LUISS

Dipartimento
di Impresa e Management

Cattedra di Asset Pricing

A Joint Macroeconomic and Term Structure Model: a FAVAR approach to the Ang and Piazzesi Model

Prof. Paolo Porchia

RELATORE

Prof. Marco Pirra

CORRELATORE

Lorenzo Amati - Matr. 693851

CANDIDATO

Anno Accademico 2018/2019

TABLE OF CONTENTS

Introduction

This thesis analyses the different approaches used to describe the term structure and yield curve dynamics, taking into consideration their advantages and disadvantages. The interest in this kind of models is somehow changeable, as it increases during crisis and decreases during prosperous times, with few exceptions. The starting point of the discipline find its foundations in the model proposed by Vasicek in 1977, which looks at the similarity and the differences between the yields' and the stocks' movements. The fundamental intuition on the model's dynamics led to a vast number of articles and models, using different points of view to solve the difficult forecast of the yields.

Unfortunately, multiple issues arise, as the governments set the interest rates following different purposes. As a rule, Central Banks (CBs) generally tend to maintain price stability as their main goal, but they also tend to adjust it in favour of more specific issues or targets they want to reach. Employment is, for example, one of the main macroeconomic indicators that a CB wants to keep steady. Moreover, during a crisis the CB would follow a strategy in order to bring back the economy to their previous status, as the leading indicators tend to become unpredictable. A good example to start with is the inflation that occurred in the United States between 1970s and the first years of 1980, subdued by the Federal Reserve approach. Setting higher yields helped to restore the situation in few years, averting the risk of an excessive inflation, thus confirming their usefulness. Nonetheless, investors do not know the CBs intentions and usually tends to guess, basing their intuitions on different factors. This is an issue for those who are interested, for example, in building a portfolio that includes bonds of different maturities to hedge the risk.

At this point, term structure models become useful, as they use observable inputs in order to find plausible results for the expected yields. The models presented in this thesis help both investors and Central Banks to provide good forecasts using different methodologies and presenting both advantages and disadvantages,

The core of this work uses a specific model in order to provide a more in-depth analysis, the Ang and Piazzesi (2003). Their approach is different from previous works on the topic, incorporating observables macroeconomic extracted factors in combination with latent ones. Nonetheless the approach needs a two-step estimation which requires multiple constrictions. One way to avoid this issue is to use a similar two-step procedure as the one proposed by Mönch (2005) and use a modified version of the FAVAR approach defined by Bernanke et al. (2005). The thesis provides an analysis on the advantages of the model and the forecasts for the 6, 12 and 36 months implied by the model.

Chapter 1 – Bonds: Components and Models

1.1 Bonds definition and Bond Yield Curve

Bonds are amongst the most valuable instruments used by public and private institutions to obtain liquidity from third party investors. By buying a bond, the holder agrees to lend a pre-determined amount of money to the issuer in exchange of interests that can be paid at multiple dates until maturity, or in a single solution together with the lent capital. As this thesis will move forward into an affine term structure model with macroeconomic variables, it is adequate to start from the basics, briefly defining first what a bond is, what is the yield curve and how these can relate to the thesis' purposes.

A first distinction must be made, since the bonds hereby analysed are only US discount bonds and are assumed to be default-free. Zero-coupon bonds are extremely useful in this context and for such reason in future paragraphs the latter will be simply referred to as bonds unless it is explicitly said otherwise. The reason for their use comes from their straightforwardness as they can be treated in a more mathematically convenient way than coupon bonds. In a more simplistic way, a zero-coupon bond returns a notional made up by the price of the bond and its interests at maturity. To be more thorough and see this from a more formal point of view, a zero-coupon bond is a financial security instrument sold at a discount that pays a unit of cash at a predefined maturity – using a variable identified as T - without paying any interests during its life. By looking at it in this way, there is a single coupon paid at maturity which is a fixed amount equal to the notional 1 minus the price P_t^T. From this, by defining y_t^T as the yield requested for a zero-coupon bond to pay its unit price, the following formula is obtained:

$$P_t^T = e^{-y_t^T(T-t)} \tag{1}$$

Which can also be written as:

$$y_t^T = -\frac{1}{T-t}\log(P_t^T) \tag{2}$$

These formulas are usually not priced in real-world conditions, but in a constructed environment defined as risk-neutral world. In this space, it is possible to assert that the price of a security today is equal to its expected future price discounted at the risk-free rate. The risk-neutral world is usually defined with the letter $\mathbb{Q}$, while real-world conditions with $\mathbb{P}$[1]. Considering a stochastic process *r*, the

[1] To go from Real world to neutral risk it is usually used the Radon-Nikodym derivative. It is widely used in Affine Term structure models, for example in Ang and Piazzesi (2003).

price of a discount bond of maturity T is equal to its conditionally expected payoff discounted for the interest rates between t and T. So, the formula to express the price in continuous time is the following:

$$P_t^T = E_t\left[e^{-\int_t^T r_s ds}\right] \tag{3}$$

All the calculations that follow will be then based on neutral world measures. When this is true, the prices that derive are arbitrage-free[2].

Duration is the last measure introduced in this paragraph. It is a measure of the sensitivity of the price of a given in bond in relation to a change in the interest rate. It can be defined as the derivative of the price respect to the yield, divided by minus the price. So that:

$$Dur_t^T = -\frac{1}{P_t^T}\frac{dP_t^T}{dy_t^T} \tag{4}$$

Calculating it would give out a formula for the duration equal to $T - t$. This result comes in handy for the definition of convexity in paragraph (1.2.3).

From a macroeconomic point of view, yields are extremely important. Central Banks (CB) usually have at disposal different ways in order to follow their objectives, from reserves' requirements amounts to the quantity of circulating money. Usually they find their main role in fulfilling price stability and the control of the yields proved to be the most valuable tool they can use. Setting them usually follows the gathering of a massive amount of macroeconomic data in order to present clear analysis and model them in order to forecast the impact of a yields' change in the real economy. As keeping prices' level stable is not an easy task, usually CBs – like the FED – tends to adjust yields constantly. For example, in the United States, the Federal Reserve Board's Open Market Committee takes these decisions in order to meet the FED's targets, which can be not limited to inflation alone[3]. In fact, when signs of a possible slowdown start to appear, the FED could decide to lower the yield to incentivize borrowing and lending in the market. The work of CBs is then to stimulate the economy, keeping the price level growing slowly and smoothly at the same time.

Having models that predict how interest rates should be set according to macroeconomic inputs appears to be fundamental then.

[2] A good definition comes also from Duffie (2001) who states that: "Working in $\mathbb{Q}$ is equivalent to markets being complete".
[3] Piazzesi (2001) analyses the change in the yields following the BOMC.

1.1.2 Forward Rates

The concept of forward rate comes from the spot rate defined in equation (1). In continuous time, it is the instantaneous rate that is applicable by entering in a financial contract - going long or short - with a transaction that takes place in the future. Equation (5) provides the forward rate f_t^T for a maturity T such that:

$$F_t^{T+\tau} = \frac{1}{\tau}\left(\frac{P_t^T}{P_t^{T+\tau}} - 1\right) \tag{5}$$

With $\tau = (T - t)$

It is worth to notice that the equation does not present any expectations and as the maturity T approaches t, the forward rate approaches the spot rate (6).

$$r_t = \lim_{T \to t} f_t^T \tag{6}$$

1.1.3 The Yield Curve

Having briefly defined bonds' prices and forward rates, it is possible to start with a more complete framework for the following analysis. To begin with, it is important to also delineate what a term structure is made of. Namely, it represents the relationship that links the different interest rates for the same instrument – with the same credit quality - at different maturities. The graphic representation of the term structure is defined as "yield curve" and it is quite useful as it comes to provide an understanding of the bonds' market behaviour. Hereby is reported an example of a curve with maturities between one and fifteen years.

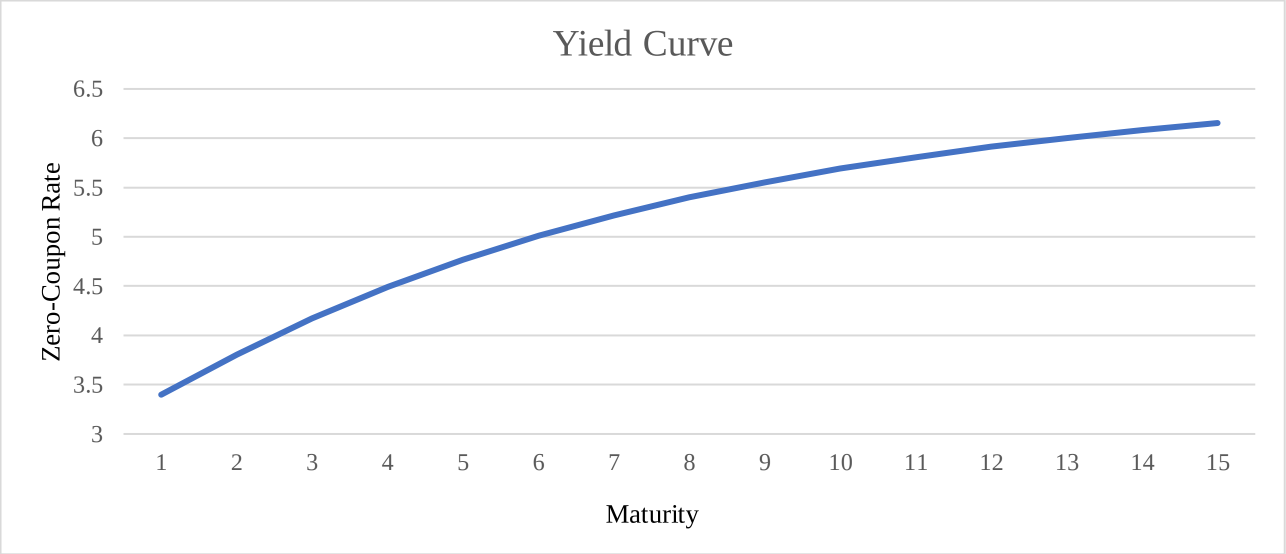


Figure 1. A zero-coupon series of rates observed in a single day

The yield curve provides a snapshot at a certain time *t* of how the rates behave considering different maturities. It can take different shapes, but the most common is the one presented above, with the curve upward sliding. Having it in this way it is considered to be the normal scenario, in which investors get higher compensations for longer maturities. Usually, before a recession, it has been noticed that the longer yields become closer to the shorter ones, resulting in a flattening of the curve, defined precisely as a flat curve. During a recession or when one is approaching, the longer yields tend to get lower than the shorter ones, resulting in an inverted curve.

The question regarding why the term structure is so important appears to be crucial at this point. Mainly, it is extremely useful to those concerned with investment decisions and in the assessment of the policy adopted by the Central Bank.

Banks and other financial institutions are primarily influenced by it as the treasury rates are a perfect benchmark on which it is possible to set up lending and saving rates. This is also linked to interest-rate-contingent claims, as it plays a crucial role in the determination of prices for multiple financial securities, like caps, floors, swaptions and others. As in the following pages, this work will be based on a joint macroeconomic and term structure model, it is clear that this is even more significant.

1.2 The drivers of the Yield Curve

Mainly there are three forces that move the yield curve: expectations, risk premia and convexity. The three influence the curve with different intensities: expectations drive it in the short term, risk premia in the medium and convexity in the long end.

1.2.1 Expectations

Starting from the beginning, it is better to introduce what the expectation hypothesis (EH) states. At the same time, it is worthy to introduce one of the most important variants for the purposes of this thesis, the local expectations hypothesis or LEH[4]. To quickly define it, the expectations hypothesis states that a long-term bond's yield with a maturity T equals the one of a series of shorter period bonds such that the last short bond ends in T. In the formula, it is equivalent to state that:

$$y_t^T = \left(1 + y_t^{(1)}\right)\left(1 + y_{t+1}^{(1)}\right) \dots \left(1 + y_T^{(1)}\right) \tag{7}$$

Another way to put it is to consider the value of a yield Y_t^T as the expectation in time *t* of average future yields.

[4] Cox, Ingersoll, Jr., and Ross (1981) consider the application of both hypothesis in their paper.

$$EH: y_t^T = \frac{E_t[S]}{T-t} \tag{8}$$

Where S $=\left[\int_t^T r_s ds\right]$.

If the expectations hypothesis is correct, it is then possible to use the slope of the term structure to forecast the future path of the interest rate. While the EH states that the interest rate at a time t is equal to the expected values of future short rates, the LEH focuses on stating the same for a short-term horizon. The expectation hypothesis is often used, but empirical shreds of evidence provide little support on its behalf and in some works the LEH is considered to be more useful.

$$LEH:\ y_t^T = -\log\frac{E_t[e^{-S}]}{(T-t)} \tag{9}$$

which becomes:

$$y_t^T = \frac{E_t[S]}{(T-t)} - \frac{1}{2} var_t\left(\frac{[S]}{(T-t)}\right) \tag{10}$$

due to Jensen inequality and thus an added variance term.

1.2.2 Risk Premia

Risk Premia are the second driver of the yield curve. A risk premium is a form of compensation to investors that bear uncertainty by investing in financial security for a pre-determined period of time *(T-t)*. There are different risks linked to possessing a bond. The first risk comes from inflation changes. An intuitive example of inflation risk involves a bond emitted by a country A that pays a fixed amount of money in nominal terms. Between the time from its emission and maturity, assume there is a sudden increase in the inflation of that country which is not accounted in the payoff guaranteed by the bond. This would simply result in a loss of purchase power for the acquirer. The difference between the nominal rate and the inflation rate is usually defined as real risk-free rate and Fisher equation describes it in terms of expected inflation and nominal rate such that:

$$1 + i = (1 + r)(1 + \pi_e)$$

Risk premia cannot be set by Central Banks or other financial institutions, but are indeed influenced by the communications, the intentions and above-all their credibility regarding price control. From Piazzesi and Cochrane (2008), there are different ways to formally define risk premium and these definitions can be interlaced to the expectation hypothesis. The two linked to the previously seen equations ~~states~~ state that[5]:

- The long-term yield can be considered as the average of all the expected future short term rates, as seen in equation (7), but with an added term: the risk premium.
- Starting from equation (5), a forward rate can also be seen as the conditional expected value of the future short rate with – again – an added risk premium.

1.2.3 Convexity

Convexity is the last driver of the curve and dominates the long end of the yield curve. It is considered as a measure of the curvature, interlacing bonds' duration and rate: as duration increases with yields, it is said that the convexity is negative. From a mathematical point of view, it is the second derivative of the price respect to the yield such that:

$$Conv_t^T = -\frac{1}{P_t^T}\frac{\partial^2 P_t^T}{\partial {y_t^T}^2}$$

Its effect on the curve is more difficult to observe compared to the other two drivers, as its influence emerges only in the last part of the curve, for bonds with long term maturities. This is caused by the greater effects that expectations and risk premia have on the curve. However, this measure still has relevance for modelling, especially for zero-coupon bonds.

1.3 The no-arbitrage condition

The no-arbitrage condition is usually specified in some of the presented models, like Vasicek, through theoretical foundation and appears in numerous others. The assumption is important, as having a no-arbitrage condition suggests that the stochastic process above-mentioned is strictly positive. It is worth to mention though, that some academics have raised some concerns about the no arbitrage condition. Empirical evidences have led to infer that imposing a no-arbitrage condition could not really be that significant since in real-world condition an arbitrage rarely occurs[6] and imposing it could not really improve the model and it affects its outputs little. Nonetheless, multiple empirical studies have concluded that it improves the forecast and the explanatory of the model and thus giving a good reason to keep imposing it.

[5] The two statements can be considered as equivalent.

[6] Usually it is possible to observe only pseudo arbitrage strategies, as there are risky arbitrage and not riskless as the definition for pure arbitrage requires.

1.4 Modeling the yield curve

The problem that arises from having a snapshot of the yield curve is that it is not a scalar value but instead a vector quantity. In other words, it is not limited to a single instant in time[7], but it varies for each *t*, adding the time dimensionality factor to the problem. After the quick introduction to bond pricing, it is now worth to add some notions about other conditions needed for a model to work.

There are different ways to model the term structure. Each of them is designed to observe the problem of the term structure from a different point of view, resulting in multiple advantages and disadvantages. A good classification has been made by Rebonato[8], who differentiates the models considering their structure and follows.

- Statistical models, like the one that will be proposed in the next chapter, rely on their strength in the so-called Vector Auto-Regressive or VAR models, which are extensively used to forecast yield and risk premium estimates. This model has a great predictive power in contrast to the other models here mentioned, but bases the whole analysis to time series data, bringing large error in the forecast, especially for longer maturities. Its flexibility still makes it a good choice to model term structure, in addition to the possibility of easily using the implied impulse response functions and variance decompositions to get an additional understanding of macroeconomic and yields interactions[9]. From a mathematical point of view, they are basically Vector Auto-Regressive models that have their foundations in the AR (1) process, an autoregressive model of order 1[10]:

$$x_{t+1} = \mu + x_t\varphi + \nu\eta_{t+1} \tag{11}$$

In this context, μ represents the intercept for the regression and the φ its slope[11].

- Structural no-arbitrage models. These include the no-arbitrage conditions from their assumptions and explains the three components of the yield curve: expectations, risk premia, and convexity. The most known models from this group are Vasicek and Cox-Ingersoll-Ross. Both use a single factor model to estimate the curve yield, but the last incorporates a square root factor in order to avoid negative interest rates, as pre-crisis academics were not interested in considering the case.

[7] Like in Graph (1), which considers a snapshot of the yield curve at a given moment.
[8] Bond price and yield curve modelling, Rebonato
[9] Cochrane and Piazzesi (2008), Diebold and Rudebusch (2013).
[10] with $\eta_{t+1} \sim N(0,1)$.
[11] If the $\varphi > 1$ the model AR (1) could not be used. In the case of $\varphi = 0$ it would be a random walk.

- Snapshot model: such as the well-known Nelson-Siegel. They, as the name suggests, are cross-sectional models that give a glimpse of how the curve behaves in order to interpolate the yields that are unobservable using observable data. It is also worth to notice that this kind of model gives as outputs the discount bonds or – to say it in another way - the inputs for the other model types here described.

The following paragraphs present some of the most renown models of yield curve estimation, providing examples for the above-mentioned categories[12].

1.4.1 Affine Term Structure Models (ATSMs)

The most appreciated approach to the yield curve issue considers using an affine term structure model, which works by linking the term structure of interest rates with a time-invariant linear function made up by a set of variables, which can be latent or observable. This distinction is important, as from the 90s the augmentation with latent variables has brought some advantages to the model. Indeed, even if they cannot be compared to the other variables, they own an intrinsic explanatory capacity. This has led the research to include different numbers of latent variables in affine term structure models, but defining these factors with different names, such as real inflation (Dewachter, Lyrio and Maes, 2005) or real short rate (Pearson and Sun, 1994), even if their data did not include those data.

Another issue is related to the number of variables that should be included, but empirical studies (Knez, 1994) have noted that three latent are enough to explain much of the changes. Their labels change between different studies and paper, but recently they were linked to their effect on the curve instead of arbitrary names, specifically: level, slope, and curvature.

