Erasmus University Rotterdam
Erasmus School of Economics
Bachelor Thesis Econometrics

Different PCA Approaches for Forecasting Bond Risk Premia

Sam Lakmaker (580050)

Supervisor:	Gustavo Freire
Second assessor:	Daan Opschoor
Date final version:	2nd July 2023

Abstract

Forecasting bond risk premia is important for every investor. Forecasting bond risk premia can be done with different types of data such as with forwarded bond rates but also with macroeconomics. In this research three different principal component analysis (PCA) methods are used to find the best forecasting method. For the in-sample forecast, scaling PCA (sPCA) provided the best results. When we perform an out-of-sample forecast, normal PCA performs the best. sPCA and PCA perform neck and neck. Sparse PCA (SPCA) creates factors which makes variables too sparse, and thus misses too much information. Although for SPCA improvement is achievable by making it less sparse and having access to more computer power .

1 Introduction

Forecasting bond risk premia is of interest for a long time. Predicting the bond risk premia is use full for portfolio management such that portfolios can be balanced in an optimal way. Bond risk is also associated with economical situations. By accurately predicting the bond risk premia, you may out smart the market and thus for example see an economical downturn before the rest. Forecasting the bond risk premia is also important for the valuation of stocks, since valuations are often discounted on the risk free bond returns.

Multiple papers have been written about this topic and used different methods. One of the newest ways is by machine learning, but in this research we will focus on a rather classic method, namely principal component analysis (PCA). PCA is an old technique which was proposed by Karl Pearson in 1901. The aim of PCA is to create just a few new factors that include information from multiple variables. There have been done multiple researches on bond risk premia which make use of PCA such as the research of Ludvigson & Ng (2009). They use a large macroeconomic dataset and try to forecast the bond risk premia, and found that their estimated factors contain substantial predictive power.

The biggest drawback of PCA is that all variables are included in the principal components, even though a variable might not explain anything. The research of Zou et al. (2006) introduced a new variant of the PCA, namely sparse PCA (SPCA). SPCA can set loads to zero and thus excludes variables from the estimated factors. As the name suggests the main difference between PCA and SPCA is that some loads can become sparse. Next to that there is also a method called scaling PCA (sPCA) which scales the data before performing PCA. The advantage is that the data is centered, which could make the PCA computation more accurate.

The research question is which PCA approach can most accurately explain the bond risk premia? Where the researched approaches are SPCA, sPCA and PCA.

The hypothesis is that SPCA should perform the best since it is able to make loads zero for variables with less or none predictive power. Also Goh et al. (2012) found that it is likely in a big dataset to have certain types of variables with more predictive power than others. Furthermore, did Huang & Shi (2011) also conclude that the predictability of the bond risk premia is mainly explained by variables of the group employment, price indices and housing market. So previous researches have clearly shown that there are a few variables with more predictive power, it makes us wonder if the other variables have a substantial impact at all.

The used data contains 127 transformed macroeconomic variables, of the sample period 01:1964-12:2002. In addition the yield data used by Ludvigson & Ng (2009) contains small inaccuracies, that is why in this research more accurate yield data will be used as created by Liu & Wu (2021). The used yield data range is 01:1965-12:2003.

The exciting literature did already find some interesting results. The research of Rapach & Zhou (2021) found that SPCA is especially better for cross-sectional asset pricing since it can better recognise signals in macroeconomic data. Since asset pricing and bond risk premia are somewhat correlated it is really interesting to see if the SPCA can also better forecast the bond risk premia. Also the research of Zhu et al. (2023) found similar results, but they concluded that macroeconomics alone might not include enough information to precisely forecast stock returns. They suggest that more microeconomic information is needed for stock returns, since bond risk premia relies heavily on macroeconomics it seems that extra microeconomic information is not that important in case of predicting the bond risk premia. This is partly in contrast to the research of Ludvigson & Ng (2009), they concluded that macroeconomic data have predictive power and they find strong predictable variation in bond returns. Also did they conclude that macroeconomic data has even without the factor of Cochrane & Piazzesi (2005) substantial predictive power.

2 Data

Previous research of Cochrane & Piazzesi (2005) and Ludvigson & Ng (2009) is based on the sample period 1964:1–2003:12. For this research, we examine the same sample period to make sure that we can compare the researches. The period does include multiple recession and also economically stable periods. This period includes all kinds of market changes, so we have a sample with different economical situations.

The monthly macroeconomic data will be collected from the economic research federal reserve bank of St. Louis by McCracken (2021). These 127 monthly economic series have to be individually transformed to ensure stationarity. In order to perform these transformations we use the code of McCracken (2021). This code also includes an algorithm to insert missing values. The data sample used from this database is 11:1963-12:2002. The first two month of these data are only included because we take first and second differences, after transforming the data the first two months are than excluded. This data set does differ to the one of Ludvigson & Ng (2009), for example their macroeconomic data set included five more variables. These differences could influence the results.

The monthly bond yield data is gathered from Liu & Wu (2021), this is in contrast to the research of Ludvigson & Ng (2009). The advantage of this data set is that there are less pricing errors, which should make the bond data more accurate than the one used by Ludvigson & Ng (2009). The used maturities are one, two, three, four, and five years. The used bond yield return range is 01:1965-12:2003, this differs by a year of the sample for the macroeconomics. The reason is that the excess return are forecasted a year ahead, this will be explained in section 3.2

3 Methodology

In this section the methods used for the research are discussed. First, the different PCA methods will be explained. Than the construction of the excess return and regression models are presented. Lastly, the out-of-sample forecast including a test statistics will be discussed.

3.1 PCA

PCA is a dimension reduction method and is widely used. First certain variables in the data are transformed to ensure stationarity, after which the all data is standardized. The covariance matrix of the standardized data will be computed to find the relations between variables. Next we find the eigenvalues (explained variance by eigenvector) and eigenvectors (principal componets) of the covariance matrix by performing a eigendecomposition. In the research of Ludvigson & Ng (2009) they chose to optimize the amount of factors by the BIC criteria, in this research this is simplified by choosing the same amount of factors as Ludvigson & Ng (2009). After this a projection is performed by multiplying the standardized data with the principal components.

This research is based on the research of Ludvigson & Ng (2009), so for an in depth explanation about PCA we refer to their paper.

3.1.1 Scaling PCA

The difference between sPCA and PCA is how the data is transformed. The idea is to scale data instead of standardizing. The advantage of this method is that the data will centered. This method is often used when the distribution of the data is not known. The disadvantage is that scaling is more affected by outliers, which can occur during periods of recession. Also the research of Palaniappan & Ravi (2006) found that the performance of PCA improved by 20% when using scaled data compared to not scaling the data before hand.

Scaling the data is done in the following way

$$tx_{ti} = \frac{x_{ti} - min_i}{max_i - min_i} \tag{1}$$

where t stand for the monthly time form 1964:1 to 2002:12, i for the variable in the range of 1 to 127. The min_i (max_i) is the lowest (highest) value of the variable, and x presents the value at a certain time of variable that is than transformed to tx. After this PCA is used in the sameway as in the research of Ludvigson & Ng (2009)

3.1.2 Sparse PCA

SPCA makes variables within a factor sparse, meaning that the influence of certain variables can become zero within a factor. An elastic net problem is applied to compute new principal component until the loads within the factor converge.

The SPCA method is based on the research of Zou et al. (2006). We will explain SPCA shortly but for more in depth details we refer to this paper. To compute the SPCA we first have to compute the ordinary principal components, which can be computed by PCA. This part can be done on the same way as in the Ludvigson & Ng (2009) paper. These principal components are the starting point of the SPCA approach.

