On the Suitability of the Longstaff–Schwartz Term Structure Model for Modelling the Cost of Government Debt

Joose Mikko Juhani Sauli
University of Helsinki
Faculty of Social Sciences
Economics
Master's Thesis
October 2013

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

Faculty of Social Sciences	Department of Political and Economic Studies

Sauli, Joose Mikko Juhani

On the Suitability of the Longstaff–Schwartz Term Structure Model for Modelling the Cost of Government Debt

Economics

Master's thesis	October 2013	68 + 21 pages

The study addresses an estimation problem faced by a large borrower, such as a government, related to an interest rate risk measure known as Cost-at-Risk, or CaR. The term denotes a threshold level of debt costs such that the actual debt costs incurred in a given time (say, one year) will be less than this threshold level with a given probability (say, 95 %).

The main obstacle to determining CaR is that the probability distribution of future levels of interest rates is unknown. For this purpose, various models of the term structure of interest rates have been developed. This study takes one particular term structure model, the Longstaff–Schwartz model, under examination in order to determine its inherent suitability for the estimation of CaR. The model is an affine two-factor equilibrium model with analytic solutions for bond and option prices.

The accuracy of the model is studied by simulating interest rate pseudo-data using a simulation program which corresponds exactly to the model's assumptions, and then recalibrating the model to the pseudo-data. Given that the properties of the data-generating process (DGP) are known exactly, this approach allows us to compare the CaR estimates implied by the recalibrated model against the CaR implied by the actual properties of the DGP.

Particular attention is paid to the methods by which the model is calibrated to data. In an effort to improve the accuracy of the Longstaff–Schwartz model, a new calibration method is developed.

In order to appraise the accuracy of the Longstaff–Schwartz model, we compare its performance to that of a simpler benchmark model based on the Nelson–Siegel decomposition of the yield curve. The accuracy of the CaR estimates given by the two models is compared both in an environment where the DGP is of the Longstaff–Schwartz type, and in another environment where the DGP is of the Nelson–Siegel type.

The results of the comparison can be summarized as follows. The new calibration method for the Longstaff–Schwartz model is highly accurate when the DGP is of the LS type, but is useless in the NS-type environment. When the LS model is calibrated using the technique proposed by Longstaff and Schwartz themselves, it turns out that the model gets no advantage from being correctly specified. When correctly specified, it fails to calibrate to data in about 25% of all cases, but when it is misspecified, the failure rate drops to 2.4%, and its accuracy improves and surpasses that of the NS model.

These results are highly unexpected. It is possible that they are specific to the parameter values used in the simulations, but this issue is left for further research.

mathematical finance
term structure of interest rates
interest rate modelling
Longstaff–Schwartz
Cost-at-Risk

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

Valtiotieteellinen tiedekunta	Politiikan ja talouden tutkimuksen laitos
Sauli, Joose Mikko Juhani	
On the Suitability of the Longstaff–Schwartz Term Structure Model for Modelling the Cost of Government Debt	
Taloustiede	

Pro gradu	Lokakuu 2013	68 + 21 sivua

Tutkimus koskee suuren lainanottajan, kuten valtion, kohtaamaa estimointiongelmaa, joka littyy nimellä Cost-at-Risk (CaR) tunnettuun korkoriski-indikaattoriin. CaR on vastaus kysymykseen, mikä on sellainen korkokustannusten taso, että on tietty todennäköisyys (esim. 95 %), että tietyssä ajassa (esim. vuosi) kertyvät korkokustannukset jäävät ko. tason alle.

Suurin este CaR:n estimoimiselle on se, ettei tulevien korkotasojen todennäköisyysjakaumaa tunneta. Tähän tarkoitukseen on kehitetty erilaisia korkomalleja. Tässä tutkimuksessa tarkastelun kohteeksi otetaan yksi korkomalli, Longstaff–Schwartz-malli, ja tutkitaan sen periaatteellista sopivuutta CaR:n estimoimiseen. Kyseessä on affiini kahden muuttujan tasapainomalli, josta velkakirjojen ja velkakirjaoptioiden hinnat voidaan ratkaista suljetussa muodossa.

Mallin tarkkuutta tutkitaan tuottamalla pseudokorkodataa simulaatio-ohjelmalla, joka vastaa täsmälleen mallin oletuksia, ja kalibroimalla malli uudelleen tähän pseudodataan. Koska dataa generoivan prosessin (DGP) ominaisuudet tunnetaan, tulee mahdolliseksi verrata uudelleenkalibroidun mallin antamia CaR-estimaatteja ”todelliseen” CaR:iin, joka voidaan johtaa DGP:n ominaisuuksista.

Mallin kalibrointimenetelmään kiinnitetään erityistä huomiota. Estimointitarkkuuden parantamiseksi tutkielmassa kehitetään Longstaff–Schwartz-mallille uusi kalibrointimenetelmä.

Jotta Longstaff–Schwartz-mallin tarkkuutta voitaisiin suhteuttaa johonkin, vertaamme sitä yksinkertaisempaan vertailumalliin, joka perustuu Nelson–Siegel-korkokäyrähajotelmaan. Näiden kahden mallin antamien CaR-estimaattien tarkkuutta verrataan keskenään sekä ympäristössä, jossa DGP on Longstaff–Schwartz-tyyppinen, että ympäristössä, jossa DGP on Nelson–Siegel-tyyppinen.

Vertailun tulokset voidaan tiivistää seuraavasti. Longstaff–Schwartz-mallille kehitetty uusi kalibrointimenetelmä toimii hyvin, kun DGP on LS-tyyppinen, mutta NS-ympäristössä se on aivan hyödytön. Kun LS-malli kalibroidaan Longstaffin ja Schwartzin omalla tekniikalla, käy ilmi, ettei malli ole yhtään sen tarkempi ollessaan oikein spesifioitu kuin ollessaan väärin spesifioitu. DGP:n ollessa mallispesifikaation mukainen, kalibrointi epäonnistuu noin joka neljännellä yrittämällä; mutta kun DGP on NS-tyyppinen, LS-mallin epäonnistumisprosentti putoaa 2,4:ään ja sen tarkkuus paranee, voittaen NS-mallin.

Nämä tulokset ovat erittäin yllättäviä. On mahdollista, että ne ovat seurausta simuloinneissa käytetyistä parametrinarvoista, mutta tämä kysymys jätetään myöhemmin tutkittavaksi.

matemaattinen rahoitusteoria
korkojen aikarakenne
korkomallit
Longstaff–Schwartz
Cost-at-Risk

Abstract

This study examines the performance of the Longstaff–Schwartz model of interest rates from a narrowly specified interest rate risk-management perspective, namely the estimation of a risk measure known as Cost-at-Risk. We study the accuracy of the estimates given by the model by calibrating it to pseudo-data, which has the advantage that the properties of the data-generating process are known exactly. The performance of the model is compared to that of a benchmark model based on the Nelson–Siegel decomposition of the yield curve. Opposite to what one might expect, we find that the Longstaff–Schwartz model's accuracy is not at all impaired when the model is, in fact, misspecified.

Contents

Introduction

This study addresses a problem faced by a large borrower, such as a government, who is following a well-defined debt management strategy (or trying to choose one) and needs to form a picture of the interest rate risk implied by that strategy, i.e. the uncertainty regarding the cost of debt arising from the possibility of adverse changes in the general level of interest rates.

In particular, the borrower would like to know the range of possible outcomes; or rather, their probability distribution, given that there is no limit to how high the rate of interest required from a new loan can be. Therefore a risk manager might want to determine, say, the 95th percentile of annual debt costs—that is, a level of debt costs which will be exceeded with a probability of only 5% in any given year. Such a statistic is referred to as *Cost-at-Risk*, or CaR (in keeping with *Value-at-Risk*, or VaR).

Given sufficient assumptions, it is possible to calculate that figure. In the assumptions, one must specify three things: the debt-management strategy, which defines what kinds of financial transactions are to be made and when; the cost accounting basis, which defines when funding costs are recognized; and finally, a probability distribution for the market variables which determine the costs and returns of the debt-management strategy. In practice, one addresses the third question by specifying an interest rate model.

For example, one might assume that debt is issued (and redeemed) semiannually in the form of ten-year bonds; that *costs* are interpreted as net nominal cash flows; and that interest rates behave in the fashion described by the so-called Vasicek yield curve model with such-and-such parameter values.

Whereas the funding strategy and cost accounting basis would usually follow directly from the practical problem at hand, the third question—picking a suitable interest rate model—is more involved. There are many interest rate models and modelling approaches, with different flaws and merits. Some are simple but unrealistic; others are more realistic but unwieldy. Some describe the immediate future

elegantly, but cannot be used for modelling the long term without becoming either unrealistic or too complicated.

The problem of choosing an appropriate model for the estimation of CaR is what inspired this study. Clearly, the problem is too broad to be settled here, but we can try to make a modest contribution. What we will do is take one attractive model—namely, the Longstaff–Schwartz equilibrium model—under scrutiny, and assess its usefulness for the estimation of Cost-at-Risk. We do this by way of experiments with simulated pseudo-data. This allows us to make a comparison between the actual CaR implied by the properties of the data-generating process (DGP), and the CaR estimates obtained by calibrating the model to the pseudo-data.

In order to appraise the Longstaff–Schwartz model's performance, we compare it to that of a cruder benchmark model based on the Nelson–Siegel decomposition of the term structure of interest rates. The performance of each model is measured both when the model is correctly specified and when it is misspecified. It is natural to anticipate that a correctly specified model will have an advantage against a misspecified model; yet we might find a difference between the models regarding their inherent suitability for empirical CaR estimation. To find out whether the Longstaff–Schwartz model can beat the benchmark in this respect is thus the objective of this study.

Contrary to expectations, we find that neither model gains an edge from being correctly specified. Indeed the Longstaff–Schwartz model performs much better when the DGP is as described by the benchmark model than when the DGP is of the Longstaff–Schwartz type, and the benchmark model itself performs roughly as badly under both types of DGP.

This outcome is sufficiently surprising as to suggest a certain skepticism; although we set out to draw conclusions regarding the accuracy of the Longstaff–Schwartz model, we now believe it would be premature to do so on the basis of the results of this study. Instead, we suspect that our results may be largely attributable to the parameter values used in the DGPs: perhaps the accuracy of each model depends much more heavily on the parameter values of the DGP, and much less heavily on its form, than the approach taken in this study assumes. However, this conjecture cannot be substantiated without further research.

1 Fundamentals

1.1 Interest rate relationships and terminology

Term structure modelling is concerned with how the interest on a loan depends on the term of the loan. This definition simplifies the subject matter by glossing over all other possible determinants of interest rates, notably the liquidity of the debt securities in question and the credit risk attached to them.

The basic building block of term structure models is the *default-free discount bond with unit nominal value*. 'Default-free' means that the bond carries no credit risk; as a discount bond it pays no coupons; and 'unit nominal value' means that the holder receives 1 unit of cash when the bond matures. Lending and borrowing transactions are assumed to be conducted by means of issuing, buying and selling such bonds. Furthermore, we assume that there are no transaction costs, and no restrictions on short-selling, so that issuing, selling, and shorting a bond are considered equivalent.

The central variable is the price at time t of a bond maturing at time T. We call this the T-bond and denote its price by $P(t,T)$. Since the return on a discount bond consists of capital gain alone, its continuously compounded yield y is easy to calculate:

$$P(t,T)e^{\tau y} = P(T,T) = 1 \quad \Longleftrightarrow \quad y = \frac{-\ln\left(P(t,T)\right)}{\tau} \tag{1.1}$$

where $\tau \equiv T - t$, a shorthand which we use throughout this study.

The yield on a coupon-paying bond would be harder to calculate and, what is worse, it would depend on the timing of the coupons. Therefore, to keep things simple, term structure models focus on zero-coupon rates, represented by the *zero-coupon yield curve*, or *zero curve* for short. The zero rate for maturity T is the discount rate by which investors calculate the present value of a cash flow to be realized at time T. We denote this by $R(t,\tau)$. The zero curve at time t shows how $R(t,\tau)$ depends on τ. From (1.1),

$$e^{-\tau R(t,\tau)}P(T,T) = P(t,T) \quad \Longleftrightarrow \quad R(t,\tau) = \frac{-\ln\left(P(t,T)\right)}{\tau} \tag{1.2}$$

Although in the real world bonds longer than a year in maturity virtually always pay coupons, zero rates can nevertheless be inferred from the market prices of bonds when we know the timing of the cash flows generated by each bond. The method used here is called bootstrapping: we build the curve from the short end up, discounting coupons as we go and eliminating their present values from the bonds' market values. This will work fine if there is a steady stream of bonds maturing at a sufficient frequency. Throughout the rest of this study, by 'yield curve' we will always be referring to the zero curve.

Another technicality to be aware of is the *compounding frequency*. Annual (semi-annual) compounding is when interest is added to principal once (twice) per year, accelerating the accrual of interest. For mathematical simplicity, we assume that interest accrues continuously, as in equations (1.1) and (1.2).

The left end point of the yield curve at $\tau = 0$ represents the return on a bond maturing in an infinitesimally short time. It is called the *short rate*, or *instantaneous* or *risk-free* rate of interest. For concreteness, one may think of this as the overnight rate, but strictly speaking there is no satisfactory real-world equivalent for this. The instrument which gives this return is also fictional; we may call it the *cash bond*. Formally, we denote the short rate by r_t and define it as

$$r_t \equiv \lim_{\tau \to 0} R(t, \tau) \tag{1.3}$$

The function $T \mapsto P(t, T)$ is called the discount curve. As equation (1.2) implies, the zero curve and the discount curve contain exactly the same information: whereas the zero curve indicates the discount rate applicable over a term, the discount curve gives the discount factor. Useful as that may be, the discount curve is a poor way to visualize the term structure, given that it always starts from 1 and has a downward-sloping shape which is relatively insensitive to changes in interest rates.[1] Figure 1.1 illustrates this point.

A *forward rate agreement* is a contract where one party commits, at time t, to lending to another at time T_1 in return for repayment at time T_2, and the other

[1] If interest rates over any maturity interval become negative, then the discount curve will become upward-sloping over that interval. This can happen in the real world: the Eurozone bond and interbank markets have witnessed negative nominal rates during late 2012 and early 2013. Yet many yield curve models are specifically designed to rule this out.

party commits to borrowing at these terms. Given a yield curve at time t, arbitrage dictates what the rate on the loan must be, since the borrower could replicate the deal by issuing a T_2-bond and using the proceeds to buy $\frac{P(t,T_2)}{P(t,T_1)}$ units of the T_1-bond. The rate which makes the borrower indifferent between this and a forward rate agreement is the *forward rate*, denoted by $f(t, T_1, T_2)$:

$$f(t, T_1, T_2) = \frac{\tau_2 R(t, \tau_2) - \tau_1 R(t, \tau_1)}{T_2 - T_1} \tag{1.4}$$

(where $\tau_i \equiv T_i - t$). If the forward rate were higher than this, one could make a risk-free profit by executing the two-bond strategy and lending forward, and if lower, by doing the converse.

Another useful concept for term structure modelling is the *instantaneous forward rate for maturity* T, defined as

$$f(t, T) \equiv \lim_{T_2 \to T} f(t, T, T_2) \tag{1.5}$$

It follows that

$$f(t, T_1, T_2) \;=\; \frac{1}{T_2 - T_1} \int_{T_1}^{T_2} f(t, T)\, dT \tag{1.6}$$

$$R(t, \tau) \;=\; \frac{1}{\tau} \int_t^T f(t, s)\, ds \tag{1.7}$$

$$f(t, T) \;=\; R(t, \tau) \;+\; \tau \frac{\partial}{\partial \tau} R(t, \tau) \tag{1.8}$$

The function $T \mapsto f(t, T)$ is called the forward curve. Figure 1.1 shows, in the left panel, the zero-coupon spot rates (solid line) and the corresponding forward rates (dashed line). The shape of these curves illustrates that spot rates are just cumulative averages of the instantaneous forward rate, as expressed by (1.7). In the right panel we have the discount curve. All three graphs have exactly the same information content.

The price of a bond is the more sensitive to yield curve shifts, the longer its maturity. From (1.2),

$$\frac{\partial}{\partial R(t, \tau)} P(t, T) = -\tau e^{-\tau R(t,\tau)} = -\tau P(t, T) \tag{1.9}$$

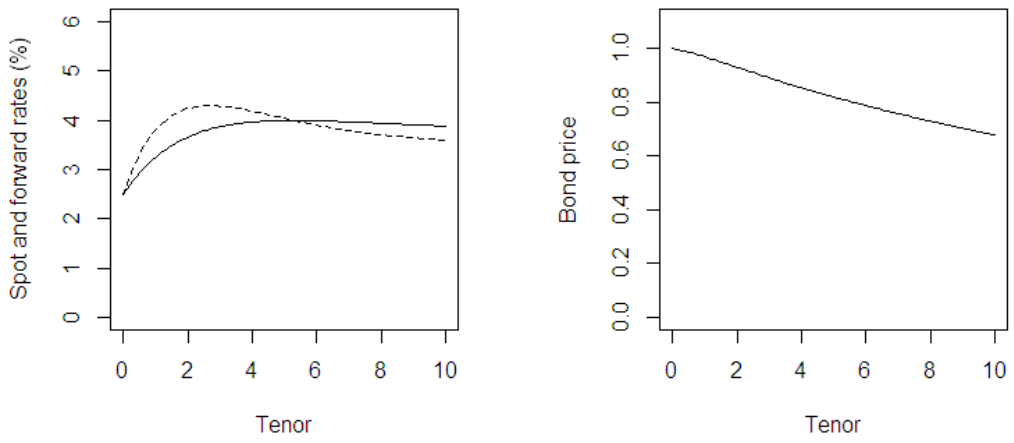

Figure 1.1: On the left: the zero-coupon curve (solid line) and the corresponding forward curve (dashed line). On the right: the discount curve that corresponds to the curves on the left. All three curves have exactly the same information content, but the discount curve is less appropriate for visualization.

In terms of Figure 1.1, an upward shift of the yield curve would be reflected in the discount curve as a steepening of the slope with the intercept still at 1. The implication for investment strategy is that the bond investor who expects an increase (fall) in yields should seek to reduce (increase) the average maturity, or *duration*, of their bond portfolio by selling long bonds and buying short-term bonds. When many investors do this, short-term yields will be pushed down and long-term yields will be pushed up, increasing the attractiveness of long-term bonds relative to short-term bonds, until the market reaches the new equilibrium.

1.2 Stylized facts

Now that we have a grasp of the concepts, we move on to consider the typical yield curve shapes and movements. A good term structure model should have features as similar as possible to the behaviour of interest rates which is actually observed in the market. But what exactly are those features?

Diebold and Li (2006) offer an explicit list of stylized facts of yield curve dynamics; another such list is found in an unpublished paper by Brandt and Chapman (2002). Most authors, however, never address this question explicitly. Those who do something of the sort, tend to focus on a fairly restricted set of historical data,

making no attempt at a general-purpose list of stylized facts. Campbell and Shiller (1991), Chapman and Pearson (2001) and Piazzesi (2010) are cases in point. The *History of Interest Rates* by Homer and Sylla (1991) is very detailed, but offers no generalizations of the kind we are looking for.

Yet there seems to be a number of generally accepted notions which are frequently cited in the literature as if common knowledge, without appeal to sources or data. The following list of stylized facts, modelled after the aforementioned two, outlines my perception of what these are.

1. The yield curve is upward-sloping most of the time and on average.
2. The yields on all maturities are strongly but imperfectly correlated. The smaller the difference in maturity, the higher the correlation between two yields.
3. Shocks to yield levels are highly persistent.
4. Long-term yields are less volatile than short-term yields.
5. The conditional volatilities of yield changes are time-varying and persistent.
6. The conditional volatilities of yield changes are positively correlated with the level of yields.

The accuracy of these statements is examined in the Appendix in in the light of actual market data from years 1953–2013, using basic statistical analysis. The result is that all these claims are defensible, with two caveats. Firstly, whether item #4 is entirely up to date is debatable: it is a good description of the data up to around 1985, but after that, it depends on how volatility is measured. Figure 1.2 illustrates this: since the late 1980s, the one-year rate has been either more or less volatile than the long rate, depending on your concept of volatility. Secondly, Stylized Fact #6 is rather discredited by the most recent 20 years of data.

The list is missing two potential items: nonnegativity and mean reversion of nominal yields. These are not on the list, because they may be a little controversial. Regarding the first, it is safe to say that negative nominal yields are very rare if not impossible. One might conjecture that this depends on whether the monetary system permits agents to hold cash risklessly at zero cost.

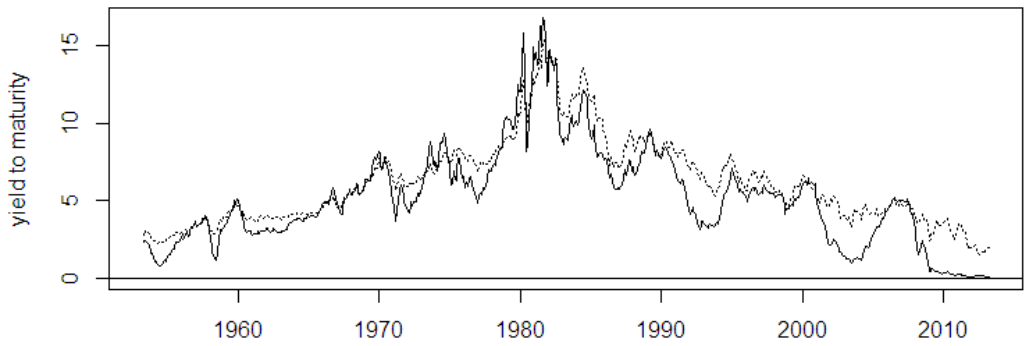

Figure 1.2: US government bond yields, Apr 1953–Mar 2013, 1-year (solid line) and 10-year (dotted line), yield-to-maturity.

What comes to mean reversion, most dynamic term structure models assume it, and in those which do not it is regarded as a major flaw. Any economist will quickly point out that a high cost of funding will curb economic initiative, which will tend to bring down returns to capital, whereas a low cost of capital will have the opposite effect. But this argument refers to real returns, and to what extent it applies to nominal yields is not obvious.

1.3 General theories of the term structure

A financial *model* is not quite the same thing as a *theory* which tries to make sense of why financial asset prices behave as they do. In this section we review some general or 'traditional' theories on what causes the yield curve to take the shape it takes.

1.3.1 The expectations theory

This theory has many possible formulations, many of them mutually contradictory (e.g. Cox et al. 1981). The general idea, however, is that forward rates are determined by expectations of future spot rates. That is, an upward-sloping yield curve means that interest rates are expected to rise; an inverted yield curve means they are expected to fall; and a humped curve would indicate expectations of a temporary rise. Such expectations might be based on projected inflation, real productivity and consumption preferences.

Campbell and Shiller (1991) summarize the literature by saying that while almost all studies statistically reject the expectations theory, some studies reach the opposite conclusion, as different studies use different econometric methods, test different implications of the expectations theory, and look at different maturities.

As Cairns (2004) points out, the pure expectations theory is difficult to square with Stylized Fact #1. If the theory were true, the average slope of the yield curve ought to be zero. Thus, if forward rates are expectations (in some sense) of future spot rates, they appear to be biased, and it remains to work out why this might be.

1.3.2 The liquidity preference theory

Whereas the expectations theory cannot be attributed to any individual author, the liquidity preference theory was first put forward by Keynes in his *General Theory* and subsequently developed by Hicks in *Value and Capital*. This theory postulates that investors do not like to tie up their capital for a long period of time, i.e. they prefer liquid assets to illiquid ones. Thus they will not buy long-term bonds unless they are rewarded for this with a liquidity premium.

As such, this explanation seems shaky, given that there exists a highly liquid market for bonds of all maturities. However, there is a variant of it which is more plausible. The returns on long-term bonds are more volatile than the returns on short-term bonds, and as this risk cannot be diversified away, investors require a risk premium to be willing to bear it. The longer the maturity of a bond, the higher its sensitivity to interest rate risk, and the higher the risk premium. It follows that the yield curve must be upward-sloping.

Since the yield curve is not always upward-sloping, this cannot be the whole story; but some combination which takes its cues partly from the expectations theory and partly from the (misnamed) liquidity preference theory might explain much of the behaviour of the yield curve.

1.3.3 The market segmentation theory

This theory is due to Culbertson (1957). It postulates that each market participant has strong maturity preferences—pension funds are a prime example—so that bonds of different maturities trade in separate markets. Then there is no reason why

yields on different maturities should have any particular relation to each other; an upward-sloping yield curve merely indicates different equilibria in different markets, i.e. long-term funding is more scarce relative to demand than short-term funding.

On a more moderate note, Modigliani and Sutch (1966) postulate that individuals have a 'preferred habitat' out of which they can only be lured by higher expected returns, i.e. they are willing to compromise on their maturity preferences, for a price.

1.3.4 Arbitrage pricing theory and term structure modelling

The above theories became outmoded in the 1970s as a new array of tools of mathematical finance, collectively referred to as *arbitrage pricing theory*, were applied to the pricing of bonds, leading to the formulation of the first term structure models. Rather than explaining security prices with economic fundamentals, arbitrage pricing theory studies the pricing of securities relative to each other.

