}, X_{2t}….Xn_t as an independent variable. We also assume that $TFP_t \sim I(1)$ and independent variables are either I(1) or I(0). The mathematical representation of the ARDL model is given below.

$$\Delta Y_t = \alpha + \sum_{i=1}^{n} \beta_{1i}\Delta Y_{t-i} + \sum_{i=1}^{n} \beta_{2i}\Delta X_{1t-i} + \sum_{i=1}^{n} \beta_{3i}\Delta X_{2t-i} + \sum_{i=0}^{n} \beta_{4i}\Delta X_{3t-i} + \dots$$
$$+ \sum_{i=0}^{n} \beta_{ni}\Delta X_{nt-i} + \delta_1 Y_{t-1} + \delta_2 X_{1t-1} + \delta_3 X_{2t-1} + \delta_4 X_{3t-1} + \delta_5 X_{4t-1} + \cdots$$
$$+ \delta_n X_{nt-1} + \varepsilon_t \qquad \dots\dots\dots\ 3.9$$

3.5.5 Long run Relationship

For the existence of long run cointegration, we use Bound testing approach. The following null hypothesis is tested against alternative hypothesis. The lagged variables' coefficients being equal to zero is the null hypothesis. It indicates that there is no long-term relationship between the variables. Alternatively, it may be said that at least one of these coefficients is not equal to zero. For bound test, null hypothesis as

$$H_0: \delta_1 = \delta_2 = \delta_3 = \delta_4 \dots \delta_n = 0$$

It mean, there is no long run relationship exist.

And alternative hypothesis as,

$$H_1: \delta_1 \neq \delta_2 \neq \delta_3 \neq \delta_4 \dots \delta_n \neq 0$$

It mean, long run relationship exists.

F-statistic is used to identify whether long run relationships exist among variables or not. The values of F-statistic is compared with the critical value, which provided by Pesaran et al. (2001). If the value of F-stat are greater than the upper bound I(1) critical value, then its mean rejection of the null hypothesis. It indicates that the long run exists there. If the F-statistic value is less than lower bound I(0) critical value, then null hypothesis will be accepted. It shows that the long run relationship does not exist and if it occurs between critical value I(0) and I(1), then its mean results are inclusive.

3.6 Error Correction Mechanism (ECM)

After any short-term shock or drift, the ECM captures the rate of adjustment or, in the long run, convergence. For convergence ECM coefficient must be negative and significant.

$$\Delta Y_t = \alpha + \sum_{i=1}^{n} \beta_{1i} \Delta Y_{t-i} + \sum_{i=0}^{n} \beta_{2i} \Delta X_{1t-i} + \sum_{i=0}^{n} \beta_{3i} \Delta X_{2t-i} + \sum_{i=0}^{n} \beta_{4i} \Delta X_{3t-i} + \dots + \sum_{i=0}^{n} \beta_{ni} \Delta X_{nt-i} - \omega \varepsilon_t + v_t \qquad \dots\dots\dots\dots 3.10$$

In the above equation (3.10), (ω) is the speed of adjustment which should be negative and significant for convergence toward the long run equilibrium.

3.7 Diagnostic Test

Diagnostic tests help to identify both strengths and weaknesses of the models. During model selection procedure different tests will be used to ensure that the residuals are free from the problem of heteroscedasticity, non-normality, and autocorrelation problem. For identifying such kind of problems, we use White heteroscedasticity (ARCH) LM test (F-stat.), Jarque Berra test (1980) ($\chi 2$) of normality and Breusch-Godfrey Serial Correlation LM test (1978). Moreover, we detect the stability of the parameters of estimated dynamic ECM with the help of CUSUM and CUSUMSQ which were proposed by Brown, Durbin, and Evans in 1975.

3.8 Earlier Models

M: 01

$$TFP = f\ (HC, TO, TT, DE, FD) \qquad (a)$$

Tufail and Ahmed (2015) used the model, where TFP is represent total factor productivity, HC for human capital, TO for openness of the economy, TT for transfer of technology, DE for development expenditure and FD for financial development. Here we know that transfer of technology is a function of foreign direct investment and imports of intermediate goods and machinery so,

$$TT = f\ (FDI, IMM) \qquad (b)$$

FDI stand for foreign direct investment and IMM for the imports of machinery and intermediate goods. Now substituting equation (b) into equation (a) we get the final equation of the study.

$$TFP = f\ (HC, TO, FDI, IMM, DE, FD)$$

Empirically the given model we can write as

$$TFP = \alpha_0 + \alpha_1 HC_t + \alpha_2 TO_t + \alpha_3 FDI_t + \alpha 4 IMM_t + \alpha_5 DE_t + \alpha_6 FD + \mu_t$$

M: 02

$$TFPt = \gamma_0 + \gamma_1 TFP_{t-1} + \gamma_2 PTR + \gamma_3\, TO + \gamma_4\, INFt + \gamma_5\, EDUt + \beta_6 SEC_t + \gamma_7\, IMMt + \gamma_8\, FDIt + et$$

TFP for total factor productivity, PTR for number of patents, TRD for trade openness, INF for inflation, EDU for education, IMM for imported machinery, FDI for foreign direct investment and et for error term. Moreover, TFPG is related to total factor productivity growth (Saleem et al., 2019).

M: 03

$$TFP_t = + \beta_0 + \beta_1 TO_t + \beta_2 FDI_t + \beta_3 EDU_t + \beta_4 SEC_t + \beta_5 CPI_t + \beta_6 UMP_t + \mu_t$$

TFP stand total factor productivity, TO for trade openness, FDI for foreign direct investment, EDU for expenditure on education, CPI for consumer price index, UMP for the unemployment rate and μ for error term (Akinlo & Adejumo, 2016).

M: 04

$$TFP = \alpha + \beta_1 FDI_t + \beta_2 ICT_t + \beta_3 R\&D_t + \beta_4 EXT_t + \beta_5 EDU_t + \beta_6 SEC_t + w_t$$

Gehringer et al. (2016) analyzed the factors influencing TFP in the European Union. Their model's dependent variable is the TFP. FDI estimates, FDI as a proportion of GDP. The ICT quantifies the contribution of ICT capital services to specific sectors. Research and development expenditures as a proportion of sectoral. As proxies for trade openness, the sum of imports and exports as a proportion of GDP was utilized. Human capital is approximated by the proportion of individuals having a secondary education and the error term (wt).

M: 05

$$TFP_t = \beta 0 + \beta_1 CR_t + \beta_2 CPI_t + \beta_3 GDPG_t + \beta_4 UMP_t + \beta_5 URN_t + \mu_t$$

Ajide (2021) investigated the asymmetric effect of crime rate on TFP. In the model CR are crime rate, CPI is consumer price index, GDPG are (GDP per capita growth rate), and UMP are unemployment and URN for urbanization.

M: 06

$$TFP_t = \beta_0 + \beta_1 TO_t + \beta_2 FDI_t + \beta_3 FD_t + \beta_4 IR_t + \beta_5 CPI_t + \beta_4 SG_t + \beta_5 EDU_t + \beta_5 HET_t + \mu_t$$

Here TFP is used for total factor productivity, TO is proxy for trade openness, FDI is foreign direct investment, FD is financial depth, IR is interest rate, CPI is consumer price index, SG is size of the government, EDU and HET are proxies for human capital which are Education and Health respectively

3.9 Data and variables Description

This research is based on the annual time series data over the period of 1982-2020 for Pakistan. Different sources are used for data collection. The detailed description of all variables with data sources is available in the following Table 3.1.

Table 3. 1 Data and variable Description

S.N	Variable	Description	Source
1	Total factor productivity	$gTFP = gY - \alpha gL - (1 - \alpha)$, GDP (HBS 1.5 2020), labor, and gross capital formation (constant US dollars), capital depreciation rate obtained from Penn World table 9.0. For capital stock series used perpetual inventory method. The data for TFP is not available, it is computed by different economists. The data used in this paper is computed by Khan (2005). Two approaches are used for the computation of TFP. The Growth Accounting Approach and The Index Number Approach. In our study we used the Growth Accounting Approach. The basic framework for this approach is provided by Solow (1957). In this approach TFP is computed as residual. The share of some specific factor (input) to total output growth is subtracted from the total output growth. The difference between the two is known as ―Solow Residual‖. The share of capital is assumed to be 0.52 and that of labour 0.48.	HBS 1.5 2020

2	Foreign direct investment	FDI is the type of investment in which the people or organization of one country invested in the company or property of another country. Foreign direct investment, net inflows (BoP, current US$)	WDI
3	Trade Openness	The Openness trade is calculated by taking the sum of import and export to divided by total GDP in million US dollars	WDI
4	Education expenditure	Government expenditure on education, total (% of GDP)	WDI
5	Secondary education	School enrollment, secondary (% gross). Secondary school enrolment is defined as the number of students which are enrolled in secondary school.	WDI
6	Development Expenditure	Development Expenditure **(million RS),**	HBS **(3.7)**
7	Unemployment Rate	Unemployment is defined as the people who want to work but do not have a job. This is the rate of unemployment and use as, unemployment % of total labour force	WDI
8	Human Capital	it is measured as the secondary school enrollment rate	WDI
9	Consumer price index	Consumer price index is used as a proxy for Inflation. Inflation is the increase in price of goods and services over time in general level. The inflation rate is measured by. $\mathbf{CPI_t - CPI_{t-1} / CPI_t * 100}$ Consumer price index (2010 = 100)	WDI
10	Patents	Used number of patents application by nonresidents (per thousand population) and number of patents by residents (per thousand population) as the proxies of innovation	WDI
11	Urbanization	Urban population as a percentage of total population	WDI
12	Size of the government	Increased government expenditures are expected to grow total factor productivity, Used budget deficit as proxy for size of the government	

13	Interest rate	Interest rate is taken as lending interest rate as this the only variable relating to interest rate. Lending rate is the bank rate that usually meets the short- and medium-term financing needs of the private sector	WDI
14	Financial Development	Financial Development Index is use for Financial Development	IMF
15	Information and communication technology	The ICT variable measures the sector-specific contribution of ICT capital services to growth	WDI
16	Financial depth	m2/GDP	WDI
17	import of machinery	**Imports of Principal Commodities** Machinery and Transport equipment’s	HBS
18	Research and Development	R&D refers to sectoral research and development expenditure as a percentage GDP	WDI

HBS* Handbook of statistics 2020

IMF* international monetary fund

WDI* World Development Indicator

.

