"Testing the Stability of the Components Explaining Changes of the Yield Curve in Mexico. A Principal Component Analysis Approach"

Víctor M. Jiménez Escobar
Mexico

Summary

A number of research works in finance rely on the selection of factors affecting the behaviour of one or more underlying variables. Factor Analysis (FA) and Principal Components Analysis (PCA) are statistical tools allowing for selecting a reduced number of factors explaining the variations of a set of variables. Of particular interest has been the study of the factors explaining the behaviour of the yield curve. It is well known that at least three factors are relevant for explaining variations in bonds and money market returns; these factors are level, slope and curvature. Financial practitioners use the set of components for portfolio hedging strategies. This document aims to test the stability of the factors over time. It is assumed that the factors change in a continuous time basis. Another hypothesis is that they are the same for a period but the variance explained for each one of them changes constantly. Technically, it means that the off-diagonal elements of a number of covariance matrices remain are the same but the diagonal elements are specific to each one of them. Three approaches were use to test the hypothesis of common principal components for the Mexican experience. The results are mixed and though in some cases they appear to support the hypothesis of stability of some of the factors, I found a number of problems, particularly computational and numerical, for appropriately testing it.

"Verificando la Estabilidad de los Componentes que Explican las Variaciones en la Curva de Rendimiento en México con base en Análisis de Componentes Principales"

Víctor M. Jiménez Escobar
Mexico

Resumen

Un número importante de trabajos de investigación en finanzas se basan en la selección de factores que afectan el comportamiento de una o más valores subyacentes. Análisis de Factores y Análisis de Componentes Principales son herramientas estatísticas que permiten la selección de un número reducido de factores o componentes que explican las variaciones de un conjunto dado de variables. De interés particular han sido el estudio de los factores que determinan el comportamiento de la curva de rendimiento. Knez, Litterman y Sheinkman en su trabajo "Common Factors Affecting Bond Returns" (1991) nos muestran que al menos tres factores son relevantes para explicar variaciones en precios de Bonos y retornos sobre instrumentos del mercado de dinero, nivel, pendiente y curvatura. Financieros utilizan este conjunto de factores para implantar estrategia de cobertura. Este trabajo tiene como objetivo probar la estabilidad de los componentes en el tiempo. Generalmente se asume que los factores cambian en el tiempo. Alternativamente, se asume que se mantienen constantes excepto que las varianzas explicadas por cada uno de ellos cambian constantemente. Tcnicamente esto significa que todos los elementos de la matriz de covarianzas permanecen constantes exceptuando por la diagonal. Tres fueron los métodos utilizados para probar la hipótesis de componentes principales comunes para la curva en México en el periodo 1996-99. Los resultados no son contundentes, sin embargo en algunos casos soportan dicha hipótesis.

C O N T E N T S

1. Introduction

A number of research works in finance rely on the selection of factors affecting the behaviour of one or more underlying variables which are considered relevant for a researcher or a finance practitioner. For instance, a portfolio manager is interested in hedging against shifts of the yield curve; in a case like this one, the factor is a vector which determines the level of the curve and a duration technique could be sufficient to cover against the mentioned risk. But what would be the result if the curve moves in a different way? In other words, what would happen if only the long-end would move up or down? What if only the short-end does it? Is it traditional duration enough to cover these cases? Moreover, is it possible for the interest rates to move in different fashions?

Factor Analysis (FA) and Principal Components Analysis (PCA) are statistical tools allowing for selecting a reduced number of factors, in order to explain the variations of a set of underlying variables. Although, these tools are similar in spirit, they are very different one another. The differences will be explained in section four; however, from now on the words factor or component will be used indistinctly during the entire document.

The first works I know directly related to finance using this type of technique were used to test multifactor-pricing models such as Arbitrage Pricing Theory (APT). Roll and Ross (1980); Cho, Elton and Gruber (1988); Connor and Korajczyk (1988) are some of the examples of the use of FA for testing the APT.

Regarding money and bond markets, the paper by Litterman and Scheinkman "Common Factors Affecting Bond Returns" (1991) seems to be one of the most influential works. Thereafter in 1996, Knez, Litterman and Scheinkman (KLS) used the same technique, i.e., FA, in order to develop models of factors influencing bond returns.

These authors analysed weekly returns of 38 money-market instruments with maturities from 1 to 12 months, representing five different risk-segments in the United States for the period January 1985 to August 1988. They are able to identify and interpret at least three factors. The first one, which they name the "level" factor, corresponds to shifts of the whole yield curve; this component explains 62% of the total variation on returns. The second one, "stepness", equivalent to the slope of the curve, explains 11% of changes in returns. Finally, the "Treasury or curvature" factor named this way because as they point out "its effect is to increase the curvature of the yield curve in the range of maturities below 20 years". The explanatory power of the third factor is close to 13%. As it can be directly inferred from these numbers, the three factors account for 86% of the variation in money market returns.

Barbel and Copper (BC) in their "Immunisation Using Principal Component Analysis" (1996), present the alternative tool PCA for obtaining the components affecting bond returns. These authors used a different data set, selecting spot rates with maturities ranging from 1 month to twenty years from August 1985 to February of 1991. Despite the differences, these authors are able to interpret factors in a similar way that those of KLS. In addition, BC use the eigenvectors and the eigenvalues of the covariance matrix of the yield curve to demonstrate that it is possible to hedge against movements in different directions.

Both analyses take us to the question, why do we need to get these factors? Hedging is one answer. The works by Reitano (1996), "Non-Parallel Yield Curve Shifts and Stochastic Immunisation", Duarte and Mendes (1998), "Robust Hedging Using Futures Contracts with an Application to Emerging Markets" are good illustrations of how to use PCA for hedging portfolios. Likewise, the paper by Golub and Tillman, "Measuring Yield Curve Risk Using Principal Components Analysis, Value at Risk, and Key Rate Durations" allows us to contrast PCA against other hedging techniques widely used.

A second application of the components is to determine parameters of the volatility functions in a model of the term structure. This makes it possible to model the term structure of interest rates under a Heath, Jarrow and Morton (HJM) or a Bruce, Gatarek, Musiela (BGM) frameworks.

One of the implicit assumptions when hedging using PCA is that the components are stable through time, but is it a sensible assumption? A method known as Common Principal Component Analysis allows comparing variance – covariance matrices in cross sections or time series and thus to test a kind of "statistical or econometric structural change".

The main objective of this work is to study the behaviour of the components of the yield curve in Mexico in four different periods. One aim is to analyse the stability of the components of the yield curve and if possible to identify the possible causes of change.

2. An Overview of the Mexican Economy

During the last five years, the Mexican economy has experienced a number of difficulties, some of them emerging from its own economic and financial structures and others because of its links with the international community. Among others, it is possible to mention the "Tequila crisis" in 1995, the "Asian - Tigers" and "Russian-Bear" crises in 1997-98 and more recently the "Samba" crisis in Brazil starting the last year and extending itself to the first quarter of 1999.

2.1 The "Tequila" crisis of 1995-96

In 1994, a number of violent political events triggered a profound economic destabilisation in Mexico. At the beginning of that year, a guerrilla rebel group arose in the southern state of Chiapas. In March, the governmental candidate to the presidency of Mexico was assassinated in the northern state of Baja California. In August, the most actively participated presidential elections in the history of the country were carried out in calm and peace, nevertheless there was a latent fear that opposition groups, with the pretext of an electoral fraud, would fiercely impugn the outcome. In September, an eminent politician was killed; he was likely to become the head of the official party and he planned to completely reform its structures. Because of this crime, in recent days, the brother of the former president was sentenced to 28 years of prison accused of being the intellectual responsible. From the economic point of view, although the inflation was 7% in 1994, the lowest in more than 25 years, there were important pressures over main economic indicators. A huge deficit in the trade balance / current account was pointed out by economists as a possible source of a financial turmoil. In addition, the Tesobonos problem, debt instruments issued by the Mexican government in pesos but denominated in dollars[1], became an unbearable burden for the Government once the exchange rate was put under pressure due to the events previously mentioned.

On the 21 of December of 1994, with the purpose of alleviating the constant speculative attacks over the exchange rate, the newly appointed Minister of Finance ordered a shift of 15% of the exchange rate target zone. This measure was not accompanied by an adjustment of interest rates to compensate investors for the depreciation of the currency causing a massive outflow of money, which rapidly dried out the foreign reserves[2]. This was the so-called December mistake. The currency depreciated more than a hundred percent in the following months, inflationary perspectives filled the economic environment, interest rates moved wildly reaching their peak in October; as a result the economy fell 7% in real terms that year. The worst economic crisis of the last decades had started. Despite of the financial aid of international financial institutions, foreign countries, and a number of measures of stabilisation implemented by the government, the country went into a financial recession, which in turn was aggravated by the weakness of the Financial System. Tight monetary and fiscal policies were among others the traditional recipes recommended by the IMF and the World Bank as the major stabilisation policies. Although the economy grew in 1996, the first signs of recovery are not evident until the last quarter of that year. In 1997, the economy started to expand again.

2.2 The succeeding crisis

The economic environment in Mexico has been characterised by intermittent periods of stability combined with financial pressures coming from the exterior, to mention a few, the Asian crisis, the turmoil in Russia, the reduction of the petroleum prices and finally the problems in Brazil.

2.2.1 The "Tiger" and "Bear" crisis of 1997-98

It is well known that the financial crisis emerged on Thailand, Indonesia, South Korea, Malaysia, Taiwan, and on a minor scale in China, Hong Kong and Singapore, caused an important depreciation of their respective currencies and a sharp fall on their respective level of output. Mexico was not an exception, the crisis increased the volatility of the exchange and interest rates, and provoked a downward trend in the Mexican Stock Exchange. Nevertheless, it does not seem to have created an important economic instability mainly because of stronger economic fundamentals or because of the vigour of the American economy.

2.2.2 The "Samba" crisis in Brazil 1998-99

The remains of the crisis in Asia and Russia suddenly contaminated other regions, and Brazil was the most vulnerable economy of the big Latin American countries. At that time, this country was running a huge fiscal deficit of 8% of the GDP, in contrast the deficit of Argentina and Mexico was 1.3% and 1.4% respectively[3]. After almost four months of speculative attacks over the Brazilian real, the Brazilian government took the decision of releasing the currency to a floating regime on the 13 of January and by the end of February, the currency had already depreciated 40%.

Volatility was the main consequence for the Mexican finances. The effects of the Brazilian crisis overcame those caused by the Asian crisis. For instance, interest rates rose to more than 40% in annual terms, substantially increasing the past-due loan portfolio of the banking system. In the market, it is estimated that the Mexican banking system requires US$13 billion in order to reach the 8% capital to risk weighted assets ratio according to the rules of the BIS of Basle. Consequently, the government has started the intervention of banks in precarious financial situation such as Serfin, the third largest bank in the country and more recently Bancrecer.

In the second quarter of 1999, the Mexican economy seems to be experiencing a new time of financial stability. However, it is possible to foreseeing new challenges in the near future, for example the elections in Argentina, Chile and Mexico during the present and the next year.

From this summary of events, it is possible to identify four periods for the Mexican economy, which will be retaken on next sections of this work. i) The Tequila crisis in 1995-96, ii) A period of stability in 1997 and first half of 1998, iii) the Samba crisis on the last quarter of 1998 and first quarter of 1999, and finally iv) a new period of economic peacefulness nowadays.

3. The Term Structure of Interest Rates in Mexico

This section describes the data, its sources and the methodology for fitting the yield curve.

3.1 Data Description

The data series comes from different sources including the Banco de Mexico (Mexico's Central Bank, known as Banxico), Operadora de Bolsa one of the largest brokerage firms of the country and Datastream. Data corresponds to non-default risk interest rates for maturities of one day, 28 days, 90 days, 182 days and 360 days traded daily in the secondary market. The day count convention is 365 days. The period of analysis starts on June of 1996 and finishes on June of 1999. The total number of observations between these dates is 785. Table 1 shows a sample of the series.

Table 1. A sample of Interest rates (%)

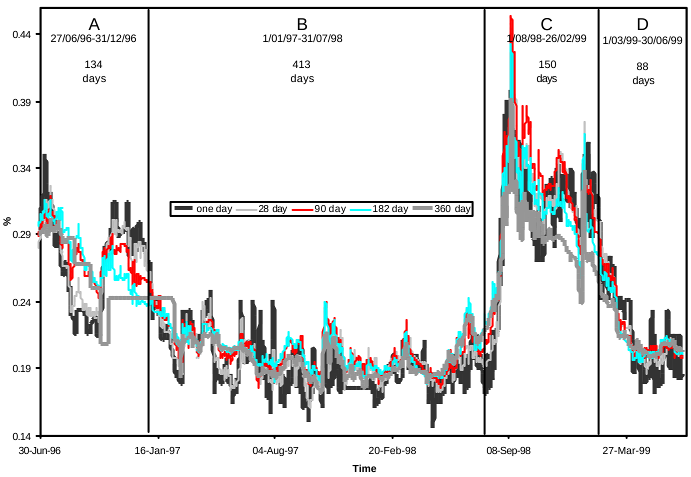
Date	Maturity in years 0.00274	0.07671	0.24658	0.49863	0.98630
27-Jun-96	29.20	28.29	30.25	32.25	33.75
31-Oct-96	31.25	30.50	29.00	29.10	27.50
30-Jun-97	22.90	21.00	20.90	21.20	21.40
30-Sep-98	31.25	35.00	38.10	37.00	34.25
31-Mar-99	24.00	22.30	22.00	21.40	23.05
30-Jun-99	18.75	20.00	20.25	21.25	22.40

Source: Banxico
OBSA
Datastream

According to table 1, rates on the 27 of June, 31 of October of 1997 and 30 of September of 1998 are higher than the others. This is not casual, the dates are chosen to be "representatives" of each one of the periods of section 2.2.

Figure 1, shows the process followed by the interest rates mentioned above. The interest rates are highly correlated and it can be seen the significant volatility particularly of the short rate. Contrarily the 360 days rate seems to have less fluctuation. Again, the four stages of the Mexican economy are evident, i.e., crisis of 1995-96, the stability of 1997-first half of 1998, Brazilian crisis in the third quarter of 1998 and first quarter of 1999 and the steady state in the second quarter of the present year. These sub-periods will be label as A, B, C and D respectively from now on.

Figure 1. The process followed by interest rates in Mexico



3.2 Fitting the Yield Curve

For extracting the components explaining interest rate changes, it is necessary to get more points of the curve, for this reason, I will start the analysis applying Spline interpolation, informally testing its goodness of fit and comparing it against other methodologies. Firstly, I will start explaining Spline Interpolation.

