Principal Component Analysis: Chile nominal and real interest rate swaps relative value strategy

Research · December 2020

DOI: 10.13140/RG.2.2.32613.01767

2 authors:

Genaro Salinas
Trinity University

Sebastian Piscoya
Trinity University

TRINITY UNIVERSITY

SAN ANTONIO • 1869

Principal Component Analysis: Chile nominal and real interest rate swaps relative value strategy

By: Genaro Salinas & Sebastian Piscoya

December 16, 2020

Trinity University – San Antonio, Texas

Table of Contents

I. Abstract

The financial swap market is one of the largest and most liquid in the global economy. These financial derivatives allow market participants to exchange a series of cash flows over a specified time frame. The Chilean swap market is the third most liquid in Latam after Mexico and Brazil, making it a relevant study to better understand the dynamics, structure, and statistical properties of its historical swap rates across multiple tenures to identify relative value trade opportunities. In order to accomplish this goal, a Principal Component Analysis (PCA) will be used to simplify and reduce the complexity between the relationships within swap rates to see precisely what its driving forces are. By using and interpreting the residuals of PCA, relative value positions across the swap curve can be found as well as cheap expressions of macro views through swaps. This paper can additionally serve as a basis for the construction of relative-value strategies in other swap markets and hedging ratios to limit a portfolio's exposure to the driving forces found by the PCA.

II. Introduction

Interest rates are closely related to every single financial security, and arguably very much everything else in the economy. They are used by individuals, companies, and governments alike for day-to-day operations from the simplest to the most complex transaction. For this reason, many market participants either hedge or speculate against the uncertain development of rates in the future, that being short, medium, or long term. One way of visualizing the term structure of interest rates is through the yield curve which is a collection of interest rates changing through different maturity periods. The line plotted through the interest rate levels for every maturity becomes the yield curve used by market participants to assess risk in financial markets and in the economy. Therefore, understanding and modeling the yield curve is quite important for investors.

Market participants are concerned about changes in the level of rates since they might affect an investment position(s) and also twist the shape of the curve that exposes to yield curve risk (changes in the slope of the curve). So what exactly affects interest rate changes? The simple answer is almost everything going on in the world at any given time from policy shifts to unemployment levels to GDP changes and expectations. For an investor, it is essential to properly isolate the effect that each risk poses to their investments so that each can be managed accordingly. A financial swap is a product offered through banks that allow two parties to exchange a series of cash flows to one another: one with a fixed interest rate and the other with a variable interest rate. As an example, say an investor is tied up with a bond paying a fixed coupon rate every period and is concerned about rates going up. This investor might want to engage in a swap agreeing to exchange that fixed amount to another party that is willing to provide a variable rate to the investor. With this product, for example, the investor can hedge against adverse interest rate movements. As with every financial asset, a swap, quoted in the fixed rate payment, is also exposed to interest rate, slope, and curvature risk. Similar to the yield curve, the line that goes through every swap rate for every possible maturity becomes the swap curve.

The Chilean swap market is the third most liquid in Latin America and offers different swap products to investors that need to hedge or speculate against movements in rates. One of them is the *Cámara swap*, that is a nominal interest rate swap that is used between market participants to exchange a fixed vs floating rate on a predetermined notional amount. Due to historical periods of very high inflation, Chile in order to aid investors tied up in long-term agreements, created a unit of account called the *Unidad de Fomento (UF)* which accounts for changes in inflation. Its currency, the *Chilean peso*, acts as a means of exchange in order to make day-to-day transactions. In this same idea, the Chilean swap market also has a real interest rate swap, where the fixed leg is CPI-linked (Consumer Price Index) to adjust to changes in inflation. In essence, the Chilean real interest rate swap (Real IRS) is the nominal swap rate minus expected inflation.

In this paper, we will examine how the Chile nominal (2011-2020) and real (2006-2020) IRS historical swap rates ranging from 1-year to 20-year tenures behave with a focus in understanding the dynamics and structure of the swap curves and also statistical properties such as their distribution. More specifically, the study will attempt to isolate the true sources of variations across swap tenures and identify relative value trade positions with Chilean nominal and real swaps considering the risks described previously such as interest rate risk and term premia. To achieve this, we propose conducting a Principal Component Analysis (PCA) to describe the variation in the data in a better vector space. PCA is a dimensionality reduction technique that constructs from the tested data principal components as linear combinations of all the variables, in this case all of the studied swap tenures (1Y-20Y).

Going back to the risks that affect interest rates and thus swap rates, the PCA will turn the expected highly correlated sway rate tenures into a set of uncorrelated orthogonal risk factors (vectors) that can explain the real relationship between the rates and the risks. According to other literature on the subject, the driving forces (principal components) behind the swap curve, yield curve, and in general term structure financial data are known to be level, slope, and curvature[1]. These PC's can therefore be used to assess how rates are exposed directly to interest rate changes, term premia (first derivative of level changes) and also curvature referring to the changes in the term premia across tenures (second derivative of level changes).

III. Data

We are using Chile nominal swap rates from 05/12/2011 until 08/06/2020 and Chile real swap rates from 01/03/2006 until 08/06/2020. The data for the analysis was obtained from the Bloomberg Terminal (tickers CHSWP, CHSWC) and cleaned in order to remove date rows that did not contain a swap rate for all tenures and also tenures with little amount of historical data.

The datasets are structured as follows:

1) The first row of the data indicates the name of the variable starting with the date column and then the rates for the 1 year, 2 year, 3 year, 4 year, 5 year, 6 year, 7 year, 8 year, 9 year, 10 year, 15 year, and 20 years swaps.
2) The data can be viewed as a 2178 x 12 matrix (nominal) and 3397 x 12 matrix (real) for the purposes of the PCA analysis. The matrix has 12 dimensions given the 12 different tenures that contain the variation of the swaps data.
3) Each value is described in percentage terms (a swap rate of 3.6%, for example, is written as 3.6).

[1] Lord, Roger and Pelsser, Antoon "Level-slope-curvature - fact or artefact?" (2006). The authors introduce how the first three principal components derived from PCA on term structure data can be interpreted as the level, slope, and curvature of the swap rates themselves.

