UNIVERSITY OF GOTHENBURG

SCHOOL OF BUSINESS, ECONOMICS AND LAW

Graduate School

An Independent Dynamic Latent Factor Approach to Yield Curve Modeling

Robin Rohlén

June, 2018

A thesis submitted for the degree of Master of Science in Finance

Supervisor: Marcin Zamojski

Abstract

Understanding the yield curve characteristics and dynamics is important for many tasks such as pricing financial assets, portfolio allocation, managing financial risk, and conducting monetary policy. Therefore, it is important to use models that are interpretable, fits well, and make useful forecasts. In this paper, I introduce a dynamic yield curve model with latent independent factors based on Independent Component Analysis, which is a statistical method used successfully in other fields than finance. I find that one can interpret the factors as level, slope, and curvature of the yield curve. I also find that the ICA-based model fits the yield curve well and produce good forecasts. In particular, it shows significantly better out-of-sample forecasts for the short-term maturities than the commonly used dynamic Nelson-Siegel model. I find that the factors correlate with macroeconomic variables such as monetary policy instrument, real economic activity, and inflation. Finally, I find that the curvature factor seems to be more important than the previous literature state.

Acknowledgements

First of all, I want to thank Marcin Zamojski for his guidance throughout this period. Not only being flexible, but he also has a passionately curious mind which I believe is an essential ingredient in a successful scientific mind. Ever since I took the course Financial Econometrics, which probably is the best course at Graduate School, he has introduced many interesting topics on this subject. Finally, I want to thank the Graduate School and Centre for Finance (CFF) for giving me invaluable knowledge throughout this Master's programme.

Table of Contents

Section 1

Introduction

In this paper, I introduce a dynamic yield curve model with latent independent factors based on a statistical method called Independent Component Analysis (ICA). ICA is used successfully in, e.g., signal processing and biomedical engineering applications (e.g., McKeown et al., 2003; Östlund et al., 2006; Mukamel et al., 2009). However, ICA is unexplored in the finance literature, in particular, to yield curve modeling, although the concept became known over 30 years ago (Hyvärinen et al., 2001). The advantage of using ICA, rather than the commonly used Principal Components Analysis (PCA), is that the ICA factors are independent, whereas PCA factors may still have higher-order dependence that may mask useful information. Also, it is more useful to study independent factors because one can analyze them one at a time.

Bonds are important fixed-income securities because governments, municipalities, and companies commonly use them to finance various activities and projects. For instance, the U.S. government regularly funds its operations through bills, notes, and bonds that are issued by the U.S. Treasury Department and backed by the U.S. government. Therefore, these securities are considered risk-free and used as benchmarks for various interest rates, such as savings and mortgage rates. Moreover, it is essential to understand these bonds characteristics and dynamics for many tasks such as pricing of other financial assets, portfolio allocation, managing financial risk, and conducting monetary policy.

The income return on the bond is called yield, and preferably we want to examine a basket of bond yields. Consequently, we construct a so-called yield curve, which is a function through all the yields, i.e., the yield curve describes yields across multiple maturities.[1]

Previous research shows that it is possible to decompose the yield curve into three unobservable factors, which one often refer to as level, slope, and curvature (Litterman and Scheinkman, 1991). These factors are related to changes in the yield curve shape. For instance, a shock to the level

[1] It is common to construct the yield curve by using splines, Nelson and Siegel (1987), or Svensson (1994) model. Also, there are many different representations of these bonds, such as the zero coupon curve, the discount curve, the forward curve, and the par yield curve. For yield curve construction, see Gürkaynak et al. (2007).

factor induces a shift of the yield curve. A positive shock in the slope factor makes the yield curve less steep and means that the short and long-term interest rates differ less than before the shock. Finally, a positive shock to the curvature factor results in a more curved yield curve because the factor loads more on the mid-term interest rates (Wu, 2003).

The level, slope, and curvature factors are a standard product of many yield curve models. A standard model is to use parametric shapes for the factor loadings (Nelson and Siegel, 1987; Svensson, 1994; Diebold and Li, 2006), which many central banks use (BIS, 2005; ECB, 2008) because they fit the yield curve well. Besides fitting the yield curve, these models produce accurate yield curve forecasts, which are important in many areas of finance. However, parametric models may suffer from limited ability to fit irregular yield curve shapes. Therefore, another standard approach is to use dimension-reduction methods which extract statistically uncorrelated factors that explain the maximal variance of the yield data (Steeley, 1990; Baygün et al., 2000; Joslin et al., 2014). However, uncorrelated factors do not imply independent factors, i.e., the uncorrelated factors may have higher-order dependence between them. Therefore, by extracting independent factors one can use these to model the yield curve, and they can also be used to improve forecasts because they may unmask important information hidden in the uncorrelated factors. Also, it is more useful to analyze independent factors because one can examine them one at a time.

Here, I introduce a dynamic yield curve model based on independent factors extracted from ICA. I find that one can (also) interpret ICA factors as level, slope, and curvature and that they explain 97% of the variation in the yield curve data. I also find that the estimated ICA-based factors are highly persistent with their dynamics, and these results are in line with previous literature (e.g., see Diebold et al., 2006; Christensen et al., 2011).

I compare the ICA and PCA yield models against the Diebold and Li (2006) model, which is a dynamic Nelson and Siegel (1987) model and well-known for its forecasting ability. The results show that both models fit the yield curve well and they also show significantly better out-of-sample forecasts for the short-term maturities and a long forecast horizon, which is more relevant to institutional agents such as central banks.

I also find that the estimated ICA-based factors correlate with relevant macroeconomic variables and these are in line with previous literature (e.g., Diebold et al., 2006). In particular, the slope factor is highly correlated with the Federal Funds Rate and Capacity Utilization during the period 1997-2017. Although the level factor does not correlate with any macroeconomic variable during the whole period, there are periods of high correlation with inflation-related macro variables. These correlations occur around the years related to the financial crisis of 2008 and also in more recent years. There are also indications that the curvature factor correlates with Unemployment Rate, Consumer Sentiment Index, and Trade Weighted U.S. Dollar in recent years. Finally, most of the impulse response functions are in line with previous literature although some differences may be the result of unconventional monetary policy such as quantitative easing since a majority of the data in the studied period has rates close to the zero lower bound. Another interesting finding is that the curvature factor seems to be more important than previous literature states.

The remainder of the paper is structured as follows. In section 2, I give a literature review associated with term structure models. In section 3, I present some basic notation related to the term structure of interest rates and, furthermore, present theoretical models associated with the yield curve shapes. In section 4, I present estimation methods along with an introduction to relevant dimension-reduction methods. I finish the section to introduce performance measures regarding in-sample and forecasting fit. In section 5, I present the yield curve models I use in this paper. In section 6, I present the data and its sources. In section 7, I present and analyze the results. Finally, in section 8, I summarize the significant findings and provide a discussion for future research.

Section 2

Literature Review

There are different types of term structure models, and the literature of term structure models is extensive. One model-type is both econometric-based and non-arbitrage-free, while another is arbitrage-free. Arbitrage-free models tie the dynamics of interest rates at longer maturities to the short rate, through a no-arbitrage condition under a risk-neutral probability measure. Furthermore, the short rate models capture dynamics of the instantaneous interest rate that typically specify an affine function of several underlying factors.

Arbitrage-free term structure models were introduced by Vasicek (1977), who derives a general form of the term structure of interest rates. Vasicek argues that short-term interest rates drive bond prices and therefore propose a one-factor short rate model. This model has a non-zero probability that the short-rate becomes negative and considering that negative rates were unheard of before the 2000s, Cox et al. (1985) propose a modified one-factor model restricted to positive rates since the goal of these models is to model the underlying process(es) that drive the prices. However, since both of these models have a finite number of free parameters, it is difficult to specify the parameter values such that the model can calibrate well to observed market prices. Thus, Ho and Lee (1986) and Hull and White (1990) propose one-factor short rate models with time-varying parameters.

One-factor models may be not flexible enough to capture the dynamics of many maturities and different curve shapes since they can only capture the dynamics using one source of uncertainty, i.e., the short rate. To allow multiple sources of risk, Longstaff and Schwartz (1992) propose a two-factor model while Chen (1996) proposes a three-factor model, which has a stochastic mean and volatility for the short rate.

Typically, one is interested in studying a basket of bonds, rather than a single bond. Therefore, Duffie and Kan (1996) propose an affine multi-factor model of the term structure, and Dai and Singleton (2000) formulate a standard framework for the canonical representation of affine term structure models, in which they describe the latent factors of the yields as the level, slope, and curvature.

The above-described term structure models are theoretically appealing, however, describing the joint dynamics of the yield curve and macroeconomic variables is important for the economic interpretation of level, slope, and curvature factors. In other words, it is easier to understand these factors by examining how changes in these factors influence macroeconomic variables, which often is more relatable. In this way, it is possible to see how changes in the factors influence the ability to control macro variables. For example, how changes in reference rates may impact inflation through the yield curve. Therefore, adding macroeconomic variables to these models is an increasingly popular concept (e.g., Cochrane and Piazzesi, 2005), and has introduced the macro-finance term structure models, whose goal is to understand economic forces that drive changes in interest rates, by jointly modelling macroeconomy and the yield curve (e.g., Ang and Piazzesi, 2003). These models imply macro spanning, i.e., that all relevant information about the economy is in the yield curve, and macro variation is spanned by (perfectly correlated with) the yield curve.

Arbitrage-free models are well-studied and theoretically rigorous. However, many practitioners and central banks use simpler approaches. The simple models leverage the high persistence, and they are empirically successful. Among these, Nelson and Siegel (1987) model (NSM) is the most popular (BIS, 2005; ECB, 2008). It is a parametric model that fits the yield curve well with four time-invariant hyperparameters with the first three having the interpretation as the level, slope, and curvature of the yield curve, while the fourth parameter relates to a decay parameter for these previous parameters. To improve the yield curve fit, Svensson (1994) suggests an extension of the NSM by including an additional hyperparameter.

The NSM has time-invariant parameters, which Diebold and Li (2006) extend to a dynamic setting and show that it can produce accurate term structure forecasts. Also, Diebold et al. (2006) use the NSM to study the interactions between the macroeconomy and the yield curve. There are other extensions such as score-driven time-varying parameters (Koopman et al., 2017), interaction with unconventional monetary policy (Mesters et al., 2014), and using shadow-rates (rather than nominal rates) respecting the zero lower bound (Christensen and Rudebusch, 2016). Although most of these models are empirically successful, they are theoretically lacking. Both Björk and Christensen (1999) and Filipović (1999) show that the NSM is not arbitrage-free. Christensen et al. (2011) resolve this by deriving a class of affine arbitrage-free dynamic term structure models that approximate the NSM yield curve specification, by adding a yield-adjustment term. Also, Coroneo et al. (2011) find that the NSM parameters are not statistically different from those derived by no-arbitrage affine-term structure models, which suggests that the yield-adjustment term is small.

There are also non-parametric methods to extract the level, slope, and curvature factors. Among these, Principal Component Analysis (PCA) is a particularly common alternative (e.g. Steeley, 1990; Baygün et al., 2000; Joslin et al., 2014). PCA transforms a set of possibly correlated observations into a set of linearly uncorrelated variables that explain the highest possible variance. They are called principal components (PCs). Lekkos (2000) find that the first three PCs have no natural interpretation of level, slope, or curvature. However, Lord and Pelsser (2007) argues that they, in fact, have a natural interpretation. Also, Baygün et al. (2000) give comprehensive details on how to use PCA for trading and hedging by managing the exposure to these factors.

PCA is one of many approaches to decomposing a dataset, where another approach is Independent Component Analysis (ICA). ICA is a multivariate statistical method that decomposes a multidimensional signal into additive independent factors (Bell and Sejnowski, 1995). The primary assumption for ICA is that the factors are non-normally distributed and independent. In essence, PCA and ICA differ in that the former decomposes the data such that the factors are uncorrelated, whereas the latter decomposes the data such that the factors are independent.

An early application of ICA to financial data is in Back and Weigend (1997), who use daily data from the Tokyo Stock Exchange, and tries to extract structure from returns. Cha and Chan (2000) show the relation between ICA and the factor model as a data mining tool to extract the underlying factors and obtain sensitivities for the factor model. Kumiega et al. (2011) investigate the factors that drove the U.S. equity market returns from 2007-2010, applying ICA to the returns of exchange-traded funds, and analyzed the factors volatility clustering. More recently, Fabozzi et al. (2016) identify three interpretable factors driving the changes in credit default swap spreads. In general, these approaches are unsuccessful because of efficient markets and low predictability of returns and the extensive literature extending the CAPM. However, the yield curves are more persistent and should (theoretically) be driven by similar underlying factors, but the literature related to ICA and fixed-income securities is sparse, especially concerning yield curve modeling. Hence, I introduce ICA into yield curve modeling. In particular, I contribute to the literature by investigating ICA's role in yield curve forecasting.

Section 3

Theory

In this section, I briefly describe bonds, yields, and the term structure of interest rates. The aim is to get some intuition related to these concepts. Finally, I describe theories that explain the influence of the yield curve.

3.1 Term Structure of Interest Rates

The primary bond market is where the supply and demand of bonds meet. The investors may enter the bond market to maximize their yield to maturity y, whereas the bond suppliers may enter the bond market to minimize their costs of funding by getting the lowest possible interest rate r_t at time t. However, bonds trade in the secondary markets and these bonds are used to derive the yield curve.

The price P of a fixed coupon bond with principal equal to 1 at time t and maturing at time T is expressed as:[1]

$$P(t,T) = \sum_{i=t+1}^{T} \frac{c_i}{(1+y)^{i-t}} + \frac{1}{(1+y)^{T-t}}, \tag{3.1}$$

with yield y and coupon payments paid at time $t+1,\dots,T$ with coupon rates $c_{t+1},\dots,c_T$. For an infinitely small interest period, the price is expressed as:[2]

$$P(t,T) = \sum_{i=t+1}^{T} c_i e^{-y(i-t)} + e^{-y(T-t)}. \tag{3.2}$$

There is an important type of bonds called zero-coupon bonds, which have no cash flows except the principal amount at the bond's maturity. Thus, let the coupon rates equal 0. Then, the zero

[1] The bond price can be split into a so-called clean and dirty price, depending on whether to include the accrued interest or not.

