

# Affine term structure models

Monika Piazzesi\*

March 31, 2003

## 1 Introduction

### 1.1 Overview

The quest for understanding what moves bond yields has produced an enormous literature with its own journals and graduate courses. Those who want to join the quest are faced with considerable obstacles. The literature has evolved mostly in continuous time, where stochastic calculus reigns and partial differential equations spit fire. The knights in this literature are fighting for different goals, which makes it often difficult to comprehend why the quest is moving in certain directions. But the quest is moving fast, and dragons are being defeated. This chapter wants to report some of these victories made by those working on *affine term structure models*.

Bond yield movements over time can be captured by simple vector autoregressions in yields and maybe other macroeconomic variables. Several aspects of bond yields, however, set them apart from other variables typically used in VAR studies. One aspect is that bonds are assets, and that bonds with many different maturities are traded at the same time. Bonds with long maturities are risky when held over short horizons, and risk-averse investors demand compensation for bearing such risk. Arbitrage opportunities in these markets exist unless long yields are risk-adjusted expectations of average future short rates. Movements in the cross section of yields are therefore closely tied together. These ties show up as *cross-equation restrictions* in a yield-VAR. Another aspect of yields is that they are not normally distributed, at least not until recently. This makes it difficult to compute the risk-adjusted expected value of future short rates.

Term structure models capture exactly these aspects of bond yields. They impose

---

\*UCLA and NBER. Anderson School, 110 Westwood Plaza, Los Angeles CA 90095, piazzesi@ucla.edu. Thanks to Pedro Santa-Clara and Martin Schneider for many helpful discussions. Thanks to David Chapman, Pierre Collin-Dufresne, Mike Johannes, Bob Kimmel, Antonio Mele, Bruno Miranda, Juha Seppala and Selale Tuzel for comments.

the cross-equation restrictions implied by no-arbitrage and allow yields to be nonnormal. The word ‘affine term structure model’ is often used in different ways. I will use the word to describe any arbitrage-free model in which bond yields are affine (constant-plus-linear) functions of some state vector  $x$ .<sup>1</sup> Affine models are thus a special class of term structure models which write the yield  $y^{(\tau)}$  of a  $\tau$ -period bond as

$$y^{(\tau)} = A(\tau) + B(\tau)^\top x$$

for coefficients  $A(\tau)$  and  $B(\tau)$  that depend on maturity  $\tau$ . The functions  $A(\tau)$  and  $B(\tau)$  make these yield equations consistent with each other for different values of  $\tau$ . The functions also make the yield equations consistent with the state dynamics.

Tractability is the main advantage of affine models. Tractability is important, because otherwise yields need to be computed with Monte Carlo methods or solution methods for partial differential equations. Both approaches are computationally costly, especially when model parameters are estimated using panel data on bond yields. The literature on bond pricing starting with Vasicek (1977) and Cox, Ingersoll, and Ross (1985, CIR hereafter) therefore has focused on closed-form solutions. The riskless rate in these early setups was the only state variable in the economy, so that all bond yields were perfectly correlated. A number of extensions of these setups followed, both in terms of the number of state variables and the data-generating processes for these variables. Duffie and Kan (1996) finally provided a complete characterization of models with affine bond yields.

Tractability has to be paid with restrictive assumptions. The functional form of bond yields is obtained from computing risk-adjusted expectations of future short rates. Therefore, restrictive assumptions have to be made on the risk-adjusted dynamics of the state vector. More concretely, the risk-adjusted process for the state vector needs to be an affine diffusion, a process with affine instantaneous mean and variance. There are no functional form assumptions on the data-generating process for the state vector. Now the question is whether this assumption implies data-generating processes for yields that can be rejected in the data. The answer seems to be ‘yes’ when risk premia imply either constant or time-varying but strictly positive expected excess returns. Recent research, however, has made more flexible assumptions on risk premia. The answer now seems to be ‘maybe not.’ A lot more research is needed before the answer is clear - exciting times lie ahead!

The rest of this chapter is organized as follows. The remainder of this introduction argues the importance of understanding bond yields in Subsection 1.2 and the need for cross-equation restrictions in Subsection 1.3. Section 2 explains the general technique of how to price bonds in continuous time. Section 3 explains how to specify the short rate, the dynamics of the state vector and the risk premia in an affine model. Section 4

---

<sup>1</sup>New terms such as ‘completely affine,’ ‘essentially affine,’ ‘semi-affine’ and ‘generalized affine’ have appeared in the literature. The use of ‘affine’ in this paper refers to the way yields depend on the state variables, not on the data-generating process of the state variables themselves.

links them to the fundamentals of an economy. Section 6 explains how to estimate affine models. Section 7 discusses the empirical performance of affine models.

## 1.2 Why care about bond yields?

Understanding what moves bond yields is important for at least four reasons. One of these reasons is forecasting. Yields on long-maturity bonds are expected values of average future short yields, at least after an adjustment for risk. This means that the current yields curve contains information about the future path of the economy. Yield spreads have indeed been useful for forecasting not only future short yields (Fama and Bliss (1987), Campbell and Shiller (1991), Cochrane and Piazzesi (2002)), but also real activity (Harvey (1988), Estrella and Hardouvelis (1991), Hamilton and Kim (2002), Ang, Piazzesi, and Wei (2002)) and inflation (Mishkin (1990), Fama (1990)), even though these forecasting relationships may be unstable (Stock and Watson (2001)). These forecasts provide a basis for investment decisions of firms, savings decisions of consumers, and policy decisions.

Monetary policy is a second reason for studying the yield curve. In most industrialized countries, the central bank seems to be able to move the short end of the yield curve. What matters for ‘aggregate demand,’ however, are long-term yields. For example, U.S. households base their decision on whether to buy or rent a house on long-term mortgage rates and not on the rate in the federal funds market which seems to be controlled by the Federal Reserve Bank. For a given state of the economy, a model of the yield curve helps to understand how movements at the short end translate into longer-term yields. This involves understanding both how the central bank conducts policy and how the transmission mechanism works. The expectations hypothesis is at work in most papers in this area (for example, Balduzzi, Bertola, and Foresi (1996)). Little work has been done with more flexible risk premia (Babbs and Webber (1993), Piazzesi (2001), Evans and Marshall (1998, 2001)).

Debt policy constitutes a third reason. When issuing new debt, governments need to decide about the maturity of the new bonds. For example, the Kennedy administration actively managed the maturity structure of public debt in the early 1960s in what is known as ‘operation twist.’ The treasury at the time was trying to flatten or invert the yield curve by selling short maturity debt and buying long maturity notes. The outcome of such operations depends crucially on how bond yields depend on the supply of bonds with different maturities. Real yields in models with nondistortionary taxation and perfect markets are independent of the maturity structure of public debt. The reason is that Modigliani-Miller in these models applies to how the government finances its budget deficit. Cochrane (2001) characterizes the dependence of the nominal term structure on debt policy in a frictionless economy. Missale (1997) considers distortionary taxation, while Angeletos (2001) assumes that markets are incomplete.

Derivative pricing and hedging provide a fourth reason. For example, coupon bonds are priced as baskets of coupon payments weighted by the price of a zero-coupon bond that matures on the coupon date. Even the price of more complicated securities, such as swaps, caps and floors, futures and options on interest rates are computed from a given model of the yield-curve (see the references in Duffie, Pan, and Singleton (2000)). Banks need to manage the risk of paying short-term interest rates on deposits while receiving long-term interest rates on loans. Hedging strategies involve contracts that are contingent on future short rates, such as swap contracts. To compute these strategies, banks need to know how the price of these derivative securities depends on the state of the economy.

### 1.3 Why care about cross-equation restrictions?

Some of the issues just mentioned, such as forecasting and the impact of Fed interventions on long-term yields, may be addressed without imposing the cross-equation restrictions implied by no arbitrage. I can add measurement error  $\varepsilon_t^{(\tau)}$  to each yield equation

$$y_t^{(\tau)} = A(\tau) + B(\tau)^\top x_t + \varepsilon_t^{(\tau)}$$

select specific variables for  $x$ , and then run an *unrestricted* regression of yields  $y^{(\tau)}$  on  $x$  for each maturity  $\tau$  separately. Least squares is easy, fast and delivers consistent estimates of parameters, at least conditional on the linear structure. For example, Fama and Bliss (1987) forecast changes in short rates without imposing the cross-equation restrictions implied by the absence of arbitrage. Evans and Marshall (1998) estimate the impact of policy shocks on long-term bonds outside of a yield-curve model.