IV. Methodology

We start by first identifying the key relationships within Chilean nominal and real swap rates. Once we obtain basic statistical properties of the data, we can now proceed to determine if a PCA is adequate and optimal for the datasets. The main assumptions of PCA are that all variables (tenures) are measured at a continuous level, are highly correlated with one another, and that the sample size used for the analysis is large enough. For this last assumption, the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy test will be used (Stewart, 57). The test yields the KMO statistic that indicates the proportion of variance in the variables that might be caused by underlying factors. A result close to 1 would indicate that the dataset can be analyzed through PCA or factor analysis. To test for stationarity in the data, the Dickey-Fuller test will be used to test the null hypothesis that a unit root is present and that the variables are non-stationary. A p-value of > 0 means that the process is not stationary while a p-value of 0 indicates that the process is stationary (McCabe, 1015).

Once determined if our dataset is prime for PCA, the analysis can be performed. Principal Component Analysis will help us exploit the structure and the high correlation within the different swap rates across the tenures by finding the projections that maximize the variance of each of our data points into a lower dimensional representation (subspace). The method is an orthogonal transformation where each of the original variables transform into new principal components. Each component accounts for the maximum amount of variance of the original variables, respectively. For example, the first principal component is the linear combination of the variables that yields the highest variance of the data. The second principal component is the next component that yields the highest residual variance, which is also completely uncorrelated with the first component. The next principal components can be described in a similar manner, each completely uncorrelated with the others. Once the principal components are calculated, the direction of the component entries can be reflected without affecting the results (Harman, 135).

By maximizing the variance of our data, we are able to capture the most important information and ignore the noise or redundant aspects in the data.

In PCA, we are interested in finding the projection, $\underline{x}_{n,}$ of our data each observation, $x_{n,}$ that are as similar to the original data points as possible, but that have a significantly lower dimensionality, so our data is more compact and easier to work with (Desenroth, Faisal, Ong, 318).

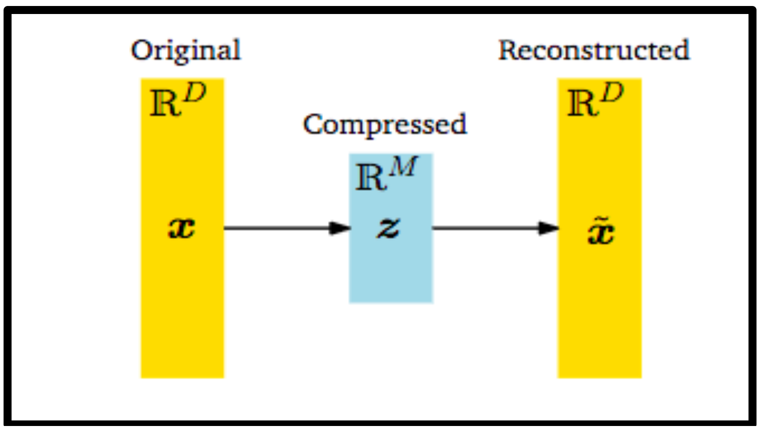


Figure 1 Graphical representation of a PCA

We start with our two *original* data sets as in the left-most part of Figure 1. The first one for the nominal interest rate swaps containing 2178 observations represented by $X_{nominal} = \{x_1, x_2, x_3, \ldots, x_{12}\}$ and the second one for the real interest rate swaps containing 3397 observations represented by $X_{real} = \{x_1, x_2, x_3, \ldots, x_{12}\}$, each x containing swap rates for 12 different maturities. For the sake of simplicity and understanding of the PCA, let's assume there is an original dataset $X \;\; = \{x_1, x_2, x_3, \ldots, x_N\} \in R^D$.

Following to the middle section of Figure 1, we seek to find a compressed version of our data in a lower dimension than that of our data set called Z_n.

$$Z_n = B^T x_n \;\in\; R^M$$

We define the projection matrix $B = [b_1, b_2, b_3, \ldots, b_N]$, whose columns form the basis of the M dimensional subspace. The compressed version of our original dataset becomes $Z_n$which lies in the subspace M such that $M \subseteq D$, meaning that the M-dimensional subspace is smaller and contained in the D-dimensional subspace of our data set X. We seek to find this projection matrix B which contains the vectors that maximize the variance of the low-dimensional data set Z_n. We proceed to define the covariance matrix S of the original dataset X.

$$S = \frac{1}{N} \Sigma x_n {x_n}^T$$

Then we choose the basis vector $b_1 \in B$ associated with the largest eigenvalue of the data covariance matrix S. This eigenvector is called the first principal component. For the purposes of our project, we then chose the two following largest eigenvalues of the covariance matrix S, $b_2, b_3 \in B$. These two eigenvectors will be the second and the third principal components that explain the remainder of the variance explained from the first principal component. For our particular dataset. The three principal components cumulatively explain 99% of the total variance. Having found this projection, matrix $B = [b_1, b_2, b_3]$ can now reconstruct our data as in Figure 1.1 in the right-most part. We define the reconstructed data as $\underline{x}$.

$$\underline{X_n} = BB^T X_n$$

The next step would be to compute the residuals by subtracting the reconstructed data from the original dataset. Since the PCA residuals are expected to balance out, have a sum of zero, and show no correlation, they can be used to find relative value trades.

V. Results & Interpretation

The following tables (Table 1 and Table 2) show descriptive statistics of Chilean real and nominal IRS. As expected, both swap curves are historically upward sloping and shorter maturities have more variation (volatility) than longer tenures. The likely reason for this outcome is due to the nature of short-term rates, which are more highly impacted by policy shifts by Chile's Central Bank than longer tenures that might be more closely related to risk management and demand from large corporations and market participants.

