Stockholms universitet

Prediction of inflation based on high dimensional time series

Martina Sandberg

Kandidatuppsats i matematisk statistik
Bachelor Thesis in Mathematical Statistics

Kandidatuppsats 2015:8
Matematisk statistik
Juni 2015

www.math.su.se

Matematisk statistik
Matematiska institutionen
Stockholms universitet
106 91 Stockholm

Stockholm University

Mathematical Statistics
Stockholm University
Bachelor Thesis **2015:8**
http://www.math.su.se

Prediction of inflation based on high dimensional time series

Martina Sandberg*

June 2015

Abstract

Forecasting of macroeconomic variables has for many years been done by models with quite few variables. Today we have many different time series that can contain relevant information. In this paper we want to forecast one series using many predictor time series. We will do this by applying the dynamic factor model with the factors estimated by principal component analysis. To forecast the inflation in Sweden we use 39 yearly macroeconomic time series for 1991-2013. We find that three estimated factors can make a good prediction.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail: Martina145@gmail.com. Supervisor: Martin Sköld and Jan-Olov Persson.

Preface

This thesis of 15 ECTS will lead to a Bachelors Degree in Mathematical Statistics at the Department of Mathematics at Stockholm University.

I would like to thank my supervisors Martin Sköld and Jan-Olov Persson, Department of Mathematic Statistics, for all the help and guidance through this thesis. I would also like to thank my friends for taking time to read and give constructive criticism.

Contents

1 Introduction **1**

2 Methods **2**
- 2.1 Principal components and factor models 2
- 2.2 Principal component analysis 3
- 2.3 The linear factor model 4
- 2.4 The dynamic factor model 4
  - 2.4.1 The forecasting model 5
  - 2.4.2 Estimation 6
  - 2.4.3 Determination of the number of factors 7
  - 2.4.4 h-step ahead forecasts 7
  - 2.4.5 Alternative methods for forecasting 7
- 2.5 AIC 7
- 2.6 R^2 8

3 Modeling **8**
- 3.1 Data 8
- 3.2 Prediction 8
- 3.3 Results 9

4 Discussion **10**

References **11**

Appendix **13**
- Data 13
- Tables 14
- Figures 14

1 Introduction

Economic forecasting of macroeconomic variables, such as inflation, has for many years been done by models with quite few variables. Today we have a lot of different time series that can contain relevant information. These time series of potential predictor variables (n) can be so many that they exceed the number of observations in the time series. What we want to do is to forecast one series using many predictor time series. According to [James H. Stock, 2006] this can be done by modeling the covariability of the series in terms of a relatively few number of unobserved latent factors.

The aim of this thesis is to find a model that can be used to forecast the inflation in Sweden. One way of doing this is to estimate a vector of factors from the predictor time series and use it in a dynamic factor model (DFM). We will make this estimate using principal component analysis. We then get the prediction model by regressing the inflation time series against these estimated factors. This regression is linear so we therefore get a forecast that is linear in its predictors.

In section 2 we will look into how principal components and factor models are connected, how to compute principal components and how the classical linear factor model is defined. We will also examine the dynamic factor model and how to estimate and forecast with it. Section 3 contains results from applying the preceding methods to forecast the Swedish inflation.

There are many studies that have applied these methods. For example [Kunovac, 2008] use 144 macroeconomic variables to construct a factor analysis model by PCA to forecast the inflation in Croatia. Their conclusion is that the factor model forecast is better than the autoregressive (AR) forecast which is the benchmark model. [Figueiredo, 2010] try to find forecast models that are better than models typically used by the monetary authorities (vector autoregressive models). The forecast of the inflation is done by using macroeconomic data from the Central Bank of Brazil. As alternative models they use factor models with PC and partial least squares (PLS). Their conclusion is that the factor model outperforms the other models. [Banerjee et al., 2008] compare both empirical, where they use macroeconomic variables of the Euro area and Slovenia, and simulated forecasting performance between factor models and the benchmark AR-forecast. Here they use relative short time series and their conclusion is that factor-based forecasts are good also in short samples. This is however not true for vary small samples. [Marcellino et al., 2003] aims at comparing forecasting models such as autoregressions, vector autoregressions (VAR) and DFM by forecasting the Euro-area variables real GDP, industrial production, price inflation, and the unemployment rate. Their conclusion is that within the multivariate methods the factor models are better than VARs.