[2] This is called continuous compounding.

coupon bonds yield to maturity is expressed as:

$$y = P(t,T)^{-1/(T-t)} - 1, \tag{3.3}$$

where $P(t,T)$ is the price of the zero-coupon bond. Also, under continuous compounding the zero coupon bonds yield to maturity is expressed as:

$$y = -\frac{\log P(t,T)}{T-t}. \tag{3.4}$$

One constructs a yield curve by mapping the yield to maturity against different maturities. Similarly, the term structure of interest rates is a function of the interest rate $r_t(\tau)$ with respect to the maturity τ, where a bond with a price $P(t,T)$ implies the whole set $\{r_t(\tau_i)\}_{i=t+1}^{T}$. This set is related to the bonds cash flows. However, a single yield for the investor is determined directly. Since each interest rate is related to individual cash flows, it is suitable to use the term structure of interest rates, and that is why I use zero-coupon bonds.

In Figure 3-1, I show a yield curve (bold line) and how the level, slope, and curvature affect the yield curve in the short, mid, and long-term with a positive shock (dotted line). As mentioned before, the result of most studies are that these three factors tend to affect the different maturities. As shown in the figure, a shock to the level factor leads to a shift of the yield curve. A positive shock to the slope factor makes the yield curve less steep, which results in less spread between the short and long-term interest rates than before the shock. Finally, a positive shock to the curvature factor makes the yield curve more curved since it loads more on the mid-term rates.

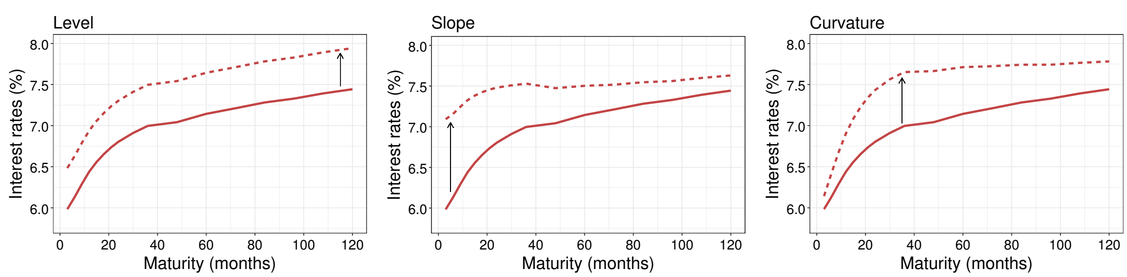

Figure 3-1: The yield curve response to shocks on the level, slope, and curvature.

3.2 Yield Curve Theories

In this section, I present theories that try to explain the influence of the term structure of interest rates, and how they vary with maturity. For example, the US Treasury yield curve has been upward sloping nearly 90% of the time in last decades, a fact that may reflect that the market has been expecting rising rates or that they require a positive bond risk premia (Ilmanen, 1995). I give a brief description of four theories, i.e., (i) expectations hypothesis, (ii) liquidity premium theory, (iii) segmented market hypothesis, and (iv) preferred habitat theory. After that, I provide a short view of Ilmanen (1995) who describes the main influences of the term structure of interest rates in terms of three factors; market's rate expectations, bond risk premia, and convexity bias. Finally, I try to link these theories and macroeconomic variables to the level, slope, and curvature factors.

Expectations Hypothesis

Under expectations hypothesis, observed forward rates are an unbiased estimator of the future spot rates and suggest that the shape of the yield curve depends on market participants' expectations of future interest rates. Thus, there is lack of arbitrage opportunities in the bond market since the long-term bonds return, holding it for n years, is equal to rolling over short-term bonds n times. Hence, the bonds have the same expected return, and higher-yield bonds are expected to suffer capital losses that offset their yield advantage.

When the market expects an increase in bond yields, the current term structure becomes upward-sloping so that any long-term bond's yield advantage and expected capital loss, due to the expected yield increase, exactly offset one another. In contrast, expectations of yield declines and capital gains decrease the current long-term yields below the short-term rate, making the term structure inverted. In other words, the market's rate expectations influence the yield curve steepness and relate to the slope factor. However, the market's expectations about the steepness of the yield curve affect the curvature of the yield curve and should refer to the curvature factor. For example, if the market expects a flatter curve, they need to offset the expected capital gains, and this makes the yield curve more curved.

Liquidity Premium Theory

A fundamental assumption in the expectations hypothesis is that all bonds have the same expected rate of return, regardless of maturity. However, empirical evidence suggests that expected returns vary across bonds, i.e., there is a risk premium associated with nominal holding period. Therefore, Keynes (1936) considers a constant risk premium associated with the maturities. As the long-term bonds are related to increased risk exposure, investors demand a higher risk premium. Thus, positive bond risk premia make the yield curve slope upward and affect the slope factor such that the spread between the short and long-end are larger, whereas a negative bond risk premia tend to make the yield curve inverted. Hence, the bond risk premia possibly relate to the slope factor. Also, there is some evidence that the risk premia relates to the curvature factor. For example, Campbell et al. (2017) show a theoretical link between curvature and the level of the term premia, which is driven by the covariance between the real interest rate and inflation. Also, Abbritti et al. (2018) provide empirical support for the relationship between the curvature factor and term premium dynamics in an international context.

Segmented Market Hypothesis

The segmented hypothesis assumes that investors at various maturities are strictly different and that as a result the supply and demand for short and long-term bonds are different (Culbertson, 1957). For instance, if investors prefer liquid portfolios, they may buy short-term bonds, increasing the need for short-term bonds, which results in higher prices and lower yields. In other words, there is no implicit relationship between the interest rates for short, mid, and long-term bonds. Hence, one should view the different rates separately, and therefore this theory may suggest that one should study independent factors and that the level, slope, and curvature factors may correspond to the supply and demand of different investor types.

Preferred Habitat Theory

The preferred habitat theory is closely related to the segmented market hypothesis. It states that, besides interest rate expectations, investors have distinct investment horizons and require a premium to buy bonds with maturities outside their preferred maturity (Modigliani and Sutch, 1966). Short-term investors would appear more frequently in the fixed-income market, and therefore longer-term yields tend to be higher than short-term yields. Thus, this should reflect the shapes of the level and slope factors. Also, that the mid-term yields are more relevant for hedging and certain investors appear at these maturities and therefore relevant for the curvature factor.

Market's Rate Expectations, Bond Risk Premia, and Convexity Bias

Ilmanen (1995) argues that three economic forces influence the term structure of the forward rates; the market's rate expectations, the bond risk premia, and the convexity bias. First, market's rate expectations coincide with the previous theories. Second, past theories assume either a non-zero or constant risk premia which are inconsistent with empirical evidence that suggests a time-varying risk premium. Third, convexity bias refers to the fact that different bonds have different convexity[3] and may reflect the different yields. Rather than yields, an investor is primarily interested in expected returns, and therefore they tend to demand less yield to improve their returns as a result of convexity.

A steep yield curve that slopes upwards may reflect either the market's expectations of rising rates or high required risk premia and relates to the slope factor. A humped curve can reflect the market's expectations of either a flatter yield curve or high volatility, which makes the convexity

[3]Convexity is a measure of the curvature between bond prices and its yields that show how the duration of a bond changes as the interest rate changes.

more valuable because of its property of increasing the expected return and may reflect the shape of the curvature factor.

Level, Slope, and Curvature: Links to Macroeconomic Variables

All these theories try to explain the influence of the term structure of interest rates, and how they vary with maturity. They all have in common that there are underlying factors that affect the yield curve. Yield curve models often use level, slope, and curvature as their underlying factors and they explain the majority of the yield curve variation (Litterman and Scheinkman, 1991). However, it is not trivial to connect the various theories to these factors since they are likely to represent combined influences of other factors that affect different parts of the term structure. For example, following Ilmanen (1995), it may be that the level, slope, and curvature is a mixture of the market's rate expectations, the bond risk premia, and the convexity bias. In fact, the literature says that there is a connection between the level factor and inflation expectations, and also between the slope factor and monetary policy actions (e.g., Diebold et al., 2006; Rudebusch and Wu, 2008). Finally, the curvature factor has received less attention in the literature, but there are links with the term premia as pointed out earlier in this section (Campbell et al., 2017; Abbritti et al., 2018).

Section 4

Methods

In the previous section, I discuss different yield curve theories and how they relate to that yields often vary with maturity. The various theories agree there are underlying factors that affect the yield curve, and it is common to use level, slope, and curvature as underlying factors in yield curve models. Before I describe the models, which I introduce in the next section, I present factor extraction and dimension-reduction methods I use for the different models.

4.1 Factor Extraction Methods

Often one has two different processes; a measurement process that observes some phenomena (e.g., yields) and an underlying transition process that tries to capture the underlying dynamics. Under the assumption that the latent process has dependence over time, it is common to model it by a vector autoregressive process of first order. This approach refers to a state-space model, which I specify as:

$$\begin{aligned} \mathbf{y}_t &= \mathbf{\Lambda}_t \mathbf{f}_t + \boldsymbol{\varepsilon}_t, \\ \mathbf{f}_t &= \mathbf{\Gamma}_t \mathbf{f}_{t-1} + \boldsymbol{\eta}_t, \\ \boldsymbol{\varepsilon}_t &\sim N(\mathbf{0}, \mathbf{H}_t), \\ \boldsymbol{\eta}_t &\sim N(\mathbf{0}, \mathbf{Q}_t), \end{aligned} \tag{4.1}$$

where $\mathbf{y}_t$ are the observed data, $\mathbf{f}_t$ are the underlying factors, $\boldsymbol{\varepsilon}_t$ are the measurement errors, $\boldsymbol{\eta}_t$ are the innovations to the latent factors, $\mathbf{\Lambda}_t$ is the observation matrix mapping latent processes to the observations, $\mathbf{\Gamma}_t$ is the transition describing the evolution of the latent processes in time, $\mathbf{H}_t$ is a measurement error covariance matrix, and $\mathbf{Q}_t$ is a covariance matrix related to the innovations to the latent factors. The measurement errors are assumed to be independent across time, and it is common to assume that the measurement error process and transition error process are mutually independent.

Given this linear and normally-distributed state-space model, Kalman filter (KF) efficiently computes the joint likelihood of both the state and the observation. Now, let $\mathbf{y}_{1:t} = \{\mathbf{y}_1, \ldots, \mathbf{y}_t\}$, and I define the conditional expectation for the filtering and forecast distributions as:

$$\begin{aligned} \mathbf{f}_{t|t} &= E(\mathbf{f}_t|\mathbf{y}_{1:t}), \\ \mathbf{f}_{t|t-1} &= E(\mathbf{f}_t|\mathbf{y}_{1:t-1}). \end{aligned} \tag{4.2}$$

Furthermore, I define the conditional error covariance matrices for filtering and forecasting as:

$$\begin{aligned} \mathbf{\Sigma}_{t|t} &= E[(\mathbf{f}_t - \mathbf{f}_{t|t})(\mathbf{f}_t - \mathbf{f}_{t|t})^\top|\mathbf{y}_{1:t}], \\ \mathbf{\Sigma}_{t|t-1} &= E[(\mathbf{f}_t - \mathbf{f}_{t|t-1})(\mathbf{f}_t - \mathbf{f}_{t|t-1})^\top|\mathbf{y}_{1:t-1}]. \end{aligned} \tag{4.3}$$

The forecast distribution is:

$$\mathbf{f}_t|\mathbf{y}_{1:t-1} \sim N(\mathbf{f}_{t|t-1}, \mathbf{\Sigma}_{t|t-1}), \tag{4.4}$$

where $\mathbf{f}_{t|t-1} = \mathbf{\Gamma}_t\mathbf{f}_{t-1|t-1}$ and $\mathbf{\Sigma}_{t|t-1} = \mathbf{Q}_t + \mathbf{\Gamma}_t\mathbf{\Sigma}_{t-1|t-1}\mathbf{\Gamma}_t^\top$. The filtering distribution is:

$$\mathbf{f}_t|\mathbf{y}_{1:t} \sim N(\mathbf{f}_{t|t}, \mathbf{\Sigma}_{t|t}), \tag{4.5}$$

where $\mathbf{f}_{t|t} = \mathbf{f}_{t|t-1} + \mathbf{K}_t(\mathbf{y}_t - \mathbf{\Lambda}_t\mathbf{f}_{t|t-1})$ and $\mathbf{\Sigma}_{t|t} = (\mathbf{I} - \mathbf{K}_t\mathbf{\Lambda}_t)\mathbf{\Sigma}_{t|t-1}$, where $\mathbf{I}$ denotes the identity matrix, and $\mathbf{K_t} = \mathbf{\Sigma}_{t|t-1}\mathbf{\Lambda}_t^\top(\mathbf{\Lambda}_t^\top\mathbf{\Sigma}_{t|t-1}\mathbf{\Lambda}_t + \mathbf{H}_t)^{-1}$ is called the Kalman gain.

Given some initial conditions $\mathbf{f}_{0|0} = \mu_0$, $\mathbf{\Sigma}_{0|0} = \mathbf{\Sigma}_0$ (e.g., see Durbin and Koopman, 2012), and assuming the parameter matrices $\mathbf{\Lambda}_t$, $\mathbf{\Gamma}_t$, $\mathbf{Q}_t$, $\mathbf{H}_t$, $t = 1, \ldots, T$ are known, I obtain sequential estimates of the state by the following algorithm:

for $t = 1$ to T

1. Obtain the forecast-distribution mean $\mathbf{f}_{t|t-1}$ and covariance matrix $\mathbf{\Sigma}_{t|t-1}$.
2. Obtain the gain $\mathbf{K}_t$, the filtering-distribution mean $\mathbf{f}_{t|t}$, and covariance matrix $\mathbf{\Sigma}_{t|t}$.

end

The parameters that go into the KF are estimated with maximum likelihood (ML). ML is a method for estimating unknown parameter vector $\boldsymbol{\theta} \in \mathbf{\Theta}$, where $\mathbf{\Theta}$ is a parameter space. Given a joint density $\mathbf{Y} \sim f(\mathbf{y}|\boldsymbol{\theta})$ for a parametric distribution $f(\cdot)$ and observed data $\mathbf{y}$, the idea is to choose parameters that maximize the probability of generating the observed sample, i.e., the parameter values that maximize a likelihood function $\mathcal{L}(\boldsymbol{\theta}|\mathbf{y})$. In practice, it is convenient to work with the logarithm of the likelihood function, $\ell(\boldsymbol{\theta}|\mathbf{y}) = \log \mathcal{L}(\boldsymbol{\theta}|\mathbf{y})$.[1] Hence, assuming the existence

[1]In some cases it is more convenient to minimize the negative log-likelihood. Also, to work with the average log-likelihood is often more computationally efficient and stable.

of a global maximum, the ML estimator is defined as $\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \ell(\boldsymbol{\theta}|\mathbf{y})$.

Often, closed-form solutions to the maximization problem do not exist, leading to numerical optimizations by gradient-based methods. These methods find a local maximum by making use of, in this case, the log-likelihood function and its corresponding derivative(s). Furthermore, these methods can be divided into first and second order methods, where the latter implies using second derivatives (Hessian) besides the first derivatives (gradient).

