KUMAR, RAKESH

Working Paper

Towards a Deeper Understanding of Yield Curve Movements

Towards a Deeper Understanding of Yield Curve Movements

Rakesh Kumar
Bloomberg L.P.
rkumar21@bloomberg.net

Abstract

From a modeling perspective, daily movements in risk-free yield curves deserve further exploration. In this paper, I study daily yield curve changes for three major liquid sovereign yield curves, namely US, UK and Germany. I focus on two main themes (1) characterizing the statistical distribution of the multivariate vector of daily yield curve changes, and (2) exploring efficiency of deep learning based dimensionality reduction approaches for the daily yield curve changes.

Specifically, I am able to reject the null hypothesis that daily yield curve changes are normally distributed. I also show that a multivariate student-t distribution with 2 degrees of freedom or a gaussian mixture model are more accurate at capturing the leptokurtic distribution of daily yield curve changes.

In addition, on the dimensionality reduction front, I show that a deep neural network autoencoder outperforms a PCA based linear decomposition approach.

1 Introduction

Yield curve modeling is of considerable practical interest. A number of works [Cochrane and Piazzesi, 2009], [Nelson and Siegel, 1987], [Diebold and Li, 2006] have explored yield curve models from modeling the term structure of yield curve levels.

At the same time, there is not much published work around exploring two fundamental questions:

- Can we characterize the statistical distribution of daily changes in yield curve? Specifically, is normal distribution a good approximation for such a multivariate random variable.
- How do the linear dimensionality reduction methods (such as PCA) stack up with more general non-linear approaches in exploring latent randomness in yield curve movements?

In this paper, I show that daily yield curve changes display significant excess kurtosis and therefore are not normally distributed. Further, I show that a good approximation to yield curve changes is a multivariate student-t with $\nu = 2$ degrees of freedom or a gaussian mixture model with $K = 10$ components.

Yield Curve	Period	Samples
US	02/77-07/19	11079
UK	02/96-07/19	6126
Germany	01/95-07/19	6422

Table 1: Yield Curve Data

The second focus of the paper is on studying dimensionality reduction for yield curve movements. On this front, the closest related work to this paper is [Barber and Copper, 2012]. The authors use PCA to examine the significance of covariance matrix eigenvalues for characterizing latent number of dimensions in yield curve movements for a given threshold. This still leaves an important question unanswered, namely, how efficient are linear decomposition approaches compared to more general non-linear methods?

In this paper, I show that for dimensionality reduction of yield curve movements, linear decomposition approaches (e.g. PCA) underperform the more general non-linear decomposition methods. In particular, I compare PCA to a deep autoencoder neural network - a proxy of generalized non-linear dimensionality reduction methods. I show that deep autoencoder approach outperforms PCA in out -of-sample testing.

1.1 Data

This paper uses daily yield curve data for U.S, U.K and Germany. This yield curve is characterized by $k = 8$ tenors (1y,2y,3y,5y,7y,10y,20y,30y). Table 1 show the time period spanning the observations and the number of observations in my dataset. Note germany yield curve data corresponds to pre-euro time period as well.

For the rest of this paper, I use the notation of matrix $X_{n,p}$ to denote the dataset for daily change in yield curve for a particular country. Here, n is the number of observations of the daily change in yield curve and $p = 8$ is the dimensionality of the yield curve. Therefore, each row of this matrix, $x \in \mathbb{R}^p$ corresponds to the day over day change in the respective sovereign yield curve.

2 Distribution of Yield Curve Changes

From a risk management and general financial modeling perspective, it is of keen interest to characterize the statistical distribution of yield curve changes.

In this section, first I show that the daily yield curve change data does not follow a normal distribution. This is primarily due to highly leptokurtic nature of the underlying data samples. Next, I propose two alternatives to a normal distribution: (a) multivariate student-t (b) Multivariate gaussian mixture model. I also show the calibrated values for data for these distribution families.