Going back to the foundations, the starting point for an affine model is the stochastic process that drives the dynamics of the variables involved[13]. This kind of process can vary, but a single factor generic geometric Brownian motion can give an idea of how it works. The one presented in equation (12) is made up of two terms - the drift and the diffusion – and is commonly used in asset pricing in combination with a Monte Carlo simulation. Considering only the first part of the equation, the dynamic process results in a constant growing yield as it is non-random (we say defined as yields, but usually in such a context an underlying asset price is used). In order to add randomness, the second term is needed, as it contains a Weiner Process for the term z_t over an interval dt and the resulting

[12] Rebonato also introduces a last class of models, defined as derivatives models. Basically, they are no-arbitrage models that try to fit the term structure of interest rates by adopting IRS models or similar, like the Hull and White model (1990).
[13] Mathematically speaking, affine means linear plus a constant More formally, a function like F : Rn → R is called affine if there exist a ∈ R and b ∈ Rn such that $F(x) = a + b^{\mathrm{T}\,\mathrm{x}}$ for all x ∈ Rn. (Piazzesi, 2010).

graph would now appear as an erratic path. This is exactly how a share price, for example, moves in a time-varying graph.

$$dr_t = r_t dt + \sigma_r dz_t \tag{12}$$

The next few paragraphs introduce four well-known models that use different approaches to solve the term structure estimations' issues. Starting from the founding Vasicek model, the focus then moves onto more complex models that consider advanced dynamics and combination of data in order to better forecast the yield curve and define what drives it.

1.4.2 Vasicek Model

The idea of using this type of dynamics in a term structure was firstly introduced by Vasicek in 1977 with a single factor model, in order to describe the movements of interest rates through time. The differential equation (13) figures an Ornstein–Uhlenbeck process with a drift. In this case, a constant long-term mean-reversion level *theta*, to be achieved through time and a speed reversion k. The idea of an interest rate mean reversion was correct, as empirical data proved right differentiating interest yields from other financial securities. This is due to the fact that interest yield cannot exceed certain values and this model was able to display negative yields, an eventuality considered as a disadvantage before the 2008 crisis and fixed with the so-called extended Vasicek models[14]. For the normal Vasicek model the equation follows:

$$dr_t = k_t(\theta_t - r_t)dt + \sigma_r dz_t \tag{13}$$

Its discrete-time version is[15]:

$$x_{t+1} = k\theta + x_t(1 - k\Delta t) + \sigma\sqrt{\Delta t}\varepsilon_t \tag{14}$$

Considering the above process, it is then possible to link it with the AR (1) introduced before. Indeed, using the same equation (14) it is possible to drive parallelism since the discretized version of the Ornstein-Uhlenbeck process is just a special AR (1) with:

$$\mu = k\theta\Delta t$$

$$\varphi = (1 - k\Delta t)$$

$$\nu^2 = \sigma^2\Delta t$$

By knowing this it is then possible to assert that the model is solvable through a simple regression. Thus, the coefficients – the values needed – can be estimated.

[14] Exponential Vasicek and Cox Ingersoll Ross.

[15] With $\varepsilon_t \sim N(0,1)$.

Following Piazzesi[16], affine models can be then considered as a class of term structure models. If this is true, the expression for a bond price P_t^T in Vasicek is then:

$$P_t^T = e^{\left(A_t^T + B_t^T r_t\right)} \tag{15}$$

A_t^T and B_t^T are coefficients that depend on T. Their values can be found through:

$$B_t^T = \frac{1 - e^{-k(T-t)}}{k} \tag{16}$$

$$A_t^T = -\left(\theta' - \frac{\sigma_r^2}{2k^2}\right)\left[B_t^T + (T-t)\right] - \frac{\sigma_r^2}{4k}\left(B_t^T\right)^2 \tag{17}$$

This approach is considered a breakthrough in term structure modeling, as its mean reversion properties differentiated bonds' pricing different from other financial securities. Moreover, its simplicity and the fact that it can be analytically solved made it the preferred approach to the subject.

This can be found out by knowing that the dynamics for the short rate are the one expressed in equation (14) and assuming or knowing the value of a starting rate namely r_0. It is then possible to derive the value of the short rate at time t r_t, resulting in:

$$r_t = \left(e^{-k(t-t0)}\right) r_0 + \left(1 - e^{-k(t-t0)}\right)\theta + \sigma_r \int_{t_0}^{t} e^{-k(t-s)} dz_s \tag{18}$$

The solution states that the short rate distribution at any time in *t* will be normal and stationary distributed. Considering other quirks of the model, the duration of a bond in Vasicek presents some particularities. Starting from equation (4), duration becomes:

$$Dur_t^T = -\frac{1}{P_t^T}\frac{1}{\beta_t^T}\frac{dP_t^T}{dy_t^T} = -\frac{1}{P_t^T}\frac{k(T-t)}{1 - e^{-k(T-t)}}\frac{dP_t^T}{dy_t^T} \tag{19}$$

1.4.3 Vasicek model's improvements

Even if the homonym model proposed by Vasicek had some limitations, such as a fixed reversion level, other studies made it possible to modify it in different ways. This specific issue was quickly solved with the Doubly Mean Reverting Vasicek or DMRV, which provides a second equation to add dynamics for θ.

$$\begin{cases} dr_t = k_r(\theta_t - r_t)dt + \sigma_r dz_t \\ d\theta_t = k_\theta(r_\infty - \theta_t) + \sigma_\theta dz_\theta \\ E[dz_t dz_\theta] = \rho dt \end{cases} \tag{20}$$

[16] Affine Term Structure Models, Piazzesi (2012).

Adding another equation to the model improves it and - in order to further enhance it - it is possible to add another one to describe the movement of the Beta, thus making a Trebly Mean Reverting Vasicek Model. This approach could surely improve the basic Vasicek but leads to an impasse, as providing more and more dynamic processes expand the calculations without really giving an edge to the model.

Since Vasicek and its derivation were only the primitive versions of this type of model, in the following years different approaches appeared on the scene to explain the term structure of interest rates. The first thing to be modified starting from a Vasicek model is the number of variables to be included. Vasicek uses only a single factor and adding another one requires only a few improvements.

Adding multiple factors can indeed improve the model, but again it can lead to an impasse as before. From empirical data, the total number of variables should not exceed six. Models that encompass more get perfect fit of the data, but their forecast is poor, especially for medium and long-term yields. According to this, as it is difficult to explain the yields only by using observable data, latent come to help. Choose which real-world data to use is troublesome though and numerous solutions were found. A good guide, made up by Dai and Singleton, helps discerning among them, providing a further taxonomy other than the one presented before.

1.5 Common techniques

1.5.1 Principal Components Analysis (PCA)

Before moving to the next models, some valuable techniques must be defined. Principal Component Analysis is a common technique that eases the dimensionality issue, employing an eigenvalue decomposition starting from a covariance matrix of the data[17]. Indeed, some calculations using all the data can lead to an immense hurdle due to the high dimensionality of the data and the techniques reduce the number of factors adopted. In other terms, given a series of yields, its PCs are defined as:

$$x^t = V^T y^t \tag{21}$$

Each component describes a certain degree of the variance: depending on the data matrix, the number of factors needed to explain can vary. In some models, like the Ang and Piazzesi (2004), only the first principal component is used, which accounts for a good 50-70% of the variance explained. This is a good starting point though and can be used to define the observable variable as to it is enough to understand how the entire curve moves. Adding the second PC would bring new data and give the

[17] Litterman and Scheinkman conducted a study on US treasuries PCA decomposition (1997).

chance to observe risk premia[18], but still would miss some effects on the curve. The discipline tends to use three, as it can describe a good percentage of the variance and avoiding the most dimensionality issues. The specific number of factors is therefore determined by the authors and on their purposes, as nothing forbids to use a desired number of factors to describe market yields. Usually, after five components it is useless adding others, as it would only create more dimensionality problems without adding anything to the explanatory power.[19]

It is important to notice that this technique works best when there is a signal to noise ratio is high. In fact, Shelns in its paper of 2009 states that:

"Measurement noise in any dataset must be low or else, no matter the analysis technique, no information about signal can be extracted".

Another requirement to correctly use this technique is a high correlation among data, which is true for yields. In their paper, Gurkaynak, Sack and Wright (2000), assume that in real-world conditions the form of the reversion speed is a diagonal matrix such that:

$$\begin{bmatrix} dx_1 \\ dx_2 \\ dx_3 \end{bmatrix} = \begin{bmatrix} k_{11} & 0 & 0 \\ 0 & k_{22} & 0 \\ 0 & 0 & k_{33} \end{bmatrix} \left(\begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix} - \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \right) dt + \begin{bmatrix} s_{11} & 0 & 0 \\ 0 & s_{22} & 0 \\ 0 & 0 & s_{33} \end{bmatrix} \begin{bmatrix} dz_1 \\ dz_2 \\ dz_3 \end{bmatrix} \tag{22}$$

1.5.2 Filtering techniques

Modeling the term structure of interest rates brings some issues regarding the data themselves. Considering a dataset used for this approach, it is common that these contain some sort of inaccuracies and relevant statistical noise. In order to solve the issue, some common techniques have been developed, leading to their massive application in the field. This approach is generally called filtration. In term of affine term structure modelling, thanks to the studies of Duan and Simonato (1995), Lund (1997), Geyer and Pichler (1998), de Jong (1998), and Babbs and Nowman (1999), the most common approach is the Kalman Filtering, also known as Linear Quadratic Estimator (LQE).

The reasons for its extensive use can be found within the possibility to use it in linear systems – which is our case – and both in discrete and continuous time with few modifications. The approach is most valuable when the time series used to perform calculations contain a lot of noise – as in this case [20] – and for yield curve applications, when the underlying state variables are unobservable.

In fact, Kalman Filtering produces an estimate of unknown variables starting from the time series data considering a joint probability distribution for each timeframe. Specifically, it is a recursive

[18] Piazzesi and Cochrane (2005).
[19] Gurkaynak, Sack and Wright (2000) have done an extensive research on this method.
[20] It is not uncommon that financial data have high correlation.

algorithm that works by doing a two-step filtering. In the first step it assumes certain starting values for the state variables, which can be their mean and variance for example (Bolder, 2001). By using these values, it moves on by guessing the value of the measurement equation through the so-called prediction step. Obtaining these permits to go to the next step, which is the observation of the effective values and to update the previously found values, eliminating part of the initial error. Recursively, the filter moves on the next timestep repeating the process and ending up with a more precise output for the estimation. A downside of the Kalman Filtering is that the errors are not assumed to be Normally distributed.

The Hodrick-Prescott is another type of filter widely used in economics research, especially in macroeconomic researches which concerns business cycles.

$$\min_{\{g_t\}_{t=-1}^{T}} \left\{ \sum_{t=1}^{T} (y_t - g_t)^2 + \lambda \sum_{t=1}^{T} [(g_t - g_{t-1}) - (g_{t-1} - g_{t-2})^2] \right\} \tag{23}$$

Compared to the Kalman Filter, which recursively weakens the statistical noise, the HP filter works by removing the cyclical components from the data. This is mainly intended for the studies in macroeconomics that intend to decompose time series, but it presents multiple disadvantages and issues. The main disadvantages can be found in the work of Hodrick and Prescott itself, as they state multiple conditions for the filter to work properly. A proper condition is that the filter needs historical data to work and cannot be used in a dynamic context of forecast. Moreover, the filter does not discern certain events from others: for example, a single shock – big enough – could lead the filter to generate a non-existent trend in the dataset. Some of these issues were underlined by Hamilton (2017), who affirm that the HP filter needs multiple adjustments in order to provide de-cycled data. For Hamilton, the use of this filter leads to spurious regressions and thus to misleading interpretations.

This is firstly noticed by Harvey and Jaeger (1993) and later by Cogley and Nason (1995). Thus, for a process such that: $y_t = y_{t-1} + \varepsilon_t$, where ε_t is white noise and $(1 - L)y_t = \varepsilon_t$, the function that drives a cyclical component for the series would result in:

$$c_t = \frac{\lambda(1 - L)^3}{F(L)} \varepsilon_{t+2} \tag{24}$$

Usually, depending on the period of the data and on the interests of the researchers, this value lambda should be adequately modified. As Hamilton states, different researches have linked the lambda to a value of 1600[21], which is an empirically determined and adjusted value, not a theoretically justified one.

Other studies, like the one of Ravn and Uhlig (2002), produced different values considering other time frequencies, but assuming the value given by Hodrick and Prescott as correct. Definitively, using it – or other values – would only create cycles or trends that are entirely created by the filter itself rather than already being present in the data. This is even worse in a dataset that encompasses a random walk. In the same paper though, Hamilton provides a solution. Considering an eventual process, his HP enhancements are based on possible different cases, for example, that the growth rate is nonstationary and the change in the growth rate is stationary. To solve all these cases, he proposes a filter based on regression, considering that the typical economic time series is best approximated by a random walk. The regression can be executed using an OLS of the observed non-stationary time series on a four period plus a constant.

$$y_{t+h} = \beta_0 + \beta_1 y_t + \beta_2 y_{t-1} + \beta_3 y_{t-2} + \beta_4 y_{t-3} + v_{t+h} \tag{25}$$

The cyclical component is instead given by the error term. Its decomposition is:

$$v_{t+h} = y_{t+h} - \hat{\beta}_0 - \hat{\beta}_1 y_t - \hat{\beta}_2 y_{t-1} - \hat{\beta}_3 y_{t-2} - \hat{\beta}_4 y_{t-3} \tag{26}$$

Like Hodrick and Prescott, Hamilton too suggests different values in relation to the analysis' purposes: h = 8 in case of quarterly data concerned with business cycles analysis and h = 20 for financial cycles.

Even if it should improve the HP filter, a recent study made by Schüler (2018) proves that some of the disadvantages could still raise some concerns, such as its tendency in creating spurious regressions, emphasizing cycles that endure more than expected. At the same time, this issue almost erases short term variations.

"However, when applied to a random walk, Hamilton's (2017) regression filter reduces to a difference filter. In the case of difference filters, we know that certain cycle frequencies are canceled, and others emphasized".

Nonetheless, Schüler eventually considers the Hamilton filter as a better instrument compared to the HP.

[21] Hodrick and Prescott motivated their choice of λ = 1600 assuming a change of 5% in the cyclical component, having quarterly data. This is justified by saying that a change trend component would be around $\left(\frac{1}{8}\%\right)$, Giving out a $\lambda = \frac{\sigma^2 c}{\sigma^2 v} = \left(\frac{5}{\frac{1}{8}}\right)^2 = 1600$.

1.6 Ang and Piazzesi Model (2003)

The Ang and Piazzesi model combine the techniques seen in the first paragraphs in order to forecast bond yields, describing joint dynamics for macroeconomic variables and bond yields in a VAR. The model is Gaussian and consists of five variables organized in the observable and unobservable categories. The two observables are built upon macroeconomic data opportunely reorganized, while the three orthogonal latent ones encompass the yield curve's movement that cannot be forecasted by observable data alone. The model uses the no-arbitrage rule in order to set restrictions. Including macroeconomic variables is useful to understand how yields move: from the data used in the paper, they explain up to 85% of yields' movement at short and medium maturities. The VAR approach is also useful for the reasons seen at the beginning of paragraph (1.4) as it is possible to compute IRs and Variance decomposition easily in order to clearly see how the macro shocks impact the term structure. Latent variables impact can be seen in the same way and then compared to macro variables.

The model is set up starting from the observable variables. They are defined as Inflation and Real activity and both encompass different useful indexes. Inflation is made up by Consumer Price Index (CPI), Purchase Price Index (PPI) and Spot Market Commodity Prices (PCOM), while Real activity by HELP (Help Wanted Advertising in Newspapers), Unemployment (UE), growth rate of Employment (EMPLOY) and Industrial Production (IP). At this point there is a problem of dimensionality and PCA can help to solve it out. In this case, just the first components can be used to perform the analysis, as they explain enough variability for the model's purposes. The next step is normalizing these PCAs and stacking them into two separate vectors of dimension 3X1 and 4X1 named Z_t^1 and Z_t^2, represented with $Z_t^i = Cf_t^{0,i} + \varepsilon_t^i$,

Usually, other models, such as the Duffie and Kan (1996), follow a Taylor rule to specify the short rate such that the movements in *t* in the short rate are linked to macroeconomic variables movements at the same time. In a variant of this, a forward-looking version of the same Taylor rule, Clarida (2000) states that the Central Bank reacts both to the expected inflation and output gap[22], including forecast errors in the shock v_t. Ang and Piazzesi present two variations of their idea: a VAR model which encompasses macro factors plus three latent yields to forecast the model implied yields. The dynamics process follows a VAR (12) for the macro derived variables, while the unobservable ones an AR (1). The process estimation is performed in a double step: first they find the short rate and the VAR parameters through the use of an OLS, while the rest are derived using a MLE. This type of process allows to define all the factors needed for the yields' forecasts but is quite demanding in term of

[22] Output gap is the gap between the theorical output an economy can reach and the output that currently has. In some model it is used to name latent.

calculations. But their idea has led to different models that uses macroeconomic inputs to improve term structure models.

1.7 Dewachter, Lyrio and Maes

Another interesting model has been proposed by Dewachter, Lyrio and Maes in 2006. It can be considered as an evolution of the Ang and Piazzesi Model, as it is based too on macroeconomic and latent variables, but it works in continuous time, thus making it a continuous-time vector error correlation model (VECM). Again, there are two observable and three latent variables: the first is defined as output gap and inflation, while the others are considered to be the real interest rate and the central tendencies of inflation and the real interest rate.

From their results, it appears that the observable variables they used to forecast the yields do not explain the long end of the term structure. This is not a surprise, as other models had the same issues for long term forecast, but this model presents an improvement. Dewachter, Lyrio and Maes observed that central tendencies of the macroeconomic variables offered better performance, but also confirmed that the latent factors have an important role in the formulation of the interest rate policy rule. Moreover, as already confirmed in Ang and Piazzesi, they help to describe better the yield curve over time.