On these first eight principal components an elastic net problem will be applied to compute new principal components, this process will be repeated until the loading of the principal components converge. The elastic net problem that will be applied is the following

$$\beta_j = \arg\min_{\beta} (\alpha_j - \beta)^T X^T X (\alpha_j - \beta) + \lambda \|\beta\|^2 + \lambda_{1,j} \|\beta_1\| \qquad j = 1, ..., 8 \tag{2}$$

The betas that flow out of this elastic net problem will be used to construct the singular value decomposition. In this equation the α_j are the computed principal components. After the principal components have been optimized they have to be normalised as a final step.

In theory this method is promising but in reality it has some bottlenecks. Computing the principal components by the SPCA method requires a lot of computer power, especially for data sets with a lot of variables. That is possibly also the reasons why there is a lack of research on macroeconomical variables with SPCA. Performing SPCA as how it is intended is simply not possible with the available equipment for this research. To still perform a SPCA variant we have to simplify the procedure. We fix certain criteria such as how sparse the variables can be, we choose a value of one for λ which is definitely more sparse than PCA but on the other hand does not make the factor rely on one or two variables. This value was chosen after trying out some different values for λ. Also do we make use of a so called mini-batch optimization. This takes smaller subsets of the data to make the process compute faster. It takes the smaller subsets on random bases but still this is not optimal, since certain data may not be included and thus this method is less accurate. Even with this process it takes a lot of computer power to compute the factors.

3.2 Regression

To forecast the excess return we create multiple regressions. The generalised form of the regression is

$$rx_{t+12}^{(n)} = \beta_0 + \beta_1^{'}\hat{F}_t + \beta_2 CP_t + \epsilon_{t+1} \tag{3}$$

where $rx_{t+12}^{(n)}$ is the excess return of a n-year bond a year later. How the excess return is exactly calculated is explained in section 3.2.1. $\hat{F}_t$ contains the estimated factors which can be $\overrightarrow{F6}_t = (\hat{F}_{1t},\ \hat{F}_{1t}^3,\ \hat{F}_{2t},\ \hat{F}_{3t},\ \hat{F}_{4t},\ \hat{F}_{8t})$ or $\overrightarrow{F5}_t = (\hat{F}_{1t},\ \hat{F}_{1t}^3,\ \hat{F}_{3t},\ \hat{F}_{4t},\ \hat{F}_{8t})$. Also can it included the single factor $F6_t$ or $F5_t$ as explained in section 3.2.2. Finally the CP_t factor can also be included, which estimation is explained in the section 3.2.2. A constant is always included in each regression. Since there is serial correlation in the standard errors, we make use of an 18 lag Newey and West structure which is also used by Cochrane & Piazzesi (2005).

3.2.1 Excess return

The yield data is gathered from Liu & Wu (2021), to obtain the excess return from this data we have to transform the data. The data used contains only zero coupon bonds. First we calculate the log price by applying this formula

$$logP_t(n) = -ny_t(n) \tag{4}$$

where n is the bond maturity in years and y_t the yield at time t, t is in months. After this we calculate the return on this bond when we hold it for one year and then sell it. We do this by subtracting the log price of a bond from the log price of a bond one year later which thus has one year of maturity less. This equation shows this holding period return

$$r_{t+12}(n) = logP_{t+12}(n-1) - logP_t(n) \tag{5}$$

where $r_{t+12}(n)$ is return of the one year holding period. To correct this return and obtain the excess return we have to subtract the yield of an one year bond in the following way

$$rx_{t+12}(n) = rx_{t+12}(n) - y_t(1) \tag{6}$$

where $rx_{t+12}(n)$ is the excess return that is used in this paper.

3.2.2 Single factor

To find out if a linear combination can also forecast the excess return, we create a single factor for the subset $\vec{F6}_t$ in the following way

$$\begin{aligned} \frac{1}{4}\sum_{n=2}^{5} rx_{t+12}^{(n)} &= \gamma_0 + \gamma_1 \hat{F}_{1t} + \gamma_2 \hat{F}_{1t}^3 + \gamma_3 \hat{F}_{2t} + \gamma_4 \hat{F}_{3t} + \gamma_5 \hat{F}_{4t} + \gamma_6 \hat{F}_{8t} + u_{t+1} \\ F6_t &\equiv \hat{\gamma}' \vec{F6}_t \end{aligned} \tag{7}$$

where $\hat{\gamma}'$ contains the six estimated coefficients. The same is done for $\vec{F5}_t$,

$$\begin{aligned} \frac{1}{4}\sum_{n=2}^{5} rx_{t+12}^{(n)} &= \delta_0 + \delta_1 \hat{F}_{1t} + \delta_2 \hat{F}_{1t}^3 + \delta_3 \hat{F}_{3t} + \delta_4 \hat{F}_{4t} + \delta_5 \hat{F}_{8t} + v_{t+1} \\ F5_t &\equiv \hat{\delta}' \vec{F5}_t \end{aligned} \tag{8}$$

where $\hat{\delta}'$ contains the five estimated coefficients.

We also construct the CP model by Cochrane & Piazzesi (2005). For this model the forwarded rates have to be calculated. In contrast to research of Cochrane & Piazzesi (2005) and Ludvigson & Ng (2009), we use again the revised yield data of Liu & Wu (2021). The n-year forwarded rate at time t is calculated in the following way

$$f_t(n) = logP_t(n-1) - logP_t(n) \tag{9}$$

where $logP_t(n)$ represents the log price of the n-year bond. With these forwarded rates, we can than construct a single factor in the same procedure as done for the other single factors

$$\begin{aligned} \frac{1}{4}\sum_{n=2}^{5} rx_{t+12}^{(n)} &= \theta_0 + \theta_1 y_t^{(1)} + \theta_2 f_t^{(2)} + \theta_3 f_t^{(3)} + \theta_4 f_t^{(4)} + \theta_5 f_t^{(5)} + w_{t+1} \\ CP_t &\equiv \hat{\theta}' \vec{CP}_t \end{aligned} \tag{10}$$

where $\hat{\theta}'$ contains the five estimated coefficients.

3.3 Out of sample forecasting

An out of sample forecast can tell us more if the bond risk premia can be predicted. There are two forecast samples we use, namely 01:1985-12:2003 and 01:1995-12:2003. The data before the start of this sample is used the construct the starting factors and are then updated every iteration. For example, the forecast of the excess return of 04:1985 uses the variable data of 01:1960-04:1984 and the excess return data of 01:1961-03:1985. The forecast is recursively and adds every month newly available data, this also means that each month the factors have to be re-estimated.

Forecast are done with an unrestricted model ($\vec{F5}_t$ and $\vec{F5}_t$ + CP) and compared to a restricted model (constant and constant + CP). $\vec{F5}_t$ is the subset of the factors, note that this is not the linear combination. CP is the linear combination as presented in 10. This makes the forecasts including CP a little bit more complex. Since we first have to forecast the average excess return, before we can forecast the excess return with the model $\vec{F5}_t$ + CP and constant + CP. To overcome this issue we first forecast the average excess return for the CP model (lineair combination as in 10), afterwards we can forecast the model as specified. Essentially meaning we do a double forecast. All the above mentioned out-of-sample forecast variants are done for the PCA variant.

We also compare different PCA methods, but only for the 2- and 5-year excess return variant with $\vec{F5}_t$ as the unrestricted model and the constant as the restricted model. The forecast sample is the short one, 01:1995-12:2003. The reason for this select comparison is that performing SPCA takes a lot of computing power, especially when the factors have to be re-estimate every iteration. By selecting the closest and furthest excess return we try to get an image of the accuracy of the the other PCA variants.