Table 1: Chilean Nominal Interest Rate Swaps (2011-2020) summary statistics

Variable	Mean	Std. Dev.	Min	Max
One	3.394275	1.185358	.36	5.74
Two	3.48938	1.121666	.47	5.75
Three	3.655082	1.080357	.605	5.815
Four	3.815579	1.036064	.788	5.857
Five	3.971466	.9915453	1.038	5.92
Six	4.102938	.9512807	1.295	5.925
Seven	4.217261	.9184365	1.527	5.945
Eight	4.31583	.8900335	1.69	5.965
Nine	4.399411	.8673801	1.86	5.995
Ten	4.463612	.85396	1.955	6.047
Fifteen	4.598127	.8282847	2.14	6.07
Twenty	4.691354	.8245477	2.16	6.11

Table 2: Chilean Real Interest Rate Swaps (2006-2020) summary statistics

Variable	Mean	Std. Dev.	Min	Max
One	.900295	1.276945	-2.575	4.776
Two	1.112785	1.11671	-1.78	3.73
Three	1.300821	1.056557	-1.55	3.6
Four	1.451072	1.021473	-1.42	3.7
Five	1.574074	.9880261	-1.28	3.86
Six	1.670601	.9572346	-1.16	3.92
Seven	1.759034	.934564	-1.07	3.96
Eight	1.820659	.9108629	-.96	3.96
Nine	1.87796	.900862	-.88	3.96
Ten	1.922384	.8926356	-.8	4.1
Fifteen	2.00919	.87858	-.73	4.13
Twenty	2.08791	.8689405	-.71	4.23

From Table 1 and Table 2 we also see that the minimum historical real rates are all negative, indicating periods in which inflation was larger than the nominal swap rate across all tenures. Another interesting take away from the Chilean real IRS statistics is that the highest swap rate, contrary to the general shape of its swap curve, is for the 1 year tenure. For the nominal swap, the highest rate is the 20 year which is the longest dated maturity in this analysis. In order to see if PCA works for both swaps, we correlated the tenures (Table 3 and Table 4) to determine if the variables have a high degree of correlation and also performed a KMO test as described in our methodology.

Table 3: Chilean Nominal Interest Rate Swaps (2011-2020) correlation matrix

	One	Two	Three	Four	Five	Six	Seven	Eight	Nine	Ten	Fifteen	Twenty
One	1.0000											
Two	0.9931	1.0000										
Three	0.9856	0.9980	1.0000									
Four	0.9795	0.9942	0.9987	1.0000								
Five	0.9729	0.9885	0.9949	0.9986	1.0000							
Six	0.9657	0.9812	0.9885	0.9937	0.9967	1.0000						
Seven	0.9584	0.9753	0.9843	0.9913	0.9965	0.9965	1.0000					
Eight	0.9518	0.9688	0.9788	0.9870	0.9935	0.9946	0.9994	1.0000				
Nine	0.9465	0.9635	0.9740	0.9829	0.9904	0.9924	0.9982	0.9995	1.0000			
Ten	0.9416	0.9589	0.9701	0.9798	0.9879	0.9905	0.9970	0.9988	0.9997	1.0000		
Fifteen	0.9280	0.9475	0.9601	0.9711	0.9806	0.9842	0.9923	0.9952	0.9969	0.9981	1.0000	
Twenty	0.9123	0.9362	0.9513	0.9637	0.9738	0.9781	0.9870	0.9905	0.9925	0.9943	0.9979	1.0000

Table 4: Chilean Real Interest Rate Swaps (2006-2020) correlation matrix

	One	Two	Three	Four	Five	Six	Seven	Eight	Nine	Ten	Fifteen	Twenty
One	1.0000											
Two	0.9485	1.0000										
Three	0.8786	0.9808	1.0000									
Four	0.8182	0.9509	0.9909	1.0000								
Five	0.7863	0.9315	0.9805	0.9971	1.0000							
Six	0.7632	0.9165	0.9712	0.9924	0.9982	1.0000						
Seven	0.7369	0.8981	0.9589	0.9852	0.9943	0.9983	1.0000					
Eight	0.7208	0.8860	0.9507	0.9802	0.9910	0.9955	0.9986	1.0000				
Nine	0.7024	0.8727	0.9414	0.9741	0.9865	0.9923	0.9969	0.9989	1.0000			
Ten	0.6899	0.8632	0.9347	0.9694	0.9827	0.9892	0.9947	0.9974	0.9990	1.0000		
Fifteen	0.6814	0.8565	0.9292	0.9651	0.9793	0.9870	0.9933	0.9957	0.9975	0.9978	1.0000	
Twenty	0.6641	0.8433	0.9193	0.9573	0.9729	0.9814	0.9889	0.9920	0.9943	0.9950	0.9985	1.0000

For the nominal swap data, we obtained a KMO result of 0.9286 and for the real swap data a result of 0.9303, meaning that most of the variance in the data can be explained by underlying factors (principal components). Since the variables are measured continuously, we are now certain that a PCA is a suitable tool after complying with the three assumptions of PCA as described in our methodology. The PCA for both swap products was conducted in *Stata* and *Python* to compare the outputs from both of these programs[2]. The results, shown in Table 5 and Table 6, show the first three principal components for the swaps, respectively, since the cumulative variance explained by these components is over 99%, leaving the rest of the components from the analysis insignificant.

Table 5: Chile Nominal IRS (2011-2020) Principal components (eigenvectors)

Variable	Comp1	Comp2	Comp3	Unexplained
One	0.2825	0.5256	0.7160	.0001483
Two	0.2867	0.4054	-0.0572	.00089
Three	0.2886	0.2923	-0.3281	.0006272
Four	0.2900	0.1848	-0.3636	.0002857
Five	0.2909	0.0785	-0.2873	.0005251
Six	0.2906	0.0025	-0.2133	.004287
Seven	0.2909	-0.0966	-0.0720	.0008044
Eight	0.2905	-0.1621	0.0353	.0008049
Nine	0.2899	-0.2100	0.1410	.00077
Ten	0.2895	-0.2493	0.1858	.000529
Fifteen	0.2879	-0.3413	0.2334	.0009285
Twenty	0.2861	-0.4187	0.0300	.003193

Table 6: Chile Real IRS (2011-2020) Principal components (eigenvectors)