The equations that define the dynamics of the model follows:

$$\begin{cases} dy_t = \left[k_{yy}y_t + k_{yy}(\pi_t - \pi_t^*) + k_{y\rho}(\rho_t - \rho_t^*)\right]dt + \sigma_y dW_{y,t} \\ d\pi_t = \left[k_{\pi y}y_t + k_{\pi\pi}(\pi_t - \pi_t^*) + k_{\pi\rho}(\rho_t - \rho_t^*)\right]dt + \sigma_\pi dW_{\pi,t} \\ d\rho_t = \left[k_{\rho y}y_t + k_{\rho\rho}(\pi_t - \pi_t^*) + k_{\rho\rho}(\rho_t - \rho_t^*)\right]dt + \sigma_y dW_{y,t} \\ d\pi_t^* = \sigma_\pi dW_{\rho,t} \\ \rho_t^* = \gamma_0 + \gamma_{\pi^*}\pi_t^* \end{cases} \tag{27}$$

The model, apart from including the above-mentioned variables, includes a filtering method in order to recover the time series of the factors from the data. The idea of filtering is not new in the panorama of ATSM, as since the 90s others – such as Chen and Scott (1993) and Pearson and Sun (1994) – have done studies regarding the different methods and their issues. One of the most feared is to obtain erroneous filtered data that include distortions created by the filter. In order to avoid it, the authors use a Kalman filter, previously explained in detail in paragraph (1.4.6) and it is usually very efficient in a model where all the factors are assumed to be latent[23]. The model avoids a couple of problems that Ang and Piazzesi solve differently, starting with the incorporation of the macroeconomic factors inside the model. As the unobservable variables – real output gap and real inflation – need to be analyzed,

[23] De Jong (1991, 2003) analyse the use of a diffuse Kalman filter with MLE and how to smooth its output.

the state vector of the model needs to be updated in order to consider them as non-latent. To estimate these parameters, a one-step-ahead prediction is used to fit the output gap and inflation with the measurement equation (26). This is not a real issue in Ang and Piazzesi, as they did not use any filtering technique, but instead a double-step regression. The second one regards the restrictions of the dynamics: Dewachter, Lyrio and Maes opt not to make any restrictions[24], contrary to Ang and Piazzesi.

$$\begin{bmatrix} \widehat{y_{1}{}_t^{\tau 1}} \\ \cdots \\ \widehat{y_{m}{}_t^{\tau m}} \\ y_t \\ \pi_t \end{bmatrix} = \begin{bmatrix} a \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} B \\ e'_2 \\ e'_3 \end{bmatrix} \begin{bmatrix} y_t \\ \pi_t \\ x_3 \\ \ddot{\pi}_t \end{bmatrix} + \varepsilon_t \tag{28}$$

$$\text{With } a = \left(\left(\frac{a_{\tau 1}}{\tau 1}\right), \dots, \left(\frac{a_{\tau m}}{\tau m}\right) \right)' \text{ and } B = \begin{pmatrix} \left(\frac{b'_{\tau 1}}{\tau 1}\right) \\ \cdots \\ \left(\frac{b'_{\tau m}}{\tau m}\right) \end{pmatrix}$$

1.8 Nelson Siegel Model

The Nelson Siegel model is here reported as an example of snapshot models. This model is a perfect candidate, since it is as simple as the Vasicek model described before and can be quite adaptable to the different shape of the yield curve. In the years of its formulation, there was a debate regarding the use of polynomial or exponential splines to estimate the present value function, as both presented different advantages and disadvantages. In brief, polynomial functions were already used in the 70s, but they presented some issues, such as a bad fitting considering the whole curve, creating artifices. Moreover, there is a propensity in creating, as Chambers (1984) states: *"explosive tendencies…toward the end of the fitted maturity range"*.

Nonetheless, the use of exponential spline – also recommended by Vasicek and Fong (1982) – would present similar concerns.

The model presents the same no-arbitrage[25] results of Ang Piazzesi (2004) and is based on a transformation of the non-linear estimation issue into a linear one. In order to solve the model, the common approach is minimizing the sum of squared errors (SSE). This can be done by using an OLS, specifically a series of OLS in order to estimate model's conditionals to maximize the R^2. This procedure is defined by other authors (Annaert, Claes, De Ceuster and Zhang) as a "grid search".

[24] "In other words, one should not prevent the possibility of imaginary eigenvalues with respect to the spectral decomposition of the matrix K", Dewachter, Lyrio and Maes.

[25] Coroneo, Nyholm and Vidava-Koleva (2008) find that the parameters of the Nelson-Siegel are "not statistically different from those obtained from the 'pure' no-arbitrage affine-term structure models".

Otherwise, it is always possible to use an OLS already having a grid of pre-specified values[26] or even assigning a fixed shape parameter[27].

The model, though, has some disadvantages: being nonlinear, its results depend mostly on the technical capacity of the researcher's using it, on the initial values chosen for the optimization process[28] and on the shape parameter[29]. Other issues comprehend large estimated errors and the chance to obtain negative long-term interests.

Formulating it in matrix form, the spot rate curve can be shown as:

$$r_t^T = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}' \begin{bmatrix} 1 \\ \dfrac{\lambda\left(e^{-\frac{T}{\lambda}}\right)}{T} \\ \dfrac{\lambda\left(e^{-\frac{T}{\lambda}}\right)}{T} - e^{-\frac{T}{\lambda}} \end{bmatrix} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} \begin{bmatrix} r_0 \\ r_1 \\ r_2 \end{bmatrix} \tag{29}$$

Or alternatively:

$$y_t^T = \beta_0 + \beta_1 \frac{\left[1 - e^{\left(-\frac{T}{\lambda}\right)}\right]}{\frac{T}{\lambda}} + \beta_2 \left(\frac{\left[1 - e^{\left(-\frac{T}{\lambda}\right)}\right]}{\frac{T}{\lambda}} - e^{\left(-\frac{T}{\lambda}\right)} \right) \tag{30}$$

The term r_0, r_1 and r_2 represent respectively the level, the slope and the curvature of the model. The level is clearly assumed to be constant in order to consider a given value for the short rate and it is usually referred to as the long-run yield level. The slope respectively follows the form of an exponential decay function. β_1 is then a function of time to maturity with a weight λ and thus it exponentially decays to zero as the maturity grows, resulting in a bigger impact on the curve in the medium term. Lastly, the curvature is a Laguerre function[30]. It is again a weighted function, but its shape presents results in a curve starting from zero when maturity is zero, rising and then decreasing as maturity increases. β_2 is usually defined as the hump term.

However, even if it has been used for numerous studies, the model presents some insurmountable obstacles. A way to address them is a modification proposed by Svennson in 1994. Calibration remains

[26] This is the approach that Nelson and Siegel followed in their 1987 first version of the model.
[27] Diebold and Li (2006).
[28] Cairns and Pritchard (2001).
[29] In their paper Diebold and Li (2006) empirically derived an annualized fixed parameter equal to 1.37.
[30] Mathematically speaking, Laguerre polynomials are the solution of a second-order linear differential equation properly named Laguerre's equation.

difficult, like in the original Nelson Siegel model, but adding another hump term in equation (30) helps to improve the curve fit. Then equation (30) becomes:

$$y_t^T = \beta_0 + \beta_1 \frac{\left[1 - e^{\left(-\frac{T}{\lambda}\right)}\right]}{\frac{T}{\lambda}} + \beta_2 \left(\frac{\left[1 - e^{\left(-\frac{T}{\lambda}\right)}\right]}{\frac{T}{\lambda}} - e^{\left(-\frac{T}{\lambda}\right)} \right) + \beta_3 \left(\frac{\left[1 - e^{\left(-\frac{T}{\lambda}\right)}\right]}{\frac{T}{\lambda}} - e^{\left(-\frac{T}{\lambda}\right)} \right) \qquad (31)$$

Chapter 2: A Joint Macroeconomic and Term Structure VAR

2.1 A short introduction on Macro models

In the first chapter different models were presented in order to give a glimpse of the term structure modelling literature. The models were just briefly introduced, presenting their perks, their advantages and disadvantages compared to others, but it appears that the problem of modelling the term structure of interest rates and presenting valid forecasting is not easy to solve. It is worthy to say though, that recently the literature has progressed very quickly thanks to the improvements in calculation power and estimation techniques. Moreover, the financial crisis occurred in 2008 has led to a change of direction for some studies[31]. The importance of this discipline has also seen a renowned interest from policymaker and funds, as different authors started considering using as inputs macroeconomic variables when not the role of the monetary policy currently ensued by Central Bank. Different approaches are adopted to determine correctly a possible value for the future interest rate. The most commonly used models, with the notable exception of the CIR due to its limits, are reported in chapter one or at least named. The reason why a Central Bank and researchers want to use more models is due to their intrinsic structure and their different approaches to the problem, for example using cross-section or time-series data instead of panel. All the different results can then help to define new policies in case of CBs studies or investment opportunities in case of investment or pension funds. In the last two decades, the literature has acknowledged that the use of good data could not be enough to explain the whole variations of the yield curve. As can be seen in chapter one, this problem has been partially solved by adopting latent or unobservable variables derived from – for example – the yields themselves in order to improve the explicatory power of the model. This approach is now widely used, and the literature is abundant of examples. The one that is analysed in this chapter follows the “tradition” now well-established of adopting three latent variables in order to explain some otherwise inexplicable movements in the yield curve. This approach is the Ang and Piazzesi Model (2003) and constitutes the basic for the analysis conducted in this and in the next chapter. It is then used as a base for further enhancements that follow in the same chapter. To be more thorough, this first part looks at the model from a theoretical and estimation perspective, looking at the ideas that lead its creation and finding the needed parameters to obtain the results. This will follow much of the original paper, but in a more detailed way to arrive to the estimation process with a proven ground of mathematical results.

[31] As specified before, numerous models were considered faulty or not well specified if they allowed for negative interest rates before the crisis.

Moreover, it also serves as a stage from which the modification starts, introducing another model that comes in handy for what regards the use of the macroeconomic data.

2.2 The Ang and Piazzesi Model: a complete derivation

2.2.1 Data setup

The concept at the base of the Ang and Piazzesi model comes from the intuition that using macroeconomic variables, namely a set of data series, can improve the forecasting performance. In their paper, the variance decomposition show that these factors can explain up to 85% of variation in bond yields, especially in the short term. The restrictions are derived from the no-arbitrage condition and seems to improve the forecast performance. The model, as stated before, uses a VAR structure with a combination of observable macro factors and latent variables as inputs in order to explain the movements of the yield curve. This is done through a double step estimation, using first an OLS method and at last an MLE approach. Each step involves the estimate of four distinct variables: the first step looks for the values for the macro factors and two coefficients of the short rate equation, while the second look into the parameters for the market price of risk and the remaining coefficients for the short rate. From a theoretical point of view, it combines the original Taylor's Rule and a modified version of the same with the short rate equation for affine models, which in this case incorporates both types of factors. The paper makes also a distinction between two models analysed: one with a policy rule that contains only contemporaneous factors and one that also contains lags. The yields used are five and specifically the 1, 3, 6, 12 and 60 months from the CRSP Fama-Bliss dataset. Starting from the Macro Factors, seven series are used to define two different factors, namely Inflation and Real Activity: three series are used to determine inflation and four for the real activity. In order to obtain these two factors, the data are normalized, and an eigenvalue decomposition is performed in order to extract the first principal component from the series and form the factors. The general formula is:

$$Z_t^i = C f_t^{0,i} + \varepsilon_t^i \tag{32}$$

In this equation, Z represents the vector (3 X 1) or (4 X 1) – depending on the factor – at time *t*, C is the vector of the factor loading vector with the same dimensions of Z, with *i* being equal to Inflation or Real Activity. The error term ε and the factor *f* have: $E\left[f_{\mathrm{t}}^{0,\mathrm{i}}\right] = 0$, $cov\left[f_{\mathrm{t}}^{0,\mathrm{i}}\right] = I$, $\mathrm{E}[\varepsilon_t^i] = 0$ and $\mathrm{cov}[\varepsilon_{\mathrm{t}}^{\mathrm{i}}] = \Gamma$, with Γ diagonal.

2.2.2 The dynamics of the model

By knowing the two macro factors and organizing them into a vector $\mathrm{f}_{\mathrm{t}}^{0} = \left(\mathrm{f}_{\mathrm{t}}^{0,1}\ \mathrm{f}_{\mathrm{t}}^{0,2}\right)$. In this way it is possible to consider a bivariate VAR (*12*) process in the form[32]:

$$f_t^0 = \rho_1 f_{t-1}^0 + \rho_2 f_{t-2}^0 + \cdots + \rho_{12} f_{t-12}^0 + \Omega u_t^0 \tag{33}$$

in the general form, the VAR(p) is:

$$f_t^0 = \sum_{k=1}^{p} \Phi_k\ f_{t-k}^0 + \Omega\, u_t^0 \tag{34}$$

Considering the VAR (12) used in the model, the literature and the authors themselves consider that using too much data could lead to overfitting, as by doing a regression on the lags for the Macroeconomic data it is clear that only a few lags are significant. In this case though, it is possible to pick different lags in order to avoid missing some eventually useful data, as Ang and Piazzesi decided to do.

Following equation (33) it is then possible to stack all the observable variables $\mathrm{f}_{\mathrm{t}}^{0}$ into a vector $\mathrm{X}_{\mathrm{t}}^{0}$ such that:

$$X_t^0 = \begin{bmatrix} X_t^0 \\ X_{t-1}^0 \\ \cdots \\ X_{t-p-1}^0 \end{bmatrix} \tag{35}$$

Knowing that X_t^0 has dimension (*np* x *1)*, the process follows equation (36). In order to bring more clarity, it is useful to write in companion form.

$$X_t^0 = \Phi^0 X_{t-1}^0 + \Sigma_0 \varepsilon_t^0 \tag{36}$$

Using this form can help avoiding some calculations' issues, as the matrix of Φ of dimension *(np* x *np*), now contains the coefficients needed along the first row, such that:

$$\Phi = \begin{bmatrix} \Phi_1 & \Phi_2 & \cdots & \Phi_{(\mathrm{p}-1)} & \Phi_p \\ I & 0 & \ldots & 0 & 0 \\ 0 & I & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & I & 0 \end{bmatrix}$$

[32] With each coefficient ρ of dimension *(2x2)*.

The matrix for Σ has the same dimension, but only its first element is a non-zero value. Specifically, the Ω, which is a lower triangular matrix. The error term has a similar structure, with a single non-zero value u_t^0 as the first element and the rest composed by zeros. This is a first setup concerning the observable macro factors.

Moving to the latent factors, the idea of using three unobservable variables is quite common in the literature and allow to explain much of the variations of the yields. As they are assumed to be unconditionally orthogonal to the macro factors, it is possible to follow a normalization of these factors as the observable variables. Following Ang and Piazzesi[33]:

"*The idea of these normalization in a VAR setting is that affine transformations and rotations of the unobservable factors lead to observationally equivalent yields*"

These independence assumption between the observable and unobservable factors has some theoretical contradictions. Following it, there should not be any link between them, which means that the monetary policy has no influence on the future macroeconomic state. This goes against the numerous empirical evidences and the fact that CBs operate in a context where changing the policy would not have any impact on the data it is observing to define the policy.

In order to find the latent factors, the setting follows, again, a common approach. It is defined as an AR (1) process, such that:

$$f_t^u = \rho f_{t-1}^u + u_t^i \tag{37}$$

The coefficient ρ has dimension (*3* x *3)* and is lower triangular. This form permits to estimate the unobservable factors that have to be included in the system, stacking them into a single factor as $F_t = \left(f_t^{o'} f_t^{u'}\right)$. In this way, the state dynamics resemble the previous equations, starting with a Gaussian AR(p):

$$F_t^0 = \Phi_1 F_{t-1}^0 + \Phi_2 F_{t-2}^0 + \cdots + \Phi_{12} F_{t-12}^0 + \theta u_t$$

Again, in companion form, the equation that comprises both observable and unobservable variable is given by:

$$X_t = \mu + \Phi X_{t-1} + \Sigma \varepsilon_t \tag{38}$$

[33] For a more complete discussion Dai and Singleton (2000) conduct an analysis on normalizations in affine term structure models.

With:

$$\Phi = \begin{bmatrix} \Phi^o & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \Phi^u \end{bmatrix}$$

While Σ follows a similar structure and the same dimensions, with Σ^0as the first element of the matrix and a *(3* x *3)* matrix of I in the bottom right corner, the error term $\varepsilon_t = \left(u_t^{o\prime}, 0, \dots, 0, u_t^{u\prime}\right)'$

The non-observable or latent ones need to be derived considering a policy rule, like the one proposed by Taylor (1993). In this situation, the movements in the short rate r_t can be found in the factors derived before f_t^0 in addition to a non-observable component, assumed to be an orthogonal shock v_t [34]. As Ang and Piazzesi state, this policy is handy since the macro factors proposed in the original Taylor Rule were two, simplifying the process and avoiding the need to adapt it to the model.

$$r_t = a_0 + a_1' f_t^o + v_t \tag{39}$$

In the same paper, the authors set up the model considering a second version of this rule proposed by Clarida (2000), in which the Taylor Rule is said to be *forward-looking*. In this context, the CB does not respond to the actual macroeconomic factors, but to their expected values. This is also referred to as lagged Taylor Rule and it is used to arrange a second version of the model.

$$r_t = b_0 + b_1' X_t^0 + v_t$$

Following the literature, ATSM follows a close equation to the one provided in (39) but using the latent factors X_t^u. So, it becomes:

$$r_t = c_0 + c_1' X_t^u$$

Imposing orthogonality for the latent factors respects to the macro factors and adding those, the final equation for the short rate is:

$$r_t = \delta_0 + \delta_{11}' X_t^o + \delta_{12}' X_t^u \tag{40}$$

2.2.3 Pricing Kernel

The pricing kernel of the model constitutes the heart of the model and implies a no-arbitrage condition necessary to set up the bond prices in their affine form. This assumption is very popular in the literature and is interpreted in different ways[35], but it generally improves the quality of the results.

[34] This shock can be interpreted in different ways. Christiano, Eichenbaum and Evans (1996) suggest that it could be interpreted as a monetary policy shock and this assumption is considered valid in this thesis and in Ang and Piazzesi.
[35] Duffee (2002), Dai and Singleton (2002) for example.

The starting point in the Ang and Piazzesi model is the notion that under risk-neutral expectations, the price of an asset V_t is given by:

$$V_t = \mathbb{E}_t^Q[e^{-r_t}(V_{t+1})] \tag{41}$$

In order to convert the risk-neutral into the real-world measure, a solid method concerns the use of the Radon-Nikodym derivative. The model then, considering that $E_t^Q\ [Z_{t+1}] = \frac{E_t[\xi_{t+1}Z_{t+1}]}{\xi_t}$, uses it in order to include the no-arbitrage condition by linking it to the condition of existence of the term ξ_{t+1}. Hypothesizing that this term follows a log-normal dynamic, it becomes:

$$\xi_{t+1} \ = \ \xi_t e^{\left(-\frac{1}{2}\lambda_t'\lambda_t \, - \, \lambda_t'\varepsilon_{t+1}\right)} \tag{42}$$

With $\varepsilon_{t+1} \sim \mathrm{N}(0, \mathrm{I})$. The λ_t represents the market price of the risk at time *t* and is time varying. It can be modelled into an affine model by using the same one-factor model seen before. To be more thorough, the process uses a common approach in ATSMs as one of the most renown paper of Dai and Singleton (2002) suggests using it due to the advantage of having a simple form to model the shocks for the factors[36]. This is commonly referred to as essentially affine form and accounts for only contemporaneous movements for the market price of risk.

$$\lambda_t = \lambda_0 + \lambda_1 X_t \tag{43}$$

As a classic one-factor model, the constant follows the usual form of a vector of dimension *(K* x *1)*, while the coefficient of the X_t, λ_1, has a matrix of dimension *(K* x *K)*. This is useful in order to put eventual restrictions for the macro or macro-lag models' coefficients. It is also important to notice that the source of randomness for equation (42), as it follows the same as the Radon-Nikodym derivative, as ε_{t+1}, so it is shared by both equations. Going back to the structure of the price of the risk, Ang and Piazzesi describe the number of free parameters in both the vector and the matrix. In the first, it is possible to find a number equal to K_1+K_2 such that there is an upper *(*K_1 x *1)* row and a lower *(*K_2 x *1)* row. The matrix presents a $(K_1$+$K_2)^2$ number of free parameters, sharing a similar structure to the vector λ_0, presenting corners of different shapes based on the dimension of $K_1\ and\ K_2$. This leads to the concept of linking the market price only to the variables included in the model, namely the macroeconomic and the latent factors. Following the construction, only five elements are non-zero values for the vector λ_0 and thirteen for the matrix λ_1, simplifying again the task of estimation.

[36] Duffee (2002) too suggests using this form.

It is also interesting to notice that this equation is used in different models in order to easily set time-varying risk premia through the use of a non-zero λ_1 matrix' coefficients[37]. Otherwise, to simplify the estimation process, it is possible to derive only the constant risk premia, assuming that the matrix λ_1 contains only zeros.

