ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ΣΧΟΛΗ ΟΙΚΟΝΟΜΙΚΩΝ ΕΠΙΣΤΗΜΩΝ

ΤΜΗΜΑ ΟΙΚΟΝΟΜΙΚΗΣ ΕΠΙΣΤΗΜΗΣ

MSc in Applied Economics and Finance

Διπλωματική Εργασία

Forecasting economic series using data reduction methods

ΜΑΡΓΚΑΣ ΚΩΝΣΤΑΝΤΙΝΟΣ

ΑΘΗΝΑ 2020

Επιβλέπων καθηγητής : κ. ΔΕΝΔΡΑΜΗΣ ΙΩΑΝΝΗΣ, Οικονομικό Πανεπιστήμιο Αθηνών

Εξεταστής καθηγητής : κ. ΔΙΟΙΚΗΤΟΠΟΥΛΟΣ ΕΥΑΓΓΕΛΟΣ, Οικονομικό Πανεπιστήμιο Αθηνών

Εξεταστής καθηγητής: κ. ΤΖΑΒΑΛΗΣ ΗΛΙΑΣ, Οικονομικό Πανεπιστήμιο Αθηνών

Contents

1. Introduction

Historically, macroeconomic methods used for forecasting have been focused on a small number of predictors. Both univariate autoregressions and vector autoregressions have been widely use in macroeconomic forecasting, and today they serve as standard benchmarks used to evaluate economic forecasts. Recently, a high variety of explanatory variables are available, since the data pool is enriched, and often highly collinearly related. While the majority of the above data has a limited history causing the number of predictors to exceeds the number of available observations, their massive growth has led to the development of multiple data shrinkage techniques. Thus, dynamic factor models have become the most popular approach on economic forecasting.

The structural forecasting on the field of econometry is based on explicit theory. As a result, it is closed related to its theory but often appears lagging. Structural Keynesian forecasting in the macroeconomics' field is based ot postulated systems of decision rules which has risen between 1950 – 1960 after having a strong background on the advances in Keynesian theory back in 1930 – 1940. Although these two were declined in 1970-1980. Some conditional forecasts when they involve one or more conditional variables are often being used by a bunch of forecasting situations in which they maintain assumption which are related for example with the behavior of policy makers. The use of macroeconomic theory in structural econometrics shows that there are some lags regarding the developments in theory. 20th century's macroeconomics have been influenced by the Keynesian theory, which was flowered in the period 1930 – 1940. This wave was brought major advances regarding the structural microeconomic forecasting. When John M. Keynes published in 1936 the pioneering for that time General Theory, there wasn't even enough space for measurement back in '30s. Mesurements came to light by Klein's (1946) Keynesian Revolution as well as Klein and Goldberger's (1955) Econometric Model of the United States: 1929-2952 in the form of the systems of equations. The publication of the General Theory altered the following research and helped the development, estimation and analysis of Keynesian structural econometric models. Moreover, since statistics have been influenced by the advances of major statisticians such as Fisher, Neyman and Pearson, the branch of econometrics was equally developing.

The economic side, was highly formed by Keynes' important contribution which had a great positive result into offering workable solutions for several economics problems of the period. The growth of the Economic Society and its journal, Econometrica brought into light the relation of statistics and economics theory. This relation was purified in the work of the Cowles Commision for Research in Economics at the University of Chicago in the 1940s and early 1950s. The academician focus and depth of the talent combined, were extraordinary for the standards of the economics. This pneumatic phenom was formed by Cowles researchers included T.W. Anderson, K. Arrow, G. Debreu, T. Haavelmo, L. Hurwicz, L.R. Klein, T. Koopmans, H. Markowitz, J. Marshak, F. Modigliani, H. Simon, A. Wald, and others. Cowles research program was mainly based on the recognition and evaluation of differentiation of the stochastic systems in order to simulate the posulated decision, unless of Keynesian macroeconomic theory. Both the combination of statistics and economics related to the Cowles commission, and the confidence of solving pressing macroeconomic problems have been historically known. In the macroeconomics sector, gradually performed large-scale Keynesian macroeconomic forecasting models led to the bloom of the structural econometric forecasting in the late 1950s and 1960s. Even if there was conflict even if there was disagreement on details such as the relative slopes of ISLM curves, a great unanimity was arisen about the general paradigm, resulting the extentinitely application of the forcasting models and policy analysis in both academic and governmental level. In the late 1960s and early 1970s, academic annoyance with the underpinnings of Keynesian macroeconomic systems of equations led to breaches in the base. Absence of a background for the disequilibrium nature of the Keynesian model was initially disturbing among the economists. Microfoundations for Keynesian macroeconomic theory specifically for the core assumptions of sticky wages and prices sought because of multiple research programms. Phelps et al. (1970) reffered to many key early additions, and more recent additions are gathered in Mankiw and Romer (1991).

Second, just as macroeconomists turned rapidly disappointed with the ad hoc treatment of sticky prices in traditional models, they turned in the same way, disappointed with ad hoc treatment of expectations. Accordin to the early Muth's works (1960, 1961), who was the first one ever used the idea of rational expectations, schemes such as adaptive expectations were rational only in unprobable situations, when the "rational expectations revolution" quickly took hold based on Sargent and Wallace's (1975) simple but important early paper. Third, generally speaking economists started feeling annoyed not only due to ertain parts of the Keynesian macro-econometric program, such as the assumptions about price behavior and expectations formation, but also due to the way the model works in general. Prescott's approach (1896) was titled "system-of-equations", due to the fact that it was focused into estimating the parameters of equation systems representing ad hoc hypothesized decision rules ("consumption functions," "investment functions," etc.) as antithetic to more crucial parameters of interests and technology.

Technological advances and other trends back in late 1960s and early 1970s, led to the late shown-up macroeconomic work. Lucas and Prescott (1971) and Lucas (1972) are remarkable examples. After the publication of Lucas' (1976) formal appraisal of the system-of-equations approach, a work on the tastes-and-technology boosted briskly, positioned on the perspective that decision rules are an essentially defective model for producing conditional forecasts. This approach was based on the ristrictions of decision rules that will finally be changing along with policies. Although breaches in the Keynesian base forecacting started as academic annoyance, they were broaden by the economic facts of 1970's. More specifically the contemporaneous presence of high inflation and unemployment, naturally led economists to question the so-called inflation/unemployment trade off embedded in the Keynesian systems of equations. Finally, based on a series of published studies, simple statistical estimations creating no expectations in any factor of economic structure, and often the macroeconomic activity can be forecasted just as well as large-scale Keynesian macroeconomic models; Nelson (1972) still being a landmark. Keynesian macroeconomics shortly after, got declined, and Keynesian structural econometric forecasting followed in the footsteps of it as well.

2. Literature Reviews

2.1. Macroeconomic Forecasting

2.1.1. Definitions

Macroeconomic forecasting in someform has a longhistory. The ancient Egyptians foretold harvests (a large part of what we would call their GDP) from the level reached by the Nile in the floodseason. Economic activity refers to the macroeconomic situation, especially in relation to its current positive or negative trend, but the expression is also used for a particularly good macroeconomic situation (high economic situation), i.e. when most of the companies have a large turnover. In economic terms, economic fluctuations in the degree of utilization of the production potential of an economy are understood to be cyclical. Furthermore, more or less regular fluctuations of economic variables such as production, employment, interest rates and prices may occur, with the result that cyclical fluctuations in macroeconomic activity may arise. This can be measured by the degree of capacity utilization. The main indicator of this is real gross domestic product.

More precisely, the term "economic cycle" is used for the in demand and production fluctuations, which can lead to changes in the management of production capacities. The term is used contrary to the development of capacities in the sense of economic growth itself, and if they have a certain regularity (in delimitation of one-off special items).

Economic theories examine and explain the wave-shaped changes in the level of economic activity within a market economy, which recurs in business cycles. Different methods are used to obtain from the observed data the regular pattern of a business cycle, in particular its upper and lower reversal point.

In contrast, growth theory balances the long-term trend of an economy's growth. In order to look at the economic cycle in and of itself, it is therefore necessary to abstract from the underlying growth trend. The economic theory analysis overlaps in part with employment theory.

Independently of each other, several economic researchers have identified fluctuations of different lengths. The Kitchin cycle is used to assess business production and production planning and warehousing. In the first phase, more is produced than sold to fill the warehouses. As soon as slower growth appears, production is curtailed in the second phase. For this cycle, a length of 3 to 4 years is empirically demonstrable. The Juglar cycle describes investment phases. It lasts between 6 and 10 years. The triggers for the Kondratyev cycle, which lasts from 40 to 50 years, are technological innovations.

The period attributed to a business cycle depends largely on whether the level of economic activity (usually measured by aggregate production, i.e. gross domestic product) is taken as a measure, or whether the level of economic activity is taken as a measure. Growth rates. If the beginning and end of a business cycle are delimited according to whether economic output has been in absolute decline ("classic business cycles"), longer cycles can be found. A classification based on growth rates leads to a greater number of shorter "growth cycles". Alternatively, economic cycles can be measured by the extent to which companies' production capacities are utilised. Here, fluctuations between about 70% (recession) and 100% (boom) are conceivable. Accordingly, the business definition of the economic cycle according to the conventional doctrine is: fluctuations in the degree of utilization of the production potential of an economy. It describes a macrocycle of several years to decades of duration as well as microcycles of a few years, but not under-year seasonal cycles.

2.1.2.Macroeconomic Variables

2.1.2.1. GDP

Gross domestic product (GDP) is onen of the primary indicators used to gauge the health of a country’s economy. It is a measure of aggregated output in the national in come accounts, and it is counted by the value of all currently produced financial goods and services over a given period, typically a year. Notice, that only final goods and services enter the definition of GDP. Intermediate goods and services are not included because they contribute to the value of the final goods. So, adding intermediate goods directly in GDP would lead to a double counting. However, two types of goods are used in the production process associated with intermediate goods, and are included in GDP. Firstly, currently produced capital goods are included in GDP, as only a depreciation of these goods contributes to the value of the final goods in each period. Hence, not including capital goods in GDP is equivalent to assuming full depreciation in the present period. Second, inventory investments are included, as inventory stock of final goods are currently produced output.

