Investment Forecasting with Multivariate Linear Regression in the Construction Industry of Pakistan

Mehdi, Masood

24 October 2021

INVESTMENT FORECASTING WITH MULTIVARIATE LINEAR REGRESSION IN THE CONSTRUCTION INDUSTRY OF PAKISTAN

Masood Mehdi

Table of Contents

CHAPTER NO. 01
BACKGROUND OF THE SUBJECT AND STATEMENT OF THE PROBLEM

INTRODUCTION

This study is about the investment forecasting in the construction industry of Pakistan. The main task of this study is to identify the key variables, which are responsible for the changes in the investment scenario in this sector. A research done by Sir John Thomson, is taken as the plate form on with I m going to workout my thesis. He research on the investment forecasting with the multivariate linear regression. I used the same mathematical method, which he mentioned in his research. His work is also mentioned in the appendix.

According to the annual report of state bank and the economy survey of Pakistan following statistical information comes into knowledge.

The resurgent Pakistani economy has shown an impressive growth trajectory for the second year in a row (2002-2003 & 2003-2004) with real GDP growing by 6.4% during the current fiscal year against 5.1% last year. The higher growth is underpinned by a buoyant industrial sector, which grew by a record 13.1 percent, and a services sector, which grew by 5.2%. The agriculture sector has marginally under-paced real GDP growth by growing at 2.6% against 4.1% last year. The growth performance was quite impressive and mainly emanated from the industrial sector reflecting an enhanced productivity, as well as job creating capacity of the economy.

The growth performance of the overall manufacturing sector was spearheaded by unprecedented growth in the large scale-manufacturing sector, which grew by 17% against the target of 8.8% and last year's actual rate of 7.2%. The large scale-manufacturing sector, accounting for 68% of overall manufacturing, recorded an impressive and board based growth and helped overall manufacturing grow by 13.4%, against a target of 7.8% and the last year's growth of 6.9%. The small-scale manufacturing continued to grow at an estimated 7.5%rate in 2003-2004. The construction sector was grew by 7.9% against 3.1% last year and a yearly target of 5.4%.

The electricity and gas distribution sector registered a massive increase of 22.5% against a decline of 2.6% last year and a yearly target of 5.3%. the services sector grew by 5.2% against 5.3% last year. The wholesale and retail trade and transport, storage and communication sub-sectors grew by 8.0% and 3.9% respectively against 5.9% and 4.0% last year.

This encouraging information brings me the idea of forecasting the investment scenario in the construction industry of Pakistan.

STATEMENT OF PROBLEM

Investment forecasting with multivariate linear regression in the construction industry of Pakistan.

SIGNIFICANCE OF THE STUDY

This study is helpful to both of the sectors i.e. public and the private sector. It helps the public sector in a way that it provides information about the future investment growth in this sector, which eventually creates many job opportunities for the general public. By using this study, government can easily estimate the revenue from this sector.

As far as private sector is concerned, it will guide them about the future scenario of this sector, their core decisions are also highly interrelated with this study. i.e. Investment and supply.

DEFINITION

Regression	A technique use to find out the trend.
R square.	It shows the effect of independent variables to the dependent variable.
R	It shows the relation of independent variables to the dependent variable.
Standard error of estimate	it shows the difference between the actual and the estimated figures.
T	It represents the significance of the independent variables.
Variance	It shows the difference between the values.
Standard deviation	statistical measurement used to mark the accidental error.

CHAPTER NO. 02
RESEARCH METHOD AND PROCEDURE

RESEARCH DESIGN

I have made full efforts to go through the research process as it is required. I used all the techniques to gather the accurate and up-to-date information for the accuracy of the results.

RESEARCH METHOD

The nature of the topic puts on view that it can be studied in artificial environment. The topic was analyzed in natural environment, so the research is a fundamental research and the suitable design for the research is descriptive.

The purpose of the study is descriptive in nature as it is aimed at learning about the activities of the setups, and how do they differ from one another in their functions. The study is conducted in natural environment.

The population reviewed for the study comprised aviation authorities and agencies.

The elements involved in the study for the purpose of research and obtain information were mostly air force personnel who were retired and as well who are currently serving.

The method of conducting this research was based on primary as well as secondary data. The more emphasis was on secondary data. Every thing was available because researchers are doing extensive researches in this area. I went through newspapers , magazines and internet for literature review.

RESPONDENTS OF THE STUDY

Respondents of my study were mostly Pakistan armed forces personnel.

RESEARCH INSTRUMENT

For collecting information, I went through Internet and government publications.