4.2 Dimension-Reduction Methods

Sometimes it is interesting and practical to determine typical characteristics of many variables. For example, if only a few factors are in common and explain the majority of the variables variation, it is easier to interpret the driving factors in the data. Thus, this gives the motivation to decompose a dataset, and I present two methods widely used in different applications. The first method decomposes a dataset such that factors are uncorrelated to the other factors, and also such that they have a maximal variance. The second method decomposes a dataset such that factors are independent.

4.2.1 Principal Component Analysis

Principal Component Analysis (PCA) is a statistical method that transforms a set of observed variables into a set of uncorrelated variables called principal components (PCs). The first PC has the maximal variance, i.e., accounts for maximal variability in the data, and each following factor has the maximal variation given that it is uncorrelated to the previous factors. More formally, consider an observation matrix $\mathbf{Y}$ with dimension $T \times m$ and centered columns, where T is the number of observations (or yields), and m is the number of variables (or maturities). The PCs of $\mathbf{Y}$ is expressed as:

$$\mathbf{P} = \mathbf{Y}\mathbf{W}, \tag{4.6}$$

where $\mathbf{P}$ is a $T \times m$ matrix, $\mathbf{W}$ is a $m \times m$ matrix with columns as loading vectors (and eigenvectors of $\mathbf{Y}^\top\mathbf{Y}$). PCA finds the direction that maximizes the sample variance, i.e., for each loading vector $\mathbf{w}$, the first loading vector $\mathbf{w}_1$ is defined as:

$$\mathbf{w}_1 = \arg\max_{\|\mathbf{w}\|_2=1} \{\widehat{\mathrm{Var}}(\mathbf{Y}\mathbf{w})\} = \arg\max_{\|\mathbf{w}\|_2=1} \left\{\mathbf{w}^\top \frac{\mathbf{Y}^\top\mathbf{Y}}{T}\mathbf{w}\right\}. \tag{4.7}$$

The resulting projection $\mathbf{p}_1 = \mathbf{Y}\mathbf{w}_1$ is called the first PC of $\mathbf{Y}$ and the elements of $\mathbf{w}_1$ are called the PC loadings. To obtain the k^{th} PC, the first $k-1$ PCs are subtracted from $\mathbf{Y}$, i.e.,

$$\hat{\mathbf{Y}}_k = \mathbf{Y} - \sum_{i=1}^{k-1} \mathbf{Y}\mathbf{w}_i\mathbf{w}_i^\top, \tag{4.8}$$

and to find the loading vector that extracts the maximum variance from this new data matrix one need to calculate:

$$\mathbf{w}_k = \underset{\|\mathbf{w}\|_2=1}{\arg\max} \left\{ \mathbf{w}^\top \frac{\hat{\mathbf{Y}}_k^\top \hat{\mathbf{Y}}_k}{T} \mathbf{w} \right\}. \tag{4.9}$$

An efficient and standard way to do PCA is through singular value decomposition (SVD), i.e.,

$$\mathbf{Y} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{W}^\top, \tag{4.10}$$

where $\boldsymbol{\Sigma}$ is a $T \times m$ diagonal matrix of decreasing positive numbers (singular values of $\mathbf{Y}$), $\mathbf{U}$ is an orthogonal $T \times T$ matrix, and $\mathbf{W}$ is an orthogonal $m \times m$ matrix. Using the SVD, the PCs can be written as:

$$\mathbf{P} = \mathbf{U}\boldsymbol{\Sigma} = \mathbf{Y}\mathbf{W}. \tag{4.11}$$

Note that the principal scores $\mathbf{P}$ has dimension $T \times m$ and the loadings $\mathbf{W}$ has dimension $m \times m$. Recall that PCA is useful because it can take a large dataset and a small set of factors that explain a large fraction of the variation in this dataset. PCA is used in many fields, as a pre-processing step for dimension reduction, or it is used to find predictors. PCA is also common in the yield curve literature because it is well-known that the first three PCs relates to the level, slope, and curvature factor. This corresponds to taking the first three columns of $\mathbf{P}$ and $\mathbf{W}$, which has dimension $T \times p$ and $m \times p$ respectively, with $p = 3$.

4.2.2 Independent Component Analysis

PCA decomposes a dataset such that the set of observed variables form a new set of uncorrelated variables. Independent Component Analysis (ICA) decomposes a dataset such that the set of observed variables form a new set of independent variables. The advantage to studying independent variables is that one can examine them one at a time. The model specification for ICA is:

$$\mathbf{Y} = \mathbf{S}\mathbf{A}^\top, \tag{4.12}$$

where $\mathbf{Y}$ is a $T \times m$ observation matrix and columns are assumed to have zero mean, $\mathbf{S}$ is a $T \times p$ matrix where columns are the latent factors, and $\mathbf{A}$ is a $m \times p$ mixing matrix. Note that both $\mathbf{A}$ and $\mathbf{S}$ are unknown and have to be estimated. Thus, the goal is to find the unmixing matrix $\tilde{\mathbf{A}} = \mathbf{A}^+ = (\mathbf{A}^\top\mathbf{A})^{-1}\mathbf{A}^\top$ such that the columns of $\mathbf{S}$ are independent, where $\mathbf{A}^+$ denote the pseudoinverse of $\mathbf{A}$. It is possible to express the latent factors $\mathbf{S}$ as (Bell and Sejnowski, 1995; Hyvärinen et al., 2001):

$$\mathbf{S} = \mathbf{Y}\mathbf{Q}^\top\mathbf{R}, \tag{4.13}$$

where $\mathbf{Q}$ is a $p \times m$ whitening matrix[2] and $\mathbf{R}$ is a $p \times p$ orthogonal rotation matrix. It is also possible to express the factor loadings $\mathbf{A}$ as:

$$\mathbf{A} = \mathbf{Q}^{+}\mathbf{R}, \tag{4.14}$$

where $\mathbf{Q}^{+} = (\mathbf{Q}^{\top}\mathbf{Q})^{-1}\mathbf{Q}^{\top}$ denotes the pseudoinverse of the whitening matrix. To make sure that the $\mathbf{A}$ and $\mathbf{S}$ matrices are identifiable, I assume that the factors are statistically independent and have non-normal distributions (e.g., see Hyvärinen et al., 2001). Note, there are several ambiguities. For example, it is not possible to determine the independent factors variances since both $\mathbf{A}$ and $\mathbf{S}$ are unknown, and any scalar in $\mathbf{A}$ may offset the inverse scalar in $\mathbf{S}$, and vice versa. Consequently, it is a common practice to standardize the factors such that they have unit variances. However, the problem of ambiguity of the sign is inevitable, which means that the factors may be inverted. The second ambiguity of ICA is its inability to determine the factor ordering since one can multiply an arbitrary permutation matrix with $\mathbf{S}$ and the inverse of the permutation matrix with $\mathbf{A}$. In other words, it is likely that the output, using the same dataset from two different occasions, give the same factors but they have a different ordering.

Uncorrelatedness, i.e., PCA, is often not enough to separate factors for a given dataset (Hyvärinen et al., 2001, Ch. 1). Therefore, there is extensive literature related to ICA methods and different procedures to estimate $\mathbf{S}$ (and $\mathbf{A}$) because it has proven to be effective in multiple applications. ICA can extract factors based on, e.g., skewness (Song and Lu, 2016), autocorrelation (Lee et al., 2011), or conditional heteroscedasticity (Matilainen et al., 2017). There are also different approaches such as the Infomax and ML approach (e.g., see Hyvärinen et al., 2001), which is shown to be equivalent (Cardoso, 1997). In summary, the goal is to find an orthogonal rotation matrix $\mathbf{R}$ (Equation 4.13 and 4.14) such that the latent factors $\mathbf{S}$ are independent.

Here, I focus on the ML approach as the estimation procedure and derive the likelihood under the assumption of negligible noise. This approach is based on using the result of the density of a linear transform.[3] Thus, let the joint density $f_{\mathbf{Y}}$ of the mixture vector $\mathbf{Y} = \mathbf{AS}$ be expressed as:

$$f_{\mathbf{Y}}(\mathbf{y}) = |\det \mathbf{A}^{+}| f_{\mathbf{S}}(\mathbf{s}) = |\det \tilde{\mathbf{A}}| f_{\mathbf{S}}(\mathbf{s}) = |\det \tilde{\mathbf{A}}| \prod_i f_i(\mathbf{s}_i) = |\det \tilde{\mathbf{A}}| \prod_i f_i(\tilde{\mathbf{a}}_i^{\top}\mathbf{Y}), \tag{4.15}$$

where $\det(\mathbf{A})$ denote the determinant of $\mathbf{A}$, $\tilde{\mathbf{A}} = \mathbf{A}^{+} = (\tilde{\mathbf{a}}_1, \ldots, \tilde{\mathbf{a}}_n)^{\top}$, and f_i, $i = 1, \ldots, m$ denote the independent factors marginal densities. Furthermore, let $\mathbf{Y} = (\mathbf{Y}_1, \ldots, \mathbf{Y}_T)$. Then, the likelihood

[2] A popular method for whitening is to use the eigenvalue decomposition (EVD). Thus, let $\mathbf{E} = (\mathbf{e}_1, \ldots, \mathbf{e}_p)$ be a $m \times p$ matrix whose columns are the unit-norm eigenvectors of covariance matrix $\mathbf{C}_{\mathbf{Y}} = E(\mathbf{Y}\mathbf{Y}^{\top})$ that belong to $\mathbf{Y}$. Furthermore, let $\mathbf{D} = \mathrm{diag}(d_1, \ldots, d_p)$ be a diagonal $p \times p$ matrix of p eigenvalues of $\mathbf{C}_{\mathbf{Y}}$. Then, a linear whitening transform is given by $\mathbf{Q} = \mathbf{D}^{1/2}\mathbf{E}^{\top}$, where $\mathbf{Q}$ is a $p \times m$ matrix, $\mathbf{D}$ is a $p \times p$ matrix, and $\mathbf{E}$ is a $m \times p$ matrix. For more explicit description that relate to whitening and pre-processing for ICA, see Hyvärinen et al. (2001).

[3] This is the reason why I use the determinant in Equation 4.15. For more information about this, see a standard textbook in probability theory under transformation of random variables.

is a function of $\tilde{\mathbf{A}}$ and a product of the density evaluated at T points:

$$\mathcal{L}(\tilde{\mathbf{A}}) = \prod_{t=1}^{T}\prod_{i=1}^{n} f_i(\tilde{\mathbf{a}}_i^\top \mathbf{Y}_t)|\det \tilde{\mathbf{A}}|. \tag{4.16}$$

The log-likelihood is expressed as:

$$\ell(\tilde{\mathbf{A}}) = \sum_{t=1}^{T}\sum_{i=1}^{n} \log f_i(\tilde{\mathbf{a}}_i^\top \mathbf{Y}_t) + T\log|\det \tilde{\mathbf{A}}|. \tag{4.17}$$

Finally, I use a gradient-based method to maximize the likelihood function numerically.

Note that the likelihood is a function of $\tilde{\mathbf{A}}$ and the factors probability densities f_i, $i = 1, \ldots, m$. I assume that the factors all have the same symmetric heavy-tailed densities, which I find suitable since I work with monthly data (Shah, 2013). Thus, let $f = f_i, i = 1, \ldots, m$ be the parameter-free reciprocal cosh density defined as:

$$f(\mathbf{s}_i) = \frac{1}{\pi \cosh(\mathbf{s}_i)},\ i = 1, \ldots, m. \tag{4.18}$$

In PCA, one may select the number of factors of interest based on the variance explained by each factor. To do the same in ICA, I let $\{a_{ij}\}, i, j = 1, \ldots, m$ be the elements in the mixing matrix $\mathbf{A}$. Then, I define the variance-accounted for by each factor as:

$$\gamma_i = T\frac{\|\mathbf{a}_i^\top\|_2^2}{\|\mathbf{Y}\|_F^2} \in [0,1],\ i = 1, \ldots, m, \tag{4.19}$$

where I denote $\gamma_{(1)}, \gamma_{(2)}, \ldots, \gamma_{(m)}$ as decreasing values in range $[0, 1]$ that correspond to the variance-accounted for by each factor. The numerator refers to the ℓ_2-norm, whereas the denominator refers to the Frobenius norm.

4.3 Performance Evaluation

To evaluate the performance of the models, I want to measure the fit in terms of a numerical value. Thus, I measure the in-sample fit as root-mean-square-error (RMSE) defined as:

$$\text{RMSE}(\tau) = \sqrt{\frac{\sum_{t=1}^{T}(y_t(\tau) - \hat{y}_t(\tau))^2}{T}}, \tag{4.20}$$

where τ is the maturity, y_t are the observed yields, and $\hat{y}_t$ are the model's fitted yields. Furthermore, the out-of-sample (OOS) performance measure is based on root-mean-square-forecast-error

(RMSFE) defined as:

$$\text{RMSFE}(h,\tau) = \sqrt{\frac{\sum_{t=1}^{n}(y_{t+h}(\tau) - \hat{y}_{t+h,t}(\tau))^2}{n}}, \tag{4.21}$$

where $t = t_0, \ldots, T$ is a total of n h-step forecasts for a given maturity τ, $h = \{1, 3, 6, 12\}$ is the forecast horizon in months, $y_{t+h}(\tau)$ is the observed OOS yield for maturity τ, and $\hat{y}_{t+h,t}(\tau)$ is the model's yield forecast.

To test whether the two models differ in predictive accuracy, I use the Diebold and Mariano (1995) test. I define the two models forecast errors as:

$$\begin{aligned} e^{(1)}_{t+h,t} &= y_{t+h} - \hat{y}^{1}_{t+h,t}, \\ e^{(2)}_{t+h,t} &= y_{t+h} - \hat{y}^{2}_{t+h,t}, \end{aligned} \tag{4.22}$$

where $t = t_0, \ldots, T$ is a total of n h-step forecasts. The Diebold-Mariano test is based on the loss differential, where I use a squared error loss function, $d_t = (e^{(1)}_{t+h,t})^2 - (e^{(2)}_{t+h,t})^2$, and the null hypothesis of equal predictive accuracy is:

$$H_0 : E[d_t] = 0,$$

against the alternative hypothesis:

$$H_1 : E[d_t] \neq 0.$$

Furthermore, the test statistic is defined as:

$$DM = \frac{\bar{d}}{\sqrt{\widehat{LRV}_{\bar{d}}/T}}, \tag{4.23}$$

where $\bar{d} = \frac{1}{n}\sum_{t=t_0}^{T} d_t$ and $LRV_{\bar{d}} = Var(d_t) + 2\sum_{i=1}^{\infty} Cov(d_t, d_{t-i})$.[4] The reason to include covariance terms in the test statistic is that the h-step forecasts are serially correlated due to overlapping data (if $h > 1$). Then, Diebold and Mariano (1995) show that under the null of equal predictive accuracy, $DM \sim N(0, 1)$. Harvey et al. (1997) modify the test such that it is better for smaller sample sizes and propose the following statistic:

$$DM_{HLN} = DM\sqrt{\frac{T + 1 - 2h + (h/T)(h-1)}{T}}, \tag{4.24}$$

which is asymptotically t-distributed with $T - 1$ degrees of freedom. In this paper, I use the Harvey et al. (1997) statistic because of the small sample sizes.