GDP is a proxy of the total value of all goods, i.e. goods and services produced during a year within the national borders of an economy as final products, after deduction of all intermediate consumption. Thus, all final goods, i.e. goods at the end-use level, are recorded as economic output. In the calculation, goods that are not reused directly but are put in stock are taken into account as a change in stocks. In contrast to gross national income, the calculation of GDP only covers domestic benefits and applies the so-called domestic principle; national borders are crucial. On the other hand, Gross National Income is based on the national principle. This also takes into account the services provided abroad by nationals; conversely, benefits provided by nationals are not taken into account. The residences of the persons are therefore decisive in this respect. If depreciation is deducted from GDP, the net domestic product (NIP) is the result. GDP is used as a proxy of an economy's economic performance over a period of time. The rate of change in real GDP serves as a measure of the economic growth of the economies and is thus the most important measure of national accounts (see list of countries by gross domestic product). Gross domestic product may refer to countries as well as to other administrative or geographical units. In some cases, the terms gross regional product, gross provincial product, gross world product and others are then used.

Nominal GDP is evaluated at market prices. So, this is making it sensitive to changes in the average price level. This sensitivity is caused by the re-evaluation of the market price whenever a shock in either supply or demand leads to a new equilibrium. To overcome this problem, we calculate real GDP which, in a traditional setting, is constructed as the sum of all produced final goods and services, multiplied by the constant prices from a base year. When there is more than one final good, the estimation of the real GDP is defined as a weighted average of the output of all final goods. The weights of each price correspond to their relative prices. There are two problems concerning real GDP. On the one hand, in case of base change, the weights and the GDP history alter. On the other hand, the price changes, and substitutions among products contained in GDP can be problem. Chain-weighted measure of real GDP could be used in order to avoid both these problems, where the constant base year is replaced by the average prices in the present- and previous years. This means, that the base moves forward each year, eliminating the problem caused by relative price-induced substitutions.

GDP comprises four main components: Consumption, investment, government purchase, and net exports. Consumption, which is the largest GDP component, is the house hold sector's demand for current used output, and consists of consumer durable goods, non durable consumption goods, and consumer services. The investment component covers the part of GDP, which is purchased by the business sector, plus residential construction. Investments include business fixed investments, residential construction investments, and inventory investments. The government purchases of goods and services component covers the share of the current output bought by the government sector. The government sector covers the federal government as well as state and local governments. The net export component covers the total export minus import, that is, the currently produced goods and services sold to foreign buyers.

With the data or otherwise the derivatives of a country's economy, it is given the opportunity to produce economic goods. The volume and quantity of various goods produced over a given period of time, which is usually one year, is a good economic measure of a country's standard of living. But the variety of products produced, from the simplest such as food and clothing to the more complex, such as synthetic fibers, computer accessories, etc. do not allow the comparison of the capacity of an economy from one period to another, unless before homogeneous goods converted into a common unit of measurement. It is not possible to add oranges and aspirins or computers and fabrics. Converting goods into measurable units is used to value them in monetary units.

Its total value is the product of its price by its quantity. The sum of the value of individual goods gives the total value for the economy and is called Gross Domestic Product (GDP).

Gross Domestic Product (GDP) includes the total value in monetary units of finished goods and services produced in a country in a given year. The reason why the term "Domestic" is used is very important because it indicates that production must take place within one country, regardless of whether the producer is resident in another country.

The mathematical expression of the Gross Domestic Product is:

$$GDP = C + I + G + NX$$

where: (C) consumption, (I) investment, (G) public expenditure on the purchase of goods and services and (NX) net exports minus imports.

Consumption refers to the expenditure incurred by households on the purchase of goods and services, ie their total consumption

Investment is reffered to an expenditure on the purchase of capital equipment, stocks and buildings, including expenditure on the purchase of new housing. Expenditure on the purchase of intangible goods, such as research and development expenses, may also be included in this category.

Goverment costs include expenditure on the purchase of goods and services carried out by the local government, state governments and the federal government, for example. purchase of a submarine for the navy.

Net exports are reffered to the expenditure on the purchase of goods and services produced in the domestic economy and purchased by foreigners (exports).

Imports are the cost of any goods or services where as a country of manufacture it is not the country in which GDP is calculated.

GDP Deflator or GDP Price Index: is an index that estimates the changes in all prices of both goods and services produced in an economy. Thus, GDP is the ratio of nominal GDP in the base year to nominal GDP in the year chosen as a percentage.

Gross Domestic Product is measured in three ways:

1. By the method of expenditure,

2. The rate of pay method, and

3. The value added method.

The flow of euros into a given economy is calculated by GDP and is calculated in 2 ways. GDP is the total income from bread production, which equals to the sum of wages and profits - the upper half of the euro's circular flow. GDP is also the total expense of buying bread - the bottom half of the euro circular flow. To calculate GDP, one can observe the flow of euros from businesses to households or the flow of euros from households to businesses. These two ways of calculating GDP must be equal because buyers' spending on product purchases is in line with accounting rules, income for product sellers. Therefore, every transaction that affects the expense must affect the income and every transaction that affects the income must also affect the expense.

In the inner spiral, there are labor and bread flows: households sell their work to businesses, and businesses sell households the bread they produce. The outer coil shows the corresponding flows of the euro: households pay and buy bread from businesses, and businesses pay salaries and profits to households. There are two ways that GDP can be calculated in such an economy. Either calculating the total cost of bread production or calculating the total income from bread production.

The information a country can collect for its economic prosperity through GDP are very important, as GDP per capita. is an important element that measures a country's standard of living.

GNP however, despite its importance and usefulness, it has many weaknesses and imperfections, and the most important causes of these weaknesses and imperfections are:

- At GDP does not include production value which relates to the same consumption as it is not marketed. For example, the food a housewife cooks at home is no different than the food of a restaurant. But it is not calculated in GDP. the added value, with home cooking.

- GNP not qualitative but quantitative. Quality improvement, when not expressed in price, is not counted in GDP. Quality, however, is just as important as quantity. It is worth mentioned that the quality of life, which is not included in the GNP calculations, derived, for example, from the clean atmospheric air, is an major factor affection health and life expectancy.

- Since GDP expresses the size of production but not the composition, ie. Kind of goods and services produced, it does not include the composition and distribution of production. But the prosperity of an economy is certainly affected if, for example, the percentage of production representing war goods changes at the expense of the relationship with consumer goods. Another factor that affects the economic well-being of citizens but is not included in GDP is the distribution of production (income) among members of an economy. However, this distribution, if it improves or worsens, positively or negatively affects the lives of citizens. The more equitable the distribution of GDP is. the higher the standard of living of a country is considered, because it narrows the gap between rich and poor.
- The goods and services of the underlying economy, which constitute a large volume of an economy, do not include GDP. The term 'anecdotal' refers to the part of the citizens’

hidden economic activity. The reason for hiding the economic activity lies either in taxes avoidance, or in illegal activities, such as smuggling, drugs, etc.

The above financial activities are not recorded in GDP. and it is very important to note that these deficiencies in GDP make it difficult to compare living standards across countries. The size of the economy, for example, varies from country to country, so comparisons are problematic.

2.1.2.2. Unemployment

Unemployment is an important variable in macroeconomics, because it tells somethingabout whether the economy uses its resources, in the formofmanpower, optimally. Moreover, investors and consumers become more reticent in their investment and spending patterns in cases when the unemployment rate is increased. This is due to their fear of economic recession. On the other hand, when the unemployement rate declines, investors and consumers are filled with more confidence, which is reflected in the economy as investment and spending patterns. The term "unemployment" is defined as the number of people who does not have a job and are looking for one. Hence, to be categorized as unemployed a person must meet two criteria, first he or she must be unemployed, and second he or she must be job seeking. To compare the unemployment numbers through different years, we construct the unemployment rate, which is the number of unemployed persons expressed as a percentage of the laborforce.

Unemployment concerns labour, the price of which is called wages and is formed on the labour market by labour supply and demand. In terms of market terms, unemployment is the surplus of labour supply over labour demand. Unemployment therefore presupposes labour markets and also a working society. These include workers who cannot make a living through their own means of production (land, real estate, technical means of production) and a market-based social formation. A mass of such people – the industrial proletariat – arose in early modernity with the peasant liberations, the population explosion and the industrial revolution. Consequentely the social question (Pauperism) of the 18th century has led to the adaptation of an employment policy after the first forms of state penal, educational and social systems. The wage-earners, for their part, organized themselves in the labour movement (trade unions, workers' parties, workers' associations, cooperatives, etc.) in order to cope better together with the problems associated with unemployment and external wage labour. Unemployment is one of the major economic policy's problems, because it must meet the objective of a high level of employment within the magic square. The existing volume of work must be distributed equally among the labour force potential by means of economic policy measures using the current rules on working time. This way the objectives will be achieved. Underemployment or overemployment are observed in the cases when the target is missed. The objective is regarded as achieved at full employment, even if a small level of unemployment in the context of full employment persists. Unemployment is only completely eliminated in the event of market clearance on the labour market.

2.1.2.3. Inflation

Inflation is measured either by annual changes in the price of goods and services of certain baskets of goods or by the GDP deflator, which reflects the price changes of all goods in an economy. Inflation is the object of knowledge of economics, especially macroeconomics. The consumer price index is most often used to measure inflation. The index is calculated using a shopping cart, which is set in a given year (base year) representative of an average household. Since July 2002, the Federal Statistical Office has been applying the so-called " Price Adjustment" to calculate inflation for some product groups. The aim is to take into account quality changes in price measurement. This method is used, for example, for IT products that are rapidly changing and cannot be observed in the same form over a long period of time. [6] The introduction of hedonic price adjustment led to significantly lower rates of change for the sub-indices concerned due to strong technical progress and comparatively stable prices. In addition to this purely statistical method, the cost of living index (COLI= cost of living index) has established itself in economics. It measures the expenditure that economic agents have to make in order to achieve a certain level of benefit. While inflation described in this way can be more accurately described as consumer price inflation, asset price inflation is called when the prices of assets such as equities and real estate rise.

