Royal Institute of Technology

A dissertation submitted for the award of MSc

Risk premia implied by derivative prices

Author:
Richard Steffen

Supervisor:
Prof. Boualem Djehiche

September, 2015

Abstract

The thesis investigates the potential to recover the real world probabilities of an underlying asset from derivative prices by using the recovery approach developed in (Carr & Yu, 2012) and (Ross, 2011). For this purpose the VIX Index and US Treasury bills are used to recover the VIX dynamics and the short rate dynamics under the real world probability measure. The approach implies that VIX and its derivatives has a risk premium equal to zero contradicting empirical evidence of a substantial negative risk premium. In fact, we show that for any asset unrelated to the short rate its risk premium is zero. In the case of recovering the short rate, the CIR model is calibrated to the US zero coupon Treasury yield curve. The predictions of the recovered CIR process is benchmarked against the risk neutral CIR process and a naive predictor. The recovered process is found to outperform the risk neutral process suggesting that the recovery step was successful. However, it underperforms the naive process in its predictions.

Derivatprisers antydan om tillgångars riskpremier

Sammanfattning

Uppsatsen undersöker möjligheten att utvinna den naturliga sannolikhetsfördelningen tillhörande en underliggande tillgång från dess derivatmarknad. Genom att använda tillvägagångsättet som utvecklats av (Carr & Yu, 2012) och (Ross, 2011) undersöks VIX och amerikanska statsskuldsväxlar för att om möjligt utvinna dynamiken på VIX och den korta räntan under det naturliga sannolikhetsmåttet. Metoden antyder att VIX och derivat på VIX har en risk premie som är noll, vilket motsäger empirisk bevisning att risk premien är signifikant negativ. I uppsatsen visar vi även att i alla fall då den underliggande tillgången är oberoende av den korta räntan blir risk premien noll på den underliggande tillgången och dess derivat. I appliceringen av tillvägagångsättet på den korta räntan så kalibrerar vi CIR modellen till amerikanska statsskuldväxlar. Efter att hänsyn tagits till risk premien görs prognoser över framtida förändringar i nollkupongsräntan på växeln med 1 månads löptid. Dessa jämförs med prognoser från CIR modellen med risk neutrala parameterar och en naiv modell vars prognoser över framtida förändringar är noll. Det visar sig att prognoserna från CIR modellen med naturliga parametrar är signifikant bättre än prognoserna från modellen med risk neutrala parametrar. Dock, är prognoserna sämre än för den naiva modellen.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

A fundamental tenet of finance is the notion of compensation for bearing risk. However, not all risks are compensated. This is illustrated in the classical capital asset pricing model (CAPM). The main result states that the expected excess return of an asset is proportional to the markets excess return. In other words only the market risk or systematic risk is compensated. The economic intuition for this fact is that in the situation of negative market returns, positive cash flows are scarce and thus assets providing such cash flows are priced relatively higher than assets providing positive cash flows when they are abundant.

In asset pricing models, such as CAPM, the stochastic discount factor (SDF) play a central role. The SDF is defined as the stochastic process, M such that the following equation holds true,

$$\Pi(t;\mathcal{X}) = \mathbb{E}[M_T \mathcal{X}_T | \mathcal{F}_t]. \tag{1.1}$$

For now we leave the technicalities aside and just note that Π is the price of some payoff $\mathcal{X}_T$ and the expectation is taken under the real world probability measure conditioned on a set of information. Today's price of the payoff, given by Eq. 1.1, will depend on the time value of money and the market's attitude towards risk associated with $\mathcal{X}_T$, and these factors are what the SDF captures. To compute prices of assets the *modus operandi* is to decide on the functional form of the SDF and the independent variables upon which it depends. This will give rise to a model implied risk premium i.e. compensation for bearing risk. In order to measure this model implied risk premium, the parameters showing up in the expression of the SDF have to be estimated on historical time series data. Using historical time series data for estimation could be problematic if the data fail to fulfil vital assumptions needed for the estimation method. One such violation could be that of non-stationarity. This is a common characteristic in many financial time series.

Another common way to express Eq. 1.1 in the literature is as follows,

$$\Pi(t;\mathcal{X}) = \mathbb{E}^{\mathbb{Q}}[e^{-\int_t^T r_s ds}\mathcal{X}_T|\mathcal{F}_t], \tag{1.2}$$

where r_s is the risk free rate of interest. The risk aversion is now encoded in the probability measure which the expectation is taken with respect to, denoted $\mathbb{Q}$, while the discounting due to time value of money is given by the term, $e^{-\int_t^T r_s ds}$. An alternative way of modelling the risk premium associated with $\mathcal{X}_T$ is to take it as the difference or ratio between the expectation of the payoff with respect to the real world probability measure and the expectation of the payoff with respect to the risk neutral probability measure, $\mathbb{Q}$. Traditionally, an approach like this would necessitate a model of the time series dynamics of the payoff, $\mathcal{X}_T$, in order to estimate the real world probability measure which is unobservable. In addition, a risk neutral model of the payoff would also have to be assumed, although the parameters in this model does not have to be estimated on historical data since they can be backed out from the derivatives market provided it exists.

The goal of this thesis is, in contrast, to obtain the risk premium with the latter approach described above using derivative prices only and side step the need for historical time series data in the estimation. This requires the ability to infer the real world probability measure from the knowledge of the risk neutral measure. It has been thought amongst scholars that this has been impossible. The logic can be seen from Eq. 1.1. A change in the left hand side of the equation obviously has to be matched by a change in the right hand side for the equality to hold. This change can be caused by either a change in the real world probability measure or a change in the risk aversion channelled through M_T. However, since (Ross, 2011) sufficient conditions for making this inference are known. I will here use the approach developed by (Ross, 2011), known as Ross recovery, or rather of it developed in (Carr & Yu, 2012) to study the risk premia of assets. (Carr & Yu, 2012) reformulates the theory in the framework of continuous diffusion processes instead of discrete state markov chains as in (Ross, 2011). Furthermore, (Carr & Yu, 2012) use the notion of numeraire portfolio to avoid the use of a representative agent which is claimed to clear the way for analysing assets that are not closely approximated by the the market portfolio. Or in the words of (Carr & Yu, 2012),

"...we will focus on a more flexible theory [relative to (Ross, 2011)] in which the role of X [the underlying driver or asset] can be defined according to the derivative security prices one has on hand.".

In this study the underlying drivers will be concentrated to the Volatility Index (VIX) calculated by Chicago Board of Options Exchange (CBOE) and the short rate of interest respectively. This would at least in theory enable us to obtain the premium embedded in VIX futures, compensation related to stock market volatility risk, and the term premium on US Treasury securities, compensation related to interest rate risk.

The contribution of the thesis is twofold. Firstly, I illustrate the consequence, for recovery of the real world probability measure, when assuming that the underlying driver, for example VIX, is unrelated to the short rate. We will subsequently see that one assumption made in recovering the real world measure is that the risk free rate is given as a function of the driver of the market. In order to impose independence in this situation I assume that the short rate is constant. This assumption is standard in finance. A study that makes this assumption in a similar context is (Eraker & Wu, 2014). Secondly, I show an application of the theory in a fixed income setting where the underlying is the short rate of interest and this is tested on empirical data. Any empirical tests of the theory with continuous state space has to the best of my knowledge not been published. More specifically, the theory will be evaluated in its prediction power of the 1 month US Treasury bill rate, which is considered a proxy for the short rate.

The thesis is disposed as follows. First, previous research on the risk premium related to stock market volatility is reviewed. This is relevant because in this study the potential of the recovery approach in retrieving this risk premium is studied. Secondly, studies on cross-sectional calibration of the classical one factor CIR model, (Cox, Ingersoll, & Ross, 1985), will be reviewed since this is a major step in applying the present approach in a fixed income setting. Then literature on interest rate prediction will be reviewed since that is how the recovery will be empirically evaluated in this study. Subsequently, the recovery approach with continuous state spaces will be presented with some preliminary theory. Then the data and methodology is described. After that the empirical results will be shown and lastly I conclude.

Chapter 2

Literature review

2.1 Risk premia embedded in VIX futures

In a stochastic volatility framework the prices of stocks are not solely driven by one risk factor but the diffusion coefficient is random as well. Formally this could be expressed as follows,

$$dS_t = \mu(S_t, \sigma_t, t)dt + \theta(S_t, \sigma_t, t)dW_t, \tag{2.1}$$

$$d\sigma_t = \gamma(\sigma_t, t)dt + \varphi(\sigma_t, t)dB_t, \tag{2.2}$$

where B_t and W_t are two possibly correlated Brownian motions. According to economic intuition exposure to risk should be compensated. The compensation of volatility risk, dB_t above, has been termed volatility risk premium in the literature. There is a fair amount of literature investigating if this risk is compensated and to what degree.

(Bakshi & Kapadia, 2003) constructs delta-hedged S&P-500 index-option portfolios (long call-option, short stock). The portfolios are exposed to volatility risk and the delta hedge removes the portfolios exposure to dW_t-risk. The portfolios statistical properties are then measured over a sample period ranging from 1988 to 1995. The portfolios average returns significantly underperforms zero (across most strike and maturity categories). The author's attributes this as evidence in support of a negative variance risk premia.

(Carr & Wu, 2009) synthesizes variance swap rates by a portfolio of options and define the variance risk premium as the difference between the realized variance and this synthetic variance swap rate. Taking the mean of the difference and multiply by 100 yields the average dollar profit and loss for each 100$ notional investment in the variance swap contract. This average and other summary statistics are computed for a universe of 40 stock indices (S&P, Dow Jones, NASDAQ) and stocks traded in the US between 1996 and 2003. They find that the averages are significantly negative for the stock indices and most of the individual stocks.

More in tune with the intent in this thesis, the variance risk premium can be gauged by using derivatives on VIX. (Eraker & Wu, 2014) document substantial negative return premium for both VIX futures and structured products based on VIX futures called ETNs. A 1-month constant maturity portfolio of VIX futures has a negative return of 30% per year over a sample period ranging from 2006 to 2013. They propose an equilibrium model that explains the negative returns of buying VIX futures. As in this thesis they assume a constant risk free rate, based on the observation that the risk free rate itself has no particular importance in valuing short term equity derivatives such as VIX futures. They formulate the expected return of holding VIX futures in a similar way as in this thesis as the quotient between the expectation of VIX at a future date with respect to the real world probability measure and the expectation of VIX with respect to the risk neutral measure.

(Johnson, 2015) studies the shape of the VIX term structure. The VIX term structure is composed both of conditional volatility expectations and a risk premium and the goal of the study is to determine to what degree variations in the shape reflects the changing risk premium and the changing volatility expectations respectively. The author applies principal component analysis to the term structure. The first three components are interpreted as the *Level*, *Slope* and *Curvature* of the term structure. The *Slope* component manages to summarize all information in predicting the excess returns of S&P-500 variance swaps, VIX futures and S&P-500 straddles which are all positions with increasing values in S&P-500 implied volatility.

