LUISS

Yield Curve Modelling and Forecasting Applications

Abstract

Forecasting is one of the main reason behind the importance of understanding the yield curve dynamics. In the first part of this thesis we present the theoretical framework of the yield curve modelling and its fundamental features. In the second part we show the applicability limits of what discussed in the theoretical part and in particular, we investigate three quite unrelated problems. The first analysis is an application of the Vasicek Model under different settings, to fit the Italian yield curve. We find that the best results are obtained by defining the market price of risk as an affine function of the short rate.

The second study is a predictability analysis of the Italian 10-year excess returns. We test as regressors the Cochrane-Piazzesi (CP) single factor and the output gap. In our findings, considering the output gap improves the result of the CP single factor model. However we show that in the last five years, the out put gap explanatory power has sharply decreased, pointing out the evolution of the statistical long-term regularities between business cycle and investors' behaviours.

Finally, the last analysis concerns the recession forecasting. We consider the recession periods defined by the National Bureau of Economic Research (NBER), and we test how well a determined set of regressors, composed of the yield curve slope and macroeconomic variables, can signal a looming recession 10-months ahead. In particular we compare the results obtained from two machine learning algorithms: the logistic regression and the support vector machine. In our findings, the former returns more conservative results and hence it is preferred.

Supervisor:	**Candidate:**
Prof. P. Santucci De Magistris	Nicolò Martino
Co-Supervisor:	**Snr:**
Prof. Tommaso Proietti	708631

Contents

1 Introduction

Understanding the main drivers of bond yields has become the focus of serveral studies in the last decade, producing an extensive research literature. This has evolved mainly in continuous time, around stochastic calculus regime and partial differential equations. Affine term strucure models have brought an important contribute to the study of bond yields movements. These particular class of models have emerged thanks to their unparalleled tractability, although as we will discuss in later sections, tractability is paid with restrictive assumptions.
The informations embedded in the yield curve (YC) concern a wide range of different entities. They can be relevant for several reasons; one of the most important is forecasting. If we momentarily do not consider investors' risk aversion, the long term yields are expected values of future short term yields. This means that the current YC structure, contains information about the expected future path of the economy. YC forecasts are crucial for investment decisions of firms as hedge or pension funds, for consumers' saving choices and for policy decision [8]. A second reason for studying the YC dynamics is the monetary policy. Indeed, central bankers extract the market expectations implied in the term structure of interest rates to make optimal monetary policy decisions. In particular, in recent years, with the introduction of unconventional monetary policies as forward guidance and quantitative easing (QE), clear communication between the Central Bank and the market participants is fundamental and the YC movements must be carefully interpreted. Finally, the YC is also relevant for derivative pricing and hedging, and government debt policy decisions.

Acknowledging the YC importance, in the first part of this thesis we examine the theoretical framework of the YC modelling and its fundamental features. In the second part, we investigate three quite unrelated problems, showing the empirical results and the consequent limits of the theory framework.

The first analysis is an empirical application of the Vasicek model (VM) to fit the observed Italian YC. The VM is unique for its tractability and for the intuitive understanding of the YC movements it provides. According to this simple affine model, the YC dynamics are described in terms of expectations and risk premium, neglecting other important features as convexity, liquidity risk and market segmentation. The model parameters are estimeted through the Maximum Likelihood Estimation (MLE) considering the Italy 3-month Bond Yield as the short rate. In our analysis, we build three different VM settings in which we propose three specifications for the market price of risk. We find that, by assuming a constant risk premium, the model appropriately fit the average observed YC but it cannot capture the single YC observed in each month. This model shortcoming is substantially reduced if we express the market price of risk as an affine function of the short rate. Moreover, we extract the model market price of risk implied in the observed values and show its evolution over time. The path of the estimated market price of risk is consistent with the economic and political instabilities that affected Italy in the period considered.

In the second study, we investigate the predictability of the Italian excess returns. We build a database whose columns contain the monthly historical yield for five specific maturities: 2, 4, 6, 8, and 10 years. From these data, we compute the forward term structure and the realized excess returns. Since we consider the excess returns, the inflation and the level of interest rates are net out, so we can focus directly on real risk premia in the nominal term structure [11]. In particular, we examine the relationship between financial market-based variables as the Cochrane-Piazzesi single factor[1], business cycle variables as the output gap and the 10-year Italian excess returns. The output gap is measured by the deviations of the log of industrial production from a trend that incorporates both a linear and a quadratic component. Our objective is understanding whether the significant relations found in Cochrane and Piazzesi (2005) and in Cooper and Priestley (2009) still apply. We find that both the CP and the output gap are significant predictors (5% significant level)[2]. Combining these two variables we achieve an R^2 as high as 47%, with an 1.2% improvement compared to the CP single factor model. Furthermore, we show that the output gap explanatory power has sharply decreased since 2015. This provides further evidence on the evolution of the long-term statistical regularities that used to relate business cycle and investors' behaviours. As also suggested in Rebonato and Ronzani (2020)[7], we believe that the main drivers of this evolution have been the unconventional monetary policies, as QE, adopted by the European Central Banks in response to the sovereign debt crisis.

Finally, the last analysis regards recession forecasting. A significant innovation in the forecasting recession literature came from machine learning (ML) methods. On this aspect, pioneering studies were conducted by Ng (2014) and Holopainen and Sarlin (2017). We continue this line of research and take inspiration from the approach used in Puglia and Tucker (2020) " Machine Learning, the Treasury Yield Curve and Recession Forecasting." In particular, we consider the recession period in the United States (U.S.) defined by the National Bureau of Economic Research (NBER). As in Puglia and Tucker (2020), we build a recession indicator taking value "1" if any of the following ten months falls within a recession period and "0" otherwise. We identify a set of regressors composed of the YC slope and macroeconomics-based variables and test them using two ML methods: the logistic regression (LR) and the support vector machine(SVM). We test the SVM for three different values (small, medium and large) of the regularization parameter ζ, employing also a polynomial kernel to test if accounting for non-separable data is more efficient than the classic linear case. Overall, results show that the linear SVM performs better for each ζ value. We further perform a feature importance analysis on both LR and linear SVM models, discovering that the treasury YC slope is the most important feature for the LR and ranks in the third place for the linear SVM. Despite the various SVM setting tested, we consider the LR the most reliable. Indeed, this model is both conservative and efficient since it returns few false positive values and, at the same time, a large number of true positive classified observations. In the final part of the study,

[1]Factor estimated in Cochrane and Piazzesi (2005)

[2]The CP significant level is less than 0.1%.

we investigate when the model begins to capture the recession signal. We find encouraging results, as the possibility to capture the signal early in time, but, depending on the period considered, the model results may be limited in terms of economic interpretability.

We believe that this work will give the reader a better understanding of the YC movements and the potential use of YC embedded informations. In each analysis, we focus on bringing a novel contribution respect to the existing literature. The first two analysis are based on the Italian framework. In particular, we show how we can set a widely known model as the VM to significantly keep track of the political history of the country. We further use Cochrane and Piazzesi (2005) and Cooper and Priestley (2009)'s approach to study the Italian excess returns in an extended sample period, finding evidence on the weakening of the statistical regularities between business cycle and investor's behavior hypothesized in Rebonato(2020)[7]. In the third analysis, we follow Puglia and Tucker (2020) intuition on the construction of the dataset. However, we focus on comparing the performance of two specific ML methods; Our interest is to understand if the inner flexibility due to the geometrical approach of SVM can return a better result compared to a classic LR statistical approach.

The thesis is structured as follows. In Section2, we explain the theoretical framework of the YC modelling and its fundamental features. We discuss the different kinds of YCs and the possible shapes they can assume. We then proceed by examining the main factors affecting the YC dynamics, namely the expectations, the risk premium and the convexity. Next, we make a crucial distinction between the $\mathbb{P}$ and $\mathbb{Q}$ probability measure and introduce the affine class of models, analysing three widely used models in the economic literature. In Section3, we state our research objective. In the last part, Section4, we develop our three analyses regarding the applicability potential and the limits of the YC theoretical framework. We present the YC modelling and the predictability of the excess returns in the Italian case, and the US recession forecasting model. Finally, we give our conclusive remarks in Section5.

2 Yield Curve Modelling, Theoretical Framework

2.1 Understanding the Yield Curve

A government bond is a debt instrument issued by a government to support public spending and to refinance the existing debt. Bonds can be divided in two macro categories; the regular bonds and zero-coupon bonds (ZCB). The latter, unlike typical bonds does not pay any periodical interest rate during the life of the bond but only the payment of the principal at maturity. In practice, any coupon paying bond can be considered as composed of several ZCB bonds so that each coupon represents a ZCB with a specific maturity date.

It is possible to express the theoretical price of a bond as the present value of all the cash flows that the bondholder will receive. The discount rates used to compute the cash flows present value are defined as *spot rates*. Sometimes, for simplicity bond traders use a single discount factor for all the bond cash flows. It is called *yield to maturity* the single factor that associates the cash flows present value to the bond market price.
Formally, the term structure of interest rate (also known as the *yield curve*) may be defined as a function of a continuous time parameter, which associates an interest rate to a maturity.[20]The yield curve can refer to the par curve, the spot curve or the forward curve. [21]

- The par curve gives the YTM for regular bonds at each maturity. The name "par" derives from the fact that when the YTM is equals to the bond's coupon rate, then the bond is said to be issued at par. From the par curve is possible to extract the coupon rate that a bond with a given maturity must pay to sell at par today. The par yield curve is used to compute the nominal spreads, which is the difference between the YTM of two bonds with the same maturity; typically a risky bond (i.e. a corporate bond) and a risk-free bond (i.e a Treasury). From the par curve is possible to derive the spot curve and the forward curve.

- The spot curve gives a spot rate used to discount a single cash flow at a given maturity. So, instead of discounting all the bond cash flows at the same rate (YTM), each cupon payment is discounted at the spot rate for the corresponding maturity. As we discussed earlier, each coupon can be rapresented as a ZCB with a specific maturity date. For ZCBs the YTM is equal to the spot rate, so the spot curve can also be intepreted as showing the YTM for zero coupon bonds at each maturity. Pricing bond using spot curve or par curve is equivalent (otherwise there would be an arbitrage opportunity).

- The forward curve is derived from the spot curve and similarly it discount a single cash flow at each maturity. The main difference is that the payment is not discounted back to today but to a period at some point in the future (generally, 6 months ahead). Then, the one-year forward rate is the rate used to discount a single payment one year from today, back to six months from today. In discrete time a forward rate can be expressed as $f(0, t, T)$, indicating the forward rate defined today, and used to discount the time T

payment back to time t. When we will describe how expectations affect the yield curve dynamics, it will be necessary to introduce the forward rate in the continuous time. In particular, we define the instantaneous forward rate as the logarithmic derivative of the price of a zero coupon bond expiring at time T: $f_t^T = -\frac{\partial log P_t^T}{\partial T}$.

In this research, the term "yield curve" (YC) will be used to refer to the spot yield curve for Italian bonds.

2.1.1 Yield Curve Shapes

Several different factors can influence the shape of the YC and generally, the curve can follow four type of patterns:

1. The *upward sloping* curve is the most common since long term rate investments are considered riskier than short term ones, thus demanding an higher risk premium.

2. An *inverted* YC is formed when the short term rates are higher than yields with longer maturities, which decline constantly.

3. A *flat* curve has approximately equal yields both on the short and long part of the curve.

4. Finally, a *humped* YC is formed when midium-term yields are above the short term yields and the rates on the long-term part of the curve decline to level below those for the short term and then level out.[11]

If we assume an investment-grade rating, both in the USA and Europe, the short-term yields are highly correlated to the fund rate and the deposit facility rate set by the respective Central Bank (CB). Figure1 shows the high correlation between the 1-year Treasury Constant Maturity Rate and the Effective Federal Funds Rate.

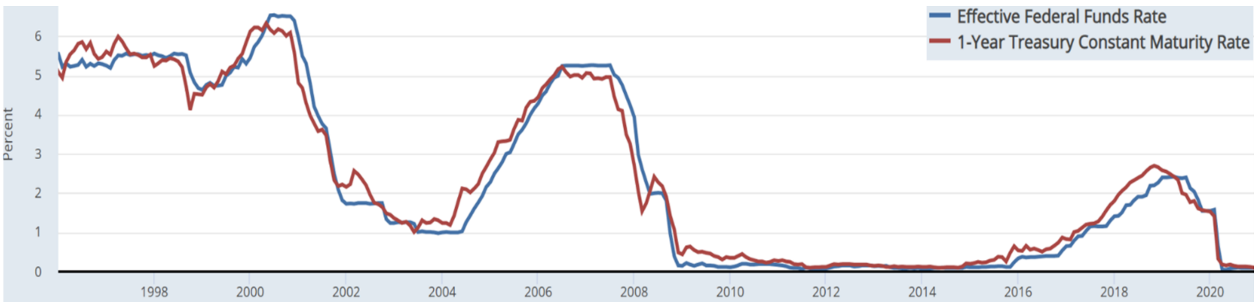


Figure 1: Evolution of the 1-year Treasury Bill and the Effective Federal Funds Rate over time. The two path are evidently highly correlated, showing the CB strong influence on short term yield expectations. Data source: FRED

Medium- and long-term yield should reflect the domestic economic growth prospect. Indeed, according to the neoclassical Solow growth model and assuming a constant saving rate, a slower economic growth will yield a lower marginal product of capital, and, ultimately, lower return from all securities (stocks and bonds). In this context, variations in the real rate ultimately reflects different market expectations on long-term economic growth[11]. The rates observed in the YC are nominal, so expected inflation plays an important role in determining the medium-long part of the curve. When forming their expectations regarding inflation and economic growth, investors strongly consider the CB forward guidance. This is a tool used by the CB to foster price stability. Through the forward guidance the CB sends a clear message to the public regarding the state of the economy and the future course of monetary policy. However, it is worth noticing that even if the investors consider the CB forward guidance, their trust and the correct interpretation of the message is not guaranteed. In fact, the CB credibility and clarity are essential in forming market expectations. There have been cases where the lack of only one of these two characteristics has caused panic and market distress.

Due to the numerous variables involved, the correct interpretation of the YC is a complex task. As an example, Figure2 depicts the Italian bond YCs on the 2-January in three different years: 2019 (green), 2020 (violet) and 2021 (orange).