More patience is required to estimate a system of yield equations in a way that ensures no arbitrage. The cross-equation restrictions have to be derived from parameters that describe the state dynamics and risk premia. Although the model is affine in the state vector  $x$ , the functions  $A(\tau)$  and  $B(\tau)$  are nonlinear functions of the underlying parameters. Using ordinary least squares is thus no longer possible. Maximum likelihood is not feasible either, because the density of yields is not available in closed form. There are a few exceptions for which the density is known, such as normal densities for yields, but they are easily rejected by the data. New econometric methods have been produced to solve these estimation problems, and this handbook shows some of these exciting developments. The implementation of these methods, however, require substantial coding and computation time. Before rolling up the sleeves and getting into the work of implementing cross-equation restrictions when distributions are nonnormal, I would therefore like to spend some time explaining why we actually need them.

Cross-equation restrictions have many advantages. First, these restrictions ensure that the yield dynamics are consistent.  $A(\tau)$  and  $B(\tau)$  make yield equations consistent with each other in the cross-section and in the time series. Most bond markets are extremely liquid, and arbitrage opportunities are traded away immediately by large investment banks. The assumption of no-arbitrage thus seems natural for bond yields.

Second, term structure models allow us to separate risk premia from expectations about future short rates. These models are therefore key to understanding to what extent investors think of long bonds as safe investments. Sargent (1979) and Hansen and Sargent (1991) are early papers that explore the expectations hypothesis under which expected excess bond returns are zero. Modified versions of the expectations hypothesis have been tested under which expected excess returns are constant. These tests compare, for example, the ratio of the likelihood function with and without restrictions implied by the expectations hypothesis (for references, see Bekaert and Hodrick (2001)). The evidence suggests that expected returns on long bonds are on average higher than on short bonds, and that they are time-varying. Cross-equation restrictions are then needed to model these risk premia.

Third, unrestricted regressions imply that the number of variables needed to describe the yield curve equals the number of yields in the regression. Lower-dimensional systems have been shown to work well in approximating true yield dynamics. Factor decompositions of the variance-covariance matrix of yield changes show that over 97% of the variance is attributable to just three principal components. Litterman and Scheinkman (1991) named these three principal components ‘level’, ‘slope’ and ‘curvature’ according to how shocks to these factors affect the yield curve. This interpretation of the driving forces of yields seems to be stable across model specifications, estimation samples and types of interest rates. Measurement errors arising from the data construction methods, data entry errors and asynchronous data sampling (of LIBOR and swap yields, for example) are responsible for at least some of the remaining variance of yields.

Fourth, the number of estimated parameters in unrestricted regressions is usually large. Imposing the cross-equation restrictions from no-arbitrage improves the efficiency of these estimates. Ang and Piazzesi (2002) show that this helps out-of-sample forecasting of yields.

Finally, ‘missing bond yields’ can be recovered from a small set of other yields in a way that is consistent with no arbitrage. Certain multifactor models predict yields that were not included in the estimation within a couple of basis points. This property of yield-curve models is important for studies of emerging markets where bonds with only few maturities are traded at any given point in time. Alvarez and Neumeyer (1999), for example, apply interpolation methods to construct yields for Argentina. The same issue arises in the context of the construction of zero-coupon bond yields. Nelson and Siegel (1987), Fama and Bliss (1987) and McCulloch and Kwon (1993) propose interpolation methods to infer these yields from observed prices of traded coupon bonds or interest-rate derivatives. These interpolation methods ignore that bond yields need to be consistent with risk-adjusted expectations of interpolated future short rates. These methods thereby admit arbitrage opportunities, which can be avoided with a term structure model.

## 2 Basics

### 2.1 Bond pricing in continuous time

Term structure modelling determines the price of zero-coupon bonds. These bonds pay a terminal payoff, usually normalized to 1 unit, without risk of default and without paying any intermediate coupons. A zero-coupon bond that matures  $\tau$  periods from now trades at price  $P^{(\tau)}$ . Buying this bond at time  $t$  and reselling it at that time  $t+n$  generates a log holding period return of

$$hpr_{t \rightarrow t+n}^{(\tau)} = \log P_{t+n}^{(\tau-n)} - \log P_t^{(\tau)}. \quad (1)$$

The holding period  $n$  cannot exceed time to maturity  $\tau$ , so we have  $n \leq \tau$ . The holding period return is usually random, because it depends on the resale value of the bond  $P_{t+n}^{(\tau-n)}$  which is generally not known at time  $t$ . The resale value is equal to its payoff when the bond matures, so that holding the bond until maturity ( $n = \tau$ ) generates a return which is known at time  $t$ . The per-period holding period return in this case is the *yield-to-maturity*:

$$y_t^{(\tau)} = \frac{hpr_{t \rightarrow t+\tau}^{(\tau)}}{\tau} = -\frac{\log P_t^{(\tau)}}{\tau}.$$

The short rate is the limit of yields as maturity approaches,  $r_t = \lim_{\tau \downarrow 0} y_t^{(\tau)}$ . Excess holding period returns  $hprx_{t \rightarrow t+n}^{(\tau)} = hpr_{t \rightarrow t+n}^{(\tau)} - hpr_{t \rightarrow t+n}^{(n)}$  are returns made in excess of the riskless return over the holding period.

Bonds are usually priced with the help of a so-called “risk-neutral probability measure”  $Q^*$ . Just like the name of this artificial measure suggests, risk-neutral pricing applies under  $Q^*$ . In other words, asset prices are the expected values of their future payoffs discounted at the riskless rate, where the expectation is computed using the probability measure  $Q^*$ . When agents are risk-neutral, this pricing result applies under the data-generating measure  $Q$ . In general, the risk-neutral probability measure  $Q^*$  will be different from  $Q$ . The payoff of zero-coupon bonds is 1 unit at maturity, so their price is

$$P_t^{(\tau)} = E_t^* \left[ \exp \left( - \int_t^{t+\tau} r_u du \right) \right]. \quad (2)$$

where  $E^*$  denotes expectation under  $Q^*$ . Standard results show that if there exists a risk-neutral probability measure  $Q^*$ , a system of asset prices is arbitrage free. The converse is also true under reasonable restrictions on trading strategies. Moreover, the uniqueness of  $Q^*$  is equivalent to markets being complete. Details and references for these powerful results can be found, for example, in Duffie (2001).

Under the risk-neutral measure, expected excess returns on bonds are zero. Put differently, the expected rate of return on a long bond equals the riskfree rate. I think

the gain in intuition justifies the following abuse in notation:

$${}^{\text{“}}E_t^* \left[ \frac{dP_t^{(\tau)}}{P_t^{(\tau)}} \right] = r_t dt^{\text{”}}. \quad (3)$$

This is abuse because “ $dP_t^{(\tau)}$ ” is not even a random variable. Expected values of instantaneous log-holding period returns are only approximately equal to (3), because of Jensen’s inequality terms.

The pricing relation (2) shows that any *yield-curve model* consists of two ingredients:

- (i) the change of measure from  $Q$  to  $Q^*$ , and
- (ii) the dynamics of the short rate  $r$  under  $Q^*$ .

In so-called *factor models* of the yield curve, (ii) is replaced by the following assumption:

- (ii)’ the short rate  $r$  is a function  $R(x)$  of  $x$  and  $x \in \mathbb{R}^N$  is a time-homogeneous Markov process under  $Q^*$ .

This means that  $x$  is the relevant state vector, a vector of *factors*. This modified (ii)’ assumption implies that the conditional expectation in (2) is some function  $F$  of time-to-maturity  $\tau$  and the state  $x_t$  at time  $t$ , or

$$P_t^{(\tau)} = F(x_t, \tau).$$

To capture certain features of yield data (seasonalities around macroeconomic news releases, for example), I will later consider functions  $R$  that also depend on time  $t$  and time-inhomogeneous Markov processes  $x$ , in which case  $P_t^{(\tau)} = F(x_t, t, \tau)$  separately depends on  $t$  and  $\tau$  (in addition to  $x_t$ ).

The big advantage of pricing bonds (or any other assets) in continuous time is Ito’s Lemma. The lemma says that smooth functions  $F$  of some Ito process  $x$  and time  $t$  are again Ito processes (see Duffie (2001), Chapter 5 for details). The lemma thus preserves the Ito property even if  $F$  is nonlinear. Ito’s Lemma allows me to turn the problem of solving the conditional expectation in (2) into the problem of solving a partial differential equation (PDE) for the bond price  $F(x, \tau)$ . The trick of computing (2) by solving a PDE is called the Feynman-Kac approach. I will first explain the local expectations hypothesis in Section 2.2 and then use it to derive the PDE for bond prices in Section 2.3. Section 2.4 derives the PDE without local expectations hypothesis.