2.2 Cross-sectional calibration of CIR model

A critical step in making the coming theory operational in an interest rate setting is the ability to cross-sectionally calibrate short rate models to the yield curve. In the present study the CIR model will be used so only the relevant literature is reviewed. The first extensive empirical study in this field is (S. J. Brown & Dybvig, 1986). They use monthly data on nominal US Treasury securities (14 maturities) over the period 1952 to 1983 to fit the CIR model. They do a least squares fit of the model to 14 bonds in the cross-section. In addition to the CIR parameters the short rate are included among the optimization variables. They focus on analysing the implied short rate and implied variance of the process. They conclude that both of these quantities are estimable. The time series estimates seems to correspond quite well with the implied variance from the cross section. The model appears to fit Treasury bills better than other issues. The same approach is taken by both (Barone, Domenico, & Emerico, 1991) and (Moriconi, 1995) but for Italian bonds between 1983 to 1990 and 1990 to

1992.

In (R. H. Brown & Schaefer, 1994) a similar study is made but for index-linked (real) UK Government bonds from 1984 to 1989. They also find that the yield curve is very well fitted by the CIR model within each cross section and that the objective function to minimize is flat in certain directions in parameter space. The parameter trajectories show a considerable instability from day to day. Many of the same conclusions are arrived at in (de Munnik & Schotman, 1994) but for Dutch government bonds from 1989 to 1990. In (Carriere, 1999) US Treasury strip data is used. Here again the fit is concluded to be good.

In (Rogers & Stummer, 2000) compares the fit of a regular least squares when all CIR parameters act as variables on each single day with the fit when all CIR parameters are kept constant and the short rate is the only optimization variable remaining. They find the fit is worsened but it is still not essentially worse.

2.3 Interest rate prediction and the term premium

The empirical evaluation of the model in the present thesis will be based on forecasting the 1 month Treasury bill rate (TBR1) which could be considered a proxy for the short rate. A study that uses it as a proxy for the short rate is (Chan, Karolyi, Longstaff, & Sanders, 1992). Other proxies could be used, such as US Effective Federal Funds Rate or the LIBOR rate. There is a vast literature on forecasting these short term and longer term interest rates. Many different methods are used. Some papers that forecasts interest rates under the real world measure is (Diebold & Li, 2006), (Diebold, Li, & Yue, 2008) and (Moench, 2008). There are also studies where forecasts are made under the risk neutral measure as in (de Munnik & Schotman, 1994) and (Bams & Schotman, 2003)

In (Fama, 1976) forward rates implicit in US Treasury bills are used as predictors of the future short rate. Additionally, he estimates the term premium i.e. the higher expected returns for longer maturity Treasury bills, by linear regression on historical returns. He finds that the forward rates performs poorly in predicting future short rates at least beyond horizons of two years unless the adjustment for the term premium is made. At which horizon one has to adjust for the term premium is unclear however. (Longstaff, 2000) finds that the term premium is small in the near term for tenors up to three months for short-term repurchase rates. (Hin & Dokuchaev, 2015) has a similar methodology as in this thesis, namely using the US Treasury yield curve to estimate the CIR model and predict the the short rate proxy, which in their case is the US Effective Federal Funds Rate. The predictions are made with the risk neutral parameters and the forecasting accuracy deteriorates after 6 months. The authors attributes this to the neglect of the term premium in their method. Other studies documenting the improvements when taking the term premium into account is (Huang & Lin, 1996), (Dai & Singleton, 2002) and (Cochrane & Piazzesi, 2005). In the following chapter the recovery approach is presented. This enables the separation of the term premium so it can be accounted for.

Chapter 3

Background

3.1 Preliminary: Sturm-Liouville theory

In this section the relevant theory of Sturm-Liouville problems will be presented. For a more complete account see (Edwards & Penney, 2008). Sturm-Liouville problems naturally occur when solving partial differential equations with the separation of variables technique. This is in fact the situation in which the Sturm-Liouville problem occurs in this thesis as well.

Definition 1. *A Sturm-Liouville problem is a differential equation of the following form*

$$\frac{d}{dx}\left(p(x)\frac{dy}{dx}\right) - q(x)y + \lambda r(x)y = 0, \quad (a < x < b);$$
$$\alpha_1 y(a) - \alpha_2 y'(a) = 0,$$
$$\beta_1 y(b) + \beta_2 y'(b) = 0,$$

with neither α_1 *and* α_2 *both zero nor* β_1 *and* β_1 *both zero.*

Here λ is a unspecified parameter called an eigenvalue. The function $y(x)$ is an called eigenfunction. It is sometimes convenient to define the spectral operator as $L(y) \equiv \frac{d}{dx}\left(p(x)\frac{dy}{dx}\right) - q(x)y$, this clarifies that the problem can be considered a continuous analogue of a discrete eigenvalue problem. The following theorem will be used in proving that the eigenfunction corresponding to the smallest eigenvalue is constant when $r(x)$ is constant.

Theorem 1. *Suppose* $p(x), p'(x), q(x)$ *and* $r(x)$ *are continuous on* $[a, b]$ *and suppose* $p(x) > 0$ *and* $r(x)$ *for all* x *in [a,b]. Then the Sturm-Liouville problem in*

Definition 1 has an increasing sequence of eigenvalues

$$\lambda_1 < \lambda_2 < \lambda_3 < \dots$$

such that

$$\lim_{n\to\infty} = +\infty$$

and such that to each λ_n *there is (up to a constant multiple) a single eigenfunction* $y_n(x)$*. The* n*:th eigenfunction has exactly* $n-1$ *zeros in* (a,b)*. Moreover, if* $q(x) \geqslant 0$ *and* $\alpha_1, \alpha_2, \beta_1, \beta_2 \geqslant 0$ *then* $\lambda_n \geqslant 0$ *for all* n*.*

A pair $(y_i(x), \lambda_i)$ for some $i \in \mathbb{N}^+$ is called the i:th fundamental or principal solution to the problem in Definition 1. If the hypotheses of Theorem 1 are satisfied the Sturm-Liouville problem is said to be regular, otherwise it is singular. We end this subsection by noting that any second order differential equation of the form,

$$A(x)y'' + B(x)y' + C(x)y + \lambda D(x)y = 0$$

can be transformed into the form in Definition 1 by multiplying by a suitable factor.

3.2 Ross recovery with continuous state spaces

This section follows (Carr & Yu, 2012) to derive the main theoretical result of this paper. It provides an extension of the finite state Markov chain framework of (Ross, 2011) to an economy with asset prices driven by a bounded univariate time-homogeneous diffusion process which we will denote by X. The former article uses an alternative set of sufficient conditions, avoiding the use of a representative agent, in deriving the same result. This yields a theory that enables us to interpret X as an arbitrary process upon which there exists a derivatives market as opposed to a process that could be considered a proxy of the market portfolio. We will next go through the sufficient conditions and then show the derivation of the result. We will then clarify the implications of the theory when assuming a constant short rate of interest.

3.2.1 Model assumptions

Consider a probability space denoted by $(\Omega, \mathcal{F}, \mathbb{F})$. Here $\mathbb{F}$ is the objective probability measure which is unknown ex ante. The goal of this section is to show that with the following assumptions one is able to identify the measure $\mathbb{F}$.

Assumption 1: There exists a money market account (MMA) with value function, $S_{0,t} = e^{\int_0^t r_s ds}$, which evolves according to,

$$\begin{aligned} dS_{0,t} &= r_t S_{0,t} dt, \quad t \geqslant 0, \\ S_{0,0} &= 1. \end{aligned}$$

The growth rate, $r_t \in \mathbb{R}$, has the standard interpretation of the risk free interest rate, also called the short rate.

Assumption 2: There exists n risky securities with spot prices denoted as, $S_1, ..., S_n$. The spot prices evolves as continuous semi-martingales over the finite interval $[0, T]$. Assume there are no dividends and no costs of holding the securities.

Assumption 3: There is no arbitrage on the market specified in Assumption 1 and Assumption 2.

Assumptions 1 through 3 guarantees the existence of a martingale measure $\mathbb{Q}$, equivalent to $\mathbb{F}$, such that the security prices on the financial market discounted by the risk free rate, $e^{-\int_0^t r_s ds} S_{it}$, are martingales under $\mathbb{Q}$. This is the first fundamental theorem of mathematical finance (see Theorem 1.1 in (Delbaen & Schachermayer, 1994)).

The $\mathbb{F}$-dynamics of the numeraire portfolio will now be derived. We mentioned that the discounted spot prices were martingales under $\mathbb{Q}$ so the martingale property yields,

$$\mathbb{E}^{\mathbb{Q}}\Big[\frac{S_{iT}}{S_{0T}}|\mathcal{F}_t\Big] = \frac{S_{it}}{S_{0t}}, \quad t \in [0, T], \quad i = 0, 1, \dots, n. \tag{3.1}$$

The previous expectation can be reformulated as an expectation under $\mathbb{F}$. Let M be the likelihood process used in creating $\mathbb{Q}$, then the following is true,

$$\mathbb{E}^{\mathbb{F}}\Big[\frac{M_T}{M_t}\frac{S_{iT}}{S_{0T}}|\mathcal{F}_t\Big] = \frac{S_{it}}{S_{0t}}, \quad t \in [0, T], \quad i = 0, 1, \dots, n. \tag{3.2}$$

See Proposition 10.21 in (Björk, 2009). It is useful to make the following definition,

$$L_t \equiv \frac{S_{0t}}{M_t}, \quad t \in [0, T]. \tag{3.3}$$

Notice that L_t is positive and grows at the risk free rate in $\mathbb{Q}$-expectation since $\frac{1}{M_t}$ is positive and is a $\mathbb{Q}$-martingale (see Proposition C.13 in (Björk, 2009)). Hence, L is the value of some self-financing portfolio. By multiplying Eq. 3.2 by M_t the follwing is obtained

$$\mathbb{E}^{\mathbb{F}}\left[\frac{S_{iT}}{L_T}|\mathcal{F}_t\right] = \frac{S_{it}}{L_t}, \quad t \in [0,T], \quad i = 0,1,\ldots,n. \tag{3.4}$$

Consequently we see that L is the value of the numeraire portfolio. Thus by choosing L as the numeraire, $\mathbb{F}$ becomes a martingale measure. Now set $i = 0$ we obviously have,

$$\mathbb{E}^{\mathbb{F}}\left[\frac{S_{0T}}{L_T}|\mathcal{F}_t\right] = \frac{S_{0t}}{L_t}, \quad t \in [0,T]. \tag{3.5}$$

Assuming that L is a continuous semi-martingale with lognormal volatility, σ_t, we have

$$\frac{d(S_0/L)}{S_0/L} = -\sigma_t dW_t^{\mathbb{F}}, \quad t \in [0,T],$$

where $W_t^{\mathbb{F}}$ is an $\mathbb{F}$ brownian motion. Itô's formula applied to $\frac{L}{S_0}$ yields,

$$\frac{d(L/S_0)}{L/S_0} = \sigma_t^2 dt + \sigma_t dW_t^{\mathbb{F}}, \quad t \in [0,T].$$

Removing the discounting we thus see that the $\mathbb{F}$ dynamics of L_t is given by,

$$\frac{dL_t}{L_t} = (r_t + \sigma_t^2)dt + \sigma_t dW_t^{\mathbb{F}}, \quad t \in [0,T]. \tag{3.6}$$

Note that the risk premium for the numeraire portfolio is its instantaneous variance. The market price of $W_t^{\mathbb{F}}$-risk is thus σ_t (see section 14.6 in (Björk, 2009)). This result is central to the solution of the recovery problem. By a no arbitrage argument all derivatives on the market, complete or incomplete, driven solely by $W_t^{\mathbb{F}}$ are related in that they have the same market price of risk (see Proposition 15.1 in (Björk, 2009)). If we can estimate the volatility of the numeraire portfolio we also know the real world dynamics of any derivative on this market, provided we have its volatility which is the same under $\mathbb{Q}$ as under $\mathbb{F}$. With this consideration in mind we assume the following connection between the assets on the market,

Assumption 4: There exists a univariate time-homogeneous bounded diffusion process X such that $S_{it} = S_i(X_t, t)$, $i = 0, 1, \ldots, n$, where $S_i(x,t) : [l,u] \times [0,T] \to \mathbb{R}$.