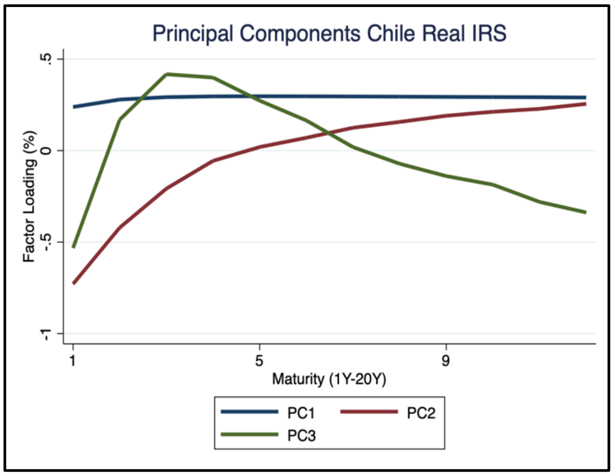
Variable	Comp1	Comp2	Comp3	Unexplained
One	0.2384	0.7295	0.5325	.0006528
Two	0.2792	0.4210	-0.1690	.00295
Three	0.2922	0.2076	-0.4177	.00172
Four	0.2963	0.0559	-0.3992	.001129
Five	0.2970	-0.0196	-0.2732	.001047
Six	0.2967	-0.0702	-0.1648	.001379
Seven	0.2958	-0.1249	-0.0193	.001392
Eight	0.2949	-0.1568	0.0717	.00124
Nine	0.2937	-0.1906	0.1392	.001191
Ten	0.2927	-0.2123	0.1861	.001906
Fifteen	0.2919	-0.2282	0.2803	.001055
Twenty	0.2901	-0.2556	0.3380	.003318

Plot 1 and Plot 2 are visual representations of the three principal components for the swap products. The next series of plots show the correlations between PC1 and the 10 Year swap rate (level), PC2 and the 10Y-2Y (slope), and PC3 and 1Y5Y10Y butterfly (curvature) for the nominal and real swap rates. For these plots, PC2 and PC3 were reflected in order to show a positive correlation between the variables[3]. Plots 3-5 show the correlations of the Chile nominal IRS and plots 6-8 for Chile real IRS.

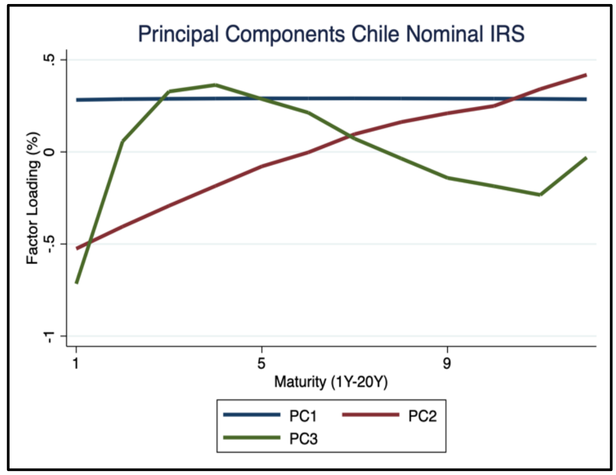
[2] Appendix 1

[3] See Harman, H. H. (1960). Modern factor analysis. Third edition revised. Univ. of Chicago Press.