Let us first develop an intuition of the fat-tailed nature of the observed data distribution. To this end, I compare probability that a data point from the observed data sample $x \in \mathbb{R}^8$ is 'larger' than expected. Specifically, I compare the observed empirical probability of a data point $P_{data}(\|x\| > d)$ for some $d > 0$ to what we would expect from a normal distribution $P_{norm}(\|x\| > d)$ with same mean and covariance as the target data distribution.

To this end, I normalize the observed data matrix X to that of zero mean and unit diagonal covariance with the following transformations.

$$X_m = X - \frac{1}{n}(\mathbb{1}\mathbb{1}^T X)$$

$$\Sigma = \frac{(X_m^T X_m)}{n} = LL^T$$

$$X_c = X_m L^{T^{-1}}$$

Here, LL^T is the cholesky decomposition of the covariance matrix Σ. As a result of the above linear transformations, X_c is a zero mean and diagonal unit covariance matrix. We now draw samples x from X_c and calculate probability ratios w.r.t a standard normal distribution. These results of probability ratio for the three daily yield curve changes are shown in figure 1. The figure shows the leptkourtik nature of the empirical distribution. At $d > 4$, $P_{data}(\|x\| > d)$ is significantly greater than probability implied by normal distribution $P_{norm}(\|x\| > d)$.

2.1 Multivariate Normality Tests

A multivariate Normal distribution is characterized by its mean and covariance matrix. In case of a symmetric positive definite covariance matrix $\mathbf{\Sigma}$, the pdf is given by

$$f_{\mathbf{x}}(x_1, \ldots, x_k) = \frac{\exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)}{\sqrt{(2\pi)^k|\mathbf{\Sigma}|}}$$

I use Mardia [Mardia, 1970] and Henze-Zirkler [Henze N, 1990] tests to test the null hypothesis that daily yield curve changes are normally distributed. Table 2 shows the results for the various yield curves. I conjecture that the main reason for rejection of null hypothesis is due to the leptokurtic nature of the observed data distribution. This motivates search for more appropriate distributions. In the rest of this section, we explore two such distributions: (a) Multivariate Student-t and (b) Multivariate Gaussian Mixture Model.

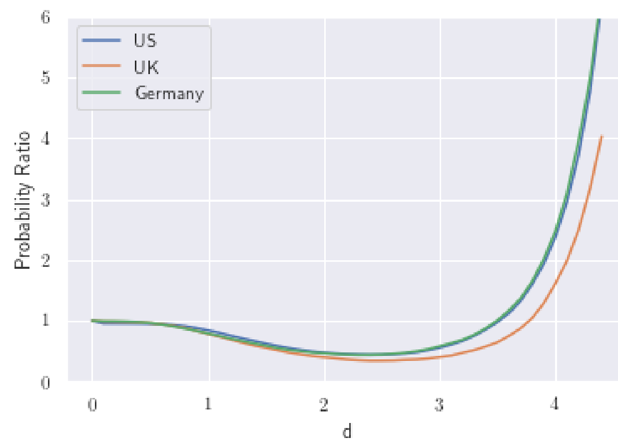


Figure 1: Leptokurtic Distribution. For a given data point $x \in \mathbb{R}^8$ and a scalar $d > 0$, the chart compares the ratio of observed data probability $P_{data}(\|x\| > d)$ to probability of a standard normal distribution $N(0, I)$, $P_{norm}(\|x\| > d)$. Daily yield curve changes for all three countries display high degree of leptokurtic behavior compared to that of a multivariate normal distribution with identical mean and covariance.