Consumer Price Index is refered in the monthly weighted average of prices for a representative basket of consumer goods and services, containing two separate indexes in order to represent different groups or populations of consumers. Consumer Price Index for All Urban Consumers is the most frequently used index, while the Consumer Price Index for Urban Wage Earners and Clerical Workers is often used for wage escalation agreements. The Consumer Price Indices are often used to adjust payments for rents, wages, alimony, child support, and otherobligations that may be afected by changes in the cost of living.

Producer Price Index is a weighted index, that estimates the average change over time in the selling prices received by domestic producers goods and services. This Index is used as a proxy of the trends within the wholesale markets, manufacturing industries, and commodity markets. Thus, all of the physical goods-producing industries that make up the U.S. economy are included. It measures the price changes from the producer's perspective, in contrast with the Consumer Price Index.

Average price changes of non-military goods and services, imported to or exported from the U.S. are measured by Import and Export Prices Indexes. This index helps to measure the inflation rate in globally traded products.

Employment Cost Trends estimate the changes in laborcostsover time, and the average costs per hour and are only published quarterly.

The most commonly used index is the consumer price index, despite the fact that various price indices can be used to determine the inflation rate.

Inflation is generally regarded as a negative phenomenon due to its adverse effects. Some of the effects of inflation are as follows:

- Decreasing of purchasing power of fixed income. There are incomes that remain stable for long periods or are adjusted very slowly. The increase in prices has the direct effect of reducing the purchasing power of these incomes. Thus, in times of inflation real incomes are falling, along with the amount of money held as savings, ie their real value decreases.
- Worsening the balance of goods and services. Increasing the price level reduces the competitiveness of the economy in the international environment because it makes domestic goods more expensive. As a result, inflation reduces exports and at the same time increases imports.
- Redistribution from lenders to borrowers. It is easily understood that inflation reduces the real value of debt that is to be repaid in the future. Therefore, the borrower is favored by inflation because the real value (purchasing power) of the money that will be repaid will be less than what he received when he got the loan. Accordingly, the lender loses because the real value of the loan repaid will be less than the original value of the loan. Of course, as long as inflation persists, the inflation rate will be incorporated into the interest rate to avoid the transfer of money due to inflation. Thus, the interest rate will rise by the rate of inflation. For example, if with an inflation rate equal to zero the interest rate is 5%, and inflation rises to 3% the interest rate will become 8%. In other words: nominal interest rate = real interest rate + inflation rate.
- Costs of dealing with inflation As mentioned above, inflation reduces the purchasing power of cash. Households and businesses that need cash for their transactions have an incentive to reduce cash to the minimum, and this requires effort and spending on resources, such as the employment of individuals in businesses to oversee the status of cash. Businesses are also often required to replace their pricelists and promotional material. There are still difficulties in business arrangements for future pricing and in general difficulties in coordinating cash expenditure and inflows, e.g. how will a product order be billed next year? It is clear that tackling inflation creates costs for businesses. The intensity of the above consequences depends, of course, on the level of inflation. When the inflation rate is low, e.g. 2-3% of the effects are mild and the functioning of the economy is not disturbed. Inflation at low rates is often characterized as herpes inflation. In contrast to creeping inflation, in many countries and at times,

galloping inflation with rates fluctuating above 25% or 30% annually. Gallup inflation has a strong redistributive effect on fixed income and lenders, and generally creates serious problems in the functioning of the economy. There are also periods of extremely high inflation rates, such as during wartime, which is characterized as hyperinflation. In such cases, the consequences are so detrimental to the financial system that economists (households and businesses) avoid holding cash and preferring, as long as they can exchange products instead of trading for money. In essence, money is abandoned and replaced with products such as gold and silver.

Combating inflation is a difficult tast, due to both the aggregate supply's and demand's interaction, and the expectations of economists, households and businesses, as well as the government. Financial situation and activity is usually affected for a long time by the decisions of individuals and the state. For example, the signing of collective agreements specifies employees' pay for the next one or two years. This has implications for both the consumption of individuals and the cost of producing businesses. Also, the government's decision to carry out a major project (eg a highway) has committed funds for several years. For these decisions, the expected rate of inflation is very important because it affects the agreements that each side will seek. If workers believe that next year's inflation rate will be 3%, they will seek an agreement with employers that will provide for a nominal wage increase of at least 3%. Also, if businesses believe that inflation will be 3% next year, they will close deals on future deliveries at prices increased by 3%. In addition, the government needs to increase the budget line against expected inflation so that it can cover future spending.

2.1.2.4. Other Variables

Economic value added refers to the financial burden of the value of a commodity as a raw material at each stage of its processing to its final form.

In order to get the good in its final form, that is, when it has been bought by the consumer, it has gone through various processes and each of them has left its mark on the form of its final form. Thus value added is expressed as a currency unit surcharge.

We can therefore say that this is always a difference of newer value with an earlier one in the manufacturing line, from the contribution of labor, capital and entrepreneurship. Thus the salaries given in this case, both to labor (wages - salaries) and to capital (interest, rents, etc.), on their contribution to the production of a business, constitute the added value.

The most important element of this concept is that if the sales of all the business activities of a society are grouped together, the transactions between them, which are automatically called "intermediaries", are eliminated. This is due to the value of the final products, such as their sales (of goods or services), and to their investment that equals the added value of the same market.

In the wider economy, the added value is essentially the value of the gross national or domestic product, that is, that of the domestic product.

The production of goods, for the most part, in addition to labor and natural resources, also requires buildings, tools and machinery, in other words capital, which is the means of acquiring capital equipment.

However, with the use of capital equipment, the wear and tear is reduced and should be replaced at some point. Correct and timely replacement of capital equipment results in the right results, as it avoids shortages of machinery and tools, which will result in less commodity production and consequently significant business problems.

The same is established for the economy in general, because if businesses fail to replace worn-out capital equipment, reducing the country's total capital equipment and its total production capacity will be severely restricted by adverse effects on the country's economic life.The depreciation of capital equipment, which is due to its use, measured in monetary units, is called depreciation.

The way companies should choose to be able to replace worn out equipment in a timely manner is by saving a sum of money from their total revenue. With this money they can, when needed, buy machinery and tools that will replace the old.

That is, some of the machinery and tools purchased each year are intended to replace those machinery and tools that have completed their productive lives. These machines do not increase the productive capacity of the country, they just keep it at the level it used to be.

As stated above, the National Product refers to the value of all finished goods produced in one year. Machinery, tools, buildings, etc. are also considered as final goods. that is to say, capital equipment. If, when measuring the National Product, we also include the value of the capital equipment that removes the damage, if we also measure the depreciation, then we have the Gross National Product (GDP). If we do not calculate these goods' value, the National Product, then we have the Net National Product.

2.1.3. Dynamic Factor Models

Dynamic and static models must be distinguished in the cases of real world analysis's with macro data. The well known text book consumption function, i.e., the relationship between private consumption expenditure (C) and households' disposableincome (Y) is an example of a static equation

$$C_t = f(INC_t),\ f' > 0.$$

Consumption in any period t is strictly increasing in income, hence the positive signed first order derivative f0 –the marginal propensitytoconsume. To be able to apply the theorytoobservationsof the real economy we have to specify the function f(INCt).

For simplicity, we use the same symbolsfor the coefficients in the twoequations but it is importanttonote that the slopecoefficient β1 has a different economic interpretation in the two cases. In (2.2), β1 it is the marginal propensity to consume (MPC forshort), and is assumed to be a constant parameter. In the log linear model (2.2) β1 represents the elasticity of consumption in period t with respect to income. β1 can be used as a proxy in the estimate of the percentage increase in Ct following a 1% increase in INC. Thus, the log-linear specification in (2.2) can imply that the marginal propensity to consume can be considered a function of income. In that context, the log-linear model is less restrictive among the two, and that specification is used as an example.

Macroeconomic textbooks usually omit the term et in equation, but for applications of the theory to actual data it is a necessary to get an intuitive grip on this disturbance term in the static consumption function. Supposing real data corresponding to Ct and Yt, and we assume the intercept β0 is quantified in a really good way and the marginal propensity to consume β1. Least-squares estimation is a way of finding the numbers for β0 and β1 can measure the average best predictionof Ct for a given value of Yt.

According to Geweke (1977) and Sargent and Sims (1977), the seminal idea of dynamic factormodels was introduced. The model I, thay presents here, is the basic dynamic factor model by Sargent and Sims (1977). The main idea of this model is that the observation t of a dataset can be modeled as the sum of a number of common factors, the lags of these common components and an idiosyncratic component. Summary of this model can be found below:

$$y_t = \Lambda_0 f_t + \Lambda_1 f_{t-1} + \ldots + \Lambda_m f_{t-m} + e_t$$

where Λ0, ...,Λm are N x r matrices and ft is a vector of r factors.

e_t is the vector of idiosyncratic components, which are to be independent stationary processes. Thus, these components are uncorrelated to both leads and lags of the common factors and to the other idiosyncratic components. The estimation of the loading matrices, of the factors and of the rest of the parameters of the model, can be performed by Kalman filter, which is a particular maximum likelihood technique.

In the last few years a series of papers had as their main objective to develop new versions of the dynamic factor model, presented in the previous section. These new models try to relax some of the restrictions of the basic model, while trying to formulate new estimation methods, that are more time-efficient and that can allow the use of larger datasets. In the next subsections I will present the models from the previously cited authors. These kind of models are also addressed as "secondgeneration" models. Finally, a brief description of model by Kapetanios and Marcellino (2006) will be done. This model unifies the state-space representation, typical of "first generation" models, together with the ability to handle large dataset.