[4] $\widehat{LRV}_{\bar{d}}$ is a consistent estimate of the asymptotic variance of $\bar{d}\sqrt{T}$ (Diebold and Mariano, 1995).

Section 5

Models

In this section, I present four different yield curve models. The first is the parametric Nelson and Siegel (1987) model (NSM), which is a model that many practitioners and central banks (BIS, 2005; ECB, 2008) use to construct yield curves. Another parametric model is the dynamic NSM (DNSM). Finally, I propose two decomposition-based models where I use ICA and PCA to extract the factor loadings.

5.1 Nelson-Siegel Model

Many practitioners and central banks use NSM, or some slight variant, for fitting bond yields (BIS, 2005; ECB, 2008). NSM expresses a set of yields of various maturities as a function of three hyperparameters, where a fourth parameter is a decay parameter. I denote the set of N yields as $y_t(\tau)$, $t = 1, \ldots, N$, where τ denotes the maturity in months. The NSM specification is expressed as:

$$y_t(\tau) = \mathbf{f}_1 + \mathbf{f}_2 \left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} \right) + \mathbf{f}_3 \left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau} \right), \tag{5.1}$$

where $\mathbf{f}_1, \mathbf{f}_2, \mathbf{f}_3$, and λ are time-invariant parameters. The decay parameter λ determines the maturity at which the loading on the medium-term, or curvature, factor achieves it maximum. For example, $\lambda = 0.0609$ is the value that maximizes the loading on the medium-term factor at precisely 30 months, i.e., the average of two- and three-year maturities (Diebold and Li, 2006). However, it is possible to estimate λ given observed data.

5.2 Dynamic Nelson-Siegel Model

NSM is a simple and effective model. However, time-invariant parameters may be infeasible in adapting to various market conditions. Therefore, Diebold and Li (2006) propose a dynamic latent factor model, DNSM, in which $\mathbf{f}_1$, $\mathbf{f}_2$, and $\mathbf{f}_3$ are time-varying level, slope, and curvature factors. DNSM is specified as:

$$y_t(\tau) = \mathbf{f}_{1,t} + \mathbf{f}_{2,t}\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau}\right) + \mathbf{f}_{3,t}\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}\right), \tag{5.2}$$

where $\mathbf{f}_{1,t}$, $\mathbf{f}_{2,t}$, and $\mathbf{f}_{3,t}$, $t = 1, \ldots, T$ are now time-varying factors. I assume that these dynamic latent factor loadings follow a vector autoregressive (VAR) process of first order and the full model can be written as:[1]

$$\begin{aligned} \mathbf{y}_t &= \mathbf{\Lambda}^{\text{DNSM}}\mathbf{f}_t^{\text{DNSM}} + \boldsymbol{\varepsilon}_t^{\text{DNSM}}, \\ \mathbf{f}_t^{\text{DNSM}} - \boldsymbol{\mu} &= \mathbf{\Gamma}^{\text{DNSM}}(\mathbf{f}_{t-1}^{\text{DNSM}} - \boldsymbol{\mu}) + \boldsymbol{\eta}_t^{\text{DNSM}}, \end{aligned} \tag{5.3}$$

where $\mathbf{y}_t$ is the observed yields at time t, $\mathbf{\Lambda}^{\text{DNSM}}$ is the factor loadings of the latent factors $\mathbf{f}_t^{\text{DNSM}}$, $\mathbf{\Gamma}^{\text{DNSM}}$ is the transition matrix, $\boldsymbol{\varepsilon}_t^{\text{DNSM}}$ and $\boldsymbol{\eta}_t^{\text{DNSM}}$ are the measurement and state-space specification errors with covariance matrices $\mathbf{H}^{\text{DNSM}}$ and $\mathbf{Q}^{\text{DNSM}}$ respectively. I assume that the white noise transition and measurement errors are orthogonal to one another:

$$\begin{pmatrix} \boldsymbol{\varepsilon}_t^{\text{DNSM}} \\ \boldsymbol{\eta}_t^{\text{DNSM}} \end{pmatrix} \sim N\left[\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{H}^{\text{DNSM}} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}^{\text{DNSM}} \end{pmatrix}\right], \tag{5.4}$$

and to the initial state where $E[\mathbf{f}_0(\boldsymbol{\varepsilon}_t^{\text{DNSM}})^\top] = 0$ and $E[\mathbf{f}_0(\boldsymbol{\eta}_t^{\text{DNSM}})^\top] = 0$.

As in the original papers (Diebold and Li, 2006; Diebold et al., 2006), I assume that $\mathbf{H}^{\text{DNSM}}$ is diagonal and $\mathbf{Q}^{\text{DNSM}}$ is non-diagonal. The assumption of a diagonal $\mathbf{H}^{\text{DNSM}}$ matrix means that the pricing errors for yields of various maturities are uncorrelated. The assumption of a non-diagonal $\mathbf{Q}^{\text{DNSM}}$ matrix allows the shocks to the three factors to be correlated. This model fits into the KF framework as described in section 4.1.

Under quadratic loss the optimal forecast is the conditional expectation, i.e., the optimal forecast at time t for time $t+h$ is:

$$\mathbf{y}_{t+h,t} = E_t[\mathbf{y}_{t+h}] = \mathbf{\Lambda}^{\text{DNSM}} E_t[\mathbf{f}_{t+h}^{\text{DNSM}}]. \tag{5.5}$$

In practice, I replace $E_t[\mathbf{f}_{t+h}^{\text{DNSM}}]$ with KF forecasts to obtain $\hat{\mathbf{y}}_{t+h,t}$.

I estimate the model parameters $\{\lambda, \mathbf{H}^{\text{DNSM}}, \mathbf{Q}^{\text{DNSM}}, \mathbf{\Gamma}^{\text{DNSM}}\}$, which corresponds to estimate $1 + N + p(p+1)/2 + p^2$ parameters (with $N = 17$ and $p = 3$). Furthermore, I estimate the parameters with MLE using quasi-Newton optimizer (L-BFGS-B). Filtered estimates of the factors are estimated with KF, which I initialize using the unconditional mean and covariance matrix of the state vector. I maximize the likelihood by iterating the L-BFGS-B algorithm, where I impose a non-

[1] I omit intercept parameters and center the observed yields to have mean equal to zero because that is needed for the other models. However, it is easy to add back the mean to the forecasts to get interpretable results.

negativity constraint on all estimated variances to estimate log variances, and I compute asymptotic standard errors using the delta method. I obtain startup parameter values by estimating a VAR(1) model for the factors to obtain the initial transition matrix and initialize entries in the covariance matrix with a value equal to 100.

5.3 PCA Yield Model

In DNSM, the parametric factor loadings require a parameter λ, which is estimated jointly in KF based on observed data. However, it is possible to use PCA to extract these factor loadings non-parametrically. Hence, PCA gives the modeler an attractive alternative to DNSM. When λ is estimated, the loading matrix is fixed. In other words, it finds the most suitable loading matrix given the data. However, in the PCA approach, I obtain the factor loadings ($\mathbf{\Lambda}^{\text{PCA}}$) and its factors ($\mathbf{f}_t^{\text{PCA}}$) directly from the observed yields ($\mathbf{y}_t$). This is based on Equation 4.11 with $\mathbf{f}_t^{\text{PCA}} = \mathbf{P}$, $\mathbf{\Lambda}^{\text{PCA}} = \mathbf{W}$, and $\mathbf{y}_t = \mathbf{Y}$. Note that I choose $p = 3$ factors, i.e., choose the three first columns of $\mathbf{P}$ and $\mathbf{W}$.

After that, following Diebold and Li (2006), I fit a univariate AR(1) model to each of the estimated factors rather than a VAR(1). The reason is that one might expect the forecasts to be superior with univariate AR(1) since unrestricted VARs tend to produce poor forecasts due to the many parameters (and small sample) and their potential for overfitting. Also, the factors are not highly correlated, so it is appropriate with a set of univariate models (Diebold and Li, 2006). Hence, I express the PCA yield model (PCAYM) as:

$$\mathbf{f}_{t,i}^{\text{PCA}} = \Gamma_i^{\text{PCA}} \mathbf{f}_{t-1,i}^{\text{PCA}} + \boldsymbol{\eta}_{t,i}^{\text{PCA}}, \tag{5.6}$$

where $i = 1, 2, 3$ denote the index to level, slope, and curvature factors respectively. Also, $\mathbf{f}_{t,i}^{\text{PCA}}$ is the latent factor i at time t, Γ_i^{PCA} is the model parameter for latent factor i, and $\boldsymbol{\eta}_{t,i}^{\text{PCA}}$ is white noise for latent factor i. The forecasting procedure is similar to Equation 5.5.

5.4 ICA Yield Model

PCA extracts uncorrelated factors with maximal variance. However, the factors may have some higher-order dependence that may mask useful information, and it is also more useful to study independent factors because one can analyze them one at a time. The standard procedure in the ICA literature is first to reduce the dimension of the dataset through PCA and then transform these variables into independent factors. The first step is important since ICA tends to produce estimates of independent factors that have a single spike and fluctuate around zero elsewhere (Hyvärinen et al., 2001, Ch. 13). This problem refers to as overlearning in the ICA literature and may be the case if one apply ICA to the full dataset. In other words, one may consider this to be overfitting.[2]

Recall from the previous section, that I assume the latent factors have a reciprocal cosh density, which has a higher kurtosis than the normal distribution. Hence, I get factors that are higher-order independent (up to the fourth order).[3] In this way, I obtain the factor loadings ($\boldsymbol{\Lambda}^{\text{ICA}}$) and its factors ($\mathbf{f}_t^{\text{ICA}}$). This is based on Equation 4.12, 4.13, and 4.14 with $\mathbf{f}_t^{\text{ICA}} = \mathbf{S}$, $\boldsymbol{\Lambda}^{\text{ICA}} = \mathbf{A}$, and $\mathbf{y}_t = \mathbf{Y}$. Note that I choose $p = 3$, i.e., I select the first three factors.

After that, I fit a univariate AR(1) model to each of the estimated factors similar as in the PCAYM. I refer this model to the ICA yield model (ICAYM), and I express it as:

$$\mathbf{f}_{t,i}^{\text{ICA}} = \Gamma_i^{\text{ICA}} \mathbf{f}_{t-1,i}^{\text{ICA}} + \boldsymbol{\eta}_{t,i}^{\text{ICA}}, \tag{5.7}$$

where $i = 1, 2, 3$ denote the index to level, slope, and curvature factors respectively. Also, $\mathbf{f}_{t,i}^{\text{ICA}}$ is the latent factor i at time t, Γ_i^{ICA} is the model parameter for latent factor i, and $\boldsymbol{\eta}_{t,i}^{\text{ICA}}$ is white noise for latent factor i.

Finally, to get an economic interpretation of these factors, I correlate the estimated factors with macroeconomic variables. Also, I fit a VAR(1) model of the factors together with the macro variables and produce impulse response functions to examine the factors response to shocks in the macro variables and vice versa.

[2] I did apply the entire dataset to ICA and found that the last two factors have precisely these single spikes, and therefore opted against using it.

[3] I want to point out that this is not the same as PCA and rotation after that, such as varimax.

Section 6

Data

I use U.S. zero-coupon yields[1] as provided by ICAP, through Thomson Reuters Datastream, for 17 maturities: 3, 6, 9, 12, 15, 18, 21, 24, 30, 36, 48, 60, 72, 84, 96, 108, and 120 months. The reported data is at a daily frequency, but I resample it at a monthly rate, which is more commonly used in the previous literature (e.g., Diebold and Li, 2006; Diebold et al., 2006; Christensen et al., 2011; Joslin et al., 2014). I present the descriptive statistics of the yield curve data in Table 6.1.

Table 6.1: Descriptive statistics of the yield curve data with maturities denoted in months.

Maturity	Mean	Std. dev.	Minimum	Maximum	$\hat{\rho}(1)$	$\hat{\rho}(12)$	$\hat{\rho}(30)$
3	2.593	2.335	0.224	7.140	0.988	0.757	0.257
6	2.625	2.327	0.229	7.441	0.990	0.763	0.267
9	2.672	2.321	0.239	7.588	0.990	0.770	0.282
12	2.725	2.311	0.259	7.652	0.990	0.778	0.301
15	2.788	2.294	0.281	7.686	0.990	0.783	0.326
18	2.852	2.273	0.303	7.706	0.989	0.788	0.350
21	2.917	2.250	0.327	7.725	0.988	0.793	0.374
24	2.983	2.228	0.345	7.733	0.987	0.796	0.397
30	3.114	2.184	0.389	7.747	0.986	0.802	0.431
36	3.246	2.142	0.431	7.766	0.985	0.806	0.464
48	3.471	2.040	0.562	7.730	0.983	0.812	0.505
60	3.676	1.964	0.752	7.742	0.982	0.811	0.531
72	3.851	1.898	0.962	7.767	0.980	0.807	0.548
84	4.000	1.844	1.129	7.766	0.980	0.803	0.560
96	4.126	1.800	1.206	7.779	0.979	0.799	0.568
108	4.233	1.768	1.277	7.792	0.978	0.795	0.574
120	4.328	1.742	1.341	7.791	0.977	0.792	0.580

The sample period is 1997:03-2017:08 (246 monthly observations), see Figure 6-1. The yield curve construction is based on a number of financial instruments, where the short-term rates are based on LIBOR[2], while the mid and long-term rates are constructed based on forward rate agreements[3] and swaps[4] respectively. They are smoothed by being fitted to a spline.

Previous literature uses unsmoothed Fama and Bliss (1987) zero-coupon yields that are based on

[1] Thanks to Marcin Zamojski for providing the data.

[2] LIBOR (London Interbank Offered Rate) is a benchmark rate that the leading banks charge each other for short-term loans.