2 Methods

2.1 Principal components and factor models

Principal component analysis (PCA) and factor analysis are closely related. Both are statistical techniques used to reduct the dimension of a set of variables. The aim of PCA is to reduce a large set of variables to a smaller set that still contains most of the information from the large set. The aim of factor analysis is to regroup the large set of variables into a limited set of clusters based on shared variance. Although the estimates of the two methods differ when n is small, we can see that when n increases the difference disappears. Because of this PCA can be used to estimate the factors in the DFM [James H. Stock, 2006].

Suppose we have p observations of correlated variables, PCA transforms these variables into a set of linearly uncorrelated variables, which is called principal components. Often a small number k of these principal components can account for most of the variability of the total set of variables [Johnson, 2014a].

The factor model analysis describes the covariance relationship among many variables in terms of a few underlying, unobservable, random quantities called common factors. The variables are thus grouped by their correlations, so within the groups the variables are highly correlated and all groups have small correlation with each other. Each group represent an underlying construct (factor) that is responsible for the observed correlations [Johnson, 2014b]. Any variance that is not explained by the factors are described by the residuals (error terms).

The first use of factor analysis was by psychologists who wanted to understand intelligence. Intelligence tests were performed and then factor analysis was used to analyse these tests, with the aim to establish if intelligence is made up of one or several factors measuring properties like mathematical ability. Other than in psychology factor analysis is often used in physical sciences and biological sciences [Chatfield, 1980].

2.2 Principal component analysis

Let $\boldsymbol{X}$ be a $p \times 1$ random vector with covariance matrix $\boldsymbol{\Sigma}$ and eigenvalues $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_p \geq 0$. Consider the linear combinations

$$Y_i = \mathbf{a}_i'\mathbf{X} = a_{i1}X_1 + a_{i2}X_2 + \cdots + a_{ip}X_p, \quad i = 1, \ldots, p.$$

Using that the linear combinations $\boldsymbol{Y} = \boldsymbol{a'X}$ have $Cov(\boldsymbol{Y}) = Cov(\boldsymbol{a'X}) = \boldsymbol{a'\Sigma a}$ we obtain

$$Var(Y_i) = \mathbf{a}_i'\boldsymbol{\Sigma}\mathbf{a}_i, \quad i = 1, 2, \ldots, p.$$
$$Cov(Y_i, Y_k) = \mathbf{a}_i'\boldsymbol{\Sigma}\mathbf{a}_k, \quad i, k = 1, 2, \ldots, p.$$

The matrix $\boldsymbol{a}$ contains the loadings of the observable vector $\boldsymbol{X}$ (also known as the eigenvectors of the covariance matrix $\boldsymbol{\Sigma}$), which tell us how the components in the new space relates to the initial variables. The principal components (also called the principal component scores, the observations coordinates in the new space) are those uncorrelated linear combinations $Y_1, Y_2, \ldots, Y_p$ whose variances are as large as possible [Johnson, 2014a]. Here the variance is the same as the corresponding eigenvalue. The first principal component is the one with maximum variance. That is the one who maximizes $Var(Y_1) = \mathbf{a}_1'\boldsymbol{\Sigma}\mathbf{a}_1$. In order to ensure uniqueness in the definition of principal components we additionally need the weight vectors $\mathbf{a}_i$ to have a unit form, therefor we set the condition $\mathbf{a}_i'\mathbf{a}_i = 1$. The principal components are now defined as:

$$\text{First principal component} = \text{linear combination } \mathbf{a}_1'\mathbf{X} \text{ that maximizes}$$
$$Var(\mathbf{a}_1'\mathbf{X}) \text{ given condition } \mathbf{a}_1'\mathbf{a}_1 = 1.$$

$$\text{Second principal component} = \text{linear combination } \mathbf{a}_2'\mathbf{X} \text{ that maximizes}$$
$$Var(\mathbf{a}_2'\mathbf{X}) \text{ given condition } \mathbf{a}_2'\mathbf{a}_2 = 1 \text{ and}$$
$$Cov(\mathbf{a}_1'\mathbf{X}, \mathbf{a}_2'\mathbf{X}) = 0.$$

i:th principal component = linear combination $\mathbf{a}_i'\mathbf{X}$ that maximizes

$Var(\mathbf{a}_i'\mathbf{X})$ given condition $\mathbf{a}_i'\mathbf{a}_i = 1$ and

$Cov(\mathbf{a}_i'\mathbf{X}, \mathbf{a}_k'\mathbf{X}) = 0$ for $k < i$.