Going back to equation (42), the authors assume a pricing kernel form for:

$$M_{t+1} = e^{-r_t}\left(\frac{\xi_{t+1}}{\xi_t}\right) \tag{44}$$

Having all these equations it is possible to combine them in order to setup both the final version of the pricing kernel adopted by Ang and Piazzesi and the specification for the discount bond price starting from the canonical exponential function together with the Radon-Nikodym derivative. Thus, the final equation that defines the pricing kernel can be found including the equation for the short rate inside the assumption on the kernel, such that:

$$m_{(t+1)} = e^{\left(-\frac{1}{2}\lambda_t'\lambda_t - \delta_0 - \delta_1 X_t + \lambda_t'\varepsilon_{t+1}\right)} \tag{45}$$

The starting point of this derivation is then:

$$P_t^{\mathrm{n}} = e^{\left(\bar{A}_n + \bar{B}_{\mathrm{n}}' X_t\right)} \tag{46}$$

As Ang and Piazzesi state:

> *"The dynamics of the short rate, in combination with the Radon-Nikodym derivative form a discrete Gaussian K-factor model with $K_1 * p$ observable factors and K_2 unobservable factors".*

The Gaussian form, as said before, is also granted by the normalization of the factors previously done. This same form can be used in combination with equation (45) and applying the no-arbitrage condition seen before, considering true the expression:

$$E_t\left(m_{(t+1)} p_{(t+1)}^n\right) = 1$$

This is then set by:

$$p_t^{(n+1)} = E_t\left(m_{(t+1)} p_{(t+1)}^n\right) \tag{47}$$

[37] As it is in Vasicek (1977).

The whole derivation for the solution of equation (46) can be determined by substituting for A and B the values obtain through the policy rate and the market price of the risk. It can be then found recursively from equations (45) and (47).

$$\bar{A}_{n+1} = \bar{A}_n + \bar{B}_n'(\mu - \Sigma\lambda_0) + \frac{1}{2}\bar{B}_n'^{\Sigma\Sigma'}\bar{B}_n - \delta_0 \tag{48}$$

$$\bar{B}_{n+1}' = \bar{B}_n'(\Phi - \Sigma\lambda_1) - \delta_1' \tag{49}$$

Reworking equation (46) the price of a bond is then:

$$p_t^{(n+1)} = e^{\left(\bar{A}_n + \bar{B}_n'(\mu - \Sigma\lambda_0) + \frac{1}{2}\bar{B}_n'^{\Sigma\Sigma'}\bar{B}_n - \delta_0 + \left(\bar{B}_n'(\Phi - \Sigma\lambda_1) - \delta_1\right)X_t\right)} \tag{50}$$

Rewriting this equation in term of A_1 and B_1, we have $A_1 = -\delta_0$ and $B_1 = -\delta_1$.

2.2.4 Estimation Process

At this point, it is useful to recap all the procedures to obtain the desired values and to finally move on to its possible modifications. Not considering the data used, as their choice constitutes a paragraph in Chapter 3, the model collects a system of yields and macroeconomic variables. The last ones are transformed with PCA, using a single principal component for each macro factors, reducing the dimensionality in order to simplify the system and normalized. The next step requires the use of a VAR (p) on the macro factors, later reduced into a VAR (1) for the previous results. The yields are transformed to define a group of latent or unobservable factors using a VAR (1) structure assuming certain yields to be known without error and inferring on them. These two groups are then inserted into a system in order to start the estimation process. Having defined all the components needed, it is then possible to define the six parameters vector θ, which is then divided into two four parameters vector θ_1 and θ_2 to procedurally follow the estimation process. The first step of the estimation involves the use of Ordinary Least Squares, as above mentioned, which permits to obtain the coefficients needed to perform the second step, the Maximum Likelihood estimation. The idea of using a two-step is not new in the literature and in this case the OLS approach is useful, as it permits to avoid *"Explosive yield dynamics"* that could occur by doing a single step estimation using ML as Ang and Piazzesi state[38].

The model can be estimated at this point: the first step involves the short rate equation seen before (40), which contains Taylor's rule plus a term that contains the unobservable variables. The coefficient δ_0 and δ_{11} can be estimated using an OLS and kept fixed to start the second estimation step. The

[38] Other examples of this procedure can be found in models that uses a similar approach, such as the ACM.

parameters to be found are then: $\Sigma, \delta_0, \delta_{11}, \Phi^o, \lambda_0, \lambda_1, \delta_{12}$ and Φ^u. This vector θ can be seen as split into a vector $\theta_1 = \Sigma, \delta_0, \delta_{11}, \Phi^o$ and a vector $\theta_2 = \lambda_0, \lambda_1, \delta_{12}, \Phi^u$.

Before that, a little step back is needed. In the preceding paragraph states that the objective of this model is to transform a system of yield and macroeconomic factors into a system of a known and unknown one. This can be useful for the last estimation step. The approach used to perform is then defined through the Maximum Likelihood function, derived starting from the reformulation of the model. Considering that the goal of the model is the forecast of zero-coupon yields, what can be found on the right side of the equation is a constant A, the term for the macroeconomic factors and the term for the unobservable.

$$Y_t = A + B^o X_t^o + B^u X_t^u \tag{51}$$

Or, in matrix notation:

$$\begin{bmatrix} y_{n_1}^t \\ y_{n_2}^t \\ \vdots \\ y_{n_1}^t \end{bmatrix} = \begin{bmatrix} A_{n_1} \\ A_{n_2} \\ \vdots \\ A_{n_1} \end{bmatrix} + \left[\begin{bmatrix} B'_{n_1}(1) & \cdots & B'_{n_1}(N-K_2) \\ \vdots & \ddots & \vdots \\ B'_{n_N}(1) & \cdots & B'_{n_N}(N-K_2) \end{bmatrix} \begin{bmatrix} B'_{n_1}(N-K_2+1) & \cdots & B'_{n_1}(K) \\ \vdots & \ddots & \vdots \\ B'_{n_N}(N-K_2+1) & \cdots & B'_{n_N}(K) \end{bmatrix} \right] \begin{bmatrix} X_t^o \\ X_t^u \end{bmatrix}$$

Also, inverting equation (51) for the estimated X of unobservable values $\hat{X}_t^u$ is then equal to:

$$\hat{X}_t^u = (B^u)^{-1} \; (Y_t - \; A - B^o X_t^o)$$

At this point, a measurement matrix of dimension *(N* x *(N - K_2)*should be added. This is needed in order to write the Maximum Likelihood function as it needs to be constructed in terms of both observable and unobservable variables.

$$Y_t = A + B^o X_t^o + B^u X_t^u + B^m u_t^m \tag{52}$$

The general idea to derive the ML then comes from the intuition that it should first define the joint conditional density of the entire system, comprising the yields, the macro factors and the observation errors. So, it should be first acknowledged that equation (53) represents the starting point in the derivation of the needed ML function:

$$L(\theta) = \prod_{t=2}^{T} f(\, Y_t \; X_t^o | Y_{t-1} \; X_{t-1}^o) \tag{53}$$

Which can be then as the log-likelihood function of the previous equation, such that:

$$log\big(L(\theta)\big) = -(T-1)log|det(J)| - (T-1)\frac{1}{2}log\big(det(\Sigma\Sigma')\big) - \frac{1}{2}\sum_{t=2}^{T}(X_t - \mu - \Phi X_{t-1}) - \frac{(T-1)}{2}log\sum_{i=1}^{N-K_2}\sigma_i^2 - \frac{1}{2}\sum_{t=2}^{T}\sum_{i=1}^{N-K_2}\frac{(u_{t,i}^m)^2}{\sigma_i^2}$$

In which the Jacobian Matrix J can be found as[39]:

$$J = \begin{bmatrix} \frac{dX_t^o}{dX_t^o} & \frac{dX_t^o}{dX_t^u} & \frac{dX_t^o}{du_t^m} \\ \frac{dY_t}{dX_t^o} & \frac{dY_t}{dX_t^o} & \frac{dY_t}{dX_t^o} \end{bmatrix} = \begin{bmatrix} I & 0 & 0 \\ B^o & B^u & B^m \end{bmatrix}$$

2.2.5 Conclusions: advantages, disadvantages and possible improvements

The idea of using the Ang and Piazzesi approach as a base to analyse the term structure is not new: different authors have already provided modifications to this model. One of the most notable works is also briefly explained in the first chapter and is the continuous-time version provided by Dewachter, Lyrio and Maes (2005), which largely diverts from the original model's specifications and estimation. At the same time, it is interesting to observe a possible change of inputs, keeping still the number of factors, but increasing their data series. Even with a different approach, Mönch (2005) partially follows the idea of using a VAR as Ang and Piazzesi do but using an "augmented" model, including more macroeconomic factors and eliminating latent variables. This kind of model has been introduced in 2005 by the work of Bernanke, Boivin and Eliasz, which has its foundations fundamentally three concepts: the elimination of latent variables, the use of a large macroeconomic dataset and the introduction of the short rate in the dynamic of the model. In order to maintain a parsimonious model, the number of macro factors should be limited, following the idea of extracting useful components that other authors ensure by adopting principal components analysis or other similar techniques. In its work Mönch (2005) states that different topics developed from the original Ang and Piazzesi work can be exploited, in particular, the idea of using PCA on macroeconomic variables to explain the yield curve and reduce the dimensionality of the problem. Also, the use of the first principal component as done by Ang and Piazzesi is not efficient in terms of explicatory power, even if it allows to avoid eventually dimensionality issues. The use of only two variables to define the whole macroeconomic state is though scarce, as using only inflation and real activity data (or the more generally used output gap) does not capture a whole economy state. In other words, using so few data could cause the loss of

[39] For simplicity it is reported in this matrix notation, as Ang and Piazzesi do, but the matrix for I has dimension $(N-K_2)$ x $(N-K_2)$, while the two zeros have dimension of $(N-K_2)xK_2$ and $(N-K_2)\ x\ m$. J is also a square matrix.

useful information to add to the model in order to present a more correct forecast. The main reason to prefer such an approach can be found in the behavior of CBs, as they prefer to set their interest rates' goals respect to the whole economy or, as Mönch suggests, in a *"data-rich environment"*. The model also uses a similar approach to the AP, but changes some of its fundamental becoming a so-called Factor Augmented Vector Auto Regression, or simply FAVAR. This can be a good starting point in order to enhance Ang and Piazzesi, increasing the macroeconomic explanatory power while maintaining the latent variables. The idea is then the same used in the FAVAR by Bernanke, Boivin and Eliasz (2005), observing that using data such as the inflation could not realistically be perfectly observable, and they can generally be noisy. Other studies, conducted in the same years in which Ang and Piazzesi were developing their model, present similarities[40], using dynamic factor models which exploits the use of large macro information dataset. This approach is then useful to separate the commonly used cross-sectional data into what they call "common components", which incorporates the most useful information regarding the macro factors and allowing a better separation among them and the noise.

2.3 Implementing a FAVAR version of the Ang and Piazzesi Model

2.3.1 Setup data and policy rule

Starting from these observations, it is worthy to develop a similar context in which the Ang and Piazzesi (AP) can be improved using some concepts used for other methods and changing the nature of latent variables. Doing this means changing the original model into a version that incorporates different factors. The Mönch model, for example, used a complex dataset to which extract different PCs, noticing that until the 10^{th} there is a variance explanatory power equal to 70% of the data. Due to dimensionality issues, again it is worthy to use fewer and the author decided to adopt four. This is difficult to implement in the Ang and Piazzesi model for its numerous restrictions, but it is a starting point. Also, depending on the data chosen, which are then well defined in the next chapter, the eigenvalue decomposition appears to be more useful compared to the original model's specifications. Considering the AP, it is though possible to change the equation that drives the short rate, specifically using the same lagged Taylor rule to define the dynamic of the short rate. This was already specified in the AP paper and it proved to be more effective for short term forecasts. The modification is then focused on considering the same "large dataset of macro variables" theorized by other authors, such as Bernanke and Boivin (2003), linking the movements of the whole economy to the short rate. This improvement would though give some issues regarding the structure of the model: having already

[40] Like the paper of Stock and Watson, (2002)

specified two different factors, latent and non-latent, adding another one increases the estimation process. Introducing a similar policy, for example, would mean changing the assumption of orthogonality put in place in the AP model to link macro variables to latent term structure factors. A macro lagged version of the Taylor rule version can be equally used. In the FAVAR specification, this equation contains an orthogonal shocks error term and with the short rate. It should be possible to substitute this error term with the latest one, but it would, again, increase the complexity of the model, increasing the terms to be estimated.

In order to simplify the whole process, it is possible to follow the FAVAR approach, introducing the modified policy rule for which the policy reaction function specification in AP can be modified to:

$$r_t = \delta_0 + \delta_{11}' X_{t-1}^o + \delta_{12}' r_{t-1} \tag{54}$$

This goes to modify the dynamic equation, as it must contain the short rate term, so that:

$$X_t = \mu + \Phi X_{t-1} + \Sigma \varepsilon_t \tag{55}$$

With: $\mathrm{X_t} = (X_t^{o'}, r_t, \dots X_{t-p+1}^{o'}, r_{t-p+1})$. The companion form of the matrix follows the same identical structure to the one in AP. Knowing equation (55) it is also possible to reduce the short rate expression in term of X_t, so it is only dependent on a matrix δ' of *(k* x *kp)* such that:

$$r_t = \delta' X_t \tag{56}$$

The approach followed here is then a mix of the FAVAR approach used by Mönch and the original Ang and Piazzesi described above. The decision of abandoning the use of latent comes from an already abundant macro parameter and the specification given by FAVAR, to which the use of latent variables can only increase complexity with few benefits.

2.3.2 Modified Pricing Kernel

Doing this also ensures that some equations are kept the same as they were in the AP model. Namely, the pricing Kernel is still the one suggested by Duffee (2002). Also, knowing this, the market price of risk is kept the same. It should be noticed that this could also be lagged, but since the term X_t already has some contemporaneous and lagged term inside, it is not a big loss not including it. Nonetheless, some adjustments must be done in terms of equations and matrix. First, the pricing kernel now does not need the term δ_0, as it stood for the first of the short rate in AP, now substituted by the term δ' plus the X_t^o. In this way, the pricing kernel becomes:

$$m_{(t+1)} = e^{\left(-\frac{1}{2}\lambda_t'\lambda_t - \delta' X_t + \lambda_t' \varepsilon_{t+1}\right)} \tag{57}$$

In order to find the prices of the bond, it can be followed the same exact process used for the AP, recursively finding the bond price starting from the canonical assumption that yields are affine in the state variables, allowing the bond to be exponential linear functions of the state vector. This classical definition represents equation (46). Using the same approach, the prices for the bonds are:

$$\bar{A}_{n+1} = \bar{A}_n + \bar{B}_n'(\mu - \Sigma\lambda_0) + \frac{1}{2}\bar{B}_n'\Sigma\Sigma'\bar{B}_n \tag{58}$$

$$\bar{B}_{n+1}' = \bar{B}_n'(\Phi - \Sigma\lambda_1) - \delta' \tag{59}$$

The new values are equal to 0 for A_1 and δ' for $B_{1.}$

2.3.3 Estimation Process: multiple approaches

At last, it comes to the estimation process. Compared to the AP model seen above, it is simplified thanks to the absence of latent factors. The estimation can then be performed in the same two steps, with few modifications or using a single-step Bayesian likelihood approach, like Bernanke et al. (2005) suggest. This last provides better results but as it must be conducted jointly for every parameter, which leads to some issues regarding the restrictions needed. The first step is conducted using a VAR(p) to obtain the first step coefficients, exactly as AP, but the determination of λ_0 and λ_1 is not performed through a GMM given the use of a VAR (p). The reason is quite straightforward, as having to determine each moment for each lag can create confusion and thus increases the number of parameters in the estimation. The procedure follows partially what AP does, looking for the values that minimize the residual squares of errors between the estimated yields and model implied observed ones. This minimization can be done in different ways, but the one that follows uses a simple approach in order to keep the estimation process quick. It starts by setting first the risk premia constant and estimating the yields, using these results as a starting point for a second estimation. Looking for the value that minimize the error considering fixed risk premia, it is possible to get a second estimate for the yields in case of time-varying risk premia. Lastly, it is possible to use the values founded for the vector λ_0 and the matrix λ_1 to perform a last regression, that sets to zero all the parameters that are non-significative. Now comes the differences that make this approach the most feasible for this kind of large dataset: as Sims (1980) reports, using a single step estimation in identifying some necessary restrictions gives an advantage in terms of the estimation itself, but the drawbacks are then defined by the correct use of these restrictions. In case they are wrong, the model is then badly specified and looking up for the mistake is generally extremely time costly. Imposing a simple normalization as done in the AP on the model leads to the use of the two-step estimation procedure, which is then dominated by the use of the OLS and the MLE, given that the dimensionality of the model are restricted, as in this case and opposed to the original FAVAR specification.

Chapter 3: Estimation Results

3.1 Setting and Data

Following the previous chapter, this section provides a modification in the Ang and Piazzesi model through the use of the FAVAR approach. Instead of using a large dataset, the approach here considered is parsimonious, using few data in order to provide a good forecast without the incumbrance of a large dataset and a quick double step estimation process. The first step is performed through the use of a simple OLS estimation, while the second can be done using a recursive minimization algorithm, such as a Kalman filter.

The data used for the estimation follow the Ang and Piazzesi approach, defining a set of variables divided into two groups: the first represents Inflation, while the second Real Activity. These groups are respectively formed from a combination of Purchase Parity Index, Consumer Parity Index, Commodities Average Price, Employment, Unemployment, Average Wage and Industrial Production. This follows the common approach in term structure analysis of considering a limited range of macroeconomic information grouped into two measures. Opting for this approach permits using a limited amount of inputs, contrary to Bernanke et al. (2004) and Mönch (2006), which use a massive information matrix in order to extrapolate eventually useful factors through principal component analysis. This thesis follows then the same set of inputs variable of Ang and Piazzesi grouped in a single matrix, assuming that the time series contained can be explained by a set of common factors plus the monetary policy instrument and an idiosyncratic noise. The span of observation included goes from January 1962 to December 2018, for a 864 observations vector for each variable and comes from the Federal Reserve database. This choice was made in order to include every meaningful macroeconomic event. The macro data are calculated considering the log return on each observation compared to the same of the previous year.

The procedure starts by extracting the principal components from the data matrix, following the procedure outlined by Mönch (2006). In this context, a normal PCA is not possible, as the inputs matrix is dependent not only by the unobservable factors but also by the policy rate, such that:

$$D_t = \Lambda_f F_t + \Lambda_r r_t + e_t$$

Where D_t is the initial data matrix, Λ are matrices of factor loadings, r_t is the short-term rate and F_t the vector of observations on the common factors. Using this assumption, prior to extract the common factors, it is mandatory to assess the effect of the short rate on the matrix D_t. This can be easily done by regressing all the variables in the matrix into r and performing a PCA on the unified residuals' matrix. There are other solutions[41] other than the one used to derive this thesis's model but require more work and the benefits are not that much.