To test the significance of the models we use the ENC-NEW test statistic of Clark & McCracken (2001). This statistic is calculated as following

$$\text{ENC-NEW} = P\frac{P^{-1}\sum(\hat{u}_{u,t+1}^2 - \hat{u}_{u,t+1}\hat{u}_{r,t+1})}{MSE_r} \quad (11)$$

where P is the amount of forecasts and $\hat{u}$ the estimated error. The subscript u stands for the unrestricted model ($\vec{F5}_t$, $\vec{F5}_t$ + CP) and subscript r notes the restricted model (constant, constant + CP)

4 Results

In this section the results are presented and discussed. First the summary statistics will be explained. Then the in-sample forecast will presented. Afterwards the factors will discussed and the out-of-sample forecast will be presented.

4.1 Summary statistics

	PCA		sPCA		SPCA	
i	AR1	R^2	AR1	R^2	AR1	R^2
1	0.692	0.157	0.841	0.184	0.785	0.155
2	0.698	0.232	0.900	0.282	0.918	0.226
3	-0.216	0.288	0.656	0.355	0.001	0.281
4	0.548	0.340	0.321	0.400	0.115	0.329
5	0.345	0.386	0.476	0.439	0.949	0.372
6	0.554	0.419	0.229	0.475	0.061	0.403
7	0.254	0.451	-0.240	0.509	0.383	0.433
8	-0.335	0.478	0.277	0.537	-0.337	0.453

Table 1: This table reports the summary statistics for each factor using different PCA. The used data contains 127 variables in the time period 01:1964-12:2002 after transformation. The AR1 model gives information about the persistence and the R^2 explains the variation.

In table 1 the summary statistics of the different pca methods are presented. For the normal PCA method the R^2 results are similar to the ones of Ludvigson & Ng (2009). There is a small difference in the first factor, where the first factor of Ludvigson & Ng (2009) captures 0.02 more variation. Next to that we see that the sPCA captures around 10% more variation compared to PCA. In contrast SPCA seems to perform the worst. This can be argued by the fact that SPCA makes variables sparse within a factor. A lot of variables could be missing in a factor which might explain the lack of explained variation. For the first five factors the variation is between 37-44%. For the AR1 coefficient we see that the PCA and SPCA give results in a similar way. In both cases the 8th factor is negative, the main difference is for the 5th factor. sPCA does have some considerably different AR1 coefficients but for all three models the range of the coefficients is around the same.

4.2 In-sample forecast

We compare each n-year excess return of a treasury bond by different PCA methods. Since CP is constructed by the forwarded rates and does not involve PCA computations it is noted separately in table 2. The coefficient of CP is significant and can explain 27% of two-year excess return of next year. For the PCA methods we see that sPCA tends to predict the excess return the best by explaining 26% when including all the presented factors. PCA follows closely but performs a bit less. SPCA give way worse results, and is only able to explain 10%. Also none of the coefficients of SPCA are significant when all factors are included (row b of SPCA). When making a linear combination of the factors we see that the explanation of all models decreases by around 1%. What we also see is that the CP factor can influence the significance of other factors, this can come due to the fact that the CP factor already captures the information that is also presented in a certain factor. Surprisingly it also has this effect the other way around, we see this in row e of SPCA. Including CP turns factor 1 and 3 significant. The explanation behind this could be that the CP factor overestimated it and that this is correct by the other factors.

Table 2: Results for $rx^{(2)}_{t+1}$. The coefficient are presented in bold if they are significant at a 5% level, and the corresponding t-statistics are presented in parentheses. CP_t, $F5_t$ and $F6_t$ are linear combinations as explained in the mythology. The used data is from 1964:1 till 2003:12

		$\hat{F}_{1t}$	$\hat{F}^3_{1t}$	$\hat{F}_{2t}$	$\hat{F}_{3t}$	$\hat{F}_{4t}$	$\hat{F}_{8t}$	CP_t	$F5_t$	$F6_t$	R^2
	a							**0.45**			0.27
								(7.033)			
	b	**0.18**	**-0.00**	-0.10	**-0.06**	**0.19**	**0.08**				0.22
		(4.13)	(-2.22)	(-1.83)	(-3.03)	(3.73	(2.29)				
	c	**0.15**	**-0.00**	**0.03**	-0.00	**0.12**	**0.03**	**0.41**			0.38
		(4.05)	(-2.16)	(0.69)	(-0.01)	(2.75)	(1.11)	(4.92)			
PCA	d	**0.17**	**-0.00**		**-0.06**	**0.19**	**0.08**				0.19
		(4.02)	(-2.24)		(-2.96)	(3.44)	(2.20)				
	e	**0.15**	**-0.00**		-0.00	**0.12**	0.03	**0.39**			0.38
		(4.16)	(-2.18)		(-0.18)	(2.97)	(1.13)	(5.02)			
	f								**0.53**		0.19
									(4.99)		
	g									**0.49**	0.21
										(5.46)	
	h							**0.38**	**0.39**		0.37
								(4.91)	(4.15)		
	b	**1.33**	-0.24	**-1.29**	-0.32	**1.19**	**-1.23**				0.26
		(3.58)	(-1.78)	(-3.16)	(-0.78)	(4.06)	(-2.09)				
	c	**1.25**	**-0.28**	-0.19	-0.25	**0.98**	-0.63	**0.39**			0.39
		(3.96)	(-1.99)	(-0.47	(-0.75)	(3.25)	(-1.34)	(4.52)			
sPCA	d	**1.32**	-0.24		0.33	**1.19**	**-1.24**				0.17
		(3.34)	(-1.51)		(0.70)	(3.16)	(1.99)				
	e	**1.24**	-0.28		0.24	**0.97**	-0.59	**0.41**			0.39
		(3.93)	(-1.94)		(0.75)	(3.14)	(-1.34)	(5.73)			
	f								**0.54**		0.16
									(4.99)		
	g									**0.48**	0.25
										(6.55)	
	h							**0.40**	**0.44**		0.38
								(5.92)	(5.09)		
	b	-0.09	0.00	0.06	-0.7	0.44	0.21				0.10
		(-1.67)	(-1.42)	(1.66)	(-1.65)	(-1.60)	(-1.56)				
	c	**-0.11**	0.00	-0.04	**-0.09**	0.11	0.06	**0.45**			0.35
		(-2.99)	(1.72)	(-0.92	(-2.23)	(-1.53)	(-1.30)	(6.56)			
SPCA	d	-0.10	0.00		-0.05	-0.19	-0.07				0.10
		(-1.83)	(1.47)		(-1.32)	((-1.66)	(-1.27)				
	e	**-0.11**	0.00		**-0.10**	0.11	0.07	**0.44**			0.35
		(-2.65)	(1.66)		(-2.35)	(-1.50)	(1.76)	(6.68)			
	f								**0.55**		0.09
									(2.48)		
	g									**0.53**	0.10
										(2.84)	
	h							**0.42**	**0.45**		0.34
								(6.17)	(3.03)		

For the other n-year returns we see similar patterns, these tables can be found in the Appendix A. Row c, which includes the presented factors and CP, explains most of the excess return for every model. sPCA performs the best and is closely followed by PCA. SPCA performs the worst and misses a bit of explanation compared to the other models. This is probably the case because the variables within the factors become too sparse. It is also worth noting that when the maturity of the bonds increases, SPCA seems to perform relatively more worse than the other methods.

Comparing the results to the results of Ludvigson & Ng (2009), we see that Ludvigson & Ng (2009) performs slightly better. Since the used variables as well as the bond yield is gathered from different sources, this could explain the small differences.