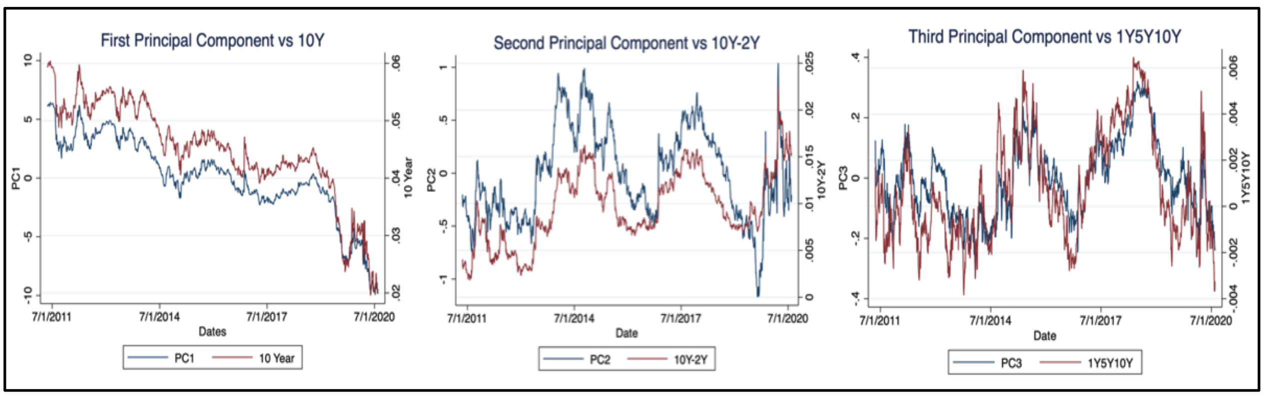
Plot 1: Nominal Swap Rates



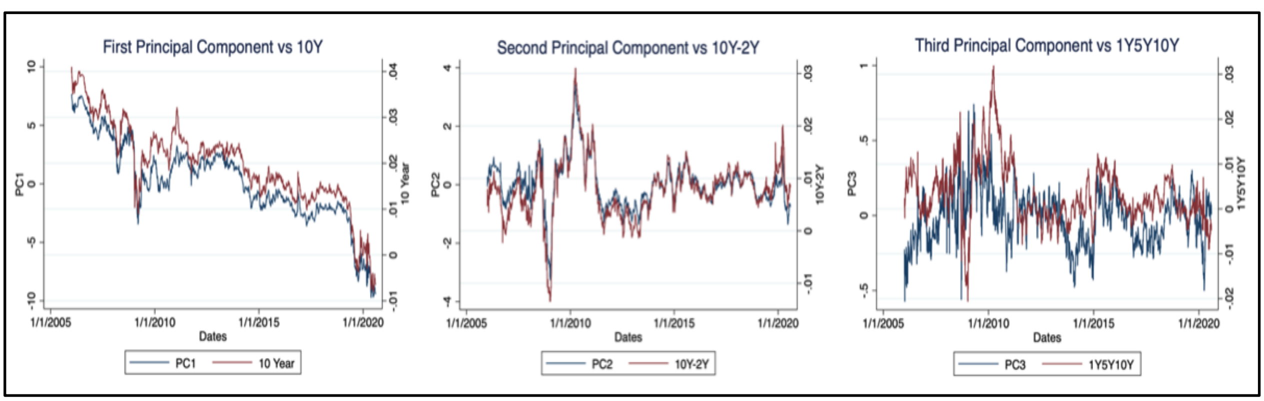
Plot 2: Real Swap Rates



Plots 3.1 - 3.3: Nominal Swap Rates



Plots 4.1- 4.3: Real Swap Rates

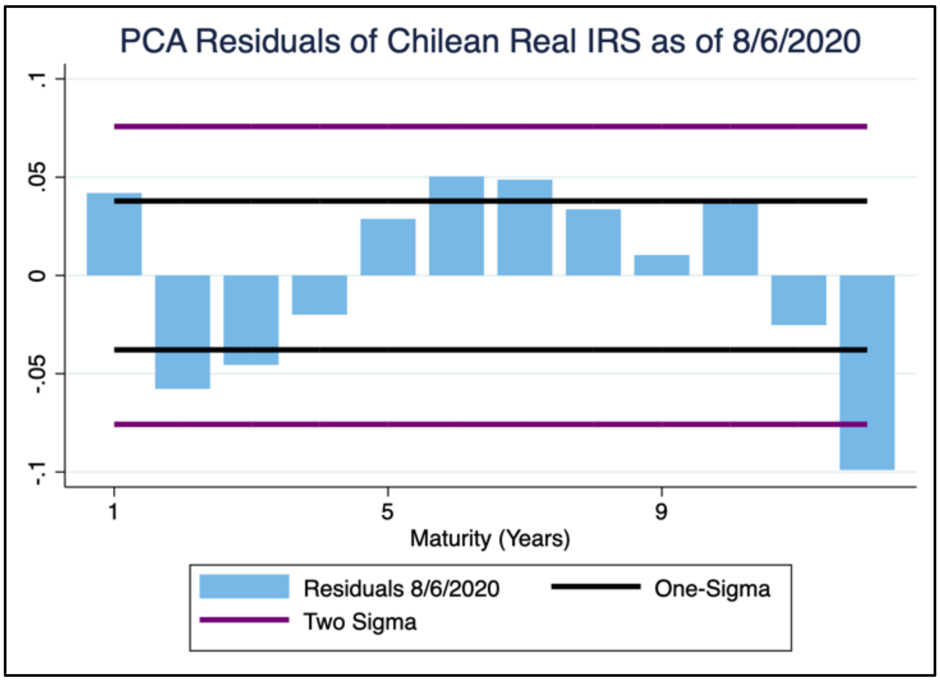


The results from the PCA indicate that the first three principal components have an actual interpretation for both swaps. For the nominal swap rates, PC1 has a 99% correlation with the 10 Year swap rate (represents the level of rates), PC2 has a 74% correlation with the 10Y-2Y swap rates (represents slope by subtracting a shorter maturity rate from a longer maturity rate), and PC3 has an 86% correlation with the 1Y5Y10Y butterfly (represents curvature). PC1 (as seen in Plot 1) has a positive effect for all tenures and is slightly upward sloping. PC2 has negative loadings until the 5Y tenure and positive loadings for all other swaps. This indicates that higher slope on the curve would cause the short end of the curve (1Y-5Y) to go lower, given the negative loadings of PC2 for these tenures, and the longer end (5Y-20Y) to go higher. PC3 has negative loadings on the shorter (1Y-3Y) and longer end of the curve (10Y-20Y), pointing out the effect that curvature has on swap rates. For the real swap rates, PC1 has a 98% correlation with level, PC2 has a 94% correlation with slope, and PC3 has a 36% correlation with curvature. The three PC's have the same interpretation as the nominal swap rates PC's.

Now in order to see if the PC's of the analysis maintain overtime, we performed another PCA on a training subset of our data to then project onto the remainder data and compute the residuals. For both datasets, we used 80% of the data for the training model and projected, with the PC's of this analysis, the remaining 20% of the data onto the PCA subset of the training data. In order to do this, we performed a matrix multiplication as described in our methodology. The first multiplication was the three PC's eigenvectors of the training model multiplied by the matrix of the test data rates that were first normalized, then the second matrix mutilation was between the result of the first multiplication multiplied by the transpose of the PC's eigenvectors matrix of the training model. Afterwards the residuals were calculated by subtracting the projected matrix of the test data (the result of the matrix multiplications) from the actual test data (the swap rates) and ended with residuals very close to 0. This outcome is a great indication that the effects of the principal components interpreted as level, slope, and curvature on Chile swap rates maintain really well over time. Therefore, the PCA results can be used to have a better estimation and understanding of how swap rates can behave in the near future. Most importantly, the residuals of the PCA can be effectively used to find relative value strategies since we now know that the PC's will maintain over time.

Plot 5 shows the residuals of the last day in the PCA analysis (complete analysis with all the data) from the real swap rates dataset as an example. The plot contains one and two sigma moves in the residuals to better visualize the strength of the residuals. A large positive or very negative residual can indicate which tenures are more expensive or cheap and thus provide a relative value indication[4]: receive the expensive rates and pay the cheap rates.