	PC1	PC2	PC3	PC4	PC5	PC6	PC7
Standard Deviation	1.8641	1.6503	0.63968	0.4380	0.34342	0.22854	0.17434
Proportion of Variance	0.4964	0.3891	0.05845	0.0274	0.01685	0.00746	0.00434
Cumulative Variance	0.4964	0.8855	0.94394	0.9714	0.9882	0.99566	1

From the table here above it appears that much of the variance is explained by the first and the second component, which can be used to develop the same one factor model described in Ang and Piazzesi[42]. This can help the estimation procedure that will be conducted in the next step, keeping the model parsimonious and thus avoiding the dimensionality issues in the calculations that can occur with more factors.The extracted two factors can be then used to derive the components used to define the first two variables to be used in the model. Using the one factor model $Z_t^i = Cf_t^{0,i} + \varepsilon_t^i$ as in equation (32), two vector of dimension *(k* x *1)* can be defined to be as factors for Inflation and Real Activity. The correlation between the input data and the macro factors obtained is high among them, with an average correlation of 0.9. The plot for the two Macro Factors resembles the one provided by Ang and Piazzesi and is reported below.

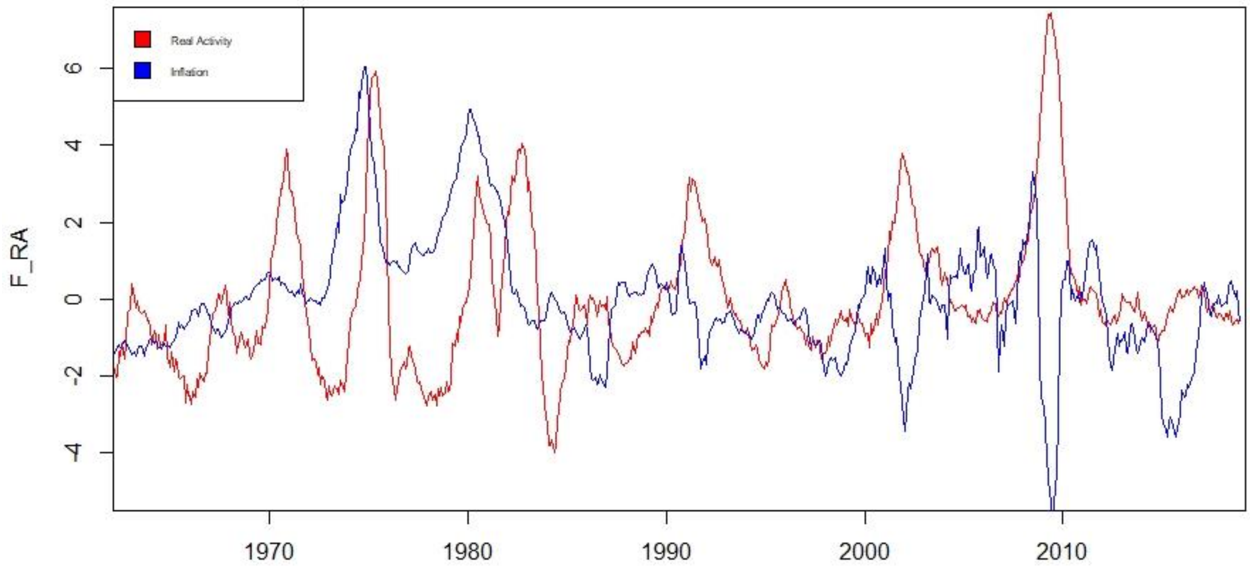


Figure 2. The plot represents a visual presentation of the dynamics of the macro factors. The biggest movements appear to be during crisis, as expected.

[41] Bernanke et al. (2006)
[42] Equation (32)

Nonetheless, principal components factors are here improperly defined as Real activity and Inflation, but this distinction is maintained for more clarity in the references. The reason for this is due to the rotation indeterminacy issues that occur in PCA and makes almost impossible to trace them to the original inputs. However, some information can be deducted by analyzing the correlation between the data and the factors and by running multiple regressions to determine the R^2 for each regression. The tables below report both the correlation and the regressions' results for the factors.

	CPI	PPI	PCOM	EM	UNEM	IP	WAGE
Inflation	0.0682	0.204	0.273	0.931	-0.873	0.894	0.962
Real Activity	-0.911	-0.946	-0.924	-0.00415	-0.331	0.173	-0.331

Figure 3. A table representing the correlation among the different inputs and the two macro factors.

	CPI	PPI	PCOM	EM	UNEM	IP	WAGE
Inflation	0.83076	0.89511	0.85316	0.00002	0.10953	0.02991	0.04922
Real Activity	0.00465	0.04147	0.07464	0.86727	0.76197	0.79923	0.92571

Figure 4. A table representing the regressions on the two factors for each variable. They provide a good point of view in understanding what moves each yield. The distinction in Inflation and Real Activity for Factor 1 and Factor 2 follow Figure 3 and 4.

For what concerns the yields, the 1-month, 6-month, 1 year and 3 year are taken into consideration for the estimation process and for the results. The yields appear to be very correlated among them, with the highest level between the 6-month and 1 year. They become handy in the next parameter estimation in order to minimize the market price of risk's factors.

The last parameter needed to start the estimation is the short rate. Following the literature, it is possible to infer that the lagged short rate can be important to improve the estimation. Following equation (60), it is possible to assume that each of the macroeconomic data series can be driven not only by the Macro Factors extracted, but also by the monetary policy instrument *r*[43]. Incorporating the short rate, which can be proxied using a short maturity yield[44], the matrix of the state variables X_t is then equal to (*3* x *n*).

[43] There is also an idiosyncratic noise that drives the data, as there can be some unobservable phenomena. It is reported in equation (60).
[44] In this thesis it is proxied by using the 1-month yield.

Having defined the state variables, the next step is defining a dynamic for the whole system, which follows equation (55) for its companion form. Instead of using 12 lags, as Ang and Piazzesi do in their model, it is better to use only 4 lags for this variation: including more does not bring any significative improvement and only enlarge the estimation matrix. Obviously, the most significant variable appears to be the short rate, but the significance first two lags of inflation and real activity seem to be equally high, but it steadily decreases until the fourth lag. The VAR (4) is then the preferred approach to define the dynamics and follows equation.

$$X_t = \mu + \Phi X_{t-1} + \cdots + \Phi_4 X_{t-4} + \varepsilon_t$$

The table reports all the coefficient obtained through the VAR (4), which are used to perform a first estimates for the bond prices needed for the next steps. Notice that these are the top left values of a bigger matrix which account also for more lags whose elements are set to zero. The matrix Σ represents the variance covariance matrix of the VAR and is a triangular top left matrix.

	μ		Φ_1			Φ_2	
RA	0.03759	1.13457	-0.04943	-0.03996	0.11688	0.03152	0.05672
INFL	-0.0152	0.01523	1.33841	0.02511	-0.14953	-0.36224	-0.02129
y_t	0.08587	-0.19385	0.12531	1.37287	0.15958	-0.18782	-0.60889
			Φ_3			Φ_4	
RA		-0.07401	0.02138	0.05724	-0.21383	0.01543	-0.07411
INFL		0.14936	0.10452	-0.00211	-0.02201	-0.11353	0.00041
y_t		0.02608	0.05887	0.19907	-0.01477	0.01876	0.03338
			Σ				
RA		0.07361					
INFL		-0.00123	0.08289				
y_t		-0.01865	0.00462	0.13001			

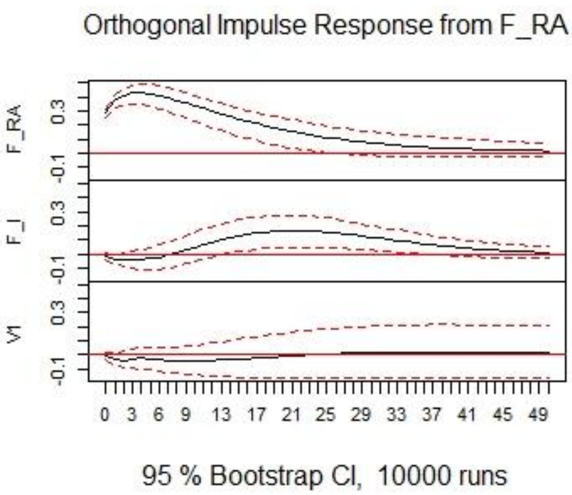
Figure 5. A table representing all the values found for the first three parameters needed for the estimation. Notice the shape of the Σ. Notice that all the matrices hereby are partially reported, as their dimension is (kp x kp), with all the other values set to 0, except for the φ matrix, which contains also a diagonal of 1 elements.

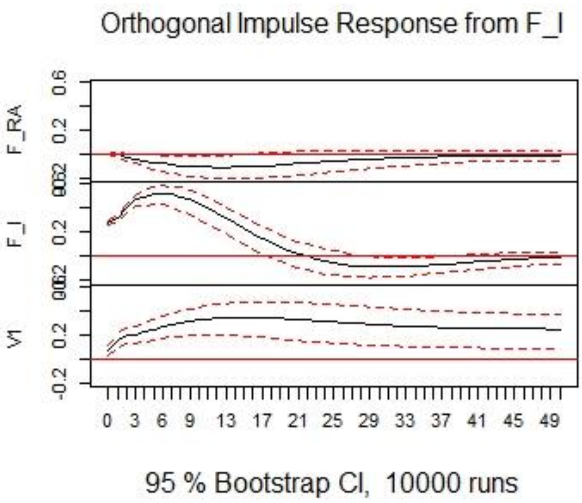
3.2 VAR Analysis

3.2.1 Impulse Response

An impulse response analysis can help describing the reaction of the variables that constitutes the model to exogenous impulses, commonly defined as shocks. These shocks can derive from a multitude of effects, such as a change in the government that creates a different spending policy and tax rates or a shock in productivity determined by unforeseeable events or, at least, not considered in the model. Impulse response considers the reaction of the model’s inputs after a shock, looking for their behavior through time and the persistence of the same shock. Below are reported the results from the impulse response of the VAR. Considering F_RA as the factor for the Real Activity and F_I as the factor for inflation and V1 for the short rate. Time is represented on axis x in month, stopping at 50. The plots that follow report the impulse responses derived from a Cholesky one standard deviation innovation to each variable. They are then grouped in order to be easily understood by the reader. The humped shape that both Inflation and real Activity have in relation to each other is expected, as they are macro variables with a certain degree of correlation among them, especially since the factor derives from a single matrix of multiple data series.

The response from the short rate respect to inflation seems to be the most interesting relation among the ones observed. It seems that the inputs contained in it can explain it better than the other PC factor, Real Activity. Also, even if slightly visible the Inflation shows a change within the period interested by the model, while the Real Activity seems to be completely flat in the impulse response respect to the short rate.





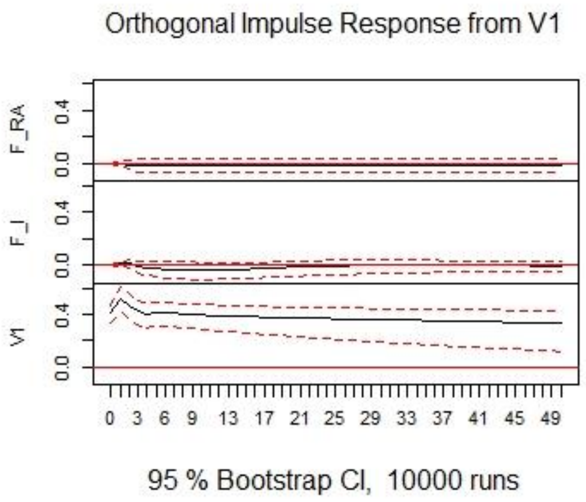


Figure 6. The plots for the Impulse responses of the VAR (4) respect to each variable. The decomposition follows the so-called Orthogonal Impulse Response, which use a Cholesky decomposition. All the calculations have been made using the R package VARS.

3.2.2 Variance Decomposition

The table here presents the analysis on a *i*-step ahead variance decomposition for the model's implied yields following the VAR (4) used for the dynamics. It appears that the lagged short rate explains much of the variations for the short period, with the Inflation factors rapidly increasing as the *i*-step approaches infinite. The factor Real Activity, which encompasses much of the "real" macroeconomic data, such as Industrial Production and Wages, does not provide any real explanation of the yields' variance. The results are calculated after the estimation of all the needed parameters but is reported before in order to present a clear view on the VAR elements.

6-Months Yield			
i-step	Real Activity	Inflation	Short Rate
6	0.02	0.09	0.89
12	0.03	0.13	0.84
36	0.06	0.34	0.57
∞	0.03	0.3	0.57

12-Months Yield			
i-step	Real Activity	Inflation	Short Rate
6	0.01	0.07	0.92
12	0.02	0.11	0.87
36	0.05	0.23	0.72
∞	0.05	0.22	0.73

36-Months Yield			
i-step	Real Activity	Inflation	Short Rate
6	0.0	0.06	0.94
12	0.03	0.1	0.87
36	0.04	0.12	0.84
∞	0.03	0.11	0.86

The Variance decomposition of the VAR respect to the implied yields present some similarities to the ones presented by Ang and Piazzesi (2003). In this case, the Short Rate assumes a bigger role, explaining much of the variance for the longer-term yields, while for the short-term Inflation seems to have an impact on the 6-months yield's variations. Real Activity is somehow poor in explaining the variance of the yields, with a significance close to zero when it comes to forecast the 36-months yield.

3.3 The short rate

A first way to observe the explanatory power of the macro economic factors can be a regression on the short rate considering a primitive matrix without the lagged short rate. The R^2 is 0.31, which means that using only the factors returns some hints on the yields' behavior. Adding lags lead to an even better R^2 of 0.48. Using the matrix X_t in the short rate regression provides a R^2 equal to 0.98. All these results are similar to the ones provided by both Ang and Piazzesi, with little variations. In order to properly set the short rate, it is useful to adopt a reduced form. Specifically, this estimation follows the short rate equation from Mönch, where there are no constant and the matrix δ' present a series of zero and one in order to make it depend only on one lag of the short rate. This is due the massive significance given by the lagged short rate compared to the Macro factors and also to ease the estimation, setting numerous constraints. Thus, the short rate assumes the form of equation (56), which is:

$$r_t = \delta' X_t$$

3.4 Market Price of Risk

Having determined the first three parameters (μ, Φ, Σ) needed to compute the bonds' prices and thus the yields accordingly to equation (58) and equation (59), it is possible to obtain a first estimate. Not knowing the market price of risk, which follows equation (43), the only way to determine it is to start by determining the yield setting both the elements in vector λ_0 and in the matrix λ_1 equal to zero. By doing this it is possible to get a first estimate for the yield. Knowing it, it is then possible to obtain a first value for the vector λ_0 by looking for the values that can minimize the sum of residual squares obtained by the estimated yield minus the model implied one. The formula is then:

$$S = \sum_{t=1}^{T} \sum_{n=1}^{N} (\hat{y}_t^n - y_t^n)^2$$

Assuming values equal to zero for the matrix λ_1 is equal to assume that the risk premia are constant throughout the time series and this first approximation gives the starting values for λ_0 that can be used for the second step. In this second estimation, both the vector and the matrix are estimated letting all the parameters freely, in order to obtain more fitting values. Lastly, as a common practice in the literature, the insignificant elements of the matrix are set to zero and a last estimation is conducted to enhance the values.

This approach is conducted using recursive minimization algorithm and does not involve the GMM used in other models[45]. This is the last step in order to find a good fitting for the estimated yield in comparison to the one used for the regression.

The different yield found can be computed recursively using equation (60), which encompasses the two previous equations in order to set the classic exponential affine form. Following this form, this thesis analyses only a few: the 6 months, 1-year and 3-year yields in order to present the result of this model compared to the observed yield.

[45] Specifically, Ang and Piazzesi use it in order to improve in sample fit. Nonetheless, since the FAVAR approach also considers lags of order higher than one, using the GMM would only lead to computation issues due to the massive amounts of variables.

The market price of risk estimates found are reported in the table below.

$\tilde{\lambda}_0$	$\tilde{\lambda}_1$		
23.672	1.532	-0.931	-2.904
10.041	-2.047	0.342	-0.425
-41.163	-	-	0.003

Figure 7. The table presents the estimates for the market price of risk. The non-significant values are set to zero.

Deriving them allows to calculate risk premia for the macro factors used in the model and the excess returns for the yields used in the model. From the data, it seems that the inflation risk premia have the highest correlation with the yields, in particular with 6-months yield.

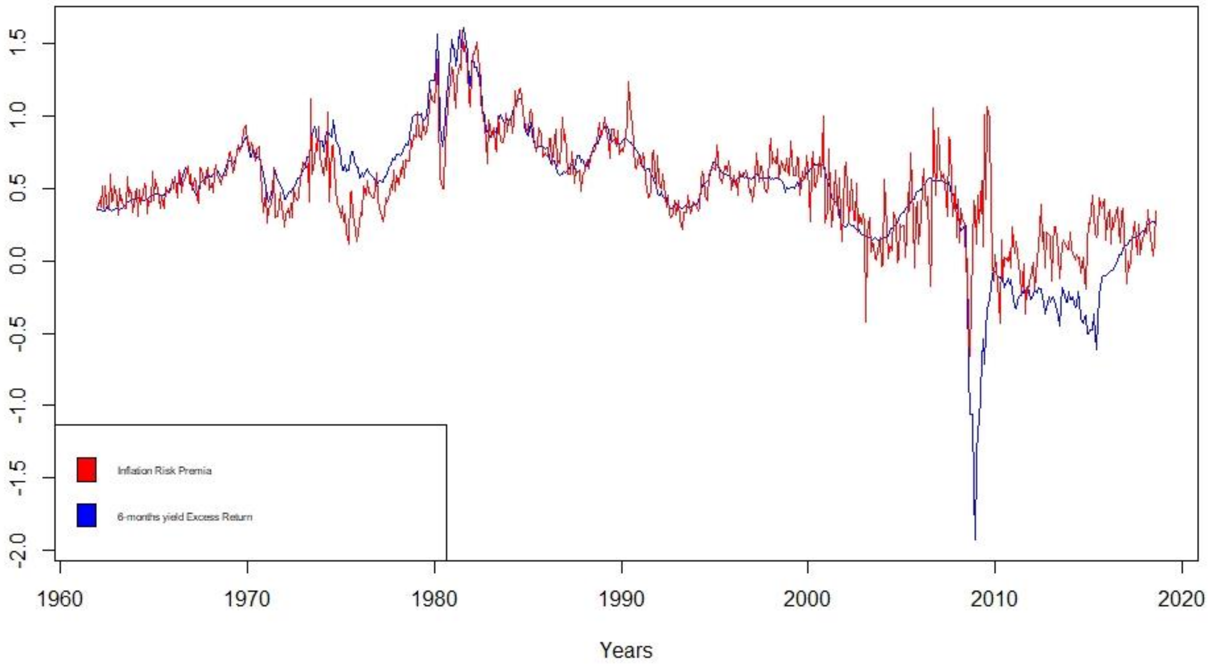


Figure 8. The plot reports the scaled risk premia for the first macro factor and the excess returns for the 6 months yield. there seems to be a high correlation between the risk premia found and the excess returns of the yields

There is a relationship between the excess returns of the yields and the term premia of the macro factors. Usually, when it comes to simple one factor models with constant risk premia and volatility, the expected excess returns for the yield is given by a simple linear relation between the volatility, the risk premia and the maturity of the bond. In this case, it is more difficult to acknowledge the exact relationship, as there are more factors and the risk premia are time varying.

3.5 Yields Estimates

The graphs that appear below present the results from an in-sample forecast along all the data period. The 6 months yield presents a good forecast of the yield's behaviour, with few spikes during financial crisis and a good fit overall. Especially during the post-Volcker inflation, the interest rates are well described by the model, but failing to describe it during the last financial crisis. It is interesting to notice that the model exacerbates the extreme movements, such as the one that followed 2008 crisis, hypothesizing negative yields. This is an interesting result considering the macro factors inputs used and the collapse of the U.S. economy that followed.