[3] A forward rate agreement is an over-the-counter contract between two counterparties that determines some rate

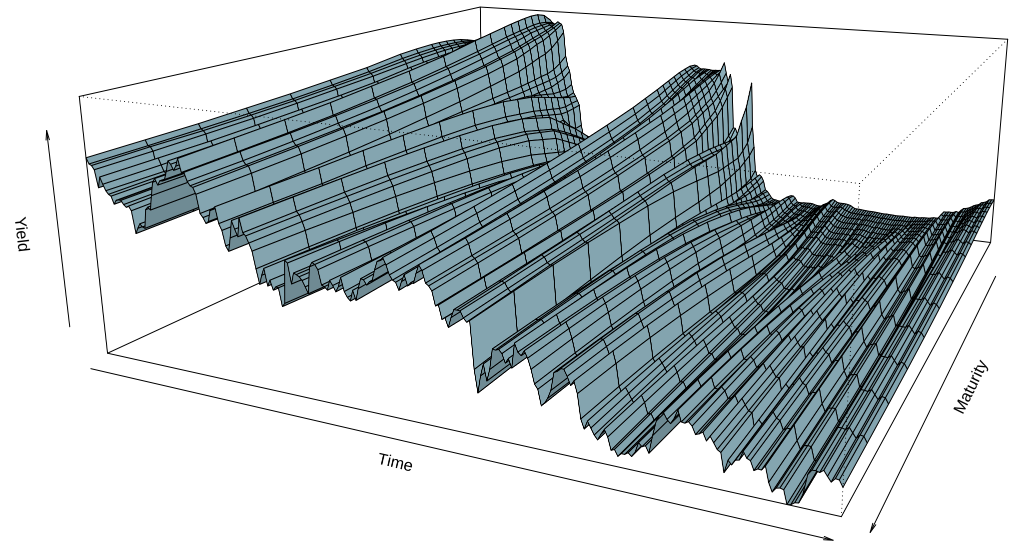

Figure 6-1: Three-dimensional plot of the yield curves where the axes are yield, maturity and time. The yield range is 0.22 to 7.79, the maturity range is 3 to 120 months, and the period is 1997:03-2017:08.

sovereign bonds and derived from bid-ask average price quotes (e.g. Diebold et al., 2006; Christensen et al., 2011). The filtered paths of the latent factors are similar to those obtained with the ICAP data. However, since the data is not treasury-based, the level is shifted upwards to reflect the interbank counterparty risk premium (Koopman et al., 2017). For this paper, the importance lies in the fact that the data contains the relevant dynamic characteristics since I examine the ability to provide proper guidance about the yield curve dynamics and how the factors interact with macroeconomic variables as long as the counterparty risk premium is stable in time.

The macroeconomic variables I use to correlate with the estimated factors are selected from the publicly accessible monthly macroeconomic database FRED-MD given by the Federal Reserve Bank of St. Louis (McCracken and Ng, 2016). These variables are chosen based on previous literature, theories that connect them to interest rates, their presumed importance for the yield curve, and also their connection to the level, slope, and curvature factors.

As the level factor relates to inflation expectations, I include variables such as Real Consumption Expenditures and CPI, which often serve as proxies for inflation. Ang and Piazzesi (2003) include three inflation measures with one of them being CPI. Both Diebold et al. (2006) and Joslin et al. (2014) use only one inflation measure. Also, the Unemployment Rate is presumed to have an inverse relationship to inflation, but it can also relate to real activity (Ang and Piazzesi, 2003). Therefore, I include Unemployment Rate. Previous literature also states that the slope factor relates to real

of interest to be paid, or received, on an obligation beginning at a future date.

[4]A swap is a derivative contract between two counterparties that specifies an exchange of payments benchmarked against some rate or index.

economic activity (Diebold et al., 2006), and therefore I include Capacity Utilization and Federal Funds Rate since it is a monetary policy instrument. Diebold et al. (2006) include these two variables, whereas Joslin et al. (2014) use another proxy for real economic activity. In addition, I include several more variables that relate to inflation or economic activity. However, other variables can be used such as those that link to unconventional monetary policy, but I do not include them due to conciseness and lack of time.

In Table 6.2, I provide the selected macro variables and give brief descriptions of them. For more information about the data transformations, I refer to McCracken and Ng (2016).

Table 6.2: Macroeconomic variables from the FRED-MD database that I use to correlate with the estimated factors.

Macroeconomic variable	Description
Real Income	The number of goods and services one can buy today compared to the price of the same goods and services in another period.
Real Consumption Expenditures	Price changes in consumer goods and services and sometimes used as a proxy for inflation.
Capacity Utilization: Manufacturing	Proportion of potential economic output that is actually realized, i.e., the real economic activity.
Unemployment Rate	Share of the labor force that is jobless, expressed as a percentage.
Housing Starts	Number of new residential construction projects that have begun during any particular month.
Real M2 Money Stock	Money supply that includes cash, checking deposits, savings deposits, money market securities, mutual funds, and other time deposits.
Commercial and Industrial Loans	Commercial and industrial loans at all commercial banks in the U.S.
Real Estate Loans, All Commercial Banks	Real estate loans at all commercial banks in the U.S.
Federal Funds Rate	The rate banks lend reserve balances to other banks overnight. It is a monetary policy instrument that influence short-term rates.
Trade Weighted U.S. Dollar	Foreign exchange value of the U.S. dollar compared against certain foreign currencies.
CPI: All Items	Weighted average of prices of a basket of consumer goods and services. It is often used as a proxy for inflation.
Consumer Sentiment Index	The economy's overall health, determined by consumer opinion.

Section 7

Results

7.1 Estimation of Factor Loadings and Latent Factors

Based on Figure 7-1, the first three PCA factors explain 97% of the total variation in the yield curve data, whereas an additional fourth PCA factor help explain the variation by an additional 1.5 p.p. This is in line with previous literature and that these first three factors relate to the level, slope, and curvature (e.g., Litterman and Scheinkman, 1991). These three factors are used to construct the ICA factors and its loadings.

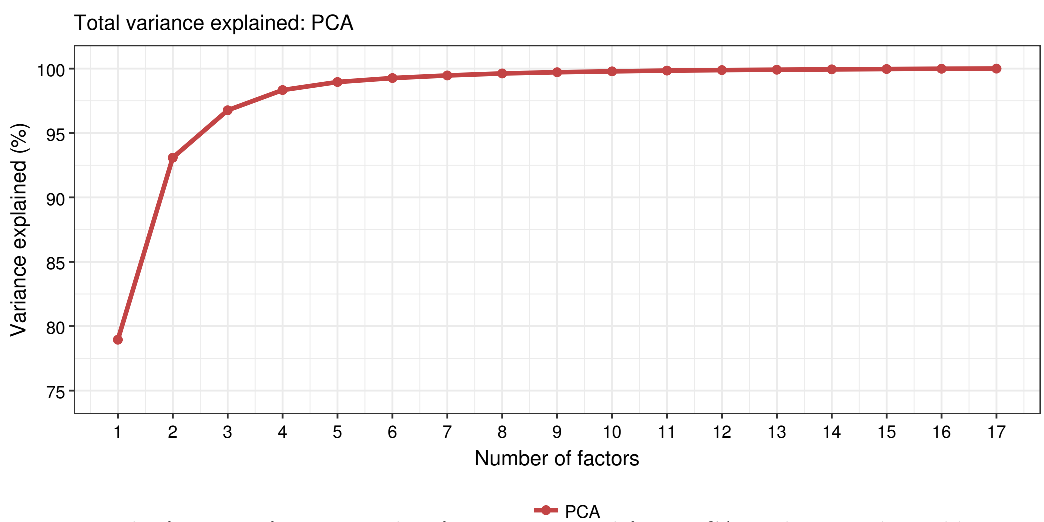

Figure 7-1: The fraction of variation that factors extracted from PCA explains in the yield curve data.

The first three factor loadings from DNSM, PCA, and ICA can be seen in the left panel of Figure 7-2.[1] The loading on the level is different between the methods because the DNSM level factor loads equally across maturities by design where one unit increase in this factor results in that all maturities increase by one basis point. The PCA and ICA level factors load more on the long

[1] Note that some of the factors are inverted and (or) shifted for visualization purposes.

end since the level factor relates to the long-term maturities. The slope loadings have similar shapes for all methods and loads more on the short end and therefore relates to the short-term maturities. Although PCA and ICA loadings may look different it is due to the scale which is an ambiguity in the ICA method.[2] The curvature loadings also have similar shapes for all methods and loads more on 20 to 30 months and therefore relate to the mid-term maturities.

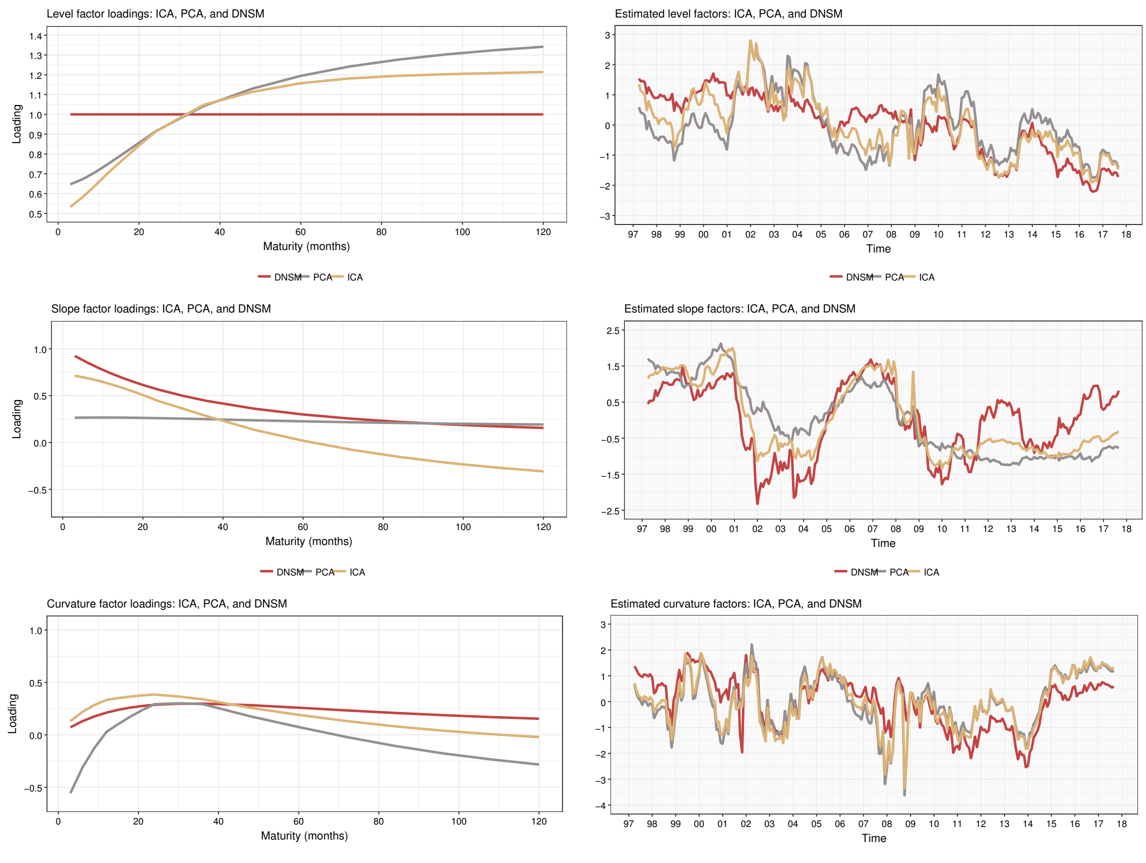

Figure 7-2: Factor loadings for the DNSM, PCA, and ICA (left column) and the estimated factors for DNSM, PCA, and ICA (right column). Note that some of the loadings and factors are inverted and (or) shifted for visualization purposes.

As can be seen in the right panel of Figure 7-2, the estimated latent factors from DNSM, PCA, and ICA are highly correlated (especially the PCA and ICA factors). The level factors match the stylized fact that the interest rates were falling in this period. There seem to be two large spikes in the dataset. The first spike is at the end of 2001 and may be due to U.S. decision to end the regular sale of 30-year bond making the yield move to three-year lows.[3] This decision followed other significant events such as the Enron scandal and the 9/11 attack. The second spike seems to relate to the financial crisis in 2008.

[2] See section 4 for clarification regarding the ambiguities of ICA. For more explicit details, see Hyvärinen et al. (2001).

[3] `http://money.cnn.com/2001/10/31/markets/longbond/index.htm`, `https://www.nytimes.com/2001/11/01/business/us-will-end-regular-sale-of-long-bond.html`, and `http://money.cnn.com/2001/11/15/markets/bonds/`

The correlation between ICA slope factor and Fed Funds Rate is very high. The same goes for the PCA level factors. This high correlation with Fed Funds Rate may be because it loads more on the short end, and perhaps also that the dataset has more short-term instruments. However, I provide a discussion that relates to the zero lower bound later in this section, but it is important to have this in mind when I forecast future yields because I use nominal rates and not shallow rates since it can reflect the forecast performance.[4]

The estimate of the parameters in ICAYM indicates highly persistent dynamics with the lagged level, slope, and curvature factors where the significant ($p < 0.001$) coefficient estimates are 0.982, 0.936, and 0.893 respectively. These results are in line with previous findings for other models and other datasets (e.g. Diebold et al., 2006; Christensen et al., 2011; Joslin et al., 2014).

Table 7.1 contains in-sample fit for the ICA, PCA, and DNSM, which fit the data well and deviate only by a few basis points[5] across all 17 maturities on average. Note that the ICA and PCA approach has the same in-sample fit which is by construction (see section 4 and section 5). Although they have the same in-sample fit, I show later on that they have different out-of-sample forecasts.

All models fit the yield curve well. The worst fit is the three-month maturity for DNSM, which deviates 32 basis points where both ICA and PCA have a relative RMSE below one. Also, DNSM seems to fit the yield curve better in the mid-term, whereas ICA (and PCA) seem to fit better in both the short and long end. Also, there are some maturities where DNSM targets three maturities and have a perfect fit, i.e., RMSE close to zero. Notably, these maturities are consistent with how the factors relate to the short, mid, and long-term yields.

Table 7.1: In-sample fit for ICA, PCA, and DNSM. RMSE (in basis points) and relative RMSE for 17 maturities (in months). The estimated λ in DNSM is 0.05348 (0.00062).