## 2.2 Local expectations hypothesis (LEH)

The local expectations hypothesis (LEH) states that the pricing relation (2) holds under the data-generating measure  $Q$ . Bond yields are thus given by

$$\text{LEH : } y_t^{(\tau)} = -\log E_t[\exp(-S)] / \tau, \quad (4)$$

where  $S = \int_t^{t+\tau} r_u du$ . The LEH therefore amounts to risk-neutral pricing: the data-generating measure  $Q$  and the risk-neutral measure  $Q^*$  coincide. This means that expected excess returns on long bonds are zero.

The LEH is not the same as the more prominent “expectations hypothesis” (EH) which states that bond yields  $y_t^{(\tau)}$  are expected values of average future short rates, or

$$\text{EH: } y_t^{(\tau)} = E_t[S] / \tau. \quad (5)$$

The difference between the two hypotheses (4) and (5) is due to Jensen’s inequality. For example, suppose that the short rate is Gaussian under  $Q = Q^*$  which implies that  $S$  is also Gaussian (as the sum of Gaussians). With this normality assumption, equation (4) becomes

$$y_t^{(\tau)} = E_t[S] / \tau - \frac{1}{2} \text{var}_t[S] / \tau,$$

which differs from (5) because of the variance term.

For example, suppose that the short rate is a random walk with normally distributed shocks. More concretely, the short rate  $r$  solves the stochastic differential equation (SDE)

$$dr_t = \sigma_r dz_t$$

where  $z$  is a standard Brownian motion (under the data-generating measure) and  $\sigma_r$  is some constant. The shocks  $dz_t$  are, loosely speaking, independently normally distributed with mean 0 and variance  $dt$ . I can solve for the short rate explicitly as

$$r_t = r_0 + \int_0^t \sigma_r dz_u = r_0 + \sigma_r z_t,$$

because the Brownian motion  $z_0$  starts at 0 with probability 1. The EH predicts a flat yield curve in this case,

$$y_t^{(\tau)} = E_t[S] / \tau = E_t \left[ \int_t^{t+\tau} r_t + \sigma_r (z_u - z_t) du \right] / \tau = r_t,$$

because  $E_t(z_u - z_t) = 0$  for all  $u \geq t$ . The LEH predicts a downward-sloping yield curve, because<sup>2</sup>

$$y_t^{(\tau)} = r_t - \frac{\text{var}_t[S]}{2\tau} = r_t - \frac{\sigma_r^2 \tau^2}{6}.$$

---

<sup>2</sup>The variance can be computed as

$$\begin{aligned} \text{var}_t \left[ \int_t^{t+\tau} z_u - z_t du \right] &= \text{var} \left[ \int_0^\tau z_u du \right] = \int_0^\tau \int_0^\tau \text{cov}(z_u, z_s) du ds \\ &= \int_0^\tau \int_0^\tau \min\{s, u\} du ds = \frac{\tau^3}{3} \end{aligned}$$

CIR (1981) argue that the EH is not consistent with no arbitrage. Counterexamples to this argument exist for some special economies (for example, Fisher and Gilles (1998)). Longstaff (2000a) argues that market incompleteness may make it impossible to actually exploit such arbitrage opportunities. Campbell (1986) finds that the Jensen's inequality terms tend to be small in the data, except in periods of high volatility such as the end of the 1970s and for bonds with long maturities.

## 2.3 Partial differential equation for bond prices with LEH

For now, I assume that the LEH holds. An advantage of the LEH is that there is no need to know how to change the probability measure in step (i). Another advantage is that we have some intuition about the parameters that determine the dynamics of the short rate under the data-generating measure, while we do not have such intuition about the parameters under the risk-neutral measure. The LEH is therefore a useful starting point. I will discuss the change of measure in the next subsection.

In continuous time, a Markov process  $x$  lives in some state space  $D \subset \mathbb{R}^N$  and solves the stochastic differential equation (SDE)

$$dx_t = \mu_x(x_t) dt + \sigma_x(x_t) dz_t, \quad (6)$$

where  $z$  is an  $N$ -dimensional standard Brownian motion under  $Q$ ,  $\mu_x : D \rightarrow \mathbb{R}^N$  is the drift of  $x$  and  $\sigma_x : D \rightarrow \mathbb{R}^{N \times N}$  is its volatility. Gaussian processes have an affine drift  $\mu_x(x)$  and their volatility  $\sigma_x(x)$  is constant. Fat tails in the distribution of the state vector can be modelled by specifying an appropriate state-dependence for the volatility  $\sigma_x(x)$ . Another way to depart from Gaussianity is to model 'large moves' in the process  $x$ , which I will add in Section 3.5. The Markov process solving (6) is time-homogenous because the functions  $\mu_x$  and  $\sigma_x$  do not depend on time. The extension to time-inhomogeneous Markov processes is straightforward.

Bond prices can now be solved using the Feynman-Kac approach. The idea is to view the conditional expected value (2) as the solution of the PDE for the bond price  $F(x, \tau)$ . The PDE can be obtained in four steps. First, the pricing equation (2) implies that the price of the bond at maturity is equal to its payoff (here the bond price is taken to be *cum-dividend*). This means that  $F(x, 0) = 1$  for all  $x \in D$ . Second, the pricing equation also shows that the bond price is the expected value of an exponential function, so  $F(x, \tau)$  is strictly positive (which makes it possible to divide by  $F$ ). Third, Ito's Lemma implies that  $F(x, \tau)$  itself is an Ito process

$$\frac{dF(x_t, \tau)}{F(x_t, \tau)} = \mu_F(x_t, \tau) dt + \sigma_F(x_t, \tau) dz_t \quad (7)$$

with instantaneous expected bond return

$$\mu_F(x, \tau) = -\frac{F_\tau(x, \tau)}{F(x, \tau)} + \frac{F_x(x, \tau)^\top}{F(x, \tau)} \mu_x(x) + \frac{1}{2} \text{tr} \left[ \sigma_x(x) \sigma_x(x)^\top \frac{F_{xx}(x, \tau)}{F(x, \tau)} \right], \quad (8)$$

where  $F_\tau$ ,  $F_x$  and  $F_{xx}$  are partial derivatives of  $F$  and  $tr$  denotes trace. Finally, the LEH implies that the expected return  $\mu_F(x, \tau)$  is equal to the short rate  $r = R(x)$ . The following *Cauchy problem* summarizes these steps:

$$\begin{aligned}\mu_F(x, \tau) &= R(x) \\ F(x, 0) &= 1\end{aligned}\tag{9}$$

for all  $x \in D$  and  $\tau > 0$ . A number of regularity conditions are needed for the Feynman-Kac approach to work. For example,  $F(x, \tau)$  needs to be smooth enough for Ito's Lemma. These conditions are stated in Duffie (2001, Appendix E).

Bond prices can now be computed in different ways. The conditional expected value in (2) can be computed using Monte-Carlo methods. The PDE in (9) can be solved numerically. For small dimensional systems ( $N \leq 3$ ), solving the PDE is precise and relatively fast. For larger dimensional systems ( $N \geq 3$ ), Monte-Carlo methods tend to be more attractive. The alternative is to make strong functional form assumptions on the coefficients  $\mu(x)$  and  $\sigma(x)$  and the short-rate function  $R(x)$  so that the PDE has a closed form solution. The broad class of exponential-affine solutions for  $F(x, \tau)$  is called affine term structure models. The requirements on the coefficients and the short-rate function are laid out next.

## 2.4 Without LEH

The last step in the derivation of the PDE (9) for the bond price invoked the LEH to conclude that the expected return on long-term bonds  $\mu_F(x, \tau)$  is equal to the riskless rate  $R(x)$ . I will now derive the PDE for the (empirically relevant) case where *the LEH does not hold*. The key is to realize that expected returns are always equal to the riskless rate under the risk-neutral measure  $Q^*$ , or

$$\mu_F^*(x, \tau) = R(x).$$

Instead of the state-dynamics (6) under the data-generating measure, the state vector  $x$  solves

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

for a Brownian motion  $z^*$  under the risk-neutral measure  $Q^*$ . To get some intuition about risk-neutral coefficients, consider the case of a single state variable equal to the riskless rate,  $x = r$ . Risk-neutral pricing then applies after appropriately adjusting the distribution of the short rate. For example, the conditional density of the short rate may be needed to be shifted right, towards higher values of  $r$ . This would make the risk-neutral mean of the short rate higher than its actual mean. In this case, yields are roughly equal to the expected values of average future short rates  $r$ , but the expectation is computed

using a twisted distribution, a distribution with a higher mean. A unique feature of the continuous-time setting is that the volatility turns out to stay the same under both probability measures:  $\sigma_x^* = \sigma_x$ . In other words, changes of probability measure do not affect the variance of innovations to  $x$ , unless we allow for jumps.