4.3 Interpretation of factors

To provide a more in dept explanation of the factors, we use the variable groups as constructed by McCracken (2021). We calculated the marginal R^2 of each variable within a factor. The figures below presented these marginal R^2 for the factors that are also used for the in-sample forecast. The presented figures contain the results of the normal PCA method. Since the interpretation for the different PCA methods is similar we do not present them. The marginal R^2 of factor one by the SPCA method is also presented to visualize how the SPCA method functions.

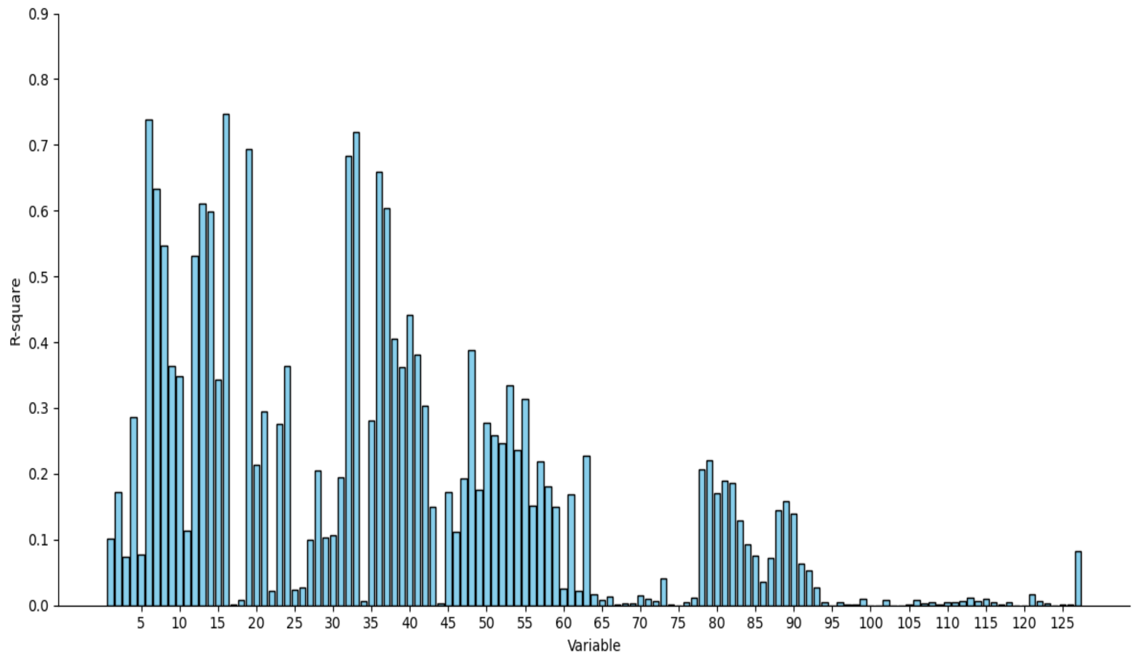

Figure 1: Marginal R^2 for factor 1 by using PCA

In figure 1 we see that R^2 are the highest for the output and income group. Variables as IP growth and manufacturing score the highest. Also some variables of the labor market group are included, such as total employment and good producing industries. The factor also contains some variables of the consumption and orders group and stock market group but this is relatively less presented.

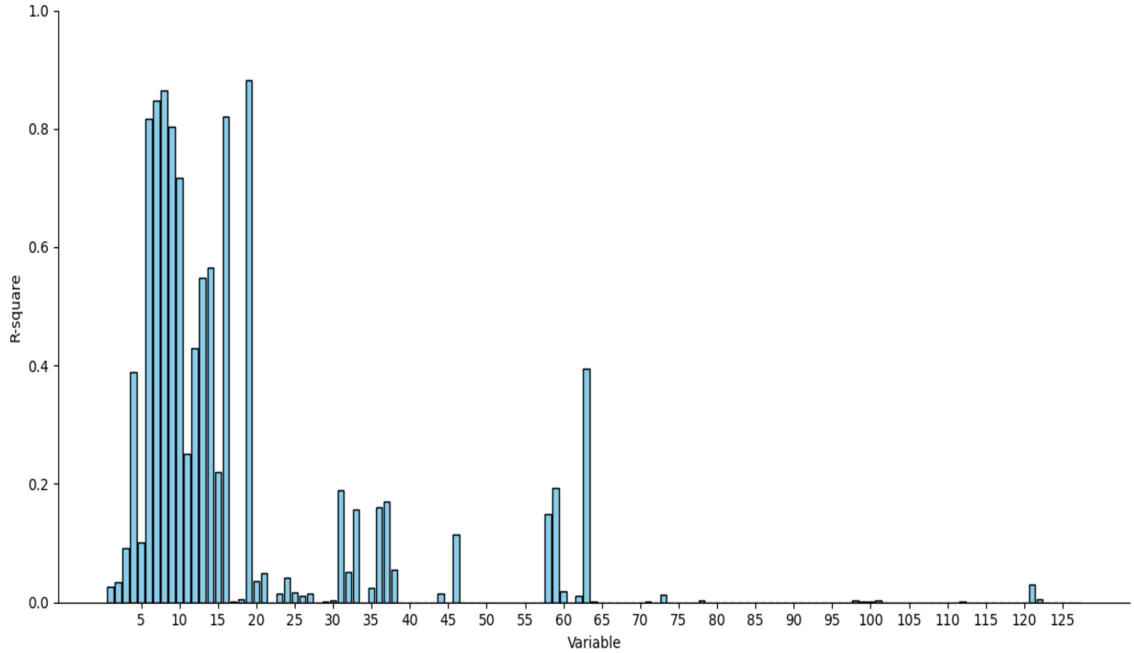

Figure 2: Marginal R^2 for factor 1 by using SPCA

In figure 2 we see a more extreme effect of the PCA. We see that by using SPCA the load shifts to the group output and income. By making the other variables sparse, the factor is more focused on a certain group but the trade off is that there is possibly useful information missing. As we saw in figure 1, the labor market group contains some significant explanation but is not included in the SPCA variant, and thus might miss some important information.

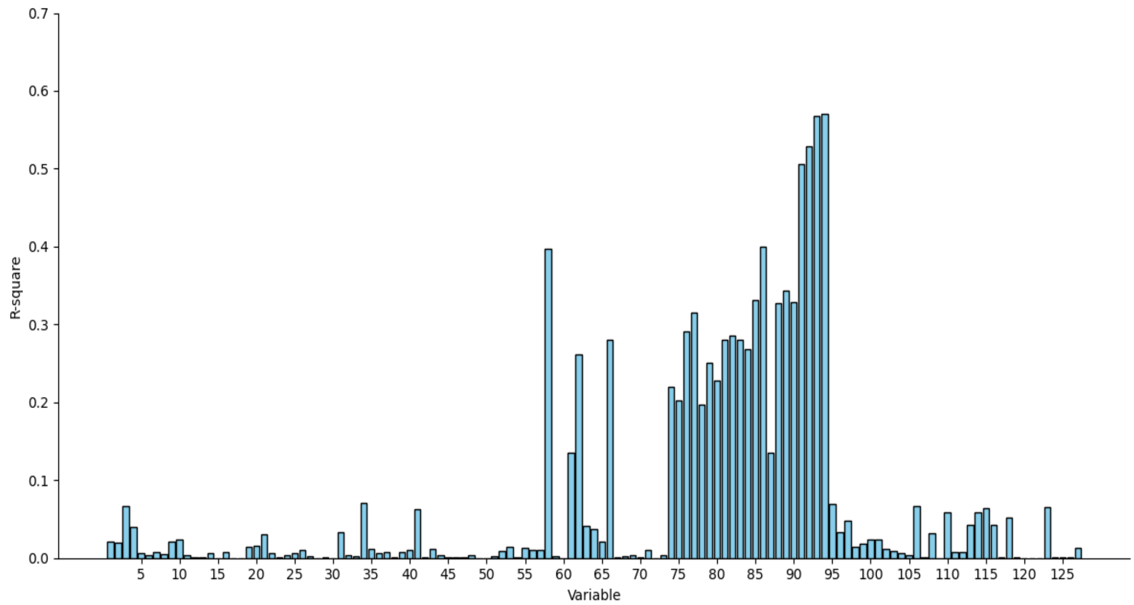

Figure 3: Marginal R^2 for factor 2 by using PCA

Factor 2 explains mainly variables of the interest rate and exchange rates group as can been seen in figure 3. It explains a huge portion of the bond spread variation, the same was concluded by Ludvigson & Ng (2009). Besides that we see that almost all variables explain a bit of the variation. Since it explains so less it could also cause some noise in the regression and a preferred choice could be the SPCA factor.