The main idea of this model is tocombine the approximate factor model and the dynamic factor model. The approximate factor model is an extension of the static model, which allows heteroskedasticity and weak serial and cross-correlationof the idiosyncratic terms. Publication's authors use the estimated factors to create out-of-sample forecasts. The first step includes the defining of yt+1, the series to be forecast, and Xt, the N-dimensional series of predictor variables. These series are observed for time t = 1, ..., T. Mean zero values are assumed for y and X variables. The model with r ∗common dynamic factors ft is shown in the next equations,

$$y_{t+1} = \beta(L)f_t + \gamma(L)y_t + \epsilon_{t+1}$$

$$X_{it} = \lambda_i(L)f_t + e_{it},$$

fori = 1, ..., N, where et = [e1t , ...eN t] 0 is a N x 1 idiosyncratic term and λi(L), β(L) and γ(L) are lag polynomials in non-negative powersof L. It is assumed that E (t+1|ft,yt , Xt , ft−1, yt−1, Xt − 1, ...) = 0.

The assumption of finite lags for the factors average values that the true number of factors underlying the dataset is finite. Thus, they can be gathered in a vector. Thanks to this static representation (which is a notational artifact the allows us to write the model in terms of static factors), it is possibleto use principal component (PC) estimation. Of course, it is important to remember that this model would be inconsistent with infinite distributed lags of the commonfactors. Under a set of asymptotic rank conditionson Λ and moments conditions, the model allows for serial correlation of error.

2.1.4. Shrinkage Methods

An alternative to the dynamic factor model approach is to use shrinkage estimators in ourregression. Unlike dynamic factor models, the shrinkage estimators does not reduce the dimensionality of the date by extracting commonfactors. Instead they focuson estimating a regression model over a constrained parameter space. The basic idea of the shrinkage methods is to use them for linear regression in order to modify the least squares estimate by imposing a penalty on their size. A linear regression models the relationship between the dependent variable and the predictors. The linear regression model has the form

$$y_i = \beta_0 + \sum_{j=1}^{p} x_{ij}\beta_j,$$

and can be solved by ordinary least squares, which finds the coecients = (0, 1,..., p) that minimize the residual sum of squares,

$$RSS(\boldsymbol{\beta}) = \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2$$

The solution to the ordinary least squares is found by differentisl Equation with regard to, and leads to a closed-form expression for the estimated value of the unknown parameter

$$\widehat{\boldsymbol{\beta}}^{OLS} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$$

The ordinary least squares estimation performs rather well in cases when the number of observations, n, is much higher than the number of predictors, p. However, since we focus on forecasting with several predictors, we have p>n. This causes a problem for the ordinary least squares estimation because, when p>n some of the predictors might be colinear, which means that X doesnot have full rank. If X is not of full rank, then X0 X is singular and there is no longer a unique least squares estimate. Furthermore, when the number of predictors are large, the least squares estimates will often have low bias but high variance. Therefore, we introduce some methods which shrinks the coecient estimates towards zeroor setting some equal to zero. These methods introduce some bias but reduce the variance of the predicted values, and thus may improve the overall prediction accuracy measured in terms of the mean-squared error. There exist several shrinkage methods, the ones that we implement are the ridge regression, lasso, grouped lasso, and elastic net method.

2.1.5. Vector Autoregressive Models

As stated above, we have dealt with methods that are designed to handle large number of predictors. These methods are usually combined with a linear regressions in which the relationship between the dependent variable and the estimated factors, dynamic factor models, or the constrained predictors, shrinkage methods, are modeled. In this chapter we use these method in the context of the vector auto regressive (VAR) models, in order to compute models that not only regress the dependent variable at the predictors, but also includes the variable’s own lagged values. VAR models, first introduced as a methodfor estimating economic relationships by [Sims, 1980], is widely used for structural analysis and simultaneous forecasting of a number of temporally observed variable. Each variable has an equation explaining its progression based on its own lags and the other model's variables' lags, in VAR models. So, if we measure three different time series, yt,1, yt,2, yt,3, and want to explain their progression over one period, we have the following VAR model.

$$y_{t,1} = v + \phi_{11} y_{t-1,1} + \phi_{12} y_{t-1,2} + \phi_{13} y_{t-1,3} + \varepsilon_{t,1}$$

$$y_{t,2} = v + \phi_{21} y_{t-1,1} + \phi_{22} y_{t-1,2} + \phi_{23} y_{t-1,3} + \varepsilon_{t,2}$$

$$y_{t,3} = v + \phi_{31} y_{t-1,1} + \phi_{32} y_{t-1,2} + \phi_{33} y_{t-1,3} + \varepsilon_{t,3}.$$

The conventional VAR models have quadratically growing parameter spaces, e.g. in our case we have 80 time series of 31 observations, thus we want to compute a VAR (30) model for the 80th time series, which requires estimating 900 parameters. However, such large number of stationary observations are not available in practice which means that the conventional VAR models users from a dimensionality problem. To over come this problem, introduced the Bayesian VAR approach, in which the problem with dimensionality is solved by shrinking the variables by imposingpriors. Another method that overcomes the dimensionality problem is the factor-augmented VAR model, where the variable that we would like to forecast together with estimated factors of the predictors, see Section 4.1.1, are arguments in a conventional VAR model.

The factor-augmented VAR (FAVAR) context allows us to include much of the information stored in the 80 macroeconomic variables in the model, by including the dependent variable and the estimated factors, which describes much of the variance in the remaining 79 predictors. Let yt be a k→1 vectorofobservable variables, and Ft be a r→1 vector of unobservable factors. Assume that the joint dynamics of (Ft, Yt) are given by

$$\begin{bmatrix} \boldsymbol{F}_t \\ \boldsymbol{y}_t \end{bmatrix} = \Phi(L)$$

$$\begin{bmatrix} \boldsymbol{F}_{t-1} \\ \boldsymbol{y}_{t-1} \end{bmatrix} + \boldsymbol{v}_t$$

where (L) is a lag polynomial of finite order p, and vt is the error term with zero mean. Equation (6.1) is the FAVAR model, in which the factors can be estimated by PCA as described in Section4.1.1. Note, that the dynamic factor models described in Chapter 4 and the FAVAR models are estimates in the same way. The dynamic factor models are preferred when the corrpurpose is forecasting, whereas the FAVAR models are preferred when conducting a structural analysis of the variables. The dynamic factor model is used when forecasting because it only produces the model for the desired forecast variable. The FAVAR model is used to conducting a structural analysis, because it produces a grid of models, one for each incorporated variable.

2.1.6. Bayesian Model Averaging

Using Bayesian inference to this purpose has been suggested as a framework capable of achieving these goals. Bayesian Model Averaging (BMA) is an extension of the usual Bayesian inference methods in which one doesnotonly models parameter uncertainty through the priordistribution, but also model uncertainty obtaining posterior parameter and model posteriors using Bayes' theorem and therefore allowing for allow for direct model selection, combined estimation and prediction.

Let each model in consideration be denoted by Ml , l = 1, . . . , K representing a set of probability distributions income passing the likelihood function L(Y |θl , Ml) of the observed data Y in terms ofmodel specific parameters θl and a set of prior probability densities for said parameters, denoted in general terms by π(θl |Ml) on which we omit eventual prior hyperparameters for the sake of clarity. Notice that both the likelihood and priors are conditionalon a particular model. Given a model, one then obtains the posterior distribution using Bayes' theorem, resulting in

$$\pi(\theta_l|\boldsymbol{Y}, M_l) = \frac{L(\boldsymbol{Y}|\theta_l, M_l)\pi(\theta_l|M_l)}{\int L(\boldsymbol{Y}|\theta_l, M_l)\pi(\theta_l|M_l)d\theta_l}$$

where the integral in the denominator is calculated over the support set for each prior distribution and represents the marginal distribution of the dataset over all parameter values specified in model Ml . This quantity is essential for BMA applications as we will show momentarily and is called the model's marginal likelihoodor model evidence and is denoted by

$$\pi(\boldsymbol{Y}|M_l) = \int L(\boldsymbol{Y}|\theta_l, M_l)\pi(\theta_l|M_l)d\theta_l$$

Bayesian model averaging then adds a layer to this hierarchical modeling present in the Bayesian inference by assuming a prior distribution over the set of all considered models describing the prior uncertainty over each model's capability to accurately describe the data. If there is a probability mass function over all the models with values π(Ml) for l = 1, . . . , K, then Bayes' theorem can be used to derive posterior model probabilities given the observed data by

$$\pi(M_l|\boldsymbol{Y}) = \frac{\pi(\boldsymbol{Y}|M_l)\pi(M_l)}{\sum_{m=1}^{K}\pi(\boldsymbol{Y}|M_m)\pi(M_m)},$$

resulting in a straight forward posterior model probability, representing the backing of each considered model by the observed data. There is also a link between these posterior model probabilities and the use of Bayes Factors. Given two models l and m, the Bayes factor of model l against model m is given by

$$BF_{lm} = \frac{\pi(M_l|\boldsymbol{Y})}{\pi(M_m|\boldsymbol{Y})},$$

it is clear that equation can be written in terms of Bayes Factors by simply dividing by the baseline model's evidence, resulting in

$$\pi(\Delta|\boldsymbol{Y}) = \sum_{l=1}^{K} \pi(\Delta|\boldsymbol{Y}, M_l)\pi(M_l|\boldsymbol{Y}),$$

which is an average of all posterior distributions weighted by each posterior model probability. Therefore, BMA allows for a direct combination of models to obtain combined parameter estimates or predictions (Roberts; 1965). This practice leads to predictions with lower risk under a logarithmic scoring rule (Madigan and Raftery; 1994) than using a single model. However, the implementation and applicationof BMA is not without difficulties. A prior distribution over the considered models must be specified, which is non trivial in most applications. Additionally, calculating each model evidence (equation 2) is non-trivial. Except in simple settings like in some generalized linear models with conjugate distributions, the evidence does not present a closed form and must be approximated, which presents plenty of challenges and is an active research field (Friel and Wyse; 2012). Despite these difficulties, BMA was extensively applied in the last 20 years, mostly in combining multiple modelsfor predictive purposes and selecting models, particularly covariate sets in regression models or network structure in Bayesian Network models. The latter application induces another pitfall in the formof large model spaces. For instance, consider a regression model with p covariates. The number of possible models without any interaction coefficients is 2p, which represents a large number of models even formoderate values of p. This difficulty can be mostly addressed by prior filtering of all possible models or through stochastic search algorithms over the model space.