Note that according to the Meta-theorem 8.3.1 in (Björk, 2009) the defined market is now complete. Furthermore, we should note that the assumption that X is bounded is not necessary but merely sufficient in order for recovery to work. This is of course a major restriction in the general case but with our present interpretations of X (VIX and the short rate) it is not an unnatural assumption.We also map the function $S_0(x,t)$ into another function $r(x,t) = \frac{\partial}{\partial x}\ln S_0(x,t)$.

The driver, X, evolves as a continuous bounded time-homogeneous diffusion under $\mathbb{Q}$ so there exists a $\mathbb{Q}$ brownian motion, $W^{\mathbb{Q}}$, a drift function $b(x), x \in [l,u]$ and variance function $a^2(x), x \in [l,u]$ such that X solves,

$$dX_t = b(X_t)dt + a(X_t)dW_t^{\mathbb{Q}}, \quad t \geqslant 0.$$

We have the following generator,

$$\mathcal{G}_{xt} = \frac{\partial}{\partial t} + \frac{a^2(x)}{2}\frac{\partial^2}{\partial x^2} + b(x)\frac{\partial}{\partial x}. \tag{3.7}$$

The risky securities satisfies the linear parabolic partial differential equation,

$$\mathcal{G}_{xt}S_i(x,t) = r(x,t)S_i(x,t), \quad x \in [l,u], \quad t \in [0,T]. \tag{3.8}$$

This is a fact we will use in the derivation in the next section.

Assumption 5: The functions $r(x,t), b(x)$ and $a^2(x)$ are known ex ante. $a(x)$ is positive on (l,u).

We will comment on how to handle this assumption in the methodology section below.

Assumption 6: Assume,

$$L_t \equiv L(X_t,t), \quad x \in [l,u], \quad t \in [0,T], \tag{3.9}$$

where $L(x,t)$ is a positive function. Also assume that r_t depends only on the driver X and not on time t,

$$r(x,t) = r(x), \quad x \in [l,u], \quad t \in [0,T]. \tag{3.10}$$

We finally assume that the volatility , σ, of L depends on X only. Thus we have,

$$\frac{dL_t}{L_t} = r(X_t)dt + \sigma(X_t)dW_t^{\mathbb{Q}}, \quad t \in [0,T]. \tag{3.11}$$

We will now go on to the next section and show that under these assumptions it is actually possible to find out what $\sigma(x)$ looks like. This implies that the market price of risk is determined and we can appeal to Girsanov's theorem in changing the probability measure from $\mathbb{Q}$ to $\mathbb{F}$ and thus obtain the dynamics of the driver X and all the spot prices S_i under $\mathbb{F}$.

3.2.2 Deriving the result

First we derive an expression for $\sigma(x)$ in terms of $L(x,t)$ by using Assumption 6. Then we will determine $L(x,t)$, and the market price of risk is completely known.

Itô's formula applied to Eq. 3.9 with derivative terms evaluated in (X_t,t) yields,

$$dL_t = \left[\frac{\partial L}{\partial t} + b(X_t)\frac{\partial L}{\partial x} + \frac{a^2(X_t)}{2}\frac{\partial^2 L}{\partial x^2}\right]dt + a(X_t)\frac{\partial L}{\partial x}dW_t^Q. \tag{3.12}$$

Setting the diffusion coefficients in the previous equation equal to the diffusion coefficient in Eq. 3.11 we obtain,

$$\begin{aligned}\sigma(x)L(x,t) &\equiv a(x)\frac{\partial L}{\partial x}(x,t) \Leftrightarrow \\ \sigma(x) &\equiv a(x)\frac{1}{L(x,t)}\frac{\partial L}{\partial x}(x,t) = a(x)\frac{\partial}{\partial x}\ln L(x,t).\end{aligned} \tag{3.13}$$

Because $a(x) > 0$ on (l,u) we can without problems divide by it in the previous equation. Then integrate w.r.t. x to get,

$$\ln L(x,t) = \int^x \frac{\sigma(\gamma)}{a(\gamma)}d\gamma + f(t).$$

Hence, after exponentiation it is clear that the value function of the numeraire portfolio separates multiplicatively into two functions dependent on x and t respectively positive on (l,u) and $t \geqslant 0$,

$$L(x,t) = \pi(x)p(t).$$

We know from the self-financing condition on the numeraire portfolio it must satisfy the PDE given in Eq. 3.8 so we have,

$$\pi(x)p'(t) + \frac{a^2(x)}{2}\pi''(x)p(t) + b(x)\pi'(x)p(t) = r(x)\pi(x)p(t).$$

Rearranging this equation gives,

$$\frac{a^2(x)}{2}\frac{\pi''(x)}{\pi(x)} + b(x)\frac{\pi'(x)}{\pi(x)} - r(x) = -\frac{p'(t)}{p(t)}, \quad x \in (l,u), \quad t \in [0,T].$$

The equality can only be true if both sides are equal some constant, say $-\lambda \in \mathbb{R}$. This yields two ordinary differential equations,

$$\frac{p'(t;\lambda)}{p(t;\lambda)} = \lambda, \quad t \in [0,T] \tag{3.14}$$

with solution,

$$p(t;\lambda) = p(0;\lambda)e^{\lambda t} \propto e^{\lambda t},$$

and

$$\frac{a^2(x)}{2}\pi''(x;\lambda) + b(x)\pi'(x;\lambda) - r(x)\pi(x;\lambda) = -\lambda\pi(x;\lambda), \quad x \in (l,u). \tag{3.15}$$

As far as boundary conditions go for this ODE we have assumed that the driver X is bounded. This implies some restrictions on the domain of the infinitesimal generator in Eq. 3.7. The eigenfunction π has to obey these restrictions. If one allows boundry conditions to be separated, that is of the form,

$$\begin{aligned} A\pi(l) - B\pi'(l) = 0, \quad A^2 + B^2 > 0 \\ C\pi(u) + D\pi'(u) = 0, \quad C^2 + D^2 > 0, \end{aligned}$$

we obtain a regular Strum-Liouville problem (see section 3.1). Depending on the behaviour of the driver X near it's boundaries l, u the boundary conditions can be determined, that is the coefficients A, B, C and D are specified. For example, to enforce a completely reflexive behaviour of X we set $A = 0$ and $C = 0$ while setting $B = 0$ and $D = 0$ enforces a killing behaviour at the boundaries (see Item 7, Chapter 2 in (Borodin & Salminen, 2002)).

We denote the principal solution by $\rho, \phi(x)$, that is the smallest eigenvalue λ for which there exist an eigenfunction $\pi(x)$ which solves Eq. 3.15 and it's corresponding eigenfunction. From section 3.1 we know that this principal solution is unique (up to positive scaling in $\phi(x)$) and it is in fact the only solution which yields a positive value function, $L(x,t)$, on the entire interval (l,u) because all other eigenfunctions contain at least one zero in the interval (l,u). Thus it is only this principal solution we are interested in since the numeraire portfolio must be positive. The principal solution can be found with numerical methods but in this study it will be found analytically.

Once the principal solution is found the expression for $L(x,t)$ is,

$$L(x,t) = \phi(x)e^{\rho t}, \quad x \in (l,u), \quad t \in [0,T], \tag{3.16}$$

which is determined up to a scaling constant due to the factor in $\phi(x)$. We will see that the scale factor will cause no problem when finding the real world density.

Inserting the right hand side of Eq. 3.16 into right hand side of Eq. 3.13 the following expression for the function $\sigma(x)$ is obtained,

$$\sigma(x) = a(x)\frac{\partial}{\partial x}(\ln\phi(x) + \rho t) = a(x)\frac{\partial}{\partial x}\ln\phi(x). \tag{3.17}$$

Hence the market price of risk is determined. By appealing to Girsanov's theorem (see Theorem 11.3 in (Björk, 2009)) the $\mathbb{F}$-dynamics of X_t is given by,

$$dX_t = [b(X_t) + \sigma(X_t)a(X_t)]dt + a(X_t)dW_t^{\mathbb{F}}, \quad t \geqslant 0. \tag{3.18}$$

The $\mathbb{F}$-dynamics of the spot prices S_{it} can also be determined. Recall that in Assumption 4 we assumed that $S_{it} = S_i(X_t, t)$ so using Itô's lemma we have,

$$dS_{it} = \Big[\frac{\partial S_{it}}{\partial t} + b(X_t)\frac{S_{it}}{\partial x} + \frac{a^2(X_t)}{2}\frac{\partial^2 S_{it}}{\partial x}\Big]dt + a(X_t)\frac{\partial S_{it}}{\partial x}dW_t^{\mathbb{Q}}.$$

In an arbitrage free market the drift is equal to $r(X_t)S_{it}(X_t, t)$ under $\mathbb{Q}$ so we obtain after a change of measure,

$$dS_{it} = [r(X_t)S_{it}(X_t, t) + \sigma(X_t)a(X_t)\frac{\partial S_{it}}{\partial x}]dt + a(X_t)\frac{\partial S_{it}}{\partial x}dW_t^{\mathbb{F}}. \tag{3.19}$$

Now the real world probability density, $d\mathbb{F}$, can be determined in the following way. The change of numeraire theorem (Géman, El Karoui, & Rochet, 1995) states that the Radon-Nikodym derivative is given by,

$$\frac{d\mathbb{F}}{d\mathbb{Q}} = \frac{S_{0,0}}{S_{0,T}}\frac{L_T}{L_0} = e^{-\int_0^T r(X_t)dt}\frac{L(X_T, T)}{L(X_0, 0)}.$$

Inserting the right hand side of Eq. 3.16 into the previous equation yields,

$$\frac{d\mathbb{F}}{d\mathbb{Q}} = e^{-\int_0^T r(X_t)dt}\frac{\phi(X_T)}{\phi(X_0)}e^{\rho T}.$$

Solving for the real world density we obtain,

$$d\mathbb{F} = e^{-\int_0^T r(X_t)dt}\frac{\phi(X_T)}{\phi(X_0)}e^{\rho T}d\mathbb{Q}. \tag{3.20}$$

All the quantities on the right hand side are known and notice that the unknown positive scaling factor in $\phi(x)$ mentioned previously cancels out thus the real world probability density are completely determined.