Multivariate Normal Test		
Yield Curve	Mardia	Henze-Zirkler
US	No	No
UK	No	No
Germany	No	No

Table 2: Multivariate Normality Tests for the daily yield curve changes

Yield Curve	Normal	Student-t	Mixture
US	16.7	19.2	20.2
UK	21	24.5	24.6
Germany	20.8	24	24.6

Table 3: Mean log likelihood for different Yield curves with various model choices

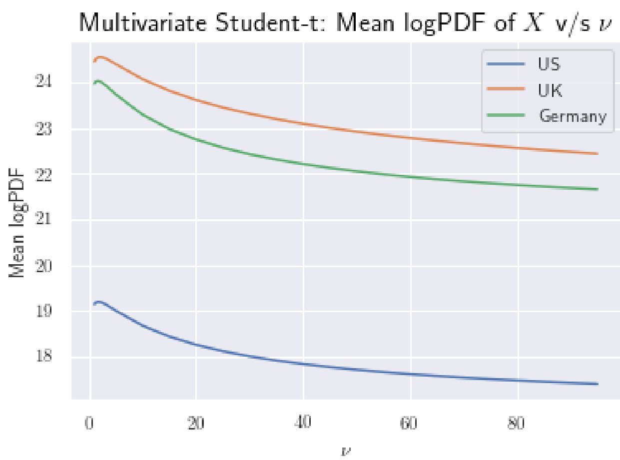


Figure 2: Multivariate Student-t: Mean log likelihood for different ν degrees of freedom showing presence of no local minimas. Also, the fat tailed nature of the observed data is poorly explained by large values of ν.

2.2 Multivariate Student-t: A more realistic choice

Student-t distribution is a family of distributions parameterized by degree of freedom parameter ν, mean μ and covariance matrix Σ. The pdf is given by

$$\frac{\Gamma\left[(\nu+d)/2\right]}{\Gamma(\nu/2)\nu^{d/2}\pi^{d/2}\left|\mathbf{\Sigma}\right|^{1/2}}\left[1+\frac{1}{\nu}(\mathbf{x}-\boldsymbol{\mu})^T\mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right]^{-(\nu+d)/2}$$

Student-t distribution family is a rich set of distributions with multivariate normal being a special case with $\nu \rightarrow \infty$.

I fit daily yield curve change data to the multivariate student-t distribution. To this end, I use the expectation maximization algorithm as outlined in [Liu and Rubin., 1995]. For a given ν, I estimate the μ and Σ parameters and subsequently calculate the mean log likelihood for the samples in data matrix X.

EM based estimation procedures may not always lead to global optimal fits. Figure 2 shows the mean log likelihood for different choices of degrees of freedom ν. Due to the leptokurtic nature of the observed data distribution, large values of ν lead to poor fits.

2.3 Gaussian Mixture Model Approach

Another rich family of distributions is of Gaussian Mixture models. A gaussian mixture is characterized by a linear combination of component gaussian distributions.

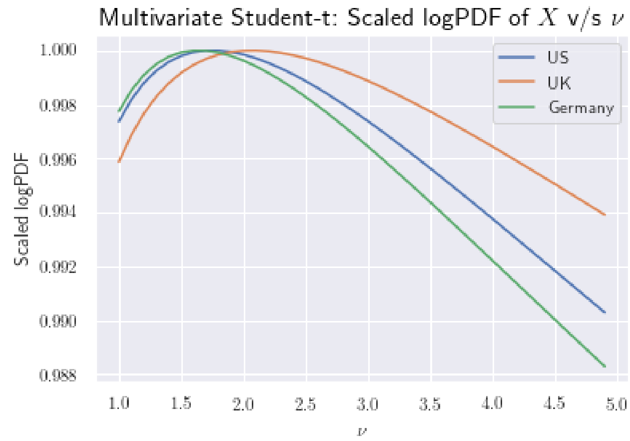


Figure 3: Scaled logPDF for different ν degrees of freedom for the yield curves. A value of $\nu = 2.5$ is a decent choice. This figure is zoomed in and scaled version of 2