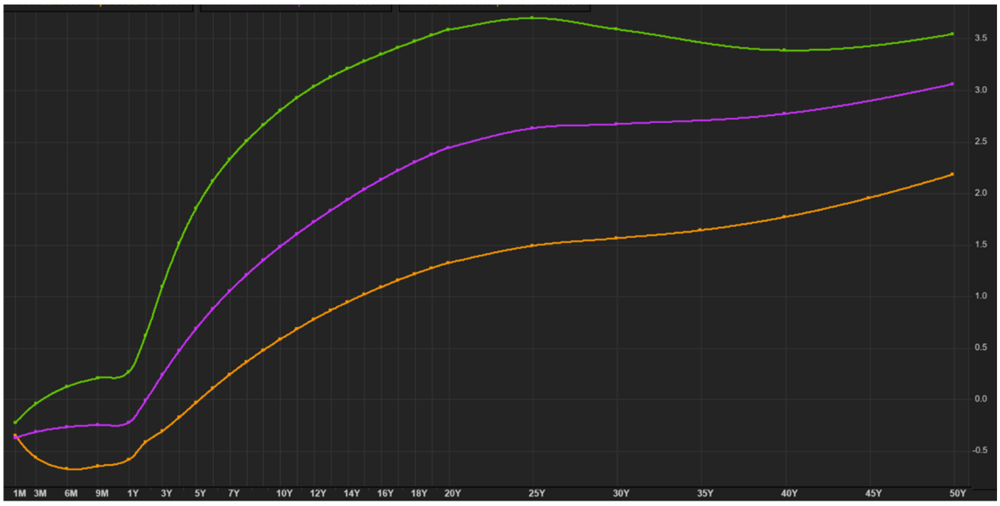


Figure 2: Italian Government Bond Zero Curve, 2/01/2019, 2/01/2020, 2/01/2021. Data from Thomson Reuters

Overall, it is possible to observe how the general curve level has decreased since 2019. A possible interpretation of the downward parallel movement is the decrease in the demanded risk premium from investors, who now consider Italy a safer country to invest in, than three years ago. But if this was the case, then the spread between the 10-year BTP and the 10-year German BUND should have clearly shown a decreasing trend over time. Looking at the data this does not match our scenario since the spread has been high volatile in the last three years. The risk premium explanation does not find any support from the rating agencies as well. In the last years, the Italian credit rating evaluation has been fixed at the base of the investment Grade class, with a slight variation in the outlook. A more convincing explanation could be derived by considering the decreasing of interest rates as a global phenomenon. In the last decade, European countries as well as USA and Japan experienced a downtrend in longer maturity yields. It could be argued that this was due to the aftereffects of the expansions monetary policies adopted to stimulate the economy in the recovery's early years[3]. In Europe, the ECB has set for the first time a negative deposit rate of -0.1% in 2014, which has dropped to -0.5% since 2019. Moreover, unconventional monetary policies such as quantitative easing, has influenced the long part of the YC, lowering mid- and long-term interest rates. In Figure2 is worth noticing the slight parabolic shape formed in the January 2021 YC for maturities less than 1 year. This means the market is expecting the rates to fall in a short period of time (6 months) and grow again afterwards. Moreover, it could be argued that given the closure of most economic activities due to the global pandemic, investors have a pessimistic view on the inflation growth for the next year, while they could expect it to rise again after the virus will be under control thanks to the vaccinations.

2.1.2 Inverted Yield Curve as Recession Indicator

The YC slope can be considered a solid proxy for a country's state of economy. Conventionally, an inverted yield curve has been considered a recession indicator by the majority of economic literature. In the United States (US) the last inversion happened in 2020. It began on February 14th when the yield on the one-month bill rose to 1.60% while the ten-year note fell to 1.59%. The increasing concerns of the investors due to the possible consequences of the global pandemic were the main drives of the inversion. Along with the increase of COVID-19 cases the inversion steadily worsened, the demand of long-term treasuries rose, and yields hit new records low [23]. By March 9th, the 10-year note yield was 0.54% while the 1-month bill was 0.57%. The National Bureau of Economic Research (NBER) defines the recession as "the period between a peak of economic activity and its subsequent trough, or lowest point". According to NBER chronology the most recent peak occurred in February 2020 and it ended a record long lasting expansion of eleven years. In this case, the inversion has not been a clear predictive signal of the looming recession since it happened almost coincidently, but it still shows the existence of a significant relation. However, the Treasury YC has also inverted in August 2019 and slightly in March 2019 and December 2018. These inversions are due to various reasons such as the growing investors' concern for the President Donald Trump trading war and the unexpected Fed forward guidance. Then, it could be argued that the slope signalled an uncertain condition of the US economy presaging a looming distress period which actually realized with the pandemic outbreak.

The Treasury YC inverted before the recessions of 1970,1973,1980,1991 and 2001.[6] Regarding the 2008 financial crisis, the YC inversion occurred two-years earlier in December 2005 and the inversion lasted until June 2007. The monetary policy employed by the CB during periods of "exuberance" and distress followed an expected screenplay until the crisis of 2008. In fact, during a period of economic distress the CB would have engineered a steep YC by cutting rates at the short end. On the other hand, in periods of "exuberance" the

[3] After the Sovereign debt crisis.

CB would have controlled inflationary pressures by hiking short-terms rate, thus creating a inverted YC. Given this framework, investors foreseeing a looming recession also expects a cut in the rates by the CB in a short period of time. Therefore, investor would prefer to buy bonds with longer term maturities (pushing their yield downwards) in order to avoid the reinvestment risk. Consequently, the demand for shorter maturities bonds would drop (setting the yield at higher level) thus causing the formation of a flat or inverted YC. Nowadays these mechanisms are not unshakable anymore, and the CB monetary policy tools have widened. The link between the slope and business cycle has weakened since the introduction of unconventional monetary policies (such as quantitative easing). Indeed, the CB stimulates the economy buying long-maturity assets class, causing the YC to become flatter and not steeper during distress periods. Therefore, the long-term regularities existing between business cycle, investor behaviour, and the YC slope may no longer apply [7]. In Section4, in two different analyses we introduce further evidence on how these regularities have actually changed but we also show that the YC slope is still a very important variable in determining variations in the business cycle.

2.2 Yield Curves Dynamics

Knowing how to interpret the YC dynamics is essential for several reasons. One of these is forecasting. Companies, consumers, and institutional entities base crucial decisions on the future predictions of bond yields. A second reason regards the monetary authority. Central Bankers continually monitor the market expectations and the price dynamics to make optimal policy decisions. A third reason regards the sovereign debt policy. In issuing new debt, the government has to select the amount and the maturities of the debt. This decision is made considering the actual YC and the impact that the debt issue will have on market expectations. Lastly, it is crucial to consider that the fixed income assets are all derivatives of the discount rates. Then modeling the YC is essential for correct pricing and hedging with other assets.[11]In this work, expectations, term premia, and convexity are the three building blocks used to model the YC. It is important to remark that this three blocks approach cannot fully explain the YC dynamics; in particular, two critical elements are left out. The first one is liquidity, which can become very important in a period of market distress. However, depending on the securities we deal with, the liquidity factor can be neglected. For example, the US Treasury bonds are between the most liquid asset available to investors. The second missing factor is the market segmentation, i.e., the idea that investors such as pension funds have preferences for specific maturity ranges when investing in fixed income securities. Consequently, bonds with different maturities are not interchangeable, then there are different supply-demand dynamics for each maturity term that influence the shape of the curve.

2.2.1 The $\mathbb{P}$ and $\mathbb{Q}$ measures

Before going any further is important to define two kind of probabilities measure: the $\mathbb{P}$ and the $\mathbb{Q}$ measures. $\mathbb{P}$ refers to the real world probabilities which can be defined as the probabilities of the various states of the world and can be estimated using all the historical and model-driven information available about the world and its history. This information may be about past prices or states of the economy. It is assumed that everybody agree on the real-world probabilities since these are extracted from observable public data through statistical tools available to each individual[22]. Real world probabilities cannot be used for pricing assets, since price it is not the simple weighted average of future outcomes. Investors are risk averse and generally require a compensation for their risk exposure. It is possible to solve this issue remaining in the $\mathbb{P}$ measure by modifying the possible future outcomes, taking into account a risk compensations. However, there is another way to look at the same problem. Instead of modifying the future outcomes, it possible to reverse engineer a set of probabilities by adjusting the probability of good market outcomes downward and incresing the probability of bad outcomes. This new probabilities are referred as risk neutral and are indicated as $\mathbb{Q}$ measure. These probabilities can be used to price any asset in the market since they already account for risk-aversion.

Under the $\mathbb{Q}$ measure is possible to derive an important expression for the price of a ZCB. It is possible to show (AppendixB) that the price of a ZCB with maturity T can be expressed as:

$$P_t^T = E_t^{\mathbb{Q}}[e^{-\int_t^T (r_s ds)}]. \tag{1}$$

This result relates the price of a discount bond to the $\mathbb{Q}$ measure expectation of the exponential of the path of the short rate from today to the maturity T of the bond.
Equation1 will come in handy in the following subsection.

2.2.2 The role of Expectations

The expectations hypothesis (EH) has been at the centre of empirical and theoretical work in fixed income since Macaulay (1938).

If the attractiveness of an economic hypothesis is measured by the number of papers which statistically reject it, the expectations theory of the term structure is a knockout. (Froot, 1989).

The term Expectations Hypothesis refers to numerous statements that link yields, returns on bonds, and forward rates of different maturities and periods. EH in its strong form is called the pure expectation hypothesis (PEH), which postulates that (a) expected excess returns on long-term over short-term bonds are zero, or that (b) yield term premia are zero, or that (c) forward term premia are zero. Independently on which statement is considered, according to

PEH market's rate expectations are the only determinant of the YC shape. The Economic literature has robustly rejected the EH, however it does not mean that expectations play no role in determining the YC shape.

Generally is convinient to work with continuous-time or log variables for tractability and simplicity reasons, and this is also true to formally explain how the expectations formation influence the shape of the YC. We anticipated in the section 2.1 the definition of the instantaneus forward rate as the logarithmic derivative of the price of a zero coupon bond expiring at time T:

$$f_t^T = -\frac{\partial log P_t^T}{\partial T}. \tag{2}$$

Integrating both sides of the equation one gets the following:

$$P_t^T = e^{-\int_t^T (f_t^s ds)}. \tag{3}$$

This is a fundamental relationship which allows to express the price of a ZCB in function of the forward rates values. It is important to underline how there is no single expectation involved in Equation3, all the quantities are known and can be extracted from time t YC. Previously, in Equation1 we have shown the price of a bond in the $\mathbb{Q}$ measure expectations. Comparing 1 and 3 we have:

$$P_t^T = e^{-\int_t^T (f_t^s ds)} = E_t^{\mathbb{Q}}[e^{-\int_t^T (r_s ds)}] \approx e^{-E_t^{\mathbb{Q}}[\int_t^T (r_s ds)]}. \tag{4}$$

The symbol $\approx$ in Equation4 points out that the expression is an approximation. This approximation is valid if we neglect convexity and consider the bond price as a linear function of the exponential of the path of the short rate. Otherwise, as we will discuss in the convexity section, the last part of Equation4 cannot be equate because of the Jensen's inequality[4]. However it is important to consider that in Equation4 are compared two different quanties; the path of the forward rate as it has been said, is well known at time t, while the second part involves the expectation over future and unknown values of the short rates. In the continuous compounding regime, the price of a ZCB can be also expressed as $P_t^T = e^{-y_t^T (T-t)}$ where y_t^T is defined as the continuously compounded yield from time t to T of the bond. Expliciting the previous relationship for y_t^T we obtain:

$$y_t^T = -\frac{1}{(T-t)} log(P_t^T). \tag{5}$$

Now, considering 4 and 5 the bond yield can be re-expressed as:

$$y_t^T \approx -\frac{1}{(T-t)} E_t^{\mathbb{Q}}[\int_t^T (r_s ds)]. \tag{6}$$

[4]Indeed the expectation of a function is not equal to the function of the expectation except for the linear case.

Equation6 shows that the ZCB yield can be expressed as an average of the $\mathbb{Q}$ measure expectation of the path of the future rates.

Considering Equation6 is possible to see how unconventional monetary policies as quantitative easing (QE) affect the market participants' expectations. Economists and finance professional regularly provide their predictions about macroeconomic data and financial quantities. Previously, it has been discussed how Central bankers also provide 'forward guidance' about the future paths of the rates they directly control. Some of these predictions are reported in public releases of which the Fed' blue dots' and the Bank of England' inflation fans' are probably the best known. Once again, to the extent these predictions are believed by the market participants, it should be possible to extract the real-world probabilities ($\mathbb{P}$-measure). However, it often happens that market yields (which are in the $\mathbb{Q}$-measure) differ from the public guidance. These discrepancies could be explained in terms of risk premia required from investors but also by the simple statistical concept of mode and mean. The surveys report the most likely outcome according to the experts that in statistical terms is the mode. The mode coincides with the mean or expected value if the possible yield realization follows a symmetric distribution. However, if the distribution is skewed, the yield market expected value can significantly differ from the survey forecast [4]. If the forward guidance were truly believed, then the CB statement should influence the yields in the desired direction. For example, if ECB set the deposit facility rate to 1% and stated its intentions to keep it at that level for 10 years, the 10-year sport rate would automatically be 1%. The CB recurs to unconventional measures such as QE to send a believable signal to the market. Indeed, the CB aims to make its message more believable by buying long-term sovereign bonds and having them directly on its balance sheet. As a consequence of the QE, the prices of the targeted sovereign bonds have increased, forcing investors to substitute the sovereign asset class by investing in lower rating bonds or other asset classes (i.e., stocks). Through this substitution effect, the lowering rate effect has regarded the whole fixed income asset class, and individual asset preferences have consequently changed. In particular, investors used to buy stocks for capital gain and bonds for a stable income. but since QE, bond yields have drastically decreased, reaching negative levels. Then investors switched their interest on high dividend yield stocks to fulfill their desire for a stable income.

This change of preference can be seen in Europe, where the ECB has been lowering the deposit rate since 2011 and the 10-year government AAA-rating bonds yield has fallen in negative territory. Market participants expect the yield downward trend to continue and they invest in bonds for capital gain since the bond price is inversely correlated with the interest rate level. Conversely, data shows that in 2020 the FTSE dividend yield was 3.6%, while the S&P500 one was 1.5%. Therefore, nowadays, investors prefer bonds for capital gain and stocks to achieve a stable income (through dividends). What has been discussed constitutes further evidence on the QE influence in changing traditional economic relations.

What if the CB is not believed? If the individual investor does not believe the rate will be kept low as stated by the monetary authority, he would consider the long-term bonds overpriced. Then he could short the long-term bonds and invest the money at the short-term rates, taking advantage of his expectation of the short rate.

2.2.3 Risk Premium

In explaining the YC dynamics, the main deviation from the EH comes from risk premium and convexity. Term premia are the compensation above the riskless rate an investor expects to receive for bearing a particular risk. As will be addressed later, there exist different kinds of risk exposures. Each source of uncertainty has its own "market price of risk" that, as the name suggests, is the market's price for bearing the specific factor's risk. Risk factors are not directly traded, but they are reflected in bond prices. The expected return from holding a T-maturity bond will then embed different compensations for unit risk, $\alpha_i^{t,T}$. These can be formalized as:

$$\alpha_i^{t,T} = \frac{\partial P_t^T}{\partial x_i} \sigma_i^t \lambda_i^t. \tag{7}$$

Where, σ_i^t is the time-t volatility of the ith risk factor,λ_i^t is the time-t compensation for unit risk associated with that x factor and $\frac{\partial P_t^T}{\partial x_i}$ measures the variation of the bond price in relation of the same factor change. So, the compensation to each risk exposure is transmitted down the curve by the responsiveness of a given bond to the risk factor, $\frac{\partial P_t^T}{\partial x_i}$. If we describe the YC dynamics in terms of expectations, risk premium and convexity[5], , the risks to which a bond is exposed are:

- Inflation risk: is the risk that inflation will compromise the investment return through a decline in purchasing power. The more investors are uncertain about future inflation, the more inflation risk premium they will demand. For the same estimate of expected inflation, a higher inflation uncertainty will translate into higher nominal yields because of the greater inflation risk premium. Recent studies (as Bauer and Rudebusch, 2018) show that inflation risk premium has become close to zero or even negative in recent years. Moreover, taking the data from 1999, the correlation value between the 10-year nominal and real Treasury yield is 96% (see AppendixA). This is a clear evidence that from the beginning of the new millennium, nominal rates have been driven by real rates and not inflation.

- Real-rate risk: is the risk of real rate returns changing over time. As in the case of inflation risk, real rates are uncertain, so the investors must estimate real rate future values. Moreover, since investors are risk-averse, compensation on top of their estimate is considered. It has already been mentioned how changes in the real rate ultimately reflect different market expectations about long-term economic growth. Since real economic growth estimations are very uncertain, they are associated with a risk premium [22].