CHAPTER 4

RESULTS AND DISCUSSIONS

The aim of this study is to identify the potential covariates of total factor of productivity by using Encompassing and LASSO techniques. Different researchers such as Adnan et al. (2020), Tufail and Ahmed (2015), Saleem et al. (2019), Akinlo and Adejumo (2016), Kolasa and Żółkiewski (2004), Gehringer et al. (2016), M. Ajide (2004), have built various models to examine potential covariates of TFP in Pakistan which shows different determinants of TFP. These existing models impose a priory zero restriction on each other. In such a way, one regression model has omitted a relevant variable from the other regression model, so introducing bias; hence, all regression models are invalid due to misspecification. In this chapter we followed six existing models of TFP by different researchers namely, M1, M2, M3 ...M6. We have applied Encompassing and LASSO method to estimate significant determinants. In this chapter we will find the most suitable model of TFP among the different existing models by using the Encompassing and LASSO technique.

4.1 Specifying Model by Using Encompassing Technique

Using the encompassing approach, encompassing methodology which define by Dr Ateeq ur Rehman in "Determinants of Islamic Banking Industry's Profitability in Pakistan" {Siddique, 2016 #120@@author-year}. We estimate all the existing models i.e., M1, M2, M3 ...M6. To find the general model. First, we check the standard error of all the estimated models, then rank all the estimated models according to their standard error and finally we will see that which model has minimum standard error. All possible Total factors of productivity models (M1, M2, M3 ...M6) which are discussed are given as below.

M:01 Tufail and Ahmed (2015) used the model, where TFP is represent total factor productivity, HC for human capital and Edu and Sec used as a proxy for human capital, TO for openness of the economy, DE for development expenditure and FD for financial development, FDI is stand for foreign direct investment and IM for the imports of machinery and intermediate goods.

$$TFP = \alpha_0 + \alpha_1 Edu_t + \alpha_1 Sec_t + \alpha_2 TO_t + \alpha_3 FDI_t + \alpha 4 IM_t + \alpha_5 DE_t + \alpha_6 FD + \mu \quad \ldots\ldots \ (4.1)$$

The regression results of above model on the annual time series data for the period 1982-2020 are provided in Table 4.1

Table 4. 1 Regression results of M1

Dependent variable TFP

	Coefficient	Standard Error	t-value	t-prob
Constant	3.886	1.459	2.66	0.012
EDU	0.650	0.557	1.17	0.252
SEC	-0.147	0.857	-0.172	0.864
TO	0.914	0.828	1.10	0.278
FDI	-0.017	0.187	-0.094	0.925
IMM	0.482	0.397	1.27	0.212
DE	-0.584	0.276	-2.12	0.042
FD	-0.026	1.863	-0.014	0.988
Std. Error	**RSS**	**R2**	**F (7,31)**	
0.3929	4.787	0.718	11.33 (0.000) **	

M:02 TFP for total factor productivity, GDP used for real gross domestic product, while LPT for number of patents, TO for trade openness, INF for inflation and proxy used for INF is CPI, Sec and EDU for human capital, IMM for imported machinery, FDI for foreign direct investment and et for error term. Moreover, TFPG is related to total factor productivity growth (Saleem et al., 2019).

$$TFPt = \gamma_0 + \gamma_3 LPT + \gamma_4 TRDT + \gamma_6 INFt + \gamma_8 EDUt + \gamma_9 IMMt + \gamma_{10} FDIt + et \ldots\ldots \quad (4.2)$$

The regression results of the above model on the annual time series data for the period 1982-2020 are provided in Table 4.2.

Table 4. 2. Regression results of M2

Dependent variable TFP

	Coefficient	**Standard Error**	**t-value**	**t-prob**
Constant	-2.372	3.718	-0.638	0.528
PTN	-0.0261	0.036	-0.709	0.484
PTR	-0.167	0.239	-0.701	0.489
TO	0.208	0.892	0.233	0.817
CPI	-0.231	0.613	-2.01	0.053
EDU	0.882	0.553	1.59	0.121
SEC	2.056	1.767	1.16	0.253
IMM	0.113	0.419	0.270	0.788
FDI	0.032	0.200	0.163	0.871
Std. Error	**RSS**	**R2**	**F (8,30)**	
0.3837	4.417	0.740	10.71 (0.000) **	

M:03 TFP stand total factor productivity, TO for trade openness, FDI for foreign direct investment, HUM for human capital and Edu and Sec used as a proxy for HC, INF for the inflation rate, UMP for the unemployment rate and μ for error term (Akinlo & Adejumo, 2016).

$$TFP_t = + \beta_0 + \beta_1 ECO_t + \beta_2 FDI_t + \beta_3 HUM_t + \beta_4 CPI_t + \beta_5 UMP_t + \mu_t \quad \ldots\ldots (4.3)$$

The regression results of the above model on the annual time series data for the period 1982-2020 are provided in Table 4.3.

Table 4. 3. Regression results of M3

Dependent variable TFP

	Coefficient	Standard Error	t-value	t-prob
Constant	-0.627	2.344	-0.268	-0.790
TO	0.588	0.775	0.758	0.453
FDI	-0.044	0.164	-0.269	0.789
EDU	1.087	0.548	1.99	0.055
SEC	1.675	1.696	0.988	0.330
CPI	-1.123	0.511	-2.20	0.035
UMP	0.020	0.104	0.194	0.847
Std. Error	**RSS**	**R2**	**F (6,32)**	
0.3821	4.674	0.725	14.1 (0.000) **	

M:04 Ajide (2021) investigated the asymmetric effect of crime rate on TFP. In the model CR are crime rate, CPI are inflation, and UMP are unemployment and URB for urbanization.

$$TFP_t = \beta0 + \beta_1 CR_t + \beta_2 CPI_t + \beta_4 UMP_t + \beta_5 URB_t + \mu_t \ \ldots\ldots (4.4)$$

The regression results of above model on the annual time series data for the period 1982-2020 are provided in Table 4.4

Table 4. 4. Regression results of M4

Dependent variable TFP

	Coefficient	**Standard Error**	**t-value**	**t-prob**
Constant	0.053	6.844	0.007	0.993
CR	1.308	3.078	0.425	0.673
CPI	-1.457	0.951	-1.53	0.134
UMP	0.018	0.101	0.183	0.855
URB	0.206	0.289	0.711	0.481
Std. Error	**RSS**	**R2**	**F (4,34)**	
0.4144	5.840	0.657	16.29 (0.000) **	

M:05 Gehringer et al. (2016) analyzed the factors influencing TFP in the European Union. Their model's dependent variable is the TFP. FDI estimates, FDI as a proportion of GDP. The ICT quantifies the contribution of ICT capital services to specific sectors. Research and development expenditures as a proportion of sectoral. As proxies for trade openness, the sum of imports and exports as a proportion of GDP was utilized. Human capital is approximated by the proportion of individuals having a secondary education and the error term (wt).

$$TFP = \alpha + \beta_1 FDI_t + \beta_2 ICT_t + \beta_3 R\&D_t + \beta_4 TO_t + \beta_5 EDU_t + \beta_5 SEC_t + w_t \ \ldots\ldots (4.5)$$

The regression results of above model on the annual time series data for the period 1982-2020 are provided in Table 4.5

Table 4. 5. Regression results of M5

Dependent variable TFP

	Coefficient	Standard Error	t-value	t-prob
Constant	2.231	1.634	1.37	0.181
FDI	0.149	0.175	0.849	0.401
ICT	-0.522	0.301	-1.73	0.092
R&D	-1.201	0.878	-1.37	0.180
TO				
EDU	0.554	0.503	1.10	0.278
SEC	-0.882	0.943	-0.935	0.356
Std. Error	**RSS**	**R2**	**F(5,33)**	
0.3924	4.826	0.716	16.69 (0.000) **	

M:06 Now we run the regression on the same model on the annual time series data for the period of 1982-2020 and then we get the below results which are provided in the Table 4.6.

Table 4. 6. Regression results of M6

Dependent variable TFP

	Coefficient	**Standard Error**	**t-value**	**t-prob**
Constant	0.610	2.163	0.282	0.779
TO	0.614	1.101	0.599	0.580
FDI	0.123	0.162	0.716	0.452
FD	-0.011	0.018	-0.635	0.530
IR	-0.010	0.038	-0.272	0.787
CPI	-0.634	0.234	-2.71	0.011
SOG	0.023	0.071	0.332	0.742
EDU	0.620	0.847	0.732	0.469
HET	0.227	0.696	0.326	0.746
Std. Error	**RSS**	**R2**	**F (8,30)**	
0.3998	4.796	0.718	9.566 (0.000) **	

The standard errors of all the above estimated models are given in the below Table 4.7. Here we want to see which of the estimated models has minimum standard error.

Table 4. 7 Standard Errors of all Existing Models

Model	Sigma/ S.E Value
M1	0.392
M2	0.383
M3	**0.382**
M4	0.414
M5	0.392
M6	0.399

It has been observed from the above Table 4.7, that M3 has the minimum standard error i.e., 0.382187 as compared to all other estimated models. So, the M3 is our best model of all the existing models. While using encompassing approach we must check whether M3 encompasses all the other existing models or not. When M3 encompasses all the other existing models, than it means that the prediction power of all existing models which are encompassed by M3, is already presented in M3. So, we ignore all those existing models which are encompassed by M3. On the other hand, if the M3 does not encompass a model, then we cannot ignore that model. Therefore, all those existing models which do not encompass M3 will be put aside. Then we will take union of independent variables of M3 and all the other existing models which are not encompassed by M3. Finally, in this way we get a generalized unrestricted model.