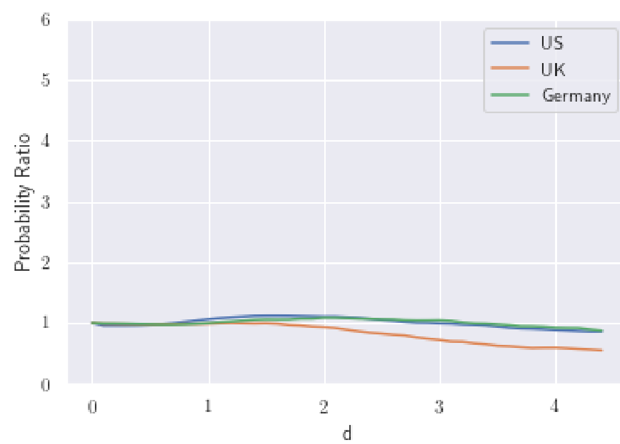


Figure 4: Leptokurtic. Using a student-t distribution fit with $\nu = 2$.

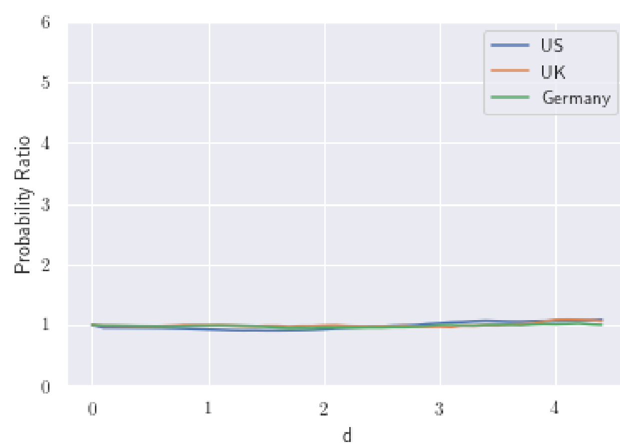


Figure 5: Leptokurtic. Using a GMM with $K = 10$ mixture components.

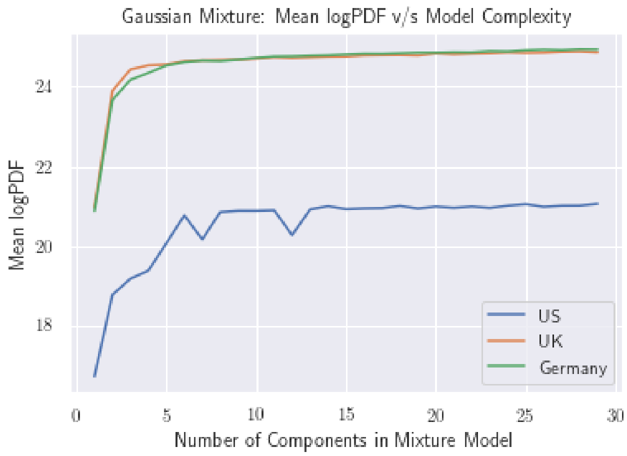


Figure 6: Mean log likelihood for different K components of gaussian mixture model.

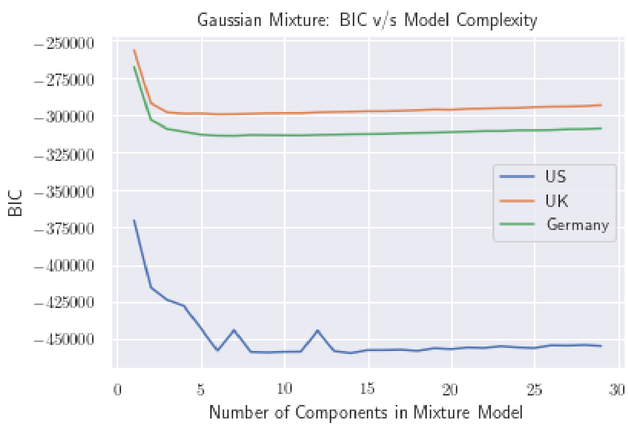


Figure 7: BIC (Bayesian Information Criterion) for different K components for gaussian mixture model.