While term structure models make no appeal to the traditional theories, it is clear that the insights expressed by these theories are echoed in the models derived by arbitrage pricing. An arbitrage-free term structure model where expectations of future yields play no role is an outright impossibility; risk premia on longer maturities are readily incorporated; and when a model has more than one stochastic factor, the risk premia related to the different risk factors can be interpreted to represent market segmentation.

2 Overview of interest rate models

In this chapter, we take a look around the marketplace of term structure models. A distinction is often drawn between *equilibrium models* and *no-arbitrage models*, but the content of this demarcation seems contested. There are two views:

1. Equilibrium models are those which are derived from an underlying equilibrium model of fundamental economic processes. No-arbitrage models take a shortcut by making assumptions concerning some aspect of the yield curve dynamics—usually the short rate—directly, and derive from these a coherent model of the entire yield curve by means of no-arbitrage arguments.

2. The defining feature of no-arbitrage models is that they give an exact fit to any initial yield curve, effectively taking it as an input. With equilibrium models, the yield curve is exclusively an output, and often gives a poor fit to the yield curve actually observed.

The first way to draw the distinction has more theoretical appeal, but we adopt the second because it is more relevant for practical purposes. By and large, models that give a perfect fit to the current yield curve tend to have implausible long-term dynamics—and realistic long-term dynamics are very important for our ultimate objective, the estimation of CaR.

The distinction does not cover all term structure models. A third group consists of what could be dubbed *purely descriptive models*, i.e. models that are not derived from assumptions of maximizing behaviour, or any other distinctively economic reasoning for that matter. They are purely instrumental and non-explanatory.

2.1 Equilibrium models

The most canonical models in this category are the Vasicek and the Cox–Ingersoll–Ross (CIR) models. The two produce quite similar yield curve behaviour, but their theoretical underpinnings are very different: the Vasicek model would classify as a

no-arbitrage model under the first definition of this term given above, as it is derived by no-arbitrage arguments from assumptions concerning the dynamics of the short rate. The CIR model is an equilibrium model in both senses of the term.

2.1.1 Vasicek

Vasicek's paper (Vasicek 1977) has two parts, the first of which consists mostly of arguments and results which are nowadays considered basic principles of mathematical finance, so we do not reiterate them here. The second part lays out a simple application of these principles which has come to be known as 'the Vasicek model', although Vasicek himself intended it only as an illustration. The model is based on the following assumptions:

1. The short rate r follows an Ornstein–Uhlenbeck process, i.e. the instantaneous change of r is described by the SDE

$$dr \;=\; \alpha(\gamma - r)\,dt \;+\; \rho\,dW_t \tag{2.1}$$

   where the mean reversion rate α, long-run mean γ and volatility rate ρ are positive constants, and W_t is a Brownian motion.[1]

2. The price $P(t,T)$ of a discount bond is determined by the assessment, at time t, of the probability distribution of the path of the short rate during the life of the bond.

3. The market is efficient: there are no transaction costs and no arbitrage opportunities.

4. The market price of risk due to W_t, denoted by q, is constant.

Figure 2.1 shows 100 sample paths of the short rate, with one path singled out. Here the long-run average short rate is 0.04, represented by the dashed line. The reversion rate is 0.5, and the volatility rate is 0.015. With these parameters, shocks

[1] A summary introduction to stochastic differential equations (SDEs) like (2.1), as well as to the Itô differentiation rule invoked below, is found in many basic mathematical finance textbooks, e.g. Hull (2006). For a more thorough introduction, see e.g. Shimko (1992), Baxter and Rennie (1996) or Mikosch (1998).

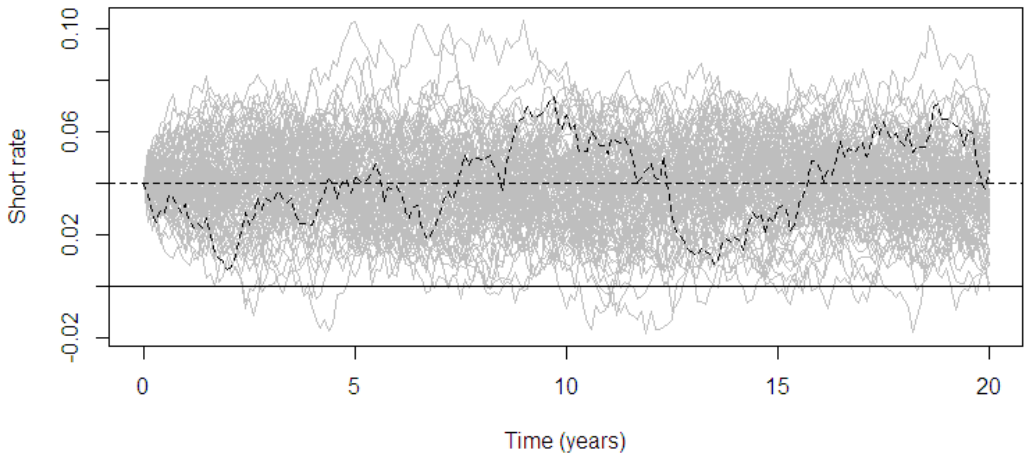

Figure 2.1: Vasicek short rate paths at parameter values $\alpha = 0.5$, $\gamma = 0.04$, $\rho = 0.015$. Showing 100 paths, with one singled out.

to r have a half-life of 1.4 years, and the unconditional standard deviation of r is 0.015. All paths start from the level γ.

From assumptions 1–4 Vasicek derives the following bond-pricing formula (where τ is our shorthand for $T - t$):

$$
\begin{aligned}
P(t,T) &= \exp(A_\tau - B_\tau r) \qquad (2.2)\\
\text{where} \quad B_\tau &= \frac{1}{\alpha}(1 - e^{-\alpha\tau})\\
\text{and} \quad A_\tau &= \big(B_\tau - \tau\big)R(\infty) - \frac{\rho^2}{4\alpha}B_\tau^2\\
\text{and} \quad R(\infty) &= \gamma + \frac{\rho q}{\alpha} - \frac{\rho^2}{2\alpha^2}
\end{aligned}
$$

From the discount function (2.2), the yield curve is readily derived using (1.2):

$$
R(t,\tau) = -\frac{\ln\left(P(t,T)\right)}{T-t} = \frac{B_\tau r - A_\tau}{\tau} \quad \begin{cases} \underset{\tau\to 0}{\longrightarrow} r \\ \underset{\tau\to\infty}{\longrightarrow} R(\infty) \end{cases} \qquad (2.3)
$$

The Vasicek model belongs to the class of *affine models*, distinguished by the feature that the yield on each maturity is an affine function of the state variables (in this case, r).

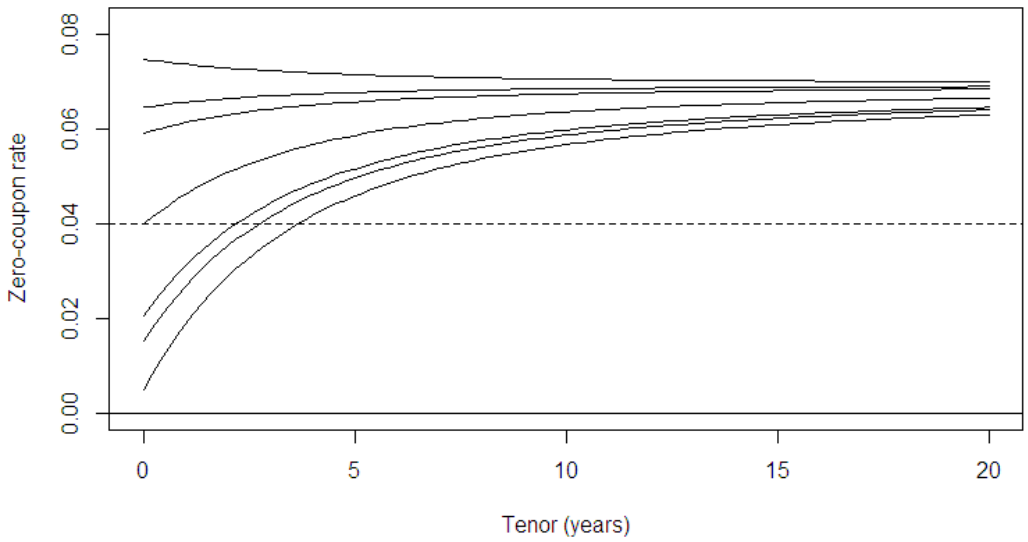

Figure 2.2: Yield curves generated by the Vasicek model at parameter values $\alpha = 0.5$, $\gamma = 0.04$, $\rho = 0.015$, $q = 1$, for various values of r representing the 1%, 5%, 10%, median, 90%, 95% and 99% quantiles of the unconditional (long-run) distribution of r.

The yield curve can take three basic shapes, depending on the current level of r: upward-sloping (when r is low), inverted (when r is high) and slightly humped (when r is close to $R(\infty)$). The hump is never very pronounced at reasonable parameter values.

Figure 2.2 gives an illustration of what Vasicekian yield curves look like. α, γ and ρ are the same as in Figure 2.1, and the market price of interest rate risk is $q = 1$. The starting point of a curve represents the short rate, which in the Vasicek model is normally distributed. The lowermost curve corresponds to the 1% quantile; rates will be below this level only 1% of the time. The other curves represent the 5%, 10%, median, 90%, 95% and 99% quantiles.

While the yield curves depicted in Figure 2.2 are quite realistic, the model as a whole is subject to the following critical observations:

1. The short rate will go negative from time to time; how often this happens depends on the model parameters. This has been viewed as a highly undesirable feature in an interest rate model, and economists have gone to a lot of trouble to devise models where this is not possible.

2. The volatility of long-term yields is unrealistically low; a realistic range cannot be achieved without allowing an excessively broad range for the short rate.

3. The rates on all maturities are perfectly correlated—a feature intrinsic to all single-factor affine models. This is inconsistent with what is actually observed in the market, and renders the model useless for many interest rate risk management applications.

4. The range of shapes for the yield curve is quite limited. While it is technically true that a curve may display a hump, with reasonable parameter values—say, $\gamma < 0.5$—the hump is very modest, hardly perceptible.

2.1.2 CIR

This model, introduced in Cox, Ingersoll and Ross (1985b), differs from the Vasicek model in two ways. The first is relatively superficial: the SDE for the short rate process is assumed to be

$$dr \;=\; \kappa(\theta - r)\,dt \;+\; \sigma\sqrt{r}\,dW_t \tag{2.4}$$

i.e. it is rather similar to Vasicek's, but the volatility of the short rate is proportional to the square root of its level, which implies that the short rate cannot go negative; the closer it gets to zero, the more its fluctuation abates, and the more the pull to mean dominates.

The other difference is more profound. The CIR model is derived from an underlying general equilibrium model of economic fundamentals, including a model of how investments in various stochastic production processes are turned into output; a model of how these production processes themselves develop over time; a model of how individual agents engage in investing, trading and consuming the production output, maximizing the expected value of an intertemporal utility function.

From this equilibrium model, which is highly abstract, the authors derive a partial differential equation which all asset prices in the economy must satisfy. The solution of the equation depends on the functional forms and parameter values of the underlying economic model.

The general framework is developed in Cox & al. (1985a). In Cox & al. (1985b) the authors illustrate the framework by specifying explicit functional forms for the underlying economic processes; from these, a dynamic term structure model is derived. The bond pricing formula is reminiscent of Vasicek's, albeit more complex; the resulting yield curves and their time-paths look very much like those produced by the Vasicek model, except that the distribution of rates is skewed upwards as a result of the introduction of $\sqrt{r}$ into the volatility term. By the same token, the CIR model is more consistent with our Stylized Facts #5 and #6 than the Vasicek model.

The one-factor CIR model is vulnerable to criticisms 2–4 levelled against the Vasicek model above, and is therefore too simple for serious interest rate risk management applications. The authors also consider multifactor extensions, and lay out in detail a two-factor version with inflation as the second state variable.

2.1.3 Other equilibrium models

While Vasicek and CIR are the most canonical equilibrium models, there are of course others; we mention them here in a chronological order.

The so-called Merton model was sketched by Robert C. Merton (1973) in a footnote to illustrate the general principles of arbitrage pricing. The short rate is simply a scaled Brownian motion with constant drift; its SDE is

$$dr \;=\; \mu\, dt + \sigma\, dW$$

—Vasicek without mean reversion, that is.

Dothan's (1978) model omits the drift term from the SDE for r, and introduces the current level of r as a multiplier, so that the short rate becomes a (scaled) geometric Brownian motion; future values of r are lognormally distributed given the current value of r. Bond prices have analytical solutions, but the implied yield curve dynamics are unrealistic.

Brennan and Schwartz (1979) propose a two-factor model. One of the state variables is the short rate r; the other is the long rate ℓ. The model allows more versatile yield curve shapes than one-factor models, but unfortunately it has no closed-form solutions for bond prices. Also, it has been shown that the process

explodes, i.e. either r or ℓ will reach ∞ in finite time with probability one.

Langetieg (1980) presents a multifactor version of the Vasicek model: the short rate is obtained as a linear combination of an arbitrary number of state variables jointly following a multivariate Ornstein–Uhlenbeck process. Alas, bond prices do not generally have analytical solutions.

Schaefer and Schwartz (1984) put forward a two-factor model whose state variables are the short rate r and the spread s between the short and the long rate, i.e. $s \equiv \ell - r$. The model has no exact analytical solutions, but does have an approximate analytical solution for bond prices.

Longstaff and Schwartz (1992) propose a two-factor model where the state variables are the short rate r and its instantaneous variance V; that is, $\mathrm{Var}(dr) = V\,dt$. The model has many attractive features, including analytic solutions for bond prices as well as bond option prices; more flexibility with regard to yield curve shapes than one-factor models are capable of; and, with suitable parameter values, fairly realistic yield curve dynamics.

Pearson and Sun (1994) develop an extended version of the two-factor CIR model. The model has analytic solutions for bond prices, but the authors find its empirical performance to be very poor.

Apart from equilibrium models proper, yet engaged in a dialogue with them, there is a wider literature concerned more with the mathematical forms of term structure models and less with their theoretical underpinnings. For example, Duffie and Kan (1996) provide a generalization of multifactor affine models, with closed-form solutions for bond prices, but their framework is not underpinned by an economic model based on maximizing behaviour. Others have studied quadratic models; a survey and analysis of this class of models is found in Ahn, Dittmar and Gallant (2002).

2.2 No-arbitrage models

2.2.1 Ho–Lee (1986)

The first model to provide a perfect fit to any given initial yield curve was due to Ho and Lee (1986). Yield curve movements are driven by the stochastic movements

of the short rate r. The model was originally formulated in discrete time, but was later shown to have a continuous-time limit characterized by the SDE

$$dr_t \;=\; \sigma\, dW_t + \theta_t\, dt \tag{2.5}$$

where σ is constant and θ_t is a deterministic function of time t. θ_t is chosen so that, under the model's assumptions, the initial yield curve is implied to have the shape it actually has. The Ho–Lee model could also be described as the Merton model with time-dependent drift.

Bond and option prices have closed-form solutions in this model. Due to the lack of mean reversion and constant volatility of yields, however, the long-term behaviour of the yield curve is highly implausible.

2.2.2 Hull–White (1990)

This model could be described as the Ho–Lee model with mean reversion, or as the Vasicek model with time-dependent drift that implies a perfect fit to the initial term structure of interest rates observed in the market. The SDE of the short rate is

$$dr_t \;=\; \sigma\, dW_t + (\theta_t - \alpha r_t)\, dt \tag{2.6}$$

The model has the same amount of analytic tractability as Ho–Lee; bond and option prices have analytic solutions. The presence of mean reversion makes this model somewhat more realistic than the Ho–Lee model. However, no-arbitrage models generally have implausible long-term dynamics, because they interpret the initial yield curve roughly in the sense of the expectations theory of interest rates. That is, when the initial yield curve is upward-sloping, the model will exaggerate the tendency of interest rates to rise. This bias could be eliminated if the market price of risk were known and could be taken into account, but in practice this is too much to ask for.

Hull and White (1994b) have also proposed a two-factor model which bears roughly the same relation to the Brennan-Schwartz model as their one-factor model bears to the Vasicek model.

2.2.3 Black–Karasinski (1991)

In this one-factor model, the logarithm of the short rate follows a process similar to that followed by the short rate itself in the Hull–White model:

$$d(\ln r_t) \;=\; \sigma\, dW_t + (\theta_t - \alpha \ln r_t)\, dt \tag{2.7}$$

As a consequence, interest rates cannot go negative. This model has no closed-form solutions for bond prices, but according to Cairns (2004), there are efficient numerical methods for this.

2.2.4 Heath–Jarrow–Morton (1992)

Heath, Jarrow and Morton put forward a general framework for building one's own customized, multifactor no-arbitrage model. The Ho–Lee and Hull–White models can be derived as special cases.

For mathematical convenience, the HJM approach focuses on the forward curve (which contains exactly the same information as the yield curve). The behaviour of the forward curve is described by the SDE

$$d_t f(t,T) \;=\; \sum_{i=1}^{n} \sigma_i(t,T)\, dW_{i,t} + \alpha(t,T)\, dt \tag{2.8}$$

where $f(t,T)$ represents the instantaneous forward rate on maturity T at time t and the $W_{i,t}$ are independent Brownian motions. Thus the forward rate on each maturity evolves according to its own drift $\alpha(t,T)$ and volatility $\sigma(t,T)$, subject to the n sources of uncertainty that drive the model.

The HJM framework does not generally lead to closed-form bond pricing formulae, and is difficult to calibrate to the prices of traded instruments.

2.2.5 LIBOR market models

Market models came to prominence in the late 1990s due to Miltersen, Sandmann and Sondermann (1997), Brace, Gatarek and Musiela (1997) and Jamshidian (1997). Their main advantage over earlier no-arbitrage models is the ease with which they can be calibrated to the market prices of standard financial instruments. Whereas

earlier models centre around unobservable variables like the short rate and instantaneous forward rates, the new approach was to model the relevant market-quoted rates (such as the three-month LIBOR rate) directly. Coupled with the assumption that the future levels of these relevant rates are lognormally distributed, the derivation of no-arbitrage prices for many interest rate derivatives becomes extremely straightforward. Thus the popularity of market models is mainly due to their attractiveness to practitioners rather than economists.

2.3 Purely descriptive models

If one's goal is to model the long-run behaviour of the yield curve realistically, is it really important for the model to present no arbitrage opportunities, or to be based on an economic model? If by relaxing these requirements one could achieve a model that reproduced the relevant features of the movement of the yield curve more faithfully, then for many applications a purely instrumental model might be preferable. Ultimately, economists would like to find a realistic model based on a *theory*, something that would *explain* why rates move as they do—but in the meantime, there is no shame in trying to accomplish the more modest task of merely *describing* those movements as accurately and as parsimoniously as possible. It turns out that even this goal is more elusive than one might expect.

The affine and quadratic term structure models mentioned above in connection with equilibrium models ought perhaps to be assigned to the present category. Apart from that, the models we consider here all have a similar structure—the yield curve, itself a function $f : \mathbb{R}_+ \to \mathbb{R}_+$ from maturities to yields, is regarded as a linear combination of component functions, the weights on which are allowed to vary through time, producing various movements of the yield curve. We now illustrate this with two examples.

2.3.1 Nelson–Siegel (1987)

In this model, the yield curve at time t is the function

$$R(t,\tau) \;=\; \beta_{1,t} \;+\; \beta_{2,t}\left(\frac{1-e^{-\lambda_t\tau}}{\lambda_t\tau}\right) \;+\; \beta_{3,t}\left(\frac{1-e^{-\lambda_t\tau}}{\lambda_t\tau} - e^{-\lambda_t\tau}\right) \qquad (2.9)$$

where the betas act as multipliers on three additive components that control the level, slope and curvature of the yield curve. λ_t is a horizontal squeeze.

This static model can be extended into a dynamic one by specifying a dynamic process for the t-subscripted variables. It could be a multidimensional diffusion process, or a VAR process, if a discrete-time model is preferred. A simpler alternative is to use an independent process for each of the betas, but observe that this will ignore any dependencies between the betas.

In practice, λ_t is often redefined as constant, because the betas can then be estimated by OLS for each observation of the yield curve—after which a dynamic model can be estimated for each of the estimated betas.

2.3.2 Principal components analysis

An alternative for the three additive components of the Nelson–Siegel model, which are fundamentally arbitrary, is to estimate the components from data using an approach pioneered by Litterman and Scheinkman (1991). The observed yield curve is treated as a random vector, of which we have a set of observations; these are subjected to a principal components analysis. The factor loadings thus obtained play a role similar to that played by the artificial level, slope and curvature components of the Nelson–Siegel model, while the factor scores play a role corresponding to that of the betas.

The analysis then continues with the modelling of factor score dynamics. The PCA method conveniently guarantees that the estimated scores on the different factors are uncorrelated, allowing one to safely assume an independent stochastic process for each.

3 Description of the empirical data set

This chapter describes the data set used in the remainder of this study. It is stitched together from two sources. The first part comes from the US Federal Reserve; it is publicly available for download at `http://www.federalreserve.gov/releases/h15/data.htm` (under the heading 'Treasury constant maturities'). These constant-maturity yields have been calculated from the prices of traded US federal government bonds using a spline interpolation method explained on the website.

The second source is the financial news service Bloomberg. Bloomberg's data has both gaps and errors, some of them rather hard to spot, so we only use this as a secondary source to complement the Fed's data. Table 1 summarizes what periods we have data for, and from which source. For example, we are using Bloomberg data on 30-year yields from March 2002 to January 2006; in February 2006, Fed data becomes available again.

Table 3.1: The available data and its sources. Column headings indicate yield maturities. Within the table, m (d) indicates periods for which we have monthly (daily) data from the Fed. B stands for daily data from Bloomberg.

From	3m	6m	1y	2y	3y	5y	7y	10y	30y
1953-04			m		m	m		m	
1962-01	B	B	d		d	d		d	
1969-07	B	B	d		d	d	d	d	
1976-06	B	B	d	d	d	d	d	d	
1977-02	B	B	d	d	d	d	d	d	d
1982-01	d	d	d	d	d	d	d	d	d
2002-03	d	d	d	d	d	d	d	d	B
2006-02	d	d	d	d	d	d	d	d	d

Most of the data is daily data (excluding non-business days), but for the years 1953–1961 only monthly data is available. We make use of both monthly and daily data. Monthly yield levels are calculated as simple averages over the business days in that month.

In the Appendix, the data is analyzed statistically in order to assess the tenability of the stylized facts of Chapter 1. For this purpose, we carve up the data into

different subperiods in order to make optimal use of the data as the coverage of the different variables varies over time.

In the chapters below, however, where the data is used for estimating the parameters of various term structure models, we consistently use a single subset of the data: we only use the yields on maturities of three and six months as well as one, three, five and ten years, from January 1962 to the present. For yield levels, we use monthly values; but as we work with the Longstaff–Schwartz model, we also use day-to-day changes in the three-month yield for estimating the instantaneous variance of the short rate.

The yields reported in the data are yields-to-maturity, i.e. they represent the internal rates of return to the buyer of a treasury bond. As the term structure models we use are actually intended for modelling zero-coupon rates, any model calibration of which we report in chapters below will be preceded by a bootstrapping exercise where yields-to-maturity are converted to zero rates.[1]

The resulting data set has 615 observations of six variables (i.e. the yields on 3-month, 6-month, 1-, 3-, 5-, and 10-year bonds). Figure 3.1 displays their evolution through time; Table 3.2 shows some basic statistics on them.

Table 3.2 confirms that the average yield curve is upward-sloping, and that short-term yields are more volatile than long-term yields—albeit volatility peaks not in the very short end of the curve, but in the maturity of one year.

[1]The procedure involves the approximating assumption that the annual par rate for each maturity is equal to the yield-to-maturity.

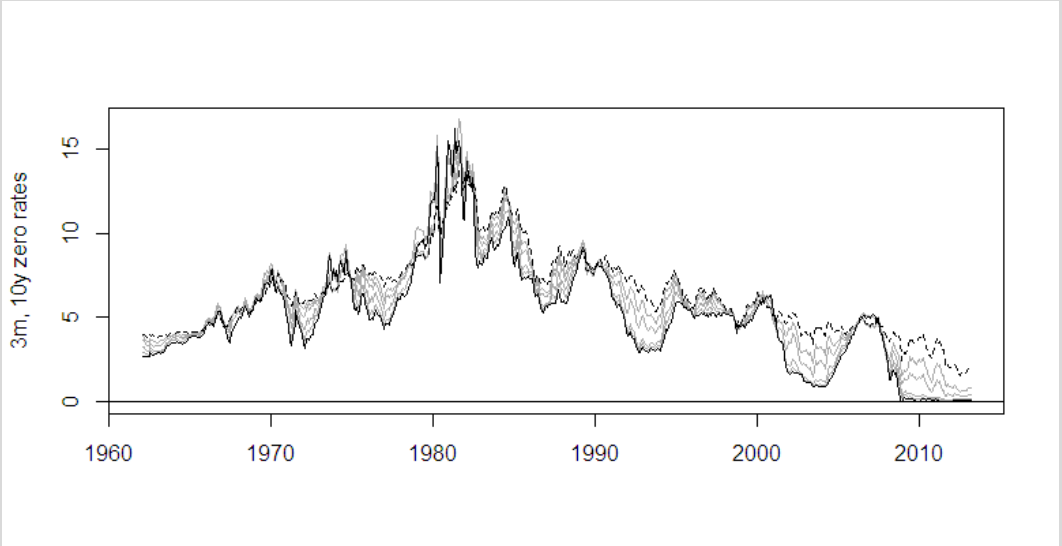

Figure 3.1: US government zero-coupon bond yields, Jan 1962–Mar 2013: 3-month (solid line), 10-year (dashed line) and 6-month, 1-year, 3-year and 5-year (grey lines).