3.2.1 Spline Interpolation

In practice, third degree polynomials are often used to fit a set of points, because they allow a "smooth" shape, which is not possible to get via either linear or squared interpolation. A general formula for a cubic Spline is given by the following expression

$$B_i(t) = \begin{cases} \sum_{j=i}^{i+4} \left(\prod_{\substack{k=p \\ j \neq k}}^{i+4} \frac{1}{(x_k - x_j)} \right) (t - x_i)_+^3 \\ 0 \quad elsewhere \end{cases} \tag{1}$$

where ξ_j is a set of interpolating knots. $B_l(\tau)$ are cubic n-3 B-Splines which are non-zero in the interval $[\xi_l, \xi_{l+4}]$. The properties of the Spline interpolation can be mathematically expressed in the following way. Writing F_k is the kth polynomials over the ξ_k knot

i) They are everywhere twice differentiable,

ii) If ξ_k is a knot, then

$$\begin{aligned} F_k(x_k) &= F_{k+1}(x_k) \\ F_k'(x_k) &= F_{k+1}'(x_k) \\ F_k''(x_k) &= F_{k+1}''(x_k) \end{aligned} \tag{2}$$

Equation 2 implies that excepting at the extreme knots, not only the slope but also the curvature must match.

3.2.2 The Bond Price Curve

The prices of the pure discount bonds are associated to spot rates through the following relation.

$$_tP_T = \frac{1}{(1+\delta(T)_t R_T)} \tag{3}$$

Equation 3 is simply the equivalent discount factor of a spot rate, where $_tP_T$ is the price of a pure discount bond starting at time t and paying 1 Mexican peso at maturity T. $_tR_T$ is the corresponding market quoted spot rate at time t and expiration at time T; $\delta(T)$ is the tenor of the interest rate on a 365 days base, so for instance, a $\delta(1) = 0.0027397$ and $\delta(360)= 0.98630$. Imposing the condition that $_tP_0$ equal to 1, we have an additional point we would not have had otherwise.

Table 2, shows a sub-sample of the pure discount bond prices calculated this way.

Table 2. A sample of the calculated pure discount bond prices using equation 3

Date	Maturity in years					
	0.00000	0.00274	0.07671	0.24658	0.49863	0.98630
27-Jun-96	1.0000	0.9992	0.9788	0.9306	0.8615	0.7503
31-Oct-96	1.0000	0.9991	0.9771	0.9333	0.8733	0.7866
30-Jun-97	1.0000	0.9994	0.9841	0.9510	0.9044	0.8257
30-Sep-98	1.0000	0.9991	0.9739	0.9141	0.8442	0.7475
31-Mar-99	1.0000	0.9993	0.9832	0.9485	0.9036	0.8148
30-Jun-99	1.0000	0.9995	0.9849	0.9524	0.9042	0.8190

Note: This sample of bond prices corresponds to the interest rates shown in table 1. Applying equation 3 to the rates of table 1 with the according maturities, we obtain table 2.

As it is illustrated in table 2, we have 6 pure discount bond prices for every day of the research period; thus, it is possible, by using the spline technique mentioned in the previous section, to fit 3 polynomials of third degree. In total, 785 splines or 2,355 third-degree polynomials were fitted. Table 2a shows the coefficients of the splines corresponding to the data in table 2.

Table 2a. Coefficients of the splines for the pure discount bond prices of table 2

	From 0 to 28 days				From 28 to 91 days				From 91 to 360 days			
Coefficient	T^3	T^2	T	Constant	T^3	T^2	T	Constant	T^3	T^2	T	Constant
Date												
27-Jun-96	*-2.094*	*0.370*	*-0.293*	*1.000*	*0.295*	*-0.179*	*-0.251*	*0.999*	*0.024*	*0.021*	*-0.300*	*1.003*
31-Oct-96	*-0.073*	*0.220*	*-0.314*	*1.000*	*-0.376*	*0.290*	*-0.320*	*1.000*	*0.070*	*-0.040*	*-0.238*	*0.993*
30-Jun-97	*-2.207*	*0.490*	*-0.231*	*1.000*	*0.095*	*-0.039*	*-0.191*	*0.999*	*0.001*	*0.030*	*-0.208*	*1.000*
30-Sep-98	*1.992*	*-0.524*	*-0.312*	*1.000*	*0.592*	*-0.202*	*-0.337*	*1.001*	*-0.131*	*0.333*	*-0.469*	*1.011*
31-Mar-99	*-1.982*	*0.450*	*-0.242*	*1.000*	*0.184*	*-0.049*	*-0.204*	*0.999*	*-0.093*	*0.156*	*-0.254*	*1.003*
30-Jun-99	*1.542*	*-0.269*	*-0.185*	*1.000*	*-0.212*	*0.134*	*-0.216*	*1.001*	*0.045*	*-0.056*	*-0.169*	*0.997*

T indicates time to maturity of the bond, so for example if we want to calculate the price of a pure discount bond maturing in 25 days and starting in 27 of June of 1996, the we would have to

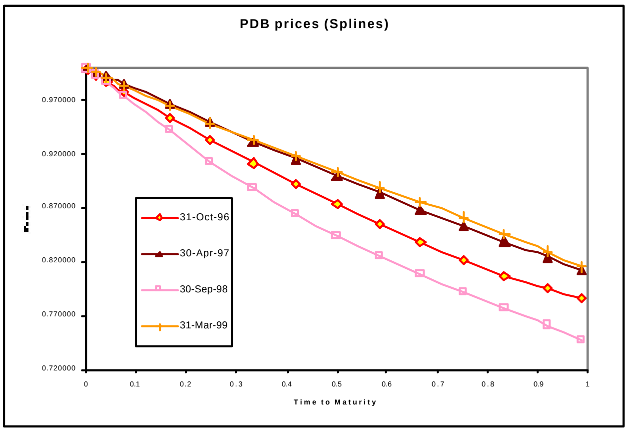
1) calculate time to maturity T in years with a base of 365 days: 25/365=0.0685
2) use the equation with the parameters corresponding to the interval from 0 to 28 days:

$$f(T)=-2.0938T^3+0.3703T^2-0.293T+1=0.980996$$

According to table 2a, we have explicit solutions for the prices of pure discount bonds for the whole period, in other words, we have explicit formulae of the bond price curves for 785 days.

I extracted 15 points from each curve corresponding to pure discount bonds maturing in 1, 7, and 15 days and from 1 to 12 months. Figure 3 illustrates a sample of the bond price curves corresponding to the same dates to those of table 2. In addition, the 15 points extracted are marked in order to show their distribution along their respective bond price curve.

Figure 3. The Bond Price Curve



As we can expect, the figure shows traditional downward sloping curves. Curves on the upper part of the graph correspond to the periods of stability B and D in which rates are lower than those of the turbulent sub-periods A and C.

3.2.3 The Yield Curve

As I mentioned in the previous section, now we have 15 points of the bond price curve and as the pure discount prices are a one to one mapping with the spot rates, hence it is possible to get more points on the yield curve by using the relationship.

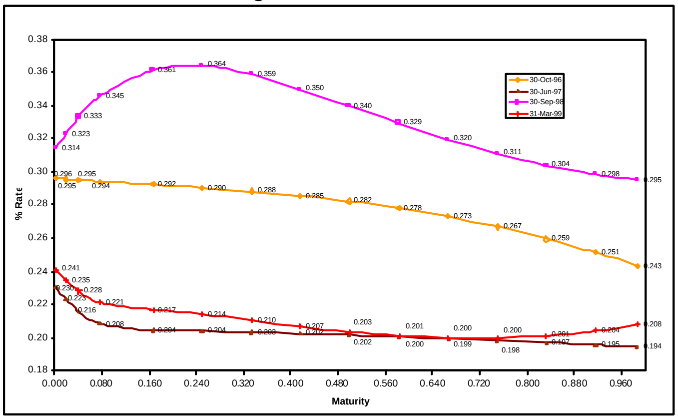
$$ {}_tP_T = Exp\{-R(t,T)\mathrm{d}(T)\} \qquad (4) $$

Where all the variables are as in equation 1 but R(t,T) is the continuous time version of the ${}_tR_T$ mentioned above. Solving for R(t,T), we have:

$$ R(t,T) = -\frac{Log\, {}_tP_T}{\mathrm{d}(T)} \qquad (5) $$

As in the case of the Bond price curve, figure 4 shows the yield curve for the same dates. The curves in June of 1997 and March of 1999 are “inverted”, while the slope on the short end of the curve in September of 1998 was very steep.

Figure 4. Yield curves



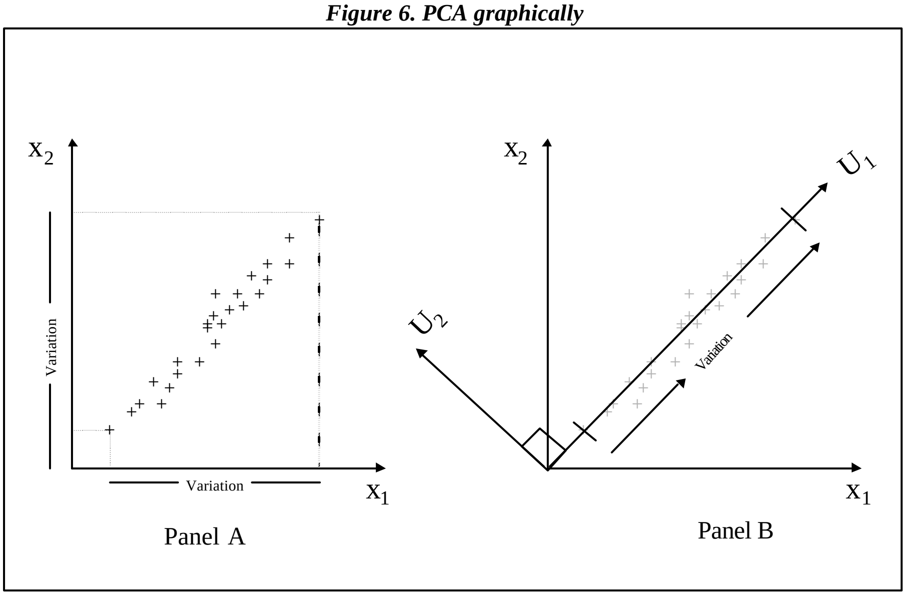
We have 785 yield curves, so it is possible to get the factors explaining the variation of the spot rates. But before doing so, it is important to check the "goodness of fit" of the model. In other words if it is possible to recover the original data.

4. *Principal Components Analysis – Theoretical Framework*

Flury (1988) affirms that PCA "can be looked at from three different points of view". i) a methodology of transforming correlated variables into uncorrelated, ii) a methodology for finding linear combinations of variables with large or small variances and iii) a data reduction technique. Although all of them seem to be a good explanation of what PCA does, it is important to mention that FA shares all these characteristics, so we need a more specific definition.

The intuition behind PCA is very simple and can be explained with the aid of figure 6.

Figure 6. PCA graphically



Panel A shows a sample of a random vector **X** composed of two random variables X_1 and X_2. The perpendicular lines to the axis define their respective variations. A point is uniquely defined by an order pair of the form (x_1,x_2). Panel B, on the other hand, shows a rotation of the axes X_1 and X_2 to a new set of variables U_1 and U_2 forming an orthogonal basis. Notice that under the new basis, the variation in U_1 is larger than under X_1, and conversely the variation of X_2 is larger than that of U_2. Furthermore, U_1 explains almost all the variation of the sample points and if the variance on U_2 is "white noise", then it is redundant and we can get rid of it. The latter explains why PCA is a data reduction technique.

In the case of R^2, it might not be necessary to reduce the number of variables but if instead, a phenomenon is explained by let's say 25 variables or more; then it might be convenient to use PCA or another technique. However, one of the main drawbacks of the methodology is that if for instance the variables X_1 and X_2 are respectively inflation and consumption, there is not straightforward interpretation of U_1.

I will follow Jolliffe (1986) in the definition of PCA because of its simplicity. Let $\mathbf{X}=(X_1, X_2,\ldots, X_p)$ a p-random vector with a known variance-covariance matrix Y and $E(\mathbf{X}) = \mathbf{0}$, where E is the expected value of the random vector **X** and **0** is a zero vector. The objective is to maximise the variance of a linear combination $a_1'\mathbf{X}$ $(a_1 \in R^p)$; however the problem is not well defined yet, because it is possible to make this variance arbitrarily high for an arbitrary a_1. Imposing the normalisation constraint that $a_1'a_1=1$, then we have

$$\begin{aligned} &\underset{a}{Max} \quad Var[a_1'\mathbf{X}] = a_1' Y a_1 \\ &st. \qquad a_1' a_1 = 1. \end{aligned} \tag{14}$$

Jolliffe (1986) solves the problem through a Lagrangean[4]

$$L = a_1' Y a_1 - l_1(1 - a_1' a_1) \tag{15}$$

Taking the first derivative of equation 15 with respect to a_1 and setting it equal to zero, yields

$$\frac{1}{2}\frac{\partial L}{\partial a_1} = Y a_1 - l_1 a_1 = 0 \tag{16}$$

from equation 16, a_1 reveals to be an eigenvector of the variance-covariance matrix Y, while l_1 is its corresponding eigenvalue. Furthermore, from equation 16 it is easy to calculate the variance of $a_1'\mathbf{X}$

$$Y a_1 - l_1 a_1 = 0 \Rightarrow Y a_1 = l_1 a_1 \tag{17}$$

multiplying both sides on the left by a_1'

$$a_1' Y a_1 = a_1' l_1 a_1 \tag{18}$$

which is what we wanted to maximise. By using the normal condition ($a_1' a_1 = 1$), we have that the maximised variance of $\alpha_1' X$ is equal to the characteristic root λ_1.

The same procedure can be repeated i-times with the additional condition that a_i must be orthogonal with the i-1 previous characteristic vectors of Y. For the case of i=2, this is

$$\begin{aligned} \underset{a}{Max} \quad & Var[a_2' \mathbf{X}] = a_2' Y a_2 \\ st. \quad & a_2' a_2 = 1. \\ & a_1' a_2 = 0. \end{aligned} \tag{19}$$

Equation 19 requires at most solving a 2*x*2 system of equations. Still a_2 is a characteristic vector of Y and l_2 is its second largest eigenvalue. Flury presents a generalisation for the case in which there are not multiple roots ($\lambda_1 > \lambda_2 > \ldots > \lambda_p$).