2.3 The linear factor model

Let $\boldsymbol{X}$ be a $p \times 1$ vector of observable random variables with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. The classical factor analysis model can be written as

$$\boldsymbol{X} = \boldsymbol{L}\boldsymbol{F} + \boldsymbol{\varepsilon},$$

where $\boldsymbol{F}$ is a $m \times 1$ vector of common factors, $\boldsymbol{L}$ is a $p \times m$ matrix of factor loadings where the coefficient ℓ_{ij} is the loading of the ith variable on the jth factor. These factor loadings explain the relationship between the variables and the common factors. $\boldsymbol{\varepsilon}$ is a $p \times 1$ vector of errors.

We assume that $\mathbf{F}$ and $\boldsymbol{\varepsilon}$ are independent and that $E(\mathbf{F}) = \mathbf{0}, Cov(\mathbf{F}) = \mathbf{I}, E(\boldsymbol{\varepsilon}) = \mathbf{0}, Cov(\boldsymbol{\varepsilon}) = \Psi$, where Ψ is a diagonal matrix [Johnson, 2014b].

2.4 The dynamic factor model

Let X_t be a high dimensional vector of observed time series variables at time t. The dynamic factor model can be written as

$$X_{it} = \lambda_i(L)' f_t + u_{it}, \qquad i = 1, \ldots, n, \tag{1}$$

where f_t is a $q \times 1$ vector of (unobserved) latent dynamic factors, u_t is a vector of mean-zero idiosyncratic (unique for each time series) disturbances that might be serially correlated. Here we assume the factors (f_t) and the idiosyncratic disturbances (u_{it}) to be uncorrelated and stationary. $\lambda_i(L)$ is a $q \times 1$ vector lag polynomial, where L is the lag operator. When multiplying this vector with the time series vector f_t we get the same vector but moved backwards one time unit, i.e $\lambda_i(L)f_t = (\lambda_{i0}L^0 + \lambda_{i1}L^1 + \cdots + \lambda_{im}L^m)f_t =$

$\lambda_{i0} f_t + \lambda_{i1} f_{t-1} + \cdots + \lambda_{im} f_{t-m}$ [Cochrane, 2005]. The vector $\lambda_i(L)$ is called the "dynamic factor loadings".

2.4.1 The forecasting model

The forecasting equation for the variable to be forecast, Y_t, can be derived from (1),

$$Y_{t+1} = \beta(L)' f_t + \gamma(L)' Z_t + \varepsilon_{t+1}. \tag{2}$$

[James H. Stock, 2006] shows that this can be done by setting

$$Y_t = \lambda_Y(L)' f_t + u_{Yt},$$

where the last term is distributed independently of f_t and u_{it} for $i = 1, \dots, n$, and supposes that it follows the autoregression

$$\delta_Y(L) u_{Yt} = v_{Yt}.$$

Now we see that

$$\delta_Y(L) Y_{t+1} = \delta_Y(L)' \lambda_Y(L) f_{t+1} + v_{t+1}$$

and we get

$$Y_{t+1} = \delta_Y(L) \lambda_Y(L)' f_{t+1} + \gamma(L) Y_t + v_{t+1}$$

where $\gamma(L) = L^{-1}(1 - \delta_Y(L))$ is a lag polynomial. Hence

$$E(Y_{t+1} | X_t, Y_t, f_t, X_{t-1}, Y_{t-1}, f_{t-1}, \dots)$$

$$= E(\delta_Y(L) \lambda_Y(L)' f_{t+1} + \gamma(L) Y_t + v_{t+1} | Y_t, f_t, Y_{t-1}, f_{t-1}, \dots)$$
$$= \beta(L) f_t + \gamma(L) Y_t,$$

where

$$\beta(L) f_t = E(\delta_Y(L) \lambda_Y(L)' f_{t+1} | f_t, f_{t-1}, \dots).$$

Here

$$\varepsilon_{t+1} = v_{Yt+1} + (\delta(L) \lambda_Y(L)' f_{t+1} - E(\delta(L) \lambda_Y(L)' f_{t+1} | f_t, f_{t-1,\dots})),$$

which has, given X_t, f_t, Y_t and their lags, mean zero. Z_t includes all the laged Y_t but can also include other observable predictors.