Now I can derive the PDE for bond prices by relying on risk-neutral coefficients and then later link the risk-neutral dynamics of the state vector to its data-generating process. The SDE for the bond price is the ‘starred’ version of equation (7)

$$\frac{dF(x_t, \tau)}{F(x_t, \tau)} = \mu_F^*(x_t, \tau) dt + \sigma_F^*(x_t, \tau) dz_t^*$$

where  $z^*$  is a Brownian motion under  $Q^*$  and the formula for the expected rate of return  $\mu_F^*(x, \tau)$  is analogous to (8)

$$\mu_F^*(x, \tau) = -\frac{F_\tau(x, \tau)}{F(x, \tau)} + \frac{F_x(x, \tau)^\top}{F(x, \tau)} \mu_x^*(x) + \frac{1}{2} \text{tr} \left[ \sigma_x^*(x) \sigma_x^*(x)^\top \frac{F_{xx}(x, \tau)}{F(x, \tau)} \right],$$

with the difference of being based on the drift  $\mu_x^*(x)$  and the volatility  $\sigma_x^*(x)$  of  $x$  under  $Q^*$ .

The easiest way to write down a pricing model is to start with a process  $x$  under  $Q^*$  and to then link  $Q^*$  to the data-generating measure  $Q$ . These two ingredients imply a data-generating process for  $x$  which can be estimated. The change of measure captures risk adjustment. The change involves the *density*  $\xi$ , which is a strictly positive martingale (so that  $Q$  and  $Q^*$  agree on probability zero events) and starts at  $\xi_0 = 1$  (so that  $Q^*$  is a probability measure). The SDE is

$$\frac{d\xi_t}{\xi_t} = \sigma_\xi(x_t) dz_t. \tag{11}$$

where  $\sigma_\xi : D \rightarrow \mathbb{R}^{1 \times N}$ . Novikov’s condition makes  $\xi$  a martingale.<sup>3</sup> Now consider the process  $z^*$  defined by

$$dz_t^* = dz_t - \sigma_\xi(x_t)^\top dt.$$

Girsanov’s theorem (Duffie (2001, Appendix D)) implies that  $z^*$  is a Brownian motion under  $Q^*$ .

---

<sup>3</sup>The solution to (11) is  $\xi_t = \exp\left(\int_0^t \sigma_\xi(x_u) dz_u - \frac{1}{2} \int_0^t \sigma_\xi(x_u) \sigma_\xi(x_u)^\top du\right)$ . The process  $\xi$  is a martingale if Novikov’s condition is satisfied:

$$E \left[ \exp \left( \frac{1}{2} \int_0^T \sigma_\xi(x_u) \sigma_\xi(x_u)^\top du \right) \right] < \infty.$$

For more details, see Appendix D in Duffie (2001).

By inserting the definition of  $z^*$  into the SDE (10)

$$dx_t = \left( \mu_x^*(x_t) - \sigma_x^*(x_t) \sigma_\xi(x_t)^\top \right) dt + \sigma_x^*(x_t) dz_t$$

it becomes clear that the volatility of the state vector is the same under both measures

$$\sigma_x(x) = \sigma_x^*(x).$$

This is often called *diffusion invariance principle*. Only the drift changes:

$$\mu_x(x) = \mu_x^*(x) - \sigma_x(x) \sigma_\xi(x)^\top. \quad (12)$$

### 3 Affine models

Affine term structure models make functional-form assumptions in step (ii)' of yield-curve modelling which lead to tractable pricing formulas. The functional-form assumptions are on the short-rate function  $R(x)$  and the process  $x$  for the state vector under the risk-neutral measure. The functional form is affine in both cases:

- $R(x)$  is affine, and
- $x$  is an affine diffusion under  $Q^*$ :
  - the drift  $\mu_x^*(x)$  is affine, and
  - the variance matrix  $\sigma_x^*(x) \sigma_x^*(x)^\top$  is affine.

These functional form are for coefficients under the risk-neutral measure. In particular, the drift  $\mu_x(x)$  is affine under the data-generating measure only when  $\sigma_x(x) \sigma_\xi(x)^\top$  is affine, which can be seen from (12). The next sections make these assumptions more precise and show that bond prices  $F(x, \tau)$  are exponential-affine in  $x$ . In this setting, yields are thus affine in  $x$  which explains the name of this class of models.<sup>4</sup>

#### 3.1 Affine short rate

The functional form of the short rate is made precise in the following assumption.

**Assumption I.** The short rate is given by

$$r = R(x) = \delta_0 + \delta_1^\top x.$$

---

<sup>4</sup>Discrete time analogous to affine diffusions are defined in Darolles, Gouriéroux, and Jasiak (2001) and applied to term structure modeling in Gouriéroux, Monfort, and Polimenis (2002). For an introduction into discrete time affine models, see Backus, Foresi, and Telmer (1998).

for  $\delta_0 \in \mathbb{R}$  and  $\delta_1 \in \mathbb{R}^N$ .

The choice of short-rate parameters  $\delta_0$  and  $\delta_1$  depends on the number of factors in the model. The short rate usually is the factor in one-factor models, which means  $\delta_0 = 0$  and  $\delta_1 = 1$ . The short rate in one-factor models is Markov. In  $N$ -factor models, the short rate alone is not Markov, but the short rate together with  $N - 1$  yields is typically Markov. The short rate often serves as one of the factors in multidimensional models. In this case, we still have  $\delta_0 = 0$  and  $\delta_1 = (1, 0_{N-1})^\top$ . Long yields still depend on the other factors because the expected future path of the short rate depends on the current state  $x$  in (2), when the short rate covaries with these other factors under the risk-neutral measure.

### 3.2 Affine diffusions

Again, I will start by imposing the LEH, which means that risk-neutral pricing applies under the data-generating probability measure. I will therefore assume that the state vector is an affine diffusion under  $Q$ , which is more restrictive than necessary to get affine solutions for yields. The more general case of an affine diffusion under  $Q^*$  with flexible risk premia will appear in the next subsection. These risk premia may introduce nonlinearities in the data-generating process for  $x$ . I start with the following two assumptions:

**Assumption II.** The process  $x$  is an affine diffusion. This means that  $x$  solves

$$dx_t = \mu_x(x_t) dt + \sigma_x(x_t) dz_t,$$

with coefficients

$$\begin{aligned}\mu_x(x) &= \kappa(\bar{x} - x) \\ \sigma_x(x) &= \Sigma s(x)\end{aligned}$$

where  $s(x)$  is a diagonal  $N \times N$  matrix with  $i$ -th diagonal element  $s_i(x) = \sqrt{s_{0i} + s_{1i}^\top x}$ , and where  $s_{0i} \in \mathbb{R}$ ,  $\bar{x}$ ,  $s_{1i} \in \mathbb{R}^N$  and  $\Sigma, \kappa \in \mathbb{R}^{N \times N}$  are constants.

Some intuition about affine diffusions is easy to get in the univariate case. The affine drift  $\mu(x_t)$  makes sure that if the current state  $x_t$  is above its mean  $\bar{x}$ , the change  $dx_t$  is likely to be negative as long as  $\kappa > 0$ . If the current state  $x_t$  is instead below its mean  $\bar{x}$ , the change is likely to be positive. In both cases, the process  $x_t$  is likely to be pulled back to its mean. The speed of this adjustment is determined by  $\kappa$ . If the speed is zero,  $\kappa = 0$ , the process is nonstationary. The autoregressive coefficient of discretely sampled observations is  $\exp(-\kappa h)$ , where  $h$  is the interval length between two observations. Time is usually measured in years, so that  $h = 1$  is one year. Monthly and weekly observation intervals then simply mean that  $h = \frac{1}{12}$  and  $h = \frac{1}{52}$ , respectively. For daily data, the choice of  $h$  is less obvious. Most papers shorten the year to an average number of 250 business days so that  $h = \frac{1}{250}$ . Few papers take weekends and holidays seriously and set  $h = \frac{1}{365}$ .

Shocks  $dz_t$  disturb  $x_t$  from moving back to its mean. These shocks are normally distributed with mean zero and variance  $dt$ . The effect of these shocks on  $x_t$  is determined by the volatility  $\sigma_x(x_t)$ . With constant volatility, the normally distributed shocks  $dz_t$  translate into a conditional normal distribution for changes  $dx_t$ . More generally, shocks  $dz_t$  may translate more into  $dx_t$  during times of high volatility  $\sigma_x(x_t)$  and less in times of low volatility. This state-dependent amplification effect introduces conditional heteroskedasticity. In bond-yield data, the pattern of this heteroskedasticity seems to positively depend on the level of yields. The half-life  $H$  of shocks solves  $\exp(-H\kappa) = 0.5$ . For example, with  $\kappa = 5$ , the half life is  $H = -\log 0.5/\kappa = 0.1386$  years, about 7 weeks.