3.3 Theoretical implications of constant short rate

The theory above makes no assumptions about the driving process X, apart from it being a univariate bounded diffusion process. On the other hand it makes an assumption about the short rate, namely that all stochastic variation is determined by the driver X. There exist many candidate processes which real world probability measure we wish to recover but at the same time do not effect the short rate. In order to facilitate that in the theory above we have to assume that the short rate is constant. Thus it is interesting to see what the theory above implies when the short rate is constant i.e. $r(x) = r$ in Eq. 3.10. We obtain the following equation,

$$\begin{aligned} &\frac{a^2(x)}{2}\frac{d^2\pi}{dx^2} + b(x)\frac{d\pi}{dx} = (r-\lambda)\pi, \quad x \in [l,u], \\ &A\pi(l) - B\pi'(l) = 0, \quad C\pi(u) + D\pi'(u) = 0. \end{aligned} \tag{3.21}$$

We are only interested in the smallest value of λ for which there exist a non trivial real valued solution to Eq. 3.21, previously called ρ, and its corresponding eigenfunction, $\phi(x)$. We can thus proceed by using the fact that the equation above can be transformed into an equivalent self-adjoint form (see (Carr & Yu, 2012)). We obtain the following equation,

$$\begin{aligned} &\frac{d}{dx}\Big(p(x)\frac{d\pi}{dx}\Big) - q(x)\pi + w(x)(\lambda - r)\pi = 0 \\ &A\pi(l) - B\pi'(l) = 0, \quad C\pi(u) + D\pi'(u) = 0. \end{aligned} \tag{3.22}$$

where $p(x) = e^{\int_l^x \frac{2b(\gamma)}{a^2(\gamma)}d\gamma}$, $q(x) = 0$, $w(x) = \frac{2p(x)}{a^2(x)}$. Since $p(x)$ and $w(x)$ are positive and $q(x)$ is nonnegative on the interval $[l,u]$ we know that every eigenvalue in Eq. 3.21 is nonnegative (See Theorem 1 in Section 3.1), that is every λ_i is such that $r - \lambda_i \geqslant 0$. We can immediately see from Eq. 3.21 that, with an ansatz $r - \rho = 0$, a solution for the corresponding eigenfunction is any constant function, call it c. By Sturm-Liouville theory we know that this solution is unique up to a scaling constant, denote this constant by K. Hence we have found the principal solution of 3.21 with enough precision to conclude the following by insertion into Eq. 3.20,

$$d\mathbb{F} = e^{-\int_0^T r dt} e^{rT}\frac{Kc}{Kc}d\mathbb{Q} = e^{-rT}e^{rT}\frac{Kc}{Kc}d\mathbb{Q} = d\mathbb{Q}. \tag{3.23}$$

In other words in the case of a constant short rate of interest the above theory implies that the real world probability measure is identical to the risk adjusted measure. This means that the pricing of the assets specified in the model are risk-neutral i.e. the risk premiums are zero.

3.4 Recovery of the short rate

The previous subsection shows that we are not able to pick any asset, make the standard assumption of constant interest rate and obtain a non zero risk premium. In this subsection I proceed by considering the short rate as the underlying driver. In this case the relationship between the short rate and the underlying driver is precisely $r(x) = x$ (see Eq. 3.10). Making the assumption that the term structure of bonds are driven solely by the short rate, the bonds play the role of $S_i(X_t, t)$ in Assumption 4 above.

One of the most popular stochastic models for the risk neutral dynamics of the short rate is the Cox-Ingersoll-Ross model introduced in (Cox et al., 1985). The SDE is given by

$$dX_t = \kappa(\theta - X_t)dt + s\sqrt{X_t}dW_t^{\mathbb{Q}}, \tag{3.24}$$

where $\theta, s > 0, \kappa \in \mathbb{R}$ and B is a standard Brownian motion under the risk neutral measure. Also set $\gamma = \sqrt{\kappa^2 + 2s^2}$. The Feller condition, $2\kappa\theta \geqslant s^2$ is also assumed. The following expectation is well known (see for example (Durfesne, 2001)),

$$\mathbb{E}[X_T|\mathcal{F}_t] = \theta + (X_t - \theta)e^{-\kappa(T-t)}. \tag{3.25}$$

Zero coupon bond prices are given by

$$S^{CIR}(X_t, t, T) = A(t, T)e^{-B(t,T)X_t}, \tag{3.26}$$

$$A(t, T) = \Big[\frac{2\gamma e^{\kappa+\gamma((T-t)/2)}}{(\gamma+\kappa)(e^{\gamma(T-t)}-1)+2\gamma}\Big]^{2\kappa\theta/s^2}, \tag{3.27}$$

$$B(t, T) = \frac{2(e^{\gamma(T-t)}-1)}{(\gamma+\kappa)(e^{\gamma(T-t)}-1)+2\gamma}. \tag{3.28}$$

The Sturm Liouville equation becomes,

$$\frac{1}{2}s^2 x\pi'' + (\kappa\theta - \kappa x)\pi' - x\pi = -\lambda\pi, \tag{3.29}$$

with $\lambda \in R$. As alluded to previously boundary conditions are not needed but only sufficient. In the above equation the Feller condition assumed above imposes a natural boundary at zero and in (Qin & Linetsky, 2014) it is shown that this is enough to obtain a unique principal solution with a positive eigenfunction. The above equation can be transformed to a *confluent hypergeometric equation* (see (Slater, 1960)) which solutions are characterized by linear combinations of *Kummer* and *Tricomi* functions (see (Abramowitz & Stegun, 1964)).

The set of solutions to Eq. 3.29 are given in Proposition 9 (i) in (Davydov & Linetsky, 2003). For our purposes only the solution for the smallest eigenvalue is needed since that is the only solution which gives a positive eigenfunction on the entire support. We have,

$$\rho = \frac{\kappa\theta}{2}(\gamma - \kappa),$$
$$\phi(x) = e^{-\frac{(\gamma-\kappa)}{s^2}x}.$$

From Eq. 3.17 and Eq. 3.18 the instantaneous risk premia for the short rate becomes,

$$\sigma(X_t)a(X_t) = -X_t(\gamma - \kappa).$$

The magnitude of the risk premia increases monotonically in both κ and s and the higher κ the less effect an increasing s has.

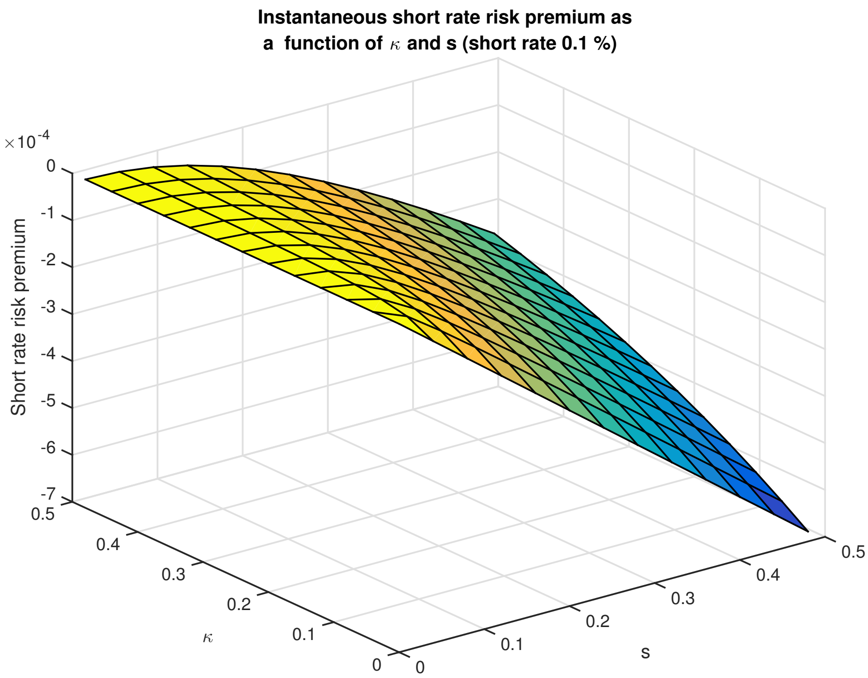


Figure 3.1: Instantaneous risk premia for the short rate as a function of κ and s.

The dynamics of the short rate under the $\mathbb{F}$-measure can now be recovered as,

$$dX_t = \gamma(\frac{\kappa\theta}{\gamma} - X_t)dt + s\sqrt{X_t}dW_t^{\mathbb{F}}$$

i.e. a CIR process with mean reversion γ and long term mean $\frac{\kappa\theta}{\gamma}$.

Chapter 4

Data & Methodology

4.1 Data

To investigate the theory with regards to VIX the VIX Index values, calculated by the Chicago Board of Options Exchange (CBOE), between February 2010 to August 2015 are used. This time series is taken from the Board of Governors of the Federal Reserve System database.[1]. The data set used in the empirical investigation for short rate recovery is the Treasury Constant Maturity Rates. These are also provided by the Board of Governors of the Federal Reserve System[2]. Four time series of different maturities are used, 1,3,6 and 12 months. Each time series is of daily frequency. This data also spans over the period February 2010 to August 2015. The estimation methodology used by the Federal Reserve to construct the yield data is a quasi-cubic hermite spline model which input is primarily yields for on-the-run U.S Treasuries. According to the U.S treasury on-the-run treasuries usually trades close to par so when these data points are used as knot points in the cubic spline model the yield curve is to be considered a par yield curve. Furthermore, the coupons are paid with semi-annual periodicity.[3] All rates are converted to continuously compounded rates. The day count convention for all US Treasury securities are based on actual day counts on a 365- or 366-day year. [4]

To transform the par yield curve into a zero coupon yield curve the Matlab®-function `pyld2zero()` is used. The function uses the bootstrap method to obtain the zero coupon rates for all maturities (see (Fabozzi, 2005) for more on the bootstrapping method). The zero coupon yield curve is shown in Figure 4.1.