Now we have to test whether M3 and M2 encompasses Mi or not where i=M3 By using Cox and Ericsson test the result are reported in Table 4.8

Table 4. 8 Encompassing Results

Encompassing Hypothesis	**Cox Test and Ericsson Test (P-values)**
H_o^1 = M3 Encompasses M1	-0.6654 [0.5058] 0.5903 [0.5550]
H_o^2 = M3 Encompasses M2	-12.16 [0.0000] ** 10.71 [0.0000] **
H_o^3 = M3 Encompasses M4	-0.5346 [0.5929] 0.4813 [0.6303]
H_o^4 = M3 Encompasses M5	-0.4625 [0.6437] 0.4125 [0.6800]
H_o^5 = M3 Encompasses M6	-0.3058 [0.7598] 0.2755 [0.7829]

Table 4. 9 Encompassing Results

Encompassing Hypothesis	**Cox Test and Ericsson Test (P-values)**
H_o^1 = M2 Encompasses M1	0.4027 [0.6028] 0.3767 [0.6852]
H_o^2 = M2 Encompasses M3	-8.26 [0.0005] 8.52 [0.0000]
H_o^3 = M2 Encompasses M4	-0.6506 [0.5237] 0.5635 [0.4307]
H_o^4 = M2 Encompasses M5	0.6813 [0.3257] 0.3250 [0.7822]
H_o^5 = M2 Encompasses M6	-0.4206 [0.5579] 0.3543 [0.5747]

The results of the above Table 4.8 and 4.9 indicates that the p-value of all the tests is greater than 0.05. Therefore, we have failed to reject it in all the cases. According to Cox and Ericsson test M2 does not Encompass M3 in table 4.8 and M3 does not Encompass M2 in table 4.9. So, M3 encompasses all the existing models except M2 in table 4.8 and M2 encompasses all the existing models except M3 in table 4.9, which prob value is less the 0.05 and we consider M3 and M2 as our general model.

$$\text{GUM:} \quad gr(TFP)t = \beta0 + \beta1\ lnFDIt + \beta2\ lnTOt + \beta3\ lnEDUt + \beta4\ lnSECt + \beta5\ lnUMPt + \beta6\ lnCPIt + \beta7\ lnPTNt + \beta8\ lnPTRt + \beta9\ lnIMMt + et \ \ldots\ldots (4.7)$$

4.1.1 General to Specific Model

In the encompassing approach we have observed that M3 has the minimum standard error and encompasses the rest of the existing models. Therefore, we consider the M3 as our best model. Finally, we make our general model with the help of M3. Now we check the significance of all variables in the general model. In the general model there may be such independent variables which may have insignificant impact on the dependent variable. We omit all those independent variables which have insignificant impact on the dependent variable.

Now we estimate the model, and the results are as follows.

Table 4. 9 Steps of General to Specific Model

Variable	Coefficient	St. Error	t-value	t-prob
gr(TFP)_1	0.084	0.139	0.607	0.563
gr(TFP)_2	0.073	0.170	0.431	0.680
Constant	-18.209	6.793	-2.680	0.032
ln FDI	0.579	0.248	2.340	0.052
ln FDI_1	-0.098	0.239	-0.410	0.694
ln FDI_2	-0.536	0.199	-2.690	0.031
Ln TO	1.110	0.891	1.240	0.253

Ln TO_1	-1.829	0.959	-1.910	0.098
Ln TO_2	-0.524	0.914	-0.574	0.584
Ln Edu	-2.418	0.861	-2.810	0.026
Ln Edu_1	0.903	0.808	1.120	0.301
Ln Edu_2	0.321	0.560	0.574	0.584
Ln Sec	4.537	2.626	1.730	0.128
Ln Sec_1	1.063	4.501	0.236	0.820
Ln Sec_2	-2.363	3.030	-0.780	0.461
ln UMP	-0.216	0.164	-1.320	0.230
ln UMP_1	0.052	0.177	0.293	0.778
ln UMP_2	-0.593	0.202	-2.940	0.022
Ln CPI	-16.981	4.486	-3.790	0.007
Ln CPI_1	-0.495	4.387	-0.113	0.913
Ln CPI_2	14.692	2.473	5.940	0.001
PTN	0.201	0.065	3.110	0.017
PTN_1	-0.001	0.001	-2.210	0.063
PTN_2	-2.552	0.001	-0.043	0.967
ln PTR	0.298	0.338	0.881	0.407
ln PTR_1	-2.279	0.478	-4.770	0.002
ln PTR_2	0.451	0.362	1.250	0.253
ln IMM	0.883	0.656	1.340	0.221
ln IMM_1	-1.647	0.825	-2.000	0.086
ln IMM_2	4.685	0.914	5.130	0.001
sigma	0.189		**RSS**	0.249

R^2	0.985		Adj.R^2	0.921
no. of observations	37		**no. of parameters**	30

It can be observed that the forecast performance of the model is good, and there is no problem with the two forecast tests. To see whether this model is adequate as a starting point, I am applying battery of tests, and the results are as follows:

Table 4. 10 Battery of tests

AR 1-1 test: F(1,6) = 3.3423 [0.1173]
ARCH 1-1 test: F(1,35) = 1.9752 [0.1687]
Normality test: Chi^2(2) = 0.32035 [0.8520]
RESET23 test: F(2,5) = 0.14568 [0.8680]

We observe that there is no autocorrelation in the model which implies that the model does not need further extension. There is no heteroskedasticity and non-normality in the model. Therefore, the output is a good starting point. Now if we look back at the output, we see that there are many variables whose p-value is larger than 5%, which means that these variables are insignificant. However, it is important to note that these t-values only indicate exclusion of one variable at a time, we cannot exclude all variables which are significant on basis of t-value.

For example, we see that p-value for LnTO in the output is 0.25and for LnTO_2 it is 0.58. The p-value for LnTO indicates that if all other variables remain in the model, we can exclude LnTO. But this does not indicate that we can exclude both LnTO and LnTO_2 simultaneously. If we want to exclude the two regressors simultaneously, we need to test joint restriction on the two variables.

We have applied the Lag Structure analysis to the model estimated above and the output is as follows:

Table 4. 11 Tests on the significance of each variable

Variable	F-test		Value [Prob]
gr(TFP)	F(2,7)	=	0.23377 [0.7975]
Constant	F(1,7)	=	7.1859 [0.0315]*
Ln FDI	F(3,7)	=	6.7606 [0.0178]*
Ln TO	F(3,7)	=	3.0256 [0.1030]
Ln Edu	F(3,7)	=	4.1092 [0.0564]
Ln Sec	F(3,7)	=	2.2054 [0.1752]
Ln UMP	F(3,7)	=	7.7033 [0.0128]*
Ln CPI	F(3,7)	=	15.375 [0.0018]**
Ln PTN	F(3,7)	=	4.2400 [0.0527]
Ln PTR	F(3,7)	=	13.067 [0.0030]**
Ln IMM	F(3,7)	=	12.520 [0.0034]**

Table 4. 12 Tests on the significance of each lag

Variable	F-test		Value [Prob]
Lag 2	F(10,7)	=	6.2334 [0.0119]*
Lag 1	F(10,7)	=	4.6372 [0.0268]*

Table 4. 13 Tests on the significance of all lags up to 2

Variable	F-test		Value [Prob]
Lag 2 - 2	F(10,7)	=	6.2334 [0.0119]*
Lag 1 - 2	F(20,7)	=	5.6223 [0.0130]*

In the top panel, there are results of joint restriction on all lags of variables used as regressors. For example, two lags of FDI.

The test for this hypothesis is summarized, as LnFDIt, with F-stat=6.7606 and P-value =0.0178

The P-value is so small, which means the rejection of null hypothesis. Thus, we have rejected the restriction. Remember that this rejection does mean that all of the betas are zero. If only one of the betas is non-zero, the restriction shall stand rejected. On the other hand, if null hypothesis was accepted, this would mean that all coefficients of lags of FDI are zero. Anyhow, the rejection of H0 implies that we cannot exclude all lags of FDI at once. However, it could be possible that we exclude a few lags of LnFDI retaining other variables. But for this purpose, we would have to apply for a separate test.

Similarly, restriction on all lags of CPI summarized that, we have 2 lags of CPI. We observe that F-stat=15.375 with P-value=0.0018 which means that the restriction is rejected, and we cannot exclude all 2 lags of CPI simultaneously. Remember that this does not deny that we can exclude some of lags of CPI retaining the other in the model. But for that purpose, separate testing would be needed.

The results also indicate that restriction on all legs of SEC could not be rejected, therefore we are allowed to exclude SEC from the model if the remaining model is unchanged.

Now if we come to second panel, "Tests on the significance of each lag". This is a restriction on specified lag of all variables. Restriction on 2^{nd} lags of all variables, we observed that F-stat for this hypothesis is 6.2334 with p-value 0.0119. Since the p-value is less than 5% benchmark, the restriction would be considered valid and the 2nd lags of all variables could not be excluded from the model,

Similarly, we observe that the 1st lag of all variables, which F-stat is 4.6372 and p-value is 0.0268, it's concluded that, we could not be excluded 1^{st} lag of all variables from the model.

Now we estimate the model and exclude those variables one by one, which are highly insignificant. The following table are the excluded variables from our model.