Therefore, we define the i-th component of **X, U$_i$** as

$$\mathbf{u}_i = a_i' \mathbf{X} \tag{20}$$

In general, if the columns of a matrix A are the characteristic vectors a_i of Y, then its components are

$$\mathbf{u} = \begin{bmatrix} U_1 \\ U_2 \\ \\ \\ U_p \end{bmatrix} = \mathbf{A}' \mathbf{X} \tag{21}$$

It is usual to arrange the columns of **A**=($a_1, a_2, \ldots, a_p$), such that a_1 is the characteristic vector corresponding to the largest eigenvalue l_1, a_2 is the eigenvector corresponding to the second largest l_2 and so on so forth. Additionally, in some of the references the authors refer to the characteristic vectors as the components of the covariance matrix. I will use both indiscriminately.

PCA has some important properties that I will just enlist in the next section. Some of them will be useful in subsequent sections.

4.1 Properties of Principal Component Analysis

First is necessary to make some definitions. If Y is symmetric matrix, we define its spectral decomposition as

$$Y = \sum_{j=1}^{p} l_j a_j' a_j = \mathbf{A}' L \mathbf{A} \tag{22}$$

where **A** as defined above, and Λ is diagonal matrix with λ_i entries in its diagonal. We define total variance of a random vector **X** as the trace of Σ, its covariance matrix, such that:

$$\begin{aligned} \sigma^2_{total} &= tr(\Sigma) \\ &= tr(A'\Lambda A) \\ &= tr(\Lambda) \\ &= \sum_{i=1}^{p} \lambda_i = \sum_{i=1}^{p} Var(U_i) \end{aligned} \qquad (23)$$

which is possible because the columns of **A** are normalised and uncorrelated and Λ is diagonal. By the same token, generalised variance can be defined as

$$\sigma^2_{gen} = det(\Sigma) = det(\Lambda) \qquad (24)$$

Total variance and/or generalised variance can be used as rules for selecting a number of components to retain. From the definition of PCA and equations 22, 23 and 24, the following properties of PCA apply to a population as well as a samples. Let Σ the covariance matrix of the p-random vector **X** then

Property 1
In order to maximise the explained variance of a random vector, we have to select the first **q** components of Σ**.**

Property 2
If we wish to predict each random variable X_j in **X,** by a linear function of **Z**, where **Z** is as before, then if σ^2_j is the residual variance in predicting X_j from **Z**, i.e.; the variance not explained by any subset of q components, then $\Sigma\sigma^2_j$ over j=1,..,p is minimised if $\mathbf{B} = \mathbf{A_q}$.

This property means that from the **q** principal components **Z**, we can recover the greatest amount of variation of the original random variables X_j. In other words, the residual variance is minimised by the principal components with respect all the possible linear combinations of **X**.

Property 3
The following property is named as of sphericity, and a sketch of a proof is presented. Let us consider the family of ellipsoids given by

$$\mathbf{X}'\Sigma^{-1}\mathbf{X} = k \qquad (25)$$

where Σ^{-1}, is the inverse of the variance-covariance matrix of **X** and k is an arbitrary constant. As per the definition of principal components we have **Z=A´X,** and because the columns of **A** are orthonormal, then multiplying both sides of the expression by **A**, we have **AZ=X**. Substituting the latter in equation 25 yields

$$(\mathbf{AZ})'\Sigma^{-1}\mathbf{AZ} = k \qquad (26)$$

By using the spectral decomposition of the variance-covariance matrix and knowing that the characteristic vectors of Y^{-1} are the same than those of Y, and that the characteristic roots of Y^{-1} are the reciprocal of those of Y, equation 26 becomes

$$(\mathbf{AZ})'\,\mathbf{A}L^{-1}\mathbf{A}'\,\mathbf{A}\,\mathbf{Z} = k \qquad (27)$$

it is possible to simplify 27 with the orthonormality of **A**,

$$\mathbf{Z}'\,L^{-1}\mathbf{Z} = k \qquad (28)$$

which is the canonical form of ellipses in p-dimensional space with main axis defined by the principal components **Z**. This property will be use to estimate the number of components to retain for the analysis.

Evidently, if all the characteristic roots are equal, then equation 28 represents and sphere in a p-dimensional space. Sphericity imposes several problems for statistical inference, one of them is named as redundancy or multi-collinearity of the data,

4.2 Statistical Inference in Principal Components

Statistical inference in Principal Components has the drawback that it assumes that the random vector **X** is multivariate normal with mean m-vector and covariance given by the matrix Y.

We start with the assumption that $\mathbf{X} \sim N_{\mathbf{p}}(m, Y)$ is a multivariate normal random vector with mean m and covariance matrix Y . Usually, under normality, the unbiased estimator of the mean and variance are:

$$\hat{m} = \bar{\mathbf{X}} = \frac{1}{n}\sum_{j=1}^{n}\mathbf{X}_j$$

$$\hat{Y} = \mathbf{S} = \frac{1}{n-1}\sum_{j=1}^{n}(\mathbf{X}_j - \bar{\mathbf{X}})'(\mathbf{X}_j - \bar{\mathbf{X}}) \qquad (29)$$

Starting from the unbiased estimator the Y, we know that the distribution of **S** is Wishart with n degrees of freedom.

$$\mathbf{S} \sim W_p(n, Y/n)$$

From this distribution, the Likelihood function of the covariance matrix Y is given by:

$$L(Y) = C * (\det Y)^{-n/2} * etr(-\frac{n}{2}Y^{-1}S) \qquad (30)$$

where C is a constant which does not depend on Y, and etr is the exponential function of the trace of a matrix. Maximising 30 is equivalent to minimising the following function g

$$g(Y) = 2\log C - 2\log L(Y) \qquad (31)$$

substituting L(Y) in 31 and simplifying

$$g(\mathrm{Y}) = n(log(det\ \mathrm{Y}) + tr(\mathrm{Y}^{-1}\mathbf{S}) \qquad (32)$$

and by using the spectral decomposition of Y we write

$$g(\mathrm{Y}) = n(log(det\ A'\mathrm{L}A) + tr(\mathrm{L}^{-1}A'\mathbf{S}A) \qquad (33)$$

notice that **A** is orthonormal and that the log of the determinant of L is equal to the sum of the log of each $\mathrm{l}_{\mathbf{i}}$, henceforth we can simplify 33 even more

$$g(\mathrm{Y}) = n\sum_{j=1}^{n}\left(log\ \mathrm{l}_i + \frac{\mathrm{a}_j' ' S \mathrm{a}_j}{\mathrm{l}_j} \right) \qquad (34)$$

Equation 34 is the equivalent expression for the Likelihood function for estimating the parameters of a covariance matrix. A similar version will be derived latter for obtaining the maximum likelihood estimators for the case of Common Principal Components.

Introducing Lagrange multipliers for the constraints $\mathrm{a}_{\mathbf{j}}{}'\mathrm{a}_{\mathbf{j}}=1$ in 34 and differentiating it with respect to $\mathrm{l}_{\mathbf{m}}$ and $\mathrm{a}_{\mathbf{m}}$, yields the following system of equations

$$\begin{aligned} &\forall m = 1,....,p \\ &\hat{\mathrm{l}}_m = \hat{\mathrm{a}}_m{}'\mathbf{S}\hat{\mathrm{a}}_m \\ &\hat{\mathrm{a}}_m{}'\mathbf{S}\hat{\mathrm{a}}_j = 0 \quad j \neq m \\ &\hat{\mathrm{a}}_m{}'\hat{\mathrm{a}}_m = 1 \end{aligned} \qquad (35)$$

and these three conditions hold if and only if[5]

$$\begin{aligned} &\hat{\mathrm{l}}_m = l_m \\ &\hat{\mathrm{a}}_m = a_m \end{aligned} \qquad (36)$$

where $\mathbf{l}_{\mathbf{m}}$ and $\mathbf{a}_{\mathbf{m}}$ are respectively the eigenvalues and eigenvectors of **S**, the variance-covariance matrix estimator of Y. Therefore, equation 36 says that the MLE of the characteristic roots and vectors of Y are the characteristic roots and vectors of **S.**

Flury's book (1988) explain in detail how to get the distribution (mean and variances) of these estimators. Again if **X**~Np(m,Y), **S**~Wp(n, Y/n) and $\lambda_1>\lambda_2>\ldots..>\lambda_p$, then asymptotically

i) the characteristic roots **l,** of **S,** are independent of all its characteristic vectors **a.**
ii) **l**'s and **a**'s are jointly normally distributed.
iii) estimators are unbiased. $E(\mathbf{l}_i)=\mathrm{l}_i$ and $E(\mathbf{a}_i)=\mathrm{a}_i$.

$$\text{iv)}\qquad Cov(l_k, l_m) = \begin{cases} 2\mathrm{l}_k^2 \big/ (n-1) & k = m \\ 0 & k \neq m \end{cases} \qquad (36a)$$

$$\text{v)}\qquad Cov(a_{kj}, a_{mj'}) = \begin{cases} \dfrac{\mathrm{l}_k}{n-1} \sum\limits_{\substack{l=1 \\ l \neq k}} \dfrac{\mathrm{l}_l \, \mathrm{a}_{lj} \, \mathrm{a}_{lj'}}{(\mathrm{l}_l - \mathrm{l}_k)^2} & k = m \\ -\dfrac{\mathrm{l}_k \mathrm{l}_m \mathrm{a}_{kj} \mathrm{a}_{mj'}}{(n-1)(\mathrm{l}_k - \mathrm{l}_m)^2} & k \neq m \end{cases} \qquad (36b)$$

The importance of these properties is that we are able to test sphericity of the last q characteristic roots. Furthermore, we can formally test for the contribution of each and all of the components to the total variance of the sample.

The test for sphericity of the last q components, consists in testing for equality of the last q characteristic roots, H_0: $\lambda_{p-q+1}=\lambda_{p-q+1}=\ldots=\lambda_p$. Using MLE it is possible to derive that the value which maximises the likelihood function under the null H_0 is

$$L(\bar{\mathrm{Y}}) = C * \left[\left(\frac{1}{q} \sum_{i=p-q+1}^{p} l_i \right) \prod_{i=1}^{p-q} l_i \right]^{-n/2} * exp(-\frac{n}{2} p) \qquad (37)$$

On the other hand, by substituting 36 in 30, we have that the value that maximises the likelihood function in the case of p different roots is

$$L(S) = C * \left[\prod_{i=1}^{p} l_i \right]^{-n/2} * exp(-\frac{n}{2} p) \qquad (38)$$

Therefore we can use the result that –2(Log (LY)-Log (S)) is a χ^2 with (q(q+1)/2)-1 degrees of freedom. The likelihood ratio test is then

$$\mathrm{c}^2 = nq \, log \frac{\left[\frac{1}{q} \sum\limits_{i=p-q+1}^{p} l_i \right]}{\left[\prod\limits_{i=p-q+1}^{p} l_i \right]^{1/q}} \qquad (39)$$

The following result is useful for calculating the contribution of a subset of characteristic roots to the total variance of a covariance matrix. Let **w** such that

$$W = \frac{\mathrm{l}_{p-q+1} + \mathrm{l}_{p-q+2} + \Lambda + \mathrm{l}_p}{\mathrm{l}_1 + \mathrm{l}_2 + \mathrm{l}_3 \Lambda + \mathrm{l}_{p-1} + \mathrm{l}_p} \qquad (40)$$

Obviously, **w** is the proportion of the total variance explained by the last q components. By using the distribution of the l **’s** (36a), it is possible to infer

$$Z(W) = \frac{\sqrt{n/2}\left[(1-W)\sum_{j=q+1}^{p} l_j - W\sum_{j=1}^{q} l_j\right]}{\left[W^2\sum_{j=1}^{q} l_j^2 + (1-W)^2\sum_{j=q+1}^{p} l_j^2\right]^{1/2}} \qquad (41)$$

is distributed N(0,1). So if we substitute l, by its corresponding estimators, we can estimate the percentage of variance lost as a consequence of eliminating the last **q** components.

4.3 Comparison between PCA and FA

At the beginning of section 4.1, the intuition behind PCA was described as a rotation of the original basis in order to get a set of components such that the first one explains the larger proportion of the variance of a p-random vector **X.**

On the other hand, factor Analysis starts with the model

$$r_i = \bar{m}_i + l_{i1}F_1 + l_{i2}F_2 + K + l_{ik-1}F_{k-1} + e_i \quad i = 1, K, n \qquad (42)$$

where m is the expected value of the variable $\mathbf{r_t}$, $\mathbf{F_k}$ stands for the kth factor, l's are the factor loadings and e is the specific risk. Equation 42 gives us the first difference between PCA and FA. Whereas in FA we pre-specify the number of factors we require for explaining the variation of **r**, in PCA we calculate as many factors as number of variables and then we decide the number to retain.

The model is not complete until we impose the conditions of a linear model

$i)E(\mathbf{e}) = \mathbf{0}$ $\quad ii)E(\mathbf{F}) = \mathbf{0}$ $\quad iii)E(\mathbf{e'e}) = \mathbf{?}$

$iv)E(\mathbf{Fe'}) = \mathbf{0}$ $\quad v)E(\mathbf{F'F}) = \mathbf{I}$

notice we have changed to matrix notation. The first two are only normalisation of variables whereas the last three are the conditions of normality and orthogonality between errors and factors. If we set **R** equal to the difference between **r** and m and L as a matrix of loadings $l_{\mathbf{ij}}$, then equation 42 in matrix notation is

$$\mathbf{R} = L\mathbf{F} + e \qquad (43)$$

Calculating variance of both sides of 43 yields

$$\mathbf{S} = \mathbf{?'?} + \mathbf{?} \qquad (44)$$

where S is the total variance of **R** and Y is the variance of the error term in equation 43. Equation 44 gives us the second main difference between PCA and FA. While the objective of PCA is to get linear combinations of **A´X** that maximises the diagonal of S**,** FA's objective is the off-diagonal elements. In equation 44, Y is diagonal so the term

$L'L$ accounts for all the variation of the off-diagonal elements of S, but notice that $L'L$ is the variance of $L\mathbf{F}$, which are the variables of interest for FA.

There is another difference. In general, the components from a covariance matrix are very different to those of a correlation matrix and scale measure of the variables seems to be the cause[6]. On the contrary, the factors obtained from FA are invariant of the scale of the data.

Sometimes the first **k** principal components are used as a first approximation for the factor loadings. Researchers used this technique named Principal Components Regression (PCR) in order to eliminate multicollinearity. As it was mentioned before, the principal components are constructed in such a way that they are orthogonal; consequently, if we want to explain the variation of **Y** by using a p-random vector **X**, highly collinear, then we can substitute the original variables for the orthonormal components $\mathbf{U=A'X}$.