SOURCES OF DATA

As almost all the data is gathered through the secondary means and there is no primary data because it is out of the reach of my scope.

TREATMENT OF THE DATA

The data that were gathered during secondary data collection has only been analyzed in verbal context that is qualitative and not in numerical context that is quantitative circumstance. Multivariate linear regression is used for calculating the estimated model for the forecasting. Other instruments related with the study are "R square" and significance test i.e. t test and f test.

CHAPTER NO. 03
REVIEW OF RELATED LITERATURE & STUDIES

LOCAL LITERATURE

For gathering the relevant data, government publications have been seen. These publications include "Annual Report of State Bank 2004" and "Year Book of Statistics 2004". Some pages of the related material are also attached in the appendix A.

FOREIGN LITERATURE

A research study done by Sir James Thomas is also reviewed for making this thesis. It is also attached in the appendix B.

GAPS TO BE BRIDGED BY THE STUDY

This study is helpful to both of the sectors i.e. public and the private sector. It helps the public sector in a way that it provides information about the future investment growth in this sector, which eventually creates many job opportunities for the general public. By using this study, government can easily estimate the revenue from this sector.

As far as private sector is concerned, it will guide them about the future scenario of this sector, their core decisions are also highly interrelated with this study. i.e. Investment and supply.

AREAS FOR FURTHER STUDIES

Same principal can be used to make different other forecasting models like growth in agriculture, manufacturing and other industries.

CHAPTER NO. 04
PRESENTATION ANALYSIS

PRESENTATION OF DATA

For making the forecasting model I have selected many in dependable variables. But after calculating their relevance with the growth in construction industry I have selected five important variables which have significant relevance with the growth of construction industry. These variables are describes with the help of graph below.

Bar charts are being used to give the idea about the trend of the variable.

KEY VARIABLES USE IN THE MODEL

YEARS	GROWTH IN INVESTMENT	GNP AT CONSTANT	NNP AT CONSTANT	GFCF AT CONSTANT	WHOLESALE RATE OF BUILDING MATERIAL	PREVIOUS GROWTH	YEARS INFLUENCE
1995	-22.00	21253	20722.00	2176.00	81.04	4.10	1.00
1996	17.00	21944	21395.00	2541.00	87.33	-22.00	2.00
1997	-40.60	22183	21628.00	2154.00	98.62	17.00	3.00
1998	-7.20	22462	21900.00	1998.00	98.62	-40.60	4.00
1999	-38.70	21356	20822.00	1227.00	99.62	-7.20	5.00
2000	-4.80	22456	21895.00	1200.00	97.15	-38.70	6.00
2001	17.40	22374	21815.00	1270.00	100.00	-4.80	7.00
2002	-19.00	23326	22743.00	1059.00	101.18	17.40	8.00

Above are the some key variables which have influence on the growth in investment of construction industry. They are written according to their times. I have selected variables just for 2001 because the new "year book" of statistics is not published till now so I have selected the data till 2001.

Figure 1
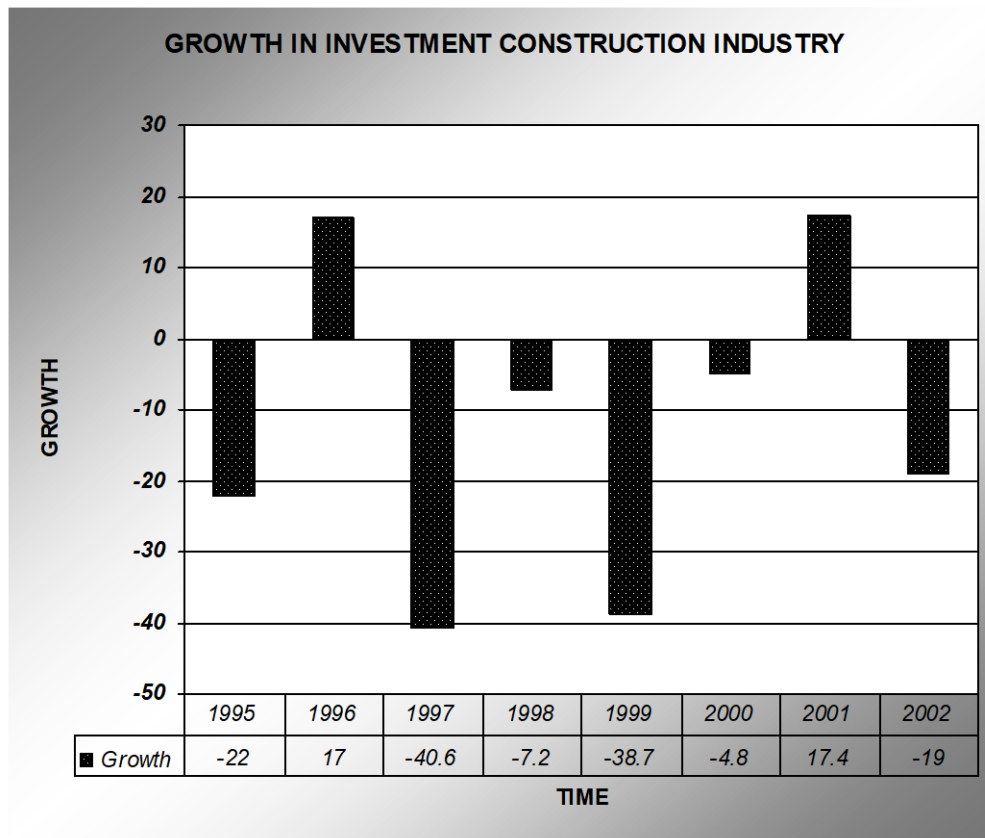
A high level of uncertainty can be seen by just have a view of the graph.