Gaussian processes and square-root processes are the best known examples of affine diffusions. The two classes differ with respect to their assumptions about the variance matrix  $\sigma_x(x)\sigma_x(x)^\top$ . *Gaussian processes* have a constant variance matrix which requires that  $s_{1i} = 0$  for  $i = 1, \dots, N$ . Without loss of generality, I can set  $s(x)$  equal to the identity matrix ( $s_{0i} = 1$ ), because the variance parameters  $\Sigma$  are free. The SDE (6) then becomes a so-called linear SDE (Karatzas and Shreve (1988), Chapter 5.6),

$$dx_t = \kappa(\bar{x} - x_t)dt + \Sigma dz_t.$$

Existence and uniqueness of solutions to linear SDEs are not problematic. The solution  $x$  is Gaussian and thus can take on negative values with positive probability.

*Square-root processes* introduce conditional heteroskedasticity by allowing  $\sigma_x(x)$  to depend on the state. Now additional restrictions are needed to ensure that the variance matrix  $\sigma_x(x)\sigma_x(x)^\top$  is positive definite. A univariate square-root process solves

$$dx_t = \kappa(\bar{x} - x_t)dt + \Sigma\sqrt{x_t}dz_t,$$

where  $\kappa, \bar{x}, \Sigma$  are now all scalars. For arbitrary parameter values  $(\kappa, \bar{x}, \Sigma)$  the conditional variance  $\Sigma^2 x_t$  may not be positive. The Feller condition  $\kappa\bar{x} > \frac{1}{2}\Sigma^2$  makes zero an entrance boundary. In other words, this condition makes sure that zero is never reached. This is important, because once the process hits zero, its conditional variance  $\Sigma^2 x_t$  collapses to zero as well. Intuitively, the parameter restriction ensures that the drift term is strong enough to always pull the process  $x$  away from the zero boundary. Note that the parameter restrictions rule out unit roots ( $\kappa = 0$ ). The solution of the last SDE only takes on positive values (which makes it possible to compute  $\sqrt{x}$ ). The conditional variance of square-root processes is thus proportional to the level of the process. The larger  $x$ , the higher its variance. For multidimensional but independent square-root processes, the Feller condition can be imposed equation-by-equation.

More generally, the coefficients  $\mu_x(x)$  and  $\sigma_x(x)$  need to satisfy regularity requirements to guarantee the existence of a unique solution to the SDE (6). These solutions  $x$  are called strong solutions, which means that any other Ito process that solves (6) is equal to  $x$  almost everywhere. The regularity requirements make sure that the solution does

not explode (*growth conditions*) and is unique (*Lipschitz conditions*).<sup>5</sup> While these conditions may sound like technical details, they severely restrict the correlation structure of affine diffusions. Moreover, they are not satisfied in even simple cases like square-root diffusions. (The volatility  $\Sigma\sqrt{x}$  does not satisfy the Lipschitz condition, which is why we need the Feller condition.) The following two standard examples are not affine, but they illustrate that these conditions are natural in the context of deterministic differential equations ( $\sigma_x(x) \equiv 0$ ). The first example is  $\mu_x(x) = x^2$  and  $x_0 = 1$  which does not satisfy the growth condition. The unique solution is  $x_t = \frac{1}{1-t}$ ,  $0 \leq t \leq 1$ , which explodes for  $t \rightarrow 1$ . The second example is  $\mu_x(x) = 3x_t^{2/3}$  and  $x_0 = 0$ , which does not satisfy the Lipschitz condition at  $x = 0$ . The equation has many solutions which are indexed by a scalar  $a > 0$ :  $x_t = 0$  for  $t \leq a$  and  $x_t = (t - a)^3$  for  $t > a$ .

Duffie and Kan (1996) provide a multidimensional extension of the Feller condition. Their condition handles general correlated affine diffusions. The condition ensures that only positive factors enter the volatility  $\sigma_x(x)$ . This involves restrictions on the correlations between state variables which prevent a potentially negative variable from pulling a variable that enters  $s(x)$  into the negative orthant. The condition is sufficient for the existence of a unique solution to the SDE (6). For necessary and sufficient conditions, see Theorem 2.7 in Duffie, Filipovic, and Schachermayer (2001).

**Condition A (sufficient for the existence of a solution to the SDE):**

1. For all  $x$  such that  $s_i(x) = 0$ ,  $s_{1i}^\top \kappa (\bar{x} - x) > \frac{1}{2} s_{1i}^\top \Sigma \Sigma^\top s_{1i}$ .
2. For all  $j$ , if  $(s_{1i}^\top \Sigma)_j \neq 0$ , then  $s_i(x)$  and  $s_j(x)$  are proportional.

The following examples illustrate how condition A restricts the admissible cross-correlations between state variables.

**Example 1.**  $x = (x_1, x_2)$  with  $s_{01} = 1$ ,  $s_{11} = (0 \ 0)^\top$  for  $x_1$  and  $s_{02} = 0$ ,  $s_{12} = (0 \ 1)^\top$  for  $x_2$ . Suppose first that  $\Sigma$  is diagonal and  $\kappa$  is unrestricted. In particular,  $\kappa_{21} \neq 0$  where  $\kappa_{21}$  is the (2, 1)-th element in  $\kappa$ , which means that the drift of  $x_2$  (which enters the volatility) depends on  $x_1$ . For condition A.1. to be satisfied, we need that  $\kappa_{21}(\bar{x}_1 - x_1) + \kappa_{22}\bar{x}_2 > \frac{1}{2}\Sigma_{22}^2$ . This inequality cannot hold for all  $x_1 \in R$  (which is a direction in which  $s_2(x) = 0$ ), so that the drift of  $x_1$  is not allowed to depend on  $x_2$  or

---

<sup>5</sup>A measurable function  $f$  satisfies Lipschitz and growth conditions in  $x$  if there exist constants  $c$  and  $k$  such that

$$\begin{aligned} \|f(x) - f(y)\| &\leq c \|x - y\| \\ \|f(x)\|^2 &\leq k (1 + \|x\|^2) \end{aligned}$$

for all  $x, y \in \mathbb{R}^N$ . The norm on matrices used here is  $\|A\| = \text{tr}(AA^\top)^{1/2}$  where “tr” denotes trace.

$\kappa_{21} = 0$ . Suppose now that  $\kappa$  is diagonal and  $\Sigma$  is unrestricted. In particular,  $\Sigma_{21} \neq 0$ . For the process to satisfy A.2., we need that  $s_1(x)$  and  $s_2(x)$  are proportional, which is not true. This implies that it must be that  $\Sigma_{21} = 0$ . Analogous considerations for  $x_2$  show that it is possible to have  $\kappa_{12} \neq 0$  and  $\Sigma_{12} \neq 0$ .

**Example 2.**  $x = (x_1, x_2)$  with  $s_{01} = 0$ ,  $s_{11} = (1 \ 0)^\top$  for  $x_1$  and  $s_{02}$ ,  $s_{12}$  as in Example 1,  $\Sigma$  is diagonal. Suppose  $\kappa$  is unrestricted with  $\kappa_{21} \neq 0$ . Again, condition A.1. requires  $\kappa_{21}(\bar{x}_1 - x_1) + \kappa_{12}\bar{x}_2 > \frac{1}{2}\Sigma_{22}^2$  for all directions in which  $s_2(x) = 0$ . The difference to example 1 is that  $x_1$  can only take on positive values. If  $\kappa_{21} < 0$ , it is now possible to choose parameters such that A.1. is satisfied for all  $x_1 \in \mathbb{R}_+$ . Condition A.2. rules out any off-diagonal terms in  $\Sigma$ .

The examples show that the main restriction coming from condition A is on the dependence of variables entering the volatility  $s(x)$  on other variables. These volatility-determining variables may not be correlated through  $\kappa$  with other variables that do not enter  $s(x)$  (as shown in example 1). Volatility-determining variables may, however, be correlated with each other through  $\kappa$ , provided this correlation is *positive* ( $\kappa_{12} < 0$  and  $\kappa_{21} < 0$  in example 2). Variables in  $s(x)$  cannot be conditionally correlated through  $\Sigma$  with any variable (examples 1 and 2). Other variables (that do not determine the volatility) are free to be correlated with variables in  $s(x)$ .