Table 4. 14 Excluded variables

Variable	**Lags**	**t-prob**	**Remark**
FDI	ln FDI_1	0.721	Excluded
EDU	Ln Edu_1	0.494	Excluded
FDI	ln FDI	0.126	Excluded
EDU	Ln Edu_2	0.409	Excluded
UMP	ln UMP	0.545	Excluded
UMP	ln UMP_1	0.492	Excluded
PTN	PTN_1	0.127	Excluded
PTN	PTN_2	0.300	Excluded
PTR	ln PTR	0.295	Excluded
PTR	ln PTR_2	0.662	Excluded

After the exclusion of insignificant variables, the final specific model under encompassing is given in the following equation.

$$\mathbf{Gr(TFP)_t = \beta_0 + \beta_1\, lnFDI_t + \beta_2\, lnEDU_t + \beta_3\, lnUMP_t + \beta_4\, lnCPI_t + \beta_5\, lnPTR_t + \beta_6\, lnPTN_t + \beta_7\, lnIMM_t + e} \quad \ldots\ldots (4.8)$$

4.2 Specifying Model by Using LASSO Technique

In this section our objective is to use the LASSO technique to find the best suitable model of TFP in Pakistan. LASSO regression is a powerful method which performs two main responsibilities such as the regularization and the feature selection. We use this technique when there is more independent variables and high multicollinearity in the model. In this model when the variables are insignificant or do not have relationship with the response variable then the lasso makes their coefficient approximately equal to zero and finally drop them from the model. The objective of the LASSO is to minimize the prediction error. Moreover, the main purpose of feature selection process is to omit those variables which are redundant, to make the model easier to interpret and to reduce the overfitting. The feature selection is a very important task because here the number of variables are very high and sometime the number of variables are greater than the number of observations. In this case it is not easy to say which one of the variables is relevant and which one is irrelevant. Therefore, the feature selection process has a great importance.

In this study our aim is to identify the potential by using the LASSO method. We have considered the covariates of TFP provided by the five models(namely, , M1, M2, M3, ...M6.) in case of Pakistan to check whether these all determinants have significant or insignificant impact on covariates of TFP of Pakistan.

4.2.1 Computations

To perform the computation of the model we have used R. The glmnet package is used for the LASSO computation. To use coordinate descent method, we fit the command of glmnet. To choose the non-zero coefficients and best lambda(λ) by cross-validation, we use the following codes.

```
out=glmnet(x,y,alpha=1,lambda=grid)

lasso.coef=predict(out,type="coefficients",s=bestlam)[1:28,]

lasso.coef

bestlam.
```

4.2.2 The Value of Tuning Parameter

The tuning parameter controls the strength of the penalty. The parameter coefficient (β) correlates with the value λ. As increasing the value of lambda(λ), the more coefficients are set to be equal to zero and in this situation only few variables are selected for our model. The range of tuning parameters is between zero to infinity and it is a crucial value for the identification of the true model. Moreover, LASSO is a feature selection process which helps us to make a general model. In our case we choose the value of λ and select the non-zero coefficients with the help of cross validation (Cross validation is often used to choose the value of λ for the LASSO estimator). In such a way we get the general model which is shown in the following Table 4.15. In the general model the lambda(λ) value is equal to 0.014 and fourteen non-zero coefficients have been selected.

Table 4. 15 non-zero coefficients & values (λ=0.014)

Variable	Coefficient	Variable	Coefficient
(Intercept)	3.405773	**ln DE1**	-0.0033004
Ln CPI	-0.0404004	**Ln CPI**	-0.0404004
ln PTR	-0.0617154	**ln PTR**	-0.0617154
URN	-0.0030810	**URN**	-0.0030810
ICT	-0.1105329	**ICT**	-0.1105329
ln M2	-0.00500069	**ln M2**	-0.0050006
CR	-0.000000025		

4.2.3 General to Specific Model

In the general model we check the significance of all variables and there may be such independent variables which may have insignificant impact on the dependent variable. We omit all those independent variables which have insignificant impact on the dependent variable.

In general, to specific methodology, we exclude the variables based on joint restrictions. First, we choose the highly insignificant level or lagged level variable based on t- value and p-value then impose joint restrictions via F- test. With null hypothesis both level and lagged level variables are insignificant against alternative at least one of these is significant. If we fail to reject the null hypothesis, then we retain the variable in the model otherwise drop it from the model. After the exclusion of insignificant variables, the final specific model of LASSO is given.

$$\mathbf{Gr(TFP)_t = \beta_0 + \beta_1\, lnCPI_t + \beta_2\, lnPTR_t + \beta_3\, URN_t + \beta_4\, lnICT_t + \beta_5\, lnFD_t + \beta_6\, lnCR_t + \beta_7\, lnDE_t + e} \quad \ldots\ldots (4.9)$$

4.3 Final Model for TFP

Now we choose the final model for TFP from the above selected specific models from encompassing and LASSO in equation (4.8) and (4.10). First rank the models according to their minimum standard error provided in the Table 4.16.

Table 4. 16 Standard. Errors of Models

Model	Standard Error
TFPM $_{\textbf{Encompassing}}$	2.081
TFPM $_{\textbf{LASSO}}$	2.286

It has been observed from the above Table 4.16, that TFP model from Encompassing has the minimum standard Error i.e., 2.081 as compared to the TFP model from LASSO . So, the TFP model from Encompassing is our best model of the above two models. While using encompassing approach we must check whether model from Encompassing encompasses the model from LASSO or not. When model from Encompassing encompasses model from LASSO, then it means that the prediction power of model from LASSO, which are encompassed by model from Encompassing , is already presented in model from Encompassing . So, we ignore the model from LASSO. By using COX test the result are reported in Table 4.17

Table 4. 17 Encompassing Results (COX test)

Model	Test statistics	Value
TFPM $_{\textbf{Encompassing}}$ **Encompasses TFPM** $_{\textbf{LASSO}}$	(COX test)	-0.316 [0.751]

The results of the above Table 4.17 indicates that the p-value of the COX test is greater than 0.05.

Therefore, we are failed to reject H_0. So, $TFPM_{Encompassing}$ Encompassas $TFPM_{LASSO,}$ and we consider TFPM $_{Encompassing}$ is our final model for TFP in Pakistan which is given in the equation.

$$\mathbf{Gr(TFP)_t = \beta_0 + \beta_1 \ lnFDI_t + \beta_2 \ lnEDU_t + \beta_3 \ lnUMP_t + \beta_4 \ lnCPI_t + \beta_5 \ lnPTR_t + \beta_6 \ lnPTN_t + \beta_7 \ lnIMM_t + e} \quad \ldots\ldots \ (4.10)$$

4.3.1 Unit Root Test (Stationary Test)

To estimate any regression model, it is necessary to describe the order of integration of variables. The series will be non-stationary if it has unit root problem. Whenever we continue and estimate those variables which have problems of unit root then it produces meaningless or spurious regression. Therefore, in this study we use Augments Dickey Fuller(ADF) test to check the stationary properties of the data. ADF procedure was established by Dickey and Fuller in 1981 to test for non-stationarity.

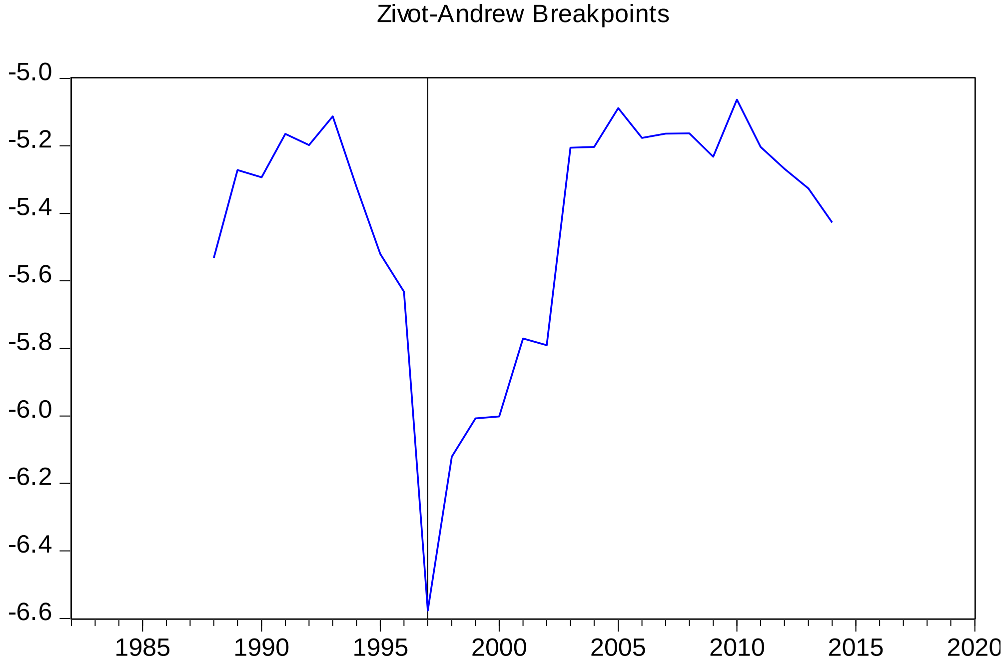
Table 4. 18 Unit Root Test

Variable	**At level**		**At 1st difference**		**Conclusion**
	t-stat	**Prob value**	**t-stat**	**Prob value**	
gr(TFP)	-1.219	0.655	-11.27	0.000	1(1)
ln FDI	-1.642	0.451	-5.845	0.000	1(1)
Ln Edu	-2.842	0.061	____	____	1(0)***
ln UMP	-1.935	0.313	-6.059	0.000	1(1)
Ln CPI	-0.156	0.935	-2.904	0.054	1(1)
ln PTR	0.721	0.991	-7.052	0.000	1(1)
Ln PTN	-5.487	0.000	____	____	1(0)
ln IMM	-0.713	0.830	-4.992	0.000	1(1)

After checking the order of integration of variables in the Table 4.18. It is confirmed that variables are stationary at different levels. Some are stationary at level and others are at first difference.