2.2. Dimension Reduction

2.2.1. Importance of High Dimensional Data

In traditional statistical data analysis, we think ofobservations of instances of particular phenomena (e.g. instance ↔ human being). Being a vector of values, these observations are measured on several variables (e.g. blood pressure, weight, height, ...). In traditional statistical methodology, we assumed many observations and a few, well chosen variables. The trend today is towards more observations but even more so, to radically larger numbers of variables – voracious, automatic, systematic collection of hyper-informative detail about each observed instance. We are seeing examples where the observations gathered on individual instances are curves, or spectra, or images, or even movies, so that a single observation has dimensions in the thousands or billions, while there are only tens or hundreds of instances available for study. Classical methods are simply not designed tocope with this kind of explosive growth of dimensionality of the observation vector. We can say with complete confidence that in the coming century, high-dimensional data analysis will be a very significant activity, and completely new methods of high-dimensional data analysis will be developed; we just don'tknow what they are yet. Mathematicians are ideally prepared for appreciating the abstract issues involved in finding patterns in such high-dimensional data. Twoof the most influential principles in the coming century will be principles originally discovered and cultivated by mathematicians: the blessings of dimensionality and the curse of dimensionality. The curse ofdimensionality is a phrase used by several subfields in the mathematical sciences; I use it here to refer to the apparent intractability of systematically searching through a high-dimensional space, the apparent intractability of accurately approximating a general high-dimensional function, the apparent intractability of integrating a high-dimensional function.

The blessings of dimensionality are less widely noted, but they include the concentration of measure phenomenon (so-called in the geometryof Banach spaces), which means that certain random fluctuations are very well controlled in high dimensions and the success ofasymptotic methods, used widely in mathematical statistics and statistical physics, which suggest that statements about very high-dimensional settings may be made where moderate dimensions would be toocomplicated. There is a large bodyof interesting workgoingon in the mathematical sciences, both to attack the curse of dimensionality in specific ways, and to extend the benefits of dimensionality.

2.2.2. Dimension Reduction Definitions

Dimensionality reduction is a machine learning (ML) or statistical technique of reducing the amoun to frandom variables in a problem by obtaining a set of principal variables. This process can be carried out using a number of methods that simplify the modeling of complex problems, eliminate redundancy and reduce the possibility of the model overfitting and thereby including results that do not belong.

The process of dimensionality reduction is divided into two components, feature selection and feature extraction. In feature selection, smaller subsets of features are chosen from a set of many dimensional data to represent the model by filtering, wrapping or embedding. Feature extraction reduces the number of dimensions in a dataset in order to model variables and perform component analysis.

2.2.3. Methods

Methods of dimensionality reduction include:

Factor Analysis

Low Variance Filter

High Correlation Filter

Backward Feature Elimination

Forward Feature Selection

Principal Component Analysis (PCA)

Linear Discriminant Analysis

Methods Based onProjections

t-Distributed StochasticNeighbor Embedding (t-SNE)

UMAP

Independent Component Analysis

Missing Value Ratio

RandomForest

Dimensionality reduction is advantageous to AI developers or data professionals working with massive datasets, performing data visualization and analyzing complex data. It aids in the processof data compression, allowing the data to take up less storage space as well as reduces computation times.

2.2.4. Dimensional Reduction for Regression and Classification

Most of the general dimension reduction methods belong to the unsupervised learning category because no label information is used. The other two traditional machine learning categories are supervised learning and semi-supervised learning, which use all or a part of the label information. In most real applications, dimension reduction is just an intermediate step toward the final goals, like classification or regression. Separating the dimension reduction and model learning may not be optimal for classification or regression. For example, in the task of document classification, feature selectionor feature extraction methods are used first to get a low-dimensional text representation, and then, a classifier is trained to make a prediction. Lacking supervision, some important words may be filtered before training the classifier, which affects the final performance. To tackle this problem, supervised dimension reduction methods have emerged and attracted growing attention. Based on the underlying techniques adopted, we categorize the supervised dimension reduction methods into three classes: PCA-based, NMF-based, and manifold-based dimension reduction methods. Among them, mostof PCA-based and NMF-based methods are linear methods, while most of manifold-based methods are non-linear methods. By analyzing the means of exploiting the label information, we find that there are two main ways: LDA and directly integrating the loss function for classification or regression. LDA minimizes the distance within class and maximizes the distance between classes. To integrate the lossfunction directly for classification or regression, the commonly-used loss functions (e.g., L2 loss, L1 loss, and hinge loss) are mainly adopted in logistic regression, SupportVector Machine (SVM), linear regression, polynomial regression, etc. We will elaborateon them in the subsequent sections.

In the past few decades, dimension reduction had been extensively explored, and several reviews on dimension reduction already exist. However, different from those that mainly reviewed existing unsupervised dimension reduction methods, our review focuses on the supervised dimension reduction. To the best of our knowledge, this is the first review to target this direction. We provide a taxonomyto systematically categorize the methods and list important open problems to guide the further developmentof this topic. Due to the greater popularity of feature extraction compared with feature selection, in our paper, we mainly focus on feature extraction for supervised learning. With regard to feature selection for supervised learning, we refer the reader to.

2.2.5. Applications and Possible Outcomes

From the inception of NMF, it had been successfully applied to face recognition due to its ability to produce interpretable bases. Naturally, Face recognition becomes the typically successful application of supervised NMF. Discriminative NMFs are the earlier successful attempts of supervised NMF methods at face recognition, and then, many direct NMF methods also demonstrated superior performance in this task. Apart from face recognition, all this object or action recognitionals involves the application of supervised dimension reduction. Wu et al. proposed a supervised Laplacian eigenmap torecognize visual objects. Kumar adopted supervised dictionary learning torecognize the actions and locationsof the objects in the images. Santiago-Mozos et al. applied supervised PCA to object detection in tinfrared images and demonstrated good performance. Recently, Xinfang et al. proposed a semi-supervised local discriminant analysis by combing the idea of LDA and LLE for polarimetric SAR image classification.

picture of the data is quite helpful; thus, visualization is very important, and it is also an important application of supervised dimension reduction. Barshan et al. provided a supervised PCA to conduct visualization, while Vlachos et al. gave another supervised dimension reduction method by borrowing the LDA idea for visualization. Geng et al. proposed a supervised Is omap to visualize. Compared with visualization from general unsupervised dimension reduction, visualization from supervised dimension reduction has clear separability due to its supervised learning property. Apart from all the above applications, text mining is probably another good application of supervised dimension reduction. Although there are already many works on unsupervised dimension reduction, there are few workson supervised dimension reduction.

Speech recognition is another successful application of NMF, and thus, supervised NMF is naturally successfully used in this kind of application. Lee et al. used discriminative NMF to classify the emotional difference in speech. Bisot et al. applied supervised NMF to acoustic scene classification and obtained rather good performance. Sprechmann et al. and Weninger et al. solved the audio source separation with supervised NMF, while Nakajima et al. and Kitamura et al. adopted supervised NMF for music signal separation. Although there exist an amount of successful applications in speech recognition, more attempts can be made in the future. As we can see that almost all of the existing supervised dimension reduction methods are NMF-based, both PCA-based and manifold-based methods can be investigated and compared with the existing methods.

2.3. Partial Least Square

The goal of partial least squares is to predict Y from X and to describe the common structure underlying the two variables (Abdi, 2003). Partial least squares is a regression method allows for the identification of underlying factors, which are a linear combination of the explanatory variables or X (alsoknown as latent variables) which best model the response or Y variables (Talbot, 1997). Although similar to principal components analysis (PCA) regression and canonical analysis and alternating least squares, it is considered to be a better alternative to multiple linear regression and PCA regression methods since it provides for more robust model parameters that donot change with new calibration samples from the population (Falk & Miller, 1992, Geladi&Kowalski, 1986). Furthermore, partial least squares is an improvement on PCA since the solution derived from partial least squares is constrained to the part of the covariance matrix that is directly related to the experimental manipulation or that relates to behavior (McIntosh, Chau, &Protzner, 2004). The term partial least squares specifically means the computation of the optimal least squares fit to part of a correlation or covariance matrix (McIntosh, Chau, &Protzner, 2004, Wold, 1982). The part of the correlation or covariance matrix that the least squares are fit to is the “cross-block” correlation between the exogenousor X variables and the dependent measures or Y variables. Partial least squares measures covariation between two or more block of variables and creates a new set of variables that is optimized for maximum covariance (not maximal correlation) using the fewest dimensions (McIntosh, Bookstein, Haxby, & Grady, 1996). Partial least squares is sometimes called soft modeling because while OLS regression makes hard assumptions such as no multicollinearity in the independent variable, soft modeling refers to softening of these assumptions. Partial least squares is a linear technique. Partial least squares is preferred as predictive technique and not as an interpretive technique except for exploratory analysis before using interpretive techniques such as multiple linear regression or SEM. Partial least squares optimal linear relationships are computed between latent variables and can be interpreted as the best set of predictions available for a study given all the limitations (Falk & Miller, 1992). Soft modeling is a way of estimating the likelihood of an event given information about other events.

As an extension of multiple linear regression, partial least squares regression has many of the same assumptions. For example, ones hould be concerned with outliers and nonlinear data relationships when using partial least squares. Because the distribution of partial least squares is unknown, there is no conventional significance test. However, significance can be tested through bootstrap methods such as jack knife which is a resampling method. The problem with using a resampling method to determine significance is that although there are no specific sample size requirements, the smaller the sample, the more likely that fitted confidence limits will be fitted to noise in data instead of true distribution.