Table 3.2: Basic descriptive statistics on the zero-coupon yield data set.

	3m	6m	1y	3y	5y	10y
Min	0.01	0.04	0.10	0.33	0.62	1.56
Mean	5.19	5.37	5.64	5.84	6.10	6.40
Max	16.18	15.46	16.72	14.87	14.61	13.94
StDev	3.09	3.11	3.23	2.87	2.73	2.50

4 Specification of the Cost-at-Risk measure

We now introduce a particular application of term structure modelling in the management of public debt. Consider a government which seeks to adopt a systematic debt issuance strategy defining the maturities of new bond issues and the issuance schedule. The government wants its borrowing costs to be both predictable and low, but there is a trade-off between these objectives: on average, short-term borrowing brings savings, since the yield curve is upward-sloping on average, but short-term yields are also more volatile, so the short-term borrower may end up paying more.

In order to make an informed choice of strategy, the government needs a risk measure that can be weighed against the projected mean cost of funding. *Value-at-Risk* (VaR) is a widely used measure of the risk of loss on a portfolio of assets; something similar could perhaps be used for the problem at hand. For a given portfolio, probability p and time-frame τ, VaR is defined as the threshold value $X_{p,\tau}$ such that the probability that the cumulative loss on the portfolio during τ exceeds this threshold is p. For example, if a portfolio has a one-day 5% VaR of \$1 million, then the probability of a one-day loss in excess of \$1 million is 5%.

We can define an analogous measure, *Cost-at-Risk* (CaR), to describe the distribution of debt costs: let the (say) 95% one-year CaR for a debt issuance strategy denote the threshold value such that there is a 5% probability that the debt costs over a given year exceed this threshold.[1]

Note that whereas VaR refers to the potential loss on a given *portfolio*, we define our CaR measure to refer to the potential annual debt costs generated by a given debt issuance *strategy*. And whereas VaR is calculated for an imminent time-frame

[1]For clarification, the term *Cost-at-Risk* was not coined by us. While its origin is unclear to us, we know that it has been used by risk analysts at many central banks and governmental debt management offices already in the 1990s. Yet there appears to be a fair amount of ambiguity in its use; sometimes it refers to a quantile of the steady-state distribution of the cost of debt, sometimes to a quantile of the next period's debt costs given current market conditions. We use the term in the first sense. So far, the concept does not seem to have attracted any academic interest, as we were unable to find any peer-reviewed papers on the topic. Instead, the reader is referred to e.g. Danmarks Nationalbank (2000) and Jönsson and Sangarabalan (2010). Regarding the advantages and shortcomings of CaR as a risk management tool, the literature on Value-at-Risk may be of some use.

under current market conditions (so that it describes the *conditional* distribution of portfolio returns), we define CaR as a quantile of the long-run, or steady-state distribution of debt costs (so that it describes their *unconditional* distribution).

Calculating VaR for an equity portfolio is simple—stock returns are so erratic that for the purpose of VaR they are quite well modelled by the multivariate normal distribution. The VaR of a bond portfolio is harder to determine, because bond yields move together in a complicated pattern. To capture this behaviour and calculate VaR, a term structure model is necessary.

By the same token, calculating CaR for a debt issuance strategy requires term structure modelling. To do this, one will have to decide which model to use; and to decide this, one needs to know which one of all the term structure models out there is best for this purpose. Although there is some literature on the empirical performance of different term structure models, there is no ready answer to this question available. In this study we examine one particular model—the Longstaff–Schwartz model—by comparing its performance on pseudo-data against that of a simple benchmark model.

In order to do such comparisons, we must specify a debt issuance strategy whose costs we will be trying to estimate. A related issue is how costs should be measured—at cash flow basis or some other basis.

Measuring costs on a cash flow basis would require us to define CaR as a quantile of the annual net nominal cash flow. We also need to decide where this cash comes from—reduced net spending or extra borrowing—because the interest payments of any given year depend on the amount of debt incurred in preceding years. With using the cash flow basis for cost accounting there also comes the slight nuisance of having to calculate par rates for bonds from zero rates.[2]

Alternatively, one could assume that debt is issued in the form of zero-coupon bonds, and that redemptions are funded by new borrowing, but aside from that, no new debt is issued. Debt costs to be allocated to a given year would be defined as the change in the market value of the debt portfolio over that year. This approach has two advantages: it is budget-neutral, and calculation of par rates is avoided. Unfortunately, there are disadvantages. Firstly, the cost of debt is excessively sen-

[2]The par rate for a bond is the coupon rate such that if the bond paid this coupon, its present value (given the current zero rates) would be equal to its redemption value.

sitive to the level of rates prevailing at the end of the period, making CaR larger. Secondly, low interest rates are associated with high costs (because the market value of debt increases), which is counterintuitive. Thirdly, a rise in interest rates may generate a negative cost of debt, which is also counterintuitive.

Therefore, we adopt a third method. We assume budget neutrality and funding by zero coupon bonds, as in the previous method, but we decompose each bond into a series of zero-coupon forward loans. The debt cost to be allocated to a given period is the cost of the forward loans belonging to that period.

To illustrate, suppose a T-bond is issued at time t, and we want to calculate the part of its cost belonging to a time segment $(T_1, T_2) \subset (t, T)$. Using the proposed accounting basis, by time T_1 the book value of the bond will have grown to $\frac{P(t,T)}{P(t,T_1)}$ (which is the forward price of a T-bond at time t for delivery at time T_1). By time T_2, the book value will have grown to $\frac{P(t,T)}{P(t,T_2)}$. Therefore, the book value will grow by a factor of $\frac{P(t,T_1)}{P(t,T_2)}$, so that the continuously compounded growth rate g is

$$g \;=\; \frac{1}{T_2 - T_1} \ln\left(\frac{P(t,T_1)}{P(t,T_2)}\right) \;=\; f(t,T_1,T_2) \tag{4.1}$$

i.e. the cost generated by a single bond over a given period is the forward rate. Thus the cost of debt is fixed when a bond is issued, and is immune to later yield curve movements. This cost allocation basis makes economic sense in that the forward rate represents the true opportunity cost of not having to obtain new funding for the period (T_1, T_2) at going rates at time T_1.

The continuously compounded cost rate $c(T_1, T_2)$ of the entire debt portfolio is

$$c(T_1,T_2) \;=\; \frac{1}{T_2 - T_1} \ln\left(\frac{\Lambda(T_2)}{\Lambda(T_1)}\right) \tag{4.2}$$

where $\Lambda(T_i)$ is the sum of the book values of the bonds at time T_i.

We are now ready to complete our CaR specification. We calculate the CaR for a one-year period $t \in (0, 1)$. We examine a government which issues debt in the form of 1-, 5- and 10-year zero-coupon bonds at the start of every year. For simplicity, we assume that each bond in the debt portfolio has unit book value at time $t = 0$, so that $\Lambda(0) = 16$. The redemption of each maturing bond is funded with a new

issue of the same maturity. At $t = 1$, the book value of the portfolio will be

$$\Lambda(1) \;=\; \frac{1}{P(0,1)} + \sum_{t=-4}^{0} \frac{P(t,0)}{P(t,1)} + \sum_{t=-9}^{0} \frac{P(t,0)}{P(t,1)} \tag{4.3}$$

and the continuously compounded cost rate is

$$c(0,1) \;=\; \ln \Lambda(1) - \ln 16 \tag{4.4}$$

This issuance strategy has the flaw that the composition of the debt portfolio is unstable. If redemptions are funded with new issues of the same maturity, and long-term yields tend to be higher than shorter-term yields, then the share of long-term bonds in the portfolio will grow over time.

In principle, the flaw is not hard to fix, as nothing prevents new bond issues from being sized to preserve the desired maturity profile. We could develop a mechanistic redemption-refinancing rule which would take care of this. In practice, however, introducing such a rule would sacrifice the simplicity of formulas (4.3) and (4.4), and given our one-year time horizon, the effect on the results would be small. For the sake of mathematical tractability, therefore, we stick with the issuance strategy already described.

For the sake of readability, we choose to express the debt cost rate, as well as CaR, in percentage points: a CaR of, say, 12.95 really means 0.1295, but is more readable due to the convention of expressing interest rates in percentage points rather than fractions.

5 The Longstaff–Schwartz model

In this chapter we examine the model whose suitability for CaR estimation is the subject of this study. First we explain the structure of the model to the extent necessary for understanding its practical application. Then we calibrate the model to empirical data, experiment with CaR estimation, and develop an improved method of calibrating the model to data.

5.1 Model specification

The Longstaff–Schwartz (1992) model is based on a two-variable version of the Cox–Ingersoll–Ross (1985a) general equilibrium asset pricing model. Its differences compared to the CIR term structure model stem from differences in the stochastic processes describing real production possibilities in the economy. The realized rate of return on capital Q is described by

$$\frac{dQ}{Q} = (\mu X + \theta Y)\, dt + \sigma\sqrt{Y}\, dZ_1 \tag{5.1}$$

where μ, θ and σ are positive constants with $\theta > \sigma^2$, X and Y are state variables, and Z_1 is a Brownian motion. The expected rate of return thus depends on the state variables, one of which is related and the other unrelated to uncertainty concerning immediate production. The dynamics of these factors are governed by

$$dX = (a - bX)\, dt + c\sqrt{X}\, dZ_2 \tag{5.2}$$

$$dY = (d - eY)\, dt + f\sqrt{Y}\, dZ_3 \tag{5.3}$$

where $a, ..., f$ are positive constants and Z_2, Z_3 are Brownian motions, with Z_2 independent of Z_1 and Z_3. The form of the SDEs guarantees that X and Y stay positive.

The economy is inhabited by a fixed number of identical individuals with time-

additive preferences of the form

$$\mathbb{E}_t\Big[\int_t^\infty \exp(-\rho s)\ln(C_s)\,ds\Big] \tag{5.4}$$

where $\mathbb{E}_t$ signifies conditional expectation, ρ is the utility discount factor, and C_s represents time-s consumption. Frictionless and perfectly competitive markets exist for trading in both riskless and risk-bearing financial claims.

Rescaling the state variables into $x = X/c^2$ and $y = Y/f^2$ gives

$$dx = (\gamma - \delta x)\,dt + \sqrt{x}\,dZ_2 \tag{5.5}$$

$$dy = (\eta - \xi y)\,dt + \sqrt{y}\,dZ_3 \tag{5.6}$$

where $\gamma = a/c^2$, $\delta = b$, $\eta = d/f^2$ and $\xi = e$. These SDEs are easier to work with, and Longstaff and Schwartz derive from them a formula for the risk-free, instantaneous rate of interest:

$$r = \alpha x + \beta y \tag{5.7}$$

where $\alpha = \mu c^2$ and $\beta = (\theta - \sigma^2)f^2$, guaranteeing that r will be positive for all feasible values of the state variables. The variance of the instantaneous rate of change of r is

$$V = \mathrm{Var}\left(\frac{dr}{dt}\right) = \alpha^2 x + \beta^2 y \tag{5.8}$$

While we work with the model, our attention will be on r and V rather than the underlying state variables X and Y. Note, however, that there is a one-to-one mapping between (r, V) and (X, Y). All of these variables are considered observable in the model economy.

α, β, γ, δ, η and ξ are collectively referred to as the *stationary parameters* of the model. The dynamics of r and V are exhaustively determined by the stationary

parameters, and can be expressed in the form of the following SDEs:

$$\begin{aligned} dr &= \left(\alpha\gamma + \beta\eta - \frac{\beta\delta - \alpha\xi}{\beta - \alpha} r - \frac{\xi - \delta}{\beta - \alpha} V\right) dt \\ &\quad + \alpha\sqrt{\frac{\beta r - V}{\alpha(\beta - \alpha)}}\, dZ_2 \;+\; \beta\sqrt{\frac{V - \alpha r}{\beta(\beta - \alpha)}}\, dZ_3 \end{aligned} \tag{5.9}$$

$$\begin{aligned} dV &= \left(\alpha^2\gamma + \beta^2\eta - \frac{\alpha\beta(\delta - \xi)}{\beta - \alpha} r - \frac{\beta\xi - \alpha\delta}{\beta - \alpha} V\right) dt \\ &\quad + \alpha^2\sqrt{\frac{\beta r - V}{\alpha(\beta - \alpha)}}\, dZ_2 \;+\; \beta^2\sqrt{\frac{V - \alpha r}{\beta(\beta - \alpha)}}\, dZ_3 \end{aligned} \tag{5.10}$$

The shape and movements of the rest of the yield curve, however, depend not only on r and V and the stationary parameters, but also on the pricing of risks in the market. In the CIR asset valuation framework, market prices of risk are determined endogenously insofar as they exist. In the Longstaff–Schwartz model, it turns out that the market price of risk related to changes in the uncertainty of production, governed by Y, is λy for some constant λ (i.e. it is proportional to y and Y), while risk stemming from X cannot be hedged and therefore has no price.

Given the values of r, V, the stationary parameters, and λ, the zero-coupon yields on finite maturities can be calculated analytically using the following formula:

$$R(\tau) \;=\; -(\kappa\tau + 2\gamma \ln A(\tau) + 2\eta \ln B(\tau) + C(\tau) r + D(\tau) V)/\tau \tag{5.11}$$

where

$$\begin{aligned} A(\tau) &= \frac{2\phi}{(\delta + \phi)(e^{\phi\tau} - 1) + 2\phi} \\ B(\tau) &= \frac{2\psi}{(\nu + \psi)(e^{\psi\tau} - 1) + 2\psi} \\ C(\tau) &= \frac{\alpha\phi(e^{\psi\tau} - 1)B(\tau) - \beta\psi(e^{\phi\tau} - 1)A(\tau)}{\phi\psi(\beta - \alpha)} \\ D(\tau) &= \frac{\psi(e^{\phi\tau} - 1)A(\tau) - \phi(e^{\psi\tau} - 1)B(\tau)}{\phi\psi(\beta - \alpha)} \end{aligned}$$

and

$$\begin{aligned}
\nu &= \xi + \lambda \\
\phi &= \sqrt{2\alpha + \delta^2} \\
\psi &= \sqrt{2\beta + \nu^2} \\
\kappa &= \gamma(\delta + \phi) + \eta(\nu + \psi)
\end{aligned}$$

Since x and y both follow a process similar to the one that the short rate follows in the one-factor CIR model, the distributional properties of that process also apply to x and y. The unconditional distribution of each is the gamma distribution: $x \sim \Gamma(2\gamma, \frac{1}{2\delta})$ and $y \sim \Gamma(2\eta, \frac{1}{2\xi})$.[1] Given (5.7) and (5.8), the means and variances of r and V are easily calculated using the properties of the gamma distribution:

$$\mathbb{E}(r) = \frac{\alpha\gamma}{\delta} + \frac{\beta\eta}{\xi} \tag{5.12}$$

$$\mathrm{Var}(r) = \frac{\alpha^2\gamma}{2\delta^2} + \frac{\beta^2\eta}{2\xi^2} \tag{5.13}$$

$$\mathbb{E}(V) = \frac{\alpha^2\gamma}{\delta} + \frac{\beta^2\eta}{\xi} \tag{5.14}$$

$$\mathrm{Var}(V) = \frac{\alpha^4\gamma}{2\delta^2} + \frac{\beta^4\eta}{2\xi^2} \tag{5.15}$$

Getting the unconditional distribution of r and V right is useful in view of simulating the process, because if one cannot draw the starting values randomly from the correct distribution, it is difficult to ensure that one's simulation outcomes are not distorted by the way the starting values are determined. An alternative to drawing the starting values directly from the correct distribution is to use a so-called burn-in period, letting the simulation process run idle for a long time before making any observations of it. With an adequate simulation program, r and V will eventually converge-in-distribution to their unconditional joint distribution. Obviously, this approach will be computationally expensive.

[1] There are different parameterizations in use for the gamma distribution; we are using the one where $x \sim \Gamma(k, \theta)$ implies that $\mathbb{E}(x) = k\theta$ and $\mathrm{Var}(x) = k\theta^2$.

5.2 Parameter estimation

How one should go about estimating the model parameters from data depends on whether the variance series V is observable. Longstaff and Schwartz assume that it is not, so they start by estimating $\hat{V}$ from r. Using a series of one-month Treasury bill rates as a proxy for r, they estimate a GARCH model of the following form:

$$r_{t+1} - r_t = \alpha_0 + \alpha_1 r_t + \alpha_2 V_t + \epsilon_{t+1} \tag{5.16}$$

$$\epsilon_{t+1} \sim N(0, V_t)$$

$$V_t = \beta_0 + \beta_1 r_t + \beta_2 V_{t-1} + \beta_3 \epsilon_t^2 \tag{5.17}$$

Due to the two-way causality between the two difference equations, this model is much less straightforward to calibrate than a vanilla GARCH model, where V would influence r only through ϵ and r would not influence V at all. Instead of following the example of Longstaff and Schwartz, we are going to estimate our variance series from the variance of day-to-day changes of our empirical proxy for the short rate, the three-month US treasury bill yield.

Given the time series for r and $\hat{V}$, the stationary parameters can be estimated with a simple method proposed by Longstaff and Schwartz. It can be assumed without loss of generality that $\alpha < \beta$, which implies that $\alpha < \frac{V}{r} < \beta$ at all times. This condition will be (almost) satisfied by our data if we set $\hat{\alpha} = \min\left(\frac{\hat{V}}{r}\right)$ and $\hat{\beta} = \max\left(\frac{\hat{V}}{r}\right)$.

Once we have an estimate for both α and β, we can solve the system of equations (5.12)–(5.15) for the remaining stationary parameters:

$$\hat{\delta} = \frac{\hat{\alpha}(\hat{\alpha} + \hat{\beta})(\hat{\beta}\overline{r} - \overline{V})}{2\big(\hat{\beta}^2 \mathrm{Var}(r) - \mathrm{Var}(\hat{V})\big)} \tag{5.18}$$

$$\hat{\gamma} = \frac{\hat{\delta}(\hat{\beta}\overline{r} - \overline{V})}{\hat{\alpha}(\hat{\beta} - \hat{\alpha})} \tag{5.19}$$

$$\hat{\xi} = \frac{\hat{\beta}(\hat{\alpha} + \hat{\beta})(\overline{V} - \hat{\alpha}\overline{r})}{2\big(\mathrm{Var}(\hat{V}) - \hat{\alpha}^2 \mathrm{Var}(r)\big)} \tag{5.20}$$

$$\hat{\eta} = \frac{\hat{\xi}(\overline{V} - \hat{\alpha}\overline{r})}{\hat{\beta}(\hat{\beta} - \hat{\alpha})} \tag{5.21}$$

Table 5.1: Estimation results for the Longstaff–Schwartz model parameters.

	L&S (1993)	Sauli (2013)
Mean(r)	0.06717	0.05186
Var(r)	7.157×10^{-4}	9.543×10^{-4}
Mean($\hat{V}$)	7.658×10^{-4}	2.521×10^{-4}
Var($\hat{V}$)	1.526×10^{-6}	4.811×10^{-7}
Min($\hat{V}/r$)	0.001149	3.525×10^{-5}
Max($\hat{V}/r$)	0.1325	0.2116
$\hat{\alpha}$	0.001149	3.525×10^{-5}
$\hat{\beta}$	0.1325	0.2116
$\hat{\gamma}$	3.0493	1.3608
$\hat{\delta}$	0.05658	9.466×10^{-4}
$\hat{\eta}$	0.1582	0.0651
$\hat{\xi}$	3.998	11.648
$\hat{\lambda}$	−3.663	−10.692

where $\overline{r}$, $\overline{V}$, Var(r) and Var($\hat{V}$) refer to the sample means and variances of r and $\hat{V}$.

Finally, λ can be estimated by minimizing the sum of squared differences between the (currently) observed yields and the theoretical yield curve given by (5.11).

Longstaff and Schwartz (1993) estimate all six stationary parameters from data using this approach. Their results are reported in Table 5.1.

Now we are at a position to examine what Longstaff and Schwartz yield curves look like under the parameter values obtained by them. Figure 5.1 gives you an idea. The three alternative short-rate levels depicted correspond to the 0.001, 0.5 and 0.999 quantiles of the steady-state distribution of r under these parameter values. For each value of r, two borderline-possible curves are displayed. These correspond to the supremum and infimum values of V allowed. The infimum of V, given r, is αr (solid lines), and the supremum is βr (dotted lines). For intermediate values of V the curve would lie somewhere between the solid and dashed curves.

It is interesting to note that when both the level and volatility of the short rate are high, the yield curve becomes inverted. This feature is related to the dynamic behaviour of the yield curve under the parameter values obtained by Longstaff and Schwartz, and we will come back to it as we discuss the simulation of the process.

Let us now estimate the values of the stationary parameters using our data set, consisting of monthly US government bond yields from 1962–2013. The shortest maturity we have is three months, so we will be using this as a proxy for the short

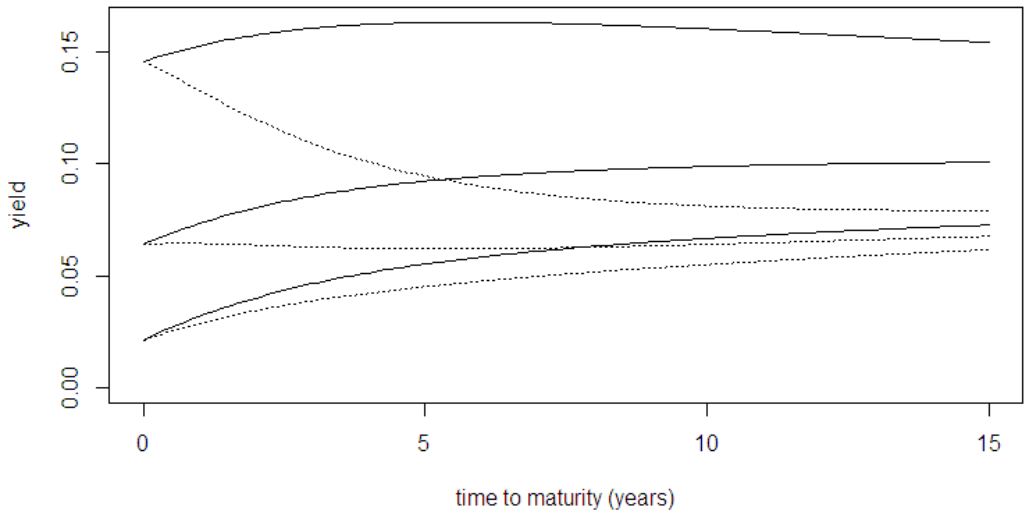

Figure 5.1: Yield curves for different (r, V). The three values of r are for the 0.01, 0.5 and 0.99 quantiles of the unconditional distribution of r. The solid lines are for $V = \alpha r$, i.e. the infimum values of V given r; the dotted lines are for $V = \beta r$, i.e. the supremum values of V given r.

rate. Also, as we have daily data available as well, we can calculate a variance of daily changes for each month in the data; this way we do not have to use a GARCH model to estimate the variance, and the observed variance is likely to be a much more reliable indication of the true variance than a GARCH estimate.[2]

Following the steps outlined above, we get the parameter estimates given in the second column of Table 5.1. They differ considerably from the estimates obtained by Longstaff and Schwartz. Notably our estimates for α and δ are almost two orders of magnitude smaller. This has consequences for the dynamics of the process and the shapes of the yield curves; we discuss them in the next section.

One possible reason why the estimates might differ is that Longstaff and Schwartz had a smaller data set—from the years 1964–1989—so we tried estimating our parameters using the corresponding subset of our data, but the discrepancy was not

[2]The variable V represents the *instantaneous* variance of r, meaning that the variance of the yield change over the infinitesimally short time interval dt is equal to $V\,dt$. The variance of the yield change over a short finite interval Δt is approximately $V\Delta t$. Now, there are roughly 250 trading days in a year, and yields are known to change no more during a weekend than between any two consecutive business days (cf. e.g. Hull, 2006). Therefore, when time is measured in years and the data consists of daily observations, $\Delta t \approx 0.004$, so that V for a given month can be estimated as $250 \times \mathrm{Var}(\Delta r)$ over values of Δr recorded in that month.

eliminated. Our guess is that the discrepancy is due to Longstaff and Schwartz's method of estimating V from r using a GARCH model. We doubt the ability of the GARCH model to appreciate the extremely low volatilities that sometimes occur; this leads to overestimating V during times of very low volatility, and thus to overstating α. However, without carrying out the GARCH estimation this remains but a conjecture.