Figure 2

Gross National Product

TIME	1995	1996	1997	1998	1999	2000	2001	2002
Growth	21253	21944	22183	22462	21356	22456	22374	23326
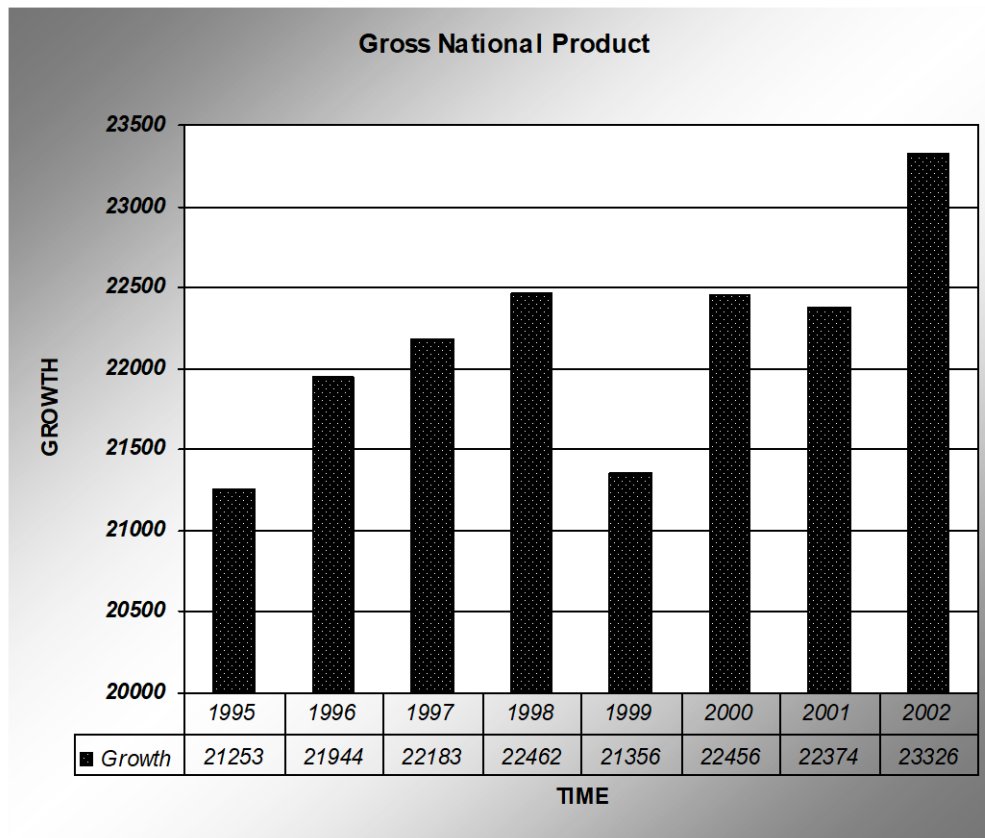
This graph shows upward trend in GNP growth.