The following are the key advantages of partial least squares:

Able to model multiple dependent as well as multiple independence variables

Can handle multicollinearity in IVs

Robust despite data noise and missing data

Creates independent latents directly on the basis of cross products involving response variable(s) = stronger predictions

Allowsfor reflective and formative latents

Applied to small sample

Distributional free •

Handle range of variables: nominal, ordinal, continuous

Disadvantages of Partial Least Squares:

Difficulty in interpreting loadings of independent latent variables (based on cross product relations with response variables not, as in conventional factor analysis, on correlations among manifest independents)

Distributional properties of estimates not known

Can't get significance unless run boots trap

Lack of model test statistics

Partial Least Square Structural Equation Modelling (PLS-SEM) is the most SEM technique used in IS research. PLS is regarded as the most fully developed and general system (JörgHenseler, Hubona, & Ash, 2016). IS was identified as the primary user of PLS (Evermann& Tate, 2014). Rönkkö et al. (2012) argue that the use of partial least squares path modelling as a tool fortheory testing has been increasing in the late 90's and PLS is currently one of the most common quantitative data analysis methods in the top IS journals. However, they emphasise that reliance on PLS method has possibly resulted in producing and publishing a large number of studies, whose results are invalid. These critics have been addressed by the literature (J. Henseler et al., 2014). The technique has been subject to many reviews (Evermann& Tate, 2012; JörgHenseler et al., 2016; Rouse&Corbitt, 2008; Urbach&Ahlemann, 2010). That has resulted in the production of guidelines for the use of PLS-SEM in IS research. Mostof these guidelines focuson either explanatory (confirmatory) orexploratory research. For instance, Henseler et al.

(2016) propose an updated guideline for the use of PLS in IS research in confirmatory settings. On the other hand, Urbach&Ahlemann (2010) come up with a guideline for the utilisation of the technique in exploratory contexts. The literature provides three purposes of any research: exploratory, descriptive or explanatory (confirmatory). An exploratory study is a valuable means of finding out what is going on; to look for new insights; to ask questions and to evaluate phenomena in a new light (Saunders, Lewis, &Thornhill, 2009). Exploratory research goes with a predictive model (Evermann& Tate, 2014).The object of descriptive research is to portray an accurate profile of persons, events or situations (Saunders et al., 2009). Studies that establish causal relationships between variables may be termed explanatory research (Saunders et al., 2009). Explanatory research goes with the causal model (confirmatory) model.

Nevertheless, Evermann& Tate (2014) argue that the causal and predictive modelling are dualities. Rather, there is a middle-ground between the two extreme positions. It is easier fordecision makers and others to easily accept a predictive model if it is plausibly interpreted. Further, they state that it may be simpler to determine the prediction boundaries, i.e. determine what situations the model will hold and under what situations the model will break, when a plausible substantive interpretation is available. Users of predictive models have more trust in its results, especially for unexpected or counter intuitive predictions, when there is a plausible interpretationpossible (ibid.). In contrast to explanatory modelling, the plausible interpretations in this contextdo not entail a rigorous formal statistical testing of all posited relationships and model constraints as in causal- explanatory modelling.

2.4. Principal Component Analysis

The origins of statistical techniques are often difficult to trace. Preisendorfer and Mobley (1988) note that Beltrami (1873) and Jordan (1874) independently derived the singular value decomposition (SVD) (see Section 3.5) in a form that underlies PCA. Fisher and Mackenzie (1923) used the SVD in the context of a two-way analysis of an agricultural trial. However, it is generally accepted that the earliest descriptions of the technique nowknown as PCA were given by Pearson (1901) and Hotelling (1933). Hotelling's paper is in two parts. The first, most important, part, together with Pearson's paper, is among the collection of papers edited by Bryant and Atchley (1975). The two papers adopted different approaches, with the standard algebraic derivation given above being close to that introduced by Hotelling (1933). Pearson (1901), on the other hand, was concerned with finding lines and planes that best fit a set of points in p-dimensional space, and the geometric optimization problems he considered also lead to PCs, as will be explained in Section 3.2 Pearson's comments regarding computation, given over 50 years before the widespread availability of computers, are interesting. He states that his methods'can be easily applied to numerical problems,' and although he says that the calculations become 'cumbersome' for four or more variables, he suggests that they are still quite feasible.

Based on "Real-Time Measurement of Business Condition" the writers using four ingredients for their framework. The first one was a dynamic model that acts as if the business conditions like an not observed variable but as the business cycle (GDP, sales, industrial production, employment) is interactions of many variables. The second ingredient was to subsume the conditions where business base on as a different differences. For example they use quarterly GDP, monthly industrial production weekly employment and continuously asset pricing and through this they tried to give continuously updated measurements. The third ingredient use incorporate index's at high frequencies and what they wanted to achieve was to find the frequency evolution of real activity. The fourth ingredient they forecasted, using linear procedures, the business conditions. The model that they propose

Application of Principal Component Regression Analysis in power load forecasting for medium and long term . This paper manages the power load forecasting for medium and long forecasting utilizing Principal Component Regression Analysis. The paper first audits the exploration accomplishment of the heap determining and its relationship with financial advancement, at that point presents the fundamental hypothesis of the head part examination and head segment relapse examination model. At last, accepting Beijing as an model, the paper removes the key segments from the pertinent monetary variables related force utilization in Beijing, at that point sets up a multi-parameter forecast model (Principal component regression model) the key factors. The outcomes show that the error is little between expectation load and real load, demonstrating that the model is an attainable and successful strategy for load forecasting. There are two models for load forecasting.the first one is based on power load and the other one is about power consumption. But until this paper the way they model was directly on power consumption. They propose a new way to modeling through Principal Component Analysis using as factor variables that have an impact to power load. Some of the variables that affect the power load is national economic development, industrial structure, national macroeconomic policies, weather and other factors. The reason they use principal Principal Component Analysis is cause they can use all the variables . They confirm that the load forecasting model based on principal component regression meets the long-term load forecasting accuracy requirements on the error, and the model is an effective method of load forecast.

3. Research

The term research refers to the continuous effort of man to seek information and knowledge in order to solve problems and questions that arise during his existence on earth. Research is therefore a primary process and in general a key "tool" for improving people's living conditions. An example of research is the human effort to develop agriculture through nature observation and experience on earth. Of course, as we all know, today's research has moved to more modern and more advanced levels. This is due to the fact that nowadays scientists continue to seek and experiment with developments and new ideas and knowledge, extending the already existing knowledge of all these thousands of years of human experience on Earth. Research is generally distinguished from:

- Basic research that tries to clarify previously unknown objects, behavioral mechanisms, basic structures or functional relationships of an elementary nature. Basic scientific research, for example, deals with the function of organisms in biology or the interactions of substances in chemistry and physics. Basic humanities research, for example, is about the phenomenon of education. It explores historically or socially relevant laws of human behavior. This research is carried out systematically and according to the mandate, especially at scientific universities. An example of basic European research is in particular CERN (European Organization for Nuclear Research) in Geneva and European Synchrotron Radiation Facility in Grenoble. In Germany, special research institutions such as the non-profit research organisation Max Planck Society (MPG) and the institutes of the Helmholtz Association of German Research Centres (HGF) are also involved. In Austria, institutions such as the Austrian Academy of Sciences (ÖAW) work in basic research. In Italy, Trieste is considered a centre of basic research with the International Centre for Theoretical Physics (ICTP), the Elettra Sincrotrone Trieste research complex with, among others, the electron accelerator Elettra and the free-electron laser FERMI. Basic research serves to expand elementary scientific knowledge. The scope of application is not in the foreground of interest. Basic research provides a foundation for applied research and development.

- Translational research, further, targeted basic research at the interface to applied research, which is based on self-acquired scientific knowledge and is oriented towards concrete application goals and/or an economic, social or cultural benefit to be developed. This includes, for example, research carried out by the Leibniz Association
- Applied research (including special purpose research) that aims to solve a practical, often technical or medical problem. It pursues economic exploitation and takes place both at universities and in the free economy, in Germany also at the institutes of the Fraunhofer Society. Other countries also have similar, partly state-funded institutions, such as the TNO in the Netherlands, the Austrian Institute of Technology (AIT) in Austria or the AREA Science Park in Trieste, Italy. In the narrower sense, a distinction is still made between process and product research in applied research. The lessons learned are translated into technical developments.

While basic research is guided by a pure interest in knowledge and tries to find generally valid connections and legalities, applied research is oriented towards practical, useful results, as something in medical research. Each of the two lines of research can be a driving force for the other and benefit from the other. Basic research works at a higher level of abstraction, application research moves closer to practical usability. Stanford University in California, with the Stanford Linear Accelerator Center, research and studies in natural and engineering sciences, and Silicon Valley IT companies, is seen as an international model for combining basic research, application research, and economic use. The basic parts of a scientific research are no different from the general content of a university work. Thus, in order to be properly substantiated it must include the following contents: introduction, the theoretical background ie the bibliographic overview of the topic, the methodology set out, the results obtained, the conclusions and finally the literature. (Shakir) It is not uncommon for scientific research and research to make people feel awe because they are convinced that these are terms that apply only to university and research center members. In order to eliminate such a feeling we will define below the concept of scientific research and formulate the most important qualities and qualifications of the researcher.

There are many definitions of research that one can find in international literature. We find it more appropriate to adopt the most general and most accepted: "Research is the systematic attempt to find answers to questions" Tuckman (1998). The adoption of such a general definition is obviously not accidental. It happens because all the other more specific and more specific definitions have several functionality problems. More specifically, this means that these definitions leave out scientific research activities that everyone would agree to be research or that require interpretation of a term that they include and which they refer back to. In addition, a list of the most important features of a proper researcher who wishes to carry out a comprehensive and clearly true research will be presented. Perhaps the most important attribute of a researcher is honesty. This may not seem logical at first sight. However, because the researcher has the ability to modify and alter his research hypothesis retrospectively, present untrue data, alter existing data and situations, make unauthorized interventions during statistical data processing, customize work and discoveries other researchers. Thus, it is not unreasonable to advocate for honesty, since most of the time without some punishment. In earlier years, reference would be made to the scientific ethos of the researcher. But since the concept of ethics is rather vague, the term sincerity is preferred nowadays.

Depending on the aim of its study, scientific research can be divided into three categories: exploratory, descriptive and experimental. Exploratory research is carried out by highly experienced individuals and it is essential that experts be involved in their conduct as their main purpose is discovery and innovation. During investigative investigations, a problem arises and should, after all its cases have been dealt with, be prioritized and then analyzed. Descriptive research is used to identify and then evaluate the characteristics of a problem under study, and the scientists conducting it need to be impartial and carry out organized and delineated research in order to perform them properly. Finally, the experimental studies are based on the physical or technical experiments and try to check whether the assumptions they used were correct. Thus, a systematic relationship between two variables is explored. This means that they are interested in two things. First, if one variable appears together with another, and whether the changes in that variable are related to changes in another variable.