5.3 Simulation

There are at least two ways to simulate paths for the state variables r and V. The most obvious method is to set up a discrete-time version of (5.9) and (5.10):

$$\begin{aligned} \Delta r &= \left(\alpha\gamma + \beta\eta - \frac{\beta\delta - \alpha\xi}{\beta - \alpha} r - \frac{\xi - \delta}{\beta - \alpha} V\right) \Delta t \\ &\quad + \alpha\sqrt{\frac{\beta r - V}{\alpha(\beta - \alpha)}}\,\epsilon_{1,t} \;+\; \beta\sqrt{\frac{V - \alpha r}{\beta(\beta - \alpha)}}\,\epsilon_{2,t} \end{aligned} \tag{5.22}$$

$$\begin{aligned} \Delta V &= \left(\alpha^2\gamma + \beta^2\eta - \frac{\alpha\beta(\delta - \xi)}{\beta - \alpha} r - \frac{\beta\xi - \alpha\delta}{\beta - \alpha} V\right) \Delta t \\ &\quad + \alpha^2\sqrt{\frac{\beta r - V}{\alpha(\beta - \alpha)}}\,\epsilon_{1,t} \;+\; \beta^2\sqrt{\frac{V - \alpha r}{\beta(\beta - \alpha)}}\,\epsilon_{2,t} \end{aligned} \tag{5.23}$$

where $\epsilon_{1,t}, \epsilon_{2,t} \sim N(0, \Delta t)$ i.i.d. and $\mathrm{Cov}(\epsilon_{1,t}, \epsilon_{2,t}) = 0$.

The most obvious method may not be the best, however. For one thing, a very small time-step Δt is required in order to ensure that the use of normally distributed disturbances does not distort the distribution. This may not be a big problem, given that random normal deviates are fast to generate. But secondly, the use of Gaussian disturbances causes r and V to occasionally take on illegal values: either one can become negative, or the ratio $\frac{V}{r}$ can stray below α or above β. Such situations will cause the simulation to fail, unless they are handled e.g. by canceling the update of both variables whenever one or both of them would take on an illegal value. That will prevent the simulations from crashing; however, we do not know how it will affect the probability distribution of the simulation outcomes.

To ensure that the simulation results have exactly the right probabilistic qualities in every respect, it would be better to dig deeper into the foundations of the model

and simulate paths for x and y rather than for r and V directly. x and y both follow a square root process of the same form as the short rate in the one-factor Cox–Ingersoll–Ross model. The increments of this process have a non-central chi-squared distribution, and the steady-state distribution is the gamma distribution. Modern statistical software packages (such as R, which was used for this study) can generate random chi-squared and gamma deviates.

To be more precise, our simulations are generated as follows. The simulating function is passed the following arguments: a length T of the desired output series of r and V, measured in years; an output frequency f, so that the number of observations in the output series is Tf; and a parameter vector containing the stationary parameter values. At the first stage, starting values for x and y are drawn randomly from the $\Gamma(2\gamma, \frac{1}{2\delta})$ and $\Gamma(2\eta, \frac{1}{2\xi})$ distributions, respectively. Subsequent values for x and y are drawn randomly from the non-central chi-squared distribution, as described in Cox, Ingersoll and Ross (1985b). Finally, r and V are obtained as $\alpha x + \beta y$ and $\alpha^2 x + \beta^2 y$, respectively.

Figure 5.2 illustrates the steady-state distribution of r and V when the model parameters have the values estimated by Longstaff and Schwartz (1993). The grey area represents the theoretical range of r and V given these parameter values, and the dots are (r, V) pairs generated by sampling x and y from their respective gamma distributions.

Figures 5.3 and 5.4 show individual simulated paths for r (solid line) and V (black area). The ten-year yield is also displayed (dotted line). The realization graphed in Figure 5.3 was generated using the model parameter values estimated by Longstaff and Schwartz (1993); that in Figure 5.4, using parameter values estimated from our US government bond data set.

It is immediately clear that the process behind Figure 5.4 is not a realistic description of yield curve movements. The simulation spans a period of 50 years, and the base level of yields hardly changes, except that the short rate spikes up now and then. The spikes are linked to simultaneous and equally short-lived volatility bursts; when the burst subsides, the short rate returns to its pre-spike level. This base level, from which the short rate makes its brief departures, changes over time, but only extremely slowly. The unconditional distribution of r has a standard deviation of

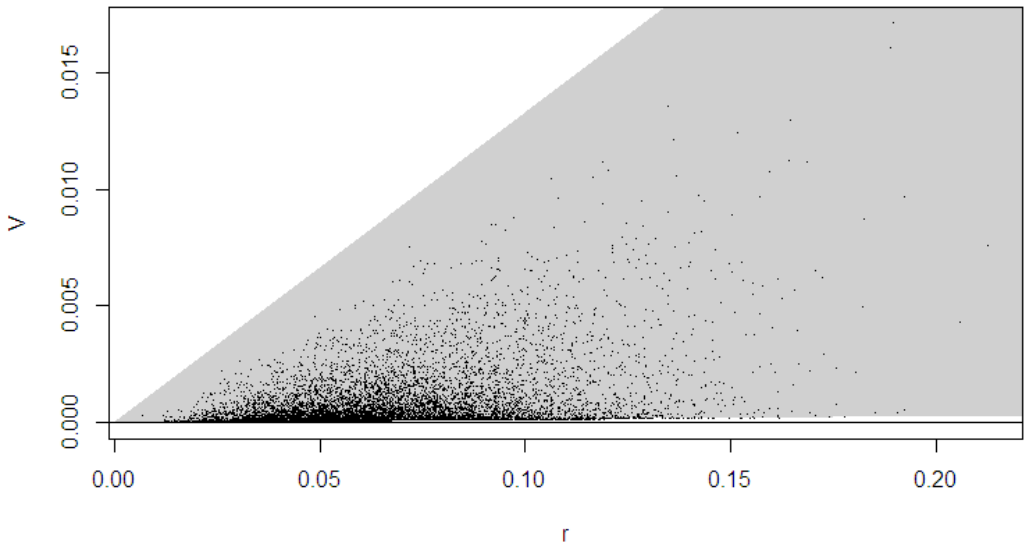

Figure 5.2: 10,000 pairs of r and V polled at random from their steady-state joint distribution under the stationary parameter values estimated by Longstaff and Schwartz. The grey area represents the theoretical range of (r, V).

about 0.03 and the middle 90% range is $[0.013, 0.111]$, but for the realization—which is *typical*—the figures are 0.003 and $[0.032, 0.042]$.

Another feature is that the spread of the ten-year yield over the short rate is almost constant. The ten-year yield moves in lockstep with the short rate otherwise, but does not shoot up at the spikes: the market seems to know that the spikes are transitory, whereas the base level of the short rate is persistent.

The simulations using Longstaff and Schwartz's parameter values seem to exhibit these same qualities, albeit to a lesser degree. The short rate spikes up when variance peaks, then promptly returns to its previous level. The ten-year yield displays only slight responsiveness to the yanks of the short rate, but otherwise mirrors it faithfully. This is where we come back to the yield curve shapes displayed in Figure 5.1 and the phenomenon that they slope up when variance is low, but down when variance is high. Namely, when a high level of r is the result of a spike, the market knows it will not last, but if r is high even though V is low, the market is in the "normal" mode, i.e. not experiencing a spike, and rates are unlikely to plunge very quickly.

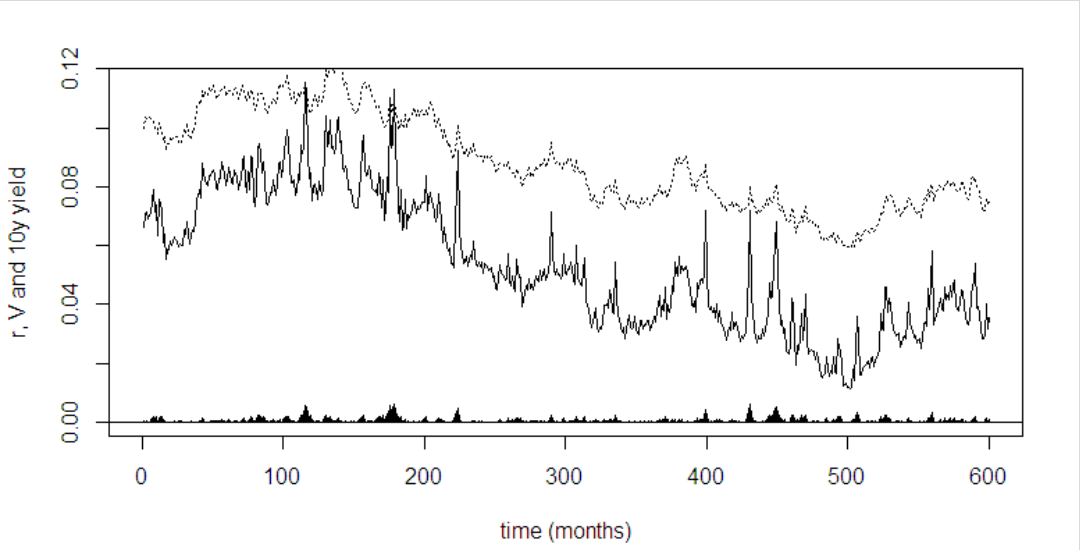

Figure 5.3: A typical realization of the Longstaff–Schwartz model using their parameter estimates: r (solid line), V (black area), and the 10-year yield (dotted line).

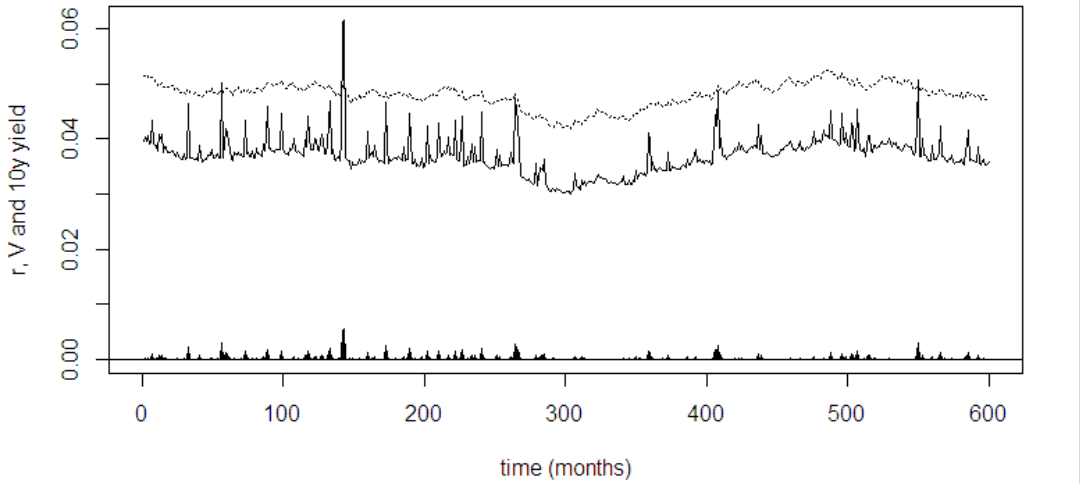

Figure 5.4: A typical realization of the Longstaff–Schwartz model using our parameter estimates: r (solid line), V (black area), and the 10-year yield (dotted line).

Although the interest rate process depicted in Figure 5.4 looks implausible, the framework underlying Longstaff and Schwartz's model guarantees that there is nothing internally inconsistent about it. Yet the model clearly fails to describe the behaviour of interest rates as reflected in our data set. What went wrong? Dismissing the possibility that our data might badly misrepresent actual market dynamics, two alternatives remain: either the Longstaff–Schwartz model might not be a good model of the historical yields-generating process, or the estimation technique might

be flawed. In what follows, we focus on the latter alternative.

5.4 Re-estimation from simulated data

We now conduct an experiment where we simulate interest rates using the Longstaff–Schwartz parameter values, then re-estimate the parameter values from the simulated data to see how close to the original parameters we would typically get.

Of course, the results will be affected by the length of the simulations: the longer the simulated data set, the closer the sample distribution of r and V is likely to be to their steady-state distribution. And conversely, if we use a small data set, the sample will not be representative of the entire unconditional distribution. Moreover, we must decide whether we take V to be directly observable, or whether we should estimate it from changes in r.

We start with generous assumptions, setting the simulation length at 200 years (2400 monthly observations) and taking V to be directly observable. Figure 5.5 displays the outcomes from 100 simulations. The graphs on the top row show the minimum and maximum values of $\frac{V}{r}$ in each simulation; the dashed lines represent the minimum and maximum feasible values of this ratio given the actual parameters of the data-generating process. The middle row shows the sample mean and variance of r from each simulation, and the bottom row those of V; again, the dashed lines represent the corresponding population means and variances.

Since α is estimated as $\min\left(\frac{V}{r}\right)$ and β as $\max\left(\frac{V}{r}\right)$, we learn from the top two plots that α is always correctly estimated and β is always underestimated. This is not surprising if one takes another look at Figure 5.2: random pairs of (r, V) cluster along the α side of the feasible sector, while the β side is never reached.

The other statistics may be either over- or underestimated. Mean values of r and V are more accurately estimated, in terms of relative variance of the estimator, than their variances. The errors in the estimated mean and variance of r are positively correlated, and likewise for V: $\mathrm{Cor}[\overline{r}, \mathrm{Var}(r)] = 0.59$ and $\mathrm{Cor}[\overline{V}, \mathrm{Var}(V)] = 0.87$.

And how does all this affect the properties of the yield curve model under the estimated parameter values? Figure 5.6 shows some more implied statistics. Continuing to use the same 100 simulations as previously, we estimate the model parameters in each case, and derive the unconditional mean of r and the unconditional 5th and

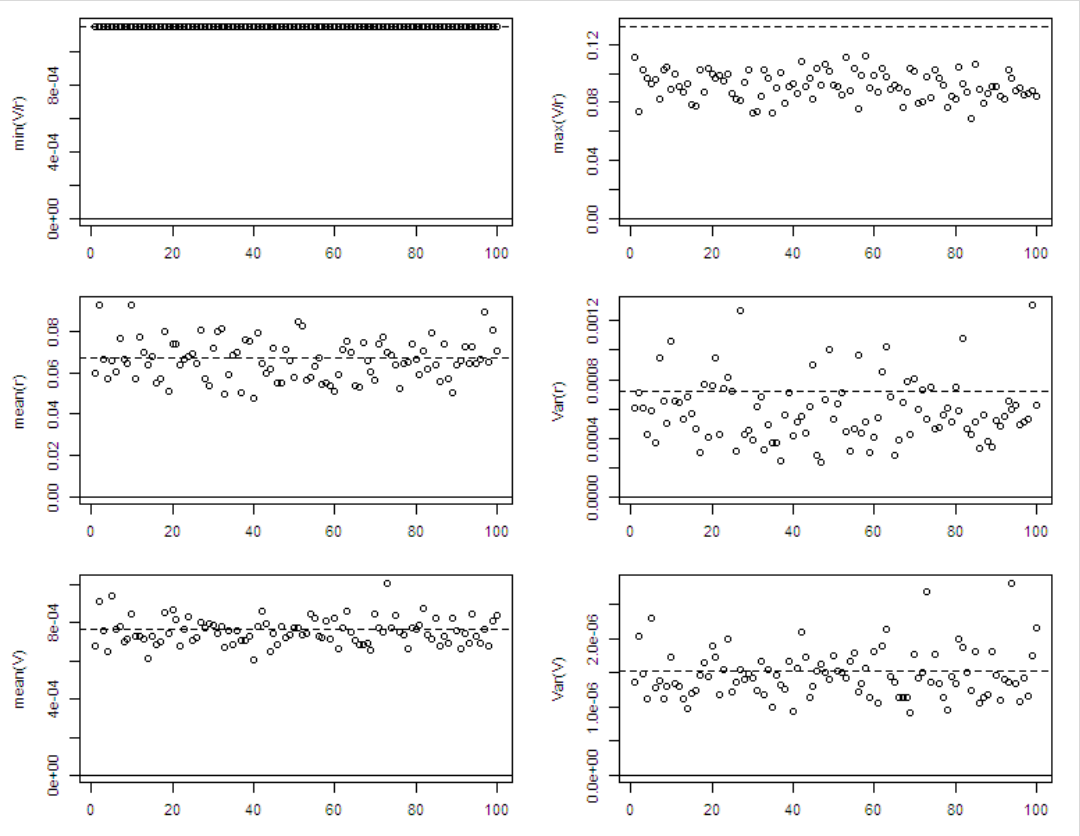

Figure 5.5: Re-estimating min $\left(\frac{V}{r}\right)$, max $\left(\frac{V}{r}\right)$, mean(r), Var(r), mean(V) and Var(V) from simulated data. The dashed line represents the population statistic, the dots represent sample statistics from 100 simulation runs.

95th percentiles of r. The dashed lines represent the 'real' values of these statistics. We find that while the estimated 90% band tends to be somewhat too narrow, the overall quality of the estimates is promising.

Of course, as Figures 5.5 and 5.6 only describe the long-run distribution of r, they say nothing about the dynamic properties of interest rates implied by the estimated parameter values.

Matters get worse when we abandon the very generous assumptions of 200 years of data and observable V. Instead, we assume we have only 615 monthly observations, i.e. a little over 51 years, like in our actual US government bond yield data set. Also, we now simulate our sets of pseudodata at a higher frequency in order to estimate the monthly variance of r from the sample variance of day-to-day yield changes.

Figure 5.7 displays the results. The accuracy of the estimation procedure is

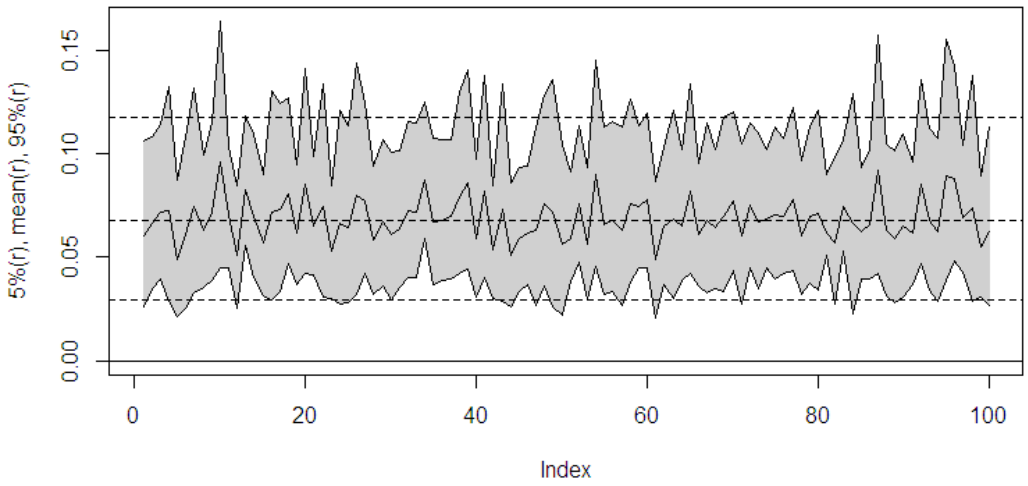

Figure 5.6: Solid lines: 5th percentile, mean and 95th percentile of r in the steady-state distributions implied by the statistical estimates from the 100 simulations. Dashed lines represent the 'real' values of these statistics.

eroded significantly under these less favourable conditions, and thus the distributions estimated for r differ much more from the population distribution than they did when there was more data and V was directly observable. The main reason for this decline is not hard to recognize: the estimation procedure advocated by Longstaff and Schwartz effectively assumes that the distribution of r and V in the sample is a reasonably good proxy for their long-run distribution. This assumption often fails when the data set spans only a short period of time, the most usual result being that the estimated range of r is too narrow, and at times severely displaced, as illustrated by Figure 5.7.

So far, we have been discussing re-estimation results when the data-generating process has the stationary-parameter values estimated by Longstaff and Schwartz. What if we used our own estimates instead? Unfortunately, this invariably leads to estimates which violate the model's assumptions: γ and δ are estimated to be negative, which makes it impossible to generate further simulations or draw any yield curves.

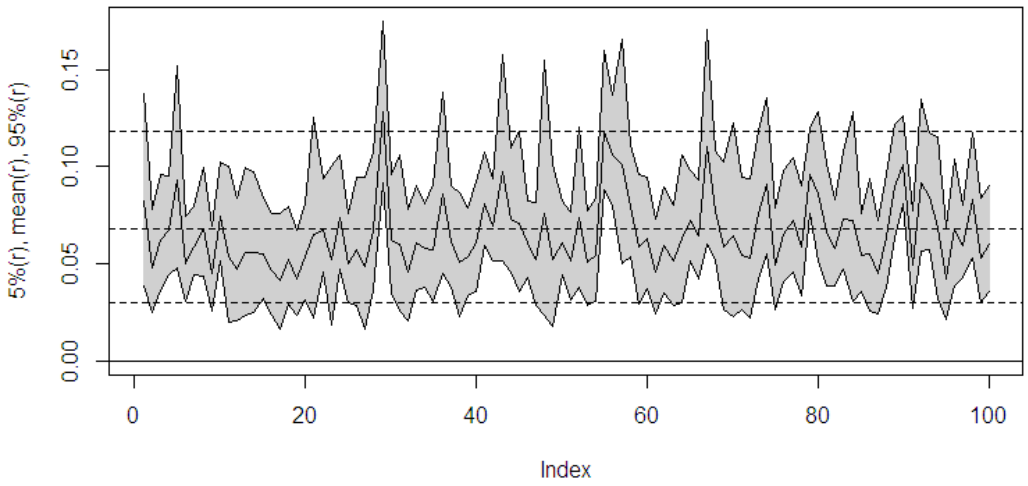

Figure 5.7: Solid lines: 5th percentile, mean and 95th percentile of r in the steady-state distributions implied by the statistical estimates from the 100 simulations when conditions for observing the process are less favourable. Dashed lines represent the 'real' values of these statistics.

5.5 Estimating Cost-at-Risk

We now move on to address Cost-at-Risk estimation using the Longstaff–Schwartz model. The question we are trying to answer is: how accurate estimates of long-run Cost-at-Risk can this model give us when calibrated using the Longstaff–Schwartz estimation method?

By now it is clear that the accuracy of the estimators for the model parameters depends on a number of factors—notably the amount of available data, but also the parameters of the actual data-generating process: for example, the lower the rate of mean reversion of the short rate, the longer a time series needs to be in order to be representative of the steady-state distribution of r. What parameter values should we use in the simulations? There are two natural possibilities available—those estimated by Longstaff and Schwartz (1993), and those estimated from our US government bond yield data set. Alas, the latter alternative is ruled out: not only does it lead to implausible interest rate dynamics, but it also turns out that the model cannot be re-estimated from the simulations it produces. Therefore we set the simulation model to use Longstaff and Schwartz's parameter values.

First we must find out the actual CaR as defined in Chapter 4. We have no closed-form solution for this, so we must use a Monte Carlo method: we sample the gamma distributions for x and y to calculate starting values for r and V, then run the simulation for nine years (once-a-year observations are enough) to calculate the cost of debt to be allocated to the tenth year. We repeat this 1,000,000 times in order to get a precise estimate of the actual 95% CaR (the 50,000th-largest cost out of 1,000,000 possible costs).

The distribution of the simulated costs, expressed in percentage points, is displayed in Figure 5.8. They have a mean of 9.51 and standard deviation of 1.86. The CaR is the 95th percentile, for which we obtain a point estimate of 12.946 and a 95% confidence interval of $[12.935, 12.957]$. This level of precision is quite sufficient for our purpose.

In the second phase we trace the distribution of the estimator ($\widehat{\text{CaR}}$) under some assumptions about how the model is calibrated to data. In order to make our results relevant for our actual conditions, we assume that the model is calibrated using 615 monthly observations. For simplicity (and speed) we assume that both r and V are observed directly. We generate 10,000 simulated data sets, from each of which we re-estimate the model parameters and calculate $\widehat{\text{CaR}}$ as the 100th-largest of 2,000 possible costs simulated using the re-estimated parameter values. Having done these 10,000 first-order simulations and 20 million second-order simulations, we are at a position to calculate a distribution for $\widehat{\text{CaR}}$ and to compare it to the actual value obtained in phase one.

Unfortunately, out of the 10,000 first-order simulations of length 615, as many as 2,589 produced data which led to a failure of the estimation method. The 7,411 successes led to CaR estimates summarized by Figure 5.9, with mean 11.24 and standard deviation 2.15 (the actual CaR being 12.95). The 5th and 95th percentiles were 8.26 and 15.16.

It appears that the CaR estimator is not very accurate, the root cause being that under the Longstaff–Schwartz parameter values, 50 years of data is not sufficient to accurately calibrate the model if the Longstaff–Schwartz calibration technique is used. We now turn to look for ways to improve the estimation procedure.

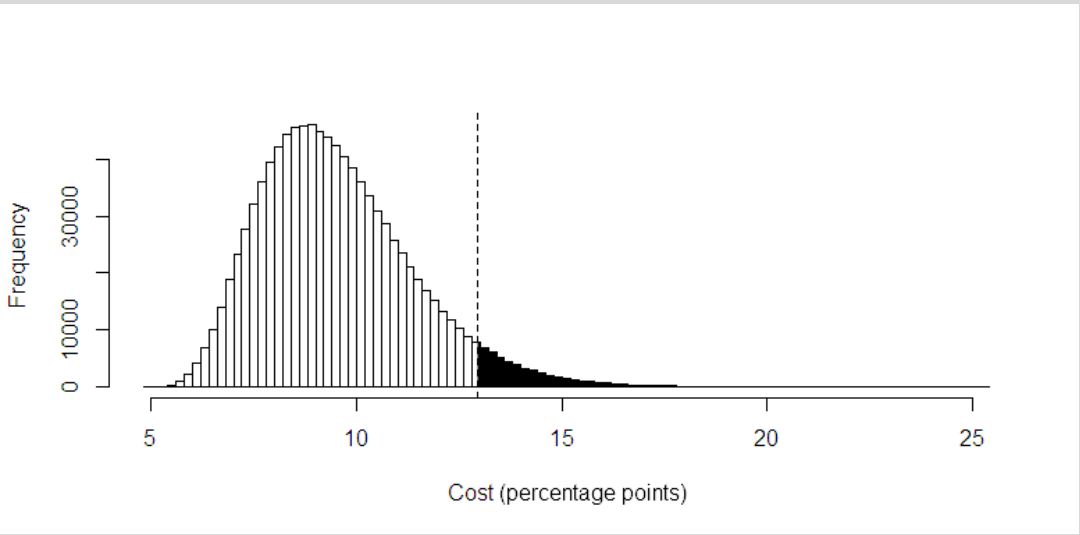

Figure 5.8: Distribution of annual debt costs under the parameter values obtained by Longstaff–Schwartz. The 95% Cost-at-Risk is 12.95, represented by the dashed line; the area to the right of the line contains 5% of the probability mass.