[5]then neglecting liquidity risk and market segmentation (as stated in the introduction of Section2)

To summarize, the yield of a nominal bond depends on four components: the expected inflation, the risk premium for the uncertainty in inflation, the real rate of return, and the risk premium for the uncertainty in the real rate. Moreover, a fifth component will be considered in the next sub-section, namely the convexity.

2.2.4 Convexity

In the earlier sections, in deriving Equation6, we have assumed the relationship between the yield and the price of the bond to be linear. Actually, the price-yield function is convex and convexity is a measure of how much a bond's price-yield curve deviates from the linear approximation of that curve. Convexity is a desirable feature and investors are willing to pay more for it. Indeed, because of the convexity of the price-yield function, the price of the bond declines slower as yield increases. On the countrary when yields decline, the rate at which the price of the bond increases becomes faster. The convexity differ among bonds depending on the maturity and the coupon of the cash flows stream. In particular the lower the coupon the higher the convexity while the longer the maturity the higher the convexity. Then, convexity would be at its highest value in low coupon and long term bonds (i.e 30 years treasuries).

In the previous section we approximated the price of the bond as $P_t^T \approx e^{-E_t^{\mathbb{Q}}[\int_t^T (r_s ds)]}$,it is then possible to express the value of convexity in price terms as

$$C_t^T = P_t^T - \widetilde{P_t^T} = E_t^{\mathbb{Q}}[e^{-\int_t^T (r_s ds)}] - e^{-E_t^{\mathbb{Q}}[\int_t^T (r_s ds)]}, \tag{8}$$

where C_t^T is the difference between the real bond price P_t^T,and the approximated bond price that assumes a linear price-yield function $\widetilde{P_t^T}$. [22][6]

[6]Formula8 links a general mathematical result as the Jensen's inequality to the concept of convexity.

2.3 Modelling the Yield Curve

In this section we will briefly introduce a popular class of strucural models used to model the yield curve: the affine class. After having explained the adventages of such class we will further introduce three of the most popular affine models used in the literature.

2.3.1 Affine Models

The term Affine model is used to describe any arbitrage-free model in which bond yields are affine (constant plus-linear) function of some state vector x. Then in this class of term structure model the yield $y^{\tau}$of a τ-period bond can be written as:

$$y^{\tau} = A(\tau) + B(\tau)^{\intercal} x, \tag{9}$$

with coefficient $A(\tau)$ and $B(\tau)$ which depend on maturity τ. The function $A(\tau)$ and $B(\tau)$ make the yield consistent with the state dynamics. The main advantage of this class of models is tractability, indeed these provide tractable solutions for bond yields that otherwise need to be calculated with computationally costly methods as Monte Carlo simulations or by solving partial differential equations. The cost of tractability are some necessary restrictive assumptions on the bond yield function and the risk-neutral dynamics of the state vector x. More concretely, the risk-adjusted process for the state vector needs to be an affine diffusion, a process with affine instantaneous mean and variance. Defining the process x as an affine diffusion means that x solves:

$$dx_t = \mu_x(x_t)dt + \sigma_x(x_t)dz_t^{\mathbb{Q}}. \tag{10}$$

Then combinaning 9 and 10 is possible to build system of ordinary differential equations. To estimate the system parameters it is important to choose an estimation method and to define a measurement error[7]. In the empirical application in section 4, we will employ the Maximum-Likelihood method to estimate the parameters of the Vasicek Model. To use the Maximum log-likelihood one must be able to compute the density of the state vector x_{t+1} given x_t, $f(x_{t+1}|x_t)$. Then the conditional density of the vector of the yields can be obtained through a change of variable as:

$$f(Y_{t+1}|Y_t) = f(x_{t+1}|x_t) \mid \frac{\partial x_{t+1}}{\partial y_{t+1}} \mid . \tag{11}$$

The log-likelihood function of observed yields is then built as the sum of the log densities $logf(Y_{t+1}|Y_t)$ over the sample [22][8]. In the next sections we will introduce three popular single factor model, which employ the short rate as the state variable. In the empirical applications the short rate is generally chosen to be closely related to the deposit fund rate set by the CB so that it can represent a proxy for the risk-free rate.

[7]for example a common choice could be to add as measurement error ε_t^{τ}in Equation9

2.3.2 The Vasicek Model

The Vasicek model is one of the simplest affine models used to explain the YC dynamics. Two are the main adventages of this model. First, it can be analytically solved, and secondly, it provides a simple, intuitive interpretation of the YC dynamic. The simplicity of the model should be seen as a virtue rather than a drawback. The model cannot recover complex shapes of the curve. However, this inability to fit could be interpreted as an alert to the modeler that something important has been left out, i.e. liquidity risk, market segmentation, or a complex pattern of the short-term expectations.

In Vasicek the short rate follows a mean-reverting process and it is attracted to a constant reversion level θ with a intensity which is measured by the reversion speed k. This is formalized in the following stochastic differential equation:

$$dr\left(t\right) = k\left(\theta - r\left(t\right)\right)dt + \sigma_r dZ\left(t\right)^{\mathbb{P}}, \tag{12}$$

where $r\left(t\right)$is the short rate level at time t, σ_r is the standard deviation of the short rate and Z is a Brownian motion under the real-world probability measure $\mathbb{P}$. This last suffix highlights that the diffusion process is not risk-adjusted yet. We will provide a risk-adjusted form in the next paragraph. It can be shown[8] that considering a short rate diffusion process as the on in 12 and given the short rate today r_0,the future value of the short rate at time t is:

$$r_t = r_0 e^{-k(t-t_0)} + \theta\left(1 - e^{-k(t-t_0)}\right) + \sigma_r \int_{t_0}^{t} e^{-k(t-s)} dz_s. \tag{13}$$

The short rate distribution is normally distributed, since in 13 the Brownian motion follows a normal distribution by definition, and the other two components are fixed once defined the parameters. This normal distribution has mean:

$$E\left[r_t|r_0\right] = r_0 e^{-k(t-t_0)} + \theta\left(1 - e^{-k(t-t_0)}\right), \tag{14}$$

and variance:

$$s^2\left[r_t|r_0\right] = \frac{{\sigma_r}^2}{2k}\left(1 - e^{-2k(t-t_0)}\right). \tag{15}$$

The diffusion process is stationary since the variance in 15 does not grow indefinitely with time but tends to a finite asymptotic level given by:$(s^k_{asympt})^2 = \frac{{\sigma_r}^2}{2k}$

Zero-coupon bond prices in Viasicek's setting are given by

$$P_t^T = A\left(t,T\right)e^{-r_t B(t,T)}, \tag{16}$$

where

$$B\left(t,T\right) = \frac{1-e^{-k(T-t)}}{k} \text{ and } A\left(t,T\right) = e^{\left(\theta - \frac{\sigma_r^2}{2k^2}\right)\left(\frac{1-e^{-k(T-t)}}{k} - (T-t)\right) - \frac{\sigma_r^2}{4k}\left(\frac{1-e^{-k(T-t)}}{k}\right)^2}.$$

[8]see Rebonato, Bond Pricing and Yield Curve Modelling A Structural Approach (2016), 168-169

Both $B(t,T)$and $A(t,T)$ expressions can be derived applying the ito's lemma to Equation16.[9]

2.3.2.1 The Risk Premium in the Vasicek Model

In order to adjust the VM bond pricing formula in the $\mathbb{Q}$ measure we need to modify the short rate diffusion process in Equation12. In general, while looking at bond returns, accounting for a risk factor means adding an extra term to the riskless rate. In Section2.2.3 we introduced how it is possible to formalize the compensation required for unit risk. Similiarly, ad time t, the expected return for a bond of maturity T can be expressed as:

$$\frac{E\left[P_{t+dt}^{T-dt}\right]-P_t^T}{P_t^T}=r_t+\sum_i\frac{1}{P_t^T}\frac{\partial P_{t+dt}^{T-dt}}{\partial x_i}\sigma_i^t\lambda_i^t, \tag{17}$$

where the term $\frac{1}{P_t^T}\frac{\partial P_t^T}{\partial x_i}$ indicates how much a variation in the risk factor changes, in percentage term, the bond return. λ_i^t is the market price of risk that we defined as the the market's price for bearing the specific factor's risk. In the VM setting the only risk factor is the short rate, so $x_i = r$. Hence, the diffusion process in Equation12 can be corrected as:

$$dr(t)=[k(\theta-r(t))+\lambda_r\sigma_r]dt+\sigma_r dZ(t)^{\mathbb{Q}}. \tag{18}$$

The subscript $\mathbb{Q}$ indicates that now the process follows a risk-neutral dynamic. Equation18 can be rearranged as:

$$dr(t)=\left[k\left(\theta-r(t)+\frac{\lambda_r\sigma_r}{k}\right)\right]dt+\sigma_r dZ(t)^{\mathbb{Q}}=\left[k\left(\theta'-r(t)\right)\right]dt+\sigma_r dZ(t)^{\mathbb{Q}},$$

with

$$\theta'=\theta+\frac{\lambda_r\sigma_r}{k}, \tag{19}$$

where θ' is the adjusted risk reversion level. If the market price of risk λ_r is positive, then $\theta' > \theta$. This means that the rates will be higher since attracted to an higher reversion level and consequently the bond prices will be lower with respect to the ones obtained in the $\mathbb{P}$ measure [22].

2.3.3 Cox-Ingersoll-Ross model

Cox, Ingersoll, and Ross (CIR) have proposed the following alternative model:

$$dr(t)=k(\theta-r(t))dt+\sigma_r\sqrt{r}dZ. \tag{20}$$

As in the VM case k,θ and $\sigma_r$are nonnegative constant. The only difference respect to the Vasicek setting lies in the diffusion part of the process where the short rate standard deviation

[9] as in Rebonato (2016), Bond Pricing and Yield Curve Modelling A Structural Approach. p. 257-261

is proportional to the square root of the short rate $\sqrt{r}$ [18]. Thanks to this adjustment the short rate in the CIR model cannot be negative. In particular, when the rate r_t is close to zero, the term $\sigma_r\sqrt{r}$ also becomes very small. This dampens the effect of the random shock on the rate. Consequently, when the rate gets close to zero, the process is guided by the drift factor, which pushes the rate upwards (towards equilibrium). Since nowadays, interest rates have fallen in negative territory, the VM model should have a better fit to the YC.

Bond prices in CIR model have the same general form of those in Vasicek model (Equation16). However, since the state variable diffusion process is different. the function $B\,(t, T)$and $A\,(t, T)$ have different expressions that we will not report since these do not give any additional insight on the tractability of the model. The process adjustment to take into account risk-aversion can be implemented using the same method explained in the Vasicek case.

2.3.4 The Hull-White Model

The Vasicek and the CIR models are called equilibrium models. These kind of models cannot automatically fit the today's term structure. Even if the modeler carefully calibrate the parameters the output will still be an approximate fit of the several curve shapes that can be encountered in the real world. Traders find these approximate results unsatisfactory since in order to value derivatives they need precise information.

A *no-arbitrage* model is a model defined to be exactly consistent with today's YC. In contrast to the equilibrium model where today's term structure is the output, in no-arbitrage model the today's term structure is used as an input. A second fundamental difference between these two class of models, is that in no-arbitrage models the drift is time dependent.

Hull and White published in 1990 an innovative paper where they proposed an extension of the Vasicek model that provided an exact fit to the YC[18]. In one version of their model the diffusion process for the short rate is specified as follows:

$$dr\,(t) = [\theta(t) - kr_t]dt + \sigma_r dZ. \tag{21}$$

This is known as the Hull-White one factor model and as it has been mentioned earlier, the drift part of the differential equation is time dependent due to the presence of a time dependent reversion level $\theta(t)$.

This model has been further extended by the same Hull and White that proposed a two factor model. The setting is extended by defining an additional diffusion process for the reversion level. This will be of the kind

$$d\theta(t) = k_\theta[\theta_\infty - \theta(t)]dt + \sigma_\theta dZ, \tag{22}$$

where θ_∞ can be seen as the ultimate, long term level of reversion. Since arbitrage model parameters are calibrated on market prices, the formula are already expressed in the $\mathbb{Q}$ measure and do not need any adjustment. In investingating the relation between the YC and macroeconomic variables, most of the YC analysis assume contant long-run means. As shown in M.Bauer and G.Rudebusch (2019) [18], accounting for long-term mean variations in these underlying variables, is crucial in understanding and modeling the YC dynamics. Hull and White provide a method to take into account long-term mean variations, considering a richer pattern of YC movements and of volatilities than the one factor equilibrium models.

3 Research Objectives

In this thesis our aim is to show the applicability limits of the YC theoretical framework previously discussed. In particular, the empirical section of this thesis is formed of three parts on quite unrelated problems.

The first analysis is an application of the Vasicek Model to fit the Italian YC. We test three different model settings to take into account the investor's risk-aversion. According to our results, the best fit is achieved by defining the market price of risk as an affine function of the short rate.

The second part is a predictability analysis of the Italian excess returns. We estimate the single Cochrane-Piazzesi factor and then use it as a regressor. Moreover following the Cooper and Priestley intuition we combine the CP factor with the output gap showing an increase in the explanatory power of the model.
The first two parts will be addressed under a common section since we use Italian data in both the analysis.

Finally the last study regards the recession forecasting. We refer to the US recession periods defined by the National Bureau of Economic Research (NBER), and we attemp to identify a set of regressors able to signal a looming recession 10-months ahead. In particular we compare the results obtained from the logistic regression and the support vector machine method, finding that the logistic regression model is more conservative and then preferred.

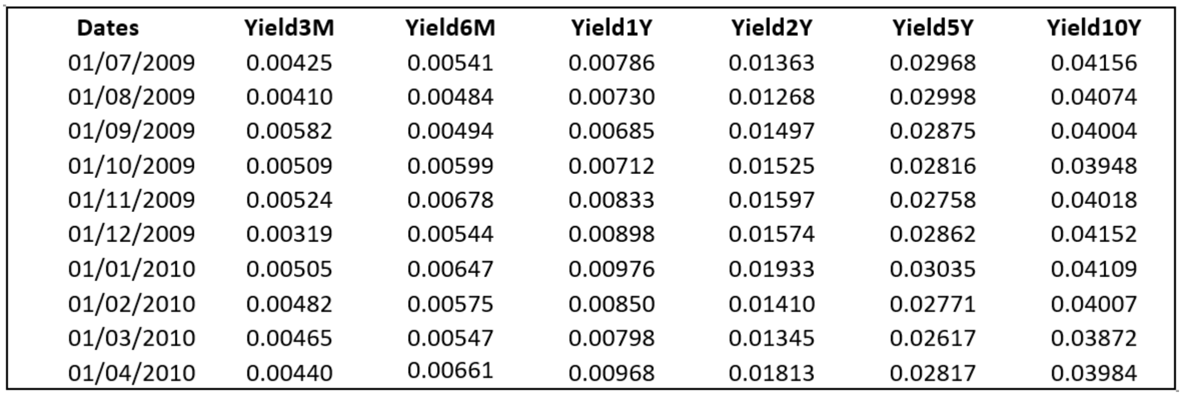
4 Emprical Analysis

4.1 Modelling the Yield Curve and Predicting the Excess Returns, an Italian Case

4.1.1 Vasicek Model

As discussed in the sub-section 2.3.2, the Vasicek Model is unparalleled for the intuitive undestanding of the YC dynamic it offers.

We use the monthly Italian bonds historic yields data for the period July 2009 – December 2020. The maturity considered are 3-months (M), 6M, 1-year (Y). 2Y, 5Y, 10Y. Next we form a $138X7$ matrix where each column contains the historic data for the specific maturity as shown in Figure3.