Figure 3

Net National Product

	1995	1996	1997	1998	1999	2000	2001	2002
Growth	20722	21395	21628	21900	20822	21895	21815	22743

This graph shows upward trend in NNP growth

Figure 4

Gross Fixed Capital Formation

	1995	1996	1997	1998	1999	2000	2001	2002
■ Growth	2176	2541	2154	1998	1227	1200	1270	1059

TIME
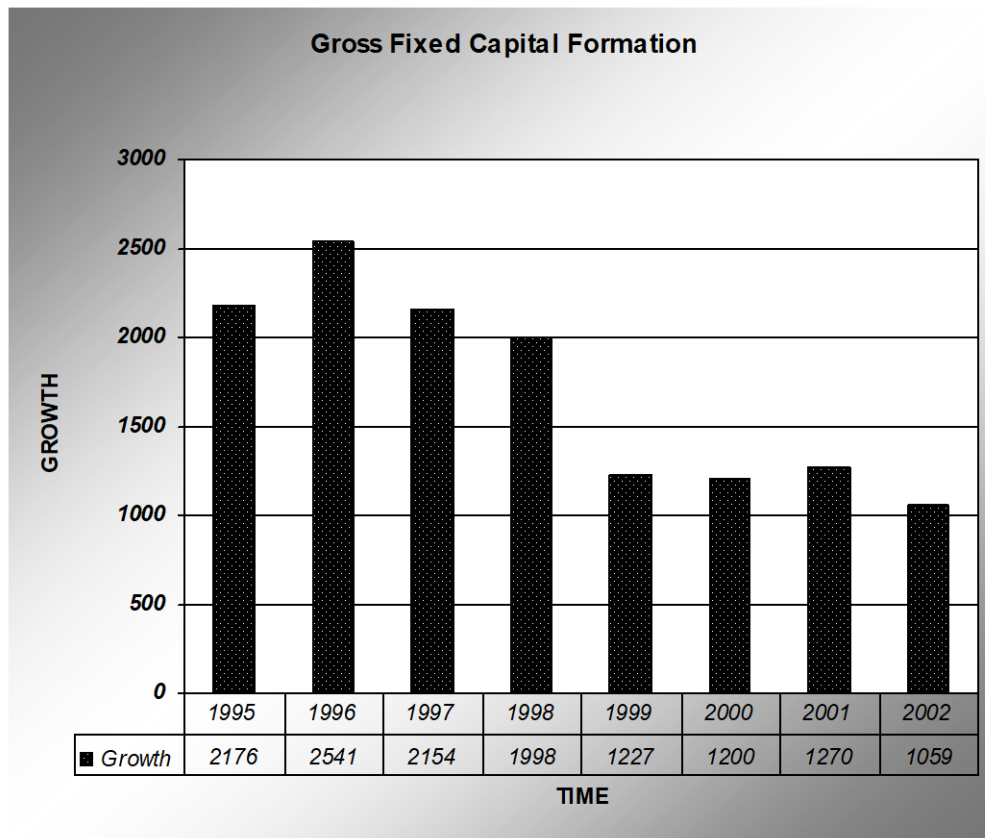
This graph shows declining growth in GFCF.

Figure 5

Whole Sale Rates of Building Material

	1995	1996	1997	1998	1999	2000	2001	2002
■ Growth	81.04	87.33	98.62	98.62	99.62	97.15	100	101.18
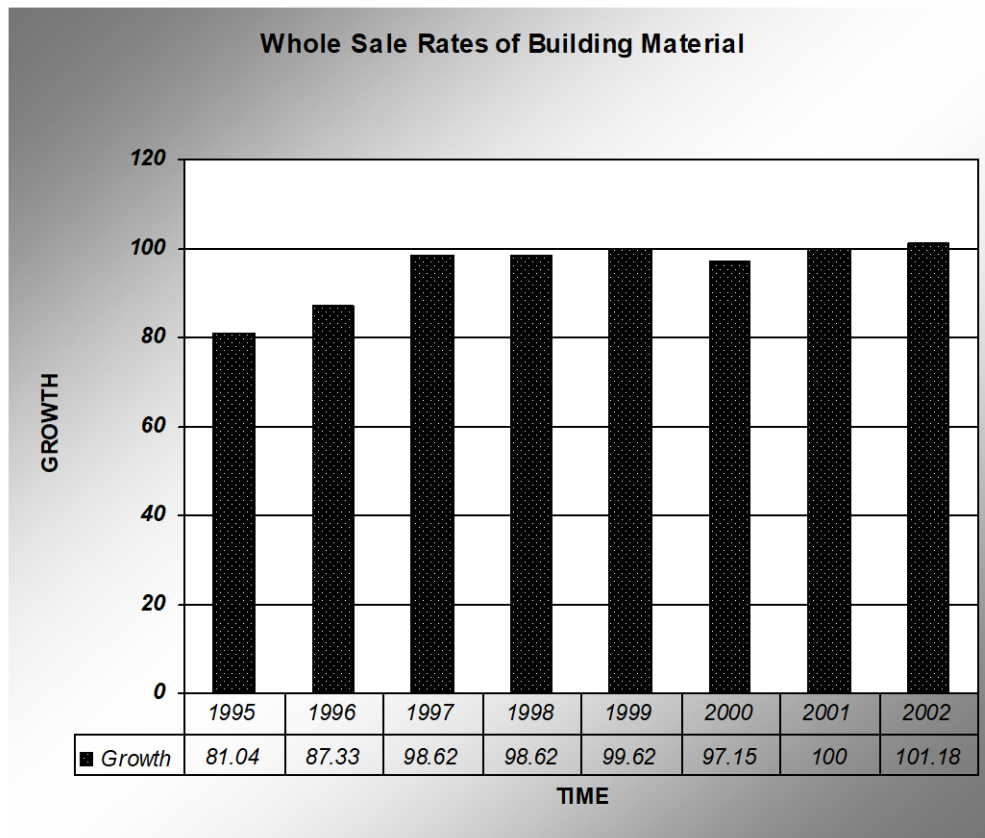
This graph shows upward trend.