	Relative RMSE			RMSE		
	ICA	PCA	DNSM	ICA	PCA	DNSM
3	0.26	0.26	1.00	8.32	8.32	32.53
6	0.13	0.13	1.00	2.35	2.35	18.65
9	0.41	0.41	1.00	3.74	3.74	9.07
12	2.52	2.52	1.00	6.38	6.38	2.53
15	67.67	67.67	1.00	4.81	4.81	0.07
18	2.89	2.89	1.00	3.36	3.36	1.16
21	1.47	1.47	1.00	2.40	2.40	1.63
24	1.15	1.15	1.00	2.82	2.82	2.45
30	212.36	212.36	1.00	2.92	2.92	0.01
36	2.01	2.01	1.00	4.54	4.54	2.26
48	1.21	1.21	1.00	7.04	7.04	5.80
60	1.24	1.24	1.00	5.52	5.52	4.45
72	1.38	1.38	1.00	3.33	3.33	2.41
84	147.14	147.14	1.00	1.30	1.30	0.01
96	0.83	0.83	1.00	1.97	1.97	2.38
108	0.86	0.86	1.00	4.03	4.03	4.67
120	0.89	0.89	1.00	6.07	6.07	6.81

[4] The models forecasts can be negative, whereas the Fed use unconventional monetary policy to stimulate the economy rather than having a (too) negative rate.

[5] 1 basis point equals 0.01%.

7.2 Forecasting the Yield Curve

To compare the different models' ability to forecast future yields, I use data starting from March 1997 to January 2010 and produce one-month-ahead, three-month-ahead, six-month-ahead, and twelve-month-ahead forecasts. After that, I extend the sample on a monthly basis until; August 2016 to obtain 80 twelve-month-ahead forecasts, February 2017 to obtain 86 six-month-ahead forecasts, May 2017 to obtain 89 three-month-ahead forecasts, and July 2017 to obtain 91 one-month-ahead forecasts. These forecasts serve as a basis for calculating RMSFEs and Diebold and Mariano (1995) statistics for different models at different forecast horizons across maturities.

In Table 7.2, I report that ICAYM and PCAYM forecast short-term yields significantly better than DNSM at a long forecast horizon ($h = 12$ months) and is relevant to central banks due to their focus on monetary policy. This improved forecast performance may be since both ICAYM and PCAYM slope factor follows the Fed Funds Rate closely. DNSM produce significantly better forecasts than ICAYM and PCAYM for a short forecast horizon ($h = 1$ month) across all maturities. This result may be due overfitting and large persistence since it does not produce as good forecasts in the longer horizons. Also, I do not intend to compare the ICAYM and PCAYM because DNSM is known in the literature for its forecasting ability (Diebold and Li, 2006).[6]

7.3 Contemporaneous Correlation with Macroeconomic Variables

To get economic interpretations of the level, slope, and curvature factors I relate them to macroeconomic variables since various economic forces change the yield curves. In other words, to determine how factors are driven by changes in macroeconomic variables (that one can directly set) and to determine how those changes propagate to other macro variables through the factors/yield curve.

A total of six macroeconomic variables seem to be non-stationary, and the correlations between estimated factors and these macro variables can be spurious, e.g., CPI only correlate with the level as the level seems to be decreasing. These variables are; Real Personal Income, Real Personal Consumption Expenditures, Real Money Stock, Commercial and Industrial Loans, Real Estate Loans, and CPI. All variables seem to fluctuate around a linear trend in the whole period (1997-2017). Therefore, I detrend these variables and make them stationary, by regressing the variables on the time index and correlate the estimated factors with the residuals from the linear regressions.[7]

In Figure 7-3, the level factor seems to be lacking strong correlations over the whole period.

[6] I do not show it in this paper, but ICA produces significantly better forecasts for the one-month-ahead and three-month-ahead forecasts across all maturities. However, there are no significant differences for the six-month-ahead and twelve-month-ahead forecasts.

[7] There are many different ways of making a non-stationary time series stationary, where another approach is to calculate differences between two adjacent time points across the whole dataset.

Table 7.2: Out-of-sample fit in terms of RMSFE (in basis points) and Diebold-Mariano statistics for ICAYM, PCAYM, and DNSM. I report both RMSFE and relative RMSFE for all models. Note that the DNSM is the reference model, and a negative sign for the Diebold-Mariano statistic is in favor of the compared model.

	Forecast horizon (months)							
	$h = 1$		$h = 3$		$h = 6$		$h = 12$	
Model	RMSFE	D-M	RMSFE	D-M	RMSFE	D-M	RMSFE	D-M
Three-month yield:								
ICAYM	2.11	6.89^{***}	1.25	1.29	0.74	-2.62^{*}	0.77	-3.71^{***}
PCAYM	2.27	7.72^{***}	1.47	2.18^{*}	0.84	-1.00	0.44	-2.94^{**}
DNSM	25.23	Ref.	34.74	Ref.	49.28	Ref.	59.52	Ref.
Six-month yield:								
ICAYM	3.12	7.06^{***}	1.44	1.79	0.83	-1.52	0.85	-1.52
PCAYM	3.40	7.78^{***}	1.77	2.67^{**}	1.00	0.01	0.53	-2.17^{*}
DNSM	17.65	Ref.	30.79	Ref.	43.94	Ref.	54.08	Ref.
Nine-month yield:								
ICAYM	4.06	7.15^{***}	1.60	2.05^{*}	0.91	-0.75	0.93	-0.56
PCAYM	4.47	7.85^{***}	2.03	2.96^{**}	1.15	0.67	0.58	-1.67
DNSM	13.93	Ref.	27.73	Ref.	39.47	Ref.	50.45	Ref.
One-year yield:								
ICAYM	4.58	7.21^{***}	1.71	2.19^{*}	0.96	-0.31	0.99	-0.05
PCAYM	5.09	7.92^{***}	2.25	3.16^{**}	1.26	1.06	0.60	-1.44
DNSM	12.45	Ref.	25.25	Ref.	35.90	Ref.	48.23	Ref.
Three-year yield:								
ICAYM	3.09	7.21^{***}	1.25	1.18	0.96	-0.25	1.18	0.69
PCAYM	3.55	8.33^{***}	1.72	2.67^{**}	0.91	-0.40	0.50	-2.07^{*}
DNSM	16.50	Ref.	28.31	Ref.	41.29	Ref.	57.23	Ref.
Five-year yield:								
ICAYM	2.43	6.41^{***}	1.06	0.38	0.95	-0.36	1.13	0.51
PCAYM	2.76	7.46^{***}	1.29	1.27	0.79	-1.02	0.67	-2.09^{*}
DNSM	20.24	Ref.	36.57	Ref.	53.03	Ref.	68.62	Ref.
Seven-year yield:								
ICAYM	2.38	6.25^{***}	1.05	0.31	0.94	-0.53	1.08	0.35
PCAYM	2.68	7.08^{***}	1.24	1.13	0.84	-0.94	0.76	-1.70
DNSM	21.41	Ref.	40.52	Ref.	58.99	Ref.	75.60	Ref.
Ten-year yield:								
ICAYM	2.26	6.19^{***}	1.04	0.26	0.92	-0.85	1.03	0.19
PCAYM	2.53	6.92^{***}	1.23	1.17	0.89	-0.77	0.82	-1.45
DNSM	23.74	Ref.	45.00	Ref.	65.13	Ref.	82.96	Ref.

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Therefore, I construct a rolling window of five-year period[8] correlations to investigate further whether there are periods of high correlation for specific macroeconomic variables. See the bottom of Figure 7-3. Around the financial crisis in 2008, there was a strong positive correlation between the level factor and the Money Stock. After that, it has been fluctuating around the zero mark. Also, there were strong negative correlations between the level factor and Real Income, Real Consumption Expenditures, and CPI. In 2009, the correlation was decreasing and fluctuate around that point

[8] I choose a five-year period because that corresponds to 60 observations and is a reasonable length to calculate a correlation between two variables.

until recent years starting to rise back to its previous levels. A remarkable fact is that while both Real Income and Real Consumption Expenditures are going back to these previous levels, CPI moved the other way from negative to a positive correlation.

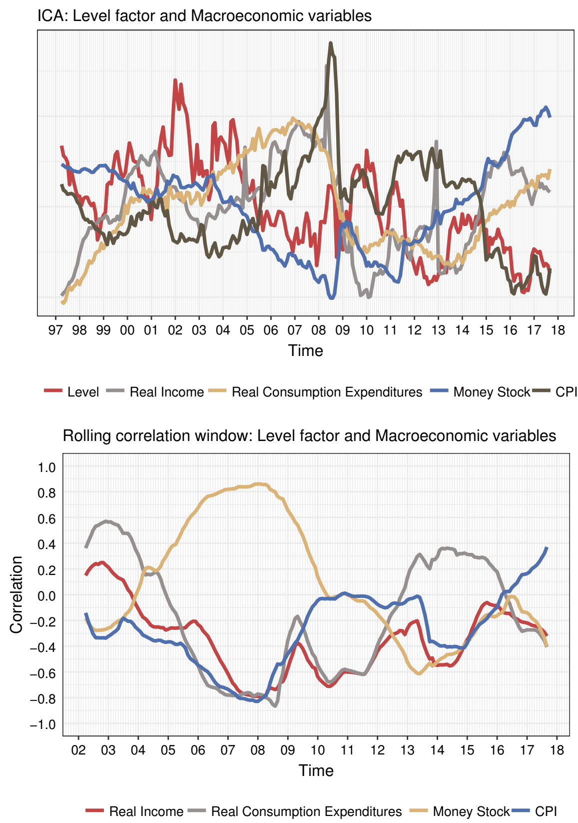

Figure 7-3: Top: Estimated level factor and Real Income, Real Consumption Expenditures, Money Stock, and CPI. Bottom: Rolling window of five-year period correlations between the estimated level factor and Real Income, Real Consumption Expenditures, Money Stock, and CPI.

As can be seen in Figure 7-4, the slope factor seems to have a strong correlation between both Capacity Utilization and Federal Funds Rate over the whole period. These results are in line with previous literature (e.g., Diebold et al., 2006). Furthermore, the zero lower bound impact from 2009 reflect that the Fed lowered Federal Funds Rate close to zero leaving them with alternative approaches to stimulate the economy because people can always hold cash rather than negative-yielding assets. Thus, the Fed stimulated the economy through quantitative easing, which is an unconventional monetary policy action where they engage in a large-scale asset buying program of government bonds to keep the short-term rates low and push down the long-term rates, which further incentives to borrowing, growth, and spending. Although it is possible to control for unconventional monetary policy (e.g., see Mesters et al., 2014), I do not include such variables.

Also, in the bottom of Figure 7-4, I construct a rolling window of five-year period correlations to investigate further whether there are varying periods of high correlation for these macroeconomic variables. There is a strong positive correlation between the estimated slope factor and both Capacity

Utilization and Federal Funds Rate from 2002 to 2012. After that, Capacity Utilization seems to have a high correlation, whereas Fed Funds Rate appears to go from uncorrelated to highly positive correlate again recovering post-2016, after Fed started increasing the interest rates. These results are in line with the zero lower bound discussion I provide above. I expect the effect to show up in regressions later and a possible way of dealing with it is to include a dummy interaction for a few years. However, I did not have time to do this.

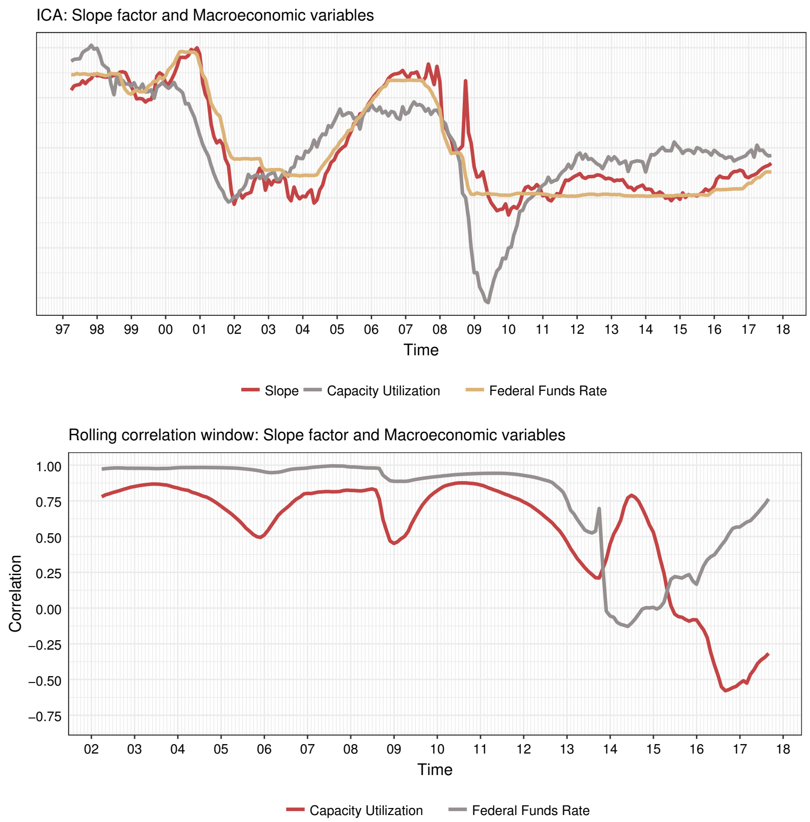

Figure 7-4: Top: Estimated slope factor and Capacity Utilization and Federal Funds Rate. Bottom: Rolling window of five-year period correlations between the estimated slope factor and Capacity Utilization and Federal Funds Rate.

In Figure 7-5, there is no strong correlations between the curvature factor and macroeconomic variables over the whole period (1997-2017). Therefore, I construct a rolling window of five-year period correlations to investigate further whether there are periods of high correlation. In fact, Unemployment Rate, Consumer Sentiment Index, and Trade Weighted U.S. Dollar Index fluctuate around the zero mark until recent years with correlation around 0.80 with Unemployment Rate negatively correlated.