[1] Can be found at the following webpage https://research.stlouisfed.org/fred2/series/VIXCLS

[2] Can be found at the following webpage https://research.stlouisfed.org/fred2/categories/115

[3] http://www.treasury.gov/resource-center/data-chart-center/interest-rates/Pages/yieldmethod.aspx

[4] http://www.treasury.gov/resource-center/faqs/Interest-Rates/Pages/faq.aspx#1

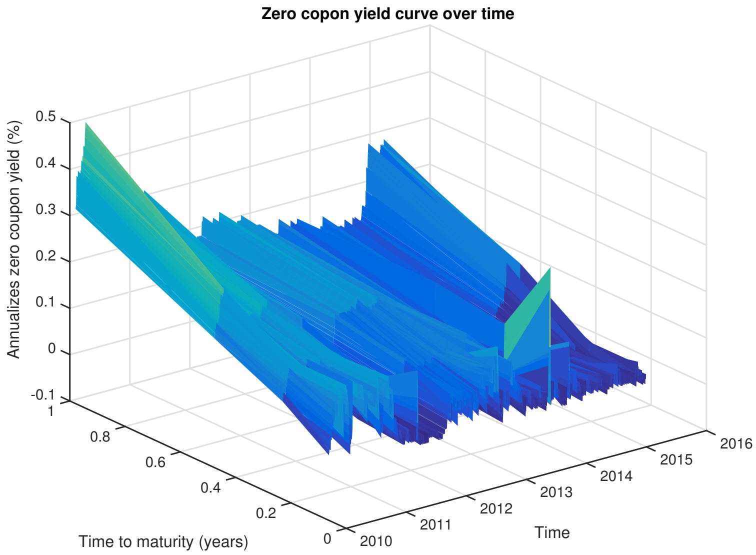


Figure 4.1: Time series and cross-sectional evolution of the zero coupon yield curve for maturities 1 to 12 months.

4.2 Relation between VIX and the short rate

A priori it is unclear what the relationship between the short rate and VIX should be. The hypothesis used is therefore that they are unrelated. This hypothesis will simply be investigated visually in a scatter plot and by a linear regression of the first difference of the 1 month zero coupon Treasury bill rate onto the percentage change in the VIX Index over the sample period.

4.3 Short rate recovery

4.3.1 Calibration

To extract the implied CIR parameters under the risk neutral measure the sum of the squared differences between the CIR term structure computed by Eq. 3.26 and the observed zero coupon bond price curve is minimized. Formally we solve the following optimization problem,

$$\begin{aligned} \underset{r,\kappa,\theta,s}{\text{minimize}} \quad & \sum_{i=1}^{4}\left(S^{CIR}(r,t,T_i) - S^{obs}(t,T_i)\right)^2, \\ \text{subject to} \quad & 2\kappa\theta \geqslant s^2, \\ & \kappa,\theta,s \geqslant 0, \end{aligned}$$

for each date in our sample period. Notice that we consider the short rate, r, as a latent state variable which we also calibrate in the optimization process. The parameters are recalibrated each day yielding new parameter estimates. Except for the parameter constraints this is the same minimization as in (S. J. Brown & Dybvig, 1986). To carry out the optimization the global optimizer provided by Matlab® named `patternsearch()` is used.

4.3.2 Principal Component Analysis

Principal component analysis (PCA) will be used in checking the calibration results. The main intuition will be stated in this section, for a more complete account of PCA see (Jolliffe, 2002). The changes in calibrated parameters should be caused by changes in the yield curve. As pointed out in (R. H. Brown & Schaefer, 1994) and mentioned in section 2.2 the objective function is very flat in some directions of the parameter space and this could lead to unstable parameter calibration results. Therefore PCA is used to check the substance in the calibrated parameters. In order to run a PCA one needs the matrix of measurements, call it $\mathbf{X}$ with dimension $n \times p$. In the present case this is the matrix with yields of the different maturities in the columns and each row correspond to a different date. When computing the principal components this matrix is centred to have zero mean, $\mathbf{X_0}$. The covariance matrix of $\mathbf{X_0}$ is then computed. The eigenvectors of the covariance matrix is placed column wise in the matrix $\mathbf{A}$ of dimension $p \times p$. The matrix $\mathbf{A}$ is orthogonal so $\mathbf{A}^{-1} = \mathbf{A}'$, where $'$ means transpose. These eigenvectors are called principal components (PC) and corresponding to each PC there is an eigenvalue. Normalizing each eigenvalue with the sum of the eigenvalues gives the percentage of the explained variance from each PC. The principal component scores, $\mathbf{Z}$, can be define as follows,

$$\mathbf{Z} = \mathbf{X_0 A}.$$

Inverting the above equation the following is obtained,

$$\mathbf{ZA}' = \mathbf{X_0}.$$

By the above equation it is evident that the scores in each column of $\mathbf{Z}$ can be interpreted as the influence of the respective components on the yield curve. Thus a change in the principal component score means that a change in the shape of the yield curve is observed.

The calibration results will be checked by running multiple linear regressions of the first difference (time series wise) of the calibrated parameters on the first difference of the principal component scores. It is desirable to see some relation between the quantities to ensure that a change in the calibrated parameter is caused by a fundamental change in the yield curve shape.

4.3.3 Prediction

The model will be evaluated based on its power to predict the changes in the TBR1. The predictions are based on the assumption that it can be used as a proxy for the short rate which is unobservable. As mentioned previously a study which uses the TBR1 as a short rate proxy is (Chan et al., 1992). The methodology presented previously allows us to recover the short rate process under $\mathbb{F}$, thus this will be the process used in forecasting changes in the TBR1. Using Eq. 3.25 the prediction of the change in the short rate from today until a specified future date is given by

$$\hat{y}^1_{t,T} = \mathbb{E}^{\mathbb{F}}[r_T - r_t|\mathcal{F}_t] = (r_t - \frac{\kappa\theta}{\gamma})(e^{-\gamma(T-t)} - 1). \tag{4.1}$$

Since $\gamma \geqslant 0$, it means the second term above will be less than 0. Thus if r_t is less than the long term mean, $\frac{\kappa\theta}{\gamma}$, the prediction will be of a positive change in the short rate. If r_t is greater than the long term mean the prediction will be that a negative change will occur.

The predictions of the recovered process will be benchmarked to predictions from two other processes. The first one is the CIR process with the calibrated risk neutral parameters. Risk neutral prediction with the CIR process is for example studied in (Hin & Dokuchaev, 2015). This comparison will give a sense of the gain from the recovery step since the prediction error corresponding to errors in the calibration step is present in the prediction errors of both processes. The prediction equation is given by,

$$\hat{y}^2_{t,T} = \mathbb{E}^{\mathbb{Q}}[r_T - r_t|\mathcal{F}_t] = (r_t - \theta)(e^{-\kappa(T-t)} - 1). \tag{4.2}$$

The logic is the same as before regarding the sign of predicted change in short rate.

The second benchmark process is a naive predictor which guess of future values is simply the present observed value. Such a process is called a martingale. This comparison gives a sense of the added value of the entire procedure in which the calibration step is included. By using the martingale property of the process the prediction equation is obtained as,

$$\hat{y}^3_{t,T} = \mathbb{E}^{\mathbb{F}}[r_T - r_t|\mathcal{F}_t] = 0. \tag{4.3}$$

As in (Hin & Dokuchaev, 2015) predictions on several horizons are made. The horizons range from 1 to 52 weeks with one week apart. So on each date in our sample 52 different forecasts are made. The number 52 is chosen with respect to the result in (Hin & Dokuchaev, 2015) that predictions on longer horizons than maturities of bonds used to calibrate the CIR model seems uninformative.

Measuring the statistical significance of the difference in the predictors the Diebold-Mariano test is used (see (Diebold & Mariano, 1995)). This test is widely used in studies evaluating forecasts, for example (Diebold & Li, 2006). Denote the change in the TBR1 between t and T by $y_{t,T}$. The prediction error for the i:th predictor for horizon $T-t$ is then $e^i_{t,T} = \hat{y}^i_{t,T} - y_{t,T}$. The test uses a loss function, in our case quadratic loss $L(e^i_{t,T}) = (e^i_{t,T})^2$, and states the null hypothesis as $\mathbb{E}[L(e^i_{t,T})] = \mathbb{E}[L(e^j_{t,T})]$. The test allows the forecast errors to have non-zero mean, be non-Gaussian, serially correlated and contemporaneously correlated (correlation between model forecasts). This is especially important in the test between recovered and risk-neutral forecasts since they use the same information set to make predicitons over the same sample period so they are most likely correlated. The fact that the test takes into account the serial correlations of the prediction errors is also important since such serial correlations are common in multi step forecasts as noted in (Diebold & Mariano, 1995).

Chapter 5

Results

5.1 VIX and the short rate

In Figure 5.1 the changes in the TBR1 is plotted against the percentage changes in the VIX Index. Here the TBR1 is assumed to be a good proxy for the short rate. There does not seem to be any particular relation between the two variables. When regressing the changes in TBR1 onto the percentage changes in VIX the p-value for the estimated regression coefficient turns out to be 0.22 and 0.94 for the intercept. The null hypothesis that the intercept and the regression coefficient is equal to zero can not be rejected as a result.

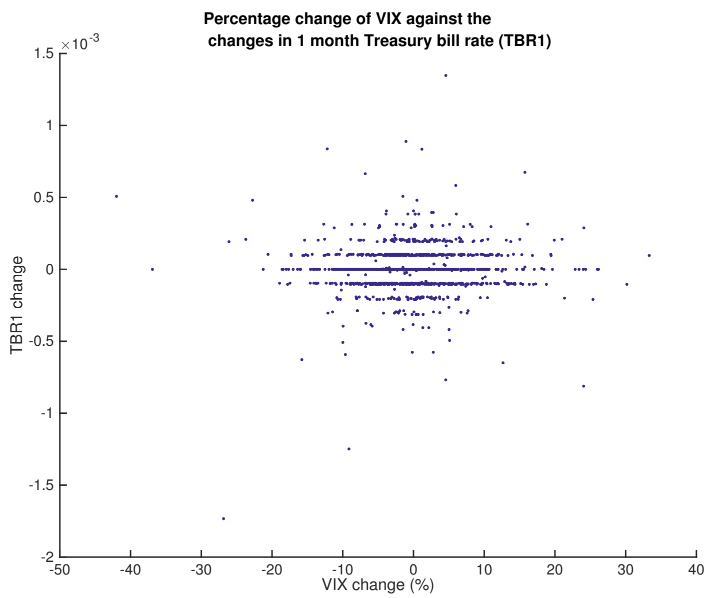


Figure 5.1: Changes in TBR1 against VIX percentage changes over the sample period.

5.2 Short rate recovery

5.2.1 Calibration

In Figure 5.2 the pricing errors of the CIR calibtration are shown. The average calibration percentage error over the sample period is approximately 3.88×10^{-5} and the standard deviation is 2.55×10^{-5}. To judge the error of the fit, this error will be compared to the average bid ask spread of the US Treasury bills. In the model the assumption of a single price is made and the fact that in the real market only the bid and ask prices are observed and this discrepancy is not accounted for. The "real" price is somewhere inside this spread. Since the fit is not made to this "real" price, which is the price that is modelled, there is limited value in obtaining a fit much less than the bid-ask spread.