Komilis (2006) states that experimenting with an independent variable is the simplest form of experimental design. Includes a two-level independent variable. The participants are randomly divided into two groups. The first group, otherwise called the experimental group, is the one that is subjected to the state representing the independent variable. The second group which does not undergo this condition is known as the control group. The researcher records individuals' performance in the dependent variable for both groups (both levels of the independent variable) and then assesses whether these two groups have statistically significant differences in their performance.

Despite the cases where the researcher has performed random sampling, the researcher is uncertain that the differences observed between two groups are due solely to the experimental intervention. Thus, the initial performance of the members of each experimental group must be known, in order to prove with great certainty that the experimental manipulation was the cause of the change in the values of the dependent variables. To achieve this, a researcher may select a modified experimental design, which will include repeated measurements (before and after the experimental manipulation). The experiment in its improved form is essentially equivalent to two studies taking place in parallel. One of them studies the changes that the individual variable has on the performance of the individual variable, while the other examines the differences between the two groups.

There are three main types of surveys; the quantitative, the qualitative and the mixed research. Below we will provide some additional information on each of these types of research.

We begin with quantitative research whose purpose is to identify relationships between various factors and is often used for the systematic investigation of phenomena. Quantitative methodology includes statistical methods, mathematical models and figures. A representative sample of observations is usually used and then generalized to a wider population. When it comes to collecting the required data, it is done with data protocols such as questionnaires, achievements and tables.

Qualitative research can work either independently or in addition to quantitative techniques. Qualitative methods can be used for the understanding of social phenomena, both in the phases of exploring and deepening. Qualitative research has the advantage that it enables the researcher to obtain rich information on the subject he is considering. Qualitative research is clearly the most appropriate methodology for answering questions related to <> and <> phenomena, since it is a mainly descriptive method, aiming more at the creation of new formulations and theoretical models. Thus, the qualitative approach is not specifically performed in order to verify single cases or generalize to a larger population, and it is preffered because of its flexibility during the research process. When looking for in-depth attitudes, perceptions, representations, motivations, emotional data, symbolic data, and general behavioral data, qualitative research is the most appropriate methodological choice. This is by no means accidental because the purpose of qualitative inquiry is not simply a description of a behavior or attitude but a holistic understanding. Qualitative research studies always focus on the broader social and cultural context in which it subscribes, whereas smaller samples have involved as well as by speech and text analysis which are some of the key meathods of the qualitative approach.

Mixed approaches, which are based on the combination of qualitative and quantitative methods, are done during the process of their methodological design in order to exploit the advantages of each method but also to deal more effectively with each other's weaknesse

4. ResearchAnalysis

In our research we perform two methods in order to shrink the dimensions and compare them with an Autoreggressive model. First, we choose the parameters that we use as dependent variable. The data set we used was large macroeconomic databases designed for the empirical analysis of ‘ big data’, Michael W. McCracken(2015).This paper introduces researchers to a set of 134 monthly macroeconomic variables based on the FRED database from 1959.

The variables we selected were unemployment and industrial production. The methods that take place arePricipal Component Analysis and Partial Least Square. As a result, we create factors for each method in every variable (unemployment and industrial production). Although, we first tried cross validation for the number of factors, the number was too big to use. Thus, we use Ng and Bai criterion and we use one factor for each model.

For the Autoregressive model we use an ARIMA (3, 1, 0) model for unemployment and an ARIMA (2, 1, 0) for industrial production. The criterion we used for the number of lags was Bayesian information criterion (BIC)**.**

We cut the data set in two parts, the firtst part was the train and the other was the test. Then, we perform one step ahead forecast for 100 periods. We repeat the same procedure for each model and in the end, we find the MSE and MAE.

$$\mathrm{MSE} = \frac{1}{n}\sum_{i-1}^{n}(Y_i - \hat{Y}_i)^2. \qquad \mathrm{MAE} = \frac{\sum_{i=1}^{n}|y_i - x_i|}{n} = \frac{\sum_{i=1}^{n}|e_i|}{n}.$$ [1]

The results that occured for uneployment were :

MODEL MSE /AR MSE

	MSE	MAE
PCA	0.945	0.964
PLS	1.015	1.089

It is odvious that PCA has better result for the unemployment compared to AR. On the other hand, the PLS does not perform so welland the AR will have better results.

The results that occured for industrial production were:

MODEL MSE /AR MSE		
	MSE	MAE
PCA	1.054	1.071
PLS	1.156	1.187

It is clear that AR is better for industrial production than PCA and PLS.

5. Conclusions

Last but not least, there are plenty of methods in literature that we can use in Machine Learning and Statistical analysis. Through them we can find r ways to improve and perform better technics with better possible outcomes. In this paper, we use two methods (PCA, PLS) and as a result we find better result in one dependent variable. That means is still room for new methods and always expand new technics.

6. References

Faust, J., Gilchrist, S., Wright, J. H. and Zakrajšek, E. (2013). Credit spreads as predictors of real-time economic activity: a Bayesian model-averaging approach. Review of Economics and Statistics, 95, 1501–1519.

Giacomini, R. and Rossi, B. (2010). Forecast comparisons in unstable environments. Journal of Applied Econometrics, 25, 595–620.

Giacomini, R. and White, H. (2006). Tests of conditional predictive ability. Econometrica, 74, 1545–1578.

Gilchrist, S. and Zakrajšek, E. (2012). Credit spreads and business cycle fluctuations. American Economic Review, 102, 1692–1720.

Gilchrist, S., Yankov, V. and Zakrajšek, E. (2009). Credit market shocks and economic fluctuations: evidence from corporate bond and stock markets. Journal of Monetary Economics, 56, 471–493.

Hall, R. E. (2011). The long slump. American Economic Review, 101, 431–469.

Hammerstein, E. A. v. (2010). Generalized hyperbolic distributions: Theory and Applications to CDO Pricing. Ph.D. thesis, PhD thesis, Universit¨at Freiburg.

Hännikäinen, J. (2015). Zero lower bound, unconventional monetary policy and indicator properties of interest rate spreads. Review of Financial Economics (forthcoming).

Hartigan, J. A. and Wong, M. A. (1979). Algorithm as 136: A k-means clustering algorithm. Journal of the Royal Statistical Society. Series C (Applied Statistics), 28(1), 100–108.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. Journal of Educational Psychology, 24(6), 417.

Kottas, A. and Krnjaji´c, M. (2009). Bayesian nonparametric modelling in quantile regression. Scandinavian Journal of Statistics 36, 297–319.

Li, L. (2007). Sparse sufficient dimension reduction. Biometrika 94, 603–613.

McNicholas, P. D. (2016). Mixture model-based classification. CRC Press.

McNicholas, P. D., Murphy, T. B., McDaid, A. F., and Frost, D. (2010). Serial and parallel implementations of model-based clustering via parsimonious gaussian mixture models. Computational Statistics & Data Analysis, 54(3), 711–723.

McNicholas, P., Jampani, K., McDaid, A., Murphy, T., and Banks, L. (2011). pgmm: Parsimonious gaussian mixture models. R package version, 1(1).

McNicholas, S. M., McNicholas, P. D., and Browne, R. P. (2017). A mixture of variance-gamma factor analyzers. In S. E. Ahmed, editor, Big and Complex Data Analysis, pages 369–385.

Meng, X.-L. and Van Dyk, D. (1997). The EM algorithm-an old folk-song sung to a fast new tune. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 59(3), 511–567.

Ng, S. and Wright, J. H. (2013). Facts and challenges from the Great Recession for forecasting and macroeconomic modeling. Journal of Economic Literature, 51, 1120–1154.

Scott, D. W. and Thompson, J. R. (1983). Probability density estimation in higher dimensions. In Computer Science and Statistics: Proceedings of the fifteenth symposium on the interface, volume 528, pages 173–179. North-Holland, Amsterdam.

Sigillito, V. G., Wing, S. P., Hutton, L. V., and Baker, K. B. (1989). Classification of radar returns from the ionosphere using neural networks. Johns Hopkins APL Technical Digest, 10(3), 262–266.

Spearman, C. (1904). The proof and measurement of association between two things. The American Journal of Psychology, 15(1), 72–101.

Springer. Meng, X.-L. and Rubin, D. B. (1993). Maximum likelihood estimation via the ECM algorithm: A general framework. Biometrika, 80(2), 267–278.

Stock, J. H. and Watson, M. W. (2003). Forecasting output and inflation: the role of asset prices. Journal of Economic Literature, 41, 788–829.

Tibshirani, R., Saunders, M., Rosset, S., Zhu, J. and Knight, K. (2005). Sparsity and smoothness via the fused lasso. Journal of the Royal Statistical Society, Series B 67, 91–108.

Tiedeman, D. (1955). On the study of types. In S. B. Sells, editor, Symposium on Pattern Analysis, pages 1–14, Randolph Field, TX: Air University. U.S.A.F. School of Aviation Medicine.

Tortora, C., McNicholas, P. D., and Browne, R. P. (2015). A mixture of generalized hyperbolic factor analyzers. Advances in Data Analysis and Classification, 10(4), 423–440.

Walentin, K. (2014). Business cycle implications of mortgage spreads. Journal of Monetary Economics, 67, 62–77.

Wang, Q. and Yin, X. (2008). A nonlinear multi-dimensional variable selection method for high-dimensional data: sparse MAVE. Computational Statistics and Data Analysis 52, 4512–4520.

Wolfe, J. H. (1963). Object Cluster Analysis of Social Areas. Master's thesis, University of California.

Wu, T. Z., Yu, K. and Yu, Y. (2010). Single-index quantile regression. Journal of Multivariate Analysis 101, 1607–1621.

Yuan, M. and Lin, Y. (2006) Model selection and estimation in regression with grouped variables. Journal of the Royal Statistical Society, Series B 68, 49–67.

Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net, Journal of the Royal Statistical Society, Series B 67, 301–320.