Dates	Yield3M	Yield6M	Yield1Y	Yield2Y	Yield5Y	Yield10Y
01/07/2009	0.00425	0.00541	0.00786	0.01363	0.02968	0.04156
01/08/2009	0.00410	0.00484	0.00730	0.01268	0.02998	0.04074
01/09/2009	0.00582	0.00494	0.00685	0.01497	0.02875	0.04004
01/10/2009	0.00509	0.00599	0.00712	0.01525	0.02816	0.03948
01/11/2009	0.00524	0.00678	0.00833	0.01597	0.02758	0.04018
01/12/2009	0.00319	0.00544	0.00898	0.01574	0.02862	0.04152
01/01/2010	0.00505	0.00647	0.00976	0.01933	0.03035	0.04109
01/02/2010	0.00482	0.00575	0.00850	0.01410	0.02771	0.04007
01/03/2010	0.00465	0.00547	0.00798	0.01345	0.02617	0.03872
01/04/2010	0.00440	0.00661	0.00968	0.01813	0.02817	0.03984

Figure 3: Fist 10-rows extracted from the analysed matrix. Each column contains the historic values for the specific maturity. The bond maturity considered are the 3 and 6 months and 1,2,5 and 10 years.

The short rate in this analysis is identified with the Italy 3-months Bond Yield.

In Section2.3.1, we discussed the Maximum log likelihood method to estimate the model parameters. We assume that the short rate follow a normal distribution with mean and variance respectively given by equations 14 and 15. So, using the statistical software R, we employ the function "*dnorm*" to compute the probability density function for each observation. We further define the average log likelihood function as

$$L(\boldsymbol{\theta}|r) = \frac{1}{N}\sum_{i}^{N} L(\boldsymbol{\theta}|r_i) = \frac{1}{N}\sum_{i}^{N} log(f_r(r_i)), \tag{23}$$

where N is the number of observations, $\boldsymbol{\theta}$ represent the vector of parameters, r indicates the short rate and f_r is the short rate probability density function. Next, we run an optimization routine to estimate the set of parameters $\boldsymbol{\theta}$ by maximizing Function23. The vector parameters $\boldsymbol{\theta}$ is composed of the reversion level θ,the reversion speed k and the standard deviation of the short rate σ_r. Table1 shows the parameters estimates,

	Reversion Level	Reversion speed	Standard Dev
Parameters	0.002	1.653	0.015

Table 1: Estimation of the Vasicek model parameters by using Maximum Likelihood estimation

In the authors' opinion, since the period covered is too short and the model simplicity allows few patterns of YC movements and volatilities, it is not possible to give a reliable interpretation to the estimated parameters. However, we can notice that the parameters confirm what we explained in the previous sections. In particular, at the beginning of the thesis, we anticipated that the short rate is positively correlated with the deposit rate set by the Central Bank. Our estimated reversion level is 0.002, and it is slightly higher than the average ECB deposit rate since 2009 that is -0.0017. This small difference reflects that the short-term part of the YC is affected by demanded risk premium for the Italian country's instability, even if for a tiny part. The reversion speed estimate is exceptionally high, and it indicates that the short rate tends to turn fast towards θ . If we interpret this result considering that once the CB sets the deposit rate, then the short rate reaches that level quite fast, then it is a plausible value. However, if we merely refer to the real world dynamics, the extremely high k value is not significant since the θ considered by the model is fictitious. In the real world, since the high degree of correlation between the deposit rate and the short rate, the reversion level changes according to CB policy decision. Finally the estimated volatility of the short rate is 1.5%, almost two times the 3M yields observed volatility that is 0.8%. In such context, a model allowing for more degree of freedom as the Hull-White Model would give a more accurate calibration of the parameters and should fit the YC better than an equilibrium model as the Vasicek.
However, since the aim of this study is not the pricing of bonds or derivatives, we choose to employ the Vasicek model since it provides a the transparent understanding about the YC dynamics. In particular, we will try three different settings to consider the investor's risk-aversion and achieve a better fit to the observed term structure. Before we proceed is necessary to define a function that computes the Vasicek theoretical yield. This function is obtained by combining equations 5 and 16 as follows:

$$y_t^T = -\frac{1}{(T-t)} \log \left(A\left(t,T\right) e^{-r_t B(t,T)} \right). \tag{24}$$

We name this as the "VasicekYieldFunction" and it takes as input: the short rate observation, the yield maturity T and the vector of parameters $\boldsymbol{\theta}$.

4.1.1.1 No - risk premium

In this first case we assume that the risk premium required by the investors is zero. So there is no need to make any adjustment to the diffusion process implying that $\mathbb{P} = \mathbb{Q}$. Generally, a good model should be able, at least, to fit the YC on average. The average observed YC is computed by taking the mean of each column of the matrix previously built. Figure4 shows the average observed YC and corresponding model fit.

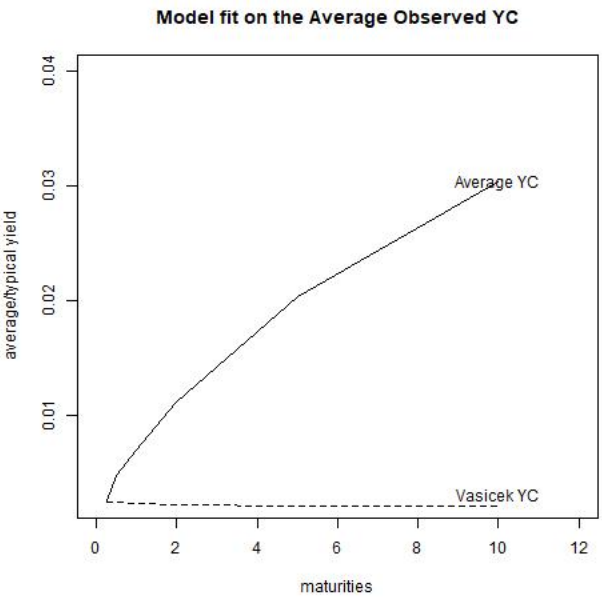


Figure 4: The continuous line represents the average observed yield curve, computed as the average of the observed yield for the specific maturity. The dot line is the model fit assuming a zero risk premium. Using this setting the model is unable to fit the curve.

In this case the output from the Vasicek model is just a stright line. Indeed, since the risk premium is assumed to be zero investors do not require any further compensation for the duration risk. However, the average observed YC is clearly upward sloping signaling that the assumptions of zero risk premium is incorrect.
As we discussed in Section2.3.2.1 is possible to adjust the short rate diffusion process to express the Vasicek formula under the $\mathbb{Q}$ measure by adding a risk compensation factor to the risk reversion level computed in $\mathbb{P}$, as in Formula19

4.1.1.2 Constant risk premium

In this case we will assumed that the investors require a constant risk compensation over time. To estimate the optimal risk-adjusted reversion level θ', we optimize the difference between the average observed yields and the Vasicek yields derived from Formula24. The θ' value obtained is 2% almost ten times higher than θ under $\mathbb{P}$. Figure5 shows the model fit to the average observed yield curve under this new model setting.

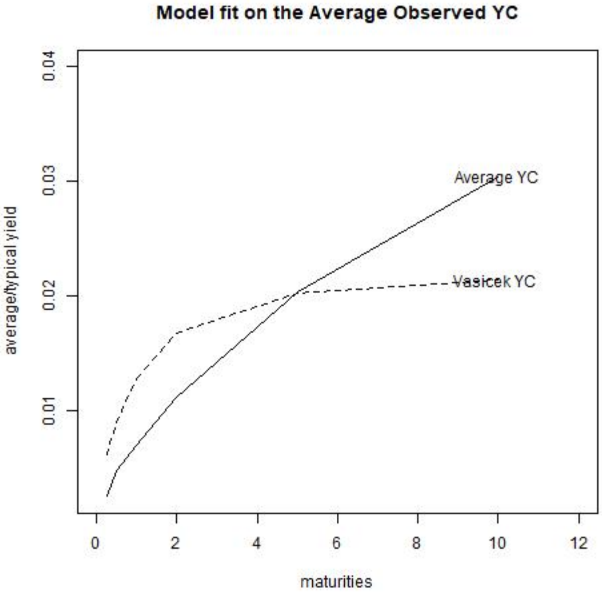


Figure 5: The continuous line represents the average observed yield curve, computed as the average of the observed yield for the specific maturity. The dot line is the model fit assuming a constant risk premium. This setting achieve a greater fit than the one shown in Figure4.

However, if instead of the average YC, we wanted to predict the curve in a particular month of the sample, the model perform poorly as it is shown is Figur6.

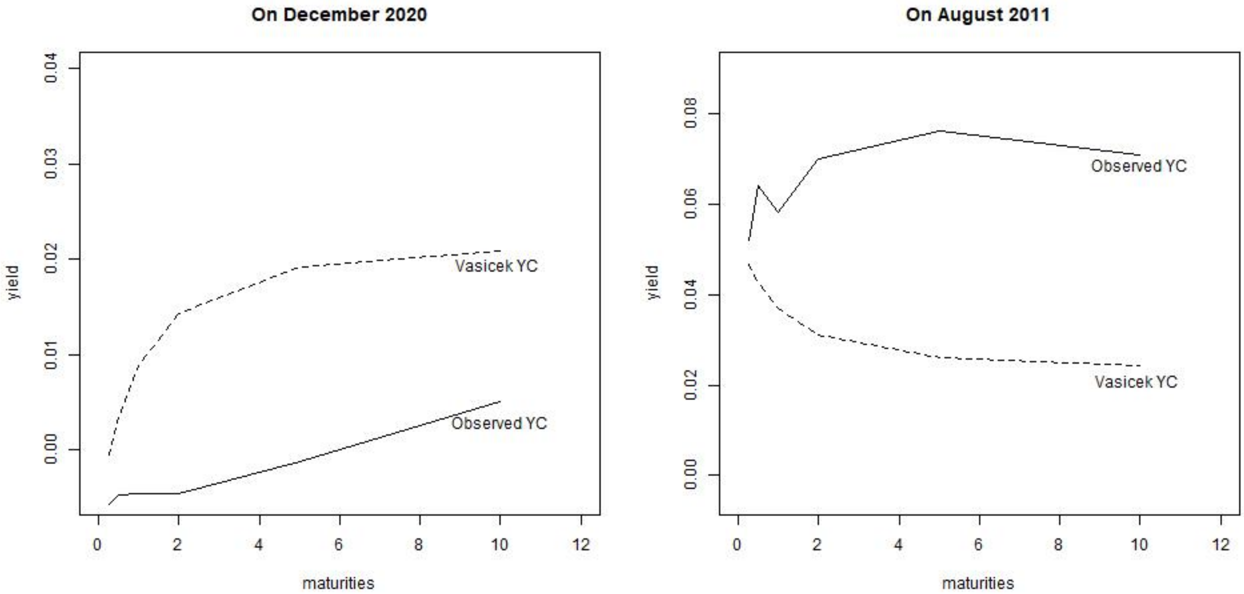


Figure 6: The figure exhibits the model fit (dotted line) to specific monthly observed yield curve (continuous line). The model fit to the observed YC in August 2011 is shown on the right, while on the left it is represented the fit to the curve observed in December 2020.

As we preannounced, this model setting is unreliable in fitting the YC in a specific date. In December 2020 the Vasicek YC is upward sloping as the observed curve. However, in August

2011, the observed YC has a complex upward sloping shape while the model fitted curve has a negative slope. We try to improve this shortcoming in the next paragraph.

4.1.1.3 Time varying risk premium

In this last case, we express the market price of risk as an affine function of the short rate. In order to achieve this, it is necessary to extract the martket price of risk $\lambda_r(t)$ implied by the observed values. We compute the risk adjusted reversion level $\theta'(t)$ for each month by optimizing the difference between the model fit and the observed values for each observation. Then we reverse Equation19 to obtain $\lambda_r(t)$ in each month. Figure7 represents the evolution of the estimated market price of risk $\lambda_r(t)$ over time.

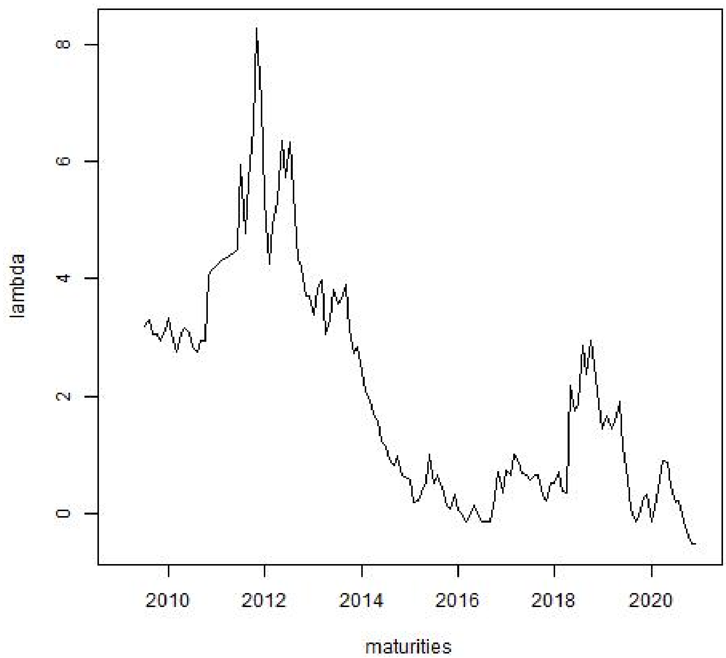


Figure 7: Evolution of the implied market price of risk $\lambda_r(t)$, in the period July 2009 – December 2020

It is possible to notice how the market price of risk reaches its peak levels in the 2011-2013 period. These were years of deep crisis for the eurozone due to the European Sovereign Debt Crisis's escalation started at the end of 2009. The modelling of $\lambda_r(t)$ also captures the systemic downtrend of the market price of risk realized after the famous ECB President Mario Draghi speech (July 2012), in which he stated the Central Bank would have done whatever it took to defend the single currency. Since then, the $\lambda_r(t)$ remained stable at low levels except for the period August 2018 – May 2019. Indeed, following the 2018 election the country remained without a government until the summer, when it was formed a coalition between the Five-Star Movement (M5S) and the Lega Nord party. The government collapsed after 14 months, plunging the country into a renewed period of crisis and uncertainty. In August

2019, Jason Horowitz published an article in the New York Times journal in which he used the following words: *"During the government's short tenure, the nationalist-populist coalition struck fear into the heart of the European establishment. It antagonized the European Union, flouted its budgetary laws, demonized migrants and embraced President Vladimir V. Putin's Russia and his strongman politics".*
These sentences stress the level of the world sentiment towards the Italian situation in that period. In March 2020 $\lambda_r(t)$ started growing again due to the outbreak of the Covdi-19 pandemic. Since June 2020, the perceived risk has been steadily decreasing and according to the model results, in September 2020 the value fell slightly in the negative territory. However, this does not mean that nowadays investors have become risk neutral or even risk taker. Indeed, there are many uncertainties for which investors would require compensations, particularly for the Italian case. We addressed this problem at the beginning of this thesis where we defined the decreasing of interest rates as a global phenomenon. The meager market price of risk does not depend on a country-specific factor, but rather is systemic. Since 2015, market stability has always become more dependent on the BCE quantitative easing. It can be argued that BCE has been trapped in the same market expectations it has led to. If these expectations are not met, deep market dislocations can be created. For example, in December 2018, there was a sharp sell-off in financial markets, ahead of the ECB announcement of ending the purchase program. Then, it can be argued that bond prices do not reflect investors' demand for risk compensation since they are kept artificially at low levels. Since rates cannot decrease indefinitely, when and how the BCE will ease its purchase program constitute a main source of instability for the eurozone countries [23].