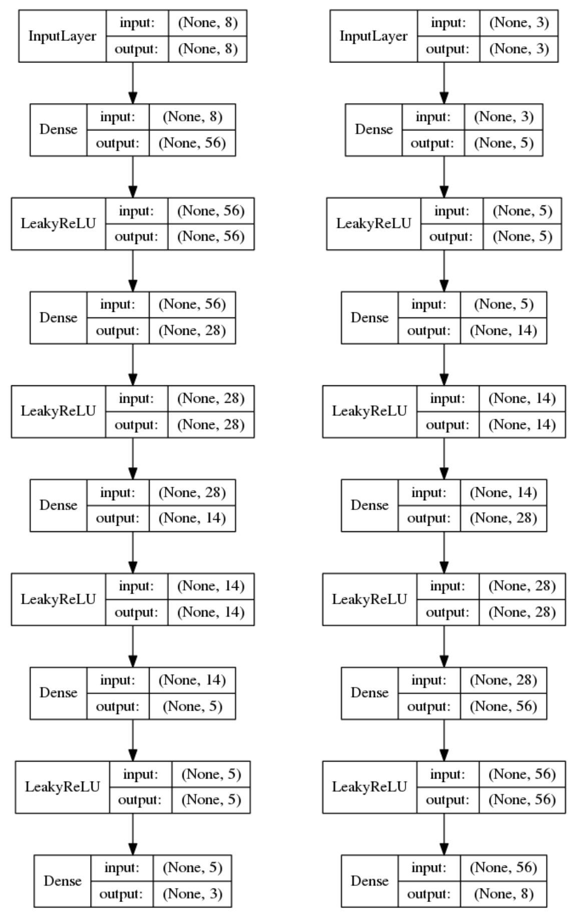
$$p(\boldsymbol{x}) = \sum_{i=1}^{K} \phi_i \mathcal{N}(x|\boldsymbol{\mu_i}, \boldsymbol{\Sigma_i})$$

In figure 6 I show the mean log likelihood for the data matrix X while varying the number of mixture components K. The mean log likelihood obtained using gaussian mixture model is superior to obtained using Student-t distribution.

A larger value of K allows higher accuracy in fitting an observed data distribution. In interest of parsimony, I use the Bayesian information criterion (BIC) [Schwarz, 1978] to explore appropriate choice of K. As shown in Figure 7, BIC fitting criterion suggests a 10 component mixture model as a good choice for modeling daily changes in the three yield curves.

2.4 Discussion of Model Choice

So far, I analyzed three different model choices: Multivariate Normal, Student-t distribution and Gaussian Mixture for



(a) Encoder (b) Decode

Figure 8: Encoder and Decoder architecture for the Autoencoder

fitting the daily changes in the yield curves. In table 3 I summarize the fitting results from the three different modeling choices for mean log likelihood. From the results, it is clear to see that a gaussian mixture with $K = 7$ components is the best choice. A multivariate student-t with $\nu = 2$ degrees of freedom falls in the middle compares well to the gaussian mixture model as well.

I also study whether the gaussian mixture model and multivariate student-t are able to capture the leptokurtik nature of the observed data distribution. The results are shown in Figures 5 and 4. The gaussian mixture model is well able to capture the fat-tailed observed distribution.

3 Dimensionality Reduction

This section is focused on studying latent dimensionality of the daily changes in the yield curve data. The general goal of dimensionality reduction is to seek to express $x \in \mathbb{R}^p$ with another vector $z \in \mathbb{R}^q$ with $q < p$ so as to minimize $\|x - T(z)\|$ where $T : \mathbb{R}^q \to \mathbb{R}^p$. The underlying idea is to express the latent randomness in the higher dimensional x with fewer dimensions expressed through z. This topic is of considerable practical interest. This is because a compact representation of $T(.)$ for the yield curve allows for more efficient forecasting and modeling of the latent driver of the randomness in the observed data.