I fit a VAR(1) model with estimated ICA factors and a subset of macroeconomic variables. I do not include all variable because I have 246 monthly observations and to use all macro variables leads to few degrees of freedom left since many coefficients have to be estimated. However, it is possible to regularize the VAR(1) model to reduce the parameter space (e.g., see Nicholson et al.,

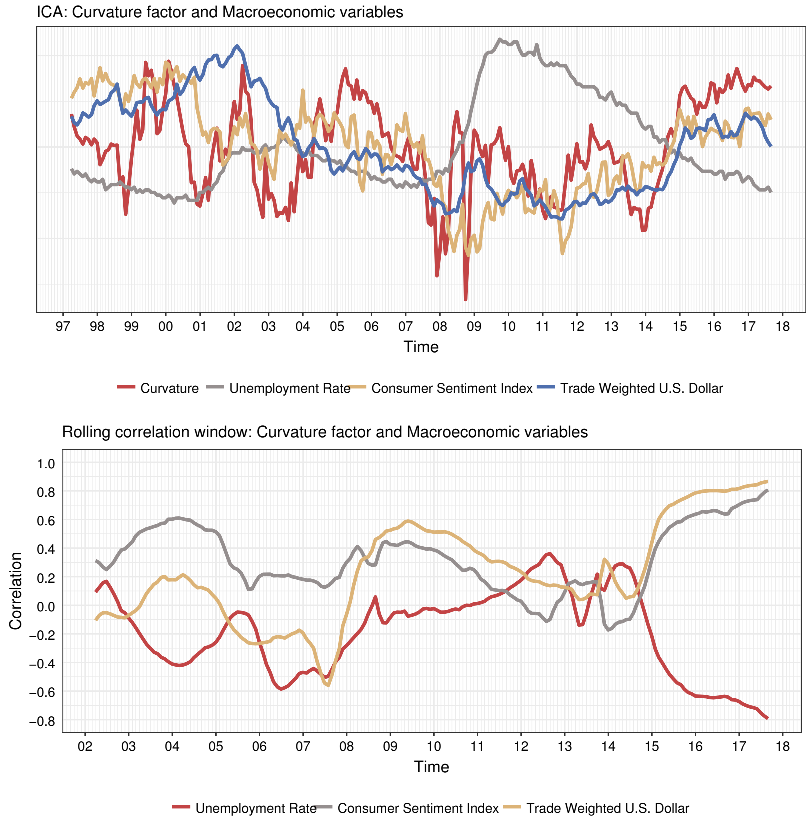

Figure 7-5: Top: Estimated curvature factor and Unemployment Rate, Consumer Sentiment Index, and Trade Weighted U.S. Dollar Index. Bottom: Rolling window of five-year period correlations between the estimated curvature factor and Unemployment Rate, Consumer Sentiment Index, and Trade Weighted U.S. Dollar Index.

2015). Although the regularization approach may be useful, I select variables based on the rolling correlations and previous literature because of the time constraint. Therefore, the variables I choose to include in the VAR(1) model are Capacity Utilization, Federal Funds Rate, CPI, Unemployment Rate, Customer Sentiment Index, and Money Stock.

In Table 7.3, I present parameter estimates for a VAR(1) model with estimated ICA factors and the macro variables. There is a high persistence in each of the factors and variables. Notably, there are two coefficient estimates above 1, i.e., Capacity Utilization and Money Stock with values equal to 1.01. However, they have standard errors equal to 0.02 and 0.01 respectively. A remarkable fact is that the largest significant off-diagonal coefficient is between the slope and curvature factor.

Based on the VAR(1) model, I examine the dynamics of the estimated factors and the macroeconomic variables via impulse response functions, which I show (along with 90% confidence intervals) in Figure 7-6 and Figure 7-7. I consider three groups of impulse responses: macro responses to factor shocks, factor responses to macro shocks, and factor responses to factor shocks. In general, the responses for level, slope, and curvature for shocks to Federal Funds Rate and CPI (inflation proxy) are in line with previous literature (e.g., Diebold et al., 2006). However, the shocks to Capacity Utilization is different and this may be due to that they study a different period that does not deal with the same issues as the period I am studying, i.e., with unconventional monetary policy due to the zero lower bound. For example, if inflation is too low or unemployment too high, the central bank may push down short-term rates to increase the incentives to spend. However, too large negative rates are not feasible since people can always hold cash. Therefore, unconventional monetary policy, such as quantitative easing, is a tool that the Fed is using to stimulate the economy. Also, in 2006 all of this was theoretical and zero-lower bound was hardly an issue in the literature. Consequently, it is likely that this tool may affect the results when comparing to literature that looks at other periods.

First, I analyze and discuss the impulses to the six macro vars and the latent factor responses. (i) An increase in Capacity Utilization immediately pushes down the level of the yield curve, whereas it pushes up the slope so that the yield curve is less positively sloped (or more negatively sloped). The latter is consistent with previous literature (Diebold et al., 2006). However, the former is not, and it is perhaps due to the extra data points as discussed above. Also, the curvature seems to have a negligible response. (ii) An increase in the Federal Funds Rate appears to increase the level of the yield curve, although it is not significant since the zero point is within the confidence bounds. The same goes for the slope factor. An increase in Federal Funds Rate depresses the curvature factor drastically so that the yield curve is less curved. This result is remarkable, since Diebold et al. (2006) did not find any significant changes and my result may be due to that half the sample has higher short and long-term rate than the mid-term rate (a so-called positive butterfly). Therefore, when short and long end increase more (or decrease less) than the mid-term rates, the convexity of the yield curve increases. (iii) An increase in the CPI, which can be considered to be a proxy for inflation, pushes up the level of the yield curve consistent with long-term inflation expectations not being firmly anchored, and raises the level factor. The slope and curvature factors seem to have negligible responses to a CPI increase. (iv) An increase in the Unemployment Rate depresses the curvature so that the yield curve is less curved. It also seems that the level decrease in the long-run, and that the slope factor is negligible. (v) An increase in Consumer Sentiment Index increase the level marginally in the short-run and also increases the slope marginally with the curvature factor being negligible. (vi) An increase in Money Stock tends to increase the level and curvature. Also, the effect on slope seems to be insignificant.

The impulses to the three latent factors and the macro vars responses show that shocks to the level and curvature factor seem to have no response in the macro variables. However, an increase in the slope factor depresses Capacity Utilization immediately. The same goes for CPI, although it seems to be only in the short run and later on appears to increase CPI. The inverse relationship is for Money Stock, where it first increase in the short term and after that decrease rapidly. An increase in the slope factor pushes the Federal Funds Rate, and the same goes for Unemployment

Rate and Consumer Sentiment Index.

Finally, most off-diagonal responses of the level, slope, and curvature own-dynamics are insignificant except the shock to slope factor which pushes up the curvature factor drastically making the yield curve more curved in the short term and making the yield curve less curved in the long term.

Table 7.3: Parameter estimates for VAR(1) model with estimated ICA factors and a subset of macroeconomic variables. The macro variables are Capacity Utilization (CU), Unemployment Rate (UR), Consumer Sentiment Index (CSI), Money Stock (MS), Federal Funds Rate (FFR), and CPI. Standard errors appear in parentheses.

	$f_{1,t-1}$	$f_{2,t-1}$	$f_{3,t-1}$	$\text{CU}_{t\text{-}1}$	$\text{UR}_{t\text{-}1}$	$\text{CSI}_{t\text{-}1}$	$\text{MS}_{t\text{-}1}$	$\text{FFR}_{t\text{-}1}$	$\text{CPI}_{t\text{-}1}$
$f_{1,t}$	0.90^{***}	0.01	-0.14^{*}	0.06^{***}	-0.02^{**}	0.05	-0.04^{***}	-0.01	0.08^{**}
	(0.04)	(0.01)	(0.06)	(0.01)	(0.01)	(0.03)	(0.01)	(0.01)	(0.02)
$f_{2,t}$	0.05	0.88^{***}	0.72^{**}	-0.26^{***}	0.13^{***}	0.01	0.16^{**}	0.05	-0.44^{***}
	(0.14)	(0.05)	(0.23)	(0.05)	(0.04)	(0.12)	(0.05)	(0.02)	(0.09)
$f_{3,t}$	-0.10^{***}	0.03^{***}	0.73^{***}	0.03^{**}	-0.01^{*}	0.05^{*}	-0.04^{***}	0.04^{***}	0.06^{**}
	(0.03)	(0.01)	(0.04)	(0.01)	(0.01)	(0.02)	(0.01)	(0.00)	(0.02)
CU_t	-0.17^{***}	0.01	0.06	1.01^{***}	-0.06^{***}	0.18^{***}	-0.01	0.05^{***}	0.04
	(0.04)	(0.01)	(0.07)	(0.02)	(0.01)	(0.04)	(0.01)	(0.01)	(0.03)
UR_t	-0.03	0.00	-0.07	0.06^{***}	0.95^{***}	0.08^{*}	-0.03^{*}	0.04^{***}	0.02
	(0.04)	(0.01)	(0.07)	(0.01)	(0.01)	(0.04)	(0.01)	(0.01)	(0.03)
CSI_t	0.12	0.04	-0.05	0.05^{*}	-0.02	0.63^{***}	0.03	0.00	-0.07
	(0.06)	(0.02)	(0.10)	(0.02)	(0.01)	(0.05)	(0.02)	(0.01)	(0.04)
MS_t	0.01	-0.02	0.00	-0.03	0.01	0.16^{***}	1.01^{***}	-0.03^{***}	-0.06^{*}
	(0.04)	(0.01)	(0.07)	(0.02)	(0.01)	(0.04)	(0.01)	(0.01)	(0.03)
FFR_t	-0.04	0.09	-0.84^{**}	0.25^{***}	-0.11^{**}	0.13	-0.21^{***}	0.94^{***}	0.51^{***}
	(0.15)	(0.05)	(0.25)	(0.06)	(0.04)	(0.14)	(0.05)	(0.03)	(0.10)
CPI_t	0.03	0.00	-0.16^{*}	-0.02	0.01	-0.10^{**}	0.03^{*}	-0.04^{***}	0.91^{***}
	(0.04)	(0.01)	(0.07)	(0.02)	(0.01)	(0.04)	(0.01)	(0.01)	(0.03)

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

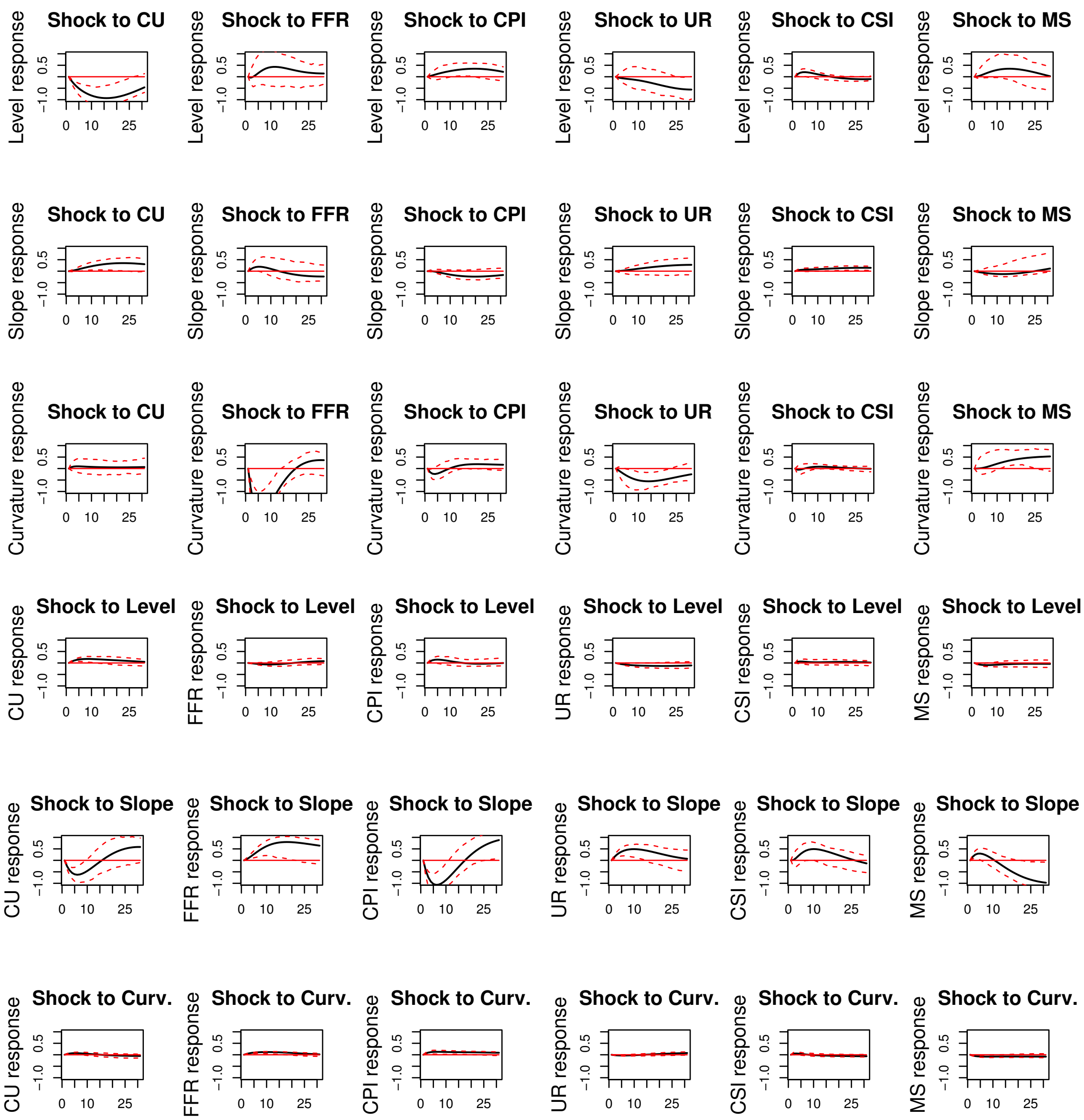

Figure 7-6: Impulse response functions for the estimated factors (level, slope, and curvature) and a subset of macroeconomic variables based on the VAR(1) model in Table 7.3. The macro variables are Capacity Utilization (CU), Federal Funds Rate (FFR), CPI, Unemployment Rate (UR), Consumer Sentiment Index (CSI), and Money Stock (MS).

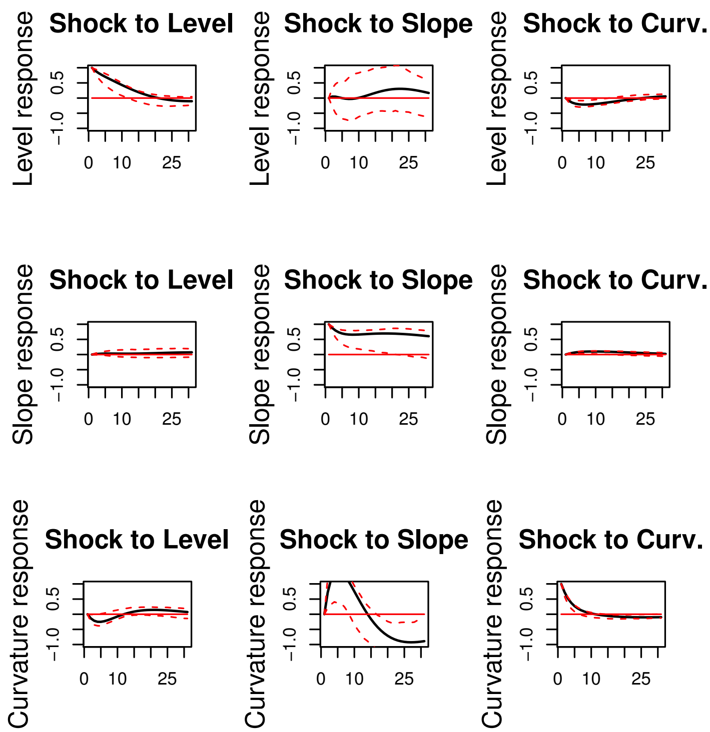

Figure 7-7: Impulse response functions for the estimated factors (level, slope, and curvature) based on the VAR(1) model in Table 7.3.