Plot 5



4 See Pelata, Giannopoulos, Haworth's *PCA Unleashed: Interest Rate Strategy* paper.

VI. Trading Strategy

The strategy will be to receive the swap rate whenever the residuals of the PCA analysis are two standard deviations from the mean up or down. For the positive residuals, the strategy will be to receive that rate given that in terms of relative value this particular tenure swap rate is expensive. For the negative residuals, the strategy will be to pay that rate since PCA indicates that the tenure's rate is cheap relative to the rest of the tenures. After backtesting this strategy[5] with the historical data, we see that the optimal holding period for this strategy is 3 trading days for both the nominal and real swaps and also works well with 1 trading day holding period. The backtesting was performed by running a PCA analysis on the first 50 days of the data for the nominal and real rates, respectively, computing the residuals of the last day of the subset (day 50), determining if whether or not there are any residuals among the tenures that are either two standard deviations above or below the mean of this subset's PCA residuals for each day. If there are multiple qualifying residuals, we pick the most negative and most positive ones. If there are any, then we check the next three days of the data set (which would be day 51-53) to see if the large and small residuals were a good indication of relative value and sense of direction of the swap rates. If there was a qualified positive residual, then this indicates that the swap rate for the specific tenure is likely to go down (the opposite for the qualifying negative residual). In the next three days of the data after the 50-day subset, we see whether or not the swap rate for the qualifying positive residual went down or not. If this rate went down, we count this first loop as a win of the strategy and loss if otherwise. The same process is followed for the qualifying negative residuals. We continue this process until the end of the historical dataset which allows for 64 loops for the real swap dataset and 41 for the nominal swap dataset.

For nominal Chile swap rates, the strategy has a win rate of 61%, being a promising trading strategy whenever there is optionality to hold specific swap contracts for a specified

[5] See Appendix 3 for Python code on the backtesting.

amount of time (optimally between 1-3 trading days). For the real Chile swap rates, the strategy for the positive residuals (whenever there is an indication that rates are likely going down in the next few days) has a win rate of 75%. However, for the negative residuals the strategy has a win rate of 37%. Given that the real rates account for expected inflation, thus reducing the risk of the product since it accounts for inflation, the increase in the win rate for the positive residual strategy makes sense. For the negative residual strategy, the low win rate could be due to the direction of swap rates themselves historically since 2006 and also the expectations of inflation. Since this specific strategy provides an indication that a specific rate is likely, in terms of relative value, to go up in the next three days while the expectation of inflation and rates themselves have an opposite force, this strategy is impacted more highly than for the nominal rates.

Hedge Ratios

Regarding hedge ratios for a swap portfolio[6], the PCA eigenvectors for PC1, PC2, and PC3 (for this section we will use the real swap rates PCA eigenvectors) can provide protection against level, slope, and curvature movements for positions along the swap curve. Looking at Table 7 we have a hypothetical swap portfolio with positions (in US dollars) along swap tenures ranging from 1-20 years:

Table 7: Hypothetical Outright Swap Portfolio

Swap Tenure (Year)	Position (millions USD)
1	-1
2	2
3	7
4	-3
5	5
6	9
7	5
8	1
9	3
10	-6
11	-2
12	1

To calculate the **risk of the portfolio** against any of the PC's, we use the following formula:

$$X_n = E_n^T * p$$

Where:

X_n: Risk exposure to PC n

n: PC being used to calculate risk exposure (1 for PC1 for example)

p: Portfolio positions vector

E^T: Transpose of the eigenvector for PC being used to calculate risk exposure

---

[6] See Pelata, Giannopoulos, Haworth's *PCA Unleashed: Interest Rate Strategy* paper Appendix.

This formula is unitless given that the calculator involves multiplying an arbitrary risk vector in US dollars by the eigenvector of a PC which is unitless. However, the further the distance from 0 of this calculation the more risk the portfolio has against the used PC since the eigenvector values are the factor loadings which simply indicate how powerful the PC impacts each tenure.

For the hypothetical portfolio in Table 7, this calculation yields a risk exposure result of 6.325 against PC1 and 1.653 against PC2.

To hedge this portfolio against **one** of the PC'S, we use the following formula:

$$H_v = (e_v * (E^T * p)) + p_v$$

Where:

H_v: Position adjustment (in millions USD) for each tenure's position

v: Tenure indicator (1 for One-year swap for example)

p: Portfolio positions vector

E^T: Transpose of the eigenvector for PC being used to calculate risk exposure

e_v: Factor loading for the tenure from the eigenvector of the PC being used

p_v: Tenure original position (in millions USD) in portfolio

For hedging against **multiple** PC's, we modify the equation:

$$H_v = (e_{1v} * (E_1^T * p)) + (e_{2v} * (E_2^T * p)) + p_v$$

Where:

H_v: Position adjustment (in millions USD) for each tenure's position

v: Tenure indicator (1 for One-year swap for example)

p: Portfolio positions vector

E_1^T: Transpose of eigenvector for first PC being hedged

E_2^T:Transpose of eigenvector for second PC being hedged

e_{1v}: Factor loading for the tenure from the eigenvector of the first PC being hedged

e_{2v}: Factor loading for the tenure from the eigenvector of the second PC being hedged

p_{v}: Tenure original position (in millions USD) in portfolio

By using the multiple PC hedging formula, the portfolio in Table 7 can be modified to hedge against PC1 and PC2:

Table 8: Modified portfolio positions to hedge against PC1 and PC2

Swap Tenure (Year)	Position (millions USD)
1	-3.4304064
2	-0.4357359
3	4.72040174
4	-5.0709709
5	3.06273961
6	7.1668186
7	3.28795793
8	-0.6375318
9	1.44779373
10	-7.493674
11	-3.4571142
12	-0.3867525

With the modified positions in Table 8 and computing the risk exposure to PC1 and PC2, the risk exposure from the portfolio to both PC's is 0, respectively and combined, meaning that the portfolio is immune to changes in level and slope.

VII. Conclusions

The study was successful in understanding the dynamics, structure, and statistical properties of the historical nominal and real Chile swap rates across multiple tenures and, through a Principal Component Analysis, be able to identify a relative value trade strategy. PCA allowed for the reduction of the complexity between the swap rates to be able to identify its primary driving forces, which we find that in accordance with other literature on the matter, the top three forces are level, slope, and curvature. Our trading strategy involves receiving the swap rate of a tenure whenever its residual is two standard deviations above the PCA residual mean from the analysis and paying the rate when its residual is two standard deviations below. After backtesting the strategy, we see that for the nominal rates, the strategy has a 61% win rate with both positive and negative residual indications. However, for the real swap rates, the strategy only works well with the positive residual indicator.