However, this procedure has problems; PCA does not say anything about the strength of a relationship among variables but for the diagonal elements Υ of equation 44. Thus it might be that a subset of PC's, completely fail in accounting for the variability of the dependant variable **Y**. A good explanation of some problems using PCR can be found in Hadi and Ling (1998).

On the other hand, FA sometimes uses a set of PC's for rotating the factor loadings in a desired direction. For example, Knez, Litterman, and Scheinkman (1996) followed a similar approach in their paper "Common Factors Affecting Money Market Returns" (1996). Firstly, the authors obtained a set of factor loadings for explaining money market returns using an iterative methodology based upon maximum likelihood estimation. Secondly, they estimate the unknown factors throughout the construction of portfolios and finally they rotate the loadings in order to make the factors consistent with the components of the 30-year zero-curve.

5. Results of the PCA over the yield curve in Mexico

From the yield curves, 15 points were chosen corresponding to interest rates maturing at 1, 7 and 15 days and from 1 to 12 months.

An equal number of components were extracted from the 15 point of the yield curves mentioned above. In addition to the components for the whole period, I also obtained the components of the four periods A, B, C, D mentioned in sections 2.2 and 3.1.

Before we go onto the PCA, I would like to show the result of testing normality and some basic trends in the series.

5.1 Principal Component Analysis for the Whole Period

The correlation matrix of table 4 corroborates what we have seen in figure 1, regarding the relationship among the interest rates, i.e., there is a strong relationship among all the variables. Actually, every rate is highly correlated with the subsequent one. With some exceptions, correlation between rates decreases with maturity, and for instance, the correlation between the one-day interest rate and the seven-day rate is 0.995 while the correlation between one day and 360 days is 0.886. This trend is similar to the one encountered by Barber and Copper for the United States experience.

Table 4 Interest rates correlation matrix for the whole period

Rate	1 day	7 day	15 day	28 day	2 month	91 day	4 month	5 month	182 day	7 month	8 month	9 month	10 month
1 day	1.000												
7 day	0.995	1.000											
15 day	0.979	0.994	1.000										
28 day	0.954	0.978	0.995	1.000									
2 month	0.934	0.961	0.980	0.991	1.000								
91 day	0.923	0.947	0.964	0.977	0.996	1.000							
4 month	0.912	0.936	0.953	0.965	0.989	0.998	1.000						
5 month	0.902	0.925	0.942	0.956	0.982	0.993	0.998	1.000					
182 day	0.890	0.914	0.932	0.945	0.971	0.984	0.993	0.998	1.000				
7 month	0.878	0.902	0.920	0.934	0.960	0.973	0.984	0.993	0.998	1.000			
8 month	0.867	0.891	0.910	0.924	0.949	0.962	0.975	0.987	0.995	0.999	1.000		
9 month	0.860	0.884	0.902	0.916	0.940	0.954	0.968	0.981	0.990	0.996	0.999	1.000	
10 month	0.859	0.882	0.900	0.913	0.936	0.950	0.964	0.976	0.986	0.993	0.997	0.999	1.000
11 month	0.862	0.884	0.901	0.913	0.936	0.950	0.962	0.973	0.981	0.987	0.991	0.995	0.998
360 day	0.866	0.887	0.902	0.913	0.937	0.950	0.960	0.968	0.974	0.978	0.981	0.985	0.991

Although it is not shown here, if we consider the periods A, B, C and D separately, the correlation among the interest rates is not as strong as it is for the whole period. For instance the correlation between the one-day interest rate and the 360-day rate for period B is 0.616 which is lower to the mentioned above.

Table 5 shows the characteristics roots of the covariance matrix for the whole period The first factor accounts in average, for 95.5% of the total variance of the yield curve, the second principal component accounts for 3.35% percent of the variation whereas the third one explains 0.73%. Therefore, the three first principal components explain 99.6% of the variance of the rates. Considering the differences between the economies of Mexico and United States, the data period and series, etc. these results are very different to those found by Knez, Litterman and Scheinkman (1996) mentioned in section 1 in which the first three factors explained 86% of money market returns.

Table 5. Characteristic roots of the covariance matrix

Component	Sample Variance	% explained	% accrued	Rescaled Variance	% explained	% accrued
1	3.364E-02	95.51	95.51	14.317174	95.45	95.45
2	1.182E-03	3.36	98.87	0.513354	3.42	98.87
3	2.565E-04	0.73	99.60	0.106162	0.71	99.58
4	8.905E-05	0.25	99.85	0.035813	0.24	99.82
5	5.229E-05	0.15	100.00	0.027489	0.18	100.00
6	7.178E-09	0.00	100.00	0.000003	0.00	100.00
7	1.242E-09	0.00	100.00	0.000001	0.00	100.00
8	9.446E-10	0.00	100.00	0.000001	0.00	100.00
9	9.237E-10	0.00	100.00	0.000000	0.00	100.00
10	8.928E-10	0.00	100.00	0.000000	0.00	100.00
11	8.431E-10	0.00	100.00	0.000000	0.00	100.00
12	8.141E-10	0.00	100.00	0.000000	0.00	100.00
13	7.966E-10	0.00	100.00	0.000000	0.00	100.00
14	7.618E-10	0.00	100.00	0.000000	0.00	100.00
15	7.281E-10	0.00	100.00	0.000000	0.00	100.00

Yet, it is more similar to Barbel and Copper's paper. In their paper the three first principal components explain 97.15% of the total variation of bond returns; however, the contribution to the variance is very different. In BC paper, the first component accounts for 81%, the second for 12% and the third for more than 4%.

Notice the sample variance values in table 5. They are at most of order two (o(2)). Actually, covariance among variables are at most of the same order, and the covariance matrix is very close to be singular, which is corroborated by the fact that, its determinant is equal to $1x10^{-106}$.

Regarding the components, the first five eigenvectors of the variance-covariance matrix for the whole period are shown in table 6.

Table 6. Five first Eigenvectors - Whole Period

Time to maturity	1	2	3	4	5
0.00274	0.278	(0.465)	0.459	(0.414)	0.107
0.01918	0.282	(0.407)	0.242	(0.057)	(0.039)
0.04110	0.285	(0.337)	0.007	0.289	(0.169)
0.07671	0.289	(0.246)	(0.242)	0.534	(0.214)
0.16667	0.293	(0.098)	(0.399)	0.164	0.132
0.24932	0.289	0.001	(0.375)	(0.183)	0.295
0.33333	0.278	0.080	(0.287)	(0.298)	0.192
0.41667	0.265	0.142	(0.179)	(0.287)	(0.000)
0.49863	0.253	0.191	(0.072)	(0.223)	(0.179)
0.58333	0.241	0.229	0.031	(0.132)	(0.304)
0.66667	0.230	0.254	0.119	(0.035)	(0.338)
0.75000	0.222	0.266	0.190	0.062	(0.268)
0.83333	0.216	0.265	0.240	0.151	(0.076)
0.91667	0.213	0.249	0.266	0.227	0.247
0.98630	0.213	0.226	0.269	0.279	0.627

By using the components as coefficients, we can illustrate the interpretation of the characteristic vectors in the following way. As per the orthogonality of the matrix of components, equation 21 can be reversed in order to express the interest rates as a linear combination of the scores U.

$$\mathbf{U} = \mathbf{A}'\mathbf{X} \quad \Rightarrow \mathbf{AU} = \mathbf{AA'X} \Rightarrow \mathbf{AU} = \mathbf{X} \qquad \text{(21a)}$$

According to 21a, the equation explaining changes (see first row in table 6) of the short rate is

$$r_{0.00274,t} = 0.278U_{1,t} - 0.465U_{2,t} + 0.459U_{3,t} - 0.414U_{4,t} + 0.107_{5,t} + e_t \qquad \text{(21b)}$$

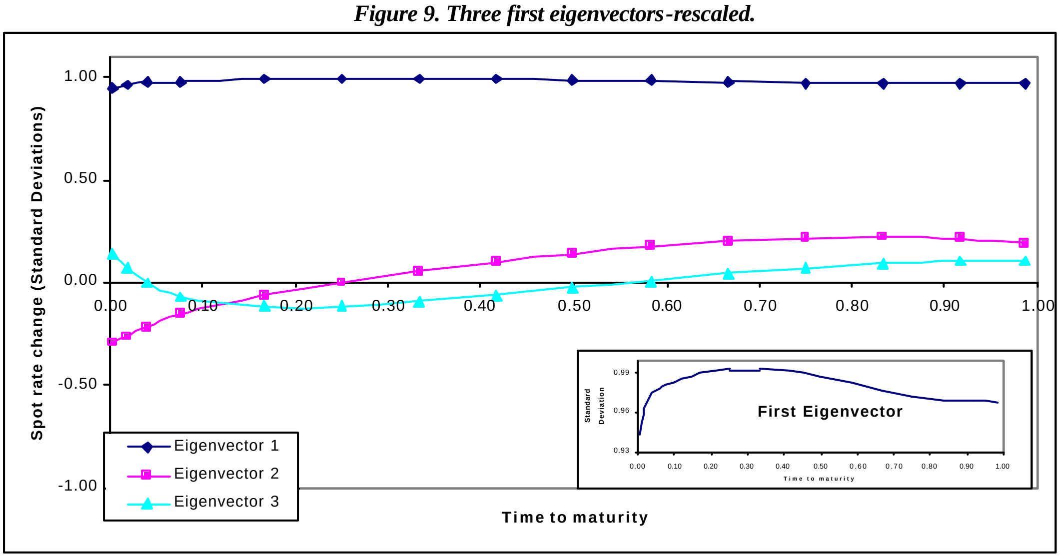
Thus, a unit increase in the component $\mathbf{U_1}$ causes the interest rate to move up 0.278 units or 27.8 basis points; a unit increase in the component $\mathbf{U_2}$ causes the rate to move down 0.465 units or 46.5 basis points and so on. This is not the whole truth, the coefficients of the components $\mathbf{U_j}$ must be rescaled by the variance explained for the respective characteristic root $l_{\mathbf{j}}$. Thus by re-scaling the coefficients, we would have that the correct equation would be

$$r_{0.00274,t} = 0.943U_{1,t} - 0.295U_{2,t} + 0.135U_{3,t} - .071U_{4,t} + 0.014U_{5,t} + e_t \qquad \text{(21c)}$$

According to table 6, it is clear that the first component causes a parallel shift of the whole yield curve. An increase in the second component makes the rates on the short-end of the curve to decrease whereas the rates on the long-end increase, consequently, the whole curve becomes steeper. The third component moves the rates such that the yield curve becomes convex. The fourth one twists the curve, and so on. These are the "level", "slope", "curvature" and so on, classification mentioned the introduction.

Figure 9 plots the three first rescaled[7] factors. It illustrates graphically what I said in the previous paragraph. For every standard deviation change in the curve, roughly 95% is a parallel shift (eigenvector 1) and the remaining change is explained by changes in the slope and curvature. A similar graphic can be found in Barbel and Copper.

Figure 9. Three first eigenvectors-rescaled.

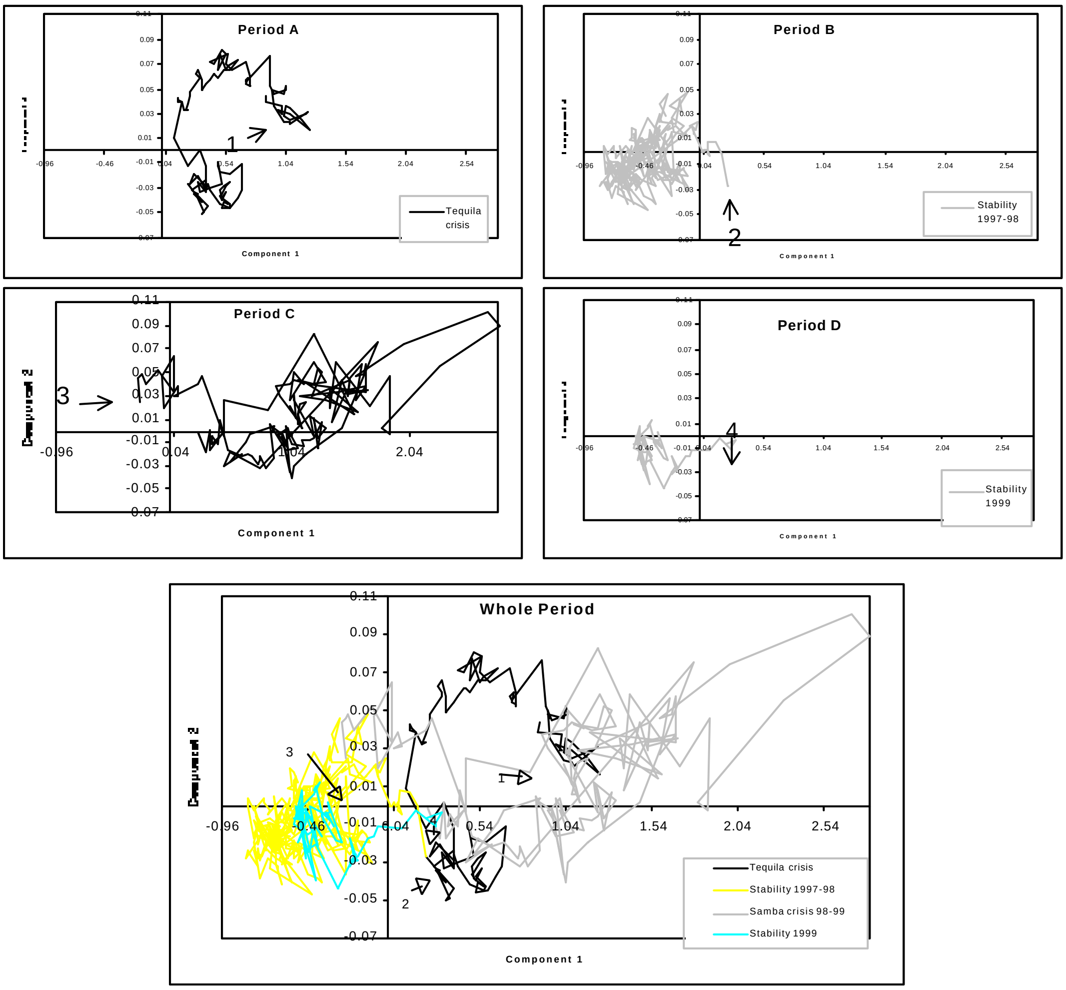


The inner graphic shows that the first eigenvector is not as flat as it seems and though component 1 shifts the curve, it does it in different proportions. This shape is similar to the findings of Knez, Litterman and Scheinkman (1996).