Section 8

Conclusion

I find that the loadings from ICA give an interpretation of ICA factors as level, slope, and curvature. They also have similar shapes as the loadings in PCA and DNSM. The estimated ICA factors have highly persistent dynamics and correlate with estimates from PCA and DNSM (especially with PCA).

All three-factor models fit the yield curve well and produce good forecasts. ICAYM and PCAYM forecast short-term yields with a long forecast horizon significantly better than DNSM, across most forecast horizons. Concerning the yield curves longer maturities, there are no statistically significant differences between ICAYM and DNSM. However, DNSM is significantly better to forecast future yields at a short forecast horizon. Now, since ICAYM has a good forecast performance for the short-term maturities and multiple horizons, one can investigate if it is possible to improve long-term forecasts to estimate the bond risk premia and apply it to forecasted short-term yields to obtain mid to long-term yields. To estimate the bond risk premia, one may use NY FED ACM model (Adrian et al., 2013, 2014) or estimate how, e.g., the 7y forward 3m rate differ from the 10y forward 3m rate. This whole approach relies on the assumption that people do not have strong expectations about interest rates that far out and that the bond risk premium is a significant yield curve influencer.

As pointed out in section 4, I assume that the factors distribution are a reciprocal cosh density. However, whether this is the best density remains to be answered. Furthermore, extensions can be considered, such as estimating ICA factor loadings jointly in a state-space model (similar to DNSM). However, this is likely to be computationally cumbersome, whereas this paper's proposed ICA-based model is fast concerning computational time. Another possible extension may be time-varying loadings, or to use a rolling window approach (e.g., Shah, 2013). An application related to the rolling window approach can be to study whether there are evident dynamic patterns regarding a coupling metric (Shah and Roberts, 2013), i.e., a dependency measure, and can be used to study how factors interact across different maturities in the yield curve for trading strategies, risk management purposes, or as an economic indicator. Also, all these approaches can be combined with wavelets to study time-scale dependencies (Shah and Roberts, 2008).

Traders may be interested in a high-frequency setting, where it may be easier to obtain independent factors because there are more observations than in the monthly case. However, one has to be cautious in selecting the period of interest because of the changes in the economic environment.

The literature for term structure models is vast, where a significant part belongs to the affine no-arbitrage models. Although ICAYM is not arbitrage-free, it gives interpretable results, good fit, and produces good forecasts. One future extension is to investigate feasibility to impose absence of arbitrage in ICAYM, similar as in the DNSM extension proposed by Christensen et al. (2011), where they add a yield-adjustment term.

The estimated factors have periods of high correlation with some of the macroeconomic variables. In particular, the correlation between the level factor and Real Income, Real Consumption Expenditures, and CPI seem to be strongly negative around financial crisis in 2008, and in recent years (except CPI with opposite sign). The slope factor appears to have a high correlation with Capacity Utilization and Federal Funds Rate over the whole period. Finally, the correlation between the curvature factor and Unemployment Rate, Consumer Sentiment Index, and Trade Weighted U.S. Dollar Index, seems to be high in recent years.

An increase in the Fed Funds Rate appears to increase the level and slope of the yield curve, whereas it depresses the curvature factor drastically so that the yield curve is less curved. An increase in the CPI pushes up the level of the yield curve, whereas the slope and curvature factors seem to have negligible responses. An increase in the Unemployment Rate decrease the level in the long-run and also depresses the curvature so that the yield curve is less curved.

In conclusion, today the most popular yield curve models are either based on parametric loading shapes or loadings for uncorrelated factors that explain the maximal variance in the yields. Here, I propose a dynamic ICA-based yield curve model that use independent factors. It fits the yield curve well and produces better forecasts for the short-term maturities and a long forecast horizon than the commonly used Diebold and Li (2006) model, with similar results for the long-term maturities. I show that the estimated factors from the proposed model relate to macroeconomic variables. Finally, I have demonstrated that ICA has a place in yield curve modeling, and many future extensions can further increase our understanding of the characteristics and dynamics of the yield curve, leading to more accurate asset pricing, portfolio allocation, risk management, and monetary policy.

Bibliography

Abbritti, M., Dell'Erba, S., Moreno, A. and Sola, S. (2018), 'Global factors in the term structure of interest rates', *International Journal of Central Banking* **14**(2), 301–340.

Adrian, T., Crump, R. K. and Moench, E. (2013), 'Pricing the term structure with linear regressions', *Journal of Financial Economics* **110**(1), 110–138.

Adrian, T., Crump, R., Mills, B. and Moench, E. (2014), 'Treasury term premia: 1961-present', *Liberty Street Economics* .

Ang, A. and Piazzesi, M. (2003), 'A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables', *Journal of Monetary Economics* **50**(4), 745–787.

Back, A. D. and Weigend, A. S. (1997), 'A first application of independent component analysis to extracting structure from stock returns', *International Journal of Neural Systems* **8**(04), 473–484.

Baygün, B., Showers, J. and Cherpelis, G. (2000), 'Principles of principal components', *U.S. Fixed-Income Research Portfolio Strategies, Salmon Smith Barney* .

Bell, A. J. and Sejnowski, T. J. (1995), 'An information-maximization approach to blind separation and blind deconvolution', *Neural Computation* **7**(6), 1129–1159.

BIS (2005), 'Zero-coupon yield curves: Technical documentation', *BIS Working Papers* **25**.

Björk, T. and Christensen, B. J. (1999), 'Interest rate dynamics and consistent forward rate curves', *Mathematical Finance* **9**(4), 323–348.

Campbell, J. Y., Sunderam, A. and Viceira, L. M. (2017), 'Inflation bets or deflation hedges? the changing risks of nominal bonds', *Critical Finance Review* **6**(2), 263–301.

Cardoso, J.-F. (1997), 'Infomax and maximum likelihood for blind source separation', *IEEE Signal Processing Letters* **4**(4), 112–114.

Cha, S.-M. and Chan, L.-W. (2000), 'Applying independent component analysis to factor model in finance', *Intelligent Data Engineering and Automated Learning–IDEAL 2000. Data Mining, Financial Engineering, and Intelligent Agents* pp. 161–173.

Chen, L. (1996), 'A bond pricing formula under a non-trivial, three-factor model of interest rates', *Economics Letters* **51**(1), 95–99.

Christensen, J. H., Diebold, F. X. and Rudebusch, G. D. (2011), 'The affine arbitrage-free class of nelson–siegel term structure models', *Journal of Econometrics* **164**(1), 4–20.

Christensen, J. H. and Rudebusch, G. D. (2016), Modeling yields at the zero lower bound: Are shadow rates the solution?, *in* 'Dynamic Factor Models', Emerald Group Publishing Limited, pp. 75–125.

Cochrane, J. H. and Piazzesi, M. (2005), 'Bond risk premia', *American Economic Review* **95**(1), 138–160.

Coroneo, L., Nyholm, K. and Vidova-Koleva, R. (2011), 'How arbitrage-free is the nelson–siegel model?', *Journal of Empirical Finance* **18**(3), 393–407.

Cox, J. C., Ingersoll, J. E. and Ross, S. A. (1985), 'A theory of the term structure of interest rates', *Econometrica* **53**, 385–407.

Culbertson, J. M. (1957), 'The term structure of interest rates', *The Quarterly Journal of Economics* **71**(4), 485–517.

Dai, Q. and Singleton, K. J. (2000), 'Specification analysis of affine term structure models', *The Journal of Finance* **55**(5), 1943–1978.

Diebold, F. X. and Li, C. (2006), 'Forecasting the term structure of government bond yields', *Journal of Econometrics* **130**(2), 337–364.

Diebold, F. X. and Mariano, R. S. (1995), 'Comparing predictive sccuracy', *Journal of Business & Economic Statistics* **13**, 253–265.

Diebold, F. X., Rudebusch, G. D. and Aruoba, S. B. (2006), 'The macroeconomy and the yield curve: a dynamic latent factor approach', *Journal of Econometrics* **131**(1-2), 309–338.

Duffie, D. and Kan, R. (1996), 'A yield-factor model of interest rates', *Mathematical Finance* **6**(4), 379–406.

Durbin, J. and Koopman, S. J. (2012), *Time Series Analysis by State Space Methods: Second Edition Time Series Analysis by State Space Methods*, second edition edn, Oxford University Press.

ECB (2008), 'The new euro area yield curves', *Monthly Bullentin February* pp. 95–103.

Fabozzi, F. J., Giacometti, R. and Tsuchida, N. (2016), 'Factor decomposition of the eurozone sovereign cds spreads', *Journal of International Money and Finance* **65**, 1–23.

Fama, E. F. and Bliss, R. R. (1987), ‘The information in long-maturity forward rates’, *The American Economic Review* pp. 680–692.

Filipović, D. (1999), ‘A note on the nelson-siegel family’, *Mathematical Finance* **9**(4), 349–359.

Gürkaynak, R. S., Sack, B. and Wright, J. H. (2007), ‘The us treasury yield curve: 1961 to the present’, *Journal of Monetary Economics* **54**(8), 2291–2304.

Harvey, D., Leybourne, S. and Newbold, P. (1997), ‘Testing the equality of prediction mean squared errors’, *International Journal of Forecasting* **13**(2), 281–291.

Ho, T. S. and Lee, S.-B. (1986), ‘Term structure movements and pricing interest rate contingent claims’, *The Journal of Finance* **41**(5), 1011–1029.

Hull, J. and White, A. (1990), ‘Pricing interest-rate-derivative securities’, *The Review of Financial Studies* **3**(4), 573–592.

Hyvärinen, A., Karhunen, J. and Oja, E. (2001), *Independent Component Analysis*, Wiley-Interscience.

Ilmanen, A. (1995), ‘Understanding the yield curve: Part 1 - overview of forward rate analysis’, *U.S. Fixed-Income Research Portfolio Strategies, Salomon Brothers, New York* .

Joslin, S., Priebsch, M. and Singleton, K. J. (2014), ‘Risk premiums in dynamic term structure models with unspanned macro risks’, *The Journal of Finance* **69**(3), 1197–1233.

Keynes, J. M. (1936), *The General Theory of Employment, Interest and Money*, Palgrave Macmillan.

Koopman, S., Lucas, A. and Zamojski, M. (2017), Dynamic term structure models with score-driven time-varying parameters: estimation and forecasting, NBP Working Papers 258, Narodowy Bank Polski, Economic Research Department.

Kumiega, A., Neururer, T. and Vliet, B. V. (2011), ‘Independent component analysis for realized volatility: Analysis of the stock market crash of 2008’, *The Quarterly Review of Economics and Finance* **51**(3), 292–302.

Lee, S., Shen, H., Truong, Y., Lewis, M. and Huang, X. (2011), ‘Independent component analysis involving autocorrelated sources with an application to functional magnetic resonance imaging’, *Journal of the American Statistical Association* **106**(495), 1009–1024.

Lekkos, I. (2000), ‘A critique of factor analysis of interest rates’, *The Journal of Derivatives* **8**(1), 72–83.

Litterman, R. and Scheinkman, J. (1991), ‘Common factors affecting bond returns’, *Journal of Fixed Income* **1**(1), 54–61.

Longstaff, F. A. and Schwartz, E. S. (1992), 'Interest rate volatility and the term structure: A two-factor general equilibrium model', *The Journal of Finance* **47**(4), 1259–1282.

Lord, R. and Pelsser, A. (2007), 'Level–slope–curvature–fact or artefact?', *Applied Mathematical Finance* **14**(2), 105–130.

Matilainen, M., Miettinen, J., Nordhausen, K., Oja, H. and Taskinen, S. (2017), 'On independent component analysis with stochastic volatility models', *Austrian Journal of Statistics* **46**, 57–66.

McCracken, M. W. and Ng, S. (2016), 'Fred-md: A monthly database for macroeconomic research', *Journal of Business & Economic Statistics* **34**(4), 574–589.

McKeown, M. J., Hansen, L. K. and Sejnowski, T. J. (2003), 'Independent component analysis of functional mri: what is signal and what is noise?', *Current Opinion in Neurobiology* **13**(5), 620–629.

Mesters, G., Schwaab, B. and Koopman, S. J. (2014), A dynamic yield curve model with stochastic volatility and non-gaussian interactions: an empirical study of non-standard monetary policy in the euro area, Technical report, Tinbergen Institute Discussion Paper.

Modigliani, F. and Sutch, R. (1966), 'Innovations in interest rate policy', *The American Economic Review* **56**(1/2), 178–197.

Mukamel, E. A., Nimmerjahn, A. and Schnitzer, M. J. (2009), 'Automated analysis of cellular signals from large-scale calcium imaging data', *Neuron* **63**(6), 747–760.

Nelson, C. R. and Siegel, A. F. (1987), 'Parsimonious modeling of yield curves', *Journal of Business* **60**, 473–489.

Nicholson, W., Matteson, D. and Bien, J. (2015), 'Varx-l: Structured regularization for large vector autoregressions with exogenous variables', *ArXiv e-prints* .

Östlund, N., Yu, J. and Karlsson, J. S. (2006), 'Adaptive spatio-temporal filtering of multichannel surface emg signals', *Medical and Biological Engineering and Computing* **44**(3), 209–215.

Rudebusch, G. D. and Wu, T. (2008), 'A macro-finance model of the term structure, monetary policy and the economy', *The Economic Journal* **118**(530), 906–926.

Shah, N. (2013), Statistical Dynamical Models of Multivariate Financial Time Series, PhD thesis, University of Oxford.

Shah, N. and Roberts, S. (2008), Hidden markov independent component analysis as a measure of coupling in multivariate financial time series, *in* 'Proceedings of the ICA Research Network International Workshop'.

Shah, N. and Roberts, S. J. (2013), 'Dynamically measuring statistical dependencies in multivariate financial time series using independent component analysis', *ISRN Signal Processing* .

Song, L. and Lu, H. (2016), Ecoica: Skewness-based ica via eigenvectors of cumulant operator, *in* 'Proceedings of The 8th Asian Conference on Machine Learning', Sheffield, pp. 445–460.

Steeley, J. M. (1990), 'Modelling the dynamics of the term structure of interest rates', *The Economic and Social Review* **21**(4), 337–361.

Svensson, L. E. (1994), Estimating and interpreting forward interest rates: Sweden 1992-1994, Technical report, National Bureau of Economic Research.

Vasicek, O. (1977), 'An equilibrium characterization of the term structure', *Journal of Financial Economics* **5**(2), 177–188.

Wu, T. (2003), 'What makes the yield curve move?', *FRBSF Economic Letter* (6).