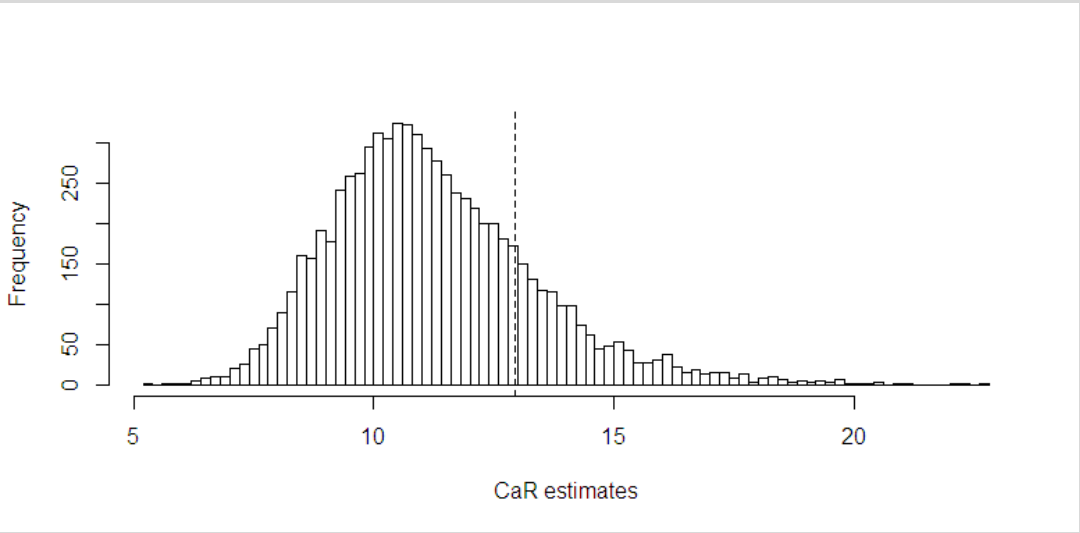

Figure 5.9: Distribution of CaR estimates, estimated from simulated yield data sets generated from the Longstaff–Schwartz parameter values. Dashed line marks the true value.

5.6 Improving the accuracy of parameter estimation

As shown by Figure 5.5, α is always accurately estimated and β is always underestimated by the Longstaff–Schwartz calibration method, at least with their parameter values. The systematic underestimation of β stems from the fact that the ratio $\frac{V}{r}$ practically never reaches β. Thus it seems questionable to squeeze the estimate $\hat{\beta}$ as low as possible by using the sample maximum of $\frac{V}{r}$ as its estimator. What

alternatives are there?

Looking at the graphs in Figure 5.3 suggests regressing ΔV on Δr and seeing what happens. This intuition can be elaborated as follows.

Note that $\Delta r = \alpha\Delta x + \beta\Delta y$ and $\Delta V = \alpha^2\Delta x + \beta^2\Delta y$. Since α is two orders of magnitude smaller than β, it follows that if the effects of Δx and Δy on Δr are of the same order of magnitude, then the effect of Δy on ΔV is about 100 times greater than the effect of Δx. Looking at Figure 5.3 suggests that this is indeed the case. r spikes up noticeably when V spikes, but even significant changes in the 'base' level of r are not reflected in V. Roughly speaking, r is affected by both x and y, but V only by y.[3]

It follows that $\Delta V \approx \beta(\beta\Delta y)$. Therefore, regressing ΔV on Δr should get us closer to the answer. And so it does. Figure 5.10 shows the results of re-estimating β from 100 simulated data sets. The dashed line marks the true value of β—i.e. the Longstaff–Schwartz value, which is what we are using in the simulations. The lowermost solid line shows the estimates of β obtained using the Longstaff–Schwartz approach where $\hat{\beta} = \max\left(\frac{V}{r}\right)$. The middle solid line represents the coefficient on r from an OLS regression of ΔV on Δr and a constant. Clearly, the regression coefficient is a better estimator of β.

But we can do better yet: since α is estimated accurately, it is possible to adjust for the bias in the regression estimator. Letting L denote the population regression coefficient, and remembering that x and y are independent,

$$L = \frac{\mathrm{Cov}(\Delta r, \Delta V)}{\mathrm{Var}(\Delta r)} = \frac{\alpha^3\mathrm{Var}(\Delta x) + \beta^3\mathrm{Var}(\Delta y)}{\alpha^2\mathrm{Var}(\Delta x) + \beta^2\mathrm{Var}(\Delta y)}$$

$$\Longrightarrow \beta^3 - L\beta^2 - \frac{\mathrm{Var}(\Delta x)}{\mathrm{Var}(\Delta y)}(L\alpha^2 - \alpha^3) = 0 \tag{5.24}$$

After we substitute our estimates of L and α into (5.24), three unknowns remain: β, $\mathrm{Var}(\Delta x)$ and $\mathrm{Var}(\Delta y)$. However, given a value of β, we can solve for x and y and calculate their variances. Thus one can use a univariate numerical optimization routine to solve (5.24) for β under the constraint that x and y must remain nonnegative. The resulting value is our new estimator for β.

[3]This line of argument depends entirely on α being dwarfed by β. Recall, however, that in our data set, the difference between the sample minimum and maximum values of $\frac{\mathrm{Var}(r)}{r}$ is almost *four* orders of magnitude (Table 5.1). Therefore α being much smaller than β seems realistic empirically.

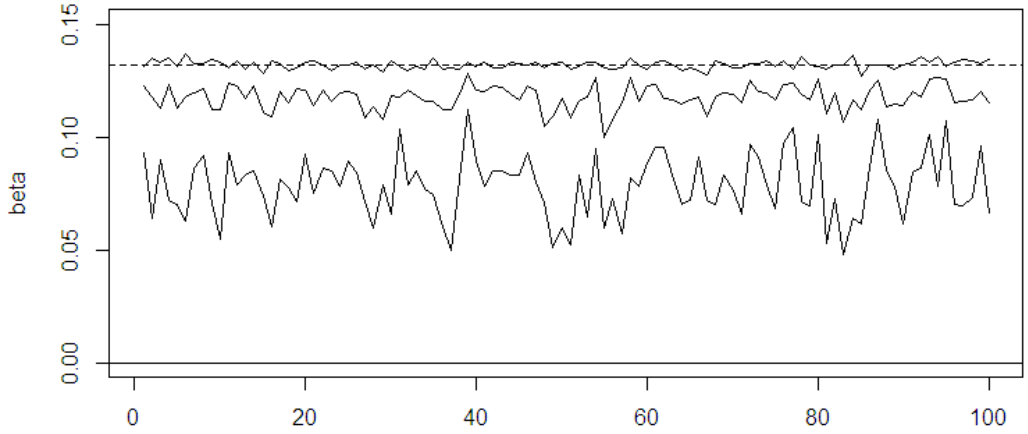

Figure 5.10: Re-estimating beta from 100 simulated data sets using the Longstaff–Schwartz method (lowermost solid line), regression of V on r (middle solid line), and bias-adjusted regression (uppermost solid line). The dashed line marks the true value.

The uppermost solid line in Figure 5.10 shows the result. This method seems satisfactory for the estimation of β.

Unfortunately this newfound accuracy in the estimation of β does not help much in pinning down γ, δ, η and ξ. As Figure 5.11 shows, these parameters remain subject to severe error, particularly γ and δ, which determine the long-run mean and rate of mean reversion of x, which in turn governs the 'base' level of r.

One approach to getting the remaining parameters right is to estimate them from (5.5) and (5.6), given that the x and y series are obtained as a by-product from estimating β. There are several ways to go about this.

- Discretize (5.5) as $\Delta x = \gamma \Delta t - \delta x_{t-1} \Delta t + \sqrt{x_{t-1}} \epsilon_t$, $\epsilon_t \sim N(0, \Delta t)$ i.i.d. Define $Z \equiv \frac{\Delta x}{\sqrt{x_{t-1}}} - \gamma \frac{\Delta t}{\sqrt{x_{t-1}}} + \delta \sqrt{x_{t-1}} \Delta t$, and choose γ and δ so as to minimize the deviation of Z's sample moments from the moments of $N(0, \Delta t)$. Similarly for y.

- As an alternative to the general method of moments, you can modify the previous method and try to match the sample distribution of Z to the $N(0, \Delta t)$ distribution some other way, such as matching some of their quantiles rather than their moments.

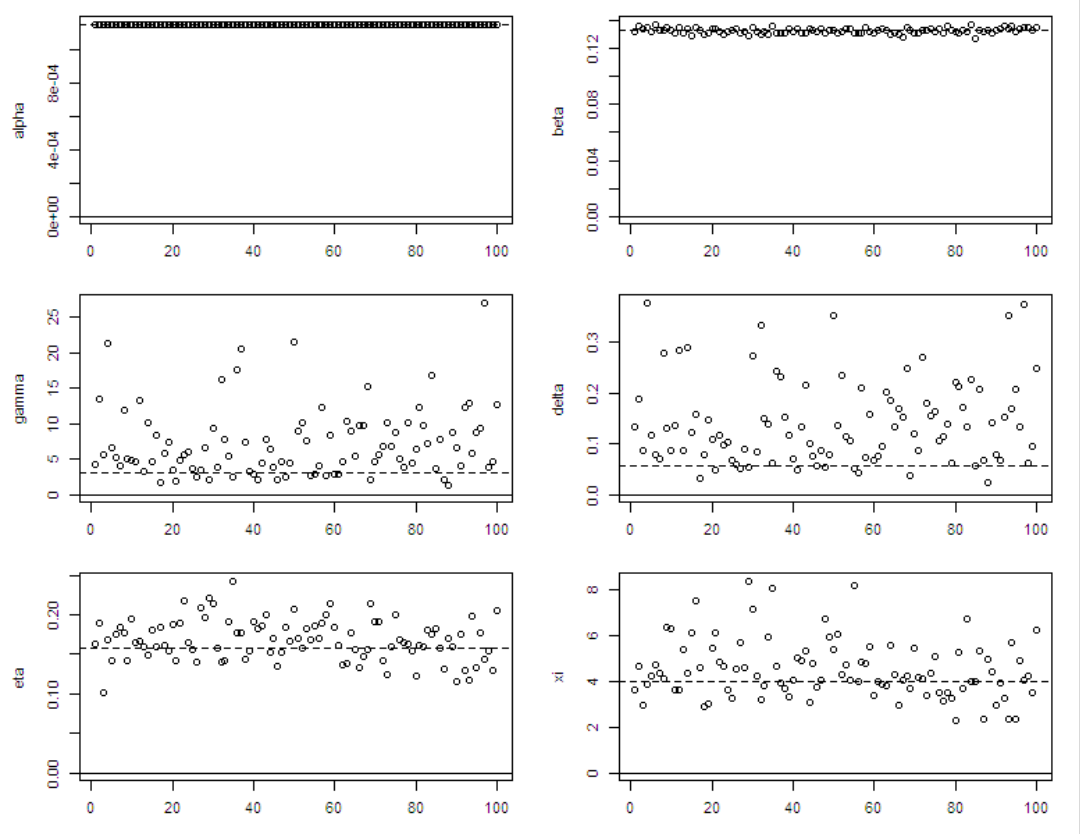

Figure 5.11: Re-estimating the stationary parameters from 100 simulated data sets using the bias-adjusted regression method for β. The dashed line marks the population value of each parameter, the points represent the estimate from each simulated data set.

- In the steady state, $x \sim \Gamma(2\gamma, \frac{1}{2\delta})$, so that $\mathbb{E}x = \frac{\gamma}{\delta}$ and $\mathrm{Var}(x) = \frac{\gamma}{2\delta^2}$. Thus γ and δ can be estimated as $\frac{\overline{x}^2}{2\mathrm{Var}(x)}$ and $\frac{\overline{x}}{2\mathrm{Var}(x)}$, respectively, using the sample mean and variance of x. Similarly for y. This amounts to assuming that the sample distributions of x and y are close to their unconditional distributions.

- Given x_t, γ and δ, we know that x_{t+1} has a non-central chi-squared distribution described in Cox, Ingersoll and Ross (1985b). Trying a certain pair of values for γ and δ, calculate the implied conditional distribution of $x_{t+1} \mid x_t$ for each x_t in the series except the last, then calculate the quantile of this distribution represented by the actual value of x_{t+1}. With the correct values of γ and δ, these quantiles should be distributed $U(0, 1)$. Matching some of the quantiles to those of the uniform distribution might be expected to yield an estimate for γ and δ. Similarly for y.

In practice, we found that none of these methods represents a significant improvement over the method used by Longstaff and Schwartz. The root cause is presumably the same in every case—the sample distributions of x, Δx, y and Δy are too different from their steady-state distributions for the parameters to be accurately recovered.

In order to improve the estimation, therefore, it seems necessary to take things a step further and exploit the information contained in the whole yield curve, not just the short rate. Equation (5.11) states that, given a set of values for the stationary parameters, the bond yield on every maturity is an affine function of r and V. Therefore, the remaining stationary parameters can be estimated by trying to match the theoretical relationship between r and V and the bond yields with the observed relationship between them, obtained by linear regression.

Therefore, we proceed as follows. With each simulation, we generate not only r and V, but also a series for the 3-month, 6-month, 1-year, 3-year, 5-year and 10-year yields using (5.11). Calibration starts from estimating α and β as described. We use the properties of the gamma distribution to obtain γ, δ, η and ξ from x and y (because this is slightly more accurate than the Longstaff–Schwartz method). Keeping these estimates fixed, we estimate ν by minimizing the sum of squared deviations of implied bond yields from the last one of the simulated yield curves. We now have a set of initial estimates.

In the second phase, we go through the simulated bond yields $R(\tau)$ for $\tau = \frac{1}{4}, \frac{1}{2}, 1, 3, 5, 10$. For each of these, we calculate a regressand $-\tau R(\tau)$, and regress it on r, V, and a constant, in order to obtain estimates of the slope coefficients $C(\tau)$ and $D(\tau)$ and intercept $\kappa\tau + 2\gamma \ln A(\tau) + 2\eta \ln B(\tau)$, respectively. Then we search for optimal values of γ, δ, η and ν that will minimize the squared deviation of the implied theoretical regression coefficients from those estimated from the sample.

Figure 5.12 shows the results. γ, δ and η (as well as ν, which is not displayed) are estimated much more accurately now. Only ξ is unaffected (because it has no effect on the yield curve except through ν).

Does more accurate estimation of the stationary parameters lead to more accurate estimation of Cost-at-Risk? Yes it does. When we repeat the earlier CaR estimation routine (with 10,000 primary simulations and re-estimations, each ac-

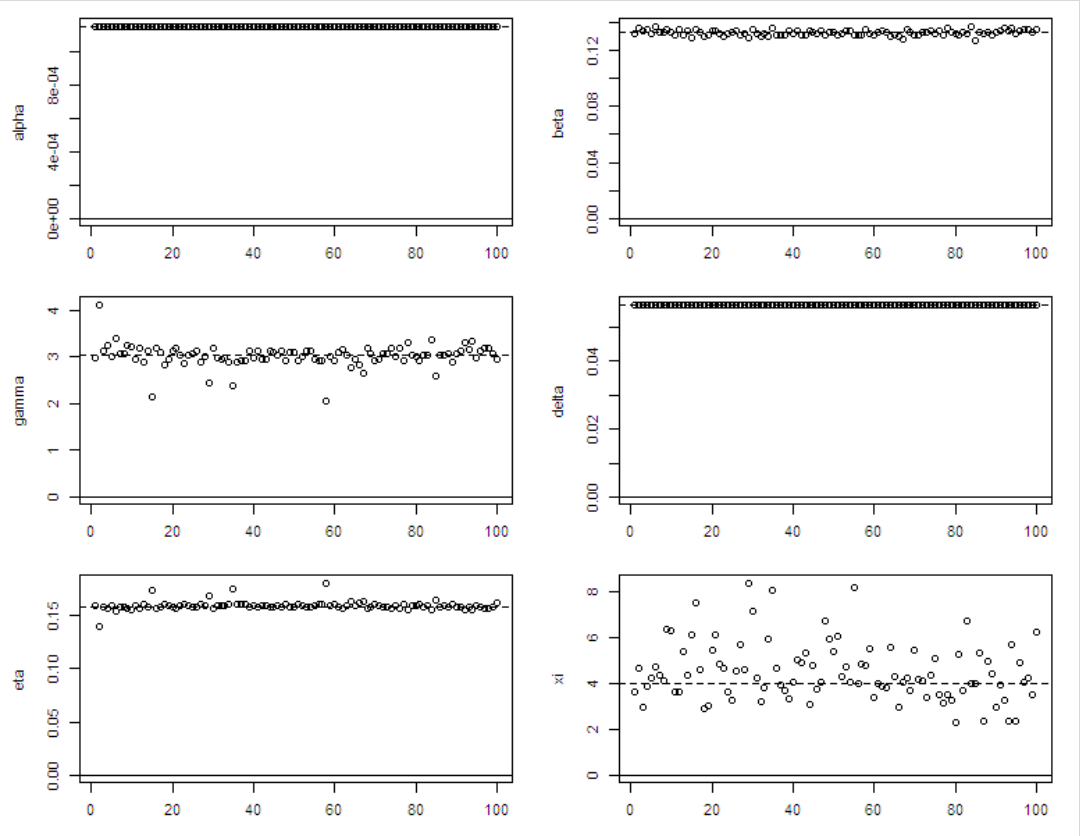

Figure 5.12: Re-estimating the stationary parameters from 100 simulated data sets using a whole-yield-curve regression method. The dashed line marks the population value of each parameter, the points represent the estimate from each simulated data set.

companied by 2,000 second-order resimulations to calculate a Monte Carlo estimate of CaR), the 10,000 estimates of CaR are distributed in the manner depicted in Figure 5.13. They have mean 12.92 and standard deviation 0.43—a remarkable improvement from the results obtained using the estimation procedure recommended by Longstaff and Schwartz themselves. The middle 90% are between 12.31 and 13.50.[4]

We also note that this new estimation technique did not fail even once, but produced coherent parameter estimates in every one of the 10,000 trials. In contrast, in the earlier experiment we found that the Longstaff–Schwartz estimation method failed in more than a quarter of all cases, producing parameter estimates which violated the model's assumptions and could not be used for estimating CaR.

However, the new estimation technique will be useful for the analysis of real-

[4]The actual CaR for this process is 12.95, as we know from Section 5.5.

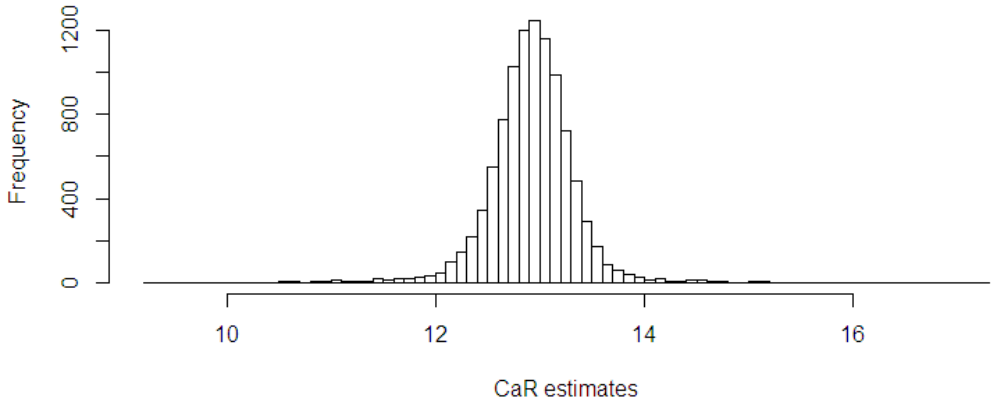

Figure 5.13: Distribution of CaR estimates, estimated from simulated yield data sets generated from the Longstaff–Schwartz parameter values, using the whole-yield-curve regression method.

world data only if the Longstaff–Schwartz model is a good model of real-world interest rates. A particularly important aspect is that long rates, the short rate and the short-rate volatility must be related in the manner predicted by the model. If not, we may expect the model to perform miserably.

The third column of Table 5.2 shows parameter estimates obtained using the new estimation technique on our US government bond yield data set. They are different from the estimates we obtained earlier using the Longstaff–Schwartz technique (displayed in the middle column), but are they any better? Apparently not, because the new values produce simulations that look even more unrealistic than those generated using the earlier values.

Table 5.2: Estimation results for the Longstaff–Schwartz model parameters using real market data. The middle column shows parameter values obtained using the Longstaff–Schwartz estimation method, the third those obtained using the new method based on whole-yield-curve regression.

	L&S (1993)	Sauli (2013) (1)	Sauli (2013) (2)
Mean(r)	0.06717	0.05186	0.05186
Var(r)	7.157×10^{-4}	9.543×10^{-4}	9.543×10^{-4}
Mean($\hat{V}$)	7.658×10^{-4}	2.521×10^{-4}	2.521×10^{-4}
Var($\hat{V}$)	1.526×10^{-6}	4.811×10^{-7}	4.811×10^{-7}
Min($\hat{V}/r$)	0.001149	3.525×10^{-5}	3.525×10^{-5}
Max($\hat{V}/r$)	0.1325	0.2116	0.2116
$\hat{\alpha}$	0.001149	3.525×10^{-5}	3.525×10^{-5}
$\hat{\beta}$	0.1325	0.2116	0.0651
$\hat{\gamma}$	3.0493	1.3608	88.591
$\hat{\delta}$	0.05658	9.466×10^{-4}	0.0640
$\hat{\eta}$	0.1582	0.0651	0.0523
$\hat{\xi}$	3.998	11.648	1.1023
$\hat{\lambda}$	-3.663	-10.692	-0.0677

6 Benchmark model: Nelson–Siegel–ARMA

In this chapter we present a simpler term structure model to serve as a benchmark against which we can compare the performance of the Longstaff–Schwartz model, allowing us to assess its suitability for CaR estimation.

The Nelson–Siegel model (Nelson and Siegel, 1987) is a static yield curve model which can easily be extended into a dynamic model. In the static part, the yield curve is modelled as a parametric curve function to be fitted to a set of yields at a single point in time. The function is

$$R(t,\tau) \;=\; \beta_1(t) + \beta_2(t)\left(\frac{1-e^{-\lambda(t)\tau}}{\lambda(t)\tau}\right) + \beta_3(t)\left(\frac{1-e^{-\lambda(t)\tau}}{\lambda(t)\tau} - e^{-\lambda(t)\tau}\right) \qquad (6.1)$$

where the betas act as multipliers on three additive components (graphed in Figure 6.1) that control the level, slope and curvature of the yield curve. λ_t is a horizontal squeeze.

The parameter values for a single time-point t can be estimated by minimizing the sum of squared residuals between observed and theoretical yields—an exercise in non-linear optimization. Restricting $\lambda(t)$ to a constant would turn this into a linear optimization problem, so that the betas could be estimated by OLS. Nelson and Siegel find that very little explanatory power is lost by doing so. They measure τ in days and fix λ at 0.02, which translates to $\lambda = 7.3$ when τ is measured in years.

The model can be extended into a dynamic term structure model by specifying a process for each of the parameters $\beta_1(t)$, $\beta_2(t)$, $\beta_3(t)$ and $\lambda(t)$. Diebold and Li (2006) do this as follows. They restrict λ to a constant and use OLS to estimate the β's for each observation in their data set, producing a time series for each β; an AR(1) model is then estimated for each series. Finally, they test the out-of-sample forecasting performance of the resulting dynamic term structure model using a 16-year data set of US government bond yields, with good results.