### Mean

For the univariate case, we can rewrite the SDE (6) for affine diffusions as

$$x_t = \bar{x} + \exp\{-\kappa(t-s)\}[x_s - \bar{x}] + \int_s^t \exp\{-\kappa(t-u)\}\Sigma s(x_u) dz_u. \quad (13)$$

for any value  $x_u$ ,  $0 \leq s \leq t$ . The same formula applies to the multivariate case, where  $e^{-\kappa(t-s)}$  is a matrix exponential. These are coded in MATLAB as “expm”. The conditional expected value can be computed immediately

$$E_s[x_t] = \bar{x} + \exp\{-\kappa(t-s)\}(x_s - \bar{x}). \quad (14)$$

The unconditional expected value  $E[x_t]$  solves

$$E[x_t] = \bar{x} + \exp\{-\kappa(t-s)\}(E[x_t] - \bar{x})$$

for stationary processes, which implies that  $E[x_t] = \bar{x}$ . Again, these are matrix exponentials.

### Variance

The conditional variance of affine diffusions is

$$\text{var}_s(x_t) = \int_s^t \exp\{-\kappa(t-u)\}\Sigma s(E_s[x_u])s(E_s[x_u])^\top \Sigma^\top \exp\{-\kappa^\top(t-u)\} du. \quad (15)$$

For Gaussian processes, the conditional variance is

$$\text{var}_s(x_t) = \int_s^t \exp\{-\kappa(t-u)\} \Sigma \Sigma^\top \exp\{-\kappa^\top(t-u)\} du.$$

For univariate Gaussians, this reduces to

$$\text{var}_s(x_t) = \Sigma^2 \frac{(1 - \exp\{-2\kappa(t-s)\})}{2\kappa}. \quad (16)$$

For univariate square root processes, the conditional variance boils down to

$$\text{var}_s(x_t) = \bar{x} \Sigma^2 \frac{(1 - \exp\{-\kappa(t-s)\})^2}{2\kappa} + x_t \Sigma^2 \frac{(\exp\{-\kappa(t-s)\} - \exp\{-2\kappa(t-s)\})}{\kappa}. \quad (17)$$

### 3.3 Affine bond pricing with LEH

To compute bond prices, I now add the assumption of risk-neutral pricing under  $Q$ .

**Assumption III.** The local expectations hypothesis holds.

Under the assumptions I, II, and III (and additional integrability conditions on the SDE coefficients for the Feynman-Kac approach to work stated in Duffie, Filipovic, and Schachermayer (2001), Section 11), Duffie and Kan (1996) guess a solution  $F(x, \tau)$  for the PDE (9) of the form

$$F(x, \tau) = \exp\left(a(\tau) + b(\tau)^\top x\right) \quad (18)$$

where the coefficients  $a(\tau) \in \mathbb{R}$  and  $b(\tau) \in \mathbb{R}^N$  solve the ODEs

$$\begin{aligned} a'(\tau) &= -\delta_0 + b(\tau)^\top \kappa \bar{x} + \frac{1}{2} \sum_{i=1}^N \left[ b(\tau)^\top \Sigma \right]_i^2 s_{0i} \\ b'(\tau) &= -\delta_1 - \kappa^\top b(\tau) + \frac{1}{2} \sum_{i=1}^N \left[ b(\tau)^\top \Sigma \right]_i^2 s_{1i} \end{aligned} \quad (19)$$

starting at  $a(0) = 0$  and  $b(0) = 0$ . This guess can be verified as follows. Given the exponential affine form (18), the instantaneous bond return for equation (8) is

$$\mu_F(x, \tau) = -a'(\tau) - b'(\tau)^\top x + b(\tau)^\top \mu_x(x) + \frac{1}{2} b(\tau)^\top \sigma_x(x) \sigma_x(x)^\top b(\tau). \quad (20)$$

The PDE in (9) and therefore equation (20) hold for all  $x$  in an open set  $D$ , so that the method of undetermined coefficients leads to the system of ODEs above.

The coefficients  $a(\tau)$  and  $b(\tau)$  can be computed in closed form only for a few cases. For example, the coefficients for a 1-factor model based on a square-root process are in CIR (1985), page 393. The coefficients for the 2-factor case with independent square-root processes are in Chen and Scott (1992), page 616. The coefficients for a 1-factor model based on a Gaussian process are in Vasicek (1977), page 186. More generally, the system of ODEs (19) can be solved fast and efficiently numerically using Runge-Kutta methods. The MATLAB command “ode45” performs the computation.

The bond-price equation (18) shows that the LEH together with a short rate which is affine in an affine diffusion (assumptions I, II, III) imply that yields are given by

$$y_t^{(\tau)} = -\frac{\log F(x_t, \tau)}{\tau} = A(\tau) + B(\tau)^\top x_t \quad (21)$$

for coefficients  $A(\tau) = -a(\tau)/\tau$  and  $B(\tau) = -b(\tau)/\tau$ .

### 3.4 Without LEH

I will now drop assumption III and modify assumption II in the following way.

**Assumption II’.** The process  $x$  solves

$$dx_t = \mu_x^*(x_t) dt + \sigma_x^*(x_t) dz_t^*$$

for a Brownian motion  $z^*$  under  $Q^*$  and coefficients

$$\begin{aligned} \mu_x^*(x) &= \kappa^* (\bar{x}^* - x) \\ \sigma_x^*(x) &= \Sigma^* s^*(x) \end{aligned}$$

where  $s^*(x)$  is a diagonal  $N \times N$  matrix with  $i$ -th diagonal element  $s_i^*(x) = \sqrt{s_{0i}^* + s_{1i}^{*\top} x}$ , and where  $s_{0i}^* \in \mathbb{R}$ ,  $\bar{x}^*, s_{1i}^* \in \mathbb{R}^N$  and  $\Sigma^*, \kappa^* \in \mathbb{R}^{N \times N}$  are constants.

To obtain exponential-affine bond-price solutions, the risk-neutral drift  $\mu_x^*(x)$  and variance-covariance matrix  $\sigma_x^*(x) \sigma_x^*(x)^\top$  need to be affine. Because of diffusion invariance, the variance-covariance matrix  $\sigma_x(x) \sigma_x(x)^\top$  under the data-generating measure needs to be affine as well. But the data-generating drift

$$\mu_x(x) = \mu_x^*(x) - \sigma_x(x) \sigma_\xi(x)^\top$$

may be nonlinear, depending on the functional form of  $\sigma_\xi(x)$ . The data-generating drift is only affine if the product  $\sigma_x(x) \sigma_\xi(x)^\top$  is affine. Many examples of affine yield-curve models in the literature described in Section 5 take the drift to be affine under both measures.

Assumptions I and II' (and again integrability conditions from Duffie, Filipovic, and Schachermayer (2001), Section 11) then imply that the ODEs for the bond-price coefficients become

$$\begin{aligned} a'(\tau) &= -\delta_0 + b(\tau)^\top \kappa^* \bar{x}^* + \frac{1}{2} \sum_{i=1}^N \left[ b(\tau)^\top \Sigma \right]_i^2 s_{0i} \\ b'(\tau) &= -\delta_1 - \kappa^{*\top} b(\tau) + \frac{1}{2} \sum_{i=1}^N \left[ b(\tau)^\top \Sigma \right]_i^2 s_{1i} \end{aligned} \quad (22)$$

where the risk-neutral parameters  $\kappa^*$  and  $\bar{x}^*$  replace  $\kappa$  and  $\bar{x}$  in (19). Finally, the drift-equation (20) has to be “starred” as well to hold under  $Q^*$ .

### 3.5 Jumps

Up to now, the state vector has been an affine diffusion under the risk-neutral probability measure. Diffusions evolve continuously through time. Large movements in yields, however, happen around macroeconomic news releases and Federal Reserve policy moves - at discrete points in time. These large movements can be modelled as discontinuous moves, or *jumps*, in the state vector. These jumps occur at arrival times  $t_1, \dots, t_n$  as in Figure 1. These arrival times are either stochastic or deterministic. Counting processes start at 0 and then record the number of jumps, as illustrated in the lower graph in Figure 1. The value of the state vector  $x$  “right before” a jump at time  $t$  is the left limit  $x_{t-} = \lim_{s \uparrow t} x_s$ . The jump in  $x$  at  $t$  is  $\Delta x_t = x_t - x_{t-}$ . The process  $x$  is right-continuous, as in the upper graph in Figure 1.

In principle, the conditional probability  $\lambda_t dt$  of a jump during the interval  $[t, t+dt]$  and the distribution of the jump size  $\Delta x_t$  conditional on a jump at time  $t$  may both depend on the state  $x_{t-}$ . In affine models, however, it turns out that there is a *dichotomy* under the risk-neutral measure between specifying the jump timing to be state dependent and specifying the jump size distribution to be state dependent. The two cannot be mixed together without giving up on tractability, so one of them has to be state independent. Either the conditional jump distribution depends on the state, but then the jump timing has to be deterministic, or the conditional probability of a jump depends on the state, but then the size distribution needs to be state independent. Taken together, these two types of jumps can be used to accommodate release calendars, central bank meetings and surprising events such as the Gulf war. Jumps at stochastic jump times have been introduced by Duffie and Kan (1996, Section 11), while jumps at deterministic jump times have been introduced by Piazzesi (2001).