Once obtained all the historic values of the market price of risk $\lambda_r(t)$, we express it as an affine function of the short rate:

$$\lambda_r(t) = \alpha + \beta r_{t-1} + \varepsilon_t, \tag{25}$$

where α and β are simple coefficient that we estimate through the Ordinary Least Square (OLS) method. Next, the estimate $\widehat{\lambda_r(t)}$ values and use them as input in Equation19 to compute the adjusted mean reversion level for each time t, $\widehat{\theta'(t)}$. Then, these values are used in the VasicekYieldFunction (Equation24). Thorugh this new setting the fit to the average observed YC is similiar to the one obtained in the constant risk premium case. However the model fit to the YC on a specific month has greatly improved as shown in Figure8.

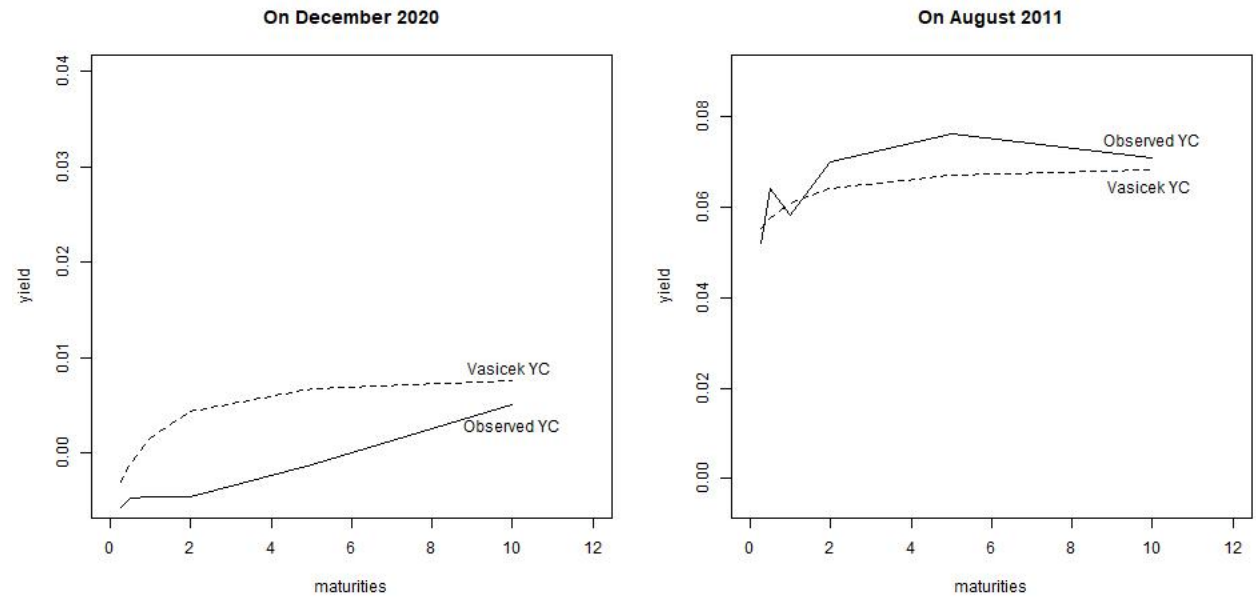


Figure 8: The figure shows the model fit (dotted line) to specific monthly observed yield curve (continuous line). The model fit to the observed YC in August 2011 is shown on the right, while on the left it is represented the fit to curve observed in December 2020. We can notice a great improvement with respect to the constant risk premium case.

In this setting the model is able to caputre paralallel shift movements of the curve with greater accuracy than the constant risk premium case. Overall, it is a great result since the few degree of freedom allowed in the Vasicek model. The main strength of this Vasicek configuration is the modelling of the market price of risk. Setting a routine that re-calibrates the parameters in Equation25 on a rolling window base, could further improve the model fit. However, the model is still very limited in recovering more complex YC shapes, as the humped YC.

4.1.2 Excess Returns Predictability

In the Section2.2 when we discussed the main reasons why it is important to understand the YC dyinamics one of those was forecasting. It is especially important for private and public institution as hedge funds or pension funds that base the duration exposure of their portfolio on the predicted future yields. In particular, shorter is the investment horizon of the company, the more accurate the yield dynamics are needed to be forecasted to capture the excess returns. In the economic literature, several research studies have investigated significant relations between macro-financial data and excess returns. Following the work of Cochrane and Piazzesi (2005)[11] and Cooper and Priestley (2009) [12], we show how well the identified predictors can explain the Italian 10-year excess returns, and if a combination of these variables can further improve the model results.

As a first step we build a $239X5$ matrix as shown in Figure9; each column contains the historic yield data for five different maturities: 2, 4, 6, 8 and 10 years. The data covers the period from February 2001 to December 2020.

Date	2Y	4Y	6Y	8Y	10Y
Feb 01	0.044199	0.046168	0.047647	0.049476	0.046546
Mar 01	0.042044	0.044725	0.046483	0.048781	0.045435
Apr 01	0.046788	0.047523	0.050474	0.051795	0.049242
May 01	0.044266	0.047675	0.050398	0.052498	0.04917
Jun 01	0.043222	0.046559	0.049457	0.052014	0.048273
Jul 01	0.041641	0.044476	0.047399	0.049704	0.04631
Aug 01	0.03949	0.042494	0.045786	0.048419	0.044768
Sep 01	0.036197	0.039874	0.045499	0.048895	0.044494
Oct 01	0.032409	0.036448	0.041334	0.044467	0.040502

Figure 9: Fist 10-rows extracted from the analysed matrix. Each column contains the rates historic values for the specific bond maturity. The bond maturity considered are the 2,4,6,8 and 10 years.

These data constitute the base from where we compute all the necessary features described below.

4.1.2.1 Defining the Excess Return:

In the theoretical framework part, we have discussed the pure expectation hypothesis (PEH). If PEH holds there is no uncertainty in the future yields realization and if we build a zero-cost portfolio at time t, it will yield a constant net return of zero over time. However, in the real world, the same portfolio evaluated at time $t + \varepsilon$ [10]could have a positive or negative return. The difference between the portfolio position a time $t + \varepsilon$ and the one at the initial time t is defined as the $t + \epsilon$ excess return.

The zero-cost portfolio is obtained by investing in a N-period bond and financing it by going short on a n-period bond, with $N > n$. In Cochrane and Piazzesi (2005), the authors use the 1-year bond as short position. Differently, since the lack of data for the 1-year yield we use the 2-year Italian bond. In AppendixC is shown how this choice slightly change the original excess return formula, that in our study is expressed as:

$$xret_{t+2} = p_{t+2}^{(N-2)} - p_t^{(N)} - \; 2Y_t^2 = Y_t^N N - Y_{t+2}^{N-2}\left(N-2\right) - 2Y_t^2, \tag{26}$$

where N is the bond maturity, p is the log of the bond price, $Y_t^N$is the zero coupon bond yield with N maturity. The yield Y used in Equation26 is intended to be continuously compounded and it is also known as the instantaneous rate. Typically, we observe the annualized rate and not the instantaneus ones, so the following conversion formula is used:

$$1 + r_{annual} = e^y \rightarrow y = \ln\left(1 + r_{annual}\right). \tag{27}$$

[10]ε is a small positive quantity.

4.1.2.2 CP factor estimation

In their research, J.Cochrane and M.Piazzesi shows how a linear combination of the forward rates can predict bonds returns at all maturities. In our study we need to change the forward rate expression used in the authors' original paper due to our choice to use the 2-year bond yield as refinancing rate. In AppendixC.0.1 is shown that in our setting the forward rate can be expressed as:

$$f_t^N = \frac{1}{2}(p_{t+2}^{(N-2)} - p_t^{(N)}) = \frac{1}{2}(Y_t^N N - \ Y_{t+2}^{N-2}(N-2)), \tag{28}$$

where the notations are the same used in Equation26.

Depending on the N-maturity of the bond on which we decide to invest, the regression model is set as:

$$xret_{t+2}^N = \beta_0^{(N)} + \beta_1^{(N)} y_t^{(2)} + \beta_2^{(N)} f_t^{(4)} + \beta_3^{(N)} f_t^{(6)} + \beta_4^{(N)} f_t^{(8)} + \beta_5^{(N)} f_t^{(10)} + \varepsilon_{t+2}^{(N)}. \tag{29}$$

In Cochrane and Piazzesi research, if we consider the estimated loadings in Equation29 for each N-excess return and we plot them against the forward maturities, the result is a perfect tent-shape. Figure10 compares the original research and our results.

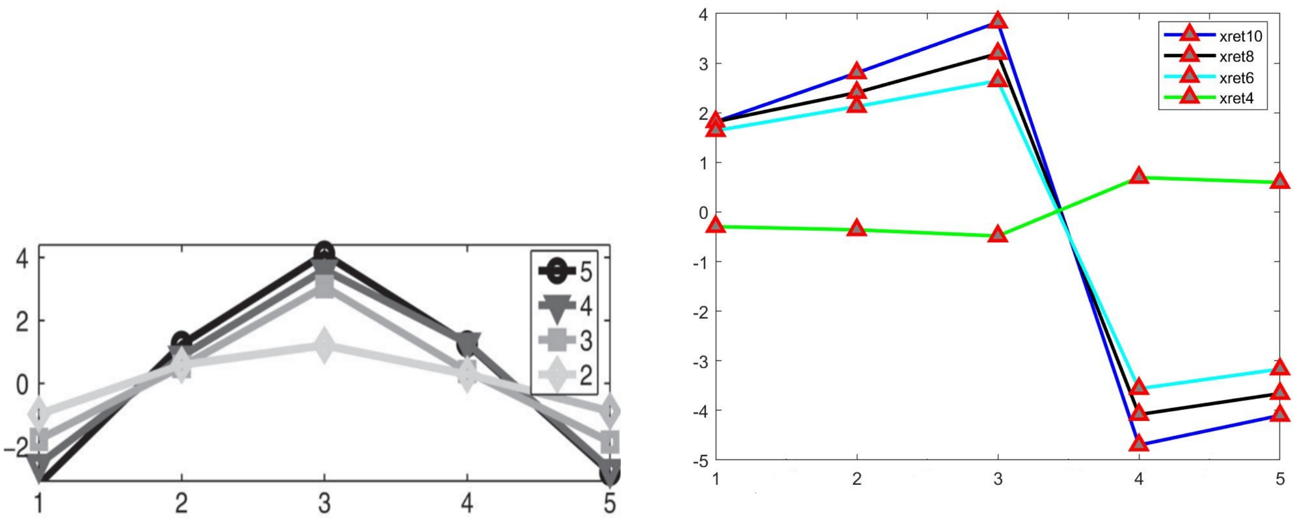


Figure 10: Cochrane perfect tent-shape estimates compared with the loadings path obtained in the Italian case (for each N-excess return considered). The N-excess returns considered in the original paper goes from 2 to 5 years. In this study, we refer to the 4,6,8 and 10 years excess returns since our choice to use the 2-year bond as short position to build the zero-cost portfolio.

Having obtained a tent-shaped linear combination of forward rate, Cochrane and Piazzesi were able to conclude that the return-forecasting factor was unrelated to pure slope movements since a linearly rising or declining forward curve would have given exactly the same return forecast. In our results, the loadings are distributed asymmetrically. If we consider $xret_{t+2}^6$,$xret_{t+2}^8$ and $xret_{t+2}^{10}$, the corresponding coefficients β_1 β_2 and β_3 follow an increasing path as in the original study. On the other side β_4 and β_5 are deeply negative. In the $xret^4_{t+2}$ case the coefficient β_1 β_2 and β_3 range from 0 to -0.5 while β_4 and β_5 range from 0.5 to 1,suggesting the end part of the forward curve to be the most informative in predicting excess returns.

Except for $xret^4_{t+2}$ Figure10 shows that the paths of the coefficients share the same trend across maturities. Then, as in Cochrane and Piazzesi (2005), it is possible to express Equation29 in terms of a single factor model, precisely the CP factor:

$$xret^N_{t+2} = b_n(\gamma_0 + \gamma_1 y^{(2)}_t + \gamma_2 f^{(4)}_t + \gamma_3 f^{(6)}_t + \gamma_4 f^{(8)}_t + \gamma_5 f^{(10)}_t) + \varepsilon^N_{t+2}. \tag{30}$$

In this specification, b_n and γ_n are not separately identified. Then, we normalize the coefficient by imposing the average value of b_n is one, $\frac{1}{4}\sum_{n=2}^{5} b_{2n} = 1$[11]Equation30 can be estimated in two steps. First we estimate the γ_n coefficient running a multivariate regression of the average excess return across maturities on all the forward rates:

$$\frac{1}{4}\sum_{n=2}^{5} xret^{2N}_{t+2} = \gamma_0 + \gamma_1 y^{(2)}_t + \gamma_2 f^{(4)}_t + \gamma_3 f^{(6)}_t + \gamma_4 f^{(8)}_t + \gamma_5 f^{(10)}_t + \bar{\bar{\varepsilon}}^N_{t+2}. \tag{31}$$

As we can notice in Equation31 the average excess return across maturities is computed by taking the average of the sum of $xret^4_{t+2}$, $xret^6_{t+2}$,$xret^8_{t+2}$ and $xret^{10}_{t+2}$.

	Gamma	Pvalue
γ_0	0.0006	0.7770
γ_1	1.5780	0.0001
γ_2	2.1804	0.0000
γ_3	2.7709	0.0000
γ_4	-3.6132	0.0000
γ_5	-3.2163	0.0000

Table 2: γ_n estimation result and corresponding pvalues

As we can notice from Table2 the path of the coefficients has the same trend across maturities of the loadings shown in Figure10.

Before proceeding to the second step, to ease the formulas explanation we define the following vectors:

$$\boldsymbol{\gamma} = [1 \quad \gamma_0 \quad \gamma_1 \quad \gamma_2 \quad \gamma_3 \quad \gamma_4 \quad \gamma_5] \qquad\qquad \boldsymbol{f} = [1 \quad y^{(2)}_t \quad f^{(4)}_t \quad f^{(6)}_t \quad f^{(8)}_t \quad f^{(10)}_t \;].$$

Both these vectors include an intercept. In particular, $\boldsymbol{\gamma}$ contains all the coefficient estimated in Table2 and $\boldsymbol{f}$ contains the 2 year spot rate and the N-year forward rates. Then Equation30 can be express as:

[11] The b subscript “2n” is necessary since our choice to consider excess returns from investing in bonds with 4,6,,8,10 years maturity while shorting the 2-year bond.

$$xret_{t+2}^{N} = b_n \left(\boldsymbol{\gamma}^T \boldsymbol{f}_t\right) + \varepsilon_{t+2}^{N}. \tag{32}$$

We are interested in analysizing the excess return realized by investing in the 10-year bond, namely to the quantity $xret_{t+2}^{N}$. Then we use the OLS method to estimate the CP factor b_n. Figure11 compares the model estimated excess returns with the observed values.