Why are these eigenvectors so important? Because knowing them allow us to immunise a portfolio against changes in the yield curve in different directions. I will not go in-depth to explain how the eigenvectors or directional vectors can be used to immunise a portfolio in several directions, the papers by Reitano and Barbel and Copper explain step by step and including examples how to do it. Instead, I rather ask, what would be the effect in a portfolio if the directional vectors change constantly through time? If it is the case, it would be interesting to estimate the potential earnings/loses derived from a change in the characteristic vectors or even from changes in characteristic roots.

Picasso could have signed figure 10 and probably he would named it as “A Turbulent History” or similar. Instead, figure 10 plots the original observations but using a new basis defined by the components $\mathbf{U_1}$ and $\mathbf{U_2}$, the 2 first principal components.

Figure 10. A Turbulent History



We can repeat the story of section 2, but getting more insight of the overall situation. The figure on the top of the left, shows how the Mexican economy was recovering from the "Tequila crisis of 1995" (period A). At point 1, we are in June of 1996. Remember this was a time of volatility, this can be inferred from the variation in both directions. Below, we will see that in this period the variance explained by the second component is larger than 30%.
At point 2 on the top-right, we are in 1997 (period B) and the fluctuations are less significant than in the previous paragraph. Now, it seems that component one explains a larger proportion of the total variation of the yield curve. When uncertainty decreases, component 2 become less important.

At point 3 in the bottom-left, the fluctuations start becoming larger and larger but notice that initially it happens only in the direction of the second component. Suddenly fluctuations in the direction of the first factor explode. Although component 2 reacted first, component 1 moves wilder. Again, component 1 regains its influence on the total variation of the interest rates. This is the period of the Samba crisis (period C). Finally, at point 4 (period D), we are in the second quarter of 1999 and the calm comes back.

Speculating with these ideas, the second component could be the "market expectations" factor. It starts fluctuating when "rumours" of a crisis fulfils the environment as previously to Brazilian crisis. Also, it goes back to stability more gradually as it was the case in the recovering period of the Tequila crisis.

On the other hand, component 1 "accommodates" itself according to expectations. It reacts later and it might be named as the "stabilisation or governmental instrument" for lessening crises.

Whereas component 1 could be useful for investigating money supply equations, component 2 might be helpful for testing rational expectations. It would be interesting to look at the plot between factors 1 and 3 and between factors 2 and 3, but for the sake of space, this is not done here.

Notice that by making this analysis we implicitly assumed that the components remain the same for the whole period but the total variance explained for each factor significantly changes through time. Is it possible that the directional vectors remain "constant"? Let us find out.

5.2 Principal Component Analysis for sub-periods

This section will go through the Principal Component Analysis for periods. but in addition, it will present the results of tests for selecting the number of components to retain.

According to table 7, the volatility in periods A and C is higher than in B and C. The contribution of the components to this volatility importantly changes. Looking at period

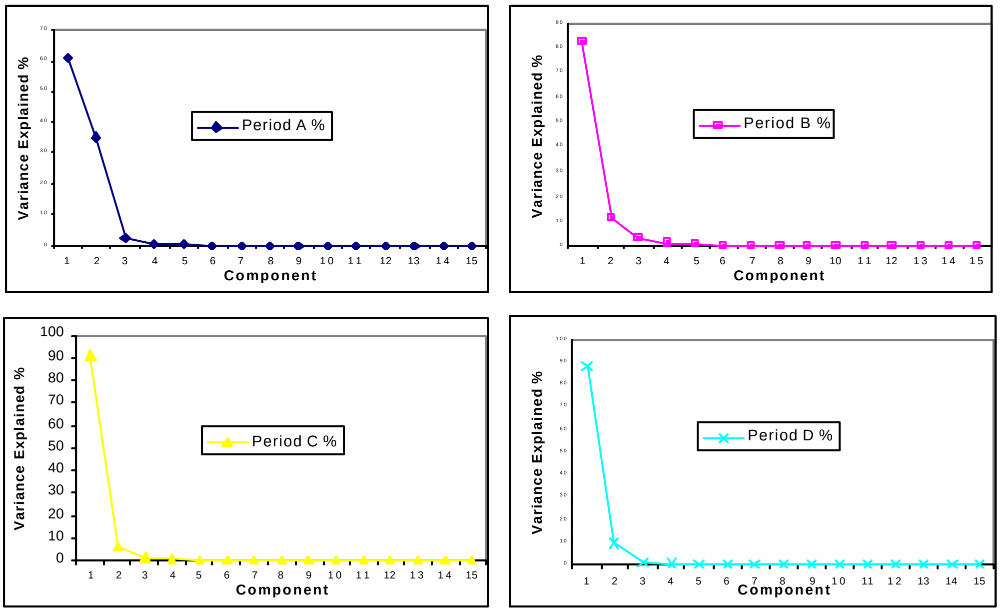
A, the first component accounts for 61% and the second for a 35% of the total variance, while in period C, the proportions are 91% and 6.5% respectively. It seems period D is very similar to C and it might be that some sequels of the Brazilian crisis remain.

Table 7. Variance of each period - Eigenvalues

Period A		Period B		Period C		Period D	
Variance	%	Variance	%	Variance	%	Variance	%
5.82E-03	60.89	3.32E-03	82.73	2.20E-02	91.40	3.22E-03	87.84
3.36E-03	35.16	4.78E-04	11.90	1.57E-03	6.52	3.52E-04	9.60
2.85E-04	2.98	1.29E-04	3.20	3.29E-04	1.37	5.67E-05	1.54
5.85E-05	0.61	5.42E-05	1.35	1.45E-04	0.60	3.05E-05	0.83
3.40E-05	0.36	3.30E-05	0.82	2.68E-05	0.11	7.00E-06	0.19
3.39E-09	0.00	1.54E-09	0.00	4.64E-09	0.00	1.34E-09	0.00
1.28E-09	0.00	9.40E-10	0.00	1.45E-09	0.00	1.09E-09	0.00
1.11E-09	0.00	9.31E-10	0.00	1.09E-09	0.00	9.81E-10	0.00
1.01E-09	0.00	8.63E-10	0.00	1.03E-09	0.00	9.60E-10	0.00
9.80E-10	0.00	8.55E-10	0.00	9.07E-10	0.00	7.80E-10	0.00
9.04E-10	0.00	8.06E-10	0.00	8.50E-10	0.00	6.51E-10	0.00
7.41E-10	0.00	7.76E-10	0.00	7.09E-10	0.00	5.59E-10	0.00
6.04E-10	0.00	7.29E-10	0.00	6.57E-10	0.00	5.11E-10	0.00
5.72E-10	0.00	7.00E-10	0.00	5.76E-10	0.00	4.12E-10	0.00
4.96E-10	0.00	6.45E-10	0.00	4.97E-10	0.00	3.55E-10	0.00

Notice the values in table 7 are consistent with what we have seen in figure 10. In period A volatility on the direction of the second component is larger, while in period C, the influence of the component 1 increases substantially. Other way of looking at the same information is through the Scree.

Figure 11. The Scree – Different periods



Despite the marginal contribution of the third component still plays a role in each period. Its contribution is very similar in periods A and B, i.e., of the order of 3%. Although in period C its contribution was reduced to 1.37%, at period D it starts to become more influential.

How many components should we select? From table 7 and figure 11, it seems we do not need more than three components to explain an important share of the total variance of the yield curve. What does it mean that the others are close to zero? What are consequences of sphericity in the last 12 eigenvalues?

Jolliffe (1986) talks about four different ways of selecting components: a) cumulative percentage of total variation, b) size of variances of principal components, c) the Scree graph, d) the number of components with unequal eigenvalues.

Whereas rules a, b and c are ad-hoc rules base on the definition of PC, rule d seems to be more statistical. For instance, we could have decided "a-priori" that we would use a number of components accounting for more than 95%, i.e., rules a and c. Under these rules we would have retained components 1 and 2. on the other hand, if the decision rule were to use all the components with a contribution larger or equal to one percent, i.e., rule b, then three components would be selected.

We could have tried to test if the eigenvalues of a covariance matrix are equal to zero. Let's try. From the distribution of eigenvalues and eigenvectors of section 4.1.2, we know that the eigenvalues of the covariance matrix are multivariate normal with the following parameters

$$E(l_m)=\lambda_m,$$

$$Cov(l_k,l_m)=\begin{cases} 2l_k^2/(n-1) & k=m \\ 0 & k\neq m \end{cases} \qquad (36a)$$

Under the null ($H_0.\ \lambda_m=0$) , we want to test that the m-th eigenvalue is equal to zero against that it is different. We set the statistic

$$z=\frac{\hat{l}-0}{\sqrt{2\hat{l}^2/(n-1)}}=\sqrt{\frac{(n-1)}{2}}\frac{\hat{l}}{\hat{l}}=\sqrt{\frac{(n-1)}{2}} \qquad (45)$$

Unfortunately, the statistic 45 will never "accept" H_0 (unless n<4) because an implicit assumption for calculating the distribution is that the eigenvalues of the covariance matrix are all different to zero; thus we can not perform this kind of tests.

Rule d is often named as the Bartlett test for sphericity. Sphericity imposes several problems to PCA, firstly, although the space defined for the components is orthogonal it is not unique, secondly, multicollinearity or data redundancy and thirdly, sphericity makes statistical inference very complicated because as I mentioned before statistics are derived from the assumption that the eigenvalues are all different.

Bartlett's test aim is to identify a subset of components with low or none contribution to the explained variance in order to eliminate them. I applied the test for sphericity to the last components using the statistic given by equation 39

$$\chi^2 = nq\log \frac{\left[\frac{1}{q}\sum_{i=p-q+1}^{p} l_i\right]}{\left[\prod_{i=p-q+1}^{p} l_i\right]^{1/q}} \tag{39}$$

Table 8. Bartlett's sphericity test results.

	Whole period	*A*	*B*	*C*	*D*
q	*8*	*9*	*9*	*8*	*9*
*χ-**squared***	*23.50*	*56.10*	*27.13*	*45.01*	*54.38*
df	*35*	*44*	*44*	*35*	*44*
p-value	***0.93070***	***0.10437***	***0.97856***	***0.11977***	***0.13582***

According to table 8, at 10% significance level, we can not reject sphericity in the last 9 components for periods A, B, and D and sphericity in the last 8 for periods C and the pooled data too. Under sphericity test we would have to retain six or seven components in order to explain 100% of the variation of the original series. However, the problem with the Bartlett test is that it tends to retain more PCs than are necessary. Thus, we need to look for another way out. We define the rule that we will retain the first three components if the statistical contribution of the last 12 components is less than 5%. According to equation 41,

$$Z(W) = \frac{\sqrt{n/2}\left[(1-W)\sum_{j=q+1}^{p}\lambda_j - W\sum_{j=1}^{q}\lambda_j\right]}{\left[W^2\sum_{j=1}^{q}\lambda_j^2 + (1-W)^2\sum_{j=q+1}^{p}\lambda_j^2\right]^{1/2}} \tag{41}$$

where as we saw, **w** is the combined contribution of the last q eigenvalues. The null hypothesis is that the last q components account for at least 5% of the variation of the series. Table 9 shows the result for a w equal 5%.

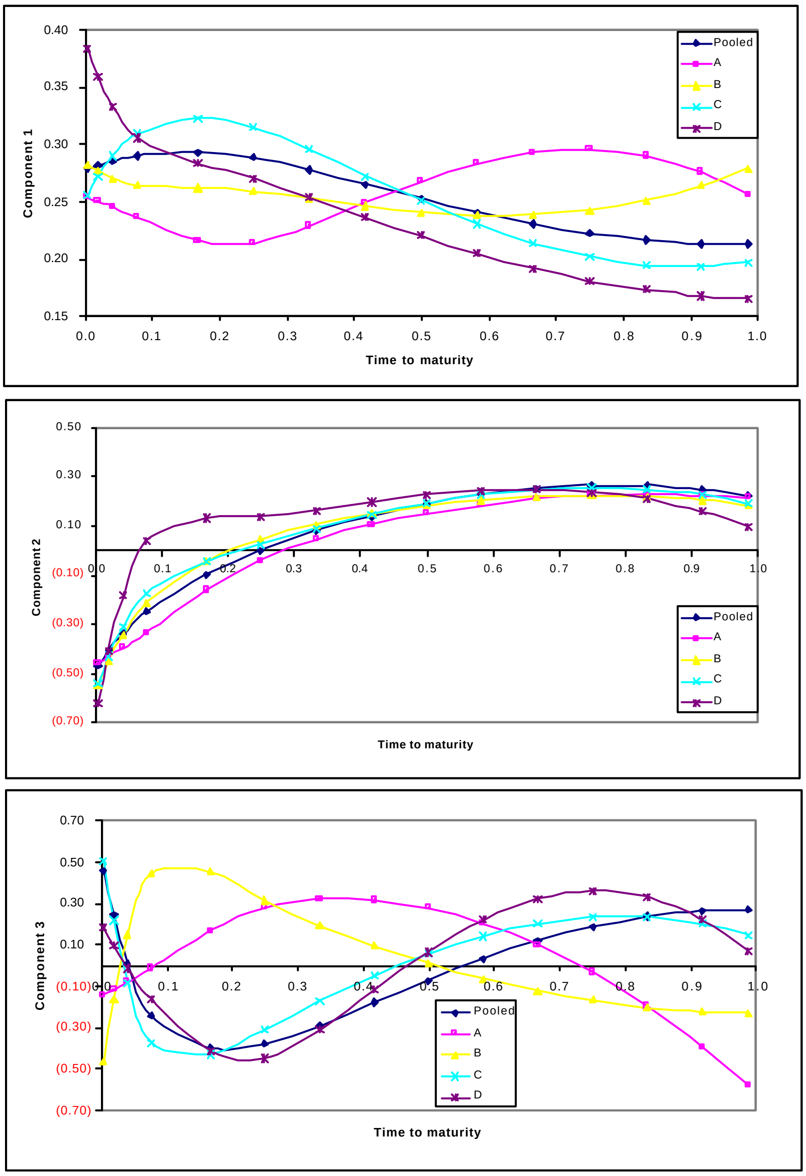
Table 9. Lost Contribution of the total variance (w) because of eliminating the last 12 components

	Whole period	A	B	C	D
q	12	12	12	12	12
w	5%	5%	5%	5%	5%
Z-statistic	**6.09**	**2.87**	**2.54**	**2.53**	**1.82**
p-value	**0.00000**	**0.00205**	**0.00560**	**0.00567**	**0.03438**

At the 5% significance level, we are sure we do not lose more than 5% of the total variance by just selecting the first three components. All these tests have a lot of puzzles, which Jolliffe (1986) explains in his book in chapter 6; nevertheless, I just wanted to illustrate some possible "formal" tests for retaining an appropriate number of components.