In (Flemming, 2003) the mean bid-ask spread in the yields for US Treasury securities are given. For the US Treasury bills with 3, 6 and 12 months to maturity they are $0.71, 0.74$ and 0.52 basis points. To convert the spreads into price units the the relation between yield and price can be used,

$$y = -\frac{\log p(t,T)}{T-t}.$$

Using the above formula the bid-ask spread in the yields can be expressed in terms of bid-ask gross percentage price spread,

$$y^{bid} - y^{ask} = -\frac{1}{T-t} \log \frac{p^{bid}(t,T)}{p^{ask}(t,T)}.$$

Inverting the above expression and converting to net percentage the following expression is obtained,

$$\frac{p^{bid}(t,T)}{p^{ask}(t,T)} - 1 = e^{-(y^{bid}-y^{ask})(T-t)} - 1.$$

Evaluating the above expression with the numbers given in (Flemming, 2003) the bid-ask percentage price spread for the 3, 6 and 12 months bonds becomes $1.78 \times 10^{-5}, 3.70 \times 10^{-5}, 5.20 \times 10^{-5}$. Thus the average spread is 3.56×10^{-5}. The CIR calibration error is 9% bigger in magnitude than this average bid-ask spread. This is deemed to be a satisfying fit.

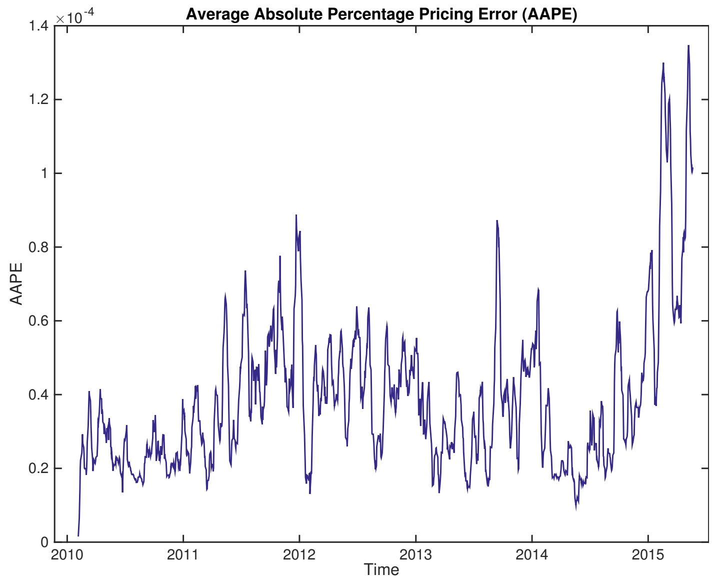


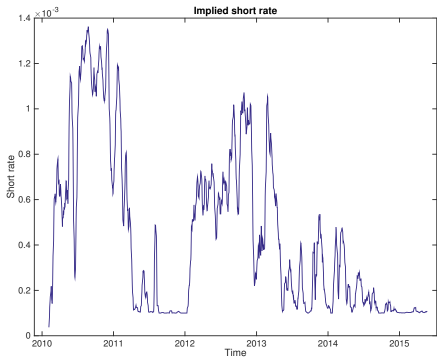
Figure 5.2: Average Absolute Percentage Error (AAPE)

In Figure 5.3 the trajectories of the calibrated parameters are shown and the means and standard deviations of the parameter estimates are given in Table 5.1. In the upper left plot the implied short rate process is shown. The other plots show the CIR process parameter estimates. The estimates are far from constant as the CIR model suggests, this is consistent with results in (R. H. Brown & Schaefer, 1994). This does not have to be viewed as problematic for the method however because this can be seen as the way that the method absorbs eventual non-stationarity in the true short rate process. This is the virtue of calibrating the process cross sectionally.

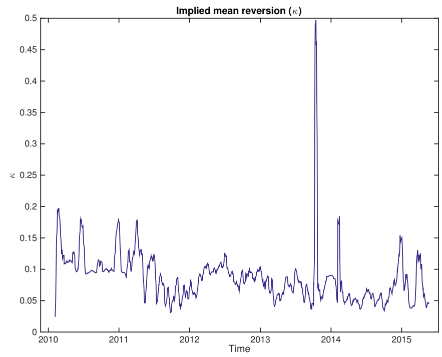
Another source of instability could stem from insensitivity of the objective function value with respect to changes in parameter values. Since a numerical optimization algorithm is used with some termination criteria, we will end up with an approximate solution. If the objective function is very insensitive to changes in any of the parameters then a sufficiently "good" objective function value can be obtained with widely varying parameter estimates. This is depicted in Figure 5.4. The plot depicts the objective function value when the parameters are scaled 50% up and down from optimum. It is evident that the most unreliable parameter is the s parameter. Its surface is almost completely flat over the entire sample period compared to the other parameters. The objective function show some sensitivity to short rate changes in the first part of the sample period but in the last two years its surface also becomes flat. The objective function response is generally greater for changes in κ and θ.

	Short rate	κ	θ	s
Sample mean	4.41×10^{-4}	0.088	0.038	0.030
Standard deviation	3.77×10^{-4}	0.059	0.0167	0.037

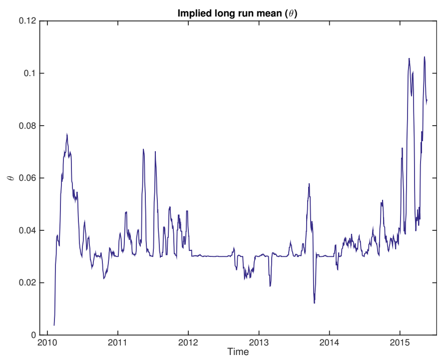
Table 5.1: Time series mean and standard deviation for the calibrated parameters.



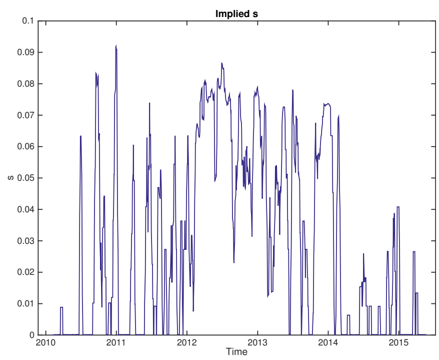
(a) Short rate



(b) κ

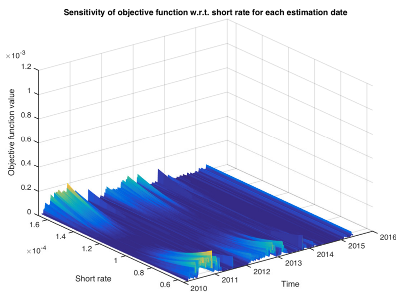


(c) θ

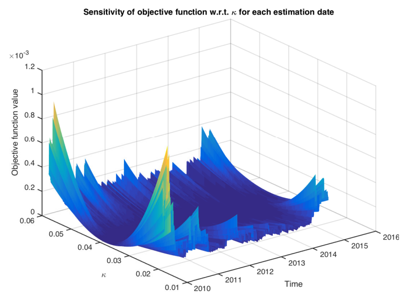


(d) s

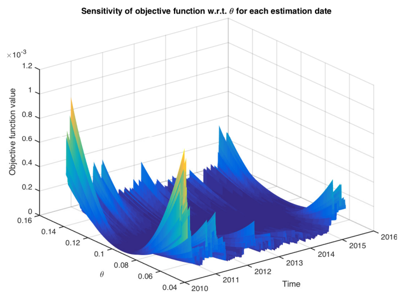
Figure 5.3: Evolution of implied short rate and calibrated CIR parameters.



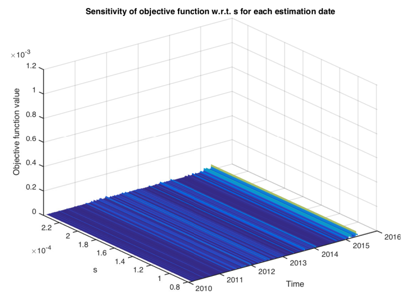
(a) Short rate



(b) κ



(c) θ



(d) s

Figure 5.4: Sensitivity of the objective function at optimal point w.r.t. short rate and CIR parameters. In each plot the the parameter of study is varied $+/-50\%$ *from optimum in the parameter-axis while the other parameters are fixed at optimal values.*

In Figure 5.5 the difference between the recovered mean reversion parameter and the risk neutral mean reversion parameter is shown. The difference is positive over the entire sample which means that a higher degree of mean reversion is exhibited in the recovered process than for the risk neutral process. With this in mind, comparing the long term mean under $\mathbb{F}$ and $\mathbb{Q}$ one sees that the long term mean under $\mathbb{F}$ will be less than the long term mean under $\mathbb{Q}$.

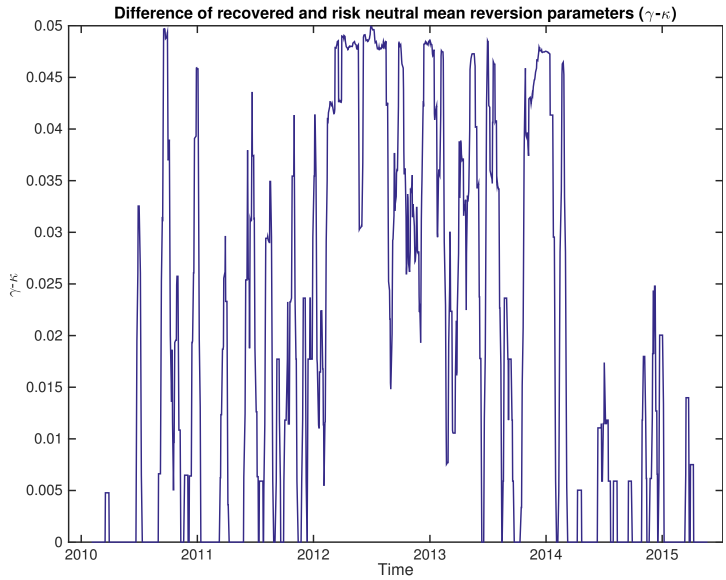


Figure 5.5: Difference between recovered- (γ) and risk neutral mean reversion (κ).

5.2.2 Correspondence with principal components

To assess the substance in the calibration of the parameters multiple linear regressions are performed where the first difference of the parameter estimate act as a dependent variable and the first difference of the three most dominant principal component scores act as independent variables. The principal components are extracted from the yield curve data. If a variable shows to be insignificant it is omitted and the regression is reran. In Figure 5.6 the component eigenvectors are plotted. It is clearly feasible to interpret changes in the scores corresponding to these components as changes in the level, slope and curvature of the yield curve. This interpretation is standard in the yield curve literature. Shifts in these risk factors explain 99.44% of the yield curve variation in the present sample. Even though we acknowledge that there is not a linear relation between the parameters and the principal component scores it would be reassuring to see some correspondence between the changes in risk factors and changes in the parameter estimates. This is because changes in parameter estimates should be caused by changes in yield curve shape. The R^2 will be dampened due to the assumed linear relation although it is nonlinear.