Table 4. 19 Unit Root Test with structural break

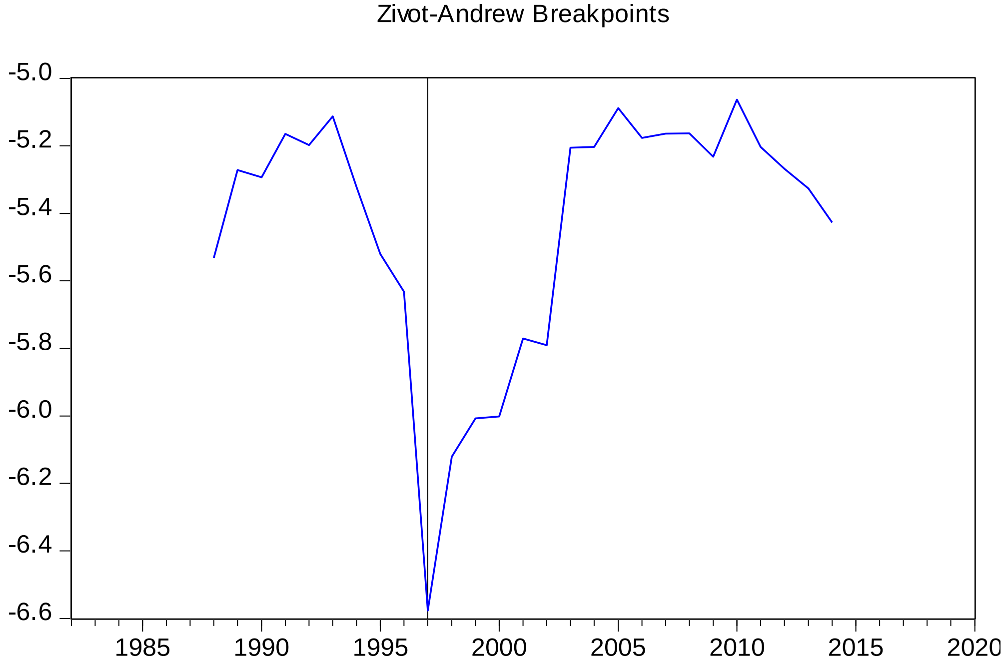
Variable	t-stat	Prob value
gr(TFP)	-6.576516	0.001340



Syed (1998) Pakistan fiscal 1997 as the most disappointing year in the country's history. A persistent recession shut down almost 7,000 industrial units, an interest rate of nearly 20% for domestic borrowing discouraged in- vectors and entrepreneurs, inflation remained high (government sources placed it at 13% and others at 20% or even higher), the budget deficit exceeded 6% of GDP, the trade deficit ranged between five and seven billion dollars, and the debt burden mounted.

Table 4. 19 Unit Root Test with structural break

Variable	t-stat	Prob value
gr(TFP)	-6.576516	0.001340

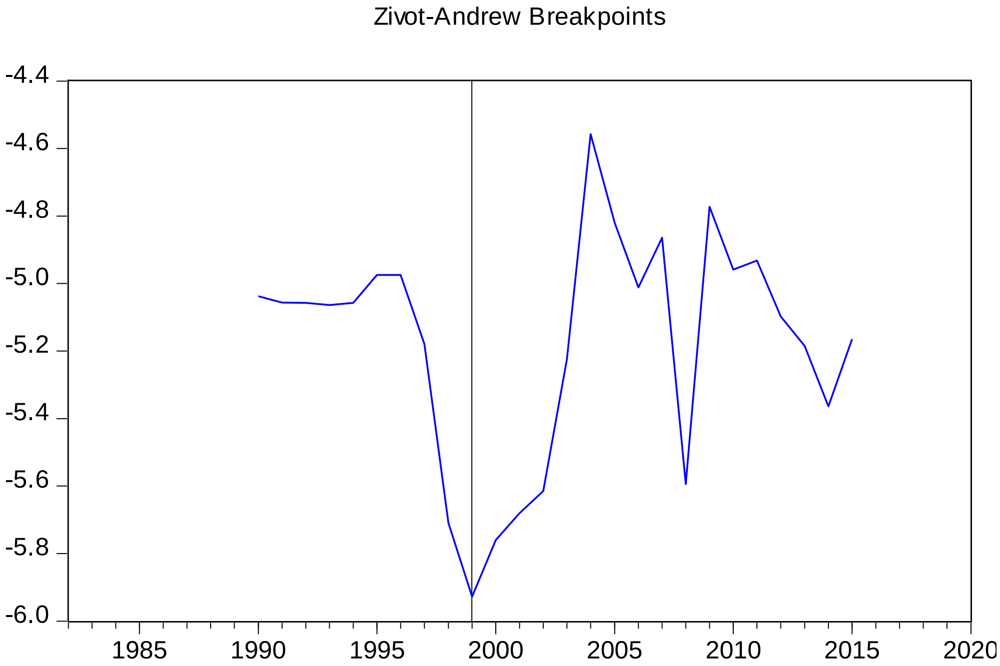


Syed (1998) Pakistan fiscal 1997 as the most disappointing year in the country's history. A persistent recession shut down almost 7,000 industrial units, an interest rate of nearly 20% for domestic borrowing discouraged in- vectors and entrepreneurs, inflation remained high (government sources placed it at 13% and others at 20% or even higher), the budget deficit exceeded 6% of GDP, the trade deficit ranged between five and seven billion dollars, and the debt burden mounted.

Table 4. 19 Unit Root Test with structural break

Variable	t-stat	Prob value
CPI	-5.927606	0.039844

Zivot-Andrews Unit Root Test (cpi)

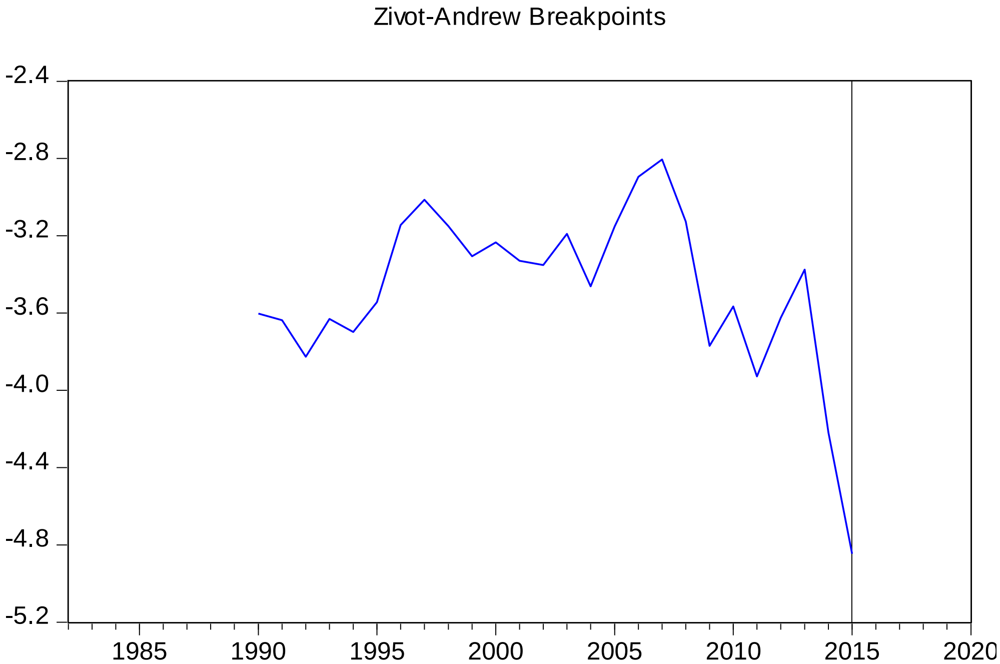


The inflation rate, which was at 5.7 percent in 1998-99, was further reduced to 3.1 percent in 2002-03 (the lowest in the last three decades). This low level of inflation was supported by strict fiscal discipline, the lower monetization of the budget deficit, an output recovery, a reduction in duties and taxes, and appreciation of exchange rate.

Table 4. 19 Unit Root Test with structural break

Variable	**t-stat**	**Prob value**
FDI	-4.846694	0.002155

Zivot-Andrews Unit Root Test (fdi)



This huge downfall in the FDI is observed due to certain reasons which are political instability, financial instability, terrorism, and energy crisis in the country.

4.3.2 ARDL Bounds Test

The ARDL Bounds test is used when the time series data is integrated of different order or integrated of I(0) and I(1). We use this procedure to study the co-integration relationship between TFP and its determinants in case of Pakistan. In our study in equation (4.10) some variables are integrated of I(0) such as expenditure on education and nonresidential patent and some are integrated of I(1) such as TFP, Foreign direct investment, unemployment rate, consumer price index, residential patent and imported machinery of Pakistan. For this reason, we use ARDL Bounds testing procedure. It should also be noted from the final step of general to specific methodology provided in Table 4.11. The appropriate lag selection of ARDL((1, 0, 2, 2, 2, 2, 2, 2, 2, 1, 2).

Table 4. 19 ARDL Bound Test Results

Test statistics	**value**	**K**
F-stat	11.984	7
Critical value		
Significance level	**Lower bound value**	**Upper bound value**
10%	1.92	2.89
5%	2.17	3.21
2.5%	2.43	3.51
1%	2.73	3.9

At 5% level of significance the calculated value is 11.984 which is greater than the tabulated value of upper bound I(1) i.e., 3.21 value. So, we reject it, and it indicates that there exists a long run relationship. In case of cointegration it is preferable to estimate the ECM of ARDL model, which is reported in Table 4.20.

Table 4. 20 ARDL Model Bound test long run result.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LN_EDU	-1.836	0.886	-2.072	0.054
LN_IMM	5.524	1.274	4.337	0.000
LN_CPI	-2.715	0.720	-3.773	0.002
LN_UMP	-0.572	0.178	-3.204	0.005
LN_FDI	-0.656	0.445	-1.473	0.159
LN_PTN	1.510	0.510	2.963	0.009
LN_PTR	-1.138	0.327	-3.481	0.003
C	-20.151	6.923	-2.911	0.010

Adjusted R2	S.E. of regression	F-statistic	Prob(F-statistic)
0.883	0.228	15.345	0.000

The above Table 4.20 shows that in the long run the expenditure on education have a negative and significant impact on TFP by looking at the t-statistics[-2.072] and probability value(0.054) . From the coefficient value a 1% increase in educational expenditure will decrease the total factor productivity growth by 1.83%. The detrimental effect of educational expenditure in Pakistan is due to the quality of education. In case of Pakistan there is not much attention to the quality of education and higher enrollment rates therefore educational expenditure does not improve TFP. The similar results have been reported by (O. Siddique, 2022) and (Khan, 2005).