DEPENDANT VARIABLE

Growth in investment in construction industry

INDEPENDENT VARIABLES

$X0$ = NNP at constant
$X1$ = GNP at constant
$X2$ = GCFC at constant
$X3$ = Wholesale
$X4$ = Growth of previous year investment
$X5$ = Year's influence

Independent Variables

Year	Wholesale $(x3)$	GNP at constant $(x1)$	NNP at constant $(x0)$	Year's influence $(x5)$	GFCf at constant $(x2)$	Growth of previous year investment "gic" $(x4)$
1995	81.04	21253	20722.00	1	2176.00	4.10
1996	87.33	21944	21395.00	2	2541.00	-22.00
1997	98.62	22183	21628.00	3	2154.00	17.00
1998	98.62	22462	21900.00	4	1998.00	-40.60
1999	99.62	21356	20822.00	5	1227.00	-7.20
2000	97.15	22456	21895.00	6	1200.00	-38.70
2001	100.00	22374	21815.00	7	1270.00	-4.80

Above are the some key variables which have influence on the growth in investment of construction industry. They are written according to their times. I have selected variables just for 2001 because the new "year book" of statistics is not published till now so I have selected the data till 2001.

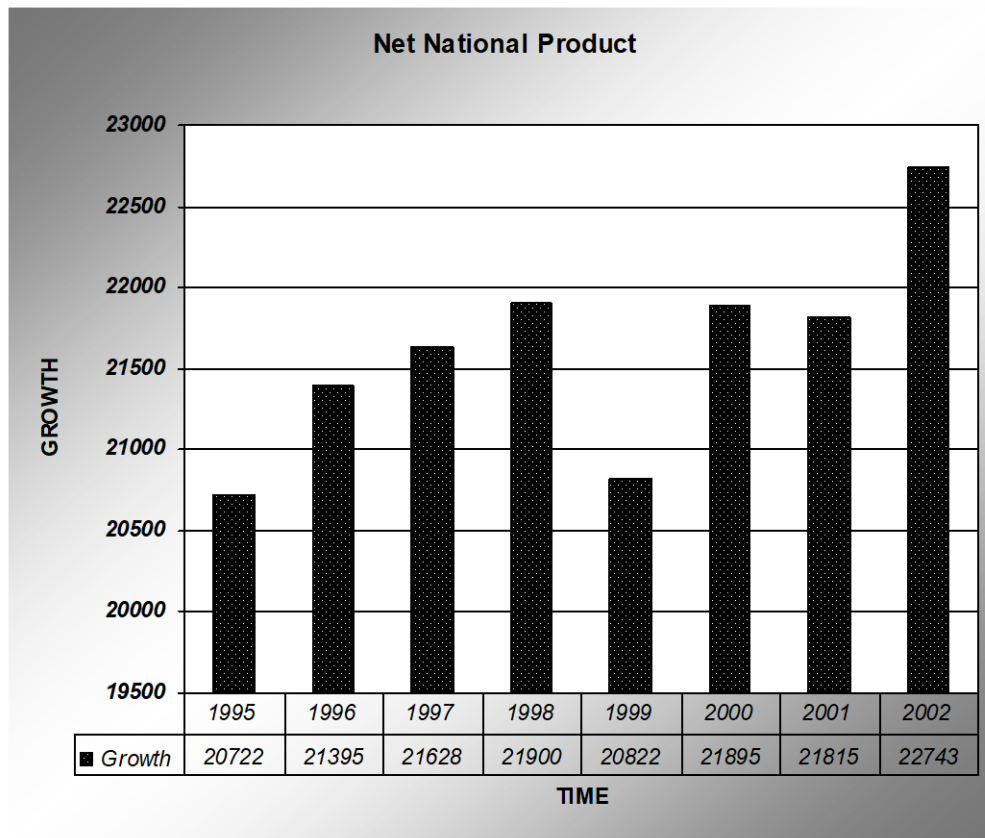## Dependant variable

Years	Next year's Growth in investment
1995	17.00
1996	-40.60
1997	-7.20
1998	-38.70
1999	-4.80
2000	17.40
2001	-19.00

For the purpose of forecasting I have planned to have in dependable variables of the previous year and the forecasted growth in construction industry of the next year. In this way we are forecasting for the next year in using the data of current year.