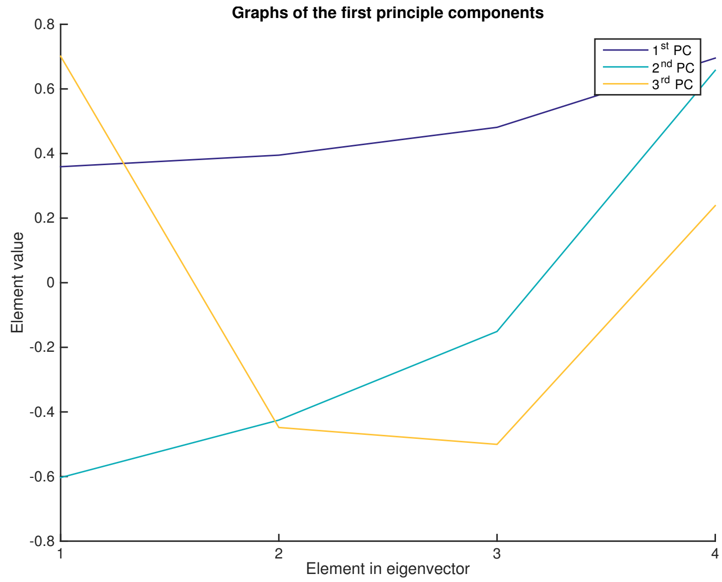


Figure 5.6: First three principal components extracted from yield curve.

In Table 5.2 the regression results are shown. The short rate is the variable that is the most related to shifts in the risk factors. Roughly 50% of its variance can be explained by parallel shifts and steepening/flattening of the yield curve. An upward parallel shift lead to an increasing short rate which is an expected effect. A steepening of the yield curve lead to a decreasing short rate. The intuition is that in order for the level of the yield curve to be unchanged a steepening of the yield curve has to be caused by the short maturity yields tipping down and the long maturity yields shooting up. This effect is reasonable as well. The parameter κ is related to parallel shifts in the yield curve although the R^2 is just 1%. 16% of the variation in θ , which is the risk adjusted long term mean short rate level, can be explained by changes in the slope and curvature. The intuition is along the same lines as for the slopes influence on the short rate. If the level is fixed then a steepening of the yield curve is caused by the short maturity yields going down and the long maturity yields going up which indicates that the long term mean short rate is increasing. When the curvature increases θ increases as well. Increasing curvature means increasing long term yields so this result is not surprising. 11% of the variation in the s parameter is explained by parallel shifts and changing slope. An upward parallel shift increases s and a steepening of the curve decreases s.

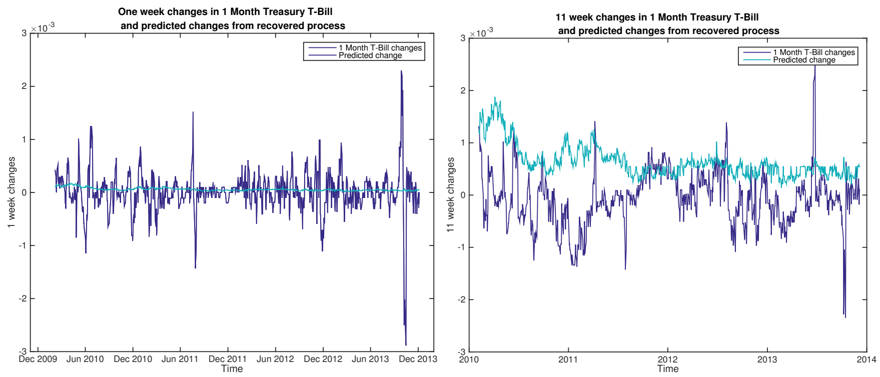
The regressions are also ran on the recovered mean reversion, γ and recovered long term mean, $\frac{\kappa\theta}{\gamma}$. The effects are almost the same as for the risk neutral parameters except for the magnitudes and that γ is decreasing for increased curvature.

	Short rate	κ	θ	s	γ	$\frac{\kappa\theta}{\gamma}$
Intercept						
PC1	0.1663**	30.4601**		15.9251**	38.4486**	
	(0.0172)	(7.8084)		(5.2726)	(9.0097)	
PC2	-0.6066**		26.4237**			27.3684**
	(0.0215)		(2.0459)			(2.3219)
PC3			29.7238**	-91.7152**	-56.9186**	40.4359**
			(2.3219)	(7.4818)	(12.7847)	(2.6352)
R^2	0.4988	0.0119	0.1609	0.1148	0.0249	0.181
N	1325	1325	1325	1325	1325	1325

*Table 5.2: Coefficients and standard errors (parentheses) from regression. First difference of variable/parameter on first difference of scores corresponding to first three principal components of the yield curve. ** denotes significance on at least 5% level.*

5.2.3 Prediction

In this section the prediction power of the recovered process is evaluated. In Figure 5.7 the predictions for the change in TBR1 with the recovered process and the actual change in TBR1 is shown. It is clear that for the 1 week prediction the variance of the changes in the TBR1 is much greater than for the predictions. This does not necessarily mean that the forecaster is bad but this signifies that the model interprets the changes in TBR1 as very noisy at this horizon. As the prediction horizon lengthens the variance of the predictor increases and it seems to become biased.



(a) 1 week *(b) 11 week*

Figure 5.7: Changes in the 1 Month Treasury Bill Rate and predicted changes by the recovered process.

In Figure 5.8 the differences in squared prediction errors for the recovered process and the risk neutral process are shown for different prediction horizons . In this specific sample the average error for the recovered process becomes progressively better relative to the risk neutral process as the horizon lengthens. This is an indication that the model is in fact able to account for the term premium to some degree.

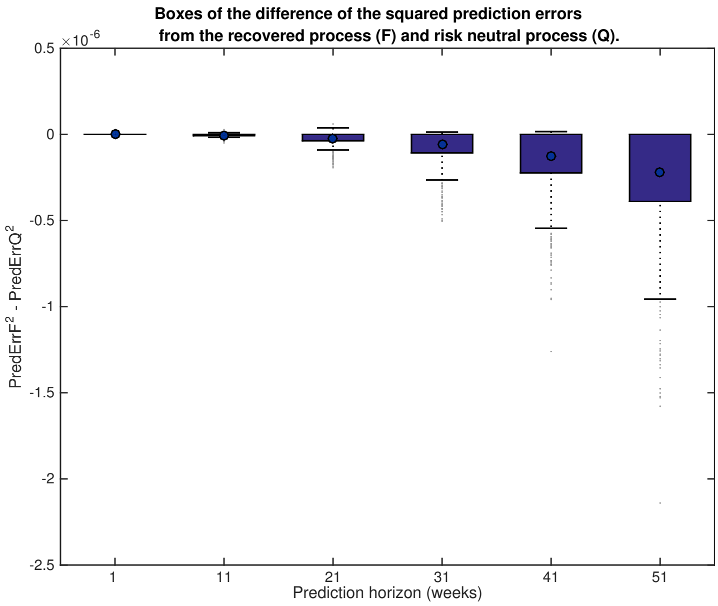


Figure 5.8: Box plot showing the difference in squared prediction errors from the recovered process and the risk neutral process. Box length denotes the difference between 3rd and 1st quantile (IQR). Upper whisker fence is equal greatest value less than or equal to 3rd quantile $+1.5\times$ IQR. Lower whisker fence is analogous.

In Figure 5.9 the same difference is shown for the recovered process relative to the martingale process. It shows that the recovered process is showing relatively worse performance as the prediction horizon increases.

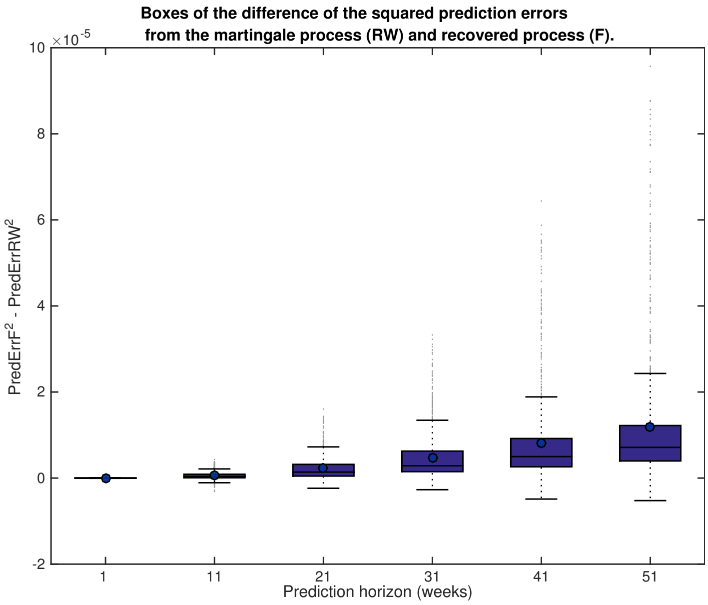


Figure 5.9: Box plot showing the difference in squared prediction errors from the recovered process and the martingale process. Box length denotes the difference between 3^{rd} *and* 1^{st} *quantile (IQR). Upper whisker fence is equal greatest value less than or equal to* 3^{rd} *quantile* $+1.5\times$ *IQR. Lower whisker fence is analogous.*

To measure the statistical significance of the above results Diebold-Mariano tests are performed at each prediction horizon for two different alternative hypothesis. One which the hypothesis is that the expected squared prediction error of the recovered process is less than for the risk neutral process. The other hypothesis is that the expected squared prediction error for the naive martingale process is less than for the recovered process. The p-values are depicted in Figure 5.10. For the first hypothesis test the p-values are below the 5% level for horizons longer than three weeks. For the second hypothesis test the p-values stays below the 5% level for horizons longer than four weeks.

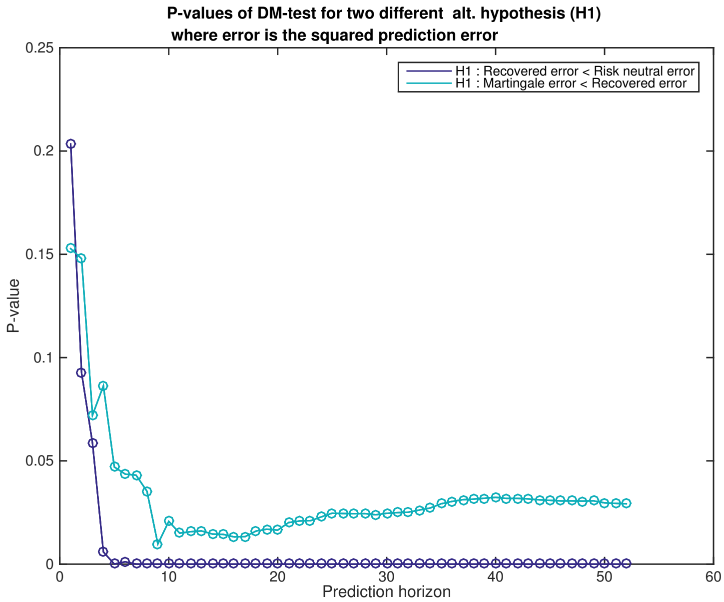


Figure 5.10: P-values of of DM-tests with two different alternative hypothesis. One with alt. hypothesis that expected squared prediction error is less for recovered process than for risk neutral process and one with alt. hypothesis that the expected squared prediction error is less for the martingale process than for the recovered process.