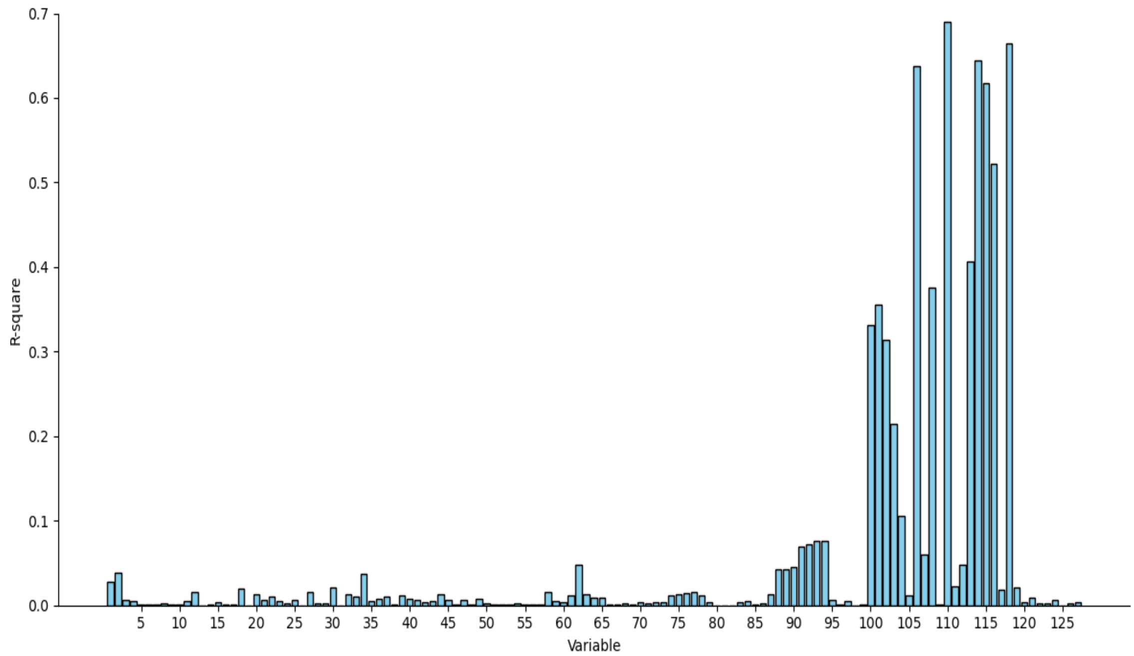

Figure 4: Marginal R^2 for factor 3 by using PCA

Factor 3 almost explains 70% of the variation of some variables in the group prices. The highest marginal R^2 is the variable of oil spot price. Factor 3 also includes some high CPI and PPI variables, which is inline with Ludvigson & Ng (2009). Again we see also in figure 4 that the other variables do not explain much of the variation.

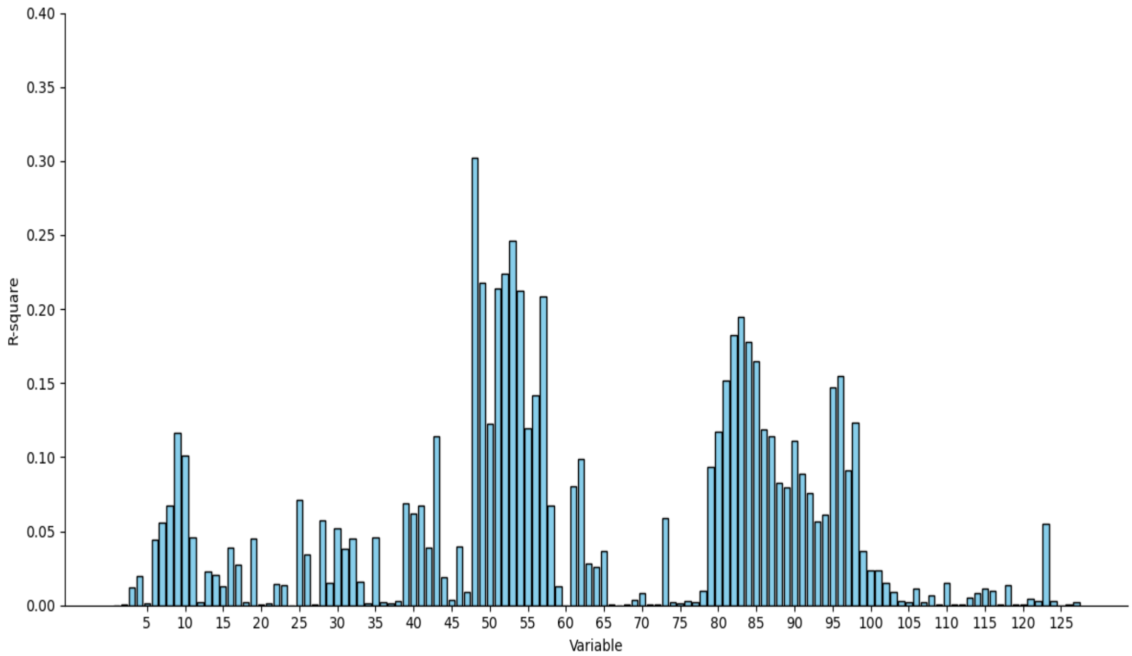

Figure 5: Marginal R^2 for factor 4 by using PCA

Just as factor one, also factor four explains the variation in the consumption and orders group. What is interesting is that the R^2 is around the same level for this group. The difference is that factor four does not explain the variation in the groups output and income, and the group labor market. One could argue that factor four is a subset of factor one, since factor one is so similar in terms of R^2 for certain variables but factor one also include high R^2 for the some other groups.