The goal of this section is to compare and contrast state of art non-linear dimensionality reduction methods with classical linear approach (PCA).

3.1 PCA

Here T is constrained to be a linear orthogonal transformation. In other words, we want to minimize $\|x - Wz\|$ where $W \in \mathbb{R}_{p,q}$ and $W^T W = I$.

PCA has been a standard workhorse for dimensionality reduction of yield curves. Popular yield curve models such as [Nelson and Siegel, 1987] are based on shift, twist and butterfly factors inspired from PCA decomposition of the yield curves.

3.2 AutoEncoders: Deep Neural Network Decomposition

In most general terms, an Autoencoder neural network [Hinton and Salakhutdinov, 2006] is a learning algorithm whose objective is to generate an output vector which is close to its input vector. In other words, it minimizes $\|x - \hat{x}\|$ where the autoencoder function can be written as $\hat{x} = A(x)$.

An autoencoder consists of two functions, the encoder and the decoder $f : \mathbb{R}^p \to \mathbb{R}^q$ and $g : \mathbb{R}^q \to \mathbb{R}^p$ such that:

$$f, g = \arg\min_{f,g} \|x - f(g(x))\|$$

The dimensionality reduction is achieved by selecting $q < p$. Autoencoders have enjoyed wide successes in representation learning particularly of images and text data (reference).

In figure 8 we show the architecture used for the encoding and decoding functions f, g for the autoencoder. This autoencoder is setup with $q = 3$ and $p = 8$.

3.3 Comparison Study

The main topic of interest herein is to compare the relative performance of PCA to that of Autoencoder. I choose the comparison metric as the reconstruction RMSE metric $\|x - f(g(x))\|$. This is motivated by the observation that RMSE is an efficiency measure of the dimensionality reduction method.

For the comparison study, I use yield curve data for the three countries with the initial 60% of the data kept as training data. The rest of the data is used for out-of-sample predictions. The training is done once and RMSE is calculated on a daily basis for a look ahead period of 500 days corresponding to an approximately 2 year trading period.

This RMSE comparison for US, UK and Germany is shown in Figures 9, 10 and 11. It is clear that autoencoders outperform PCA in the in-sample training data by having upto 50% lower RMSE than PCA. This outperformance is carried over to out-of-sample test data as well but to a more limited extent.

4 Conclusion

The first part of the paper is focused on studying an appropriate choice of statistical distribution for modeling yield curve changes for three countries (US, UK and Germany). I show that normal distribution is a poor choice for modeling this data owing to fat-tailed nature of the data samples. Instead, I propose that multivariate student-t distribution with $\nu = 2$ degrees of freedom or a gaussian mixture model with $K = 10$ components are superior choices. In particular, both student-t and gaussian mixture can fit the leptokurtic nature of the data.

In the second part of the paper, I focus on understanding how well PCA stacks up with Deep Autoencoders, a non-linear dimensional reduction technique. The results for the three yield curves show that auto-encoders outperform PCA for out-of-sample tests.

5 Future Work

It will be of interest to compare efficiency of alternative deep autoencoders (Variational autoencoders) and related architectures.

References

[Barber and Copper, 2012] J.R. Barber and M.L. Copper. Principal component analysis of yield curve movements. *Journal of Economics and Finance*, 36(3):750–765, 2012.

[Cochrane and Piazzesi, 2009] John H. Cochrane and Monika Piazzesi. Decomposing the yield curve. *AFA 2010 Atlanta Meetings Paper*, 2009.

[Diebold and Li, 2006] Francis Diebold and Canlin Li. Forecasting the term structure of government bond yields blah. *Journal of Econometrics*, 130(1):337–364, 2006.