Formally, jump-diffusions  $x$  solve

$$dx_t = \mu_x(x_{t-}) dt + \sigma_x(x_{t-}) dz_t + dJ_t$$

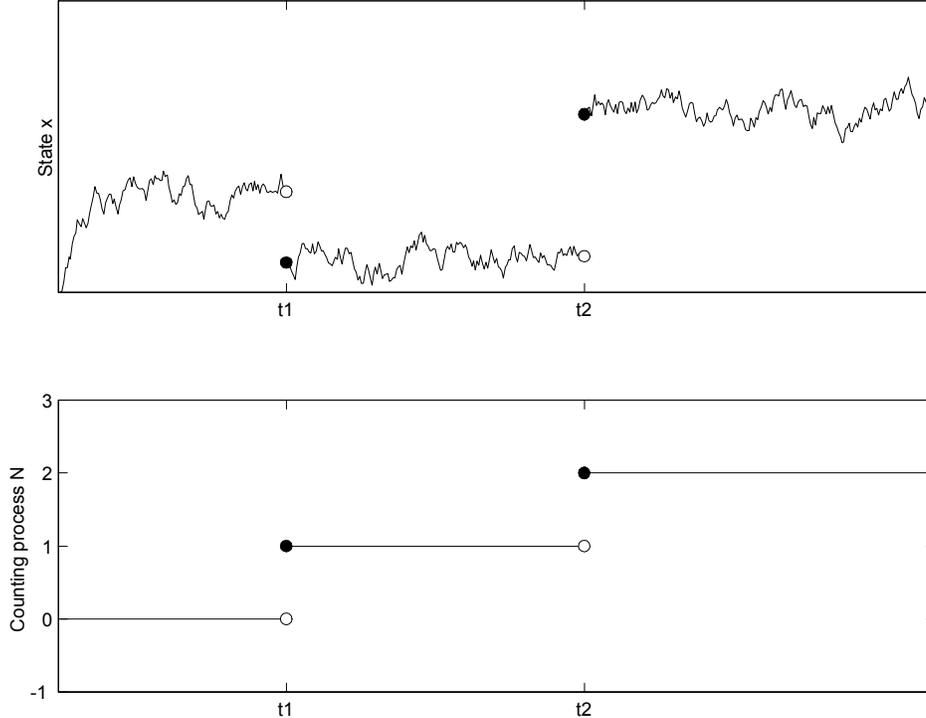


Figure 1: The upper graph illustrates jumps in the state variable  $x$  at jump arrival times  $t_1$  and  $t_2$ . The lower graph illustrates the corresponding counting process  $N$ .

where  $J$  is a pure jump process and the other terms are as before in (6). The jump process  $J$  can be activated in two possible ways. First, jumps may be caused by a Poisson process  $N^P$  with stochastic intensity  $\lambda$  (see Brémaud (1981)). Heuristically,  $\lambda_t dt$  is the conditional probability of a jump in the interval  $[t, t + dt]$ . For tiny intervals, we can therefore intuitively think of a Poisson process as a 0-1 coin flip with conditional probability  $\lambda_t dt$  of observing 1 and probability  $1 - \lambda_t dt$  of observing 0. We may observe more than one jump during longer intervals. Second, jumps may happen at deterministic points in time. These jump times are recorded by a deterministic counting process  $N^D$ . The processes  $N^P$  and  $N^D$  each start at 0 and count up in increments of 1. I use one jump process of each type to save on notation. The extension to multiple jump processes is immediate (it only involves summing up different jump processes in the formulas below).

Affine jump-diffusions make the same functional form assumptions on the coefficient  $\mu_x(x)$  and the volatility  $\sigma_x(x)$  as in the case without jumps. In addition, functional-form assumptions are needed for the jump intensities and the distribution of jump sizes conditional on information “right before” the jump. These assumptions are stated next.

#### Assumption IV.

1. (Stochastic intensity). The stochastic intensity  $\lambda$  of the Poisson process is affine

$$\lambda(x) = \lambda_0 + \lambda_1^\top x$$

for  $\lambda_0 \in \mathbb{R}$  and  $\lambda_1 \in \mathbb{R}^N$ .

2. (Conditional jump distribution). Given a Poisson jump at a stopping time  $t$ , the distribution of the jump size  $\Delta x_t$  is independent of  $x_{t-}$ . Given a deterministic jump at  $t$  counted by the deterministic counting process, the distribution of the jump size  $\Delta x_t$  conditional on  $x_{t-}$  has an exponential-affine Laplace transform. More precisely, for any given  $\alpha \in \mathbb{R}$  and  $\beta \in \mathbb{R}^N$ , there exist coefficients  $a(\alpha) \in \mathbb{R}$  and  $b(\beta) \in \mathbb{R}^N$  such that

$$E_{t-} [\exp(\alpha + \beta \Delta x_t)] = \exp\left(a(\alpha) + b(\beta)^\top x_{t-}\right).$$

*Calendar time does not matter*

Consider first the case without deterministic jump counter  $N^D$ . Let  $M$  be the compensated Poisson process  $dM_t = dN_t^P - \lambda_t dt$ . Intuitively, the compensated Poisson process is a demeaned version of the Poisson process, because we are taking out the conditional mean change  $\lambda_t dt$ . This leaves us with a mean 0 shock process  $dM$ , similar to Brownian shocks  $dz$ . Then we can rewrite

$$dx_t = \bar{\mu}_x(x_{t-}) dt + \sigma_x(x_{t-}) dz_t + \Delta x_t dM_t$$

where the drift of  $x$  is now

$$\begin{aligned} \bar{\mu}_x(x) &= \mu_x(x) + \lambda(x) E[\Delta x] \\ &= \kappa(\bar{x} - x) + (\lambda_0 + \lambda_1^\top x) E[\Delta x]. \end{aligned}$$

The new term in the drift is the expected jump in  $x$ , which is simply the probability  $\lambda(x) dt$  of a jump in the interval  $[t, t+dt]$  times the expected jump size  $E[\Delta x]$  conditional on a jump. The expectation has no subscript because the distribution of the jump size  $\Delta x$  is state-independent by Assumption IV.2. Since  $E[\Delta x]$  is a constant, and  $\mu_x(x)$  and  $\lambda(x)$  are both affine in  $x$ , the drift  $\bar{\mu}_x$  in the case of Poisson jumps is again affine.

Now suppose again that the LEH holds. Ito's Lemma for the case with Poisson jumps (Duffie (2001, Appendix F)) implies that the bond price is itself an Ito process

$$\frac{dF(x_t, \tau)}{F(x_{t-}, \tau)} = \bar{\mu}_F(x_{t-}, \tau) dt + \sigma_F(x_{t-}, \tau) dz + J_F^P(\Delta x_t, \tau) dM_t \quad (23)$$

and the size of the jump in bond returns is

$$J_F^P(\Delta x_t, \tau) = \frac{F(x_t, \tau) - F(x_{t-}, \tau)}{F(x_{t-}, \tau)}.$$

The jump size  $J_F^P$  is a function of the jump  $\Delta x$  in the state vector and the time-to-maturity  $\tau$  of the bond. The instantaneous expected bond return now is

$$\bar{\mu}_F(x, \tau) = \mu_F(x, \tau) + \lambda(x) E [J_F^P(\Delta x, \tau)],$$

where  $\mu_F(x, \tau)$  is the return in the case without jumps given by (8). The additional term reflects that bond returns now also compensate for jumps in the state vector. The compensation is equal to the probability  $\lambda(x) dt$  of a jump in the interval  $[t, t + dt]$  times the expected return  $E [J_F^P(\Delta x, \tau)]$  conditional on a jump. Again, the expectation has no subscript because the distribution of the jump size  $\Delta x$  is state-independent by Assumption IV.2.