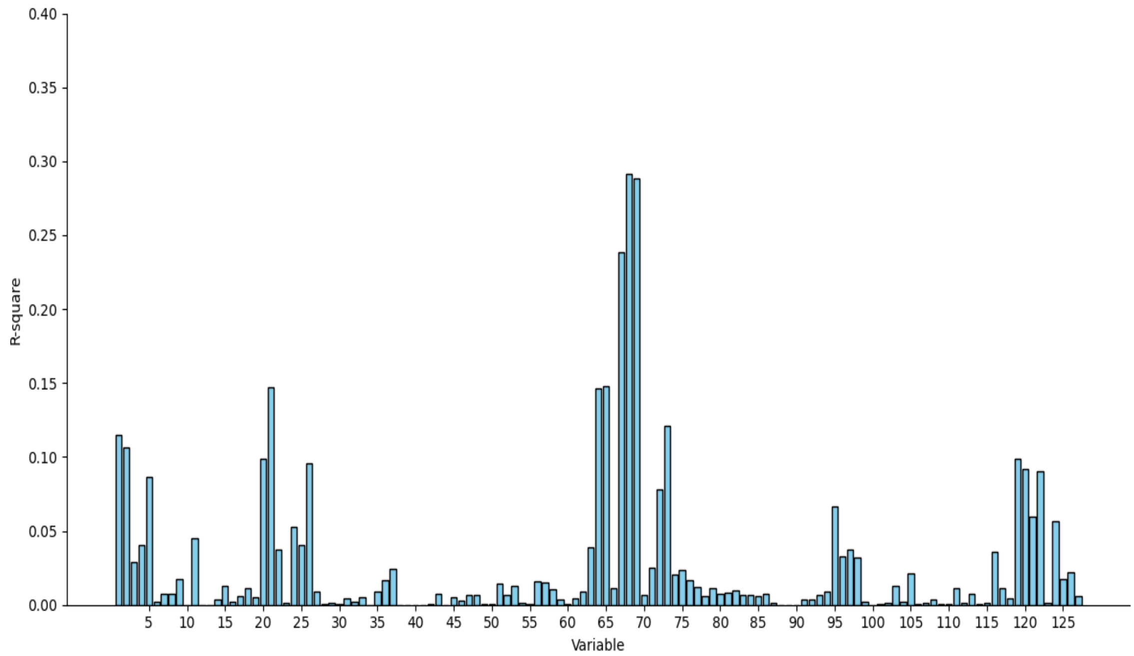

Figure 6: Marginal R^2 for factor 8 by using PCA

Factor eight is completely different compared to the one of Ludvigson & Ng (2009). Their eight factor included mainly the stock market. Also did their other variables not explain much variation. Their other variables were not complete spars but were close to 0. When we look at figure 6 we see that factor eight loads mostly on the variables business inventories, which is quite a difference.

To give a more in dept explanation of the impact of the factors, the results of $F6_t$ are presented

$$\begin{aligned}\frac{1}{4}\sum_{n=2}^{5} rx_{t+12}^{(n)} &= \underset{(2.08)}{0.98} + \underset{(3.74)}{0.31}\hat{F}_{1t} \underset{(-2.37)}{-0.00}\hat{F}_{1t}^{3} \\ &\quad \underset{(-2.19)}{-0.28}\hat{F}_{2t} \underset{(-3.06)}{-0.14}\hat{F}_{3t} + \underset{(3.31)}{0.37}\hat{F}_{4t} + \underset{(2.16)}{0.18}\hat{F}_{8t} + u_{t+1} \qquad (12) \\ R^2 &= 0.193\end{aligned}$$

where the t-statistics are presented in the parentheses. From this result we can conclude that $\hat{F}_{4t}$ is the most important factor followed by $\hat{F}_{1t}$ and $\hat{F}_{2t}$. Interestingly this differs from the results of Ludvigson & Ng (2009) where $\hat{F}_{8t}$ has a high impact and $\hat{F}_{4t}$ not. There are of course multiple reasons for this difference. The most logical reason is the difference data that is used, maybe that is also the reason why our $\hat{F}_{8t}$ includes different variables than Ludvigson &

Ng (2009). The yield curves of Liu & Wu (2021) which we use, are calculated more accurately and could also cause a difference in the estimated coefficients.

4.3.1 Correlation between factor and IP growth

For the first factor in figure 7 we see a high correlation of 0.92 with the IP growth. It does have sometimes trouble following the spikes precisely but it is really close. We see that in 1983 and 1992 when the IP growth decreases, that factor one overestimates this. Although it is clear that factor one follows growth and it goes counter cyclic with the excess return. This makes sense since excess return is often high during periods of negative IP growth. For the linear combination of F5 we see the opposite. We see also that the correlation between IP growth and F5 is negative, but it is also less correlated. Probably due to the fact that are more factors included and not only factor one that is already heavy loaded on the output and income group.

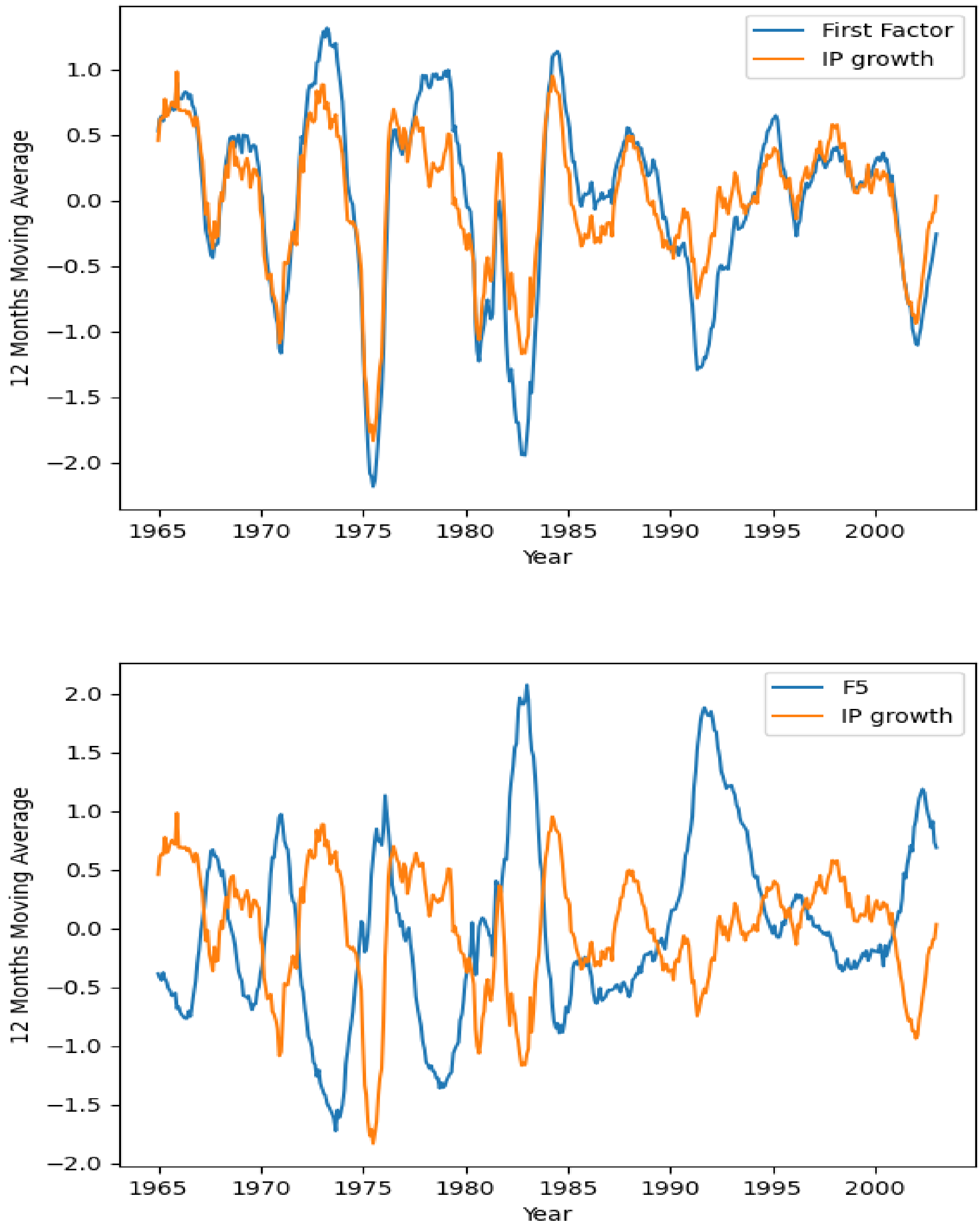

Figure 7: 12 months moving average of the IP growth compared to the first factor and F5. Correlation between the first factor and IP growth is 0.922 and for F5 and IP growth -0.583

4.4 Out-of-sample forecasting

We will first only look at the result of the out-of-sample forecast by PCA, which are presented in table 3. We see that for every n-year excess return the unrestricted model outperforms the constant model. We see that the forecasts are the most impressing for the 2-year excess return, by only having 80% of the MSE of the constant model. We also see that when n increases that the MSE comes closer to that of the constant model. More interestingly we see that when the forecast sample becomes smaller the performance increases. The difference is not huge, but what is makes interesting is that Ludvigson & Ng (2009) had worse results for all models with smaller sample compared to bigger sample. It seems more logical to perform better when the forecast sample is smaller, since the forecast can be based on more historical data. So predicting in the near future is most valuable when using PCA. Next to that, all test statistics using the ENC-NEW are significant at a 1% level.