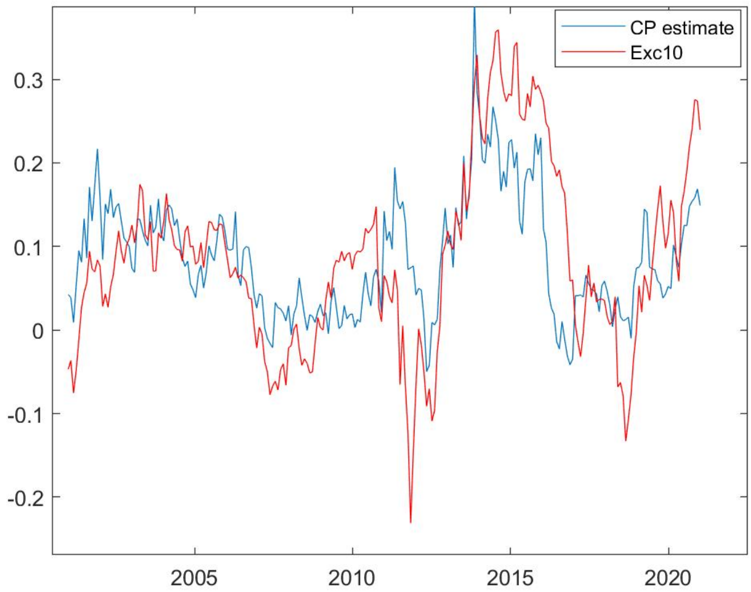


Figure 11: The red line represents the observed excess return generated by investing in the 10 year bond and financing the investment by selling the 2 year bond. The blu line shows the excess return estimated through Regression32

As we can notice in Figure11, since 2012 the model seems to underestimate the observed excess returns levels. Overall, the estimations keep track of the excess returns variation as shown by the model R^2 of 45.8%. This is close to the value of 44% found in the original paper. When Cochrane and Piazzesi developed this model, their intent were to study the time varying risk premia in U.S. governement bonds. As it is discussed in Rebonato (2020)[7], the CP model has an impressive fit in the sample data thanks to the large number of parameters employed. Given its setting, the CP model cannot capture changes in investors' expectations. However, the model high degree of flexibility allow it to chase past excess returns that were mainly due to changes in expectations, and to consider them as variations in risk premia. In further economic researches, it has been shown how the CP in-sample impressive performance does not translate well to out-of-sample and it has been suggested this could be due to the overfitting problem[12].[7][21]

[12]Overfitting happens when a model learns the detail and noise in the training data to the extent that it negatively impacts the performance of the model on new data. An overfitted model is a statistical model that contains more parameters than can be justified by the data.[22]

4.1.2.3 Output gap estimation

As it has been previously explained, our investment strategy consists of building a zero cost portfolio at time t and re-evaluating the portfolio position after 2 years to verify if any excess returns have been captured. Following the Cooper and Priestley research, we want to try to improve the goodness of fit of the CP single factor model by considering the output gap. This macroeconomic variable uses only production-related data and it is a classical business cycle indicator. In their original paper C. and P. have found significant results in the predictive power of the output-gap showing evidence regarding the risk-premium dependence over the business cycle.

In the economic literature have been used several methods to estimate the output-gap. In this study we employ a wide used technique which allows for a changing trend both linear and quadratic:

$$y_t = a + bt + ct^2 + v_t, \tag{33}$$

where y_t is the log of the industral production index[13], t is the time trend and v_t is the error term which correspond to the output gap. Figure12 shows how the Regression33 line fits the production index observations.

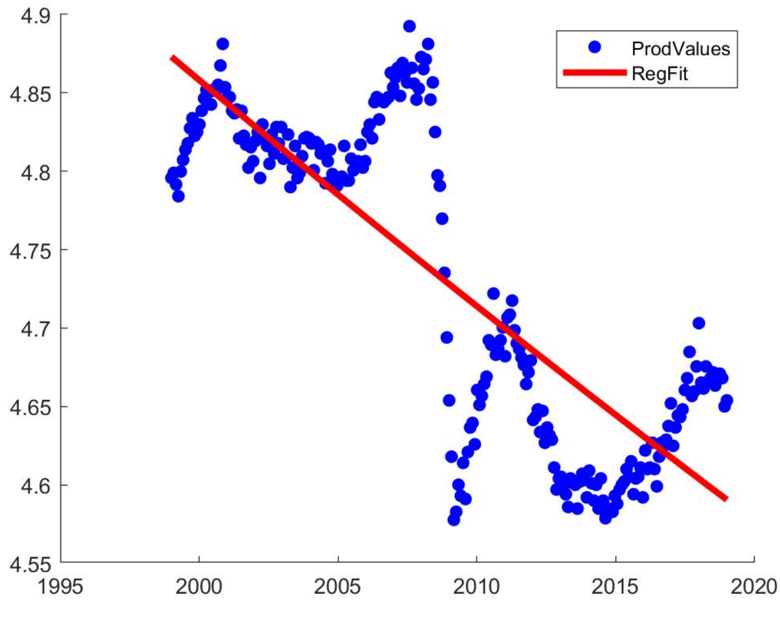


Figure 12: Output gap model fitted line across all the sample observations.

From Figure12 we can notice a general downtrend of the Italian industrial production. Moreover since the time t output gap has been defined as the distance between the observation and the fitted line, we can notice how the deepest gap values were in the period 1998-2000, 2008-2010 and 2012-2015. Finally, in the last few years 2018-2020 the model returns a positive output gap, signaling a positive conjunctural phase or that the industrial production

[13]OECD data, period considered: February 1999 – August 2020

downtrend may be coming to an end.

Once we have the output gap (OG) values for the period considered we turn the CP single factor model in a two factor model:

$$xret_{t+2}^{N} = b_n \left(\boldsymbol{\gamma}^T \boldsymbol{f}_t\right) + \beta_1 OG_t + \varepsilon_{t+2}^{N}. \tag{34}$$

As we did for the single factor model we choose $N = 10$, considering the excess return of the 10-year bond over the 2-year bond. We proceed through the OLS estimation, and in our result the OG is significant at 5% level. This suggests that the output gap is capturing risk that is independent of the financial market-based variable (CP factor). Moreover the new model R^2 is 47%, showing an improvement of about 1.2% with respect to the CP single factor model. In the original Cooper and Priestley (2009), the output gap resulted highly significant for all the excess returns maturities considered and its explanatory power ranged from 1% to 5%, depending on the estimation technique used for the OG computation.

In Section2.1.2 we have discussed how the CB's monetary policy during periods of "exuberance" and distress has followed an expected screenplay until the 2008 crisis. Furthermore, we mentioned how the long-term statistical regularities between business cycle, investor behaviour, and YC slope might no longer since the introduction of unconventional monetary policies as QE. Further evidence on this topic can be found in the OG predictive power analysis. If these economic relationships have actually changed or weakened, we would expect a consistent downtrend variation in the output-gap predictive power through time. To test this hypothesis, only the data until January 2015[14] are kept in the database. We apply the single CP factor model and the output gap extension to the new smaller database. According to our results the output gap extension model improved the single CP factor's explanatory power by 7%. So, it could be argued that in the last five years, business cycle variables have weakened their capability to track the variations of expected returns. These results constitute further evidence on the evolution of the long-term regularities that used to relate the business cycle and investors' behaviours. Nonetheless, business cycle variables as the YC slope and the output gap still significantly contribute in predicting future returns and then shall not be overlooked.

4.2 Recession Forecasting

Forecasting future state of recession has become a main research topic since the 1990's. Before then, academics focused on investigating the predicting power of financial variables on macroeconomic outcomes such as real output and inflation. The works of Estrella and Mishkin (1996,1998) are the first example in which financial and macroeconomic variables are used in a probit framework to forecast recession. According to Estrella and Mishkin's results, stock prices contribute for predicting recession within a short-time horizon, while for longer terms

[14]since we are analyzing Italy, we need to focus on the eurozone, where the quantitative easing program started in 2015

the slope perform better by it-self in out of sample prediction than in conjunction with other variables[13]. Just before the outbreak of the 2001 recession, Chauvet and Potter (2001) extended the probit method. The authors documented the instability of the YC variables predictive power and considered the existence of structural breaks. With their method, Chauvet and Potter achieved a better in sample fit compared to the original model by Estrella and Mishkin, Few years before the Great recession, Wright (2006) gave an important contribute to the recession forecasting literature. With his research, Wright re-examined the standard probit specification, in particular he looked more closely to policy instruments and other financial variables of interest to the Federal Open Market Committee (FOMC). Wright discovered that, by including the effective federal funds along a measure of the YC term spread, the predictive power improved. Following Wright intuition, few months later, King, Levin and Perli (2007) included in the probit framework credit spreads, and noticed an improvement in forecasting accuracy, plus a reduction in type 1 error. Recent studies as Favara Gilchrist, Lewis and Zakrajsek (2016) have decomposed credit spreads and showed that the main component in predicting recession is contained in a measure of the investor risk appetite called the Excess Bond Premium (EBP). In the same year Liu and Moench (2016) showed that the term spread had the highest predictive power at shorter horizons (four to six quarters ahead) and that adding lagged observations of the term spread improved the prediction accuracy. In general, the methods used in these researches have been well established in the field of econometrics (i.e probit regression, Markov switching and Bayesian techniques).
A great innovation in the forecasting recession literature came from methods that have their roots in computer science, namely machine learning (ML) classifier. On this aspect, a pioneering study was conducted by Ng (2014). In this paper, the probit framework is abandoned and a tree ensemble classifier is applied to a panel of 132 real and financial variables and their lags. This empirical analysis revealed that, despite the many variables taken into account, the predictors having a systematic and important predictive power consisted of about only 10 variables[15]. Holopainen and Sarlin (2017) used a large number of common techniques from conventional statistics and ML, with a particular focus on the problem as a classification tasks, for the purpose of creating an early-warning/crisis detection mechanism for the Euro area. They found evidence on the potential in more advanced ML approaches and they also point out the importance of using appropriate resampling techniques to account for time dependence[18]. In keeping with these results, in 2020 Michal Puglia and Adam Tucker wrote "Machine Learning, the Treasury Yield Curve and Recession Forecasting". In this paper, the authors investigated a small panel of 9 features and compared the recession forecasting power of a broad cross section of ML classifiers and the probit method. They concluded that the ML methods generally underperformed the probit regression in terms of forecasting accuracy. However, ML methods could capture important features of the joint empirical distribution of Treasury yields and other macroeconomic data over a recession indicator, which were not detectable by the probit methods. In particular ML methods thanks to their felxibility are able to capture also non-linear distribution.

4.2.1 Objectives and Methods Employed

The aim of the following study is to show how well a chosen set of regressors can predict a recession period in the United States (US). In particular, we use as the dependent variable a binary recession indicator which for any given month, is defined as a dummy variable taking the value of “1” if any of the following 10-months falls within a recession period and “0” otherwise. Moreover, we are not only interested in the classification accuracy of the model, but we also want to know at which point in time the model begins to capture the recession signal.
We focus on the US for two main reasons:

- the data granularity is higher than any other country;
- US have the highest GDP in the world and are also the top imports buyer.

Therefore, given the high degree of interconnection between the developed countries economies, a looming recession in US will most likely regards other countries as well. Then, the results of our model have a broader impact.
The analysis conducted is based on the work of Michal Puglia and Adam Tucker in “Machine Learning, the Treasury Yield Curve and Recession Forecasting” (2020). Following the same intuition of the authors, we use as predictive regressors in our model the Treasury YC slope, the Conference Board’s Leading Economic Index (LEI) and the S&P 500 index values. In addition, we consider the Conference Board’s Lag Index, Consumer Confidence Index and the US unemployment rate. In the analysis we compared the results obtained from two classes of ML models: the logistic regression and the Support Vector Machine (SVM).

4.2.1.1 Support Vector Machine

Support-vector machines were originally invented by Vapnik and Chervonenkis (1963) and grew in popularity after Boser, Guyon and Vapnik (1992) proposed the kernel trick for creating non-linear classifiers. The objective of the support vector machine algorithm is to find a hyperplane in an N-dimensional space, where N is the number of features, that distinctly classifies the data points. To separate the classes of data there are several possible hyperplans that could be chosen. The suppor vector machine algorithm focuses on finding plane that has the maximum margin which is the distance between data points of both classes. In other words, if the training data is linearly separable, we can select two parallel hyperplanes that separate the two classes of data, so that the distance between them is as large as possible. The region bounded by these two hyperplanes is called the "margin", and the maximum-margin hyperplane is the hyperplane that lies halfway between them. According to on which side of the hyrplane the data point falls, the observation is classified in the respective class. Figure13 shows the hyperplans dimensions in 2D and 3D feature space.

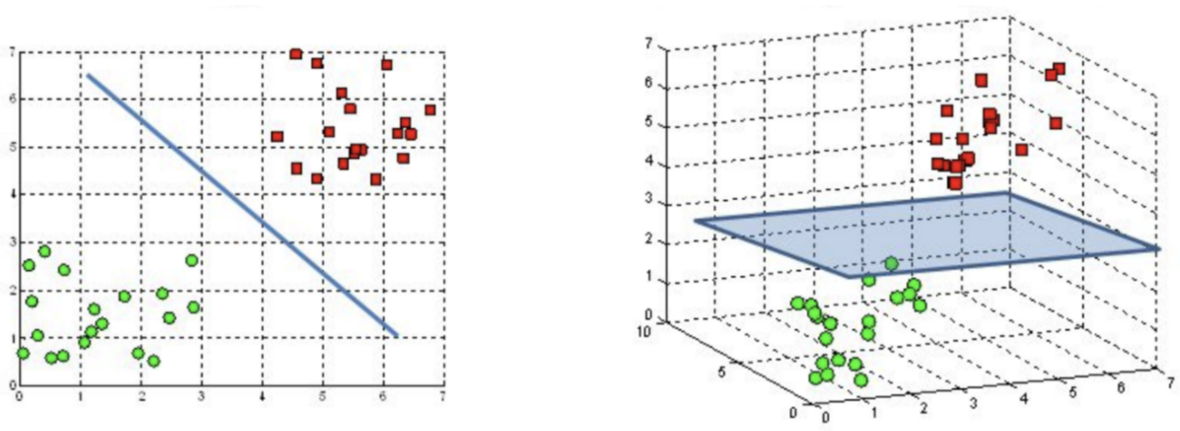


Figure 13: Hyperplanes in 2D and 3D feature space, in the 2D case the hyperplan is a simple line while in the 3D case it is a plane.

The Hyperplanes in 2D feature space is a line, the it can be expressed as $ax - y + b = 0$. If we define the vector x=(x,y) and $W = (a, -1)$ then in vector form the hyperplan is given by

$$\boldsymbol{W}\boldsymbol{x} + b = 0. \tag{35}$$

The data points closed to the hyperplan that influence its position and orientation are named as support vectors. These are used to compute the maximum margin. Differently from the logistic regression, in the SVM method we take the output of the linear function and if this output is equal 1 than we identify the observation with one class, while if the output is equal -1, we identify the observation with the other class. The aforementioned function is named hypothesis function and can be defined as:

$$y(x_i) = \left\{ \begin{array}{ccc} +1 & if & wx + b \geq 0 \\ -1 & if & wx + b < 0 \end{array} \right\} \tag{36}$$

To correctly select the hyperplan we can define the function $f = y(wx + b)$. This function is always positive when the observation is correctly classified and it will be negative otherwise. Given a dataset $D = \{(x_{i,}y_i)\,|x_i \in R^n, y_i \in -1, 1\}_{i=1}^{m}$where m is the sample size, we can compute f for each training example. We will denote as F the following function:

$$F = \min_{i=1..m} y_i(wx + b). \tag{37}$$

This is called the functional margin of the dataset, and when comparing hyperplanes, the one with the largest F will be favorably selected. Actually, the functional margin suffers from a scaling bias and one method to solve this problem is to divide the function f for the length of the vector w, $||w||$. Once the unbias metric is found, the values of $\boldsymbol{w}$ and b are computed through an optimization routine.