We now take the same approach. We use OLS to fit a Nelson–Siegel curve to each observation in our data, with λ fixed at an arbitrary value; repeating this with

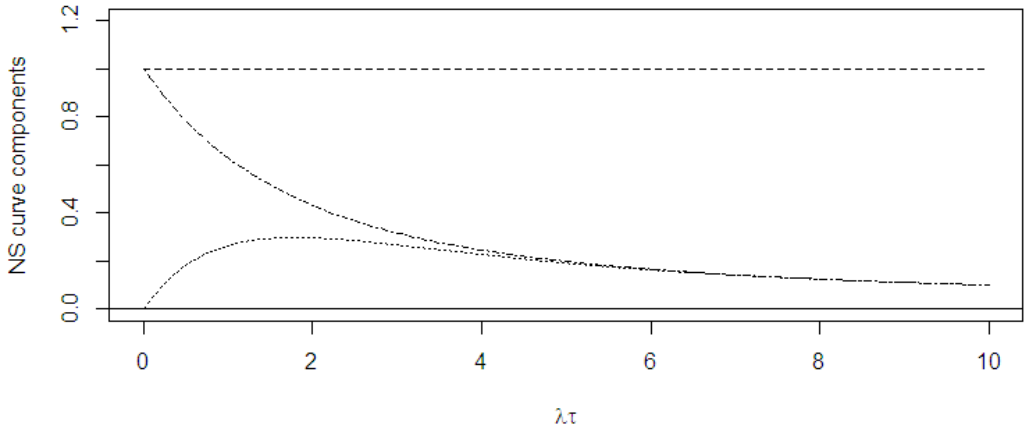

Figure 6.1: Nelson–Siegel yield curve components. From the top down, the curves correspond to the components multiplied by β_1, β_2 and β_3.

different values of λ, we find that the best overall fit in the least squares sense is achieved when $\lambda = 0.859$. The overall SD of residuals is then 12.5 bp and R^2 is 0.998.

For comparison, fitting a straight line through the yield curve at each observation gives an R^2 of 0.992, and the overall SD of residuals is 26.5 bp.

Our estimate $\lambda = 0.859$, with τ measured in years, is not far from the value used by Diebold and Li, which translates to 0.73 when τ is measured in years (they measure τ in months and fix λ at 0.0609). Both estimates are very different from Nelson and Siegel's λ of 7.3, but this is because Nelson and Siegel use data for maturities up to one year only, effectively fitting the model to but a small segment of the yield curve.

Figure 6.2 displays the estimated parameter values for each point in time. The path of $\hat{\beta}_1$ has a familiar shape; it tracks the historical ten-year yield quite closely. It is very persistent compared to $\hat{\beta}_2$ and $\hat{\beta}_3$, which fluctuate rapidly back and forth across their mean level. $\hat{\beta}_3$ has a positive correlation of 0.40 with $\hat{\beta}_1$ and of 0.63 with $\hat{\beta}_2$.[1]

Now that we have the time series for each parameter, we could follow the ex-

[1]These results are not overly sensitive to λ: repeating the regression with a of λ of 0.76 or 0.96 yields almost identical β's.

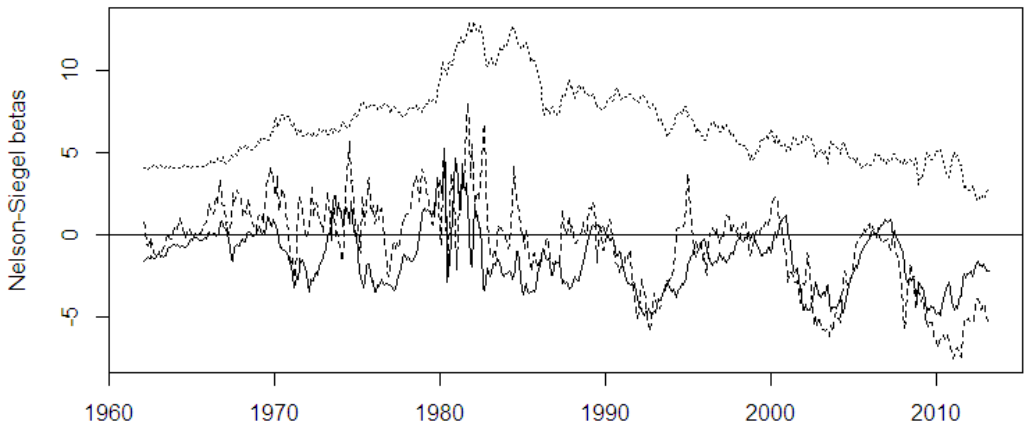

Figure 6.2: Estimated Nelson–Siegel model parameters. $\hat{\beta}_1$: dotted line, $\hat{\beta}_2$: solid line, $\hat{\beta}_3$: dashed line.

Table 6.1: Statistics for estimated Nelson–Siegel parameters, with $\lambda = 0.859$.

	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
mean	6.6	-1.4	-0.5
SD	2.3	1.8	2.7
min	2.0	-4.9	-7.5
max	13.0	5.2	8.1

ample of Diebold and Li (2006) and estimate an AR(1) model for each. However, trying various ARMA models reveals that an ARMA(2,1) model is needed in order to eliminate the autocorrelation in residuals for β_1 and β_3. AIC recommends ARMA(2,1) for β_1 and β_2, and a higher-order model for β_3.

For simplicity, we use ARMA(2,1) for each of the betas. The specification is

$$\beta_i(t) = a_{0,i} + a_{1,i}\beta_i(t-1) + a_{2,i}\beta_i(t-2) + \epsilon_i(t) + b_1\epsilon_i(t-1) \tag{6.2}$$

where $\epsilon_i(t) \sim N(0, \sigma_i^2)$ i.i.d. and $a_{0,i}$, $a_{1,i}$, $a_{2,i}$, $b_{1,i}$ and σ_i^2 are constants to be estimated. Table 6.2 summarizes the estimation results. All three processes are estimated to be stationary, albeit the process for β_1 comes rather close to a unit root process.

Now we have a hybrid model composed of the static Nelson–Siegel model and a dynamic ARMA model for each of its parameters. Since the Nelson–Siegel yields

Table 6.2: Estimated ARMA(2,1) coefficients for Nelson–Siegel betas. Significance at the 5%, 1% and 0.1% levels is indicated by *, ** and ***, respectively. The unconditional mean and variance of each process is calculated directly from the coefficients.

	β_1	β_2	β_3
$\hat{a_0}$	0.051	-0.106**	-0.063
$\hat{a_1}$	0.630***	0.688***	0.520***
$\hat{a_2}$	0.362***	0.239**	0.390***
$\hat{b_1}$	0.637***	0.663***	0.700***
$\hat{\sigma}^2$	0.054	0.217	0.637
Mean($\beta(t)$)	6.239	-1.446	-0.707
Var($\beta(t)$)	6.521	3.318	7.398

are linear in the betas, each of which has a normal unconditional distribution, the yield on each maturity is normally distributed. With the help of equation (6.1) we can calculate the implied mean/median yield curve. Figure 6.3 displays this and the 2.5th and 97.5th percentiles in solid lines. Dotted lines display the corresponding percentiles for the empirically observed rates for maturities of 1/4, 1/2, 1, 3, 5 and 10 years.

Being normally distributed, the yield on any maturity can go negative. The unconditional probability of a negative short rate is as high as 6.3%, and that of a negative 10-year rate is 1.0%.

Like spot rates, forward rates are also linear in the parameters β_1, β_2 and β_3, and thus normally distributed in the steady state. Starting out from equation (6.1), setting $\lambda(t)$ to a constant λ, and using the formula (1.4) for forward rates, we get

$$f(t,T_1,T_2) \;=\; \beta_1(t) + \beta_2(t)\frac{e^{-\lambda\tau_1} - e^{-\lambda\tau_2}}{\lambda(\tau_2-\tau_1)} + \beta_3(t)\frac{(\lambda\tau_1+1)e^{-\lambda\tau_1} - (\lambda\tau_2+1)e^{-\lambda\tau_2}}{\lambda(\tau_2-\tau_1)} \tag{6.3}$$

The intervals (T_1, T_2) we are interested in have length of one year, so $\tau_2 - \tau_1 = 1$. For the sake of notational simplicity, let

$$\phi_2(\tau) \;\equiv\; \frac{e^{-\lambda\tau} - e^{-\lambda(\tau+1)}}{\lambda} \tag{6.4}$$

$$\phi_3(\tau) \;\equiv\; \frac{(\lambda\tau+1)e^{-\lambda\tau} - (\lambda(\tau+1)+1)e^{-\lambda(\tau+1)}}{\lambda} \tag{6.5}$$

In Chapter 4 we defined Cost-at-Risk as the 95th percentile of the unconditional

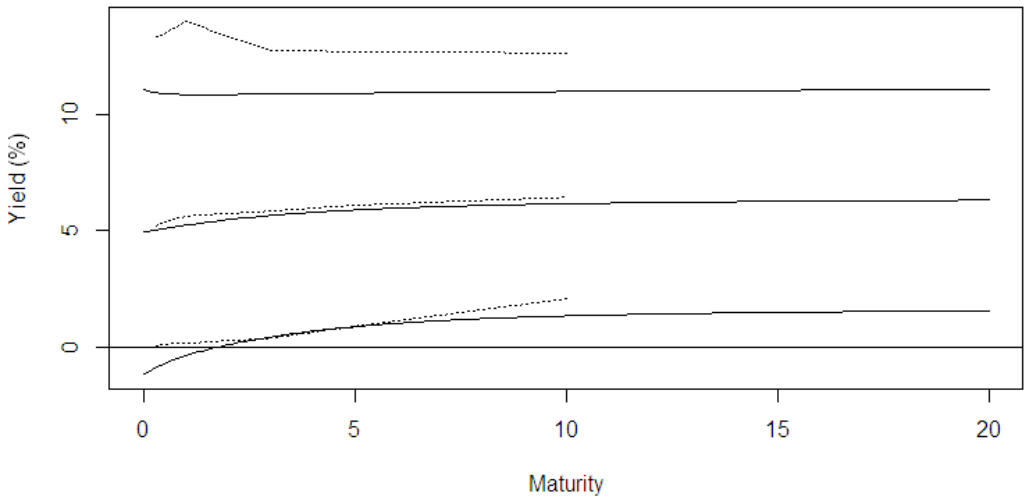

Figure 6.3: Solid lines: 2.5th, 50th and 97.5th percentiles of the unconditional distribution of the yield on each maturity, according to the estimated model. Dotted lines: the corresponding percentiles of yields in the actual data.

distribution of the annual cost of debt due to a particular funding strategy. Following the same specification, the rate of debt costs to be allocated to the period $t \in (0,1)$ is $c(0,1) = \ln \Lambda(1) - \ln 16$, where

$$\begin{aligned}
\Lambda(1) &= \sum_{t=-9}^{0} \exp\left(\beta_1(t) + \beta_2(t)\phi_2(-t) + \beta_3(t)\phi_3(-t)\right) \\
&\quad + \sum_{t=-4}^{0} \exp\left(\beta_1(t) + \beta_2(t)\phi_2(-t) + \beta_3(t)\phi_3(-t)\right) \\
&\quad + \exp\left(\beta_1(0) + \beta_2(0)\phi_2(0) + \beta_3(0)\phi_3(0)\right) \\
&= \sum_{t=-9}^{-5} \exp\left(\beta_1(t) + \beta_2(t)\phi_2(-t) + \beta_3(t)\phi_3(-t)\right) \\
&\quad + 2\sum_{t=-4}^{-1} \exp\left(\beta_1(t) + \beta_2(t)\phi_2(-t) + \beta_3(t)\phi_3(-t)\right) \\
&\quad + 3\exp\left(\beta_1(0) + \beta_2(0)\phi_2(0) + \beta_3(0)\phi_3(0)\right)
\end{aligned} \tag{6.6}$$

i.e. the cost rate, whose 95th percentile we seek, is the logarithm of the weighted average of ten lognormal deviates. While there is no closed-form method to calculate

such a percentile exactly, in practice $c(0,1)$ is fairly well approximated in distribution by the following measure:[2]

$$\begin{aligned} c^N(0,1) &= \frac{1}{16}\Bigg(\sum_{t=-9}^{-5} \big(\beta_1(t) + \beta_2(t)\phi_2(-t) + \beta_3(t)\phi_3(-t)\big) \\ &\quad +2\sum_{t=-4}^{-1} \big(\beta_1(t) + \beta_2(t)\phi_2(-t) + \beta_3(t)\phi_3(-t)\big) \\ &\quad +3\big(\beta_1(0) + \beta_2(0)\phi_2(0) + \beta_3(0)\phi_3(0)\big)\Bigg) \\ &= \frac{1}{16}\big(\boldsymbol{v}_1'\boldsymbol{\beta}_1 + \boldsymbol{v}_2'\boldsymbol{\beta}_2 + \boldsymbol{v}_3'\boldsymbol{\beta}_3\big) \end{aligned} \tag{6.7}$$

where the vectors are

$\boldsymbol{v}_1$	$\boldsymbol{\beta}_1$	$\boldsymbol{v}_2$	$\boldsymbol{\beta}_2$	$\boldsymbol{v}_3$	$\boldsymbol{\beta}_3$
1	$\beta_1(-9)$	$\phi_2(9)$	$\beta_2(-9)$	$\phi_3(9)$	$\beta_3(-9)$
1	$\beta_1(-8)$	$\phi_2(8)$	$\beta_2(-8)$	$\phi_3(8)$	$\beta_3(-8)$
1	$\beta_1(-7)$	$\phi_2(7)$	$\beta_2(-7)$	$\phi_3(7)$	$\beta_3(-7)$
1	$\beta_1(-6)$	$\phi_2(6)$	$\beta_2(-6)$	$\phi_3(6)$	$\beta_3(-6)$
1	$\beta_1(-5)$	$\phi_2(5)$	$\beta_2(-5)$	$\phi_3(5)$	$\beta_3(-5)$
2	$\beta_1(-4)$	$2\phi_2(4)$	$\beta_2(-4)$	$2\phi_3(4)$	$\beta_3(-4)$
2	$\beta_1(-3)$	$2\phi_2(3)$	$\beta_2(-3)$	$2\phi_3(3)$	$\beta_3(-3)$
2	$\beta_1(-2)$	$2\phi_2(2)$	$\beta_2(-2)$	$2\phi_3(2)$	$\beta_3(-2)$
2	$\beta_1(-1)$	$2\phi_2(1)$	$\beta_2(-1)$	$2\phi_3(1)$	$\beta_3(-1)$
3	$\beta_1(0)$	$3\phi_2(0)$	$\beta_2(0)$	$3\phi_3(0)$	$\beta_3(0)$

The $\boldsymbol{v}$'s are constant, the $\boldsymbol{\beta}$'s are multivariate normal. It follows that $c^N(0,1)$ is univariate normal, its distribution exhaustively described by its mean and variance. The mean is simply

$$\mathbb{E}(c^N(0,1)) \;=\; \mathbb{E}\beta_1\boldsymbol{v}_1'\mathbf{1} + \mathbb{E}\beta_2\boldsymbol{v}_2'\mathbf{1} + \mathbb{E}\beta_3\boldsymbol{v}_3'\mathbf{1} \tag{6.8}$$

and, since the three $\boldsymbol{\beta}$'s are driven by independent ARMA processes, the variance is

$$\mathrm{Var}(c^N(0,1)) \;=\; \mathrm{Var}(\boldsymbol{v}_1'\boldsymbol{\beta}_1) + \mathrm{Var}(\boldsymbol{v}_2'\boldsymbol{\beta}_2) + \mathrm{Var}(\boldsymbol{v}_3'\boldsymbol{\beta}_3) \tag{6.9}$$

[2]This might be described as using the weighted geometric average of the lognormals as an approximation of their weighted arithmetic average.

The kth term in (6.9) can be calculated as

$$\begin{aligned} \mathrm{Var}(\boldsymbol{v}_k'\boldsymbol{\beta}_k) &= \sum_{i=1}^{10}\sum_{j=1}^{10}\mathrm{Cov}(v_{k,i}\beta_{k,i}, v_{k,j}\beta_{k,j}) \\ &= \sum_{i=1}^{10}\sum_{j=1}^{10} v_{k,i}v_{k,j}\gamma_{k,i-j} \end{aligned} \tag{6.10}$$

where $v_{k,i}$ and $\beta_{k,i}$ denote the ith elements of the (10×1)-vectors $\boldsymbol{v}_k$ and $\boldsymbol{\beta}_k$, respectively, and $\gamma_{k,s}$ stands for the autocovariance between $\beta_k(t)$ and $\beta_k(t-s)$. We are modeling β_k as an ARMA(2,1) process, with time steps of one month; therefore, given the parameters of the ARMA(2,1) model, we can calculate the autocovariances of β_k at lags of $0, 12, ..., 108$ months.[3] These correspond to $s = 0, ..., 9$, since in the Nelson–Siegel part of our model, time is measured in years.

Once these calculations are done, using the estimated parameter values for our three ARMA models, we are ready to report the distribution of the costs $c^N(0,1)$ to be allocated to period $t \in (0,1)$, expressed as a percentage of the outstanding debt:

$$c^N(0,1) \sim N\left(5.859, 5.430\right) \tag{6.11}$$

The 95th percentile of this distribution is 9.693. However, given that $c^N(0,1)$ approximates $c(0,1)$ by using an exponential of an average in the place of an average of exponentials, Jensen's inequality implies that it has some downward bias. Fortunately the magnitude of this bias can be estimated numerically and removed: the bias is about 0.8 basis points, so the CaR is estimated to be

$$\mathrm{CaR} \;=\; \left\{x : \Pr\left\{c(0,1) > x\right\} = 0.05\right\} \;=\; 9.701 \tag{6.12}$$

This model has two main flaws. The first one has to do with negative interest rates. We already saw that the short rate would be negative about six percent of the time, and the 2.5th percentile for it (shown in Figure 6.3) was as low as -1.35, a level which, to be possible, would seem to require monetary and financial institutions quite different from currently existing ones.

Yet for many applications the possibility of negative rates may not be a problem,

[3] This can be done easily using the Yule–Walker equations; see Enders, 2004, pp. 60–64.

and the covariance structure of yields is everything. This is where our model's other major handicap comes: it pays no attention to the cross-correlation between the three Nelson–Siegel parameters, even though our estimated betas exhibit significant cross-correlation, particularly between β_2 and β_3.

A positive cross-correlation between the terms serves both to increase the variance of yields and to reduce the effectiveness of diversifying one's portfolio across maturities. Lack of attention to cross-correlation is thus potentially a very significant lapse, and before putting this kind of model into actual use one should consider using a VAR model for the betas. Alternatively, one could distribute the mass of the Nelson–Siegel 'factor loadings' on β_1, β_2 and β_3 differently to get rid of the cross-correlations.

We leave such refinements for another occasion, however, and settle for the model already estimated—matching our benchmark model to actual data is, after all, of secondary importance for our present purpose.

Moving on to address our main problem—measuring the relative efficiency of different models in estimating CaR under various yield curve dynamics—we now make a similar experiment as with the Longstaff–Schwartz model: we generate sets of interest rate pseudo-data, to each of which we recalibrate our model, in order to get an idea of the distribution of the $\widehat{\text{CaR}}$ estimator.

In the case of our Nelson–Siegel–ARMA(2,1) hybrid model, this means taking the following steps. We use an ARMA(2,1) simulation program to generate paths for the parameters β_1, β_2 and β_3, and convert these into yield curves using equation (6.1). We then calibrate our model to the simulated data just as we would calibrate it to real data.[4] This gives us a new set of parameter values, from which we calculate the CaR estimate, or $\widehat{\text{CaR}}$. Repeating this sufficiently many times, we obtain the distribution of $\widehat{\text{CaR}}$ in a world where the true yield curve dynamics would be correctly described by our model.

Again, we set our simulations to have a length of 600 observations, which corresponds to 50 years' worth of monthly yield data. 10,000 simulations will be more than enough.

[4]Admittedly it is somewhat superfluous to first convert the simulated β's into yields and then right back into $\hat{\beta}$'s. After all, the Nelson–Siegel parameters are estimated almost without error in this setting, so we might as well calibrate the ARMA models to the simulated β directly, and calculate the $\widehat{\text{CaR}}$ without working with interest rates at all.

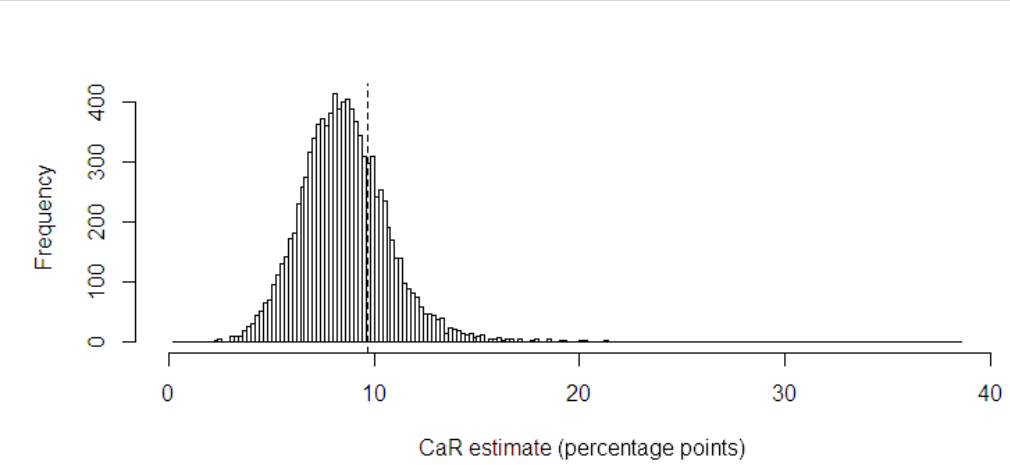

Figure 6.4: Distribution of the $\widehat{\text{CaR}}$ estimator when yield curve dynamics are as described by the model developed and calibrated in this chapter, and estimation is based on 50 years of monthly observations of the yield curve. The dashed line represents the population CaR.

The experiment shows that under the specified conditions, CaR estimates are not particularly reliable, i.e. they have a rather wide range: the first percentile is 0.29, the 99th percentile is 37.05. The very lowest estimate is -44; the highest in excess of 300; and in 22 trials, the estimation fails altogether as the process for $\beta_1(t)$ is estimated to be nonstationary. The mean is 8.69 and standard deviation is 2.42. Figure 6.4 shows the middle 98% segment of the sample distribution, with the correct value, 9.70, marked by a dashed line.

Of course, the dispersion of the estimates would be much reduced if longer data sets were available for calibrating the model. When the simulation length is increased to 2400 observations (200 years), the distribution of $\widehat{\text{CaR}}$ has mean 9.36, SD 1.17, and the 1st and 99th percentiles are 6.65 and 12.17.

That is, if we had 200 years of data, and could be confident that the process that generated it was well described by our Nelson–Siegel–ARMA model, we could estimate our Cost-at-Risk with an error margin of just a couple of percentage points. Unfortunately, our actual position is not quite so good.

7 Comparison of the LS and NS models

In earlier chapters, we studied the suitability of the Longstaff–Schwartz model (LS) and the Nelson–Siegel–ARMA(2,1) model (NS) for estimating Cost-at-Risk when the model specification is exactly correct, i.e. when the interest rate data-generating process is precisely as described by the model except that the parameter values are unknown.

In this chapter we cross-connect the models. We generate pseudo-data with a simulation program that is exactly as described by the LS model, then calibrate the NS model to this data to estimate Cost-at-Risk. After examining the performance of the NS model, we reverse the setting and repeat the procedure, calibrating the LS model to NS-type data.

It is natural to anticipate that each model will do better on its own turf, i.e. that the NS model will be less accurate than LS when the data-generating process is of the LS type, and vice versa. What we want to know is whether there is a conspicuous difference in the two models' overall performance and robustness.

7.1 When the DGP is of the LS type

Using the same simulation program as in chapter 5, we produce 2,000 sets of interest rate pseudo-data, each containing 615 observations of the yields on the maturities of $\frac{1}{4}$, $\frac{1}{2}$, 1, 3, 5 and 10 years. For the simulations, we use the model parameter values estimated by Longstaff and Schwartz (1993), as before. Therefore we know that the actual CaR for this process is 12.95.[1]

With each simulated data set, we do three model calibrations: one for the NS model; another for the LS model using the Longstaff–Schwartz calibration technique; and a third for the LS model using the regression-based calibration technique. The parameter estimates thus obtained are then used for computing the CaR estimate implied by each model. The results are summarized in Table 7.1.

[1] Cf. Section 5.5.

Table 7.1: Distributions of Cost-at-Risk estimates given by different models and calibration techniques when the DGP is of the Longstaff–Schwartz type. The actual CaR implied by the DGP is 12.95 percentage points. 'LS-R' stands for the Longstaff–Schwartz model with regression-based calibration.

Model	NS	LS	LS-R
Failure rate	3.9%	25.7%	0%
Mean($\widehat{\text{CaR}}$)	11.02	11.29	12.92
SD($\widehat{\text{CaR}}$)	3.67	2.20	0.43
Min($\widehat{\text{CaR}}$)	−12.58	6.11	9.92
Max($\widehat{\text{CaR}}$)	112.77	19.56	15.99
RMSE($\widehat{\text{CaR}}$)	4.15	2.75	0.43

As before, the Longstaff–Schwartz calibration technique fails to lead to any CaR estimate in about a quarter of all cases. The NS model does not fail as often, but the CaR estimates it produces are generally worse than those produced by the LS model. But this is partly because the LS model's precision statistics, being based on the 1,486 successful estimation results only, are embellished by the removal of the worst failures. If we take the best 1,486 CaR estimates from the NS model, we find that they have a RMSE of only 1.47. In that sense, the NS model seems to be performing better than the LS model, even though the latter has the correct specification of the DGP.

Consistently with our findings from Chapter 5, the LS model gives highly accurate CaR estimates when calibrated using the regression-based method; the robustness of this method, however, is still open to question.

7.2 When the DGP is of the NS type

Before we can calibrate the LS model to pseudo-data generated by a Nelson–Siegel–ARMA-type simulation program, we face two obstacles. Firstly, in order to calibrate the Longstaff–Schwartz model, one needs a time series of the levels of the instantaneous variance rate V of the short rate; the NS model does not, by default, produce this kind of output.