Table 3: This table presents the results of the out-of-sample forecast for the PCA method. The unrestricted model included the factors of $\vec{F5}_t$ and can also include the CP factor. This is then compared with the restricted model that contains a constant and can also include the CP factor. The test statistic is the ENC-NEW and the 95% asymptotic CV is also given.

Row	Forecast sample	Comparison	MSEu/MSEr	Test statistic	95% Asymptomatic. CV
		$rx^{(2)}_{t+1}$			
1	01/1985-12/2003	$\vec{F5}_t$ vs const	0.82	35.34	3.18
2	01/1995-12/2003	$\vec{F5}_t$ vs const	0.81	16.81	2.47
3	01/1985-12/2003	$\vec{F5}_t$ + CP vs const + CP	0.84	25.89	2.30
4	01/1995-12/2003	$\vec{F5}_t$ + CP vs const + CP	0.84	9.70	1.40
		$rx^{(3)}_{t+1}$			
5	01/1985-12/2003	$\vec{F5}_t$ vs const	0.85	27.65	3.18
6	01/1995-12/2003	$\vec{F5}_t$ vs const	0.82	14.14	2.47
7	01/1985-12/2003	$\vec{F5}_t$ + CP vs const + CP	0.87	19.27	2.30
8	01/1995-12/2003	$\vec{F5}_t$ + CP vs const + CP	0.87	7.88	1.40
		$rx^{(4)}_{t+1}$			
9	01/1985-12/2003	$\vec{F5}_t$ vs const	0.87	22.83	3.18
10	01/1995-12/2003	$\vec{F5}_t$ vs const	0.84	12.63	2.47
11	01/1985-12/2003	$\vec{F5}_t$ + CP vs const + CP	0.89	14.96	2.30
12	01/1995-12/2003	$\vec{F5}_t$ + CP vs const + CP	0.89	6.54	1.40
		$rx^{(5)}_{t+1}$			
13	01/1985-12/2003	$\vec{F5}_t$ vs const	0.89	18.88	3.18
14	01/1995-12/2003	$\vec{F5}_t$ vs const	0.86	10.60	2.47
15	01/1985-12/2003	$\vec{F5}_t$ + CP vs const + CP	0.92	11.87	2.30
16	01/1995-12/2003	$\vec{F5}_t$ + CP vs const + CP	0.91	5.40	1.40

In table 4 we see the comparison of the different PCA methods. In contrast to the in-sample forecast, PCA now outperform sPCA. The difference are again close, similarly as in the in-sample forecast. Not surprisingly the SPCA does just barely outperform the constant model. Although all models are significant at an 1% level, the effort for SPCA does not seem worth it. This can come due to the fact that the optimization for SPCA is not done optimal since we used a less accurate but faster method.

Table 4: In this table the different out-of-sample forecast for different PCA approaches can be found. See for more details about the table the note of table 3.

Row	Forecast sample	Comparison	MSEu/MSEr	Test statistic	95% Asymptomatic. CV
			$rx_{t+1}^{(2)}$		
PCA	01/1995-12/2003	$\vec{F5}_t$ vs const	0.81	16.81	2.47
sPCA	01/1995-12/2003	$\vec{F5}_t$ vs const	0.81	15.26	2.20
SPCA	01/1995-12/2003	$\vec{F5}_t$ vs const	0.97	8.84	1.73
			$rx_{t+1}^{(5)}$		
PCA	01/1995-12/2003	$\vec{F5}_t$ vs const	0.86	10.60	2.47
sPCA	01/1995-12/2003	$\vec{F5}_t$ vs const	0.88	9.83	2.20
SPCA	01/1995-12/2003	$\vec{F5}_t$ vs const	0.99	4.69	1.73

5 Conclusion

Forecasting the bond risk premia on macrovariables with different PCA methods, in the period 01:1964-12:2003 is studied in this paper. First we constructed factors on three different PCA variants. Then we also constructed two single factors based on the other factors and the CP factor was constructed. The sPCA and PCA method performed very closely in the in-sample forecast, but sPCA could slightly better explain the excess return of the bonds. Next to that, SPCA was implemented as a method to construct factors. SPCA had as problem that the computing time of the factors take a long time before they are optimized. Besides that were the factors too sparse and thus contain not enough information to explain the excess return.

By making histograms of the marginal R^2 we could conclude which variables had high loads in a factor. We found that most factors contain different groups of variables. Except factor one and four, it looks like factor four is a subset of factor one.

Lastly we performed an out-of-sample forecast. The PCA method with the smallest forecast sample and the largest training data performed the best compared to the constant model. Also forecasting bonds with a longer maturity makes the forecast stands less out compared to the constant model. Unfortunately the comparison between the different PCA variants is limited because of the computing power that is necessary for SPCA. Surprisingly we saw that PCA performed better than sPCA, which is in contrast with the in-sample forecast result. SPCA performed again disappointingly.

5.1 Limitations and extensions

Before hand SPCA seemed to be a very interesting method, since it can make variables sparse. Since the data used contains many variables of which some are not related to the bond risk premia, it seemed logical to completely exclude those variables. It turned out that the way the SPCA was implemented made the variables way too sparse, meaning that useful information was missing from the factors. This also came due tot the fact that computing SPCA demands a lot of computer power, because of this the SPCA could not be performed 100% correctly and was not as accurate as it should be. For future research it would be interesting to see if a faster computer can perform the SPCA completely and if the results differ.

Next to that is the out-of-sample forecast procedure for the CP factor not done in a great way. First the average return on the CP factor had to forecasted after which the final forecast was performed. So there is done a double forecast which includes different forecasted values for the same variables. For future research it would be interesting to rethink this process, and try to find a procedure that does not involve a forecast on a forecast.

References

Clark, T. E., & McCracken, M. W. (2001). Tests of equal forecast accuracy and encompassing for nested models. *Journal of econometrics*, *105*(1), 85–110.

Cochrane, J. H., & Piazzesi, M. (2005). Bond risk premia. *American economic review*, *95*(1), 138–160.

Goh, J., Jiang, F., Tu, J., & Zhou, G. (2012). Forecasting government bond risk premia using technical indicators. In *25th australasian finance and banking conference.*

Huang, J.-z., & Shi, Z. (2011). Determinants of bond risk premia. In *Afa denver meetings paper.*

Liu, Y., & Wu, J. C. (2021). Reconstructing the yield curve. *Journal of Financial Economics*, *142*(3), 1395–1425. doi: 10.1016/j.jfineco.2021.03.012

Ludvigson, S. C., & Ng, S. (2009). Macro factors in bond risk premia. *The Review of Financial Studies*, *22*(12), 5027–5067.