APPENDIX

For the empirical research I use as a programming language R.To creates Partial Least Square we made this process:

```
#Partial Least squares.

xtra<-model.matrix(UNRATE~.,training)[ , -25]

ytra<-na.omit(training$UNRATE)

xtest<-model.matrix(UNRATE~.,testing)[ , -25]

summary(xtest)

ytest<-na.omit(testing$UNRATE)

xtra1<-model.matrix(INDPRO~.,training)[ , -7]

ytra1<-na.omit(training$INDPRO)

xtest1<-model.matrix(INDPRO~.,testing)[ , -7]
```

```
summary(xtest)

ytest1<-na.omit(testing$INDPRO)

install.packages("pls")

library(pls)

help("pls")

t.seed(234)

pcr.fit=plsr(UNRATE~.,scale=TRUE,validation ="CV",data= testing)

summary(pcr.fit)

validationplot(pcr.fit,val.type = "RMSEP")

pls.pred= predict(pcr.fit,data=,ncomp = 1)

head(pls.pred)

head(testing$UNRATE)

mean((pls.pred-ytest)^2) #gia sum of squares

plot(RMSEP(pcr.fit))

pca.fit = pcr(UNRATE~.,scale=TRUE,validation ="CV",data= testing)
```

```
summary(pca.fit)

plot(RMSEP(pca.fit))

pca.pred= predict(pca.fit,data=,ncomp = 1)

head(pca.pred)

head(testing$UNRATE)

mean((pca.pred-ytest)^2)

pca.fit1 = pcr(INDPRO~.,scale=TRUE,validation ="CV",data= testing)

plot(RMSEP(pca.fit1),xlab = 4)

pca.pred1= predict(pca.fit,data=,ncomp = 1)

head(pca.pred1)

head(testing$INDPRO)

mean((pca.pred1-ytest)^2)
```

For the PCA:

```
# Principal Component Analysis

set.seed(111)
```

```
ind <- sample(2, nrow(my_data),

              replace = TRUE,

              prob = c(0.8, 0.2))

training <- my_data[ind==1,]

testing <- my_data[ind==2,]

library(psych)

pairs.panels(training[,-25],

             gap = 0,

             bg = c("red", "yellow", "blue")[training$Species],

             pch=21)

pc <- prcomp( training[,-25],

             center = TRUE,#kentro ajonon

             scale. = TRUE )# sto 1

attributes(pc)

pc$center
```

```
pc$scale s

print(pc)

summary( pc )# level of variance

plot ( pc )

plot ( pc$x[,1] , pc$x[,2] )

# Prediction with Principal Components

trg1 <- predict( pc ,  training )

trg1

trg <- data.frame( trg1, training [25] )

tst1 <- predict ( pc, testing )

tst1

tst <- data.frame(tst1, testing[25])

#bai and ng

install.packages("POET")

library(POET)
```

```
help("POETKhat")

K<-POETKhat(trg1)

L <- POETKhat(t (trg1) )

summary (L)



#Linear Regression

attach(trg)

ols <- lm(UNRATE~PC1,trg)

summary(ols)

olstest<-lm(UNRATE~PC1,tst)

summary(olstest)

help(plot)

"plot"#Predictions

ptrain <- predict ( ols,trg , n.ahead= 5 )

head(ptrain)
```

```
head ( trg$UNRATE )

ptest<-predict( olstest, tst )

head( ptest )

head( tst$UNRATE )
```

Last but not least, the AR process:

```
#AR models
Library ( vars)
install.packages  ("tseries") # for AR
install.packages  ("ggplot2") # for plots
library (tseries)
library  (ggplot2)
install.packages ("forecast") #for forecasts
library  (forecast)
attach ( my_data)
plot (UNRATE)
acf UNRATE)
pacf( UNRATE)
adf.test(UNRATE,alternative = "stationary") # checking stationary for uneployment
d.UNRATE <- diff(UNRATE)
```

```
plot(d.UNRATE)
adf.test(IDPPRO,alternative = "stationary") # hecking stationary for industrial production
# creating plots
acf(d.UNRATE)
pacf(d.UNRATE)
auto.arima(UNRATE,ic=c("aic"))
auto.arima(d.UNRATE,ic=c("bic"))
yyy=arima(UNRATE,c(3,0,0))
summary(yyy)
forecast=predict(arima(UNRATE,c(4,1,0)))
head(forecast)
mean((forecast-ytest)^2)
R = mu + sqrt(vr)*randn(100)
endiamesos_pinakas=matrix(,nrow=60,ncol=ncol(R))
for (i in (nrow(endiamesos_pinakas)+1):nrow(R)){
  j=i-nrow(endiamesos_pinakas)
  k=i-1
  for(beta in  1:ncol(endiamesos_pinakas)){


    endiamesos_pinakas[ ,beta]=R[j:k,beta]
mse(k,predict(k))
  }
  }
View (INDPRO)
View (IDPRO)
Acf (IDPPRO)
adf.test   (IDPPRO,alternative = "stationary")
acf(IDPPRO)
acf(IDPPRO)
summary(IDPPRO)
```

```
auto.arima(IDPPRO,ic=c("bic"))
#results to industrial production
arima(IDPPRO,c(2,0,0))
for (i in (nrow(endiamesos_pinakas1)+1) :nrow(R)){
  j= i-nrow (endiamesos_pinakas1)
  k= i-1
  for (beta in  1:nco l(endiamesos_pinakas1)){

    endiamesos_pinakas[ ,beta]=R[j:k,beta]
    mse ( k, predict(n))
  }
}
```

For industrial production :

```
attach(my_data)
remov# Partition Data
set.seed(111)
ind <- sample(2, nrow(my_data),
             replace = TRUE,
             prob = c(0.8, 0.2))
training <- my_data[ind==1,]
testing <- my_data[ind==2,]

# Scatter Plot & Correlations
library(psych)

pairs.panels(training[,-25],
            gap = 0,
            bg = c("red", "yellow", "blue")[training$Species],
            pch=21)
```

```
# Principal Component Analysis
pc <- prcomp(training[,-25],
        center = TRUE,#kentro ajonon
        scale. = TRUE)# sto 1

attributes(pc)
pc$center
pc$scale
print(pc)
summary(pc)# level of variance
plot(pc)
plot(pc$x[,1], pc$x[,2])

# Prediction with Principal Components
trg1 <- predict(pc, training)
trg1
trg <- data.frame(trg1, training[25])
tst1 <- predict(pc, testing)
tst1
tst <- data.frame(tst1, testing[25])

#bai and ng (DEN TO XO)
install.packages("POET")
library(POET)

help("POETKhat")
K<-POETKhat(trg1)
K

L <- POETKhat(t (trg1) )
```

```
L
summary (L)

hist(pc)

#Linear Regression
attach(trg)
ols <- lm(UNRATE~PC1,trg)
summary(ols)
olstest<-lm(UNRATE~PC1,tst)
summary(olstest)


#Predictions

ptrain <- predict(ols,trg,n.ahead= 5 )
head(ptrain)
head(trg$UNRATE)

ptest<-predict(olstest,tst)
head(ptest)
head(tst$UNRATE)


#Partial Least squares.


xtra<-model.matrix(UNRATE~.,training)[ , -25]
ytra<-na.omit(training$UNRATE)
```

```
xtest<-model.matrix(UNRATE~.,testing)[ , -25]
summary(xtest)
ytest<-na.omit(testing$UNRATE)

install.packages("pls")
library(pls)
help("pls")



set.seed(234)
pcr.fit=plsr(UNRATE~.,scale=TRUE,validation ="CV",data= testing)
summary(pcr.fit)
validationplot(pcr.fit,val.type = "RMSEP")
pls.pred= predict(pcr.fit,data=,ncomp = 36)
head(pls.pred)
head(testing$UNRATE)
mean((pls.pred-ytest)^2) #gia sum of squares
plot(RMSEP(pcr.fit))

pca.fit = pcr(UNRATE~.,scale=TRUE,validation ="CV",data= testing)
summary(pca.fit)
plot(RMSEP(pca.fit))
pca.pred= predict(pca.fit,data=,ncomp = 100)
head(pca.pred)
head(testing$UNRATE)
mean((pca.pred-ytest)^2)

pca.fit = pcr(~.,scale=TRUE,validation ="CV",data= testing)
summary(pca.fit)
plot(RMSEP(pca.fit),xlab = 10)
```

```
pca.pred= predict(pca.fit,data=,ncomp = 100)
head(pca.pred)
head(testing$UNRATE)
mean((pca.pred-ytest)^2)
```

Finding the importance of the variables :

Read Data

```
 attach(my_data)
str(my_data)
my_data$UNRATE <- as.factor(my_data$UNRATE)
table(my_data$UNRATE)

# Data Partition
set.seed(123)
ind <- sample(2, nrow(my_data), replace = TRUE, prob = c(0.7, 0.3))
train <- my_data[ind==1,]
test <- my_data[ind==2,]

# Random Forest
library(randomForest)
set.seed(222)
rf <- randomForest(UNRATE~ ., data=train)
print(rf)
attributes(rf)

# Prediction & Confusion Matrix - train data
library(caret)
p1 <- predict(rf, train)
confusionMatrix(p1, train$NSP)

# # Prediction & Confusion Matrix - test data
```

```
p2 <- predict(rf, test)
confusionMatrix(p2, test$NSP)

# Error rate of Random Forest
plot(rf)

# Tune mtry
t <- tuneRF(train[,-22], train[,22],
       stepFactor = 0.5,
       plot = TRUE,
       ntreeTry = 300,
       trace = TRUE,
       improve = 0.05)

# No. of nodes for the trees
hist(treesize(rf),
     main = "No. of Nodes for the Trees",
     col = "green")

# Variable Importance
varImpPlot(rf,
           sort = T,
           n.var = 10,
           main = "Top 10 - Variable Importance")
importance(rf)
varUsed(rf)

# Partial Dependence Plot
partialPlot(rf, train, ASTV, "2")
```

```
# Extract Single Tree
getTree(rf, 1, labelVar = TRUE)
# Multi-dimensional Scaling Plot of Proximity Matrix
MDSplot(rf, train$NSP)
```