Now let's turn up to the components. Figure 12 contrasts the first three components for the four periods and the pooled data.

Figure 12. Comparison of the components



Components 1 and 3 seem to behave differently in each period whereas component 2 is more stable. Regarding component 1, the behaviour of the short-end wildly changes from one period to another. Mention apart deserves period A, which is totally different to the others. The interpretation of the level component still applies since changes in this component provoke quasi-parallel shifts in the whole curve. Notice that regardless of the period, component one seems to keep inside a target zone which in average goes from 0.20 to 0.30; is there any relationship with mean reversion?

Regarding component 2, notice the similarity in all periods. The interpretation of this component is the same that above since given a positive shock to this factor, the short-end of the yield curve decreases whereas the long-end grows, i.e., the yield curve becomes steeper.

Finally, component 3 seems more difficult to interpret. It changes the curvature of the yield curve, there is no doubt; nevertheless it does it in very different fashions. For instance in periods A and B, a positive change in the component causes the short-end and the long-end of the yield curve to go down whereas the middle rates go up. On the other hand, in periods C and D, excepting for the very short rate, whereas rates in the short-end decrease, rates in the middle and long-end increase. Again, component 3 in period A behaves in a completely different fashion that in the other subperiods, whereas period B is different to C and D only in sign.

It seems that only component 2 could be common to all periods. In addition, we have seen that the interpretation of the components is the same that the one proposed by Litterman and Scheinkman (1988); nevertheless, that does not mean that they do not change through time but, how do we test for changes in characteristic vectors and roots? Common Principal Component Analysis (CPCA) answers the question, however as we will see, it is a methodology computationally intensive; hence, I explore other 2 possible alternatives. First, let us go through the intuition and theory of CPCA.

6. Common Principal Component Analysis (CPCA)

Let Y_1 and Y_2 two variance-covariance matrices. By the spectral decomposition of a covariance matrix presented in section 4.1.1, we can represent both matrices as the product of their characteristic vectors and roots

$$\mathrm{Y}_i = \sum_{j=1}^{p} \mathrm{l}_{ji}\, \mathrm{a}_{ji}'\, \mathrm{a}_{ji} = \mathbf{A}_i'\, \mathrm{L}_i \mathbf{A}_i \quad i = 1,2 \qquad (22a)$$

Let us contrast 22a by proposing the following model.

$$\mathrm{Y}_i = \sum_{j=1}^{p} \mathrm{l}_{ji}\, \mathrm{a}_{j}'\, \mathrm{a}_{j} = \mathbf{A}'\, \mathrm{L}_i \mathbf{A} \quad i = 1,2 \qquad (48)$$

Model 48 contrasts with 22a in that it allows the characteristic roots to be different, whereas the characteristic vectors have to be equal.

In addition to 48, there are other 3 types of relationship between covariance matrices. I will just mention all of them. The first relation among covariance matrices is equality of all $\mathrm{Y}_{i.}$. It does not require more explanation that all the elements of both matrices are equal.

The second level is that of proportionality and can be represented as follows

$$\mathrm{Y}_1 = \mathrm{r}_i \mathrm{Y}_i \quad i = 1,2,\mathrm{K}\,,k \qquad (49)$$

which obviously means that the all the elements of a matrix are proportional to the elements of the others.

The third level is CPCA, which has been already mentioned and finally we have Partial Common Principal Component Analysis (PCPCA. By using the spectral decomposition of a covariance matrix PCPCA is

$$\Sigma_i = \sum_{j=1}^{q} \lambda_{ji}\,\alpha_j'\,\alpha_j + \sum_{j=q+1}^{p} \lambda_{ji}\,\alpha_{ji}'\,\alpha_{ji} \qquad i = 1,2,\ldots,k \qquad (50)$$

In equation 50 the first **q** principal components are common and the others **p-q** are specific to the **k** matrices.

6.1 Maximum Likelihood Estimation of CPCA

The model is

$$H_{cpc}\!: \Sigma_i = \mathbf{A}'\Lambda_i\mathbf{A} \qquad i = 1,2,\ldots k \qquad (48a)$$

if we multiply both sides of 48b by **A** to the left and **A´** to the right, in order to take advantage of the orthonormality of the pxp matrix of components, we have

$$\mathbf{A}\Sigma_i\mathbf{A}' = \Lambda_i \qquad i = 1,2,\ldots k \qquad (48b)$$

where Λ_i are diagonal matrices which elements are the specific eigenvalues. The intuition behind 48b is that if we want to test CPC, we need to find a square matrix **A**, which simultaneously diagonalises k symmetric matrices.

For testing 48 as, as in section 4.1.2, we start with $\mathbf{S}_i$, the unbiased estimator of Σ_i,

$$\mathbf{S}_i \sim W_p(n_i, \Sigma / n_i)$$

where $\mathbf{n_i}$ is the size of the i-th sample. Assuming independence of the covariance matrices we form the likelihood function as

$$L(\Sigma_1,\ldots,\Sigma_k) = C * \prod_{i=1}^{k}(det\,\Sigma_i)^{-n_i/2} * etr(-\frac{n_i}{2}\Sigma_i^{-1}S_i) \qquad (51)$$

Equation 51 is the multivariate version of equation 30. Again, the maximisation of the likelihood function can be transformed to a minimisation problem ending with the following expression

$$g(\mathbf{A},\Lambda_1,\ldots,\Lambda_k) = \sum_{i=1}^{k} n_i \sum_{j=1}^{n}\left(log\,\lambda_{ij} + \frac{\alpha_{ij}'\,'S\alpha_{ij}}{\lambda_{ij}}\right) \qquad (52)$$

which is equivalent to equation 34. By solving 52 we get the following system of equations

$$\hat{\lambda}_{im} = \hat{\alpha}_m'\,\mathbf{S}_i\,\hat{\alpha}_m \qquad i = 1,\ldots k \quad m = 1,\ldots,p$$

$$\hat{\alpha}_m'\sum_{i=1}^{k}\left(n_i\,\frac{\lambda_{im}-\lambda_{ij}}{\lambda_{im}\lambda_{ij}}\mathbf{S}_i\right)\hat{\alpha}_j = 0 \quad j \neq m \quad m,j = 1,\ldots p \qquad (53)$$

$$\hat{\alpha}_m' \hat{\alpha}_j = \begin{cases} 1 & m = j \\ 0 & m \neq j \end{cases}$$

53 is the k-dimensional equivalent system to 35; however the solution of this system is not straightforward and Flury (1988) develops in his book an algorithm for solving it. The algorithm is denominated FG in honour to Flury and Gautschi who proposed it.

Flury points out that a unique solution always exists thanks to some technical optimisation conditions. Denoting $\hat{A}$ and $\hat{\Lambda}_i$ as the solution to 53 then we get the maximum likelihood estimator of the covariance matrices

$$\hat{\Sigma}_i = \hat{\mathbf{A}}' \hat{\Lambda}_i \hat{\mathbf{A}} \qquad i = 1, 2, \ldots k$$

Substituting in 51, the value that maximises the likelihood function is

$$L(\hat{\Sigma}_1, \ldots, \hat{\Sigma}_k) = C * \prod_{i=1}^{k} (det\, \hat{\Sigma}_i)^{-n_i/2} * exp(-\frac{pn_i}{2}) \qquad (54)$$

This is the restricted model, on the other hand, by substituting the unbiased estimator of Σ (**S**) in 54 gives the unrestricted model. Thus, the log-likelihood ratio for testing $H_{cpc.}$ is

$$\chi^2_{cpc} = -2\, ln \frac{L(\hat{\Sigma}_1, \ldots, \hat{\Sigma}_k)}{L(\mathbf{S}_1, \ldots, \mathbf{S}_k)} = \sum_{i=1}^{k} n_i\, ln \frac{det\, \hat{\Sigma}_{i,}}{S_i} \qquad (55)$$

which is distributed as χ^2 with (k-1)p(p-1)/2 degrees of freedom.

I will close this section by saying that there are explicit solutions for proportionality and PCPCA by using the MLE methodology. It is important to mention that I could not find an exact algorithm to calculate the estimators in the case of PCPCA; moreover, comparing a subset of components and allowing the others to be specific, just have the effect of complicating the calculations.

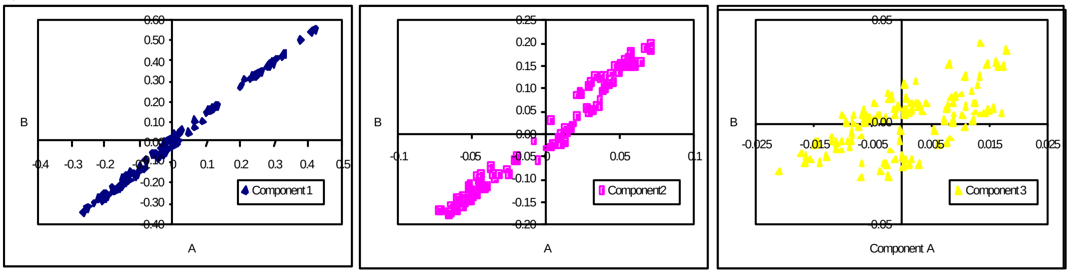
6.2 Cross-Approach for Illustrating CPCA

Denote $\mathbf{A}_A^3$, $\mathbf{A}_B^3$, $\mathbf{A}_C^3$, and $\mathbf{A}_D^3$ as the first three eigenvectors of the covariance matrices in each period. In addition, let $\mathbf{X}_A$, $\mathbf{X}_B$, $\mathbf{X}_C$, and $\mathbf{X}_D$ the sample observations where the subscripts represent the periods A, B, C and D. We say that the components of two periods are equal if and only if

$$\mathbf{A}_B^3{}' \mathbf{X}_A^3 = \mathbf{A}_A^3{}' \mathbf{X}_A^3 \qquad (56)$$

equation 56 simply means that if we applied the components of one period to the data in other, then the components are common if and only if the cross-plots are lines with a slope equal to one. Notice that this is not a formal but a graphical way of checking if a subset of components is common for two periods. In our case, we have 6 different combinations of periods and if in addition we want to compare 3 components, thus we have in total 18 graphs. I will briefly comment over the relationships.

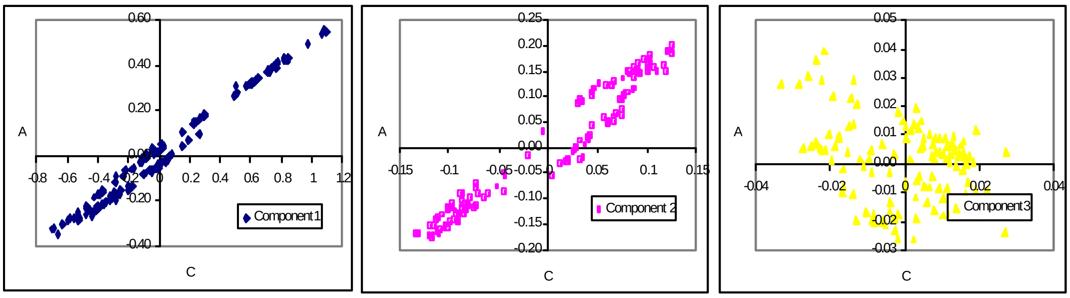
a) Periods A and B



The slope of the first component is close to 1, suggesting that this component is equal for periods A and B.

On the other hand, though there is a strong linear relationship in components 2 and a slight one in component 3, the value of the slope does not allow us to accept the hypothesis that they are common for both periods.

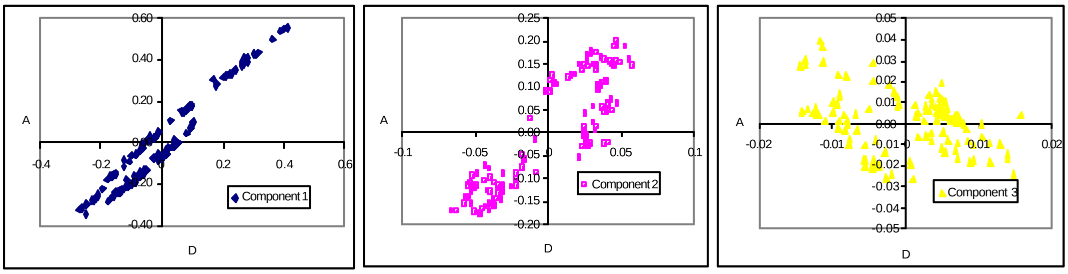
b) Periods A and C



Firstly, the linear relationships is not as strong as above; however it seems in the slope of components 1 and 2 is close to 1 and thus it does not reject the hypothesis that both are common for periods A and C.