It would be interesting to see how the strategy works in a hiking rate cycle since the analysis was conducted with historical data since 2006 in which the general trend for swap rates in Chile (and globally) since is a steady decline. Given the COVID-19 pandemic in 2020 it is likely that the strategy continues to work in the next couple of years until the effects of the pandemic in the world economy have cooled off and Central Banks around the world, primarily in the United States and in Europe, are able to raise rates once business, trade, travel, manufacturing, and energy sectors (to name a few) completely recover. Once this occurs, the strategy would have to be revised by running the same backtesting analysis but only on historical periods when rates have been going up for at least one or two months to have sufficient data for a PCA analysis in hiking rate environments.

This study was able to explore hedging ratios to limit a portfolio's exposure to the driving forces found by PCA (the eigenvectors with the highest eigenvalues). For this project, level, slope, and curvature are the top three PC's, and these are risks that investors involved in interest rate products often want to control or hedge against. By using the eigenvectors of these risks, an investor can reposition their portfolio in order to completely remove them. With further study, it would also be possible and interesting to look more closely at the relationship between hedging ratios for swap portfolios combined with a relative value strategy. This work could potentially enhance the trading strategy and increase the win rate by, for example, removing slope and curvature risk and aiming only for changes in the level of rates.

VIII. Appendix

Appendix 1: Real Swap Rates PCA output from Stata

```
Principal components/correlation                  Number of obs    =
3,397
                                                  Number of comp.  =
3
                                                  Trace            =
12
    Rotation: (unrotated = principal)             Rho              =
0.9984

    ----------------------------------------------------------------------------
-
       Component |   Eigenvalue   Difference         Proportion   Cumulative
    -------------+--------------------------------------------------------------
-
           Comp1 |      11.2946      10.6401             0.9412       0.9412
           Comp2 |      .654426      .622382             0.0545       0.9957
           Comp3 |     .0320435      .024702             0.0027       0.9984
           Comp4 |    .00734148     .00340581            0.0006       0.9990
           Comp5 |    .00393568    .000646052            0.0003       0.9994
           Comp6 |    .00328962     .00185596            0.0003       0.9996
           Comp7 |    .00143367    .000343788            0.0001       0.9998
           Comp8 |    .00108988    .000419864            0.0001       0.9998
           Comp9 |   .000670016    .000180237            0.0001       0.9999
          Comp10 |    .00048978    .000031053            0.0000       0.9999
          Comp11 |   .000458727    .000188155            0.0000       1.0000
          Comp12 |   .000270571             .            0.0000       1.0000
    ----------------------------------------------------------------------------
Principal components (eigenvectors)

    ----------------------------------------------------------
        Variable |    Comp1     Comp2     Comp3 | Unexplained
    -------------+------------------------------+-------------
             One |   0.2384    0.7295    0.5325 |    .0006528
             Two |   0.2792    0.4210   -0.1690 |      .00295
           Three |   0.2922    0.2076   -0.4177 |      .00172
            Four |   0.2963    0.0559   -0.3992 |     .001129
            Five |   0.2970   -0.0196   -0.2732 |     .001047
             Six |   0.2967   -0.0702   -0.1648 |     .001379
           Seven |   0.2958   -0.1249   -0.0193 |     .001392
           Eight |   0.2949   -0.1568    0.0717 |      .00124
            Nine |   0.2937   -0.1906    0.1392 |     .001191
             Ten |   0.2927   -0.2123    0.1861 |     .001906
         Fifteen |   0.2919   -0.2282    0.2803 |     .001055
          Twenty |   0.2901   -0.2556    0.3380 |     .003318
    ----------------------------------------------------------
```

Appendix 2: Nominal Swap Rates PCA output from Stata

```
Principal components/correlation                 Number of obs    =
2,178
                                                 Number of comp.  =
3
                                                 Trace            =
12
    Rotation: (unrotated = principal)            Rho              =
0.9989

    --------------------------------------------------------------------------
-
       Component |   Eigenvalue   Difference         Proportion   Cumulative
    -------------+------------------------------------------------------------
-
           Comp1 |      11.7857      11.6009             0.9821       0.9821
           Comp2 |      .184713      .168871             0.0154       0.9975
           Comp3 |     .0158419    .00844446             0.0013       0.9989
           Comp4 |    .00739743    .00340811             0.0006       0.9995
           Comp5 |    .00398932    .00301859             0.0003       0.9998
           Comp6 |   .000970728    .00040066             0.0001       0.9999
           Comp7 |   .000570068   .000178609             0.0000       0.9999
           Comp8 |   .000391458    .00019718             0.0000       1.0000
           Comp9 |   .000194278  .0000737715             0.0000       1.0000
          Comp10 |   .000120506  .0000284899             0.0000       1.0000
          Comp11 |  .0000920165  .0000244749             0.0000       1.0000
          Comp12 |  .0000675416            .             0.0000       1.0000
    --------------------------------------------------------------------------
-

Principal components (eigenvectors)

    ----------------------------------------------------------
        Variable |    Comp1     Comp2     Comp3 | Unexplained
    -------------+------------------------------+-------------
             One |   0.2825    0.5256    0.7160 |    .0001483
             Two |   0.2867    0.4054   -0.0572 |      .00089
           Three |   0.2886    0.2923   -0.3281 |    .0006272
            Four |   0.2900    0.1848   -0.3636 |    .0002857
            Five |   0.2909    0.0785   -0.2873 |    .0005251
             Six |   0.2906    0.0025   -0.2133 |     .004287
           Seven |   0.2909   -0.0966   -0.0720 |    .0008044
           Eight |   0.2905   -0.1621    0.0353 |    .0008049
            Nine |   0.2899   -0.2100    0.1410 |      .00077
             Ten |   0.2895   -0.2493    0.1858 |     .000529
         Fifteen |   0.2879   -0.3413    0.2334 |    .0009285
          Twenty |   0.2861   -0.4187    0.0300 |     .003193
    ----------------------------------------------------------
```