In table 4.20 imported machineries shows highly significant and positive relationship with TFP in long run, having t-statistics value is [4.337]and probability value (0.000). From the coefficient value a 1% increase in imported machinery will bring a 5.52% change in TFP. The country with higher IMM are more likely to get high benefit from technology diffusion as compared to country with lower IMM, this is because import bring machinery with advance technology which bring innovation and advancement in local domestic production. Our result is supported and line with (Saleem et al., 2019) and (Grossman & Helpman, 1993).

The consumer price index is used as proxy for inflation. Which have negative and significant impact on TFP of Pakistan in long run relationship, having coefficient is (-2.715), t-state [-3.773] and prob value (0.001). Parallel findings are found in the studies of (Akinlo & Adejumo, 2016). The result indicates that a 1% change in inflation rate will bring a 2.71% decrease in TFP growth in the long run. This is reliable with a priori prospect. The inflation will affect the TFP negatively because the high and unstable prices will cause uncertainty in the economy due to which local and foreign investor get discourage and they afraid to take risk in mega long run projects. The consequences will be observed in the long run such as a reduction in Total factor productivity growth. It could be seen from the result that even inflation is significant with negative sign, which again supports that when inflation increases it reduces the TFP of the economy. Because this money illusion reduces the labour hours worked, it even affect the investment. Low and stable inflation provide a favorable environment for TFP. In a short period of time mixed results.

Table 4.20 shows that in the long run the unemployment rate have a negative and significant impact on TFP by looking at the t-statistics[-3.203] and probability value(0.005). When there is a 1% increase in unemployment rate in the economy, it will lead to reduce the total factor productivity growth by 0.57% in long run. Our result supported by (McConnell & Brue, 2017) and (Akinlo &

Adejumo, 2016). Due to the high unemployment, the rate of productivity became low and overall output lead to be reduced.

Patent residents have negative and significant effect on TFP in Pakistan. Whereas institutions (such as Scientific research centers, Universities, and policy making organizations) play a key role in innovations and advancement in productions and technology to capture international markets. But unfortunately, in Pakistan, the innovation network between scientific research centers, firms and universities exists at a certain point, which indicates the patents are not beneficial to innovations in the production. While non-residential patent have positive and significant effect in Total productivity of Pakistan. Mirzadeh and Nikzad (2013) innovations and human creativity (patents) have greater impact on country future. Where many researchers have taken patent nonresidents as proxy for the creativity of humans and innovations in the productivity.

4.3.3 Error Correction Model

ECM also provides the rate of adjustment, or it capture convergence in the long run after any short run shock or disequilibrium. For convergence ECM coefficient must be negative and significant.

Table 4. 21 ECM Regression for short run coefficient and long run adjustment

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(LN_IMM)	2.081	0.289	7.188	0.000
D(LN_IMM(-1))	-4.529	0.391	-11.571	0.000
D(LN_CPI)	-14.252	1.726	-8.259	0.000
D(LN_CPI(-1))	-16.105	1.537	-10.476	0.000
D(LN_UMP)	0.007	0.081	0.085	0.933
D(LN_UMP(-1))	0.617	0.074	8.354	0.000
D(LN_FDI)	0.184	0.094	1.958	0.067
D(LN_FDI(-1))	0.535	0.096	5.582	0.000
D(LN_PTN)	2.399	0.321	7.482	0.000
D(LN_PTR)	0.043	0.167	0.257	0.800
D(LN_PTR(-1))	-0.530	0.147	-3.606	0.002
CointEq(-1)*	-0.917	0.073	-12.594	0.000

In the above Table 4.21, Imported machineries, unemployment, foreign direct investment, residential and nonresidential patents have positive and significant effects on total factor productivity in the short run. While Inflation have negative and significant effect on total factor productivity in short run. The ECM coefficient is negative and highly significant. So, we can say that there is convergence towards long run equilibrium after short run shock. The coefficient of is equal to -0.917 and p-value is highly significant. It means that 91% adjustment will occur in one period.

4.3.4 Diagnostic Test

The residuals of the above final model has satisfied the diagnostic tests of Breusch Pagon and Godfrey(1981) LM test of no serial correlation (F-stat= 0.770 and P-value= 0.480), Engle's (1982) ARCH test of no ARCH effect (F-stat= 0.641 and P-value = 0.8253 and Jarque-Bera normality (Jarque-Bera=2.003 and P-value= 0.367) at 5% level of significance.

4.3.5 Stability Test

Now we use cumulative sum (CUSUM) and cumulative sum of square (CUSUMSQ) tests to check the stability of the parameters of Total factor productivity. The null hypothesis is that the parameters are stable. So, we do not reject the null hypothesis because the plot of CUSUM and CUSUM square lies inside the critical bounds at 5% level of significance. Their results are given in the following Figure 4.1 and Figure 4.2. In both graphs, CUSUM and CUSUM of square we can clearly see that the lines are in between the range at 5% level of significant, which means that our model is stable in both cases.

Figure 1 CUSUM

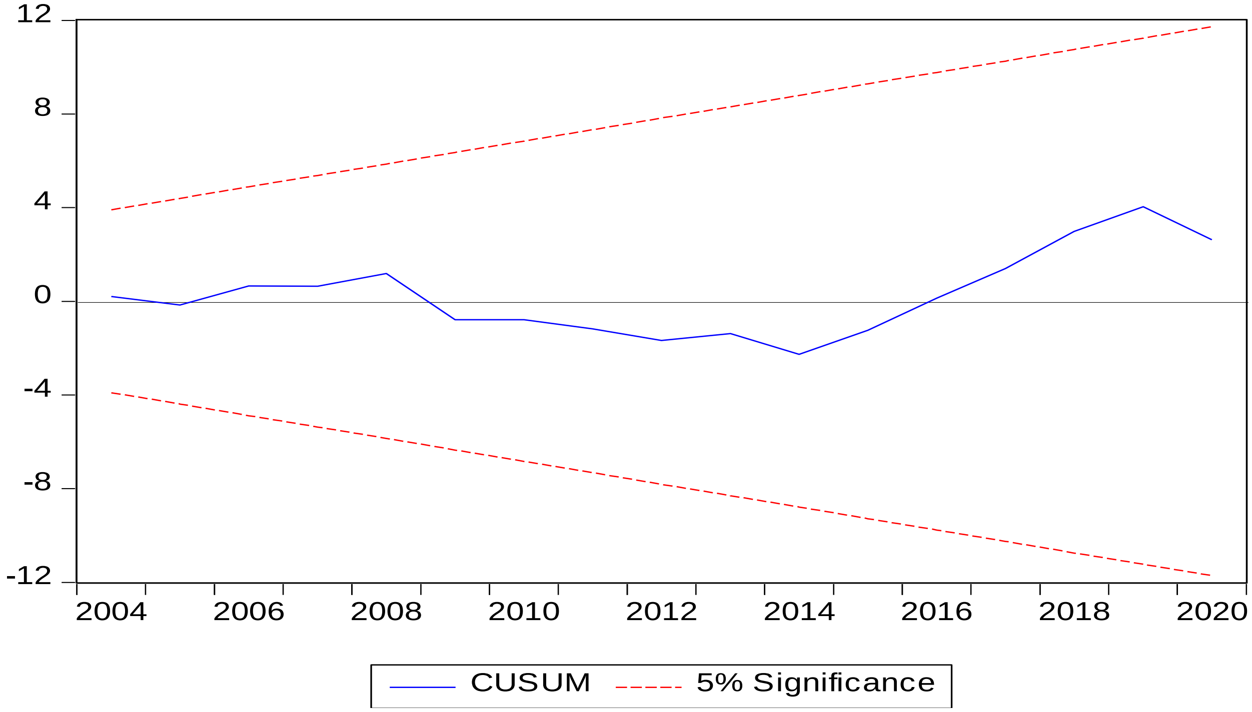
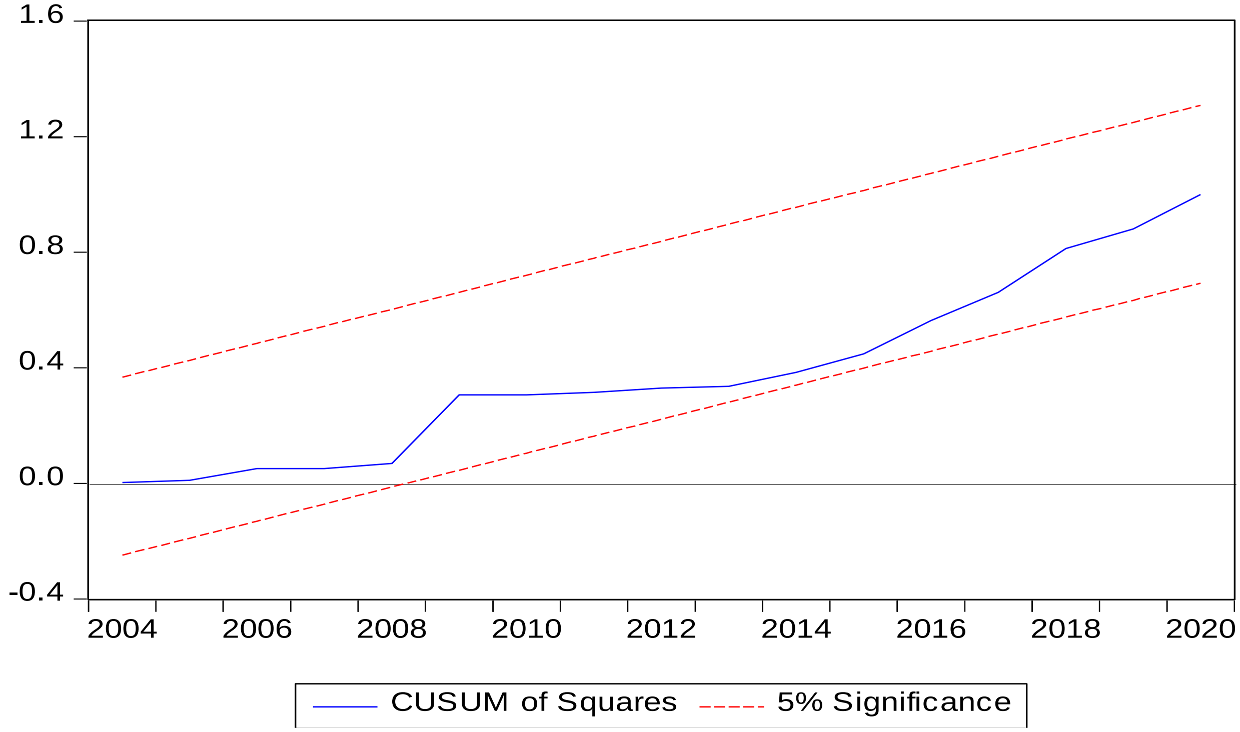