The SVM model can be set to have hard margin or soft margin. In the former case the aim of the model is to find a hyperplan that best separates the observed values such that there is no point misclassified. Conversely in the soft margin setting the model allows for some point to be missclassified. Soft margin are preferred when dealing with noisy data points so that a robust SVM classifier can be built ignoring the outlier or noisy observation. In practice, to distinguish between soft and hard margin it is necesary to operate on the constrains of the optimization problem. In particular in the hard margin case, the constrain to be respected is $y_i(wx+b) \geq 1$ so that no data point fall inside the margin. For the soft margine a regularization parameter ζ is introduced in the constraint, $y_i(wx+b) \geq 1-\zeta$. For higher values of ζ, the model will have softer margin, on the countrary for small ζ values the model will have hard margin.

In case of non linearly separable data, the simple SVM algorithm cannot be used. Intuitively this issue can be adressed by finding a way to map the data in an higher dimensional space where we are able to find the best hyperplan that clearly divides the classes. This method is effective, but not efficient since its computational burden. The kernel function addresses this problem, allowing us to operate in the original feature space whitout computing the coordinates of higher dimensional space. Kernel functions measures the similarity between two data points and can be of differen types[20]. In our application we will employ the simple linear kernel and the polynomial kernel.

4.2.1.2 Logistic Regression

In the Machine Learning context, logistic regression is named for the function used at the core of the method, the logistic function that is also known as sigmoid function. The logistic function is expressed as:

$$sigmoid(x) = \frac{1}{1+e^{-x}}. \tag{38}$$

this is an S-shaped curve that can take any real-valued number and map it into a value between 0 and 1. The logistic regression models the probability of the chosen class y. Defining with X the vector of regressors, the model equation is given by:

$$Pr(y=1 \mid X) = F(X_i\boldsymbol{\beta}) = \frac{1}{1+e^{-X_i\boldsymbol{\beta}}}. \tag{39}$$

Equation can be turned in:

$$ln(\frac{Pr(y=1 \mid X)}{1-Pr(y=1 \mid X)}) = X_i\boldsymbol{\beta}. \tag{40}$$

This re-formulation is useful since the righ side becomes just like a linear regression. The ratio on the left is called the odd ratio and is calculated as the probability of the event divided by the probability of not the event. Making prediction with a logistic regression consist of plugging the number in the logistic regression equation and computing the result. We can define the classification decision boundaries, and normally these are set as:

$$0 \; if \; Pr(y) < 0.5$$

$$1 \; if \; Pr(y) \geq 0.5$$

So the model will divide the data points in two classes following the logistic regression output. We can modify the decision boundaries, setting for example $0,7$ and 0.3, if we wanted to reduce the model error in classyfing the data in one of the two classes.

Differences and Advantages

Logistic regression and support vector machine are supervised learning algorithms, both used to solve classification problems. However, these methods present significant differences:

- The SVM method works on the geometrical properties of the data (it tries to find the optimal margin that separates the classes), while the logistic regression is based on a statistical approach (it maximizes the posterior class probability). This particularity allows SVM to handle noisy data as outliers, better than the logistic regression.

- Unlike logistic regression, SVM methods are designed to generate more complex decision boundaries. Indeed, the possibility to employ different kernel functions allow SVM to be more flexible and allow us to test non linear decision boundaries to separate the observed data.

In our analysis we will compare the results derived from the logit regressions with the ones from the SVM method. For the SVM, as we have previously anticipated, we will use a linear and a polynomial kernel. Then, first we will be able to see if the SVM geometric approach returns a better result respect to the logistic framework. Second, we will show if non linear decision boundaries are more efficient to classify the recession periods.

4.2.2 Data

As previously mentioned, we built a data panel of 6 features, that are used as regressors:

- The Treasury Yield Curve slope measured as the difference between the 10-Year Treasury Note and the 3-Months T-bill yield[15].

- The 3-month log difference of the LEI Index which is a weighted indicator whose main components are: 1. Average weekly hours, manufacturing (28% weight) 2. ISM® new orders index (16% weight) 3. Avg. consumer expectations for business conditions (14.2% weight) 4. Interest rate spread, 10-year Treasury bonds less federal funds (11.4% weight).

- The 3-month log difference of the Conference Board's Lagging Economic Index (LAG) which, as LEI Index, is a composite indicator whose main weights are: 1. Average prime rate (30.2%) 2. Consumer price index for services (20.8%) 3. Consumer instalment credit outstanding to personal income ratio (18.24%).[16]

[15]Data taken from the FRED Economic Data website.

[16]Both LEI and LAG data are taken from the Conference Board website.

- The 3-month log difference of end-of-month S&P 500 index values. [17]

- The 3-month log difference of the Consumer Confidence Index (CCI)[18], which is an economic indicator published by The Conference Board to measure the consumers' appraisal of current and prospective economic conditions.

- US unemployment rate.[19]

As it has been explained in the Subsection4.2.1 the recession indicator is built as a dummy variably. The recession periods considered are those defined by the National Bureau of Economic Research (NBER). All data is monthly and covers a period that goes from April 1970 to October 2020. Then the shape of our database is of 607 rows and 7 columns (including the dates). For each given date "t" is considered the regressor's value at time "$t-1$" to avoid forecasting biases since the regressors could embed future informations. In order to check for multicollinearity a correlation matrix is computed between all regressors, as shown in Figure14.

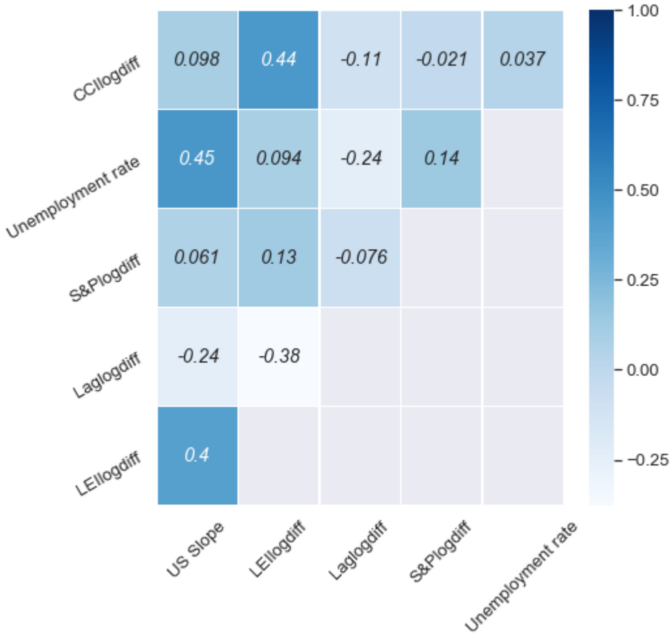


Figure 14: Correlation matrix between all the regressors, computed in Python. As shown, all the correlated values are below 0.5,so there are no highly correlated feature. The data labels are explained in Section4.2.2

The correlation btween the CCI log differences and the LEI log differences is of 0.44. This is expected since both composed index are affected by consumers' expectations. It is also interesting to notice the high degree of correlation between the YC slope and unemployment rate. Accoding to the economic theory, an upward sloping YC is an indicator that market

[17]Data taken from investing.com
[18]Data taken from the Observatory of Economic Complexity (OEC) website
[19]Data taken from macrotrend.net

partecipants expect the economic activity to likely increase in the future. However, since 2010 major currencies short rates have apporached the zero bound and some of them have fallen in the negative territory. This may distort the usual role of the YC as indicator of the market partecipants expectation. On this topic, Bomfim (2003)[19], wrote in early 2000s, *"we have a situation where an upward sloping (observed) yield curve is signaling expectations of a prolonged slump in the economy. This is exactly the opposite of the usual indicator property attributed to the yield curve".* In his research, Bomfin argues that when rates are sufficiently close to zero, a positive YC slope is more related to the expectations of a prolonged liquidity-trap situation than with forecasts of improved economic conditions. In this prospective, the positive correlation between unemployment and YC slope represents evidence on the distortion of the YC indicator as proxy of future economic activity.

Before to apply the methods discussed above, we shuffle the database rows, and we further split the dataset in two part, one for training the models (70% of the sample data) and the other part to test them. Shuffling is an important step before splitting the database, so the train and the test set are homogeneous and are not biased by specific time trends.

4.2.3 Results

We developed our study using the Python programming language[18]. In analyzing the results we could consider the accuracy of the models. However, this metric in our case is meaningless and not useful to make a comparison of the two class of model employed. This, because the recession indicator presents a disproportion between the "1", indicating the distressed economic periods and the "0". In particular, the recession indicator is composed by 461 "0" and 146 "1".[20] Therefore, a biased model which always predicts a value of zero would have about 76 accuracy. Given the fact we cannot limit our results on the accuracy metric, we will consider the confusion matrix generated from the models. The confusion matrix is a table with two rows and two columns that reports the number of false positives, false negatives on the right diagonal, and the number of true positives, and true negatives on the left diagonal. We will further compute the recall metric, considering it more useful for the purpose of the analysis. The recall metric measures the proportion of actual true predicted values, and can be expressed as follows:

$$Recall = \frac{TP}{TP + FN}, \tag{41}$$

where TP stands for "True Positive" and FN for (False Negative) values. FN are the observations the model classify as "0" when they were actually "1". We employ the Support Vector Machine method using a linear and a polynomial kernel. Moreover we specify different values for ζ , showing how the results differ if considering soft or hard margin.

[20]The difficulty to predict recessions comes from a lack of recessions (only seven since 1965)

Linear kernel

In our first attempt we consider the classical linear SVM model. We fit the model according to different values of ζ. The ζ parameter controls the tradoff between smooth decision boundary and classifing training data point correctly. For high ζ the algorithm focuses on getting the most observations as possible correctly classified. then selecting small margin. On the other side, for small value of ζ the algorithm main focus is to maximize the margin, giving less importance to the misclassification rate. We test for $\zeta = [0.1, 10, 1000]$, Figure15 shows the confusion matrices derived in the three cases:

	0	1
0	131	2
1	32	18

	0	1
0	127	6
1	21	29

	0	1
0	127	6
1	10	40

Figure 15: The figure shows the three confusion matrices derived for the three different values of the ζ parameter, $0.1, 10, 1000$. As expected the correctly classified values increase with the increment of the ζ parameter. The cost paid for this improvement is a rise in the False Positive (red circle). However, in the second and third matrix the FP value is the same, showing that choosing the ζ parameter equal to 1000 instead of 10,has only a positive effect on the model result.

Choosing an high value for the ζ parameter increase significantly the TP values. Moreover, observing the three matrices we notice that the highest recall metric is obtained for ζ =1000, with a value of 80%. Before drawning deeper conclusions on the support vector machine method we briefly show the model results obtained with the polynomial kernel.

Polynomial kernel

For degree-d polynomials, the homogeneous polynomial kernel is defined as: $K(x, y) = (x^T y)^d$, where x and y are vectors in the input space. Eventually, this function allow us to use non linear decision boundary to separate the data. We proceed as in the linear kernel case, testing the model for different values of the ζ parameter. We set the polynomial degree to 3 and train the model on the training dataset. We further test it and compute the confusion matrices for each value of ζ as shown in Figure16.

	0	1
0	133	0
1	50	0

	0	1
0	131	2
1	33	17

	0	1
0	125	8
1	19	31

Figure 16: The figure shows the three confusion matrices derived for the three different values of the ζ parameter, $0.1, 10, 1000$. Once again, as expected the correctly classified values increase with the increment of ζ. The cost paid for this improvement is a rise in the False Positive values (red circle). Differently from Figure15, for small value of the ζ parameter the model cannot classify any observation coorectly. Moreover considering ζ equal to 1000 brings a great improvement in terms of true predicted values, but the false positive are four times higher.

From these results we can observe that the polynomial kernel always undeperform the linear one for each value of ζ considered. This suggests that to analyse the data is better to use the linear SVM.

Logistic Regression

As an alternative to the SVM method we employ the logistic regression. Figure17 shows the confusion matrix computed using the logistic algorithm.

	0	1
0	130	3
1	26	24

Figure 17: Confusion matrix computed with the logistic regression method.

The model classify 24 observations correctly, presents 3 false positive values and have a recall metric of 48%. In comparing this results with the one obtained with the linear-SVM method, there is no a clear better choice. However we prefer the logistic regression because it is the most efficient and conservative model. This model is both conservative and efficient since it returns few false positive values and at the same time it has a good number of observations correctly classified. Moreover, if we imagine this model to be used to make investment decisions, money is at stake and is better to loose an investment opportunity than to make a wrong choice and to loose the capital. This approach is commonly used in the statistic literature, where it is considered better to keep the status quo unaltered than change it based on wrong results.

Given our criteria the best model is the logistic regression and the second-best is the linear SVM with ζ parameter equal to 0.1. We can now employ the feature importance analysis to verify whether the two ML models detect the same relevant features. In our study, for both models, we assess the importance of each regressor by accessing the classifier coefficients computed by the algorithm when trained. Figure19 shows the results of the feature importance analysis according to the two ML methods employed.

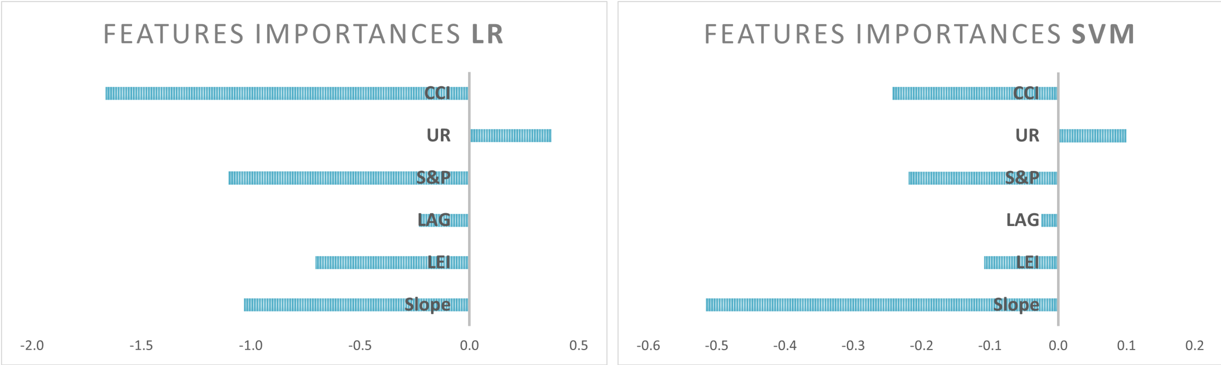


Figure 18: Feature Importance analysis result from the logistic regression (LR), on the left, and the linear SVM (Support Vector Machine),on the right. As indicated in the Data description section, "CCI" stands for the 3-month log difference of the Consumer Confidence Index, "UR" for the US unemployment rate, "S&P" for the 3-month log difference of end-of-month S&P 500 index values, "LAG" for the 3-month log difference of the Conference Board's Lagging Economic Index, "LEI" for the 3-month log difference of the Leding Economic Index and "Slope" for the Treasury Yield Curve slope. Overall, the SVM features coeffiecient have lower values than LR ones, The three most important features found thorugh the LR method are the CCI, S&P and the Slope. The SVM model returns the same feature but in different order, the Slope is the most important and it is followed by the CCI and the S&P.

From the feature importance analysis we can notice how the three most important variables are the same for both ML methods, even if in different order. In fact, the most relevant feature resulted in the LR method is the "CCI", followed by the "S&P", while for the SVM it is the "Slope", followed by the "CCI". Except for the unemployment rate, the coefficients for the other regressors are negative, showing a negative correlation with the business cycle. In particular, stock market trend, consumers' confidence and the YC slope, are all driven by consumers and investors' expectations and clearly, if these expectations are flourishing, the probability of a looming recession are very low.

When does the model begin to capture the recession signal?