For what concerns the overall precision, it seems that the model allows to find good results until a bond maturity equal to 10 years. After that, precision starts declining inevitably. Even if it does not seem a big issue, as the model is calibrated to forecast bond yields with a maximum of a three years maturity, it could still create some issues for investors interested in better long-term results. The simplest way to avoid this is to re-calibrate the model to use more long-term yield to obtain the market price of risk values needed.

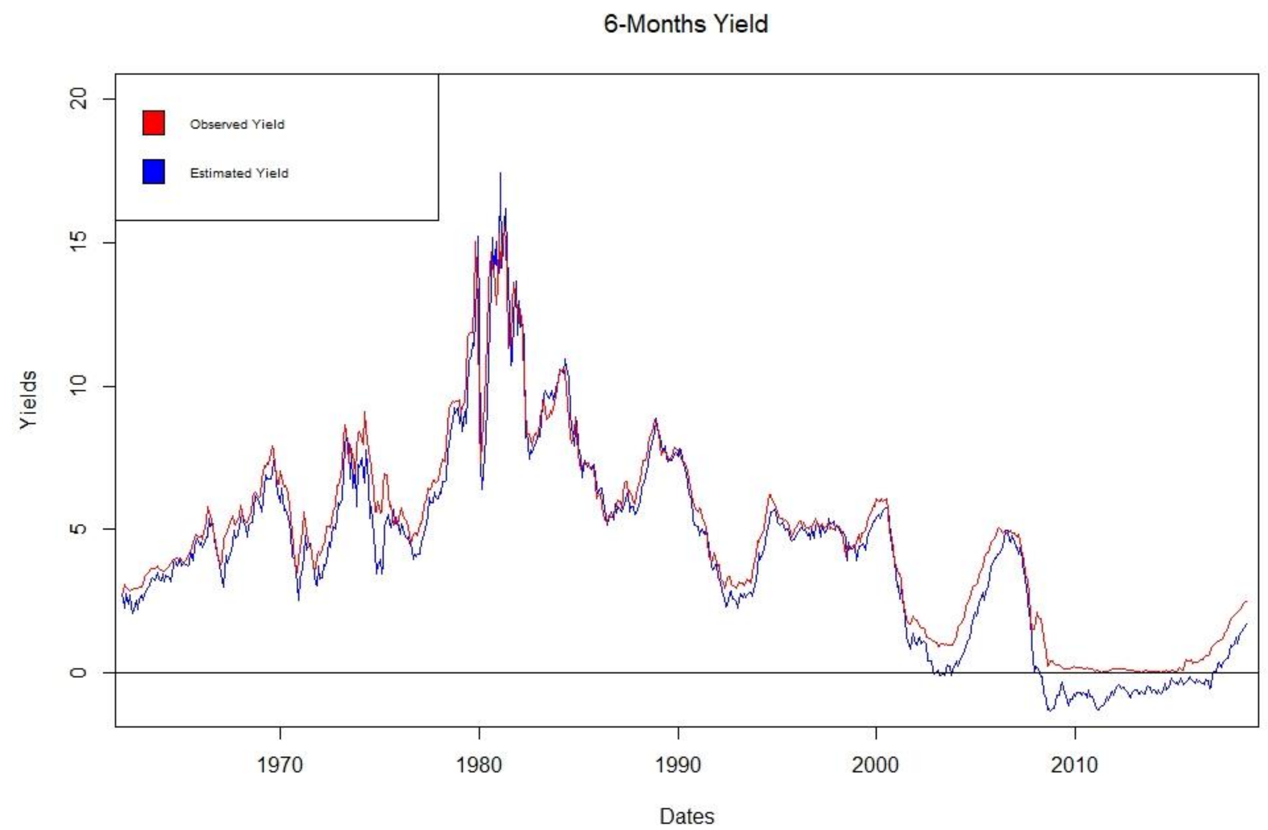


Figure 9. The plot follows the evolution of the 6-months yield modelled using three inputs variables and their lag compared to the real observed yield. The plot is in level and comprises the yield's evolution from January 1962 to December 2018.

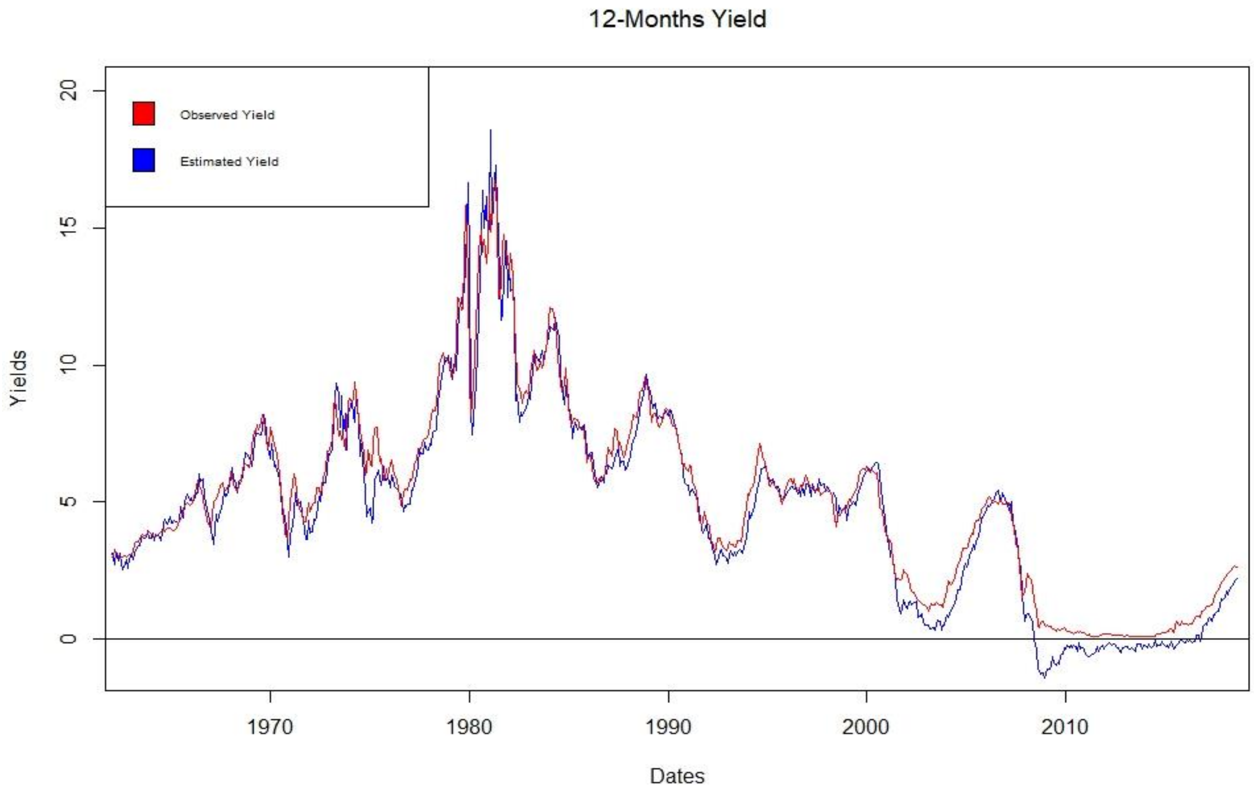


Figure 10. The plot follows the evolution of the 12-months yield modelled using three inputs variables and their lag compared to the real observed yield. The plot is in level and comprises the yield's evolution from January 1962 to December 2018.

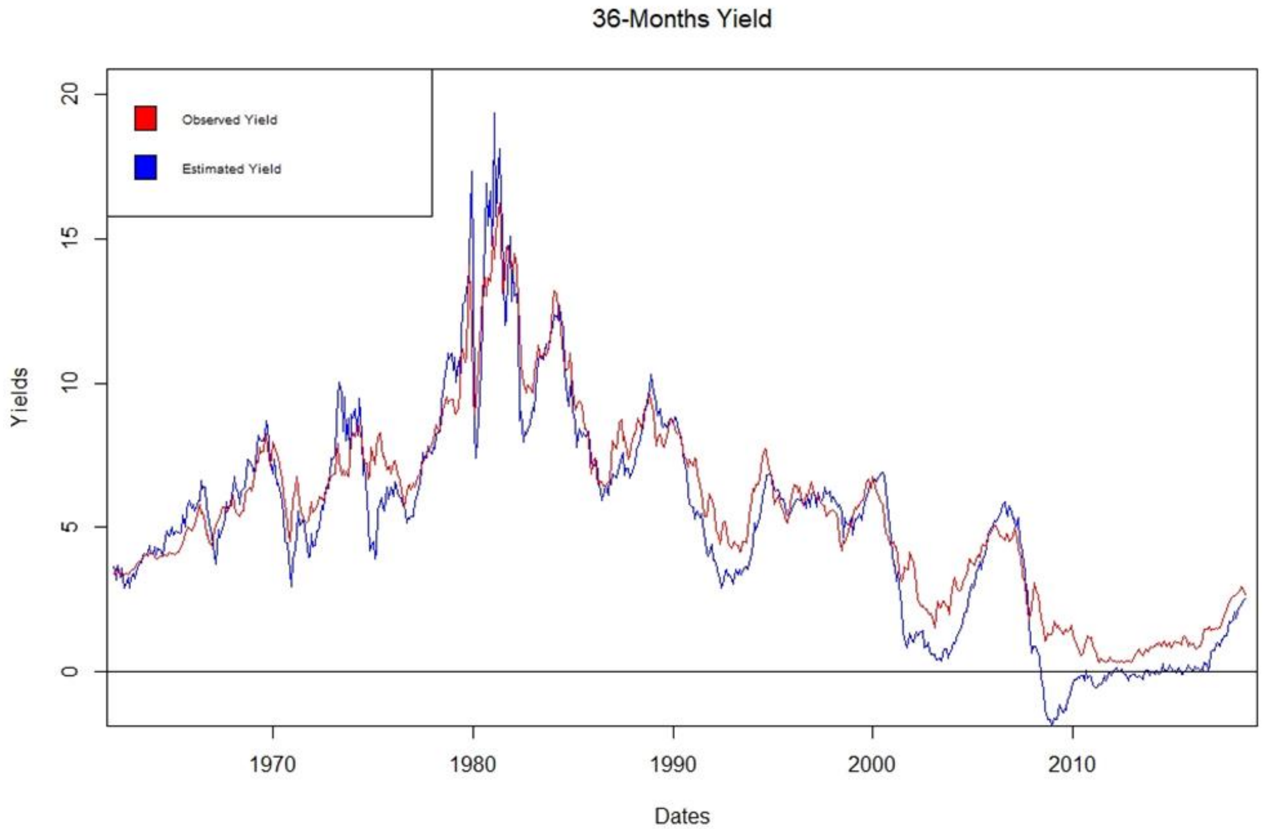


Figure 11. The plot follows the evolution of the 36-months yield modelled using three inputs variables and their lag compared to the real observed yield. The plot is in level and comprises the yield's evolution from January 1962 to December 2018.

Following the plots for the 1 and 3-years yield, the model seems to predict yields' curve inversions, that usually occur during markets' and economic tensions. It also appears that normally the model quickly recovers from these humps, converging to what is the observed yield in 20 months. This is not true for the last crisis, as the plots show that the estimated yields travel almost parallel to the observed one for a period longer than 9 years, starting first signs of convergence during 2017 and 2018.

3.6 Factor Loadings for b_n

In this section are presented the factor loadings found for the coefficient b_n for yields with increasing maturity. By construction, the two macro factors start at zero, while the short rate at one. The coefficients here presented can be also interpreted as the response of the n-month yield to a contemporary shock to each factor. As it can be seen, the short rate tends to decrease throughout the plot, trending for zero, while the two macro factors move positively after a short decrease.

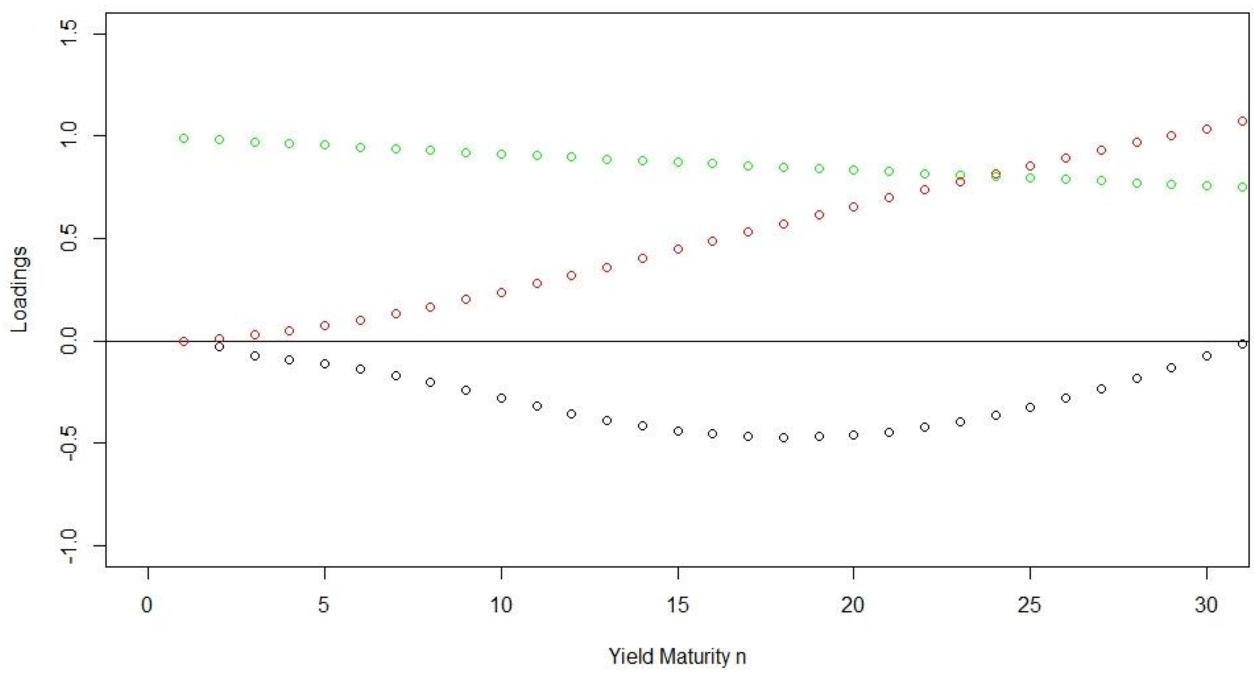


Figure 12. The plot presents the factor loadings for the term b_n. Note that by construction the starting point is set to zero for the macro factors and to one for the short rate

Considering the huge amount of time observations and the scarcity of unused data, performing an out of sample forecast does not provide any interesting results. The reasons are mainly linked to the fact that the model uses all the data from 1962 to 2018 in order to get a good fit, leaving only one year of buffer. Moreover, still trying to perform an OOS forecast does not provide good results due to the financial crisis, which has created a big discrepancy in the model. Nonetheless, since the modeling follows the one determined by Mönch, he strongly proves that avoiding a perfect fit to the data used in the modeling helps the out of sample forecasts.

Conclusions

The model hereby described presents a simple modification of the Ang and Piazzesi mode, using a close approach to the one adopted by Mönch, but with some data and estimation modifications. The data matrix used for the two macro factors only encompasses few data concerning inflation and real activity, available on the FED website. The benefits of adopting this approach comes from its simplicity in estimation and dynamics compared to others. Looking at those that uses a single step estimation procedure, which adopt maximum likelihood in order to jointly obtain all the parameters needed for the forecasts, it is quicker and returns good results. That approach still gives better results, but needs much more complexity with few benefits, needing a series of multiple restrictions and the derivation of the likelihood function.

Moreover, the adoption of a two-step estimation leads to a wide choice of techniques in order to obtain the parameters, spacing from OLS and GMM to more complex approaches. Looking at the results, it appears that the estimated yields resemble very well the behavior of the observed one, with an R_2 of way above 0.9. The forecasts seem to be very accurate considering the small amount of data used to derive them. Even if usually the FAVAR requires a large dataset to work, using less inputs can be less time consuming. Moreover, the data used are all public and can be easily accessed by the Federal Reserve official site. Using a different approach in order to get the unobservable factors would lead to slightly different results, but the approach underlined by both Mönch (2006) and Bernanke et al. (2004) seems to be the most reasonable and with the best time-results ratio.

Bibliography

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

Annaert, J., Claes, A. G. P., De Ceuster, M. J. K., Zhang, H., 2012. *"Estimating the Yield Curve using the Nelson- Siegel approach: a Ridge regression Approach"*. International Review of Economics & Finance, Forthcoming. Available at SSRN: https://ssrn.com/abstract=2054689.

Bolder, D. J. 2001. *"Affine Term-Structure Models: Theory and Implementation."* Bank of Canada: Working Paper 2001-15.

Bolder, D. J., Liu, S., 2007. *"Examining simple joint macroeconomic and Term-Structure models: a practitioner's perspective".* Bank of Canada: Working Paper 2007-49.

Cochrane, J. H., 1998. *"What do the VAR's mean? Measuring the output effects of monetary policy".* Journal of Monetary Economics 41, 277–300.

Cochrane, J. H., Piazzesi, M., 2008. *"Decomposing the Yield Curve"*. Meeting Papers 18, Society for Economic Dynamics.

Cox, J., J. Ingersoll, and S. Ross, 1985, *"An Intertemporal General Equilibrium Model of Asset Prices",* Econometrica, 53, 363–384.

Dai, Q., Singleton, K., 2000. *"Specification analysis of affine term structure models".* Journal of Finance 55, 1943–1978.

Dai, Q., Singleton, K, 2002, *"Expectations Puzzles, Time-Varying Risk Premia, and Affine Models of the Term Structure",* Journal of Financial Economics 63, 415–441.

Dai, Q., Singleton, K, 2003, *"Term Structure Dynamics in Theory and Reality"*, Review of Financial Studies 16, 631–678.

Dewachter, H., Lyrio, M. 2006. *"Macro Factors and the term structure of interest rates".* Journal of Money, Credit, and Banking, Volume 38, No. 1, 119-140.

Dewachter, H., Lyrio, M., Maes, K., 2006. *"A Joint Macroeconomic for the Term Structure of Interest rates and the Macroeconomy"*. Journal of Applied Econometrics 21(4), 439-462.

Duffie, D., Kan, R., 1996. *"A yield-factor model of interest rates".* Mathematical Finance 6, 379–406.

Duffee, G.R. 2002. *"Term Premia and Interest Rate Forecasts in Affine Models."* The Journal of Finance 57: 405–443.

Gürkaynak, R. S., Wright, J. H., 2010. *"Macroeconomics and the Term Structure"*. CEPR Discussion Papers 8018.

Joslin, S., Le, A., 2016. *"Interest Rate Volatility and No-Arbitrage Affine Term Structure Models".* Published on: *"www.marshall.usc.edu".*

Knez, P. K., Litterman, R., Scheinkman, J.A., 1994. *"Explorations into factors explaining money market returns".* Journal of Finance 49, 1861–1882.

Koijen, R. S. J., Lustig, H., Van Nieuwerburgh, S., 2017. *"The Cross Section and Time Series of Stock and Bond Returns"*. Journal of Monetary Economics, vol. 88, 50-69.