2.4.2 Estimation

If $\lambda_i(L)$ and $\beta(L)$ have order p (finit order), then (1) and (2) can be written in the static form

$$X_t = \Lambda F_t + u_t, \tag{3}$$

$$Y_{t+1} = \beta' F_t + \gamma(L)' Z_t + \varepsilon_{t+1}, \tag{4}$$

where F_t is an $r \times 1$ factor vector that include the current and lagged values of the q dynamic factors, that is $F_t = [f_t' f_{t-1}' \dots f_{t-p+1}']$, $u_t = [u_{1t} \dots u_{nt}]$, Λ is a $n \times r$ matrix of factor loadings, and from the elements of $\beta(L)$ we get a $r \times 1$ vector of parameters β. The representation (3) and (4) is called the static form of the dynamic factor model (DFM) because F_t appears without any lags. If the dimension of F_t is r we have that $q \leq r \leq qp$. (3) implies that the r factors can explain almost all of the variation of the n variables (the difference is the disturbance). (4) implies that the one step ahead forecast of the variable Y_t is formed by the factors, lags of Y_t and a disturbance.

The fact that F_t and u_t are uncorrelated gives us $Cov(X_t) = \Lambda Cov(F_t)\Lambda' + Cov(u_t)$ which is recognized as the variance of the classical factor analysis.

If we assume that the eigenvalues are $O(1)$ for $Cov(u_t)$ and $O(n)$ for $\Lambda'\Lambda$ we get that the eigenvalues of $Cov(X_t)$ are $O(N)$ for eigenvalues $1 \dots r$ and $O(1)$ for the remaining eigenvalues. From this we can see that the principal components of X that could estimate Λ is components $1 \dots r$. From this Λ we can estimate F_t.

The principal components estimator of F_t is the solution to the nonlinear least-squares problem

$$min_{F_1,\dots,F_T,\Lambda} T^{-1} \sum_{t=1}^{T} (X_t - \Lambda F_t)'(X_t - \Lambda F_t)$$

subject to $\Lambda'\Lambda = I_r$. According to [James H. Stock, 2006] the solution to this is to set $\hat{\Lambda}$ to be the first r eigenvectors of the sample covariance matrix of X_t, $\widehat{Cov}(X_t) = T^{-1}\sum_{t=1}^{T} X_t X_t'$. The estimator of the factors is $\hat{F}_t = \hat{\Lambda}' X_t$. Here $\hat{F}_t$ is a vector with the first r principal components of X_t. $T^{-1}\sum_{t=1}^{T} \hat{F}_t \hat{F}_t'$ is a diagonal matrix with the largest r eigenvalues of $\widehat{Cov}(X_t)$. Thus $\hat{F}_t$ is the

PCA estimator of the factors in the DFM.

2.4.3 Determination of the number of factors

To determine how many factors that should be included in the model when n is large we can use model selection methods by information criterion. [James H. Stock, 2002] shows that this can be accomplished given an upper bound on the dimension and lags of F_t. When estimating with PCA we can also chose the number of eigenvalues that explain the most of the variance. [Bai and Ng, 2002] prove that r, the dimension of F_t, can be estimated using a certain information criteria.

2.4.4 h-step ahead forecasts

Here h is a constant specifying how many steps ahead we want to forecast. A direct h-step ahead forecasts of the variable Y_t are computed by regressing Y_{t+h}^h against $\hat{F}_t$, Y_t and their lags. Iterated h-step ahead forecasts can be computed, according to [James H. Stock, 2010], by first estimating a vector autoregressive (VAR) model for $\hat{F}_t$, then using this VAR together with the one-step ahead forecasting equation to iterate forward h periods.