Figure 2 CUSUMSQ



CHAPTER 5

QUALITATIVE WORK

To understand the major factors affecting the productivity of growth, I conducted a qualitative interview. For that purpose, I visit planning commission of Pakistan and conducted face to face detailed interview related to growth strategy with the experts there. I asked questions, such is

1. What is the main contributing determinants of TFP?
2. If we want to improve our economic growth, which factors is more important to change it and get improvement in our productivity?
3. What are the way out if Pakistan's have low productivity?
4. What will be your suggestion to government of Pakistan related to improvement or change in policy of productivity?

Key suggestion regarding TFP and its covariates and the way out by growth expert of planning commission of Pakistan.

5.1 Human capital

Immediate investment in human capital is also essential to increase productivity. Better education and health conditions increase the physical strength of workers, decrease the number of absences, enhance their ability to concentrate on their work, and allow them to maximize the performance of machines.

In addition, a child's improved health improves his or her educational performance. A strong public health system that provides quality services to knowledge portions of the population will also encourage people to save and invest in education and training.

Similarly, providing young children with access to a high-quality education that balances theoretical scientific knowledge with practical learning at an early age will be beneficial for enhancing their creativity and future productivity.

Similarly, students studying science and technology, particularly engineering, must receive training in their respective professions through hands-on experience during their engineering coursework. To achieve this goal, labs must be modernized, course outlines must be revised to place a greater focus on the importance of learning, and strong ties to the business must be developed to help students and staff understand the skillset and level in demand.

Establishing technology parks and centers of excellence on or near university grounds that receive work contracts from the world's leading manufacturers of automobiles, construction machinery, agricultural products, and IT goods will assist students in learning and developing a professional attitude.

Such a strong partnership with the industry will also be advantageous for academia in terms of generating financial resources and, consequently, enhancing their infrastructure and faculty's skill level.

The initial financial resources required for such an upgrade of productivity and better universities can be easily created by selling loss-making government firms.

In addition, enhancing environmental circumstances and promoting sports, entertainment, art, and intellectual culture broadens perspectives and enhances creativity. This ultimately results in increased production.

To create a large pool of skilled construction workers, plumbers, carpenters, and technicians who can work in the construction industry, repair IT machines, household appliances, automobiles, etc., it is necessary to reform institutions and authorities such as the Technical Education and Vocational Training Authority (Tevta) that offer short technical courses and training.

Following the recent announcement by the government of a concessionary package for the construction and housing industries, an increase in the need for skilled labour is projected.

Given the large proportion of the population with no or very little formal education, this area must be emphasized to combat unemployment and low labour productivity. For this objective, support from technologically advanced nations such as Japan and Germany can be extremely useful.

5.2 Innovation, technology, and research & development

Low productivity is an issue impacting the output of all industries, including agriculture, manufacturing, trade, education, health care, and other service industries. Low productivity causes a rise in production costs, which decreases producers' profits.

This causes a price increase, which increases the burden on consumers and leads to a loss of international competitiveness. Therefore, it is a significant barrier to industrialization and requires urgent action if we want the manufacturing sector to expand, generate jobs for the youth, earn foreign exchange by expanding the quantity and variety of exports, and grow sustainably.

The International Labour Organization (ILO) calculates that China's output per person, a measure of labour productivity, grew by 388 percent, India's by 177 percent, and Bangladesh's by 109 percent between 2000 and 2019, while ours increased by only 32 percent.

Given the multiple causes of low labour productivity, it will take a considerable amount of time and effort in multiple areas to tackle this issue. Modern machinery and technological adaptation boost the efficiency of labour, allowing to produce higher-quality commodities in less time.

Unfortunately, our industry does not produce high-tech machines that could automate the manufacturing of goods and the delivery of services. Therefore, we must import machinery.

Given the historical issue of a massive balance of payments (BOP) deficit as well as other type of government-regulated economy we have, policymakers and finance experts frequently choose to restrict imports by increasing taxes and levies and even prohibiting some imports.

In addition, the financial, cognitive, and time costs of machinery import are excessively high due to an excessive reliance on imports for revenue generation and ineffective port and regulatory mechanisms.

As a result, companies typically prefer to retain outdated, inefficient capital, which eventually reduces labour productivity and raises the price of the final product. This delivers businesses uncompetitive, leading them to seek government protection to survive.

Therefore, it is necessary to update the local tech universities before promoting graduates to develop manufacturers who make high-quality, energy-efficient machinery and plants.

Reforming and bolstering the financial industry and key government institutions, such as the SECP and tax authorities, to ensure the smooth procedures is indeed needed.

5.3 Capitalist Economy

Last but not least, a market-based economy in which market entry and leave are easy, the cost of doing business is cheap, the financial sector is highly established, and the system supports competition may be the least expensive and most successful strategy for boosting productivity.

Such an economic system, in which subsidies, bailouts, and government exemptions for certain enterprises are absent, trade is open, institutions give a level playing field, and competition is encouraged, forces everyone to constantly improve.

Everyone has equal access to opportunities, and the market system does not favor any sector, organization, or person. Therefore, everyone strives for excellence. Those who do not improve themselves are immediately left behind, while those who work diligently, adapt to changing conditions, learn new skills, and update to better technologies are rewarded.

In a market system built on competition and non-favoritism, only the strongest and most adaptive would remain. Therefore, organizations and individuals must continually increase their productivity, which eventually increases their competitiveness on the international market. This increases exports and helps nations earn foreign currency.

A market-based economic system with unrestrained trading also allocates scarce financial and human resources to productive industries. This improved resource distribution increases production.

Due to efficient resource allocation, Mexico was able to increase productivity by 41% between

On the other hand, research indicates that Japan's lackluster productivity development after the 1990s was due to a misallocation of manufacturing sector resources.

Given Pakistan's current market conditions and massive fiscal deficit, adopting a market economy system and limiting the government's role in economic activities may be our best alternative.

CHAPTER 6

CONSLUSION AND POLICY RECOMMENDATION

6.1 CONSLUSION

Using the Automatic Model Selection techniques of PC-GeTS for covariates of TFP regressions on a set of previous studies with six models and 18 variables identifies an excessive number of significant regressors. The General-to-Simple process implemented in PC-GeTS, which has a strong performance in terms of its capacity to identify significant regressors. While omitted variables result in severely biased estimations, irrelevant regressors do not lead to bias. This is the core concept underlying the General to Simple (GeTS) technique. We began with the largest model feasible, which includes all possibly significant regressors. This is known as the Unrestricted General Model (GUM). Then, we eliminate irrelevant variables to arrive at a simple model. Our final persimmons model is superior because, it includes all those regressors that were not combinedly present in any of the six models that were used in the earlier research.

The primary goal of this research was to determine the true Covariates of Total factor productivity for Pakistan from the six existing models of TFP of empirical literatures by using encompassing and LASSO techniques. We employed the general to specific (G2S) methodology to identify the parsimonious model in both techniques., these are $TFPM_{\textbf{Encompassing}}$ and $TFPM_{\textbf{LASSO}}$. Then we chose the final model for Total factor productivity by ranking both parsimonious models according to their standard errors. It has been observed that $TFPM_{\textbf{Encompassing}}$ has the minimum standard error and has encompassed the $TFPM_{\textbf{LAsso}}$. Therefore, In Pakistan, $TFPM_{\textbf{Encompassing}}$ is regarded as the final model for TFP. The ARDL bound test was used to identify long-run cointegrating relationship among TFP determinants. The results of $TFPM_{\textbf{Encompassing}}$ suggest that the main covariates of TFP are expenditure on education, Foreign direct investment, unemployment rate, consumer price index, imported machinery, residential patent, and nonresidential patent of Pakistan. In the long run the expenditure on education, Inflation, unemployment rate and residential patent of Pakistan have negative and significant, whereas the imported machineries and nonresidential patent of Pakistan have positive and significant impact on TFP. While in the short, imported machineries, unemployment, foreign direct investment, residential and nonresidential patents have positive and significant effects on TFP of Pakistan. While Inflation have negative and significant effect on total factor productivity both short and long run.

6.2 Policy Recommendation

Based on results following policy is recommend for improving total factor productivity in Pakistan.