Hamilton, J. D. 1989. *"A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle."* Econometrica 57, 357–384.

Hamilton, J.D. 2018. *"Why You Should Never Use the Hodrick-Prescott Filter".* The Review of Economics and Statistics, vol 100(5), pages 831-843.

Hördahl, P., Tristani, O., Vestin, D. 2004. *"A Joint econometric model of macroeconomic and term structure dynamics".* ECB Working Paper Series No. 405.

Ireland, P. N., 2015. *"Monetary Policy, Bond Risk Premia, and the Economy".* Journal of Monetary Economics, Elsevier, vol. 76, 124-140.

Mönch, E., 2005. *"Forecasting the yield curve in a data rich environment. A no arbitrage factor-augmented VAR approach".* ECB Working Paper Series No. 544.

Moskowitz, T. J., Brooks, J., 2017. *"Yield Curve Premia".* Available at: https://ssrn.com/abstract=2956411.

Nelson, C. R., Siegel, A. F., 1987. *"Parsimonious Modeling of Yield Curves".* The Journal of Business, Vol. 60, No. 4, 473-489.

Piazzesi, M., 2001. *"An econometric model of the yield curve with macroeconomic jump effects".* National Bureau of Economic Research, Working paper no. 8246.

Piazzesi, M., 2010, *"Affine Term Structure Models"*, in Y. Ait-Sahalia, and L. Hansen (ed.), Handbook of Financial Econometrics. Chap. 12, 691–766.

Rebonato, R., 2018. *"Bond Pricing and Yield Curve Modeling: A Structural Approach".* Cambridge University Press.

Rogers, L. C. G. 1995. *"Which Model for Term-Structure of Interest Rates should one use?".* Mathematical Finance, IMA Volume 65, 93-116.

Rudebusch, G. D., Swanson, E.T., Wu, T., 2006, *"The Bond Yield 'Conundrum' from a Macro-Finance Perspective"*. Federal Reserve Bank of San Francisco Working Paper 2006-16.

Sargent, T. J., 1979. *"A note on maximum likelihood estimation of the rational expectations model of the term structure".* Journal of Monetary Economics 35, 245–274.

Schüler, Y. S. 2018. *"On the cyclical properties of Hamilton's regression filter".* Deutsche Bundesbank No. 03/2018.

U.S. Bureau of Labor Statistics, Multiple data series from the Federal Reserve official site. https://fred.stlouisfed.org/series/

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

Recap

This thesis analyses the different approaches used to describe the term structure and yield curve dynamics, taking into consideration their advantages and disadvantages. The interest in this kind of models is somehow changeable, as it increases during crisis and decreases during prosperous times, with few exceptions. The starting point of the discipline find its foundations in the model proposed by Vasicek in 1977, which looks at the similarity and the differences between the yields' and the stocks' movements. The fundamental intuition on the model's dynamics led to a vast number of articles and models, using different points of view to solve the difficult forecast of the yields.

Unfortunately, multiple issues arise, as the governments set the interest rates following different purposes. As a rule, Central Banks (CBs) generally tend to maintain price stability as their main goal, but they also tend to adjust it in favour of more specific issues or targets they want to reach. Employment is, for example, one of the main macroeconomic indicators that a CB wants to keep steady. Moreover, during a crisis the CB would follow a strategy in order to bring back the economy to their previous status, as the leading indicators tend to become unpredictable. A good example to start with is the inflation that occurred in the United States between 1970s and the first years of 1980, subdued by the Federal Reserve approach. Setting higher yields helped to restore the situation in few years, averting the risk of an excessive inflation, thus confirming their usefulness. Nonetheless, investors do not know the CBs intentions and usually tends to guess, basing their intuitions on different factors. This is an issue for those who are interested, for example, in building a portfolio that includes bonds of different maturities to hedge the risk.

At this point, term structure models become useful, as they use observable inputs in order to find plausible results for the expected yields. The models presented in this thesis help both investors and Central Banks to provide good forecasts using different methodologies and presenting both advantages and disadvantages,

The core of this work uses a specific model in order to provide a more in-depth analysis, the Ang and Piazzesi (2003). Their approach is different from previous works on the topic, incorporating observables macroeconomic extracted factors in combination with latent ones. Nonetheless the approach needs a two-step estimation which requires multiple constrictions. One way to avoid this issue is to use a similar two-step procedure as the on proposed by Mönch (2005) and use a modified version of the FAVAR approach defined by Bernanke et al. (2005). The thesis provides an analysis on the advantages of the model and the forecasts for the 6, 12 and 36 months implied by the model.

The issues in modelling the yield curve

Bonds are amongst the most valuable instruments used by public and private institutions to obtain liquidity from third party investors. By buying a bond, the holder agrees to lend a pre-determined amount of money to the issuer in exchange of interests that can be paid at multiple dates until maturity, or in a single solution together with the lent capital. The yields used are only U.S. zero discount bonds, because they can be modelled in an easier way compared to the ones that have multiple coupons. The yield curve represents the relationship that links the different interest rates for the same instrument – with the same credit quality - at different maturities. The graphic representation of the term structure is defined as “yield curve” and it is quite useful as it comes to provide a quick understanding of the bonds’ market behaviour. The problem that arises from having a snapshot of the yield curve is that it is not a scalar value but instead a vector quantity. In other words, it is not limited to a single instant in time, but it varies for each *t*, adding the time dimensionality factor to the problem. After the quick introduction to bond pricing, it is now worth to add some notions about other conditions needed for a model to work.

There are different ways to model the term structure. Each of them is designed to observe the problem of the term structure from a different point of view, resulting in multiple advantages and disadvantages. A good classification has been made by Rebonato[46], who differentiates the models considering their structure and follows.

- Statistical models, like the one that will be proposed in the next chapter, rely on their strength in the so-called Vector Auto-Regressive or VAR models, which are extensively used to forecast yield and risk premium estimates. This model has a great predictive power in contrast to the other models here mentioned, but bases the whole analysis to time series data, bringing large error in the forecast, especially for longer maturities. Its flexibility still makes it a good choice to model term structure, in addition to the possibility of easily using the implied impulse response functions and variance decompositions to get an additional understanding of macroeconomic and yields interactions[47]. From a mathematical point of view, they are basically Vector Auto-Regressive models that have their foundations in the AR (1) process, an autoregressive model of order 1[48]:

$$x_{t+1} = \mu + x_t\varphi + \nu\eta_{t+1} \tag{11}$$

In this context, μ represents the intercept for the regression and the φ its slope.

[46] Bond price and yield curve modelling, Rebonato
[47] Cochrane and Piazzesi (2008), Diebold and Rudebusch (2013).
[48] with $\eta_{t+1} \sim N(0,1)$.

- Structural no-arbitrage models. These include the no-arbitrage conditions from their assumptions and explains the three components of the yield curve: expectations, risk premia, and convexity. The most known models from this group are Vasicek and Cox-Ingersoll-Ross. Both use a single factor model to estimate the curve yield, but the last incorporates a square root factor in order to avoid negative interest rates, as pre-crisis academics were not interested in considering the case.
- Snapshot model: such as the well-known Nelson-Siegel. They, as the name suggests, are cross-sectional models that give a glimpse of how the curve behaves in order to interpolate the yields that are unobservable using observable data. It is also worth to notice that this kind of model gives as outputs the discount bonds or – to say it in another way - the inputs for the other model types here described.

The most appreciated approach to the yield curve issue considers using an affine term structure model, which works by linking the term structure of interest rates with a time-invariant linear function made up by a set of variables, which can be latent or observable. This distinction is important, as from the 90s the augmentation with latent variables has brought some advantages to the model. Indeed, even if they cannot be compared to the other variables, they own an intrinsic explanatory capacity. This has led the research to include different numbers of latent variables in affine term structure models, but defining these factors with different names, such as real inflation (Dewachter, Lyrio and Maes, 2005) or real short rate (Pearson and Sun, 1994), even if their data did not include those data. Another issue is related to the number of variables that should be included, but empirical studies (Knez, 1994) have noted that three latent are enough to explain much of the changes. Their labels change between different studies and paper, but recently they were linked to their effect on the curve instead of arbitrary names, specifically: level, slope, and curvature. The starting point for an affine model is the stochastic process that drives the dynamics of the variables involved. The two models analysed in the thesis both use a VAR process with different lags and are calculated in discrete time.

The Ang and Piazzesi Model (2003)

The Ang and Piazzesi model combine the techniques seen in the first paragraphs in order to forecast bond yields, describing joint dynamics for macroeconomic variables and bond yields in a VAR. The model is Gaussian and consists of five variables organized in the observable and unobservable categories. The two observables are built upon macroeconomic data opportunely reorganized, while the three orthogonal latent ones encompass the yield curve's movement that cannot be forecasted by observable data alone. The model uses the no-arbitrage rule in order to set restrictions. Including macroeconomic variables is useful to understand how yields move: from the data used in the paper, they explain up to 85% of yields' movement at short and medium maturities. The VAR approach is also useful for the reasons seen at the beginning of paragraph (1.4) as it is possible to compute IRs and Variance decomposition easily in order to clearly see how the macro shocks impact the term structure. Latent variables impact can be seen in the same way and then compared to macro variables.

Usually, other models, such as the Duffie and Kan (1996), follow a Taylor rule to specify the short rate such that the movements in *t* in the short rate are linked to macroeconomic variables movements at the same time. In a variant of this, a forward-looking version of the same Taylor rule, Clarida (2000) states that the Central Bank reacts both to the expected inflation and output gap[49], including forecast errors in the shock v_t. Ang and Piazzesi present two variations of their idea: a VAR model which encompasses macro factors plus three latent yields to forecast the model implied yields. The dynamics process follows a VAR (12) for the macro derived variables, while the unobservable ones an AR (1). The process estimation is performed in a double step: first they find the short rate and the VAR parameters through the use of an OLS, while the rest are derived using a MLE. This type of process allows to define all the factors needed for the yields' forecasts but is quite demanding in term of calculations. But their idea has led to different models that uses macroeconomic inputs to improve term structure models.

Mönch FAVAR Model

The Mönch model (2006) developed a similar approach, using a complex dataset from which extract different PCs, noticing that until the 10th there is a variance explanatory power equal to 70% of the data. Due to dimensionality issues, again it is worthy to use fewer and the author decided to adopt four. This is difficult to implement in the Ang and Piazzesi model for its numerous restrictions, but it is a starting point. Also, depending on the data chosen the eigenvalue decomposition appears to be more useful compared to the original model's specifications. Considering the AP, it is possible to change the equation that drives the short rate, specifically using the same lagged Taylor rule to define the dynamic of the short rate. This was already specified in the AP paper and it proved to be more effective for short term forecasts. The modification is then focused on considering the same "large dataset of macro variables" theorized by other authors, such as Bernanke and Boivin (2003), linking the movements of the whole economy to the lagged short rate. Nonetheless, this improvement would give some issues regarding the structure of the model: having already specified two different factors, latent

[49] Output gap is the gap between the theorical output an economy can reach and the output that currently has. In some model it is used to name latent.

and non-latent, adding another one increases the estimation process. Introducing a similar policy, for example, would mean changing the assumption of orthogonality put in place in the AP model to link macro variables to latent term structure factors. Moreover, a macro lagged version of the Taylor rule version can be equally used. In the FAVAR specification, this equation contains an orthogonal shocks error term and with the short rate. It should be possible to substitute this error term with the latest one, but it would, again, increase the complexity of the model, increasing the terms to be estimated.

In order to simplify the whole process, it is possible to follow the FAVAR approach, introducing the modified policy rule for which the policy reaction function specification in AP can be modified to:

$$r_t = \delta_0 + \delta_{11}' X_{t-1}^o + \delta_{12}' r_{t-1} \tag{54}$$

This goes to modify the dynamic equation, as it must contain the short rate term, so that:

$$X_t = \mu + \Phi X_{t-1} + \Sigma \varepsilon_t \tag{55}$$

With: $X_t = (X_t^{o'}, r_t, \ldots X_{t-p+1}^{o'}, r_{t-p+1})$. The companion form of the matrix follows the same identical structure to the one in AP. Knowing equation (55) it is also possible to reduce the short rate expression in term of X_t, so it is only dependent on a matrix δ' of *(k* x *kp)* such that:

$$r_t = \delta' X_t \tag{56}$$

The pricing kernel follows the same approach used in Ang and Piazzesi (2003), such that:

$$m_{(t+1)} = e^{\left(-\frac{1}{2}\lambda_t' \lambda_t - \delta' X_t + \lambda_t' \varepsilon_{t+1}\right)} \tag{57}$$

In order to find the prices of the bond, it can be followed the same exact process used for the AP, recursively finding the bond price starting from the canonical assumption that yields are affine in the state variables, allowing the bond to be exponential linear functions of the state vector. Using the same approach, the prices for the bonds are:

$$\bar{A}_{n+1} = \bar{A}_n + \bar{B}_n'(\mu - \Sigma\lambda_0) + \frac{1}{2} \bar{B}_n' \Sigma \Sigma' \bar{B}_n \tag{58}$$

$$\bar{B}_{n+1}' = \bar{B}_n'(\Phi - \Sigma\lambda_1) - \delta' \tag{59}$$

The new values are equal to 0 for A_1 and δ' for B_1.

The approach followed here is then a mix of the FAVAR approach used by Mönch and the original Ang and Piazzesi described above. The decision of abandoning the use of latent comes from an already abundant macro parameter and the specification given by FAVAR, to which the use of latent variables can only increase complexity with few benefits.

Estimation and Results

The data used for the estimation follow the Ang and Piazzesi approach, defining a set of variables divided into two groups: the first represents Inflation, while the second Real Activity. These groups are respectively formed from a combination of Purchase Parity Index, Consumer Parity Index, Commodities Average Price, Employment, Unemployment, Average Wage and Industrial Production. This follows the common approach in term structure analysis of considering a limited range of macroeconomic information grouped into two measures. Opting for this approach permits using a limited amount of inputs, contrary to Bernanke et al. (2004) and Mönch (2006), which use a massive information matrix in order to extrapolate eventually useful factors through principal component analysis. This thesis follows then the same set of inputs variable of Ang and Piazzesi grouped in a single matrix, assuming that the time series contained can be explained by a set of common factors plus the monetary policy instrument and an idiosyncratic noise. The span of observation included goes from January 1962 to December 2018, for a 864 observations vector for each variable and comes from the Federal Reserve database. This choice was made in order to include every meaningful macroeconomic event. The macro data are calculated considering the log return on each observation compared to the same of the previous year.

The procedure starts by extracting the principal components from the data matrix, following the procedure outlined by Mönch (2006). In this context, a normal PCA is not possible, as the inputs matrix is dependent not only by the unobservable factors but also by the policy rate, such that:

$$D_t = \Lambda_f F_t + \Lambda_r r_t + e_t$$

Where D_t is the initial data matrix, Λ are matrices of factor loadings, r_t is the short-term rate and F_t the vector of observations on the common factors. Using this assumption, prior to extract the common factors, it is mandatory to assess the effect of the short rate on the matrix D_t. This can be easily done by regressing all the variables in the matrix into r and performing a PCA on the unified residuals' matrix. For what concerns the yields, the 1-month, 6-month, 1 year and 3 year are taken into consideration for the estimation process and for the results. The yields appear to be very correlated among them, with the highest level between the 6-month and 1 year. They become handy in the next parameter estimation in order to minimize the market price of risk's factors. The table below reports the result from the PCA

	PC1	PC2	PC3	PC4	PC5	PC6	PC7
Standard Deviation	1.8641	1.6503	0.63968	0.4380	0.34342	0.22854	0.17434
Proportion of Variance	0.4964	0.3891	0.05845	0.0274	0.01685	0.00746	0.00434
Cumulative Variance	0.4964	0.8855	0.94394	0.9714	0.9882	0.99566	1

The extracted two factors can be then used to derive the components used to define the first two variables to be used in the model. Using the one factor model $Z_t^i = Cf_t^{0,i} + \varepsilon_t^i$ as in equation (32), two vector of dimension *(k* x *1)* can be defined to be as factors for Inflation and Real Activity. The correlation between the input data and the macro factors obtained is high among them, with an average correlation of 0.9.

Having defined the state variables, the next step is defining a dynamic for the whole system. Instead of using 12 lags, as Ang and Piazzesi do in their model, it is better to use only 4 lags for this variation: including more does not bring any significative improvement and only enlarge the estimation matrix. Obviously, the most significant variable appears to be the short rate, but the significance for the first two lags of inflation and real activity seem to be equally high, but it steadily decreases until the fourth lag. The VAR (4) is then the preferred approach to define the dynamics and follows equation.

$$X_t = \mu + \Phi X_{t-1} + \cdots + \Phi_4 X_{t-4} + \varepsilon_t$$

From the VAR it is possible to obtain some useful results to infer the data structure. VAR impulse response and variance decomposition can give some glimpse on how much the different components influence the estimated yields' curves.

Orthogonal Response

The plots that follow report the impulse responses derived from a Cholesky one standard deviation innovation to each variable. They are then grouped in order to be easily understood by the reader. The humped shape that both Inflation and real Activity have in relation to each other is expected, as they are macro variables with a certain degree of correlation among them, especially since the factor derives from a single matrix of multiple data series.

The response from the short rate respect to inflation seems to be the most interesting relation among the ones observed. It seems that the inputs contained in it can explain it better than the other PC factor, Real Activity. Also, even if slightly visible the Inflation shows a change within the period interested by the model, while the Real Activity seems to be completely flat in the impulse response respect to the short rate.