CHAPTER NO. 05
SUMMARY OF FINDINGS, CONCLUSION &
RECOMMENDATION

FINDINGS

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.989[a]	.978	.869	8.5559

a. Predictors: (Constant), years, gci, gnp at constant, whole sale, gfcf at constant

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-669.038	273.143		-2.449	.247
	gnp at constant	5.324E-02	.015	1.153	3.483	.178
	gfcf at constant	-.131	.024	-3.050	-5.524	.114
	whole sale	-1.609	.874	-.505	-1.841	.317
	gic"p"	.521	.198	.476	2.632	.231
	year's influence	-29.908	7.448	-2.734	-4.016	.155

a. Dependent Variable: growth in investment

Here R means the degree of association between the variable x(s) and y. Here the R is 98.9% which shows that there is high similarity between the combine trend of the in dependable variables and the dependable variable which can be easily notified by the diagram on page 26 figure 6. The other result is R Square which shows the effect of in dependable variable to the dependable variable. It is 97.8% which means that in dependable variables effect the dependable variable up to 97.8%. Adjusted R Square is 86.9% which means that with the less degree of freedom , we can say 86.9% effect on the dependable variable which is growth in investment of construction industry is due to the identified in dependable variables. Standard error of estimation shows that the forecasted value can be different up to 8.6 units.

In the coefficient chart we have different coefficients for the respective variable. Next to B column we have std. error column which means that their may be variations up to the given limit. The next column is the standardized coefficient which is used in the normal bell like curve. Then the next column is representing the t vales which are large and in front of them is the significance level of the test.

One variable is being excluded from the model because it has very low significance level which is approx equal to 0.

Excluded Variables[b]

Model		Beta In	t	Sig.	Partial Correlation	Collinearity Statistics Tolerance
1	nnp at constant	-2483.356[a]	.	.	-1.000	3.543E-09

a. Predictors in the Model: (Constant), year's influence, gic"p", gnp at constant, whole sale, gfcf at constant

b. Dependent Variable: growth in investment

PARTIAL CORRELATION

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.350ᵃ	.123	-.053	24.2485

a. Predictors: (Constant), gnp at constant

R Square shows that 12.3% variation is due to the GNP in the dependable variable.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.408ᵃ	.166	.000	23.6351

a. Predictors: (Constant), gfcf at constant

R Square shows that 12.3% variation is due to the GNP in the dependable variable.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.190ᵃ	.036	-.157	25.4145

a. Predictors: (Constant), whole sale

R Square shows that 3.6% variation is due to the whole sale rates of the building material in the dependable variable.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.282ᵃ	.080	-.104	24.8335

a. Predictors: (Constant), gic"p"

R Square shows that 8% variation is due to the growth in construction in the previous year on dependable variable.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.034[a]	.001	-.199	25.8722

a. Predictors: (Constant), year's influence

R Square shows that 0.1% variation is due to the year influence on the dependable variable.

FORECASTING MODEL

$$Y = -669.038 + 0.05324x_1 - 0.131x_2 - 1.609x_3 + 0.521x_4 - 29.908x_5$$

INTERPRETATION

1. If all other variables are zero than there will be a negative growth of 669.083 units in investment in construction industry.

2. With one unit increase in GNP, there will be increase in growth by .0532 units.

3. With one unit increase in GFCF, there will be decrease in growth by 0.0131 units.

4. With one unit increase in wholesale rate of building material, there will be decrease in growth by 0.1609 units.

5. With one unit increase in growth in investment of previous year, there will be increase in growth by 0.521 units.

6. With one unit increase in year, there will be decrease in growth by 29.908 units.

FORECASTED AND ACTUAL COMPARISON

Years	Actual	Forecasted	Difference
1996	17	19.25046	-2.25
1997	-40.6	-45.40241	4.8
1998	-7.2	-9.73566	2.54
1999	-38.7	-34.3633	-4.34
2000	-4.8	-6.36134	1.56
2001	17.4	13.39439	4.01
2002	-19	-16.97304	-2.03

N	Variance	Std. Deviation
7	12.67072	3.559595

Here I have made the comparison between the forecasted value and the actual value. I just put the values of the in dependable variables in the model and the relevant values of the dependable variable which is growth in investment of the construction industry of Pakistan. They both are much closed to each other with the little standard deviation of 3.5%. This thing can also be seen by having a sight on page 28 and figure 6.