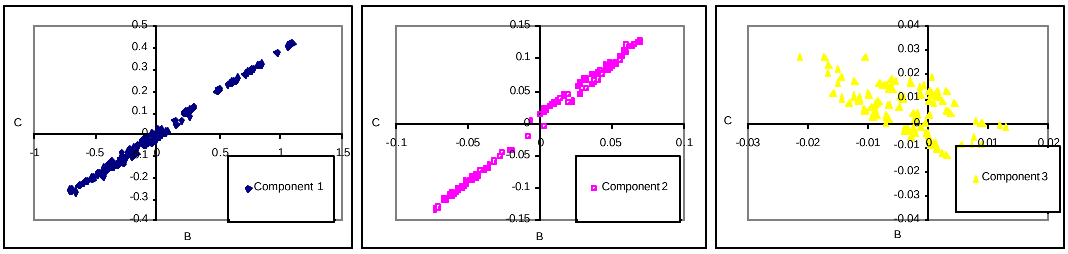
There is a weaker and negative linear relationship in the third component, which might suggest that this vector alters its direction from one period to other. This is consistent with the relationship of the third component between periods A and C shown in figure 12.

c) Periods A and D



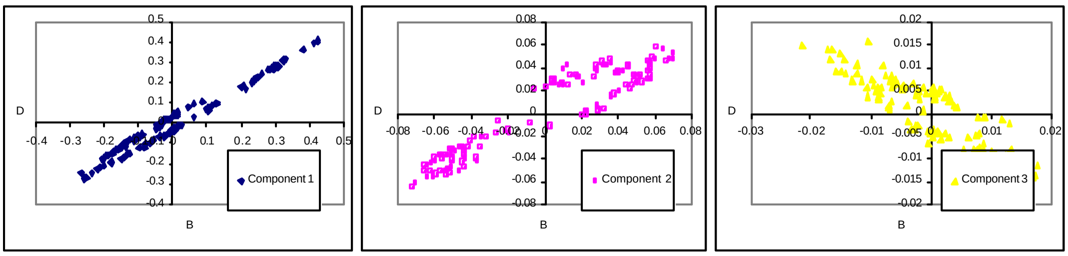
The slope of component one is close to 1 and thus this could suggest that the first component is the same in periods A and D. Graphically, it is clear that we would reject that the components 2 and 3 are common given their respective slopes. Again the relationship of component 3 is negative, meaning that this vector changes direction from period A to D. Notice that the cross plot of the first component shows what seems to be a pair of parallel straight lines and the "hole" close to the origin in the cross-plot of component 2, I will mentioned my interpretation below.

d) Periods B and C



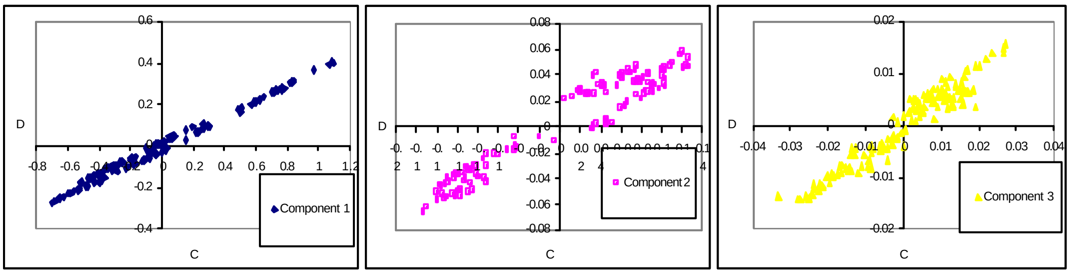
It seems none of them are common. The slope of the straight line of component 1 is close to 3 or more, whereas in the case of component 2 it is close to 2. Component 3 still shows a negative relationship.

e) Periods B and D



Again, figures 1 and 2 have slopes that could mean that the components are common. Component 3 has a negative slope suggesting that the vectors have opposite directions.

f) Periods C and D

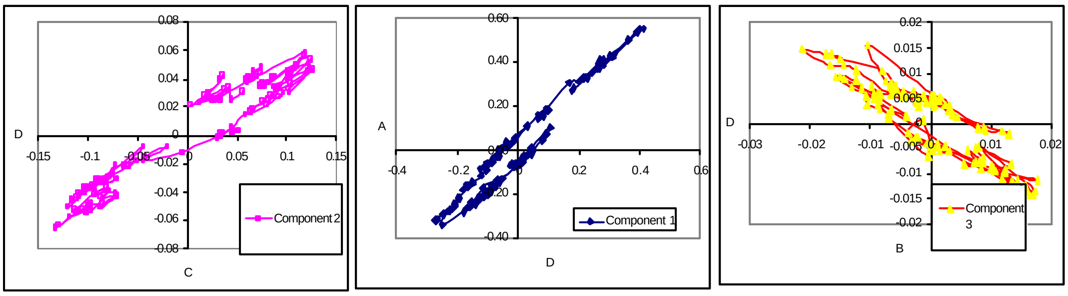


It seems that all the components could be common in this case, although the slope in the plot of component 3 is not exactly equal to 1.

From the graphs, we can infer that components change through time; however; it must be mentioned that components 1 and 2 show more stability than component 3. In almost all the cases, there is a strong linear relationship between components in different periods. OLS could be helpful in determining the strength of these linear relationships and estimating the slopes.

The “holes” and “parallel” lines, which appear in some of the figures, intrigued me. In order to interpret, I decided to joint the points of the plot looking for a pattern.

Figure 14. “Some Holes in the Common Principal Component Analysis”



As it can be seen, the lines describe the path followed by the transformation of interest rates on time. As it can be seen, lines do not behave erratically or randomly but they present an ordered pattern. Holes and parallel lines could be signs of structural breaks, meaning that we could break the periods A, B, C and D in subsequent sub-periods.

6.3 Krzanowski’s Approach for Common Principal Components

Krzanowski’s idea is to find the minimum angle between the axis of the subspaces spanned by the eigenvectors of two (or more) covariance matrices. For instance, if we have two vectors $\mathbf{a=(a_1,a_2)}$ and $\mathbf{b=(b_1,b_2)}$ in R^2 and the norm for each one of them is set equal to one, then we know that the inner product between the vectors **a** and **b** is equal to the cosine of the angle between them. Mathematically,

$$\frac{a \bullet b}{\|a\|\|b\|} = \cos\theta \qquad (57)$$

The decision rule is clear, if the angle q between **a** and **b** is close to zero then the vectors are common. The result of calculating the angle between components for each pair of periods is shown in 10.

Table 10. Angles between characteristic vectors-Comparison between periods

	AB	*AC*	*AD*	*BC*	*BD*	*CD*
			Inner product			
First component	0.9910	0.9644	0.9426	0.9851	0.9704	0.9815
Second component	0.9725	0.9653	0.8081	0.9959	0.9209	0.9311
Third Component	0.8500	0.8313	0.7552	0.8376	0.7625	0.7856
			q ***in radians***			
First component	0.13	0.27	0.34	0.17	0.24	0.19
Second component	0.24	0.26	0.63	0.09	0.40	0.37
Third Component	0.55	0.59	0.71	0.58	0.70	0.67
			q ***in degrees***			
First component	**8**	**15**	20	**10**	**14**	**11**
Second component	**13**	**15**	36	**5**	23	21
Third Component	32	34	41	33	40	38

If we define a tolerance of 15 degrees in the angle between components, the conclusions are very similar to those in the previous section. The first component is common for periods AB, AC, BC, BD, CD. In addition, periods AB, AC and BC share component 2. Regarding component 3, the angles suggest not only that they are not common but also that some of they are very close to be orthogonal.

This methodology has a number of drawbacks. How do we determine that q is close enough to zero? Instead of 15 degrees, we could have arbitrarily set 10 or 5 and the outcome would have been totally different. Jolliffe mentions that though Krzanowski simulated some values these are not statistically acceptable. Notice that in table 11 we are measuring the angle between the projection of two p-dimensional vectors in R^2; therefore we can expect that the minimum angle between two p-dimension vectors is smaller than that in table 10.

For testing in q-dimensions, Krzanowski proposes a similar. Now the idea is to find the minimum angle δ, between the subspaces spanned by q-principal components. He found that this angle is given by

$$d = cos^{-1}(l_1^{1/2}) \qquad (58)$$

where λ_1 is the largest eigenvalue of the following product

$$\mathbf{A}_q^{j}{}' \mathbf{A}_q^{k} \mathbf{A}_q^{k}{}' \mathbf{A}_q^{j} \qquad j,k = A,B,C,D \quad j \neq k \qquad (59)$$

where $\mathbf{A_q}^j$ are matrices composed of the q eigenvectors of interest in j-th period. In this case, if the angle is “small” then we can say that the q-components are common for both covariance matrices.

I performed the analysis for simultaneously test that the three first principal components are common for every pair of sub-periods. In any case, I could not reject common principal components in each case. However, I found that these test must be performed for a subset of components less than the number of variables (q<p).

From the cross-approach, we already know that the third component changes through time, therefore I performed the analysis for the remaining two significant components, i.e. 1 and 2. The results can be seen in table 11.

Table 11. Common principal components -minimum angle between characteristic vectors in degrees

	AB	AC	AD	BC	BD	CD
Minimum angle in degrees	5	8	6	4	5	3

If we consider 5 degrees as the decision rule, surprisingly the result is now that periods AB, BC, BD and finally CD share the first two components simultaneously, contrasting with the results of table 10 in which only AB and AC share the two first factors. This analysis corroborates that components 1 and 2 do not change too much.

6.4 Maximum Likelihood Estimation of common principal components

CPCA and PCPCA are undoubtedly the most formal way of testing common principal components, however they are the most difficult as well.

As we mentioned in section 4.3

$$H_{cpc}: \Psi_i = \mathbf{A}' \Lambda_i \mathbf{A} \qquad i = 1,2,\ldots k \qquad (48a)$$

is the expression for testing CPCA hypothesis, against the alternative of arbitrary matrices. As we have seen in section 6.1, the system of equations for the maximum likelihood estimation of the covariance matrix is given by

$$\begin{aligned} &\hat{\lambda}_{im} = \hat{\alpha}_m' \mathbf{S}_i \hat{\alpha}_m \qquad i = 1,\ldots k \quad m = 1,\ldots ,p \\ &\hat{\alpha}_m' \sum_{i=1}^{k} \left(n_i \frac{\lambda_{im} - \lambda_{ij}}{\lambda_{im}\lambda_{ij}} \mathbf{S}_i \right) \hat{\alpha}_j = 0 \quad j \neq m \quad m,j = 1,\ldots p \qquad (53) \\ &\hat{\alpha}_m' \hat{\alpha}_j = \begin{cases} 1 & m = j \\ 0 & m \neq j \end{cases} \end{aligned}$$

I mentioned before that there is an algorithm developed by Flury and Gautshi for solving system 53. This algorithm consists in two procedures or sub-functions the outer and the inner procedures; the outer procedure (F algorithm) set ups the maximisation problem of the likelihood test of equation 51, whereas the inner algorithm (G) is a process of rotation, which solves the conditions of the system 53.

I implemented the FG algorithm in Matlab and although the model performs well with some of the examples of Flury's book (not all), it does not converge to any solution for the series of interest. Anyway, the code in Matlab is attached.

I tried another way for solving the problem. It is based in the idea mentioned above that we have to find a square matrix which simultaneously diagonalises two covariance matrices.

The algorithm consists of the following steps: Let Y_i and Y_j covariance matrices (pxp) corresponding to the ith and j-th periods. Also, let $\mathbf{A}_i$ and $\mathbf{A}_j$, the corresponding matrices of eigenvectors (components).

a) First, calculate the diagonal matrices using the components of the covariance matrices Yk

$$\mathbf{A}_k \mathrm{Y}_k \mathbf{A}_k' = \mathrm{L}_{kl} \qquad k = A,B,C,D \quad l = 1,\mathrm{K}\ ,p$$

where A, B, C, D are the periods, p is the number of variables in the each sample and L is a diagonal matrix of eigenvalues.

b) Second, set up an initial solution. If $\mathbf{A}_{(k,j)}^{(q.p-q)}$ correspond to the matrix composed of the first q eigenvectors for the period k and the last p-q of period j, then multiply two covariance matrices by the same initial solution.

$$\mathbf{A}_{(k,k)}^{(q.p-q)} \mathrm{Y}_k \mathbf{A}_{(k,k)}^{(q.p-q)} = \mathrm{L}_k^1$$

$$\mathbf{A}_{(k,j)}^{(q.p-q)} \mathrm{Y}_j \mathbf{A}_{(k,j)}^{(q.p-q)} = \mathrm{L}_j^1$$

c) Third, define diagonal constraints. For the initial solutions L_i^1 (i=j,k), square all the off diagonal elements $\lambda^i_{m,n}$, and set

$$Objective = \sum_{m=1}^{p} \sum_{\substack{n=1 \\ m \neq n}}^{p} (1^i_{m,n})^2 + \sum_{m=1}^{p} \sum_{\substack{n=1 \\ m \neq n}}^{p} (1^k_{m,n})^2$$

d) Set up the orthogonality constraints of the component matrices. Set Orthogonal_i equal to and identity pxp matrix such that

$$Orthogonal_k = \mathbf{A}_{(k,k)}^{(q.p-q)\prime} \mathbf{A}_{(k,k)}^{(q.p-q)}$$

$$Orthogonal_j = \mathbf{A}_{(k,j)}^{(q.p-q)\prime} \mathbf{A}_{(k,j)}^{(q.p-q)}$$

e) Using solver to minimise *Objective* subject to $Orthogonal_k$ and $Orthogonal_j$ by changing simultaneously the elements of the matrices $\mathbf{A}_{(k,i)}^{(q.p-q)}$. As it can be seen in point d, the first q-columns are common, such that when solver iterates, it automatically assigns the same values to these columns in both matrices $\mathbf{A}^{(q.p-q)}_{(k,i)}$ (i=k,j). The last (p-q) eigenvectors of each matrix are specific but they are subject to orthogonality conditions.

I run the model for 2 and 3 common components. The results for 3 common principal components are not encouraging. In fact, contrary to what I said in the previous sections, I can not reject in any case that the components are common. This is strange if we consider the variability particularly in component 3. Actually, the maximum likelihood estimates of the common components are not convincing.

The results for two common components are more reasonable. The likelihood ratio for testing PCPC is exactly the same that in equation 55 but with (k-1)p(2p-q-1)/2 degrees of freedom, where p is the number of variables, q is set equal to 2 components and k is equal to the number of covariance matrices to compare (2).

$$\chi^2_{pcpc} = -2\ln\frac{L(\hat{\Psi}_1,\ldots,\hat{\Psi}_k)}{L(\mathbf{S}_1,\ldots,\mathbf{S}_k)} = \sum_{i=1}^{k} n_i \ln\frac{\det \hat{\Psi}_{i,}}{\mathbf{S}_i} \qquad (55)$$

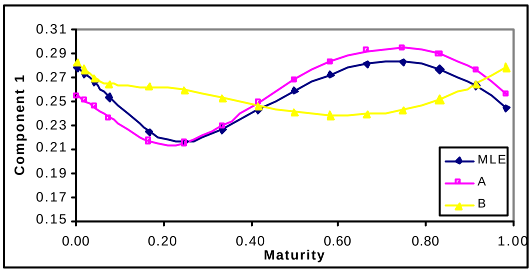
The results of the likelihood ratios are the following

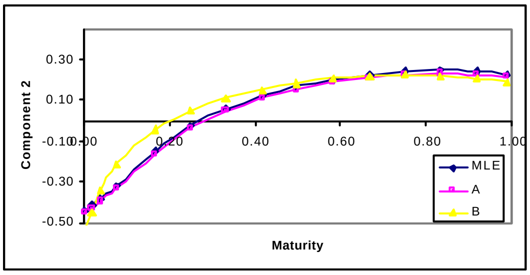
Table 12. Likelihood Ratio - 2 Common Principal Components

	A B	A C	A D	B C	B D	C D
Observations	134 413	134 150	134 88	413 150	413 88	150 88
χ**2**	**38.53**	**35.61**	**30.48**	**18.65**	**13.93**	**29.9**
df	27	27	27	27	27	27
p-value	0.0697	0.1240	0.2927	0.8822	0.9813	0.3186

The results of table 12 jointly with a graphical analysis of components will be discussed below. Notice I will plot the same components of figure 12, but in addition, the value of the maximum likelihood estimator of each component for every pair of subperiods.