We have selected the logistic regression as the most efficient and conservative model between the ones analysed. Now we wonder which are the specific values of the recession indicator that the model predicts correctly. In particular we are interested in knowing if the model is able to predict the recession in the early stages, before it actually realizes. If it were the case than the model could be used to capture the recession signal. Figure19 shows how the model behaved when applied in the Great Recession period (2007-2009).

Date	True Value	Pred_Value	Date	True Value	Pred_Value
01/08/2006	0	1	01/01/2008	1	0
01/09/2006	0	1	01/02/2008	1	0
01/10/2006	0	1	01/03/2008	1	0
01/11/2006	0	1	01/04/2008	1	0
01/12/2006	0	1	01/05/2008	1	0
01/01/2007	0	1	01/06/2008	1	0
01/02/2007	1	1	01/07/2008	1	0
01/03/2007	1	1	01/08/2008	1	0
01/04/2007	1	1	01/09/2008	1	0
01/05/2007	1	1	01/10/2008	1	0
01/06/2007	1	1	01/11/2008	1	0
01/07/2007	1	0	01/12/2008	1	0
01/08/2007	1	1	01/01/2009	1	0
01/09/2007	1	0	01/02/2009	1	0
01/10/2007	1	0	01/03/2009	1	0
01/11/2007	1	0	01/04/2009	1	0
01/12/2007	1	0	01/05/2009	1	0

Figure 19: Logit regression model applied to the observed data in the period 2006-2009. The column "True Value" represents the recession indicator while in the column "Pred_value" is reported the model output. As we wanted, the model seems to capture the recession signal in the early stage.

The model does not classify as recession the actual period in which it took place (December 2007 – June 2009). Nonetheless, the model fulfils the forecasting expectations by signalling the looming recession about 1.5 years ahead the official outbreak (December 2007).

As a second example, we consider the period from January 2019 to October 2020. During this time frame, major events as President Donald Trump's trade war with China and the global pandemic took place. Figure20 shows the logistic model behavior during this period, and this time we added a column showing the probabilities estimated before the classification of the data.

Date	True Value	Pred_value	Probabilities	Date	True Value	Pred_value	Probabilities
01/01/2019	0	0	0.45	01/12/2019	1	0	0.44
01/02/2019	0	0	0.47	01/01/2020	1	0	0.43
01/03/2019	0	0	0.43	01/02/2020	1	0	0.47
01/04/2019	1	0	0.46	01/03/2020	1	1	**0.57**
01/05/2019	1	1	**0.51**	01/04/2020	1	0	0.43
01/06/2019	1	0	0.49	01/05/2020	1	1	**0.54**
01/07/2019	1	1	**0.56**	01/06/2020	1	1	**0.51**
01/08/2019	1	1	**0.54**	01/07/2020	1	0	0.31
01/09/2019	1	1	**0.57**	01/08/2020	1	0	0.34
01/10/2019	1	1	**0.60**	01/09/2020	1	0	0.43
01/11/2019	1	1	**0.53**	01/10/2020	1	0	0.35

Figure 20: Logit regression model applied to the observed data in the period 2019-2020. The column "True Value" represents the recession indicator while in the column "Pred_value" is reported the model output. The column "Probabilities" contains the actual probability computed by the model before of classifying the observation. As happened for the Great Recession case, also in this scenario the model begins to capture signal of economic anomalies ahead of the beginning of the recession in February 2020 (as defined by the NBER committee).

As we discussed in the sub-section2.1.2 in United States the Treasury YC has inverted in

March 2019, August 2019 and started a new invesion in February 2020. Since February 2020 according to the NBER committee the US officially entered in recession. The 2019-2020 are very peculiar years that challenge our prediction model interpretability. Indeed, the end of the economic expansion period was almost entirely due to the outbreak of the epidemic, whitout which it may have not happened, at least in that moment. Moreover the epidemic is an unpredictable event, and it is not a consequence of a previous unhealthy economic situation. So in terms of predictability, training our model considering this period as a normal recession[21] could bias our results. The instability that characterizes these years is also captured from the model estimated probability path. In particular in October 2018 the estimated recession probability was about 30% while in January 2019 it became 45%. The fact that in 2019 there have been two YC inversion signals a precedent economic instability condition and maybe this helped the model to capture the recession signal also for this particular case, as shown in Figure19.

The results derived from the two setting of the SVM and from the Logistic Regression method must be carefully interpreted considering that the size and the composition of the dataset could bias the models output. The size problem, refers to the size of the test set, which is just the 30% of the whole database, so 182 observations. The database composition issue refers to the observations disproportions between the two classes. The test set randomly select from the algorithm could be composed of very few "1" and a lot of "0", than skewing the model results. Both problems can be partially addressed through the cross validation method, which allow the model to be tested multiple times on different test set originated from the sample data.

[21]The recession indicator marks with "1" also the 10 months ahead of the recession period

5 Conclusions

In the first part of this thesis, we discuss the importance of the informations embedded in the YC and describe the YC dynamic in terms of expectations, risk premium and convexity. In the Empirical section (4), we investigate three different topics related to the YC theoretical framework application on real data.

In the first analysis, we employ the Vasicek Model to fit the observed Italian term structure of interest rates. We express the VM in the $\mathbb{Q}$ measure and finding encouraging results when defining the market price of risk as an affine function of the short rate.

In the second study, we use Cochrane and Piazzesi (2005) and Cooper and Priestly (2009) framework to analyse the predictability of excess returns. Combining the CP factor and the output gap, we find an R^2 as high as 47% with a 1.2% improvement respect to the CP single factor model. Overall, our findings are in line with the cited works in terms of explanatory power of each regressor. The output gap is a significant predictor, however we notice that its explanatory power is decreasing over time. Indeed, until 2015 the output gap improved the model explanatory power of 7%. We argue that this decreasing trend is mainly due to the ECB unconventional policy, as quantitative easing, which altered the long-term statistical regularities between business cycle variables and investors' behavior.

In the third study, we investigate the research topic of recession forecasting. Acknowledging that machine learning methods have gained a widespread use acceptance in binary classification problem, we investigate their use in forecasting US recessions from the YC slope and other macro-financial variables. We use the intuition behind Puglia and Tucker (2020) "Machine Learning, the Treasury Yield Curve and Recession Forecasting" attempting to forecast recession periods ten months ahead. Next, we compare the results obtained from two machine learning methods: the logistic regression and the support vector machine. In our findings, the logistic method is preferred since it is both conservative and efficient. Furthermore, it returns few false positive and at the same time, it is able to classify correctly 48% of the observations. Besides, when we interrogate the model about at which point in time it begins to capture the recession signal, our forecasting expectations are fulfilled. Indeed, the model is able to capture the signal early in time, respect to the actual realization of the recession.

In conclusion, the different contexts analysed (Europe and US) and the adoption of several empirical methods used to model the YC and to forecast specific events and financial returns, have contributed to make this thesis a valid landmark to understand the yield curve dynamics and the potential of the information embedded in it.
Addintional results and demonstrations of the formulas used in this thesis can be found in the the Appendix section.

References

[1] Vincent Brosseau. (2002). “The Functional Form of Yield Curves”, European Central Bank, p. 9-10.

[2] Article posted by Bill Campbell III.CFA, http://www.financialexamhelp123.com/par-curve-spot-curve-and-forward-curve/ , on September 27, 2013.

[3] F. Reilly and K. Brown, Investment Analysis and Portfolio Management 10th Edition, South-Western, Cengage Learning, p. 636-645.

[4] Riccardo Rebonato, Bond Pricing and Yield Curve Modelling A Structural Approach (2016), *Cambridge University Press*, p. 3-8. 64-70. 125-138. 149-168. 535-540.

[5] Department of the Treasury, “Daily Treasury Yield Curve Rates.” Accessed Feb. 8, 2021.

[6] Federal Reserve Bank of San Francisco. (2008), "Dr. Econ: How Many Recessions Have Occurred in the U.S. Economy?”

[7] R. Rebonato and R. Ronzani. (2020). “Predicting Future Yields and Risk Premia: The Blue-Dot Affine Model”,*The Journal of Fixed Income,* p. 5-8.

[8] Monika Piazzesi 2010, “Handbook of Financial Econometrics: Tools and Techniques”,Elsevier B.V., p. 692-694 703-717.

[9] Antonios Sangvinatsos. (2008), “The Expectation Hypothesis” ,Working paper, University of Southern California, p 1-6.

[10] John C. Hull, “Options, Futures and Other Derivatives” 10th edition, p. 705-707. 717-718.

[11] John H. Cochrane and Monica Piazzes (2005) “Bond Risk Premia” ,*American Economic Review*, Vol. 95(1), p. 138-159

[12] Ilan Cooper and Richard Priestley. (2009), “Time-Varying Risk Premiums and the Output Gap”, *The Review of Financial Studies* 22, p. 2801-2831

[13] Estrella, Arturo and Frederic S. Mishkin. (1996), “The Yield Curve as a Predictor of U.S. Recessions”,Current Issues in Economics and Finance, Federal Reserve Bank of New York Research Paper Series, June 1996, 2(7), p. 1-2

[14] Michael Puglia and Adam Tucker. (2020), "Machine Learning, the Treasury Yield Curve and Recession Forecasting," Finance and Economics Discussion Series 2020-038. Washington: Board of Governors of the Federal Reserve System, p.1-30

[15] Ng, Serena (2014). "Viewpoint: Boosting Recessions," *Canadian Journal of Economics*, p. 30-32

[16] Pedregosa et al. (2011), Scikit-learn: Machine Learning in Python, Vol 12(85):2825—2830, p. 1-6

[17] Holopainen, Markus and Peter Sarlin (2017). "Toward robust early-warning models: A horse race, ensembles and model uncertainty," *Journal of Quantitative Finance*, p. 1-2

[18] Bauer, Michael D., Glenn D. Rudebusch. 2019. "Interest Rates Under Falling Stars," Federal Reserve Bank of San Francisco Working Paper 2017-16, p. 1-5

[19] Bomfim, Antulio N.(2003), "Interest Rates as Options Assessing the Markets View of the Liquidity Trap", FEDS working Paer No.2003-45, p. 8-9

[20] Christopher M.Bishop, " Pattern Recognition and Machine Learning",Information Science and Statistics, p 326-338

[21] S. Kessler, B. Scherer. (2009), "Varying Risk Premia in International Bond Markets". "Journal of Banking and Finance", 33(8), p. 1361-1375

[22] Everitt B.S., Skrondal A. (2010), Cambridge Dictionary of Statistics, *Cambridge University Press*, p. 314

[23] Laurence Fletcher. (2019), "Why the ECB's 'QE infinity' may have stolen the future", *Financial Times.*

6 List of Abbreviations

YC = "Yield Curve"
ZCB = "Zero Coupon Bond"
NBEAR = "National bureau of Economics and Research"
VM = "Vasicek Model"
CIR = "Cox-Ingersoll-Ross"
CB= "Central Bank"
ECB = " European Central Bank"
ML = "Machine Learning"
LR= "Logistic Regression"
QE = "Quantitative Easing"

A Real and Nominal 10-year Treasury yield

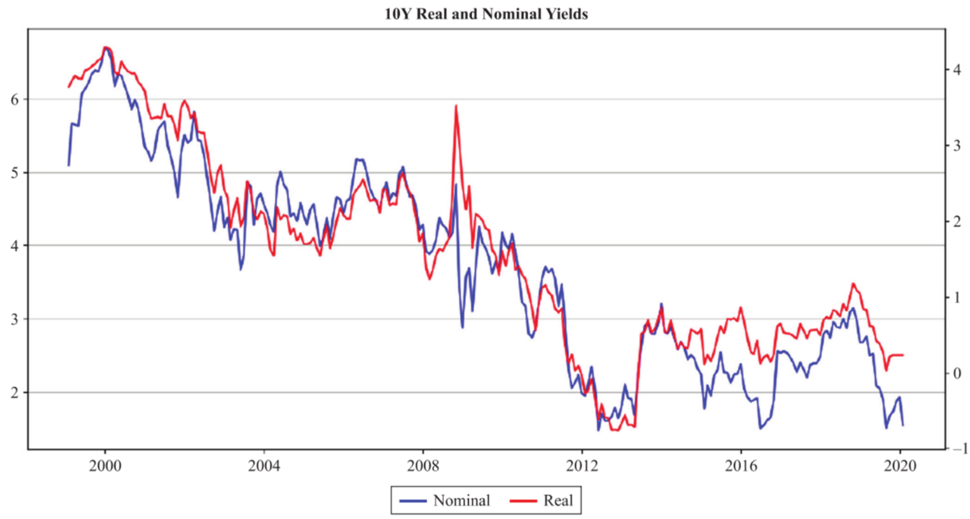


Figure 21: Evolution of the 10-year Treasury Nominal and Real Yields over time. From the plot is evident the high degree of correlation between the two measures. This shows that nominal yields have been driven by real rate and not inflation. Source: "Predicting future yields and Risk Premia, The Blue-Dots Affine Model",Riccardo Rebonato (2020)

B Bond Prices in $\mathbb{Q}$-Expectations

Let us assume a series of short rate:

- r_0 is known today and spans the period from 0 to $\triangle t$
- r_1 is not known today and refers to the period from $\triangle t$ to $2\triangle t$.

Consider now a discount bond that will pay \$1 at time $2\triangle t$. We need then to discount from time $2\triangle t$ to $\triangle t$ and from $\triangle t$ to today. Under the $\mathbb{Q}$ measure and in the continuous compounding regime,this relation can be expressed as:

$$P_0^{2\triangle t} = e^{-r_0 \triangle t} E_0^{\mathbb{Q}}[e^{-r_1 \triangle t}] = E_0^{\mathbb{Q}}[e^{-r_0 \triangle t} e^{-r_1 \triangle t}] = E_0^{\mathbb{Q}}[e^{-\sum_{n=1,2} r_i \triangle t}]$$

Extending the maturity of the bond and taking the limit of the sum for $\triangle t \to 0$ the expression can be rewritten as:

$$P_0^T = E_0^{\mathbb{Q}}[e^{-\int_0^T (r_s ds)}]$$

C Excess Return variation formula

In Cochrane and Piazzesi (2005) excess returns are defined as: $xret_{t+2} = p_{t+1}^{(N-1)} - p_t^{(N)} - p_t^1$ where p_t represents the log price of the ZCB bond. This can also be expressed as $P_t^T =$

$e^{-y_t^T(T-t)}$, then taking the log $p_t^T = -y_t^T(T-t)$. So the excess return formula can be expressed as:

$$xret_{t+2} = Y_t^N N - Y_{t++1}^{N-1}(N-1) - Y_t^1$$

Now given that in our study we finance our portfolio with the 2-year bond instead of the 1-year, the expression becomes: $xret_{t+2} = Y_t^N N - Y_{t++1}^{N-1}(N-2) - 2Y_t^2$

C.0.1 Forward rate variation formula

In discrete time the formula for the forward rate is given by:

$$F_t^n = \left[\frac{(1+r_1)^{n1}}{(1+r_2)^{n2}}\right]^{\frac{1}{(n_1-n_2)}} - 1$$

In Cochrane and Piazzesi (2005) the forward rate is defined as:

$$f_t^N = log\left(\frac{P_t^{N-1}}{P_t^N}\right) = p_{t+1}^{(N-1)} - p_t^{(N)}$$

Now given that in our study we finance our portfolio with the 2-year bond instead of the 1-year

$$f_t^N = log\left(\frac{P_t^{N-2}}{P_t^N}\right)^{\frac{1}{2}} = \frac{1}{2}(p_{t+2}^{(N-2)} - p_t^{(N)}) = \frac{1}{2}(Y_t^N N - \ Y_{t+2}^{N-2}(N-2))$$