Figure 6

Comparision
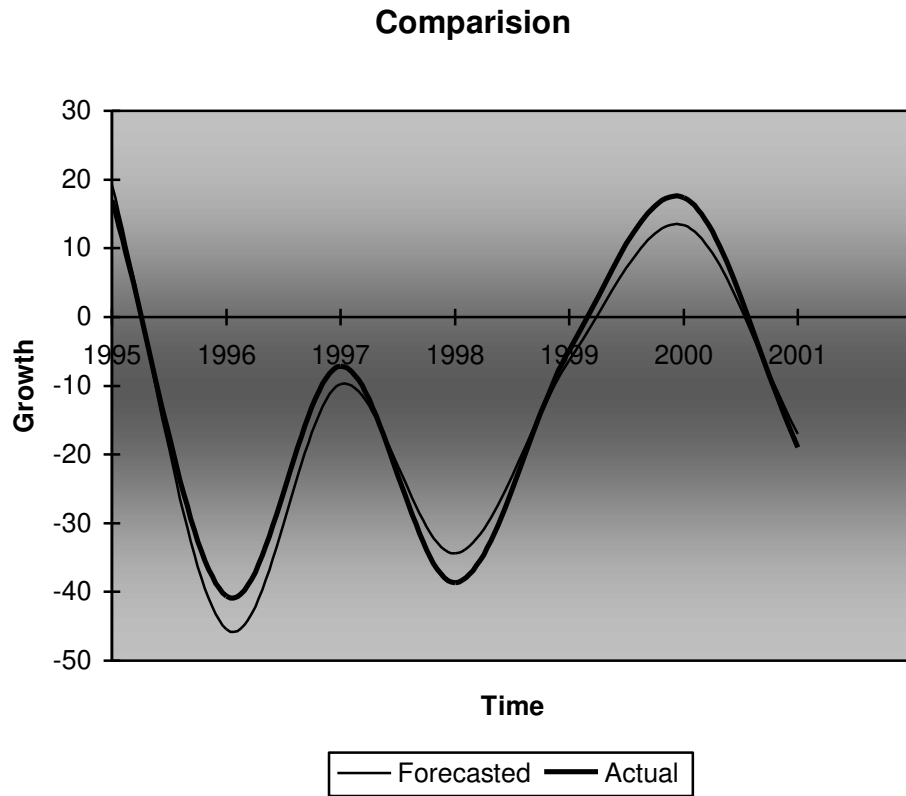
This graph is the core of my research. It shows that the forecasted values are almost moving in the same directions as the actual one with the standard deviation of 3.5%.

FORECASTING FOR COMING YEARS

For forecasting the growth in investment of construction, I find out the extrapolated values of the components on which the growth model is based. By running the regression, I got the following results. I take two independent variables. One is time and the other is the previous value of the component.

GNP

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	28984.505	11413.858		2.539	.064
	VAR00002	-.351	.533	-.300	-.657	.547
	time	206.077	126.344	.745	1.631	.178

a. Dependent Variable: VAR00001

GNP = 28984.505 - 0.351(V2) + 206.077 (time)

With every one unit increase in V2 there will be decrease in GNP by 0.351 and every one unit increase in the time, there will be increase in GNP by 206.007 units.

Whole sale

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	55.748	25.253		2.208	.092
	VAR00002	.422	.310	.672	1.363	.245
	time	.361	1.063	.167	.340	.751

a. Dependent Variable: VAR00001

Whole sale = 55.748 + 0.4220(V2) + .361(time)

With every one unit increase in V2 there will be increase in whole sale rates by 0.4220and every one unit increase in the time, there will be increase in whole sale rates by 0.36units.

GFCF

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	3169.216	1268.115		2.499	.067
	VAR00002	-9.62E-02	.424	-.091	-.227	.831
	time	-272.176	108.106	-1.008	-2.518	.066

a. Dependent Variable: VAR00001

GFCF = 3169.216 - 0.0962(V2) - 272.176(time)

With every one unit increase in V2 there will be decrease in GFCF by 0.0962 and every one unit increase in the time, there will be decrease in GFCF by 272.176 units.