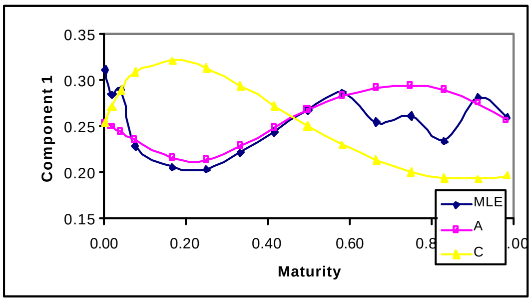
a) PCPCA-Periods A and B

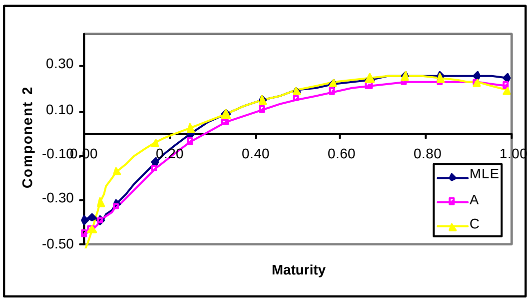




The maximum likelihood estimators of both components seem to be a weighted average of the original ones. This is consistent with the findings of Krzanowski (1984) who establishes that if common principal component holds, not only the covariance matrices but also all possible linear combinations of them have the same common components. However, according to table 12 it is possible to reject the hypothesis null of 2 common principal components at 10 % significance level.

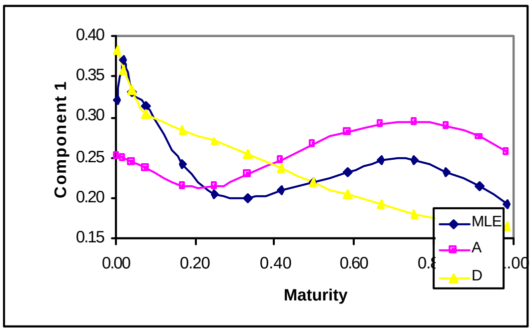
b) PCPCA. Periods A and C

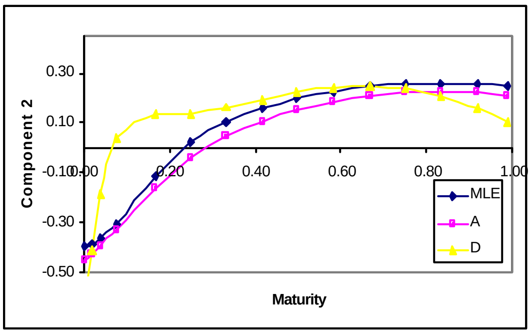




There is an erratic behaviour of the MLE of component 1 on the long-end of the curve. Regarding component 2, it behaves quiet nicely. Table 12 shows that we can not reject that the periods A and C share components 1 and 2 at 10% significance level, but the result is in the borderline of rejection.

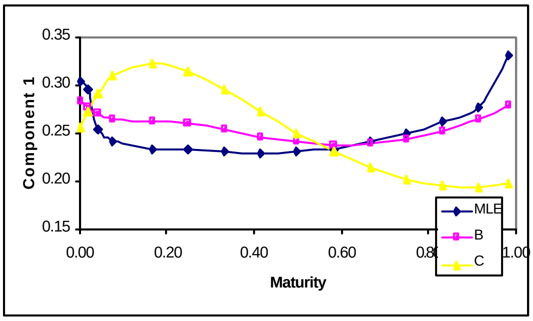
c) PCPCA. Periods A and D

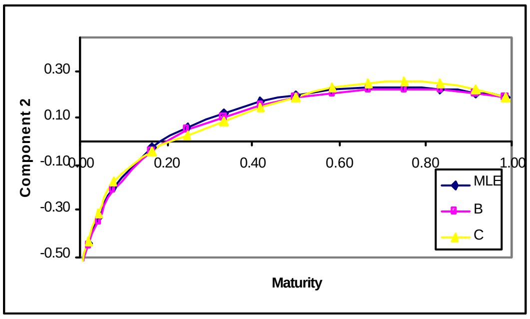




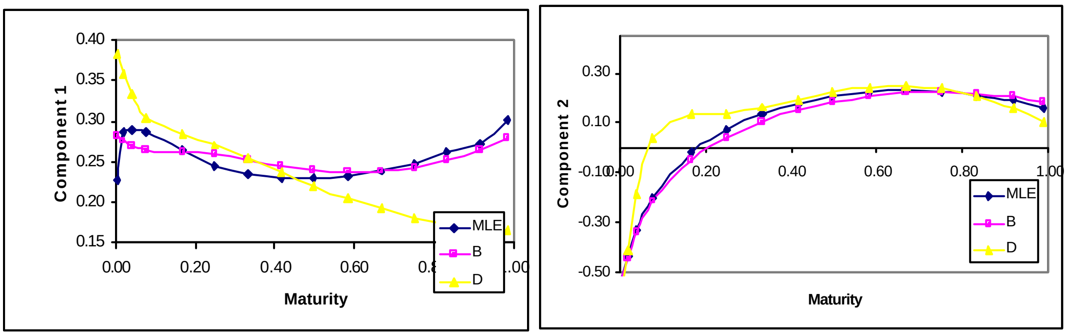
MLE has a strange behaviour in the short-end of the curve. Again, component 2 has a nice shape. According to table 12, we can not reject the hypothesis null of common principal components at almost 30% significance level.

d) PCPCA. Periods B and C and for periods B and D

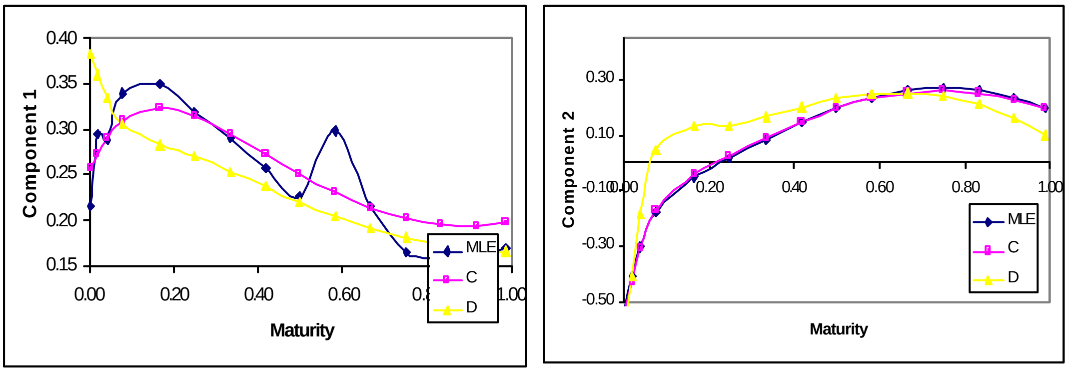




The likelihood ratio test seems to show the robustness of the hypothesis of common principal components for periods B and C and for periods B and D, their respective p-values are 0.8820 and 0.9812 that do not allow us to reject the hypothesis null at any significance level. The former was completely unexpected for me, because I had the idea that components would dramatically change from a period of stability (B) to a period of crisis (C), but it seems that only the variance explained by the components inmdede changed.



e) PCPCA. Periods A and C



The fact that we can not reject the hypothesis null of common principal components is subject to criticism considering the shape of the MLE of component 1. It can be seen in the graph for the second factor, that the MLE is very similar to the original component in period C. This is very suspicious because I used as an initial approximation the values of the components of C and it seems the algorithm did not find an alternative one.

This analysis seems to give some support to the hypothesis that the first two components are relatively stable.

Unfortunately, the results of this section are biased and I found the following drawbacks in the methodology applied. Firstly, the MLE is more closely related to the initial solution, meaning that the algorithm did not change the initial solution in most of the cases. I consider that neither the diagonalisation process nor the orthogolisation constraints of the algorithm are correct.

Secondly, as it was mention in section 4.2.2, the determinants of the covariance matrices are close to zero, which imposes numerical problem to the calculations. In fact, this is the case when sphericity is present. Jolliffe suggests that sphericity is a problem of redundancy which can be overcome by deleting some of the original variables (in our case some of the 15 interest rates extracted from the spline curves) and although he does specify some techniques for identifying them such as partial correlation among variables and some "iterative" methodologies, it is not clear that there is a formal way to do it. Remember that components are linear combinations of the original variables and deleting one of them could mean losing valuable information. I tried for example to eliminate variables highly correlated in order to eliminate multicollinearity but the results do not really improve and sometimes worsened. For example, we have seen that the correlation of the one-day rate and the seven-day rate is 0.995, by deleting the latter, the shape and proportion of the variance explained by the newly calculated factors change substantially, and even in that case the determinant is still close to zero.

Finally, I have to mention again that some of the assumptions for statistical inference in PCA do not hold either.

6.5 Final Remarks on Common Principal Component Analysis

Despite all the problems mentioned, the three different methodologies, the herein so called "crossing approach", Krzanowski and MLE/LR approach, partially support the hypothesis of common principal components for the first 2 factors (level and slope), of the yield curve. However, it is clear that the third component changes through time.

Now it would be interesting to test stability of the variance explained for the components. This can be done by testing common sphericity but I did not have time (and space) to do it. Nevertheless, it is necessary to conduct the tests again but in a more astringent environment.

I think that this hypothesis has important implications, particularly in hedging strategies. Components change, there is not doubt about, but portfolio losses can be due to changes in the influence of the directional vectors and not because these vectors change their shape or direction through time.

7. Conclusions

The aim of this work is to estimate the principal components describing the yield curve for the Mexican market and testing their stability through time. The main hypothesis was that whereas the components, at least the most important ones, are very stable or remain unchanged the variance explained for them substantially fluctuates.

Firstly, we have reviewed some of the principal economical events in Mexico in order to identify scenarios or periods. Secondly, as the money markets in Mexico are undeveloped and short-term oriented, it was necessary to perform a Spline interpolation in order to obtain more information of the yield curve. The goodness of fit of the Spline interpolation was measure throughout the mean square error. The results were particularly good. The goodness of fit of Splines was illustrated through a comparison against a Kernel method of interpolation, which in turn is the methodology used by the Mexican Central Bank.

In the following chapter, we started with a review of Principal Component Analysis an its main statistical properties. A description of the components was carried out, concluding that the classification of components proposed by Knez, Litterman and Scheinkman, "level", "slope" and "curvature" applies for the components explaining the variation of the Mexican yield curve. Common Principal Components and other levels of comparison between variance – covariance matrices were introduced in this section.

Three methodologies were applied for testing common principal component analysis. The first one, crossing approach, allows to graphically check if components of different periods are the same. Krzanowski's methodology measures the minimum angle among sub-spaces spanned by a set of principal components. If the angle is small "enough" common principal components is not rejected. Finally, MLE is a formal statistical technique based on the assumptions of multivariate normality and non-sphericity for testing common principal components. MLE is computational intensive and requires efficient algorithms for solving complicated systems of equations. Flury and Gautschi algorithm seems to perform very well for CPCA; however, it is not useful for PCPCA.

The tests applied partially support the hypothesis that the first two principal components are generally stable. The latter could have important implications on portfolio hedging strategies. It is necessary to complement the analysis with a more astringent set of statistical test accompanied by efficient algorithms of estimation. Statistically test the stability of the eigenvalues or common sphericity would be another important future task. I suspect that if the off-diagonal elements of a covariance matrix are stable, i.e., the components, then the diagonal elements should be very volatile, i.e., the eigenvalues.

Notes:

[1]This instrument was very attractive for foreign investors because of its high yield and low exchange rate risk. Low because although the Government did not have the obligation of paying back in dollars, there was a tacit promise of doing so by using the available foreign reserves.

[2]Approx. USD20 billion left the country on the 21 of December and the following days.

[3]Source Financial Times 26 of February, 4 and 5 of March 1999.

[4]Whereas Flury does it by using the spectral decomposition of a symmetric matrix and some inequalities. I will describe the spectral decomposition of a symmetric matrix on section 4.1

[5]It is important to mention that we have assumed that all the characteristic roots of Ψ are different.

[6]In addition, Jolliffe presents a good discussion of the advantages and disadvantages of obtaining the components either from the correlation matrix or the covariance matrix. If the variables have different scale measures then he recommends to use correlation matrix but the main problem is that most of the statistical results have been derived from covariance matrices.

[7]As in equation 21c.

References

Barbel J and Copper M., "Immunization Using Principal Component Analysis", The Journal of Portfolio Management ", fall 1996, 99-105.

Baxter M, and Rennie A., 1997, "Financial Calculus: An Introduction to Derivative Pricing", Cambridge, UK.

Bolton R. and Krzanowsky W, "A Characterization of PC for Projection Pursuit", The American Statisticians, May 1992, 53-2, 108-109.

Campbell J., Lo A. and MacKinlay A. G, 1992, "The Econometrics of Financial Markets", Princeton, UK.

Connor G. and Korojczyk R., "Risk and Return in an Equilibrium Asset Pricing Theory, Applications of a New Methodology", Journal of Financial Economics, 1988, 21, 255-290.

Duarte A. and Mendes B., "Robust Hedging Using Futures Contracts with an Application to Emerging Markets", The Journal of Derivatives, Fall 1998, 75-95.

Flury B., "Common Principal Components and Related Multivariate Models", Wiley Series in "Probability and Mathematics Statistics", 1988.

Golub B. and Tillman L., "Measuring Yield Curve Risk Using Principal Component Analysis, Value at Risk, and Key Rate Durations", The Journal of Portfolio Management, summer 1997, 72-84.

Hadi A. and Ling R., "Some Cautionary Notes on the Use of PC Regression", American Statisticians Associations, Feb 1998, vol 52-1, 15-19.

Heath D., Jarrow R., and Morton A., "Bond Pricing and the Term Structure of Interest Rates: A New Methodology for Contingent Claims Valuation", Econometrica, 60-1, 77-105.

Jolliffe I., 1986, "Principal Component Analysis", Springer Series in Statistics.

Knez J., Litterman R. and Scheinkman J., "Exploration into Factors Explaining Money Markets Returns", Journal of Finance, 1996, 49-5, 1861-1881.

Litterman R. and Scheinkman J., "Common Factor Affecting Bond Returns", The Journal of Fixed Incomes, 1988, 54-61.

Reitano R., "Non-parallels Yield Curve Shifts and Stochastic Immunization", The Journal of Portfolio Management, winter 1996, 71-78

Roll R. and Ross S., "An Empirical Investigation of the Arbitrage Pricing Theory", Journal of Finance, 1980, 36, 1073-1103.

Ross S., "The Arbitrage Theory of Capital Asset Pricing", Journal of Economic Theory, 1976, 13, 341-60.