In fact the conditional variance of the short rate in the NS model is constant, if we take it to mean $\text{Var}(r_t - \mathbb{E}_{t-1}(r_t))$: it is just the sum of the variances of the disturbance terms in the ARMA models for β_1 and β_2. But the Longstaff–Schwartz

model cannot accept a constant V series as input for calibration, so we have to estimate a time-varying V series from the data somehow.

Since we know for a fact that there is no useful information to be gleaned from the data in this respect, we may as well settle for a quick and dirty solution—we estimate V as a moving variance of first differences,

$$V_i \;=\; \mathrm{Var}\big\{ r_{i+j} - r_{i+j-1} \;\big|\; j = -n+1, ..., n \big\} \tag{7.1}$$

and choose $n = 6$; for the first six and last six observations, the range of j is adjusted according to the availability of data.

The second obstacle is that the NS model will occasionally generate negative yields, which cannot be digested by the Longstaff–Schwartz model. Bad as that may sound, the workaround is quite straightforward: tamper with the data. After all, if you cannot fit your model to the data, what other alternatives are there—apart from discarding the model—than fitting it to some transform of the data?

The simplest solution would be to apply a floor: if the data contains negative yields $R(t,\tau)$, use $R^*(t,\tau) = \max\left(R(t,\tau), m(\tau)\right)$ instead, where $m(\tau)$ is the floor for maturity τ. But this has the shortcoming that the resulting short rate r^* might sometimes flatline for long enough that the conditional variance would go to zero. Therefore we opt for a slightly more sophisticated transform,

$$R^*(t,\tau) \quad = \quad \begin{cases} \exp\left(R(t,\tau) - 1\right), \; R(t,\tau) < 1 \\ R(t,\tau), \; R(t,\tau) \geq 1 \end{cases} \tag{7.2}$$

This transform is smooth and strictly positive, and the conditional variance of the transformed short rate never goes to zero.

What NS model parameter values are used in the simulations may also have an impact on the results of the coming experiment. We choose these values as follows: we use the Longstaff–Schwartz simulation model, with the same parameter values as before, to generate a set of pseudo-data of 12,000 observations—the equivalent of 1,000 years. By calibrating the NS model to this data, we obtain a set of NS model parameter values which ought to produce yield curves reasonably similar to those produced by the LS model in the previous experiment. So we plug these parameter

Table 7.2: Distributions of Cost-at-Risk estimates given by different models and calibration techniques when the DGP is of the Nelson–Siegel type. The actual CaR implied by the DGP is 11.72. 'LS-R' stands for the Longstaff–Schwartz model with regression-based calibration.

Model	NS	LS	LS-R
Failure rate	6.6%	2.4%	83.12%
Mean($\widehat{\text{CaR}}$)	11.17	12.51	0.88
SD($\widehat{\text{CaR}}$)	3.33	2.08	139.93
Min($\widehat{\text{CaR}}$)	−73.35	7.03	−1670.54
Max($\widehat{\text{CaR}}$)	73.25	38.17	956.16
RMSE($\widehat{\text{CaR}}$)	3.38	2.22	140.27

values into the simulation program. The implied CaR is 11.72.

Now we can start the experiment. We generate 5,000 simulated sets of yields, each containing 615 observations of the yields on the maturities of $\frac{1}{4}$, $\frac{1}{2}$, 1, 3, 5 and 10 years. With each simulated data set, we do three model calibrations. First, we calibrate the NS model to the data in its original form. Then we apply the transform (7.2) to the data and compute V according to (7.1), so that we can calibrate the LS model to the transformed data—first using the Longstaff–Schwartz method, then using the linear-regression-based method developed in Section 5.6.

The results are summarized in Table 7.2. The LS-R model's performance is seen to be abysmal compared to that of both NS and LS, with a failure rate of 83.1%—meaning that in 4,156 cases out of 5,000, the estimation procedure failed to lead to any CaR estimate. The failure rate for the other models is much lower. The statistics in Table 7.2 were computed after excluding the cases where a model failed to deliver any CaR estimate; in the case of the NS model, one outlier was also removed.

Unexpectedly, the LS model performs somewhat better than the NS model even though the latter has the correct model specification. This effect seems to be due to the NS model's higher tendency to produce outlier CaR estimates. That is, the NS model is more accurate most of the time, but it fails more often than the LS model, and sometimes spectacularly; as a result, it has a higher RMSE than the LS model.

7.3 Conclusions

The results from the two previous sections can be summarized thus: when the interest rate data-generating process is of the LS type, the LS model performs poorly, the NS model performs only slightly better at best, and the LS model with regression-based calibration is highly accurate. When the DGP is of the NS type, the LS model outperforms the NS model if calibrated using the Longstaff–Schwartz method; the regression-based method, on the other hand, renders it completely useless.

It is surprising that the NS model has a lower failure rate when it misrepresents the DGP than when it describes it correctly. But this could be an artifact of the parameters of the NS simulation model, which were obtained from fitting the NS model to pseudo-data from the LS model.

What is really perplexing, though, is how much better the LS model (using Longstaff–Schwartz calibration) performs under a NS-type DGP than under a LS-type DGP. Its failure rate is 25.7% on its home turf, but only 2.4% when it is misspecified. On top of that, its accuracy is somewhat better in a Nelson–Siegel world. Again, this phenomenon might be specific to the simulation parameters used.

In a nutshell, what we have discovered is the inadequacy of the plan of inquiry we set out to pursue: we were hoping, more or less, to rank the models for performance and/or robustness; but now we see that before such conclusions can be drawn, we must examine the possibility that the results depend entirely on the simulation parameters used—a task we cannot fit into the limits of this study.

8 Conclusions

The objective of this study has been to assess the suitability of the Longstaff–Schwartz interest rate model for estimating the long-run Cost-at-Risk (CaR) of a well-defined debt issuance strategy. More specifically, the task was to examine the accuracy of the CaR estimates by calibrating the model to pseudo-data in order to compare the results to the actual CaR implied by the properties of the process which generated the pseudo-data.

In order to be able to calculate CaR figures, we assumed a borrower who issues debt annually in the form of zero-coupon bonds in the maturities of 1, 5 and 10 years. The borrowing costs were allocated to different years by interpreting each bond as a series of forward loans, each with a life of one year; the cost of the bond was allocated between different years on the principle that the outstanding debt grows at the forward rate.

We found that the calibration of the Longstaff–Schwartz (LS) model using the method proposed by Longstaff and Schwartz themselves has a high propensity to fail even when the data-generating process is exactly as specified by the model and the parameter values of the DGP are those obtained by Longstaff and Schwartz (1993). The model parameters would take on illegal values in about 25% of all cases, so that no CaR estimate could be computed. We went on to develop an enhanced calibration method based on exploiting the information reflected in the entire yield curve, whereas the Longstaff–Schwartz method only uses the levels and variance of the short rate. The new method works very well when the model specification is correct, never failing to deliver a CaR estimate and being more accurate than the original method.

The accuracy of the LS model was compared to that of a cruder benchmark model based on the Nelson–Siegel (NS) model of the yield curve, whose beta coefficients were modelled as ARMA(2,1) processes. The new calibration method for the LS model performed miserably when the DGP was of the NS type. Calibrated using the Longstaff–Schwartz method, the LS model was no better than the NS model

when the DGP was of the LS type, but outperformed the NS model when the DGP was, in fact, of the NS type. In other words, neither model got much advantage from being correctly specified; on the contrary, the LS model performed much better in a NS environment than on its 'home turf'.

This outcome is sufficiently counterintuitive to suggest that it may have been specific to the parameter values used in the generation of pseudo-data. The relationship between the simulation parameters and the relative performance of the two models should be studied further; until such a study is carried out, the robustness of our results remains in doubt. Caution is therefore advised in the interpretation of the above-mentioned findings.

Appendix: US government bond yields, 1953–2013

In this appendix, we analyze historical bond yield data and compare the results to the Stylized Facts of Section 1.2. We repeat them here for convenience:

1. The yield curve is upward-sloping most of the time and on average.
2. The yields on all maturities are strongly but imperfectly correlated. The smaller the difference in maturity, the higher the correlation between two yields.
3. Shocks to yield levels are highly persistent.
4. Long-term yields are less volatile than short-term yields.
5. The conditional volatilities of yield changes are time-varying and persistent.
6. The conditional volatilities of yield changes are positively correlated with the level of yields.

To examine the accuracy of these statements, we use a data set stitched together from two sources. The first part comes from the US Federal Reserve; it is publicly available for download at `http://www.federalreserve.gov/releases/h15/data.htm` (under the heading 'Treasury constant maturities'). These constant-maturity yields have been calculated from the prices of traded US federal government bonds using a spline interpolation method explained on the said website.

The second source is the financial news service Bloomberg. Bloomberg's data has both gaps and errors, some of them rather hard to spot, so we only use this as a secondary source to complement the Fed's data. Table 1 summarizes what periods we have data for, and from which source. For example, we are using Bloomberg data on 30-year yields from March 2002 to January 2006; in February 2006, Fed data becomes available again.

Table 1: The available data and its sources. m/d: monthly/daily data from Federal Reserve, B: daily data from Bloomberg.

From	3m	6m	1y	2y	3y	5y	7y	10y	30y
1953-04			m		m	m		m	
1962-01	B	B	d		d	d		d	
1969-07	B	B	d		d	d	d	d	
1976-06	B	B	d	d	d	d	d	d	
1977-02	B	B	d	d	d	d	d	d	d
1982-01	d	d	d	d	d	d	d	d	d
2002-03	d	d	d	d	d	d	d	d	B
2006-02	d	d	d	d	d	d	d	d	d

Most of the data is daily data (with non-business days removed), but for the years 1953–1961 only monthly data is available. We make use of both monthly and daily data. Monthly yield levels are calculated as simple averages over the business days in that month.

Because of the differences in data availability for the different maturities, we break the data into three periods:

1. Apr1953–Dec1961: 105 monthly observations in four maturities
2. Jan1962–Jan1977: 181 monthly and 3760 daily observations in six maturities
3. Feb1977–Mar2013: 434 monthly and 9024 daily observations in all nine maturities.

We now work through our six Stylized Facts to see whether they are supported by the data.

The slope of the yield curve

That the yield curve is upward-sloping on average is easy to verify. Tables 2–4 show this: in each of our subperiods, the average yield is a consistently increasing function of maturity. (Obviously, each subperiod must have its own table, given that the overall level of interest rates was different in different periods, so that the comparison of average yields on different maturities only makes sense when these are calculated over the same period.)

Table 2: Distribution of yield levels, Apr1953–Dec1961.

	1y	3y	5y	10y
Mean	2.79	3.16	3.30	3.41
Min	0.82	1.47	1.85	2.29
Max	5.14	5.12	5.01	4.72

Table 3: Distribution of yield levels, Jan1962–Jan1977.

	3m	6m	1y	3y	5y	10y
Mean	5.00	5.19	5.48	5.74	5.85	5.90
Min	2.68	2.78	2.98	3.33	3.56	3.83
Max	8.95	9.12	9.36	8.66	8.63	8.43

Table 4: Distribution of yield levels, Feb1977–Mar2013.

	3m	6m	1y	2y	3y	5y	7y	10y	30y
Mean	5.27	5.45	5.71	6.02	6.19	6.50	6.75	6.92	7.24
Min	0.01	0.04	0.10	0.21	0.33	0.62	0.98	1.53	2.59
Max	16.18	15.46	16.72	16.46	16.22	15.93	15.65	15.32	14.68

Tables 2–4 also show the range of variation for the yield on each maturity. In each of our three subperiods, minimum yield is an increasing function of maturity, as one would expect. Maximum yield, however, peaks at one year, and is decreasing in maturities beyond that; and the same applies to the difference between minimum and maximum.

We now turn to examine the statement that the yield curve is upward-sloping not only on average, but most of the time. This can be seen from Tables 5–7, where the excess yields on longer maturities relative to the shortest-term yield are tabulated. The first row of each table, which gives the average excess yield, gives no new information beyond what is implicit in Tables 2–4. The second row, however, shows that longer-term yields are almost always higher than short-term yields. At the same time, the third row reminds us that the yield curve *can*, at times, be steeply decreasing: in every subperiod, the minimum excess yields are negative throughout.

We also note that, with one exception (Table 7, 6m vs. 1y), the minimum excess yields are larger in absolute value, the longer the maturity. Also with one exception (Table 7, 7y vs. 10y), the maximum excess yield is an increasing function

Table 5: Distribution of excess yields over the 1-year yield, Apr1953–Dec1961.

	3y	5y	10y
Mean	0.37	0.51	0.62
% positive	94%	92%	88%
Min	-0.07	-0.13	-0.45
Max	0.85	1.23	1.74

Table 6: Distribution of excess yields over the 3-month yield, Jan1962–Jan1977.

	6m	1y	3y	5y	10y
Mean	0.19	0.48	0.74	0.85	0.91
% positive	95%	97%	93%	92%	86%
Min	-0.17	-0.35	-1.01	-1.21	-1.25
Max	0.57	1.34	2.18	2.58	2.91

Table 7: Distribution of excess yields over the 3-month yield, Feb1977–Mar2013.

	6m	1y	2y	3y	5y	7y	10y	30y
Mean	0.18	0.44	0.76	0.93	1.24	1.48	1.65	1.97
% positive	90%	92%	89%	89%	88%	89%	89%	89%
Min	-1.01	-0.94	-1.76	-2.01	-2.25	-2.46	-2.62	-3.06
Max	1.64	2.66	3.59	3.84	4.06	4.17	4.15	4.60

of maturity.

The fact that maximum yields are rather decreasing than increasing in maturity, suggests a hypothesis: perhaps the yield curve tends to slope down when the general level of yields is high. But this cannot be determined by studying Tables 2–7. Instead, we may calculate the correlation coefficients between yield levels and the excess yields they represent relative to the shortest-term yield. This is done in Tables 8–10.

In each of the Tables 8–10, the first row shows how the slope of the yield curve is related to the shortest-term yield available in the data. The slope is measured as the difference of each longer-term yield to the short-term yield; we call this the excess yield, and it is generally different for different maturities. The second row of each table shows how the excess yield on each maturity is correlated to yield on that same maturity.

If the slope and the level of the yield curve were uncorrelated, we should see

Table 8: Explaining the slope of the yield curve with its level, Apr1953–Dec1961. The first row shows the correlations between the 1-year yield and the excess yields on longer maturities (relative to the 1-year yield). The second row shows the correlation between a yield and the excess it contains.

	3y	5y	10y
1y	-0.69	-0.78	-0.87
idem	-0.55	-0.59	-0.61

Table 9: Explaining the slope of the yield curve with its level, Jan1962–Jan1977.

	6m	1y	3y	5y	10y
3m	-0.02	0.06	-0.29	-0.35	-0.43
idem	0.08	0.28	0.15	0.20	0.23

Table 10: Explaining the slope of the yield curve with its level, Feb1977–Mar2013.

	6m	1y	2y	3y	5y	7y	10y	30y
3m	0.06	0.34	0.06	-0.13	-0.33	-0.45	-0.54	-0.68
idem	0.13	0.44	0.23	0.10	-0.05	-0.13	-0.20	-0.31

correlation coefficients close to zero on the first row and positive ones on the second row. On the other hand, if the yields on different maturities were uncorrelated, we should have negative correlations on the first row and positive ones on the second row.

Instead, Table 8 shows strong negative correlations on both rows. This can only mean that from 1953 to 1961, it was indeed the case that the yield curve tended to slope down when rates were high. This is supported by Figure 1, which shows a scatterplot of the 1- and 10-year yields for this period. Plotting the 3- or 5-year yields gives qualitatively similar results.

However, the results in the two other tables do not give equally strong support for the hypothesis. The results from the middle period (1962–1977, Table 9) rather discredit it. The third period (1977–2013, Table 10) gives modest support if one looks at the long-term end of the table, but it looks as though the slope of the curve in the short-term end is positively, not negatively correlated with the level of yields.

Be that as it may; what we set out to investigate was whether Stylized Fact #1 is indeed a fact, and this has been established.

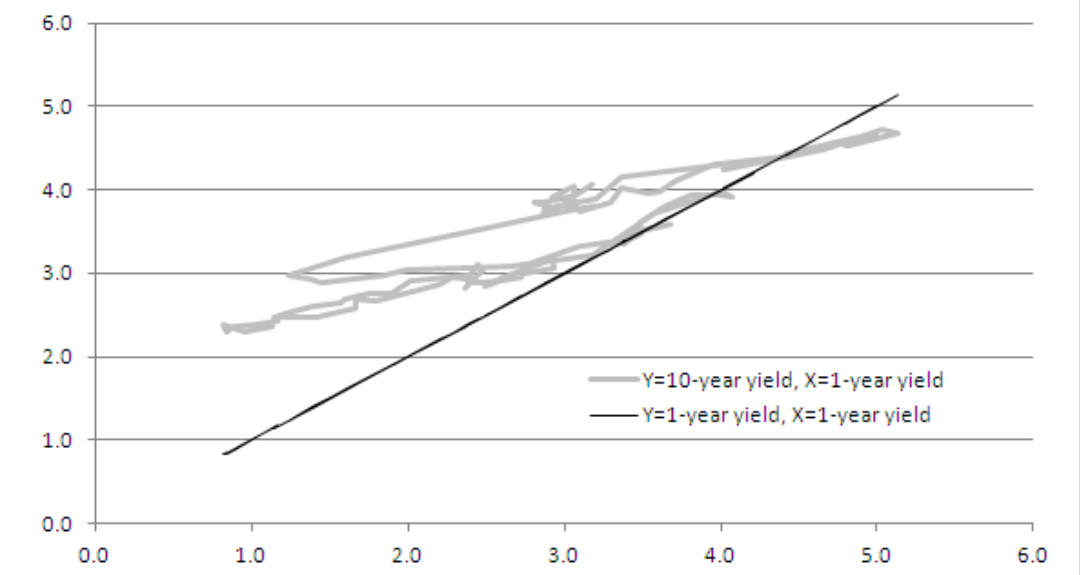

Figure 1: In years 1953–1961, the slope of the yield curve was negatively correlated with its level.

Correlation between yields on different maturities

The natural approach to testing Stylized Fact #2 would be to look at the correlation matrix for the different yields in each of the subperiods. However, that might lead to false conclusions, for the following reason. As is stated by one of the other Stylized Facts, and will be shown below, there is considerable autocorrelation in each of the yields. Consequently, each of our subperiods seems as if there was a trend in the data, which invalidates drawing conclusions from correlations between yield levels (it would be like explaining the Consumer Price Index with cumulative rainfall—the correlation is very high, because both just keep rising, but in fact they have nothing to do with each other).

Instead, we look at the first differences of monthly yield levels, i.e. the month-to-month change of the yields. The correlations between these are displayed in Tables 11–13.

We find that the correlations between the changes of yields on different maturities are very strong—the stronger, the less the difference in maturity. On the other hand, we also observe that the correlation between the far ends of the yield curve is nowhere near perfect, which completes the verification of Stylized Fact #2.

The level of correlation between any two given yields seems fairly stable from

Table 11: Correlations between month-to-month changes in yield levels, Apr1953–Dec1961.

	1y	3y	5y	10y
1y	1.00	0.90	0.84	0.76
3y	0.90	1.00	0.97	0.91
5y	0.84	0.97	1.00	0.96
10y	0.76	0.91	0.96	1.00

Table 12: Correlations between month-to-month changes in yield levels, Jan1962–Jan1977.

	3m	6m	1y	3y	5y	10y
3m	1.00	0.96	0.86	0.73	0.67	0.57
6m	0.96	1.00	0.93	0.80	0.75	0.64
1y	0.86	0.93	1.00	0.92	0.87	0.77
3y	0.73	0.80	0.92	1.00	0.98	0.87
5y	0.67	0.75	0.87	0.98	1.00	0.90
10y	0.57	0.64	0.77	0.87	0.90	1.00

Table 13: Correlations between month-to-month changes in yield levels, Feb1977–Mar2013.

	3m	6m	1y	2y	3y	5y	7y	10y	30y
3m	1.00	0.97	0.90	0.82	0.76	0.68	0.62	0.59	0.54
6m	0.97	1.00	0.96	0.91	0.85	0.79	0.74	0.70	0.64
1y	0.90	0.96	1.00	0.97	0.94	0.88	0.83	0.79	0.73
2y	0.82	0.91	0.97	1.00	0.99	0.95	0.91	0.88	0.81
3y	0.76	0.85	0.94	0.99	1.00	0.98	0.95	0.92	0.85
5y	0.68	0.79	0.88	0.95	0.98	1.00	0.99	0.97	0.91
7y	0.62	0.74	0.83	0.91	0.95	0.99	1.00	0.99	0.94
10y	0.59	0.70	0.79	0.88	0.92	0.97	0.99	1.00	0.97
30y	0.54	0.64	0.73	0.81	0.85	0.91	0.94	0.97	1.00

one subperiod to another. E.g. the correlation between the 1- and 10-year yields is 0.76, 0.77 and 0.79.

The correlations between *undifferenced* yield levels are all in excess of 0.9 for periods 1953–1961 and 1977–2013, and upwards from 0.78 for the middle period. Here, the correlation for adjacent maturities is near-perfect.

Table 14: Autocorrelation in yield levels; lags are in months.

Lag	3m	6m	1y	2y	3y	5y	7y	10y	30y
1	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99
2	0.96	0.97	0.97	0.97	0.98	0.98	0.98	0.98	0.98
3	0.95	0.95	0.96	0.96	0.96	0.97	0.96	0.97	0.97
4	0.93	0.93	0.94	0.95	0.95	0.96	0.95	0.96	0.96
5	0.91	0.92	0.93	0.94	0.94	0.94	0.94	0.95	0.95
10	0.84	0.85	0.86	0.88	0.88	0.89	0.88	0.91	0.89
20	0.64	0.66	0.70	0.74	0.76	0.78	0.76	0.81	0.79
30	0.50	0.52	0.59	0.64	0.67	0.71	0.69	0.76	0.74
50	0.35	0.36	0.45	0.43	0.53	0.57	0.54	0.62	0.56

Table 15: Autocorrelation in linearly detrended yield levels.; lags are in months.

Lag	3m	6m	1y	2y	3y	5y	7y	10y	30y
1	0.98	0.98	0.99	0.97	0.99	0.99	0.98	0.99	0.97
2	0.96	0.96	0.97	0.92	0.98	0.98	0.96	0.98	0.92
3	0.93	0.94	0.95	0.88	0.96	0.97	0.94	0.97	0.88
4	0.91	0.92	0.94	0.84	0.95	0.96	0.92	0.96	0.83
5	0.89	0.90	0.93	0.80	0.94	0.94	0.90	0.95	0.79
10	0.81	0.82	0.86	0.61	0.88	0.89	0.82	0.91	0.58
20	0.57	0.59	0.70	0.18	0.76	0.78	0.63	0.81	0.22
30	0.41	0.43	0.59	-0.07	0.67	0.71	0.51	0.76	0.03
50	0.25	0.26	0.45	-0.40	0.53	0.57	0.24	0.62	-0.29

Autocorrelation in yields

Stylized Fact #3 can be examined by looking at the autocorrelation in each variable. In this case, it is not necessary to divide the data into subperiods, because autocorrelation can be calculated for each variable independently. Table 14 shows the results. We find very persistent positive autocorrelation in all maturities.

However, some of the data series are shorter than others, and indeed look as if they had a trend; this is particularly true for the 30-year yield. As mentioned in the previous section, a trend can invalidate inferences based on autocorrelation. We therefore provide a supplementary table based on linearly detrended variables (Table 15).

Comparing Tables 14 and 15 reveals that when detrended variables are used, the autocorrelations on longer lags are weakened considerably in the 2-, 7- and 30-year yields, while 1-, 3-, 5- and 10-year yields are hardly affected (using more digits in

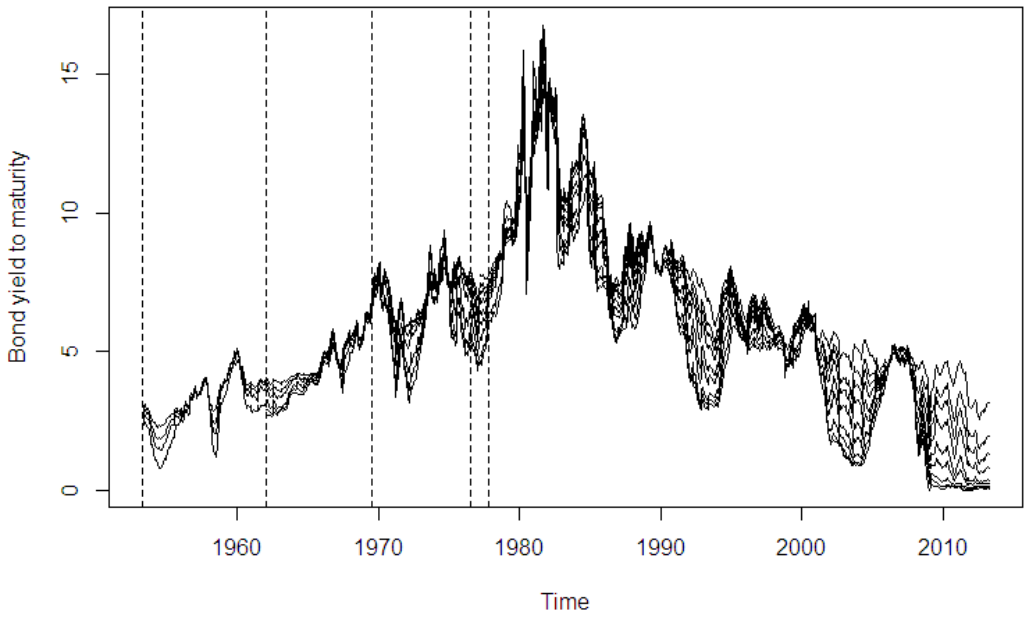

Figure 2: The variables under study are graphed in solid lines. The vertical lines show the starting points of the data series for (left to right) i) 1-, 3-, 5- and 10-year yields, ii) 3- and 6-month yields, iii) 7-year, iv) 2-year and v) 30-year yields.

the table would show the difference). Plotting these yields' time-paths (Figure 2) shows why. For those variables on which we have the longest data set, the time-path as a whole happens to have a shape which is not affected by detrending except up to a *shift*; whereas the shortest data set (30-year yield) is also *tilted*. If we drew the trendline for the 30-year yield, we would see the variable's path crossing the trendline many times. The trendline for the 1-, 3-, 5- and 10-year yields, on the other hand, would be roughly horizontal, and would be crossed fewer times.