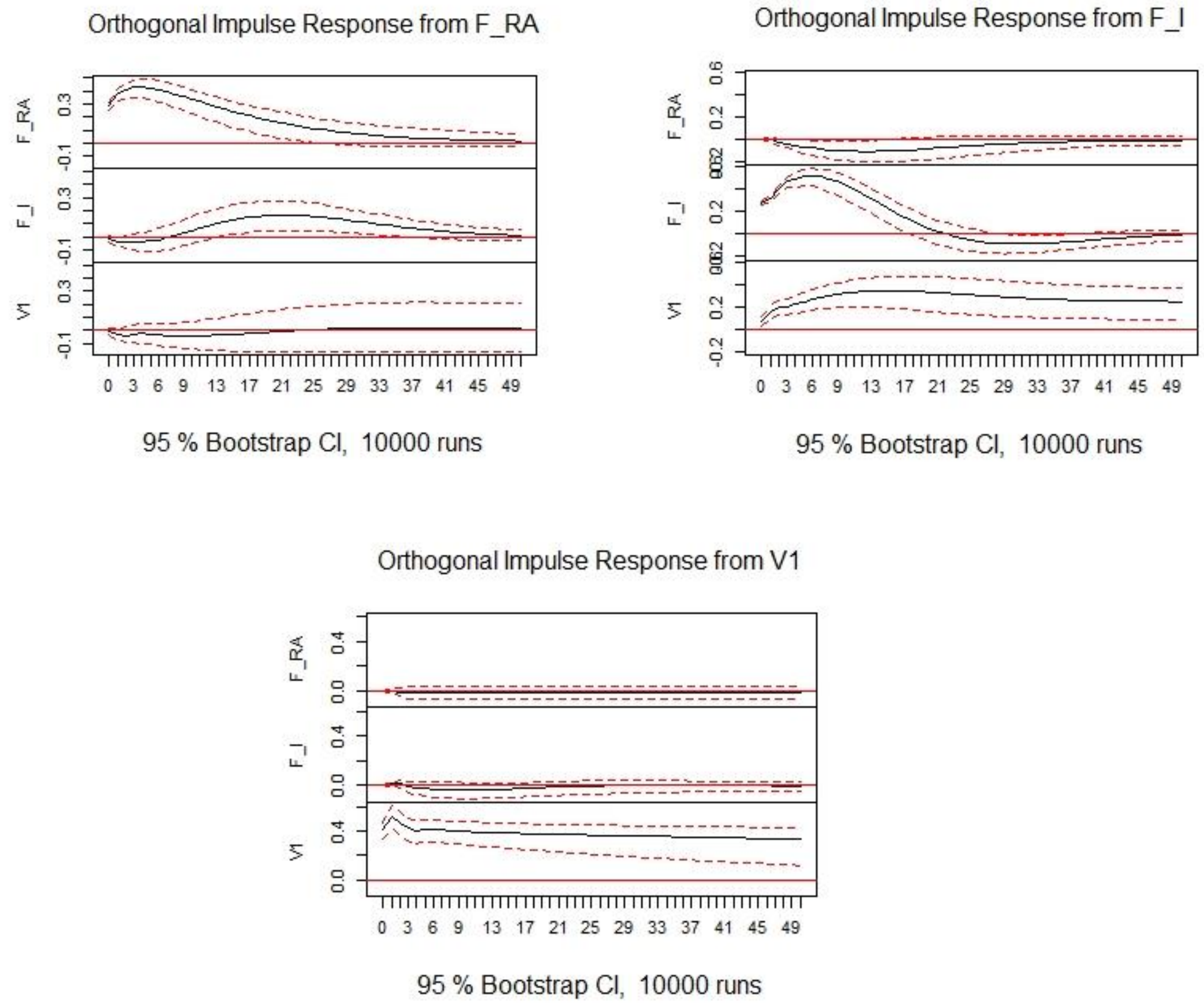


Figure 1. The plots for the Impulse responses of the VAR (4) respect to each variable. The decomposition follows the so-called Orthogonal Impulse Response, which use a Cholesky decomposition. All the calculations have been made using the R package VARS.

3.2.2 Variance Decomposition

The variance decomposition presents the analysis on a *i*-step ahead variance decomposition for the model's implied yields following the VAR (4) used for the dynamics. It appears that the lagged short rate explains much of the variations for the short period, with the Inflation factors rapidly increasing as the *i*-step approaches infinite. The factor Real Activity, which encompasses much of the "real" macroeconomic data, such as Industrial Production and Wages, does not provide any real explanation of the yields' variance. The results are calculated after the estimation of all the needed parameters but is reported before in order to present a clear view on the VAR elements. These results present some similarities to the ones presented by Ang and Piazzesi (2003). In this case, the Short Rate assumes a bigger role, explaining much of the variance for the longer-term yields, while for the short-term Inflation seems to have an impact on the 6-months yield's variations. Real Activity is somehow poor in explaining the variance of the yields, with a significance close to zero when it comes to forecast the 36-months yield.

3.4 Market Price of Risk

Having determined the first three parameters (μ, Φ, Σ) needed to compute the bonds' prices and thus the yields, it is possible to obtain a first estimate. Not knowing the market price of risk, the only way to determine it is to start by determining the yield setting both the elements in vector λ_0 and in the matrix λ_1 equal to zero. By doing this it is possible to get a first estimate for the yield. Knowing it, it is then possible to obtain a first value for the vector λ_0 by looking for the values that can minimize the sum of residual squares obtained by the estimated yield minus the model implied one. The formula is then:

$$S = \sum_{t=1}^{T} \sum_{n=1}^{N} (\hat{y}_t^n - y_t^n)^2$$

Assuming values equal to zero for the matrix λ_1 is equal to assume that the risk premia are constant throughout the time series and this first approximation gives the starting values for λ_0 that can be used for the second step. In this second estimation, both the vector and the matrix are estimated letting all the parameters freely, in order to obtain more fitting values. Lastly, as a common practice in the literature, the insignificant elements of the matrix are set to zero and a last estimation is conducted to enhance the values.

This approach is conducted using recursive minimization algorithm and does not involve the GMM used in other models[50]. This is the last step in order to find a good fitting for the estimated yield in comparison to the one used for the regression. The different yield found can be then computed recursively. Following this form, this thesis analyses only a few: the 6 months, 1-year and 3-year yields in order to present the result of this model compared to the observed yields.

Deriving them allows to calculate risk premia for the macro factors used in the model and the excess returns for the yields used in the model. From the data, it seems that the inflation risk premia have the highest correlation with the yields, in particular with 6-months yield.

[50] Specifically, Ang and Piazzesi use it in order to improve in sample fit. Nonetheless, since the FAVAR approach also considers lags of order higher than one, using the GMM would only lead to computation issues due to the massive amounts of variables.

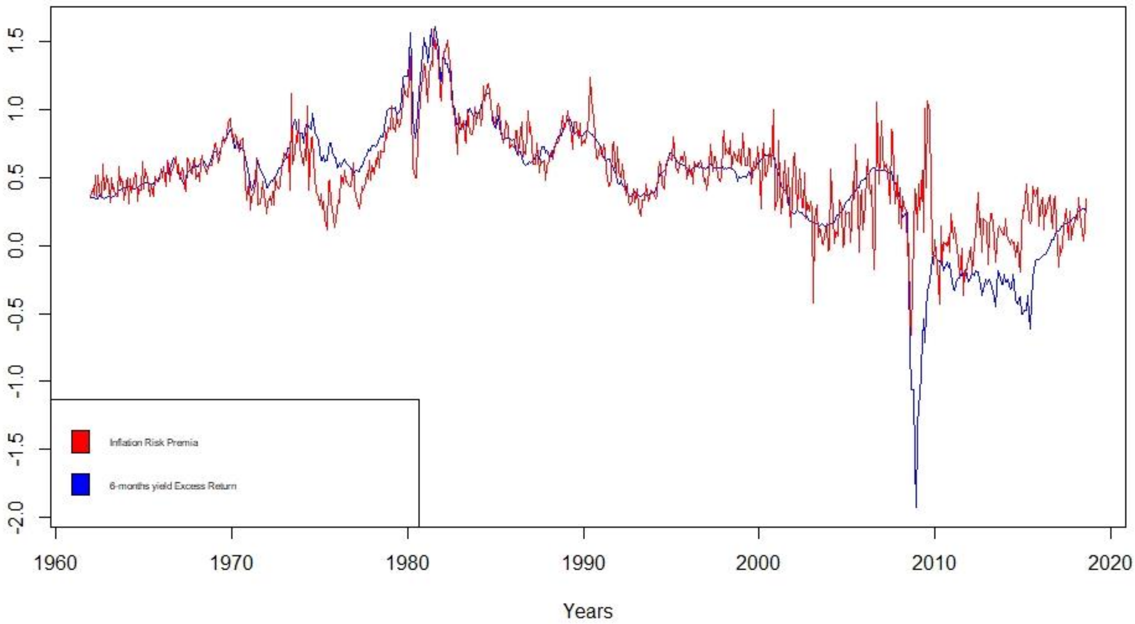


Figure 2. The plot reports the scaled risk premia for the first macro factor and the excess returns for the 6 months yield. there seems to be a high correlation between the risk premia found and the excess returns of the yields

There is a relationship between the excess returns of the yields and the term premia of the macro factors. Usually, when it comes to simple one factor models with constant risk premia and volatility, the expected excess returns for the yield is given by a simple linear relation between the volatility, the risk premia and the maturity of the bond. In this case, it is more difficult to acknowledge the exact relationship, since there are more factors and the risk premia are time varying.

3.5 Yields Estimates

The graphs that appear below present the results from an in-sample forecast along all the data period. The 6 months yield presents a good forecast of the yield's behaviour, with few spikes during financial crisis and a good fit overall. Especially during the post-Volcker inflation, the interest rates are well described by the model, but failing to describe it during the last financial crisis. It is interesting to notice that the model exacerbates the extreme movements, such as the one that followed 2008 crisis, hypothesizing negative yields. This is an interesting result considering the macro factors inputs used and the collapse of the U.S. economy that followed.

For what concerns the overall precision, it seems that the model allows to find good results until a bond maturity equal to 10 years. After that, precision starts declining inevitably. Even if it does not seem a big issue, as the model is calibrated to forecast bond yields with a maximum of a three years maturity, it could still create some issues for investors interested in better long-term results. The simplest way to avoid this is to re-calibrate the model to use more long-term yield to obtain the market price of risk values needed.

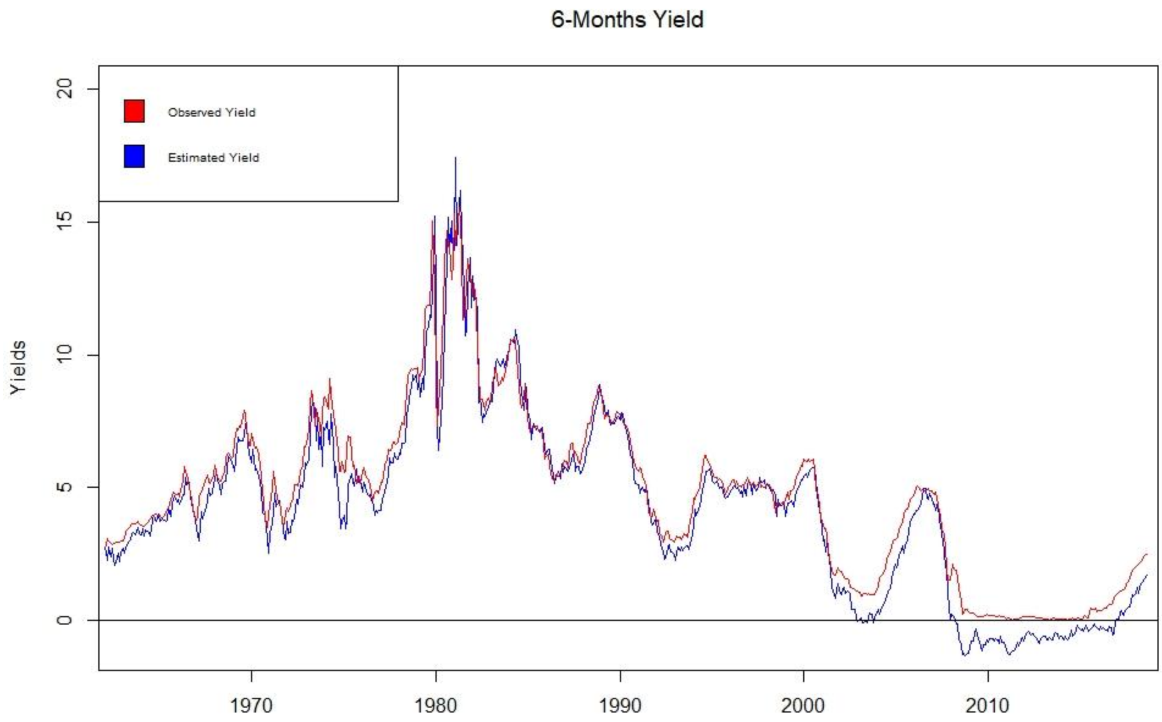


Figure 3. The plot follows the evolution of the 6-months yield modelled using three inputs variables and their lag compared to the real observed yield. The plot is in level and comprises the yield's evolution from January 1962 to December 2018.

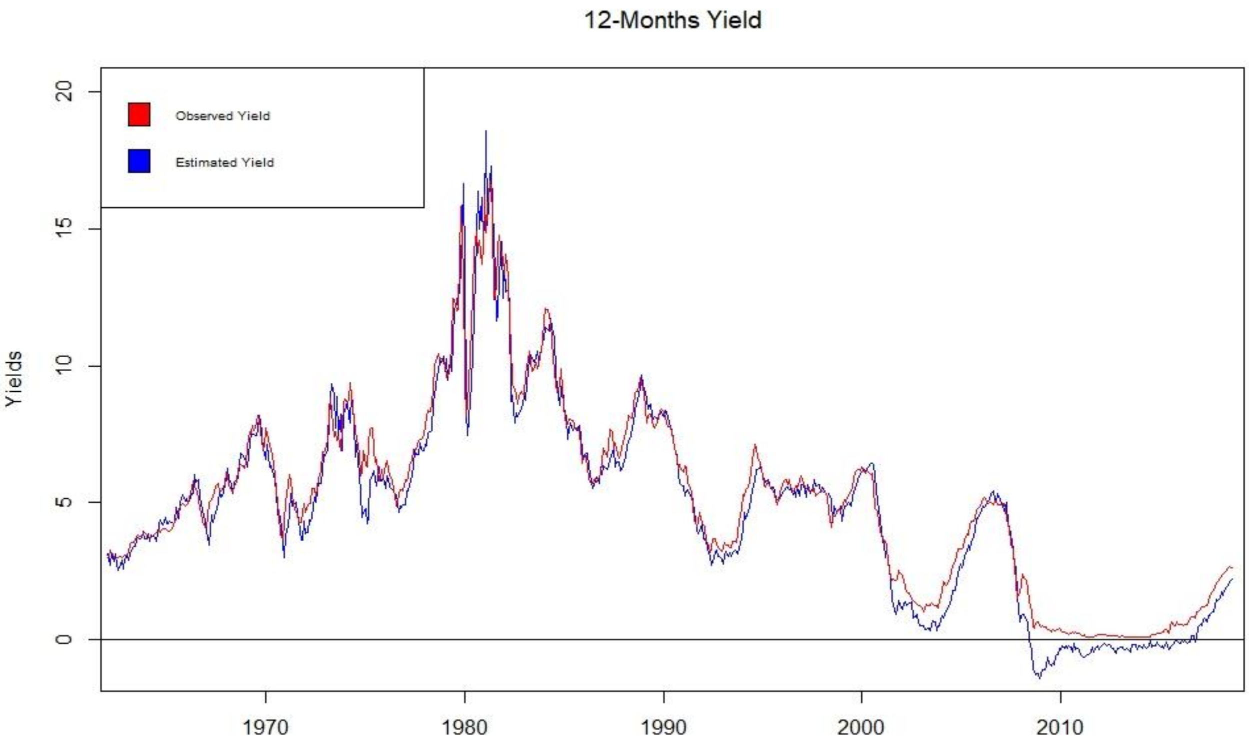


Figure 4. The plot follows the evolution of the 12-months yield modelled using three inputs variables and their lag compared to the real observed yield. The plot is in level and comprises the yield's evolution from January 1962 to December 2018.

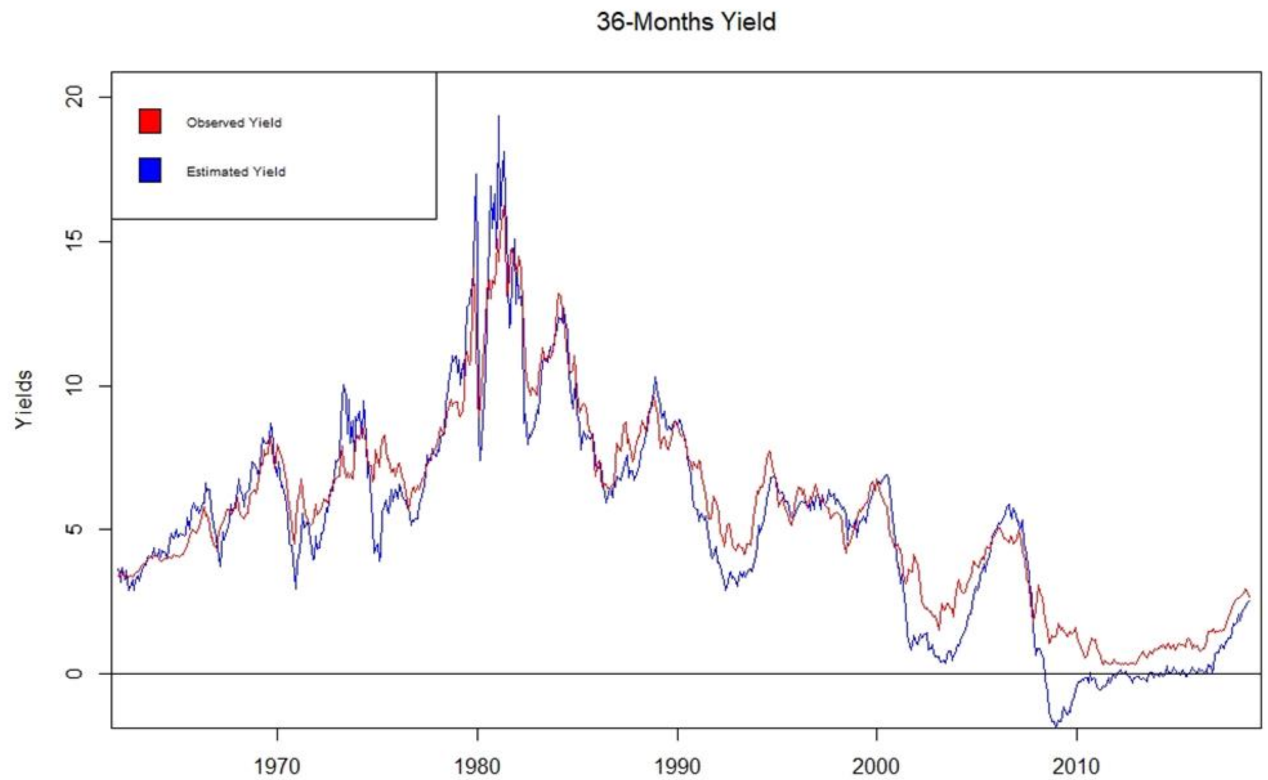


Figure 5. The plot follows the evolution of the 36-months yield modelled using three inputs variables and their lag compared to the real observed yield. The plot is in level and comprises the yield's evolution from January 1962 to December 2018.

Following the plots for the 1 and 3-years yield, the model seems to predict yields' curve inversions, that usually occur during markets' and economic tensions. It also appears that normally the model quickly recovers from these humps, converging to what is the observed yield in 20 months. This is not true for the last crisis, as the plots show that the estimated yields travel almost parallel to the observed one for a period longer than 9 years, starting first signs of convergence during 2017 and 2018.

Conclusions

The model hereby described presents a simple modification of the Ang and Piazzesi mode, using a close approach to the one adopted by Mönch, but with some data and estimation modifications. The data matrix used for the two macro factors only encompasses few data concerning inflation and real activity, available on the FED website. The benefits of adopting this approach comes from its simplicity in estimation and dynamics compared to others. Looking at those that uses a single step estimation procedure, which adopt maximum likelihood in order to jointly obtain all the parameters needed for the forecasts, it is quicker and returns good results. That approach still gives better results, but needs much more complexity with few benefits, needing a series of multiple restrictions and the derivation of the likelihood function.

Moreover, the adoption of a two-step estimation leads to a wide choice of techniques in order to obtain the parameters, spacing from OLS and GMM to more complex approaches. Looking at the results, it appears that the estimated yields resemble very well the behavior of the observed one, with an R_2 of way above 0.9. The forecasts seem to be very accurate considering the small amount of data used to derive them. Even if usually the FAVAR requires a large dataset to work, using less inputs can be less time consuming. Moreover, the data used are all public and can be easily accessed by the Federal Reserve official site. Using a different approach in order to get the unobservable factors would lead to slightly different results, but the approach underlined by both Mönch (2006) and Bernanke et al. (2004) seems to be the most reasonable and with the best time-results ratio.