2.4.5 Alternative methods for forecasting

Other methods of estimate the DFM is for example by maximum likelihood, dynamic principal components analysis and Bayes methods. The maximum likelihood estimation has been proven to be a good estimation when n is small but when n gets larger it becomes inefficiently because it takes to much time to compute. The dynamic principal component estimation has also been proven to be good but it can not be used for forecasting. Bayes methods are better than maximum likelihood estimation because it is much easier to compute but today we don't know enough to say that it is better than PCA when n is large.

2.5 AIC

The Akaike Information Criteria (AIC) is a model selection measure, which can be used to compare different models. The model we prefer is the one with the smallest AIC value. AIC is defined as

$$AIC = 2k - 2L,$$

where L is the maximized log likelihood for the estimated model with k parameters [Tsay, 2005].

2.6 R^2

R-square (R^2) is the proportion of the total variation in the data that the model explain and is thus a measure of fit. The model we prefer is the one with an R^2 value as close to one as possible. R^2 is given in procent and is defined as [Tsay, 2005]

$$R^2 = 1 - \frac{\text{Residual sum of squares}}{\text{Total sum of squares}}.$$

3 Modeling

3.1 Data

The data used to construct the factors are 39 yearly time series for 1991-2013. These macroeconomic time series were selected because they probably have an impact on the inflation and represent the macroeconomic variables: tax, wage, price index, real output, exchange rate, interest rate and money. The time series were obtained from SCB and Riksbanken (2015-02-26). For more information about the time series see Appendix. We transform all time series by taking the first difference and then standardize them to have mean zero and unit variance. An example of this can be seen in Figure 1 where the time series to the left is the original and the one to the right have been transformed.

3.2 Prediction

We will now use the previous explained methods to construct a model for the prediction of inflation in Sweden. We start by setting the time series as columns in a matrix (X_t) and compute the covariance matrix on which we can perform PCA. A summary of the first six principal components using the complete data set can be seen in Table 2. The factors we use are thus the scores in the PCA. We regress the inflation time series (Y_t) on the 1-lagged inflation time series (Z_t) and the 1-lagged scores from the principal component analysis. There are many different models that can be constructed by these principal components and the 1-lagged inflation time series. Although we should choose the first r principal components, we will examine another approach as well. We will compare the models where we in Model 1 choose the first $1, 3$ and 6 principal components and in Model 2 choose some $1, 3$ and 6 principal components. To help us decide which we should include in the model we use a model selection tool in R. In particular we will use the leap package and the function regsubsets which selects the best models with respect to R^2, residual sum of squares, adjusted R^2, Mallows' C_p or BIC. Like in Section 2.2.2 we have the model

$$\hat{Y}_{t+1} = \hat{\beta}_1 \hat{F}_t + \hat{\beta}_2 Z_t + \varepsilon_{t+1},$$

where in Model 1 we have $\hat{F}_t = (c1, c2, c3, c4, c5, c6)$, $\hat{F}_t = (c1, c2, c3)$ and $\hat{F}_t = (c1)$ and in Model 2 we have $\hat{F}_t = (c2, c3, c5, c6, c7, c8)$, $\hat{F}_t = (c2, c3, c8)$ and $\hat{F}_t = (c2)$ where $c1$ is the 1-lagged first principal component, $c2$ the 1-lagged second principal component etc.

3.3 Results

To get some sort of idea if the models are good we will compare it to the following benchmark model

$$\hat{Y}_{t+1} = Y_t, \tag{5}$$

that is, we suppose that the inflation will not change over the year. The comparison will be between the mean squared errors (MSE), that is the sums

$$\frac{1}{T}\sum_t (Y_t - Y_{t-1})^2 \tag{6}$$

$$\frac{1}{T}\sum_t (Y_t - \hat{Y}_t)^2 \tag{7}$$

where (7) belongs to the prediction using Model 1 and Model 2 and (6) to the prediction using the benchmark model (5).