The graphs suggest that in this case, the non-detrended data sets may be a better guide to the autocorrelations than the detrended ones. At any rate it is clear that Stylized Fact #3 is overwhelmingly supported by the data.

Long-term yields are less volatile than short-term yields

Volatility may refer to the degree of uncertainty regarding either the *level* of a variable or the *change in its level* over some period of time. Tables 16–18 show these for

Table 16: Standard deviation of yield levels and 1st differences, Apr1953–Dec1961.

	1y	3y	5y	10y
Level	1.05	0.91	0.81	0.66
Δ_{monthly}	0.25	0.20	0.17	0.13

Table 17: Standard deviation of yield levels and 1st differences, Jan1962–Jan1977.

	3m	6m	1y	3y	5y	10y
Level	1.52	1.52	1.57	1.47	1.45	1.41
Δ_{daily}	0.08	0.07	0.06	0.05	0.05	0.03
Δ_{monthly}	0.34	0.35	0.35	0.29	0.24	0.19

Table 18: Standard deviation of yield levels and 1st differences, Feb1977–Mar2013.

	3m	6m	1y	2y	3y	5y	7y	10y	30y
Level	3.54	3.57	3.71	3.64	3.53	3.34	3.20	3.05	2.73
Δ_{daily}	0.11	0.10	0.09	0.09	0.09	0.08	0.08	0.08	0.07
Δ_{monthly}	0.51	0.48	0.48	0.44	0.41	0.37	0.35	0.33	0.29

the different yields in each of the three subperiods separately, so that the different maturities can be compared. The standard deviations for yield levels are calculated from monthly data. The first differences are calculated from both daily and monthly data where available.

The volatilities of yield changes clearly support the notion that long-term yields are less volatile than short-term yields. By and large, the standard deviations of yield levels support the same conclusion, but it is perhaps a bit surprising that here, the volatility peaks not in the very short-term end of the maturity spectrum, but at the maturity of one year. We are reminded of a finding we made earlier, when examining Stylized Fact #1, that the range of variation of yields is the broadest at this same maturity (Tables 2–4).

At this point we make a little excursion. Casual inspection of graphed time-series of yields suggests that the yield dynamics may have changed somewhere during the eighties or nineties. Let us see what happens if we only use the last 20 years of data, starting from April 1994. Table 19 shows the results. While the standard deviation of yield level continues to support the idea of volatility being decreasing in maturity, the volatilities of daily and monthly changes lend no support to it.

Table 19: Standard deviation of yield levels and daily and monthly changes, Apr1993–Mar2013.

	3m	6m	1y	2y	3y	5y	7y	10y	30y
Level	2.13	2.17	2.16	2.12	2.03	1.82	1.66	1.46	1.19
Δ_{daily}	0.05	0.05	0.05	0.06	0.06	0.07	0.07	0.06	0.06
Δ_{monthly}	0.20	0.20	0.21	0.23	0.25	0.25	0.24	0.23	0.20

Table 20: Standard deviation of yield levels and daily and monthly changes, Jan2000–Mar2013.

	3m	6m	1y	2y	3y	5y	7y	10y	30y
Level	2.02	2.04	1.96	1.86	1.75	1.53	1.36	1.14	0.84
Δ_{daily}	0.06	0.05	0.05	0.06	0.06	0.07	0.07	0.06	0.06
Δ_{monthly}	0.21	0.20	0.20	0.22	0.24	0.24	0.24	0.23	0.21

This conclusion remains the same if we further restrict our data set to include only the 2000s (Table 20).

In other words, Stylized Fact #4 receives only somewhat mixed support from the last 20 years of data. Still, it is confirmed when we use all available data without getting selective.

Heteroskedasticity

Stylized Fact #5 says that the degree of volatility of yields is changes over time, and these changes are persistent, i.e. there is autocorrelation in the level of locally calculated volatility. We start by looking at the distribution of the monthly volatilities of day-to-day yield changes (Tables 21–22). For years 1953–1961, daily changes are not available; we could use semiannual volatilities of month-to-month changes, but they would not be comparable with the figures in the other two tables, so we just omit them.

To be more specific, the monthly volatility is calculated as $\sqrt{\text{Avg}\big((\Delta y)^2\big)}$, where Δy is the difference in yield compared to the previous business day, and the average is taken over the month in question.

Tables 21–22 show that the volatility ranges from near-zero to as much as 55 basis points per day for short rates, and 28 bp for the 10-year rate (which is enough to make bonds a risky investment, on a par with stocks). But the fact that the sample

Table 21: Distribution of monthly standard deviation of day-to-day change in yield, Jan1962–Jan1977.

	3m	6m	1y	3y	5y	10y
Mean	0.06	0.06	0.05	0.05	0.04	0.03
Min	0.01	0.01	0.00	0.01	0.01	0.00
1st qrt	0.02	0.03	0.02	0.02	0.02	0.01
Median	0.05	0.05	0.04	0.04	0.04	0.03
3rd qrt	0.07	0.07	0.07	0.06	0.05	0.04
Max	0.34	0.19	0.16	0.14	0.12	0.09

Table 22: Distribution of monthly standard deviation of day-to-day change in yield, Feb1977–Mar2013.

	3m	6m	1y	2y	3y	5y	7y	10y	30y
Mean	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.06
Min	0.01	0.00	0.01	0.01	0.01	0.02	0.02	0.02	0.01
1st qrt	0.02	0.03	0.03	0.04	0.04	0.05	0.05	0.05	0.04
Median	0.04	0.04	0.05	0.06	0.06	0.06	0.06	0.06	0.05
3rd qrt	0.07	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.07
Max	0.55	0.45	0.39	0.32	0.35	0.31	0.31	0.28	0.21

volatility estimates vary in time does not prove that the population volatility has changed; the same would be observed if the distribution of day-to-day yield changes simply had a high kurtosis with no serial dependence.

To establish heteroskedasticity, we can estimate an ARCH(1) model for the day-to-day changes of each variable. The model specification is

$$\mathbb{E}(x_t^2 \mid \mathcal{F}_{t-1}) = \sqrt{\alpha_0 + \alpha_1 x_{t-1}^2}$$

Tables 23 and 24 display the results. The α_1 coefficient is significant at vanishing p-values in every case. Some of the coefficients imply that the best-fitting model is unstable, rejecting the ARCH(1) specification; nevertheless, heteroskedasticity is established.

To get an idea of the persistence of volatility shocks, let us look at the monthly volatilities of day-to-day yield changes again. This time we examine their autocorrelations (Tables 25 and 26). For the purpose of testing against the null hypothesis that there is no serial dependence in the data, the standard error can be calculated as $(T-k)^{-0.5}$, where T is the length of the series and k is the lag. We find significant

Table 23: α_1 coefficients from an ARCH(1) model for day-to-day yield changes, Jan1962–Jan1977.

	3m	6m	1y	3y	5y	10y
α_1	1.17	1.05	0.84	0.55	0.51	0.63
p-value	$\sim$0	$\sim$0	$\sim$0	$\sim$0	$\sim$0	$\sim$0

Table 24: α_1 coefficients from an ARCH(1) model for day-to-day yield changes, Feb1977–Mar2013.

	3m	6m	1y	2y	3y	5y	7y	10y	30y
α_1	1.77	1.49	1.07	0.54	0.36	0.36	0.29	0.32	0.30
p-value	$\sim$0	$\sim$0	$\sim$0	$\sim$0	$\sim$0	$\sim$0	$\sim$0	$\sim$0	$\sim$0

Table 25: Autocorrelation in monthly volatility of day-to-day yield changes, Jan1962–Jan1977. Lags are in months. Standard error ranges from 0.074 to 0.081.

Lag	3m	6m	1y	3y	5y	10y
1	0.83	0.82	0.73	0.66	0.68	0.69
5	0.65	0.67	0.66	0.57	0.54	0.56
10	0.51	0.54	0.58	0.44	0.37	0.38
20	0.16	0.31	0.36	0.33	0.32	0.24
30	0.09	0.19	0.26	0.13	0.12	-0.06

Table 26: Autocorrelation in monthly volatility of day-to-day yield changes, Feb1977–Mar2013. Lags are in months. Standard error ranges from 0.048 to 0.050.

Lag	3m	6m	1y	2y	3y	5y	7y	10y	30y
1	0.82	0.85	0.85	0.78	0.75	0.72	0.74	0.74	0.80
5	0.68	0.73	0.75	0.64	0.58	0.54	0.53	0.51	0.58
10	0.57	0.64	0.65	0.59	0.54	0.50	0.49	0.50	0.53
20	0.40	0.45	0.43	0.30	0.26	0.23	0.25	0.26	0.30
30	0.21	0.24	0.22	0.08	0.05	-0.01	0.00	0.01	0.03

autocorrelation even at long lags. Stylized Fact #5 is therefore substantiated.

Conditional volatility is higher when yield level is high

Figure 3 displays the path of the 3-month yield together with its day-to-day changes. Visual comparison of the graphs suggests that day-to-day volatility tends to be higher when yields are high. We now test this by regressing the monthly volatility

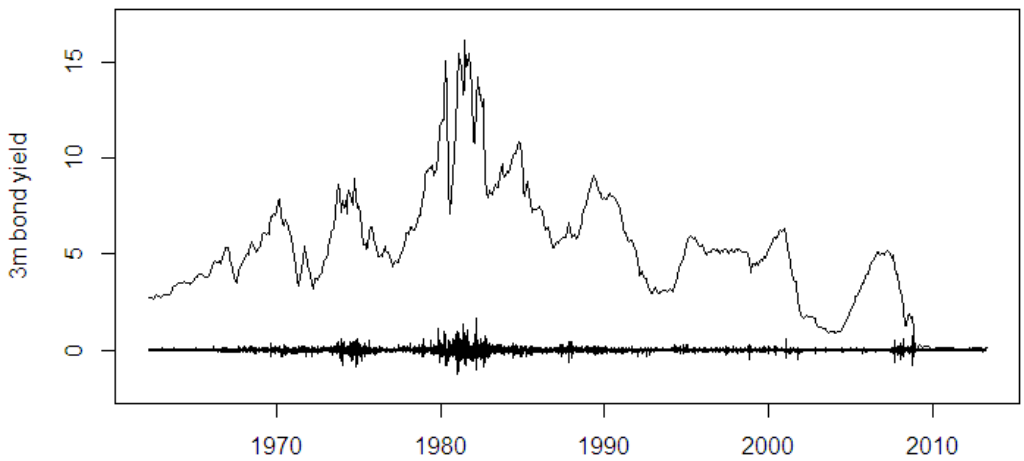

Figure 3: The 3-month yield and its day-to-day changes, Feb1962–Mar2013.

of day-to-day changes on the monthly average yield. The regression specification is

$$v_t = \mu + \beta y_t + \epsilon_t$$

where v_t is the volatility of day-to-day yield changes in month t on the maturity being examined; y_t is the average yield level in month t; μ and β are constants to be estimated; and ϵ_t is a random disturbance. We estimate this by OLS for the 3-month, 6-month, 1-, 3-, 5- and 10-year yields for the period January 1962–March 2013.

The regression results are tabulated in Table 27. Since the yields are measured in percentage points, the slope coefficients indicate that on average, a one-percentage point increase in yields is associated with a rise in day-to-day volatility of about 0.7–1.7 basis points, the increase being larger for shorter maturities. All slope coefficients are significant at convincing p-levels.

Unfortunately, because both yield levels and conditional volatility are serially correlated, as we have seen, the linear regression is not reliable. Autocorrelation means that the observations in our data set are not really independent, effectively making the data set smaller.

Table 27: Linear regression of monthly volatility on yield level, Jan1962–Mar2013.

	3m	6m	1y	3y	5y	10y
μ	-0.0185	-0.0159	-0.0111	0.0147	0.0168	0.0093
β	0.0166	0.0147	0.0131	0.0082	0.0073	0.0072
$p(\beta)$	0.00	0.00	0.00	0.00	0.00	0.00
$\mathrm{Corr}(v, y)$	0.69	0.72	0.72	0.57	0.52	0.52

Table 28: Linear regression of changes of monthly volatility on changes of yield level, Jan1962–Mar2013.

	3m	6m	1y	3y	5y	10y
μ	0.0001	0.0000	0.0000	0.0001	0.0001	0.0001
β	0.0163	0.0086	0.0071	0.0115	0.0146	0.0149
$p(\beta)$	0.00	0.01	0.02	0.00	0.00	0.00
$\mathrm{Corr}(v, y)$	0.17	0.11	0.09	0.14	0.18	0.18

To reduce the possible effect of autocorrelation, we can look at the first differences of both variables. If the two variables really are independent, then even if each one is autocorrelated, they ought to exhibit no tendency to move in the same direction. Running a linear regression with the same specification as before, but replacing v_t and y_t with their first differences, we get the results reported in Table 28. We find that the correlation between the changes of level and of volatility is weaker than the correlation between their levels; nevertheless, the slope coefficients remain significant and are of the same order of magnitude as in the previous regression.

However, there is one more concern. Looking at Figure 3, it seems possible that the early 1980s simultaneous bump in both yield level and volatility might alone be causing the positive relation we found between level and volatility. What would happen if we only used the data for the last 20 years?

Running the same linear regressions again for the period Apr1993–Mar2013, we get the results reported in Tables 29 and 30. We see that the positive correlation between yield levels and volatility levels has largely disappeared. When volatilities are regressed on yield levels, we find statistically significant positive correlation only in the short end of the curve, and the correlation coefficients are much lower than when the entire 51-year data was used.

When volatility changes are regressed on yield changes, we find *negative* corre-

Table 29: Linear regression of monthly volatility on yield level, Apr1993–Mar2013.

	3m	6m	1y	3y	5y	10y
μ	0.0233	0.0216	0.0271	0.0530	0.0629	0.0650
β	0.0050	0.0042	0.0039	0.0012	-0.0004	-0.0014
$p(\beta)$	0.00	0.00	0.00	0.15	0.59	0.12
$\mathrm{Corr}(v, y)$	0.27	0.31	0.32	0.09	-0.03	-0.10

Table 30: Linear regression of changes of monthly volatility on changes of yield level, Apr1993–Mar2013.

	3m	6m	1y	3y	5y	10y
μ	-0.0005	-0.0002	-0.0003	-0.0003	-0.0002	-0.0001
β	-0.0380	-0.0076	-0.0093	-0.0039	-0.0011	0.0013
$p(\beta)$	0.00	0.36	0.17	0.52	0.84	0.79
$\mathrm{Corr}(v, y)$	-0.20	-0.06	-0.09	-0.04	-0.01	0.02

lation in the 3-month maturity, and no significant correlation at all in the other maturities.

In summary: when we use the whole data set, excluding only the years 1953–1961 where daily data is not available and day-to-day volatilities hence cannot be calculated, we find that Stylized Fact #6 holds, since not only are volatility levels positively correlated with yield levels, but the same can be said of the first differences of both. However, this positive correlation largely disappears if we use the data for the last 20 years only.

Conclusion

The conclusion from this entire Appendix is that all of our Stylized Facts are supported by the data set as a whole, although recent data from the last 20 years gives only mixed support for #4 and hardly any support for #6.

Bibliography

[1] Ahn, Dong-Hyun & Dittmar, Robert & Gallant, Ronald (2002). Quadratic Term Structure Models: Theory and Evidence. *The Review of Financial Studies* 15, 243–288

[2] Baxter, Martin & Rennie, Andrew (1996). *Financial calculus: an introduction to derivative pricing.* Cambridge University Press, Cambridge.

[3] Black, Fischer & Derman, Emanuel & Toy, William (1990). A One-Factor Model of Interest Rates and Its Application to Treasury Bond Options. *Financial Analysts Journal* 46 (January–February), 33–39.

[4] Black, Fischer & Karasinski, Piotr (1991). Bond and Option Pricing when Short Rates are Lognormal. *Financial Analysts Journal* 47 (July–August), 52–59.

[5] Black, Fischer & Scholes, Myron (1973). The Pricing of Options and Corporate Liabilities. *Journal of Political Economy* 81, 637–659.

[6] Bolder, David (2001). Affine Term Structure Models: Theory and Implementation. Bank of Canada, Working Paper 2001-15.

[7] Bollerslev, Tim (1986). Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics* 31, 307–327.

[8] Brace, Lana & Gatarek, Dariusz & Musiela, Marek (1997). The Market Model of Interest Rate Dynamics. *Mathematical Finance* 7, 127–147.

[9] Brandt, Michael & Chapman, David (2002). Comparing Multifactor Models of the Term Structure. Working paper, Duke University and Boston College, downloaded on 13th April 2013 from `https://faculty.fuqua.duke.edu/\textasciitildembrandt/papers/working/mfcompare.pdf`.

[10] Brennan, Michael & Schwartz, Eduardo (1979). A Continuous Time Approach to the Pricing of Bonds. *Journal of Banking and Finance* 3, 133–155.

[11] Brigo, Damiano & Mercurio, Fabio (2006). *Interest Rate Models – Theory and Practice.* Springer, Berlin.

[12] Cairns, Andrew (2004). *Interest Rate Models: An Introduction.* Princeton University Press, Princeton, New Jersey.

[13] Campbell, John (1986). A Defense of Traditional Hypotheses about the Term Structure of Interest Rates. *The Journal of Finance*, 41, 183–193

[14] Campbell, John & Shiller, Robert (1991). Yield Spreads and Interest Rate Movements: A Bird's Eye View. *Review of Economic Studies* 58, 495–514.

[15] Chan, K. & Karolyi, G. & Longstaff, F. & Sanders, A. (1992). An Empirical Comparison of Alternative Models of the Short-Term Interest Rate. *Journal of Finance* 47, 1209–1227.

[16] Chapman, David & Pearson, Neil (2001). Recent Advances in Estimating Term-Structure Models. *Financial Analysts Journal* Vol. 57, No. 4, 77–95.

[17] Clewlow, Les & Strickland, Chris (1994). A Note on Parameter Estimation in the Two-Factor Longstaff and Schwartz Interest Rate Model. *The Journal of Fixed Income*, March 1994, 95–100.

[18] Cox, John & Ingersoll, Jonathan & Ross, Stephen (1981). A Re-examination of Traditional Hypotheses about the Term Structure of Interest Rates. *Journal of Finance* 36, 769–799.

[19] Cox, John & Ingersoll, Jonathan & Ross, Stephen (1985a). An Intertemporal General Equilibrium Model of Asset Prices. *Econometrica* 53, 363–384.

[20] Cox, John & Ingersoll, Jonathan & Ross, Stephen (1985b). A Theory of the Term Structure of Interest Rates. *Econometrica* 53, 385–408.

[21] Culbertson, John (1957). The Term Structure of Interest Rates. *Quarterly Journal of Economics*, November 1957, 485–517.

[22] Danmarks Nationalbank (2000). Danish Government Borrowing and Debt 1999.

[23] Diebold, Francis & Li, Canlin (2006). Forecasting the Term Structure of Government Bond Yields. *Journal of Econometrics* 130, 337–364.

[24] Duffie, Darrell & Kan, Rui (1996). A Yield-Factor Model of Interest Rates. *Mathematical Finance* 6, 379–406.

[25] Enders, Walter (2004). *Applied Econometric Time Series.* Wiley, New Delhi.

[26] Heath, David & Jarrow, Robert & Morton, Andrew (1992). Bond Pricing and the Term Structure of Interest Rates: A New Methodology for Contingent Claims Valuation. *Econometrica* 60, 77–105.

[27] Ho, Thomas & Lee, Sang-Bin (1986). Term Structure Movements and Pricing Interest Rate Contingent Claims. *Journal of Finance* 41, 1011–1029.

[28] Homer, Sidney & Sylla, Richard (1991). *A History of Interest Rates.* 3rd edition. Rutgers University Press, New Brunswick.

[29] Hong Yan (2001). Dynamic Models of the Term Structure. *Financial Analysts Journal* Vol. 57, No. 4, 60–76.

[30] Hull, John (2006). *Options, Futures and Other Derivatives.* Pearson Prentice Hall, Upper Saddle River, New Jersey.

[31] Hull, John & White, Alan (1990). Pricing Interest Rate Derivative Securities. *Review of Financial Studies* 3, 573–592.

[32] Hull, John & White, Alan (1994a). Numerical Procedures for Implementing Term Structure Models I: Single-Factor Models. *Journal of Derivatives* 2, 7–16.

[33] Hull, John & White, Alan (1994b). Numerical Procedures for Implementing Term Structure Models II: Two-Factor Models. *Journal of Derivatives* 2, 37–48.

[34] James, Jessica & Webber, Nick (2000). *Interest Rate Modelling.* Wiley, Chichester.

[35] Jamshidian, Farshid (1997). LIBOR and Swap Market Models and Measures. *Finance and Stochastics* 1, 293–330.

[36] Jönsson, Per-Olof & Sangarabalan, Sanga (2010). Modelling Cost at Risk: A Preliminary Approach for Beginners. Working paper, Oxford Policy Management.

[37] Langetieg, Terence (1980). A Multivariate Model of the Term Structure. *Journal of Finance* 35, 71–97.

[38] Litterman, Robert & Scheinkman, José (1991). Common Factors Affecting Bond Returns. *Journal of Fixed Income* 1, 54–61.

[39] Longstaff, Francis & Schwartz, Eduardo (1992). Interest Rate Volatility and the Term Structure: A Two-Factor General Equilibrium Model. *Journal of Finance* 47, 1259–1282.

[40] Longstaff, Francis & Schwartz, Eduardo (1993). Implementation of the Longstaff–Schwartz Interest Rate Model. *The Journal of Fixed Income*, September 1993, 7–14.

[41] Longstaff, Francis & Schwartz, Eduardo (1994). Comments on "A Note on Parameter Estimation in the Two-Factor Longstaff and Schwartz Interest Rate Model". *The Journal of Fixed Income*, March 1994, 101–102.

[42] Merton, Robert (1973). Theory of Rational Option Pricing. *The Bell Journal of Economics and Management Science*, Vol. 4, No. 1, 141–183.

[43] Mikosch, Thomas (1998). *Elementary Stochastic Calculus with Finance in View.* World Scientific Publishing, Singapore.

[44] Miltersen, Kristian & Sandmann, Klaus & Sondermann, Dieter (1997). Closed Form Solutions for Term Structure Derivatives with Log-Normal Interest Rates. *Journal of Finance* 52, 409–430.

[45] Modigliani and R. Sutch (1966). Innovations in Interest Rate Policy. *American Economic Review*, May 1966, 178–197.

[46] Nelson, Charles & Siegel, Andrew (1987). Parsimonious Modeling of Yield Curves. *Journal of Business* 60, 473–489.

[47] Pearson, Neil & Sun, Tong-sheng (1990) An Empirical Examination of the Cox, Ingersoll, Ross Model of the Term Structure of Interest Rates Using the Method of Maximum Likelihood. Working paper, Managerial Economics Research Center.

[48] Pearson, Neil & Sun, Tong-sheng (1994) Exploiting the Conditional Density in Estimating the Term Structure: An Application to the Cox, Ingersoll, and Ross Model. *Journal of Finance* 49, 1279–1304.

[49] Piazzesi, Monika (2010). Affine Term Structure Models. In *Handbook of Financial Econometrics*, ed. by Aït-Sahalia, Yacine & Hansen, Lars. North-Holland, Amsterdam.

[50] Piazzesi, Monika & Schneider, Martin (2006). Equilibrium Yield Curves. *NBER Macroeconomics Annual 2006*, Volume 21. Cambridge MA, MIT Press.

[51] Rogers, L. (1995). Which Model for Term-Structure of Interest Rates Should One Use? *The IMA Volumes in Mathematics and its Applications* 67, 93–115.

[52] Schaefer, Stephen & Schwartz, Eduardo (1984). A Two-Factor Model of the Term Structure: An Approximate Analytical Solution. *The Journal of Financial and Quantitative Analysis* 19, 413–424.

[53] Shimko, David (1992). *Finance in Continuous Time.* Kolb, Miami.

[54] Vasicek, Oldrich (1977). An Equilibrium Characterization of the Term Structure. *Journal of Financial Economics* 5, 177–188.