Chapter 6

Conclusion and further research

In this thesis the continuous state space version of the recovery approach of (Ross, 2011) has been studied. It was shown that making the assumption of a constant short rate leads to that the risk premia of the underlying driver of the model and its derivatives are zero. This render the empirical investigations in the case of VIX being the driver rather short. No evidence can be found that VIX is anything other than independent of the short rate, thus the above assumption has to be made. Hence the model seem to contradict empirical evidence that the risk premium of derivatives of VIX such as VIX futures is negative.

Another application that was considered was using the theory to recover the short rate dynamics. In this case the relationship between the driver of the market and the short rate was much clearer than in the VIX case. It is in fact the identity function. Proceeding by assuming that short rate evolves as a CIR diffusion and calibrating its parameters to the US zero coupon Treasury bill curve, the recovery approach was tested in its power to predict future changes in the 1 month Treasury bill rate (short rate proxy).

In the calibration step it was found to be challenging to obtain robust parameter estimates. A possible reason for this was the flatness of the objective function in certain directions of parameter space. This was especially noticeable in the CIR volatility direction. The substance of the parameter calibration was checked with principal component analysis.

The theory was tested through its prediction on 52 different horizons. The recovery step of the approach was evaluated through a comparison of the difference of the squared prediction error from the recovered process and the risk neutral process. In the sample period the mean difference was decreasing in the prediction horizon. This difference was statistically significant on a 5% level by the Diebold-Mariano test for prediction horizons longer than three weeks. This is consistent with previous research showing that predictions are improved by taking the term premium into account.

The recovered process was also compared to a naive predictor which guess on future changes in the 1 month Treasury bill rate is constantly zero. This naive benchmark saw better predictions than the recovered process as the prediction horizon lengthened in the sample used. According to the Diebold-Mariano test this was statistically significant on the 5% level for horizons longer than four weeks.

Conclusively, there seem to be some potential in retrieving the risk premium from the derivatives market by the recovery approach. The most straightforward setting in which to do this is in the classical one factor short rate setting where the relationship between the driver and the short rate is simple.

In terms of further research, the problem of constant short rate leading to zero risk premium could possibly be avoided by modelling the driver under some other probability measure than the risk-adjusted measure. For example the measure where the driver itself is used as numeraire would lead to the driver-drift showing up in the Sturm-Liouville equation instead of the constant short rate. Since the numeraire must be traded one can not use VIX as the driver in this case, but instead one has to use some closely related traded instrument such as VIX futures.

In the short rate recovery application better predictions can possibly be obtained by using some measure of historical volatility to improve the estimate of the s parameter in the CIR diffusion. This does not improve the recovery approach in itself, but might help to isolate some potential problems with practically applying it. Furthermore, in the spirit of (Rogers & Stummer, 2000) it could also be interesting to hold all parameters constant except for the short rate in the calibration. One should expect to see the fit to worsen but not worsen essentially according to (Rogers & Stummer, 2000). Then check if the predictive power of the model decreases or increases.

References

Abramowitz, M., & Stegun, I. (1964). *Handbook of mathematical functions with formulas graphs and mathematical tables*. Dover Publcations.

Bakshi, G., & Kapadia, N. (2003). Delta-hedged gains and the negative market volatility risk premium. *The Review of Financial Studies*, *16*(2), 527-566.

Bams, D., & Schotman, P. C. (2003). Direct estimation of the risk neutral factor dynamics of gaussian term structure models. *Journal of Econometrics*, *117*(1), 179 - 206. Retrieved from http://www.sciencedirect.com/science/article/pii/S0304407603001222

Barone, E., Domenico, C., & Emerico, Z. (1991). Term structure estimation using the cox, ingersoll, and ross model the case of italian trcacury bonds. *Journal of Fixed Income*(1), 87-95.

Björk, T. (2009). *Arbitrage theory in continuous time (3rd edition)*. Oxford University Press.

Borodin, A. N., & Salminen, P. (2002). *Handbook of brownian motion*. Birkhauser.

Brown, R. H., & Schaefer, S. M. (1994). The term structure of real interest rates and the cox, ingersoll, and ross model. *Journal of Financial Economics*, *35*(1), 3 - 42. Retrieved from http://www.sciencedirect.com/science/article/pii/0304405X94900167

Brown, S. J., & Dybvig, P. H. (1986). The empirical implications of the cox, ingersoll, ross theory of the term structure of interest rates. *The Journal of Finance*, *41*(3), pp. 617-630. Retrieved from http://www.jstor.org/stable/2328491

Carr, P., & Wu, L. (2009). Variance risk premiums. *The Review of Financial Studies*, *22*(3), 1311-1341.

Carr, P., & Yu, J. (2012). Risk, return and ross recovery. *The Journal of Derivatives*, *20*(1), 38-59.

Carriere, J. F. (1999). Long-term yield rates for actuarial valuations. *North American Actuarial Journal*, *3*(3), 13-22. Retrieved from http://dx.doi.org/10.1080/10920277.1999.10595819

Chan, K. C., Karolyi, G. A., Longstaff, F. A., & Sanders, A. B. (1992). An empirical comparison of alternative models of the short-term interest rate. *The Journal of Finance*, *47*(3), pp. 1209-1227. Retrieved from http://www.jstor.org/stable/2328983

Cochrane, J. H., & Piazzesi, M. (2005). Bond risk premia. *American Economic Review*, *95*(1), 138-160. Retrieved from http://www.aeaweb.org/articles.php?doi=10.1257/0002828053828581 doi: 10.1257/0002828053828581

Cox, J. C., Ingersoll, J., Jonathan E., & Ross, S. A. (1985). A theory of the term structure of interest rates. *Econometrica*, *53*(2), pp. 385-407. Retrieved from http://www.jstor.org/stable/1911242

Dai, Q., & Singleton, K. J. (2002). Expectation puzzles, time-varying risk premia, and affine models of the term structure. *Journal of Financial Economics*, *63*(3), 415 - 441. Retrieved from http://www.sciencedirect.com/science/article/pii/S0304405X02000673

Davydov, D., & Linetsky, V. (2003). Pricing options on scalar diffusions: An eigenfunction expansion approach. *Operations Research*, *51*(2), 185-209. Retrieved from http://pubsonline.informs.org/doi/abs/10.1287/opre.51.2.185.12782 doi: 10.1287/opre.51.2.185.12782

Delbaen, F., & Schachermayer, W. (1994). A general version of the fundamental theorem of asset pricing. *Matematische Annalen*, *300*, 463-520.

de Munnik, J. F., & Schotman, P. C. (1994). Special issue on european derivative markets cross-sectional versus time series estimation of term structure models: empirical results for the dutch bond market. *Journal of Banking & Finance*, *18*(5), 997 - 1025. Retrieved from http://www.sciencedirect.com/science/article/pii/0378426694000328 doi: http://dx.doi.org/10.1016/0378-4266(94)00032-8

Diebold, F. X., & Li, C. (2006). Forecasting the term structure of government bond yields. *Journal of Econometrics*, *130*(2), 337 - 364. Retrieved from http://www.sciencedirect.com/science/article/pii/S0304407605000795

Diebold, F. X., Li, C., & Yue, V. Z. (2008). Global yield curve dynamics and interactions: A dynamic nelson siegel approach. *Journal of Econometrics*, *146*(2), 351 - 363. Retrieved from http://www.sciencedirect.com/science/article/pii/S0304407608001127 (Honoring the research contributions of Charles R Nelson)

Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, *13*(3), pp. 253-263. Retrieved from http://www.jstor.org/stable/1392185

Durfesne, D. (2001). The integrated square root process. *Research paper number 90, Universiry of Montreal*.

Edwards, C. H., & Penney, E. D. (2008). *Differential equations and boundary*

value problems computing and modeling. Pearson Prentice Hall.
Eraker, B., & Wu, Y. (2014). Explaining the negative returns to vix futures and etns: An equilibrium approach. *Working Paper, Univ. of Wisconsin*.
Fabozzi, F. J. (2005). *The handbook of fixed income securities*. United States: McGraw-Hill USA.
Fama, E. F. (1976). Forward rates as predictors of future spot rates. *Journal of Financial Economics*, *3*(4), 361 - 377. Retrieved from http://www.sciencedirect.com/science/article/pii/0304405X76900271
Flemming, J. M. (2003). Measuring treasury market liquidity. *FRBNY Economic Policy Review*, *9*(3), 83 -108. Retrieved from http://www.ny.frb.org/research/epr/03v09n3/0309flem.pdf
Géman, H., El Karoui, N., & Rochet, J. (1995). Changes of numeraire, changes of probability measures and pricing of options. *Journal of Applied Probability*, *32*, 443-458.
Hin, L.-Y., & Dokuchaev, N. (2015). Short rate forecasting based on the inference from the cir model for multiple yield curve dynamics. *Working paper, Curtin University*. Retrieved from http://ssrn.com/abstract=2539556
Huang, R. D., & Lin, C. S. Y. (1996). An analysis of nonlinearities in term premiums and forward rates. *Journal of Empirical Finance*, *3*(4), 347 - 368. Retrieved from http://www.sciencedirect.com/science/article/pii/S0927539896000084
Johnson, T. (2015). Risk premia and the vix term structure. *Working Paper, The Univ. of Texas Austin*.
Jolliffe, I. T. (2002). *Principal component analysis*. Springer.
Longstaff, F. A. (2000). The term structure of very short-term rates: New evidence for the expectations hypothesis. *Journal of Financial Economics*, *58*(3), 397 - 415. Retrieved from http://www.sciencedirect.com/science/article/pii/S0304405X00000775
Moench, E. (2008). Forecasting the yield curve in a data-rich environment: A no-arbitrage factor-augmented {VAR} approach. *Journal of Econometrics*, *146*(1), 26 - 43. Retrieved from http://www.sciencedirect.com/science/article/pii/S0304407608000730
Moriconi, F. (1995). Analyzing default-free bond markets by diffusion models. In G. E. Ottaviani (Ed.), *Financial risk in insurance*. Berlin: Springer.
Qin, L., & Linetsky, V. (2014). Positive eigenfunctions of markovian pricing operators: Hansen-scheinkman factorization and ross recovery. *Working paper, Northwestern University*.
Rogers, L. C. G., & Stummer, W. (2000). Consistent fitting of one-factor

models to interest rate data. *Insurance Mathematics and Economics*, 27(1), 45 - 63. Retrieved from http://www.sciencedirect.com/science/article/pii/S0167668700000391

Ross, S. A. (2011, August). *The recovery theorem* (Working Paper No. 17323). National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w17323 doi: 10.3386/w17323

Slater, L. J. (1960). *Confluent hypergeometric functions*. Cambridge University Press.