We can see in Table 1 that regarding the MSE, all models are better prediction models than the benchmark model. The preferable model are Model 1 with 6 PCs, although we want to have a model with as few factors as possible. The R^2 is similar and about 90% in all models but Model 1: 1 PC, which also is the only one that have an positive AIC. Variation explained refer to the total variation of the original data set explained by the principal components. We can see that the factors in Modell 1 explain more of the variation than the factors in Model 2.

	Modell 1			Modell 2			Benchmark
	6 PC	3 PC	1 PC	6 PC	3 PC	1 PC	model
R^2	0.8918	0.8791	0.0961	0.9398	0.9204	0.8198	-
AIC	-22.48	-26.16	12.09	-34.79	-34.94	-21.78	-
Var. exp.	ca 90%	ca 70%	ca 35%	ca 50%	ca 35%	ca 30%	-
MSE	0.0330	0.2649	1.3397	0.3460	0.4002	0.3731	2.3600

Table 1: R-squared, AIC, variation explained (var. exp.) for Model 1 and Model 2 with 6, 3 and 1 principal components and mean squared error for all models.

4 Discussion

As this thesis do not compare different models it is hard to have a direct conclusion about the model. All models tested were better than the benchmark model, but then again, the benchmark model was not a realistic model. If we look at Figure 2-5 we can conclude that the prediction is better in the case with 6 PCs. Model 1: 1 PC should not be used to forecast. As we want a simple model the preferable model would be Model 1 with 3 PC, because it is easy to find the factors and make the prediction model and we also have a much smaller model than we started with.

According to other studies the DFM seems to outperform other similar models, such as VA and VAR. In our empirical studies we only look at the DFM but the results seem to confirm their conclusions.

To get better precision on the predictions we could use a lot more time series, as we used very few in this studie, and for a longer time period. It would be interesting to compare the DFM models we get with some VA and VAR as well to see if our DFM is in fact better than the models often used today with the purpose to predict macroeconomic variables.

References

[Bai and Ng, 2002] Bai, J. and Ng, S. (2002). *Econometrica: journal of the Econometric Society. Volume 70, Issue 1. Determining the Number of Factors in Approximate Factor Models*. Chicago, Ill.

[Banerjee et al., 2008] Banerjee, A., Marcellino, M. G., and Masten, I. (2008). Forecasting macroeconomic variables using diffusion indexes in short samples with structural change. *Vol*.

[Chatfield, 1980] Chatfield, Christopher & Collins, A. J. (1980). *Introduction to Multivariate Analysis*. London: Chapman and Hall.

[Cochrane, 2005] Cochrane, J. H. (2005). Time series for macroeconomics and finance. *Manuscript, University of Chicago*, pages 11–12.

[Figueiredo, 2010] Figueiredo, F. M. R. (2010). Forecasting brazilian inflation using a large data set. *Central Bank of Brazil Working Paper*, (228).

[James H. Stock, 2002] James H. Stock, M. W. W. (2002). *Journal of Business and Economic Statistics Volume 20, 2:nd ed. Macroeconomic Forecasting using Diffusion Index*. Taylor and Francis, Ltd. Washington: American Statistical Association.

[James H. Stock, 2006] James H. Stock, M. W. W. (2006). *Handbook of Economic Forecasting Volume 1, 1st ed. ; Chapter 10, Forecasting with Many Predictors. Edited by Elliott, Graham, Granger, C. W. J. & Timmermann, Allan (red.)*. Elsevier North-Holland, Amsterdam.

[James H. Stock, 2010] James H. Stock, M. W. W. (2010). *The Oxford Handbook of Economic Forecasting, Dynamic Factor Models. Edited by Clements, Michael P. & Hendry, David F. (red.)*. New York: Oxford University Press.

[Johnson, 2014a] Johnson, Richard A. & Wichern, D. W. (2014a). *Applied multivariate statistical analysis; Chapter 8*. Six edition, Pearson New International edition.

[Johnson, 2014b] Johnson, Richard A. & Wichern, D. W. (2014b). *Applied multivariate statistical analysis; Chapter 9*. Six edition, Pearson New International edition.

[Kunovac, 2008] Kunovac, D. (2008). Factor model forecasting of inflation in croatia. *Financial Theory and Practice*, 31(4):371–393.