[Henze N, 1990] Zirkler B. Henze N. A class of invariant consistent tests for multivariate normality. *Communications in Statistics - Theory and Methods*, 19(10):3595–3617, 1990.

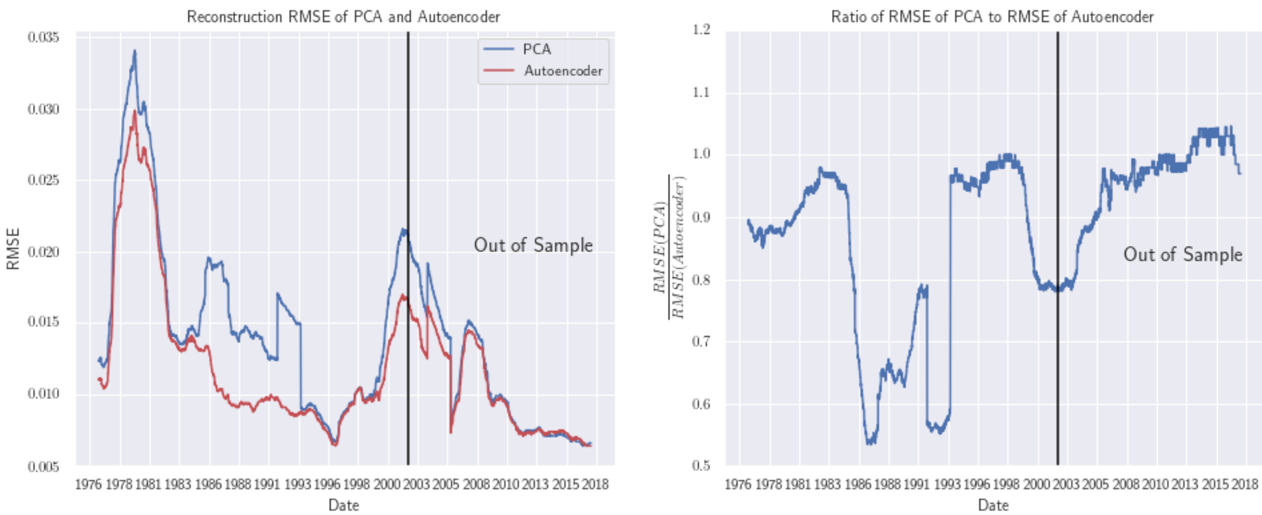


Figure 9: PCA and Autoencoder RMSE comparison: US

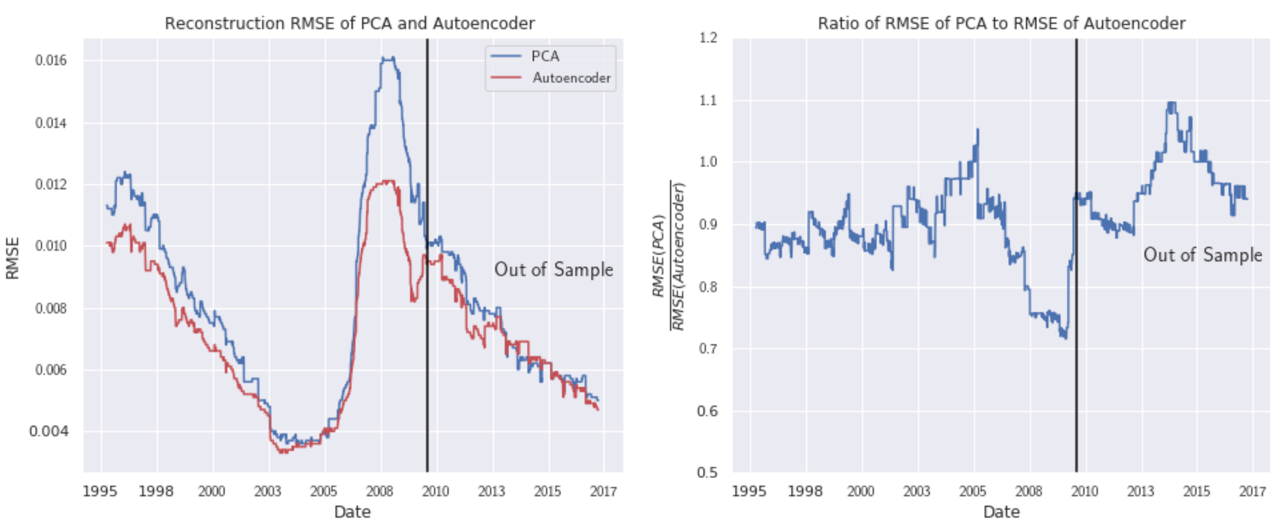


Figure 10: PCA and Autoencoder RMSE comparison: UK

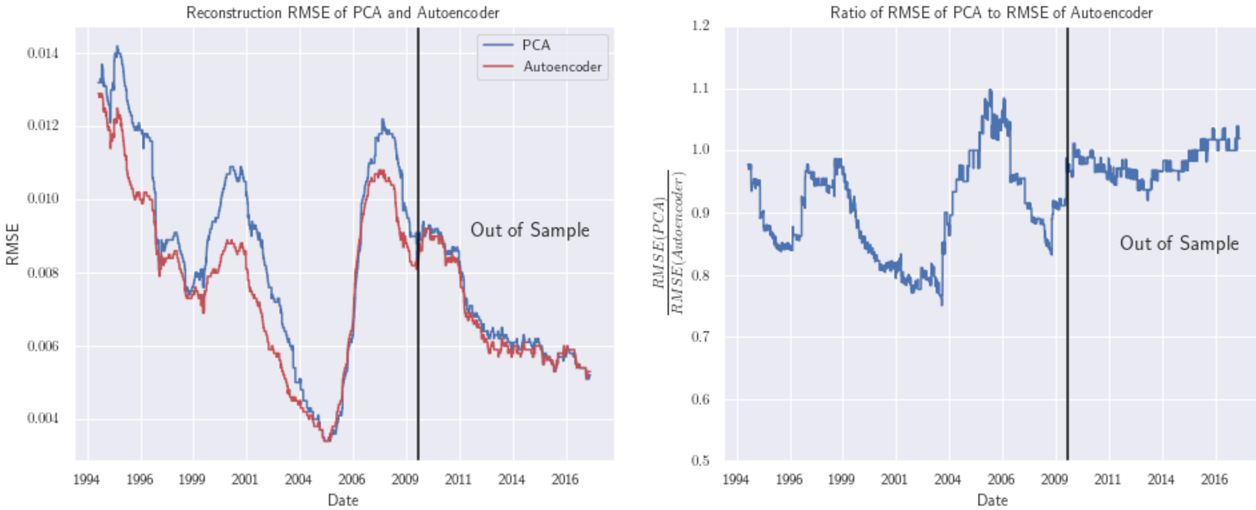


Figure 11: PCA and Autoencoder RMSE comparison: Germany

[Hinton and Salakhutdinov, 2006] G. E. Hinton and R. R. Salakhutdinov. Reducing the dimensionality of data with neural networks. *Science*, 313(5786):504–507, 2006.

[Liu and Rubin., 1995] Chuanhai Liu and Donald B. Rubin. Ml estimation of the t distribution using em and its extensions, ecm and ecme. *Statistica Sinica*, 5(1):19–39, 1995.

[Mardia, 1970] K.V. Mardia. Measures of multivariate skewness and kurtosis with applications. *Biometrika*, 57(3):519–530, 1970.

[Nelson and Siegel, 1987] Charles R. Nelson and Andrew F. Siegel. Parsimonious modeling of yield curves. *The Journal of Business*, 60(4):473–489, 1987.

[Schwarz, 1978] Gideon E. Schwarz. Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464, 1978.