An increase in the cost of inputs causes production to decline. Which ultimately leads to low revenue generation and a reduction in exports. The high cost of production is the primary obstacle in the path of industrialization. If we want to grow the manufacturing, agriculture, and services sector we need to lower the cost of inputs i-e labor and capital. Labor and capital are complementary to produce output. To increase the output of labor we need to adopt machinery (capital) which is highly effective and efficient in producing goods both in good quantity and high quality. Unfortunately, our engineering sector has not evolved to that level to make hi-tech machinery that could enhance production, that's why we need to rely on imported machinery. Given the historical issues of huge balance of payments (BOP) deficit, and regulatory authorities the policymakers and financial managers often use to squeeze imports by imposing taxes and duties, sometimes banning imports.

Moreover, the financial and time cost of machinery import is too high because of over-reliance on imports for revenue generation and inefficient mechanisms at the port and regulatory bodies. As a result, companies often opt to continue with outdated inefficient capital, which ultimately hits labor productivity and thus final product is produced at a higher cost. This makes firms uncompetitive, and they seek government protection for survival.

The evidence found in my study suggests that imported machinery has a positive impact on production. Thus, the Policymakers needs to make such policies which could increase the import of machinery. In order to increase the import of machinery they need to cut down the taxes and duties. They also need to revise the mechanism of port regulatory bodies to make the operations smooth and make lesser time to deliver that machinery to the producer. This will have positive spillover effects on the Economy as a whole.

- Therefore, it is recommended that the Government should need to improve the techniques of production by importing machinery and that in turn will lead to high productivity in every sector.

- The results also specify that long-time economic growth is remarkably dependent on the capability of nation to move up on the ladder of innovation to remain internationally competitive. This needs the allocation of suitable resources for patents activities to push key economic sectors in the country.

- Low and stable inflation provide a favorable environment for TFP. Therefore, the government of Pakistan must reduce high uncertainty inside the country. Then, in this way the foreign resident Pakistani national will be attracted more to their home country for investment and they feel their money secure in their home country and thus, our productivity will rise.

References

Abbas, Q., & Nasir, Z. M. (2001). Endogenous growth and human capital: A comparative study of Pakistan and Sri Lanka [with comments]. *The Pakistan Development Review*, 987-1007.

Abramovitz, M. (1989). *Thinking about growth: And other essays on economic growth and welfare*: Cambridge University Press.(ISBN 0-521-40774-5)

Adnan, Z., Chowdhury, M., & Mallik, G. (2020). Determinants of total factor productivity in Pakistan: a time series analysis using ARDL approach. *International Review of Applied Economics, 34*(6), 807-820.

Ajide, F. M. (2021). (A) Symmetric impacts of crime rate on total factor productivity: evidence from Nigeria. *International Journal of Social Economics*. ISSN: 0306-8293

Akinlo, A. E., & Adejumo, O. O. (2016). Determinants of total factor productivity growth in Nigeria, 1970–2009. *Global Business Review, 17*(2), 257-270.

Becker, G. S. (2009). *Human capital: A theoretical and empirical analysis, with special reference to education*: University of Chicago press. ISBN 0-226-04120-4

Danquah, M., Moral-Benito, E., & Ouattara, B. (2011). TFP growth and its determinants: nonparametrics and model averaging. (*Banco de Espana Working Paper No. 1104* ; pages 34)

Epprecht, C., Guegan, D., Veiga, Á., & Correa da Rosa, J. (2021). Variable selection and forecasting via automated methods for linear models: LASSO/adaLASSO and Autometrics. *Communications in Statistics-Simulation and Computation, 50*(1), 103-122.

Ferraro, A. M. G. Poverty in Latin America: theory and statistical application with lasso regression.

Fonti, V., & Belitser, E. (2017). Feature selection using lasso. *VU Amsterdam research paper in business analytics, 30*, 1-25.

Gehringer, A., Martínez-Zarzoso, I., & Nowak-Lehmann Danzinger, F. (2016). What are the drivers of total factor productivity in the European Union? *Economics of Innovation and New Technology, 25*(4), 406-434.

Grossman, G. M., & Helpman, E. (1993). *Innovation and growth in the global economy*: MIT press. ISBN 0-262-57097-1

Hendry, D. F., & Richard, J. F. (1987). Recent developments in the theory of encompassing. *LIDAM Discussion Papers CORE*, (1987022).

Iliyasu, A., Mohamed, Z. A., & Hashim, M. (2015). Productivity growth, technical change and efficiency change of the Malaysian cage fish farming: an application of Malmquist Productivity Index approach. *Aquaculture International, 23*(4), 1013-1024.

Isaksson, A., & Ng, T. H. (2006). *Determinants of productivity: Cross-country analysis and country case studies. Research and Statistics Branch*. Staff Working Paper 1/2006, United Nations Industrial Development Organization (UNIDO: Vienna)..

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112): Springer.

Jones, L. E., & Manuelli, R. (1990). A convex model of equilibrium growth: Theory and policy implications. *Journal of political economy, 98*(5, Part 1), 1008-1038.

Khan, S. U. K. (2005). Macro determinants of total factor productivity in Pakistan, 383-401

Kolasa, M., & Żółkiewski, Z. (2004). *Total factor productivity and its determinants in Poland-evidence from manufacturing industries: The role of ICT* (No. 64). TIGER Working Paper Series.

Koopmans, C. (1965). On the concept of optimal economic growth, in (Study Week on the) Econometric Approach to Development Planning, chap. 4. In: North-Holland Publishing

Co., Amsterdam.

Kraft, J., & Kraft, A. (1978). On the relationship between energy and GNP. *The Journal of Energy and Development*, 401-403.

Leamer, E. E., & Leamer, E. E. (1978). *Specification searches: Ad hoc inference with nonexperimental data* (Vol. 53): John Wiley & Sons Incorporated.

López-Cálix, J. R. et al. (2012, December). What Do We Know About Growth Patterns in Pakistan? (World Bank Policy Paper Series on Pakistan No. PK 05/12). Islamabad: The World Bank.

Lu, M., & Mizon, G. E. (1996). The encompassing principle and hypothesis testing. *Econometric Theory*, *12*(5), 845-858.

Lucas Jr, R. E. (1988). On the mechanics of economic development. *Journal of monetary economics, 22*(1), 3-42.

McConnell, C. R., & Brue, S. (2017). *Contemporary labor economics*: McGraw-Hill Education. http://repository.vnu.edu.vn/handle/VNU_123/90226

Mizon, G. E., & Richard, J.-F. (1986). The encompassing principle and its application to testing non-nested hypotheses. *Econometrica: Journal of the Econometric Society*, 657-678.

Muslehuddin. (2007). An Analysis of Pakistan's Growth Experience. (Asian Development Bank Pakistan Resident Mission Pakistan Poverty Assessment Update Background Paper Series). (volume 5 number 2)

Nadiri, M. I. (1993). Innovations and technological spillovers. *NBER working paper*(w4423).

Nazir, S. (2017). Encompassing Of Nested and Non-nested Models: Energy-Growth Models. (MPRA Paper No. 77487)

Pasha, H. A., A. G. Pasha, and Kalim Hyder (2002) The Slowing Down of the Growth of Total

Factor Productivity in Pakistan.(Social Policy Development Centre Research Report No. 44). Karachi: SPDC.

Pesaran, M. H., Shin, Y., & Smith, R. J. (2001). Bounds testing approaches to the analysis of level relationships. *Journal of applied econometrics*, *16*(3), 289-326.

Prescott, E. C. (1998). Lawrence R. Klein lecture 1997: Needed: A theory of total factor productivity. *International economic review*, 525-551.

Qazi, M., & Hyder, K. (2007). *Determinants of Total Factor Productivity in Pakistan*. University Library of Munich, Germany. (MPRA Paper No. 16253)

Rebelo, S. (1991). Long-run policy analysis and long-run growth. *Journal of political economy*, *99*(3), 500-521.

Romer, P. M. (1986). Increasing returns and long-run growth. *Journal of political economy, 94*(5), 1002-1037.

Romer, Paul (1990). Endogenous technical change. Journal of Political Economy, 98(5), S71–102.

Sabir, M., & Ahmed, Q. M. (2008). Economic reforms and total factor productivity growth in Pakistan: an empirical analysis. *Business Review*, *3*(1), 53-68.

Saleem, H., Shahzad, M., Khan, M. B., & Khilji, B. A. (2019). Innovation, total factor productivity and economic growth in Pakistan: a policy perspective. *Journal of Economic Structures*, *8*(1), 1-18.

Siddique, M. A., Khaleequzzaman, M., & Ur Rehman, A. (2016). Determinants of Islamic Banking Industry's Profitability in Pakistan for the Period 2004-2012. *Journal of Islamic Business and Management (JIBM), Riphah International University Islamabad, Pakistan, 6*(1), 41-61.

Siddique, O. (2020). *Total factor productivity and economic growth in Pakistan: A five decade* *overview* (No. 2020: 11). Pakistan Institute of Development Economics.

Siddique, O. (2022). The Determinants of Total Factor Productivity Growth in Pakistan: An Exploration. *PIDE Working Papers No. 2022:4*, 1-20. Retrieved from https://pide.org.pk/wp-content/uploads/wp-0215-the-determinants-of-total-factor-productivity-growth-in-pakistan-an-exploration.pdf

Solow, R. M. (1956). A contribution to the theory of economic growth. *The quarterly journal of economics, 70*(1), 65-94.

Solow, R. M. (1957). Technical change and the aggregate production function. *The review of Economics and Statistics*, 312-320.

Swan, T. W. (1956). Economic growth and capital accumulation. *Economic record, 32*(2), 334-361.

Tufail, M., & Ahmed, A. M. (2015). Measuring total factor productivity and finding the determinants of total factor productivity at sectoral level: A case study of Pakistan. *Industrial Engineering Letters, 5*(6), 38-53.

Wooldridge, J. M. (1990). An encompassing approach to conditional mean tests with applications to testing nonnested hypotheses. *Journal of Econometrics*, *45*(3), 331-350.

Zaman, A. (2017). Lessons in Econometric Methodology: The Axiom of Correct Specification. *International Econometric Review*, *9*(2), 50-68.