[Marcellino et al., 2003] Marcellino, M., Stock, J. H., and Watson, M. W. (2003). Macroeconomic forecasting in the euro area: Country specific versus area-wide information. *European Economic Review*, 47(1):1–18.

[Tsay, 2005] Tsay, R. S. (2005). *Analysis of Financial Time Series; Second edition*. John Wiley and Sons, Inc.

Appendix

Data

Price index:
Property Price Index for Rural, 1992=100, Sweden
Property Price Index for permanent smal houses (1981=100), Sweden
Property Price Index for secondary residence (1981=100), Sweden
Property Price Index for permanent houses, 1990=100, Sweden
Building Price Index for housing (BPI), inkl value added tax, 1968=100 apartment buildings, function indices
Building Price Index for housing (BPI), inkl value added tax, 1968=100 apartment buildings, quality indices
Factor price index (FPI) for residences inkl, wage adjustment, 1968=100 apartment buildings
Factor price index (FPI) for residences exkl, wage adjustment och value added tax, 1968=100 apartment buildings
Consumer price index (CPI)/Cost-of-living index, july 1914=100
Consumer price index (CPI) annual average overall, shadow index, 1980=100
Consumer price index (CPI) constant taxes, annual average, 1980 = 100
Producer Price Index (PPI), 2005 = 100 by product group SPIN 2007
Import price index (IMPI), 2005=100 by product group SPIN 2007
Export price index (EXPI), 2005=100 by product group SPIN 2007
Home sales price index, 2005=100 by product group SPIN 2007

Real output:
Completed apartments
Completed houses, number of room units
Reconstruction of apartment buildings, supplementation of apartments
Imports total, mskr
Export total, mskr
Net trade, mskr

Wage:
Average personal monthly wage, skr, all levels of education
Swedens population
Earnings for all full-year and full-time employees for persons 20-64 years, average, tsek / per cent of earnings
Earnings for all full-year and full-time employees for persons 20-64 years, median, tsek
Total income for residents in Sweden 31/12, 16+ years, average income tskr
Total income for residents in Sweden 31/12, 16+ years, median income tskr
Total income for residents in Sweden 31/12, 16+ years, total income mnskr

Tax, exchange rate, interest rate and money:
Swedish TCW index
Price base amount , skr
Tax rate, total local
Currencies: USD against Swedish kronor
Currencies: EUR against Swedish kronor
International Government Bonds, maturity 5 years: US
International Government Bonds, maturity 5 years: JP
Mortgage Bonds, 5Y
Swedish Government Bonds, 5Y
Treasury Bills, SE 6M
STIBOR (Stockholm interbank offered rate), 6M

Tables

	PC1	PC2	PC3	PC4	PC5	PC6
Standard deviation	3.2959	2.7757	2.1024	1.91527	1.67066	1.41570
Proportion of Variance	0.2785	0.1976	0.1133	0.09406	0.07157	0.05139
Cumulative Proportion	0.2785	0.4761	0.5894	0.68347	0.75504	0.80643

Table 2: Summary of the first six principal components for the total data set.

Figures

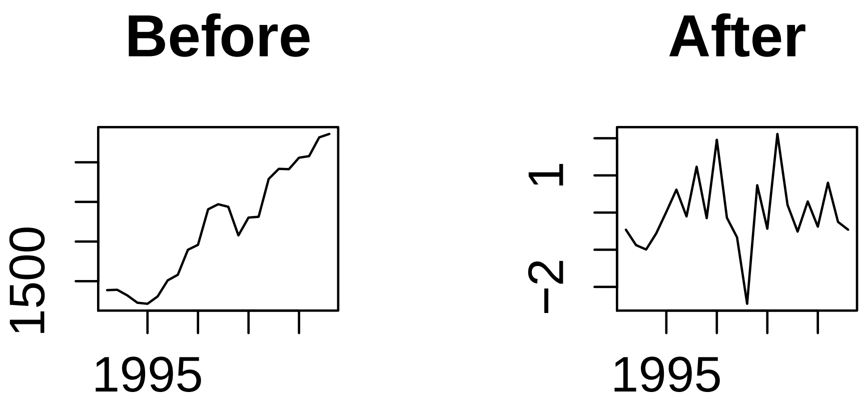

Figure 1: One of the 39 time series before and after taking the first difference and standardizing.

