Bayesian Principal Component Analysis of Interest Rates

IFoA Bayesian Methods Working Party
C. J. Paine

27th February 2025

Abstract

Under the classical model we estimate model parameters using historical data. The Bayesian approach nests the classical approach as a special case. The Bayesian framework facilitates systematic inclusion of 'expert judgements' via prior distributions . In this paper a PCA model, of the term structure of interest rates, is extended to use a Bayesian framework. This practical example demonstrates how a classical model with no adjustment would have consistently overestimated UK interest rates during the prolonged low-rate period following the financial crisis whereas a Bayesian approach that incorporates additional 'information' doesn't.

1 Introduction

Bishop [2020]

2 Principal Component Analysis

2.1 The model

The term structure of interest rates may be represented as a vector r in an m dimensional space. Each dimension $i \in \{1, 2, \ldots, m\}$ corresponds to a interest rate $\mathbf{r_i}$ at a specific maturity term.

$$\mathbf{r} = (\mathbf{r_1}, \mathbf{r_2}, ..., \mathbf{r_m})$$

Principal component analysis linearly transforms the vector $\mathbf{r}$ into a new co-ordinate system defined by principal component vectors $v_1, v_2, v_3, ..., v_m$. Consequently, the term structure may instead be reinterpreted as a linear combination of principal components:

$$\mathbf{r} = \bar{\mathbf{r}} + c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_m\mathbf{v}_m$$

where:

$\bar{\mathbf{r}}$ is an m dimensional vector of average interest rates

In practice we do not use all m principal components since PCA is a dimensionality reduction technique. Instead we select k principal components where $k \ll m$

$\mathbf{r} \approx \bar{\mathbf{r}} + c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_k\mathbf{v}_k$

Where:

$\mathbf{v_j}$ are m dimensional vectors of principal component $j \in \{1, 2, \ldots, k\}$

c_j are m are co-ordinates along the principle component axes representing data points under the transformed system

k is the number of principal components retained.

2.2 Principal Component Extraction

The principal components v_j are the eigenvectors of the covariance matrix of multiple term structure observations.

The observed co-ordinates along the principle component axes c_j (i.e. a transformed dataset) are derived using the original (untransformed) observations and the calculated principle component eigenvectors.

2.2.1 Calculate Covariance Matrix

Consider N observations of the term structure r we can denote each observation as $\mathbf{r}(\mathrm{t})$ where $t \in \{1, 2, \ldots, N\}$. We calculate the covariance matrix Σ as follows:

$\Sigma = \frac{1}{N-1}\sum_{t=1}^{N}(\mathbf{r}(\mathrm{t}) - \bar{\mathbf{r}})(\mathbf{r}(\mathrm{t}) - \bar{\mathbf{r}})^T$

insert formula for calculating r bar when multiple observations

2.2.2 Eigenvalue Decomposition

The eigenvectors of the covariance matrix are found by solving: $\Sigma\mathbf{v}_i = \lambda_i\mathbf{v}_i$

where λ_i are the eigenvalues of the covariance matrix.

2.2.3 Project original data onto principle components

The eigenvectors $\mathbf{v}_i$ facilitate expression of the original dataset $\mathbf{r}(\mathrm{t})$, $t \in \{1, 2, \ldots, N\}$ in terms of the principal components.

For a given observation at time t of the term structure $\mathbf{r}(\mathrm{t})$ we deduct the mean vector $\bar{\mathbf{r}}$ to obtain a dataset centered around zero.

$\mathbf{r}' = \mathbf{r}(\mathrm{t}) - \bar{\mathbf{r}}$

The observed co-ordinates $c_{i,t}$ on the principle component axes v_i are calculated as:

$$c_{j,t} = (\mathbf{r}(\mathrm{t}) - \bar{\mathbf{r}}) \cdot \mathbf{v}_i = r' \cdot \mathbf{v}_i$$

$c_{j,t}$ are also referred to as scalar coefficients (or principal component scores) for principal component j and observation t.

3 Modelling Changes in the Term Structure

3.1 Term Strucutre Generation vs Evolution

Insurers are more often concerned with modelling the change in a term structure of interest rates, given an initial starting point, rather than simulating the whole curve from first principles. For example, under solvency UK legislation, insurers are prescribed a risk free yield curve from which to model the changes over a 1 year time horizon.

3.2 Measure of Interest Rate Changes

Denote $\mathbf{s}(\mathrm{t})$ as the change in the term structure over the time period $(t-1, t)$. $\mathbf{s}(\mathrm{t})$ is a function of $\mathbf{r}(\mathrm{t})$ and $\mathbf{r}(\mathrm{t\text{-}1})$

We are presented with a choice of measures when modelling the evolution of interest rates.

Measure	Formula for $\mathbf{s}(\mathrm{t})$
Absolute Changes	$r_i(t) - r_i(t-1)$
Relative Changes	$\frac{r_i(t)}{r_i(t-1)}$
Log Changes	$\log r_i(t) - \log r_i(t-1)$

Table 1: Different Measures

where $r_i(t)$ is the interest rate for maturity i at time t

Each choice of measure has its advantages and disadvantages. For instance, relative changes are problematic when rates are close to zero since proportional changes will be very large.