McCracken, M. W. (2021). *FRED databases.* https://research.stlouisfed.org/econ/mccracken/fred-databases/. ([Online; accessed 8-May-2023])

Palaniappan, R., & Ravi, K. (2006). Improving visual evoked potential feature classification for person recognition using pca and normalization. *Pattern Recognition Letters*, *27*(7), 726–733.

Rapach, D., & Zhou, G. (2021). Sparse macro factors. *Available at SSRN 3259447*.

Zhu, L., Jiang, F., Tang, G., & Jin, F. (2023). The sword of damocles: Sparse macroeconomic risks and the cross-section of stock returns. *Available at SSRN 4445765*.

Zou, H., Hastie, T., & Tibshirani, R. (2006). Sparse principal component analysis. *Journal of computational and graphical statistics*, *15*(2), 265–286.

A Appendix

Table 5: Results for $rx_{t+1}^{(3)}$. The coefficient are presented in bold if they are significant at a 5% level, and the corresponding t-statistics are presented in parentheses. CP_t, $F5_t$ and $F6_t$ are linear combinations as explained in the mythology. The used data is from 1964:1 till 2003:12

		$\hat{F}_{1t}$	$\hat{F}_{1t}^3$	$\hat{F}_{3t}$	$\hat{F}_{4t}$	$\hat{F}_{8t}$	CP_t	$F5_t$	$F6_t$	R^2
	a						**0.84**			0.29
							(6.78)			
	b	**0.29**	**-0.00**	**-0.11**	**0.33**	**0.15**				0.16
		(3.81)	(-2.41)	(-2.85)	(3.10)	(2.11)				
	c	**0.25**	**-0.00**	-0.01	**0.20**	0.05	**0.75**			0.38
		(3.96)	(-2.34)	(-0.29)	(2.67)	(1.240	(5.14)			
PCA	d							**0.91**		0.16
								(4.85)		
	e								**0.88**	0.20
									(5.52)	
	f						**0.72**	**0.64**		0.37
							(5.01)	(4.07)		
	b	**2.18**	-0.43	-0.48	**2.12**	**-2.17**				0.14
		(3.15)	(-1.66)	(-0.56)	(2.91)	(-1.99)				
	c	**2.01**	**-0.53**	-0.31	**1.72**	-0.94	**0.77**			0.38
		(3.76)	(-2.16)	(-0.54)	(2.84)	(1.22)	(5.65)			
sPCA	d							**0.91**		0.14
								(4.72)		
	e								**0.86**	0.25
									(6.57)	
	f						**0.76**	**0.72**		0.38
							(5.70)	(4.81)		
	b	**-0.44**	**0.00**	-0.01	0.09	0.05				0.12
		(-3.31)	(2.36)	(-0.77)	(1.14)	(0.74)				
	c	**-0.32**	**0.00**	-0.02	**0.16**	-0.08	**0.80**			0.36
		(-2.85)	(2.00)	(-0.65)	(1.92)	(-1.56)	(5.65)			
SPCA	d							**0.92**		0.12
								(3.27)		
	e								**0.86**	0.21
									(6.92)	
	f						**0.75**	**0.65**		0.35
							(5.55)	(2.61)		

Table 6: Results for $rx_{t+1}^{(4)}$. The coefficient are presented in bold if they are significant at a 5% level, and the corresponding t-statistics are presented in parentheses. CP_t, $F5_t$ and $F6_t$ are linear combinations as explained in the mythology. The used data is from 1964:1 till 2003:12

		$\hat{F}_{1t}$	$\hat{F}_{1t}^3$	$\hat{F}_{3t}$	$\hat{F}_{4t}$	$\hat{F}_{8t}$	CP_t	$F5_t$	$F6_t$	R^2
	a						**1.22**			0.32
							(6.77)			
	b	**0.37**	**-0.00**	**-0.18**	**0.44**	**0.21**				0.15
		(3.51)	(-2.49)	(-2.98)	(2.88)	(2.03)				
	c	**0.31**	**-0.00**	-0.02	**0.24**	0.07	**1.12**			0.39
		(3.63)	(-2.41)	(-0.39)	(2.40)	(1.19)	(5.26)			
PCA	d							**1.19**		0.15
								(4.66)		
	e								**1.20**	0.19
									(5.56)	
	f						**1.08**	**0.79**		0.38
							(5.13)	(3.83)		
	b	**2,66**	-0.55	0.66	**3.01**	**-3.08**				0.12
		(2.80)	(-1.61)	(0.54)	(2.84)	(-2.04)				
	c	**2.42**	**-0.69**	0.42	**2.38**	-1.25	**1.15**			0.40
		(3.36)	(-2.19)	(0.54)	(2.88)	(1.25)	(5.69)			
sPCA	d							**1.18**		0.12
								(4.51)		
	e								**1.20**	0.24
									(6.88)	
	f						**1.13**	**0.90**		0.39
							(5.60)	(4.61)		
	b	0.00	0.02	0.00	-0.11	0.09				0.10
		(0.00)	(1.59)	(0.01)	(-0.91)	(-0.71)				
	c	-0.02	0.01	-0.00	-0.08	-0.16	**1.16**			0.36
		(-0.07)	(1.33)	(-0.09)	(-0.80)	(-1.44)	(5.78)			
SPCA	d							**1.18**		0.10
								(3.85)		
	e								**1.18**	0.12
									(3.72)	
	f						**1.12**	**0.65**		0.35
							(5.27)	(2.62)		

Table 7: Results for $rx_{t+1}^{(5)}$. The coefficient are presented in bold if they are significant at a 5% level, and the corresponding t-statistics are presented in parentheses. CP_t, $F5_t$ and $F6_t$ are linear combinations as explained in the mythology. The used data is from 1964:1 till 2003:12

		$\hat{F}_{1t}$	$\hat{F}_{1t}^3$	$\hat{F}_{3t}$	$\hat{F}_{4t}$	$\hat{F}_{8t}$	CP_t	F5_t	F6_t	R^2
	a						**1.49**			0.31
							(6.20)			
PCA	b	**0.41**	**-0.00**	**-0.22**	**0.52**	**0.26**				0.13
		(3.29)	(-2.35)	(-2.92)	(2.71)	(1.93)				
	c	**0.34**	**-0.00**	-0.03	**0.27**	0.08	**1.38**			0.37
		(3.32)	(-2.29)	(-0.47)	(2.18)	(1.13)	(4.92)			
	d							**1.38**		0.13
								(4.49)		
	e								**1.43**	0.18
									(5.73)	
	f						**1.34**	**0.89**		0.36
							(4.81)	(3.60)		
sPCA	b	**2.88**	-0.61	0.79	**3.71**	**-3.72**				0.11
		(2.53)	(-1.44)	(0.53)	(2.77)	(-1.99)				
	c	**2.58**	**-0.78**	0.50	**2.94**	-1.48	**1.41**			0.38
		(2.95)	(-1.99)	(0.51)	(2.79)	(1.19)	(5.27)			
	d							**1.18**		0.12
								(4.51)		
	e								**1.20**	0.24
									(6.88)	
	f						**1.13**	**0.90**		0.39
							(5.60)	(4.61)		
SPCA	b	-0.25	0.01	-0.01	-0.12	**0.49**				0.16
		(-1.12)	(0.96)	(-0.28)	(-0.66)	(2.28)				
	c	-0.32	0.01	-0.02	0.23	0.24	**1.50**			0.33
		(-1.08)	(0.95)	(-0.29)	(1.61)	(1.13)	(6.38)			
	d							**1.65**		0.06
								(2.21)		
	e								**1.48**	0.10
									(2.93)	
	f						**1.43**	0.56		0.32
							(5.83)	(0.77)		