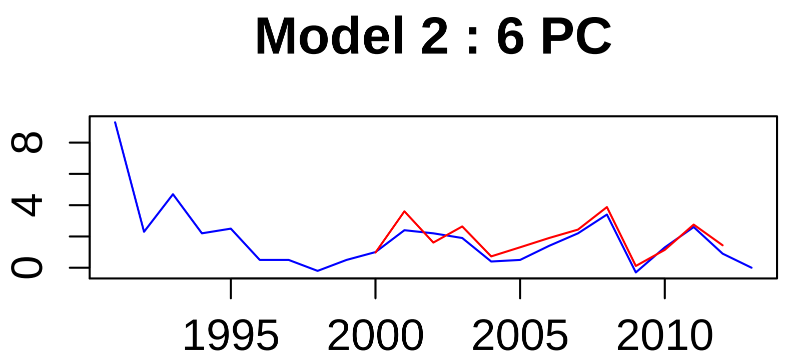

Figure 2: The inflation time series in blue and the predicted inflation in red with Model 2, six principal components.

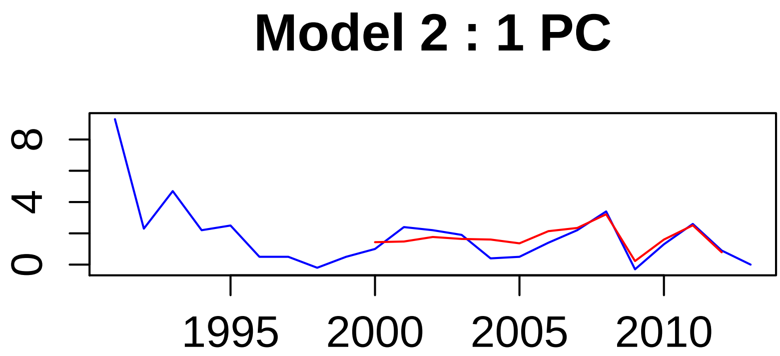

Figure 3: The inflation time series in blue and the predicted inflation in red with Model 2, one principal component.

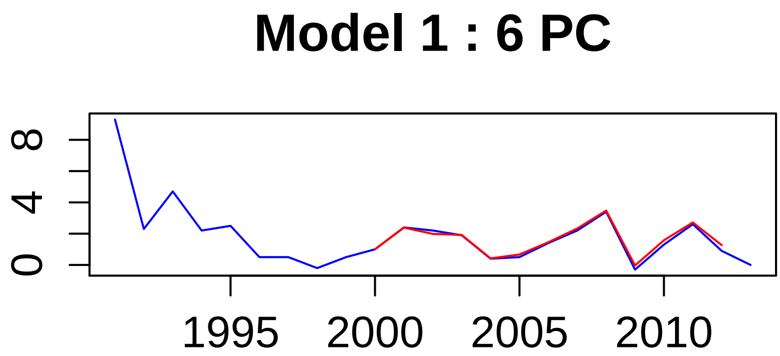

Figure 4: The inflation time series in blue and the predicted inflation in red with Model 1, six principal components.

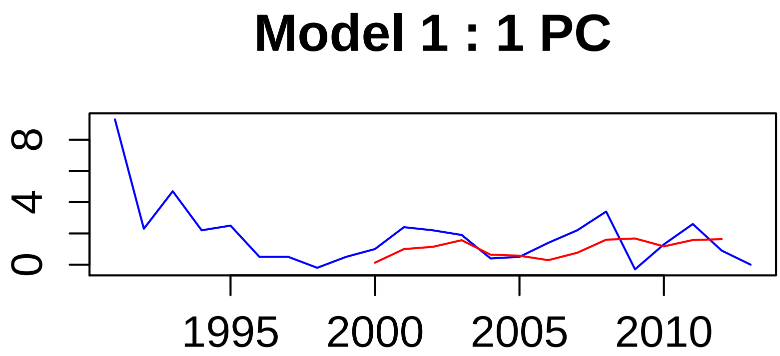

Figure 5: The inflation time series in blue and the predicted inflation in red with Model 1, one principal component.