For the purposes of the example that follows we consider log changes. Even though they are less intuitive, log changes are comparable against different levels of interest rate.

3.3 Adapting the model to measure differences

The principle component model is indifferent as to what the underlying dataset represents. It formulation is therefore inconsequential to whether we are modelling nominal rates levels or differences.

we may want to reference lags and the use of overlapping data in order to increase datapoints

We can respecify our principle component model as:

$\mathbf{s(t)} = (\mathbf{s_1(t)}, \mathbf{s_2(t)}, ..., \mathbf{s_m(t)})$

where

$s_i(t)$ is the difference in interest rates over the period $(t - t, t)$ i.e $r_i(t) - r_i(t-1)$

$\bar{\mathbf{s}}$ is an m dimensional vector of average interest rate differences

$$\mathbf{s} \approx \bar{\mathbf{s}} + c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_k\mathbf{v}_k$$

$$\Sigma = \frac{1}{N-1}\sum_{t=1}^{N}(\mathbf{r}(\mathrm{t}) - \bar{\mathbf{r}})(\mathbf{r}(\mathrm{t}) - \bar{\mathbf{r}})^T$$

3.3.1 Dataset

Assume we have N observations of the term structure $\mathbf{r}(\mathrm{t})$ where $t \in \{1, 2, \ldots, N\}$ and that we wish to model the change in interest rates from now, time N, to time $N + 1$.

We have an m x N matrix of historical observations $\mathbf{r_i(t)}$

We have a m x (N-1) matrix of historical yield differences $s_i(t) = r_i(t) - r_i(t-1)$

$\mathbf{s}(\mathrm{N+1}) = \mathbf{r}(\mathrm{N+1}) - \mathbf{r}(\mathrm{N})$
We again have an m dimensional row vector $\mathbf{s} = (\mathbf{s_1}, \mathbf{s_2}, ..., \mathbf{s_m})$

4 Related Work

Discussion of related work here.

5 Methodology

Details of your methodology here.

5.1 Determining Change in Yields from Simulated Principal Components

6 Determining Priors

Consistently positive forecast errors in the classical model, combined with knowledge of changes in central bank policies, provides insights suggesting that low interest rates are likely to persist. These insights can have be incorporated into priors, allowing for a more forward-looking assessment.

We will add back the mean but perhaps there is a financial theory that suggests the mean should be zero over time.

7 Results

Presentation of results here.

8 Conclusion

The Bayesian approach nests the classical approach as a special case. Sekerke [9999]

Expert judgements are a necessary part of the modelling process. Incoporating them in a systematic way using bayesian framework is advantageous because ...

9 Additional investigations

This section will not appear in the final paper. It is just a dumping ground for my personal thoughts:

- produce questionnaire to ask insurers if they use bayesian inference in their modelling
- STAN and BMRS
- a comparison of classical model with ad hoc adjustments against bayesian method with systematic adjustments.
- does the fact that the use case for insurers, involving a given interest rate curve, mean that it introdcues some kind of bayesian element anyway.
- terminology concerns
- linear exposures
- empirical risk drivers

References

Christopher M Bishop. Bayesian pca. 2020.

Matt Sekerke. Bayesian risk management, 9999.

Random Notes

These notes are not part of the actual paper but just a dumping ground while work is in progress.

10 Data

Consider a dataset of historical spot rates $S(t, m)$ where the time index t that details the date of each set of observed spot rates at differing maturities m

$S(t, m)$ where t is the time index and m is the maturity year

11 choice of number of principal components

A central concern in PCA is the number of principal components to retain. Bishop [2020]

explain we will use three as is standard practice and explains x percent using eigenvalues to give us this percentage.

12 nailing the terminology of empirical co-ordinates

Different terminologies used to refer to c_j reflect different perspectives. **co-ordinate** emphasises

that c_j represents a position in the transformed co-ordinate system. (the principal component space). After transformation, the datapoint is expressed as a linear combination of the eigenvectors. c_j is the co-ordinate along v_j **Principal Component Score** 'score' is a common term in statistics

to mean value of data point. In this context a score is the value of the value of the data point along a principal component axis. i.e. how much a given observation loads onto that principal component. **Scalar Coefficient** it views the data point as a coefficient in the expansion of the

eigenvector basis.... the term scalar makes sense because we are talking about numerical values data points. coefficients is confusing but they are indeed coefficients of the eigenvector...

13 principal component basis vectors

another term for principal components, or eigenvectors

14 principal component axis

15 level, slope and curvature vs eigenvector values

link to chatgpt

16 inner product and outer product

inner products and outer products are fundamental concepts in linear algebra. inner product is aka dot product. it is the projection of 1 vector onto another

an outer product is different in that it produces a matrix not a scalar

17 normalising dataset

is it necessary to normalise movements i.e. have centered around zero.

18 soundbites

only a few principal components explain most of the variance, we can approximate the data using a lower-dimensional representation taking a log transformation stabilises variance

perhaps nabsolute changes deal with negatives more easily.