Appendix 3: Python code for backtesting trading strategy

```
i=0
j=50
m=50
n=53
max_pos_count_list=[]
max_neg_count_list=[]
min_pos_count_list=[]
min_neg_count_list=[]
while i <=real.shape[0]:
   x_c=real.iloc[i:j,0:12].values
   x_m=real.iloc[m:n,0:12].values
   j=j+50
   i=i+50
   m=m+50
   n=n+50
   covariance_matrix=np.cov(x_c.T)
   #eigenvalues/vectors
   eigen_values,eigen_vectors=np.linalg.eig(covariance_matrix)
   # subset the first three columns bcs they explain >99% variation
   PCs=(eigen_vectors.T[:][:])[:3].T
   #projected data
   X_projected=x_c@(PCs)@PCs.T
   #residuals
   res=x_c-X_projected
   residuals=pd.DataFrame(res)
   # compute mean and std of the residuals and if statement
   mean_res=np.mean(res)
   std_res=np.std(res)
   std_res_above_1=mean_res+std_res
   std_res_above_2=mean_res+2*std_res
   std_res_below_1=mean_res-(std_res)
   std_res_below_2=mean_res-(2*std_res)
   x_c_df=pd.DataFrame(x_c)
   x_m_df=pd.DataFrame(x_m)
   last_day=residuals.tail(1)
   last_day_x_c= x_c_df.tail(1)
   last_day_x_m= x_m_df.tail(1)
   maxValueIndex = last_day.idxmax(axis=1)
   minValueIndex=last_day.idxmin(axis=1)
   max_last_day=last_day.iloc[:,maxValueIndex].values
   min_last_day=last_day.iloc[:,minValueIndex].values
   if max_last_day>std_res_above_2:
     pass
   elif min_last_day<std_res_below_2:
     pass
   else:
     continue  max_swap_res_substraction=(last_day_x_c.iloc[:,maxValueIndex].values-
last_day_x_m.iloc[:,maxValueIndex].values)
   # if residual is large, then receive that rate (we want the swap rate to decrease)
   # if residual is small, then pay that rate (we want the swap rate to increase)
min_swap_res_substraction=(last_day_x_c.iloc[:,minValueIndex].values-last_day_x_m.iloc[:,minValueIndex].values)
   maximum_neg_count = len(list(filter(lambda x: (x < 0), max_swap_res_substraction)))
   maximum_pos_count = len(list(filter(lambda x: (x >= 0), max_swap_res_substraction )))
   max_pos_count_list.append(maximum_pos_count)
   max_neg_count_list.append(maximum_neg_count)
   minimum_neg_count = len(list(filter(lambda x: (x < 0), min_swap_res_substraction)))
   minimum_pos_count = len(list(filter(lambda x: (x >= 0), min_swap_res_substraction )))
```

```
   min_pos_count_list.append(minimum_pos_count)
   min_neg_count_list.append(minimum_neg_count)
   if n >=real.shape[0]:
     break
max_res_count_positive=np.sum(max_pos_count_list)
max_res_count_negative=np.sum(max_neg_count_list)
print(max_res_count_positive/(max_res_count_positive+max_res_count_negative))
print('Number of loops:',max_res_count_positive+max_res_count_negative)
min_res_count_positive=np.sum(min_pos_count_list)
min_res_count_negative=np.sum(min_neg_count_list)
print(min_res_count_negative/(min_res_count_positive+min_res_count_negative))
print("Overall win rate",
(max_res_count_positive+min_res_count_negative)/(min_res_count_positive+min_res_count_negative+max_res_count_po
sitive+max_res_count_negative))
```

IX. Bibliography

1. Baygún, Bulent. Showers, Janet. Cherpelis, George. (2000) *United States Fixed-Income Research Portfolio Strategies: Principles of Principal Components A Fresh Look at Risk, Hedging, and Relative Value*. Salomon Smith Barney. Retrieved December 8, 2020, from: http://quantlabs.net/academy/download/free_quant_instituitional_books_/%5BSalomon%20Smith%20Barney%5D%20Principles%20of%20Principal%20Components%20-%20A%20Fresh%20Look%20at%20Risk,%20Hedging%20and%20Relative%20Value.pdf
2. Deisenroth, M. P., Faisal, A. A., & Ong, C. S. (2020). *Mathematics for machine learning*. Cambridge, United Kingdom: Cambridge University Press.
3. Harman, H. H. (1960). *Modern factor analysis*. Third edition revised. Univ. of Chicago Press.
4. Kelly, Richard et al. (2015) *Relative Value across the US Swap Surface: A PCA Approach*. TD Securities. Retrieved December 8, 2020, from: https://www.tdsresearch.com/currency-rates/viewEmailFile.action?eKey=36U9QCCZP3ZJIHHLM1CRYVCZA
5. Lord, Roger. Pelsser, Antoon. (2006) *Level-slope-curvature - fact or artefact?*. Tinbergen Institute. Retrieved on December 8, 2020 from: https://papers.tinbergen.nl/05083.pdf
6. McCabe, B., & Tremayne, A. (1995). Testing a Time Series for Difference Stationarity. *The Annals of Statistics, 23*(3), 1015-1028. Retrieved December 11, 2020, from http://www.jstor.org/stable/2242434
7. Pelata, Marion. Giannopoulos, Panos. Haworth, Helen. (2012). *PCA Unleashed: The application of PCA across European Rates markets*. Credit Suisse. Retrieved December 08, 2020, from: https://research-doc.credit-suisse.com/docView?language=ENG&source=emfromsendlink&format=PDF&document_id=1001969281&extdocid=1001969281_1_eng_pdf&serialid=Coz8ZUCgL92gmMydSBULHIsgm%2b9q2TPfBu%2bX1XhViIs%3d
8. Stewart, D. (1981). The Application and Misapplication of Factor Analysis in Marketing Research. *Journal of Marketing Research, 18*(1), 51-62. doi:10.2307/3151313