Following are the results of the extrapolation of the components.

| YEAR	GNP	Whole sale	GFCF
2003	22651.77	101.695	617.7562
2004	23094.5	102.2733	388.0279
2005	23145.18	102.8783	137.9517
2006	23333.47	103.4947	-110.167
2007	23473.46	104.1157	-358.474
2008	23630.4	104.7388	-606.763
2009	23781.39	105.3628	-855.053
2010	23934.47	105.9871	-1103.34

I put these results in the main equation of forecasting and compute the growth. The growth of the previous year is used for GCI (p). Place the time value which is on initial is 9 and calculate the result. Following are the results of the forecasting of my model.

FORECASTING OF GROWTH

Year	growth
2003	14.36009
2004	53.5256
2005	78.5074
2006	103.1513
2007	125.0646
2008	146.4522
2009	167.2481
2010	187.8462

This chart is being plotted on the next page which shows that the investment will grow with a high rate.

Figure 7
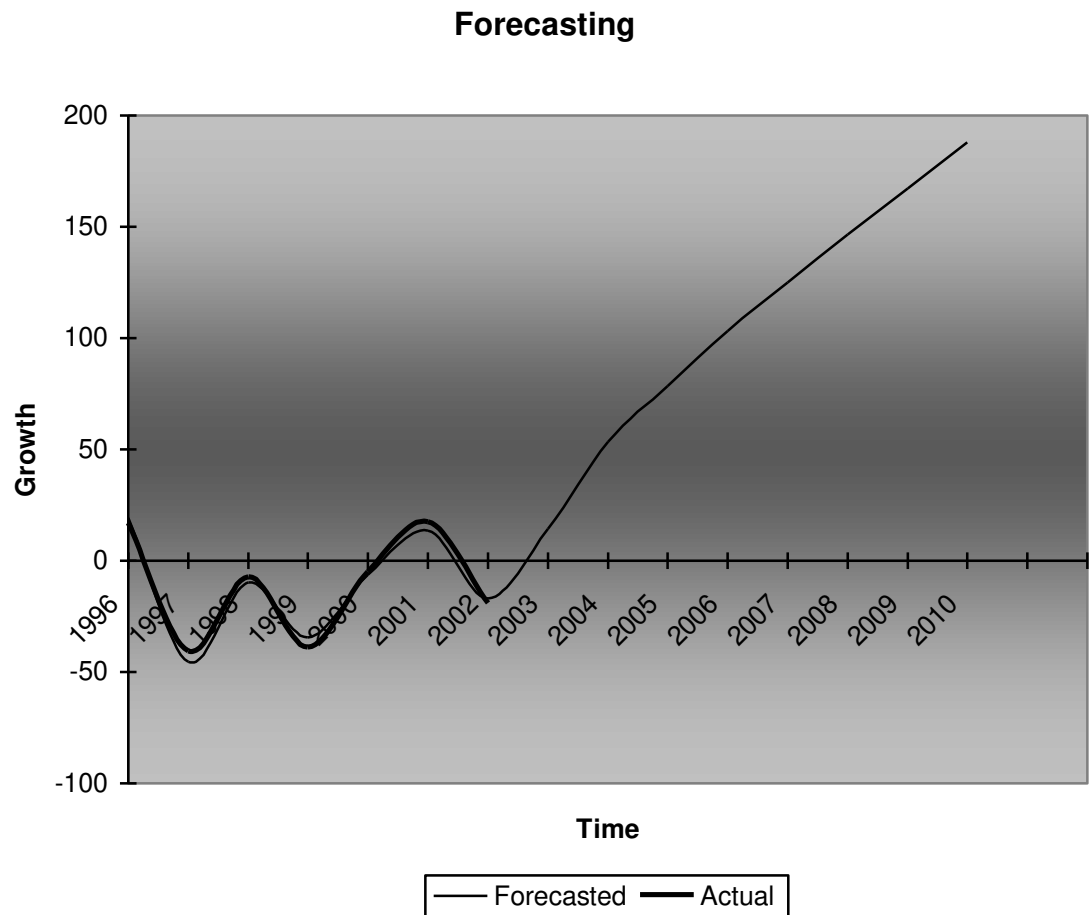
Forecasting

This graph shows there will be a boom in the construction industry of Pakistan because the forecasted line moves toward the heights which are in the favor of our country. Unlike the past it will be boom in construction industry. It is considered as the back bone of any country so I forecasted that we have a better country in future.

RECOMMENDATIONS

If government works to make the policies consistent or even improve, the growth in investment of the construction industry of Pakistan will be magnificent and this will bring change to the every corner of the country.