The guess for the bond price is again of the exponential-affine form (18). This means that the jump in returns is

$$J_F^P(\Delta x, \tau) = \exp(b(\tau)^\top \Delta x) - 1.$$

The bond-price coefficients solve the ODEs:

$$\begin{aligned} a'(\tau) &= -\delta_0 + b(\tau)^\top \kappa \bar{x} + \frac{1}{2} \sum_{i=1}^N [b(\tau)^\top \Sigma]_i^2 s_{0i} + \lambda_0 E [J_F^P(\Delta x, \tau)] \\ b'(\tau) &= -\delta_1 - \kappa^\top b(\tau) + \frac{1}{2} \sum_{i=1}^N [b(\tau)^\top \Sigma]_i^2 s_{1i} + \lambda_1 E [J_F^P(\Delta x, \tau)] \end{aligned} \quad (24)$$

starting at  $a(\tau) = 0$  and  $b(\tau) = 0$ . When  $\lambda_0 = 0$  and  $\lambda_1 = 0_N$ , these equations collapse to the ODEs for the case without jumps (19). For some special cases, the ODEs can be computed by pencil and paper. Das and Foresi (1996) compute coefficients for two such cases. The first case has jumps in a mean-reverting short rate with constant volatility. The sign of the jump size is chosen by a coin flip and the absolute value of the jump size is exponentially distributed. In the second case, the short rate reverts to a stochastic mean which is a random walk with i.i.d. jumps.

### *Calendar time matters*

Bond yields are nonstationary when there are deterministic jump arrival times counted by  $N^D$ . Thus, calendar time now matters. I therefore change the notation for the bond price in this subsection:  $P_t^{(T)}$  now denotes the price of a bond at time  $t$  for a bond that matures at time  $T$ . The price will be given by  $P_t^{(T)} = F(x, t, T)$ . The guess for the bond price is now

$$F(x, t, \tau) = \exp(a(t, T) + b(t, T)^\top x).$$

The computation of  $a(t, T)$  and  $b(t, T)$  proceeds recursively, starting at the time of maturity with boundary condition  $a(T, T) = 0$  and  $b(T, T) = 0$ . The recursive procedure

applies two main results from Piazzesi (2001, Appendix B). Result 1 computes the coefficients at a deterministic jump time, while result 2 computes the coefficients for the interim period between two deterministic jump times. More concretely, result 1 says that if the bond price at the next deterministic jump date  $t$  is exponential-affine in the state vector  $\exp(\alpha + \beta^\top x_t)$  for coefficients  $\alpha \in \mathbb{R}$  and  $\beta \in \mathbb{R}^N$ , then the price  $P_{t-}^{(T)}$  of a bond “just before” the jump date is of the same form. The proof of this result relies on Assumption IV.2. Result 2 states that if the bond price “just before” the next deterministic jump date  $t_{i+1}$  is exponential-affine  $\exp(\alpha + \beta^\top x_t)$  for some coefficients  $\alpha \in \mathbb{R}$ ,  $\beta \in \mathbb{R}^N$  and  $t = t_{i+1}-$ , then the price during the entire interim period  $[t_i, t_{i+1})$  between two deterministic jump dates is given by  $\exp\left(a(s, t) + b(s, t)^\top x_s\right)$  with coefficients  $a(s, t)$  and  $b(s, t)$  for which  $\widehat{a}(t-s) := a(s, t)$  and  $\widehat{b}(t-s) := b(s, t)$  solve the ODEs (24) with terminal conditions  $\widehat{a}(0) = \alpha$  and  $\widehat{b}(0) = \beta$ . Together, the two results guarantee that for every  $t$ , the price  $P^{(\tau)}$  is exponential affine.

### *Risk adjustment with jumps*

Changes of measure with jumps have generally effects on the jump intensity and jump size distribution. Intuitively, risk-neutral pricing applies under a probability measure under which jumps counted by  $N^P$  tend to occur more often and are on average larger in size once they occur. Jumps at deterministic times counted by  $N^D$  have the same timing under both measures, only their size distribution changes. Technically, the risk-adjustment involves a density  $\xi$  as in the case with diffusions, but now the density may jump as well. The density solves

$$\frac{d\xi_t}{\xi_{t-}} = \sigma_\xi(x_{t-}) dz_t + J_\xi^D(\Delta x_t) dN_t^D + J_\xi^P(\Delta x_t) dM_t,$$

starting at  $\xi_0 = 1$ . For notational simplicity, the jump sizes  $J_\xi^D$  and  $J_\xi^P$  only depend on the jump size  $\Delta x_t$ . The extension to dependence  $J_\xi^D(\Delta x_t, x_{t-})$  and  $J_\xi^P(\Delta x_t, x_{t-})$  on the current state  $x_{t-}$  is immediate. Assumptions on the coefficients (in addition to Novikov) are needed for  $\xi$  to be a strictly positive martingale. First, both jump sizes  $J_\xi^D$  and  $J_\xi^P$  need to be greater than  $-1$  for  $\xi$  to stay positive, because the jump size  $\Delta \xi_t$  at the deterministic jump time  $t$  is given by  $\xi_{t-} J_\xi^D(\Delta x_t)$ . The same argument holds for Poisson jump times. Second, the conditional expected value of the jump size at deterministic jump times must be zero for  $\xi$  to be a martingale. For a deterministic jump time  $t$ , the following equalities show why:

$$\begin{aligned} E_{t-}[\xi_t] &= E_{t-}[\xi_{t-} (1 + J_\xi^D(\Delta x_t))] \\ &= \xi_{t-} (1 + E_{t-}[J_\xi^D(\Delta x_t)]) \\ &= \xi_{t-}. \end{aligned}$$

**Example:** Suppose there is only one deterministic jump time  $t$ . The jump in the state is  $\Delta x_t = \mu + \sigma \varepsilon$  where  $\varepsilon \sim N(0, 1)$  is a random variable known at time  $t$ , and

$J_\xi^D(\Delta x_t) = \exp(-\bar{\sigma}\varepsilon - \frac{1}{2}\bar{\sigma}^2) - 1$  for some constant  $\bar{\sigma}$ . This jump size assumption for  $\xi$  satisfies  $J_\xi^D > -1$ . Also,  $E_{t-}[J_\xi^D(\Delta x_t)] = 0$ . Under the risk-neutral measure  $Q^*$ , the random variable  $\varepsilon$  is distributed  $N(-\bar{\sigma}, 1)$  which implies that  $\Delta x_t \sim N(\mu - \sigma\bar{\sigma}, \sigma^2)$  under  $Q^*$ .

The jump intensity  $\lambda^*$  under the risk-neutral measure is given by

$$\lambda_t^* = \lambda_t E_{t-}(1 + J_\xi^P(\Delta x_t)).$$

which is well defined because  $J_\xi^P > -1$ . Unlike at deterministic jump times, the expected jump size in  $\xi$  does not have to be zero at Poisson jump times.

**Example:** Suppose that the intensity  $\lambda$  is affine under the data-generating measure,  $\lambda(x) = \lambda_0 + \lambda_1^\top x$ . Also, suppose that  $E_{t-}(J_\xi^P(\Delta x_t)) = v$  for some constant  $v > 0$ . Then the jump intensity  $\lambda^*$  under the risk-neutral measure is again affine, but with coefficients  $\lambda_0^* = \lambda_0(1 + v)$  and  $\lambda_1^* = \lambda_1(1 + v)$ .

To see where the form of this intensity comes from, consider  $M^*$  which solves

$$\begin{aligned} dM_t^* &= dN_t^P - \lambda_t^* dt \\ &= dM_t + (\lambda_t - \lambda_t^*) dt \end{aligned}$$

I want to choose  $\lambda^*$  to make  $M^*$  the compensated Poisson process under  $Q^*$  and thus a martingale under  $Q^*$ . For  $M^*$  to be a  $Q^*$ -martingale, the product  $\xi M^*$  needs to be a  $Q$ -martingale. This can be seen from the following sequence of equations:

$$E_t^*[M_s^*] = \frac{E_t[\xi_s M_s^*]}{\xi_t} = \frac{\xi_t M_t^*}{\xi_t} = M_t^*.$$

Using Ito's lemma (see Duffie (2001, Appendix E)), the product can be written as

$$\begin{aligned} d(\xi M^*) &= M_{t-}^* d\xi_t + \xi_{t-} dM_t^* + \Delta\xi_t \Delta M_t^* \\ &= M_{t-}^* d\xi_t + \xi_{t-} dM_t + \xi_{t-} (\lambda_t - \lambda_t^*) dt + \xi_{t-} J_\xi^P(\Delta x_t) dN_t^P \end{aligned}$$

Both  $\xi$  and  $M$  are  $Q$ -martingales, so  $\int M^* d\xi$  and  $\int \xi dM$  are  $Q$ -martingales. Now if  $\lambda^* = \lambda E(1 + J_\xi^P)$ , then the last two terms are

$$\xi_{t-} (\lambda_t - \lambda_t^*) dt + \xi_{t-} J_\xi^P(\Delta x_t) dN_t^P = \xi_{t-} J_\xi^P(\Delta x_t) dM_t,$$

which gives another  $Q$ -martingale.

At deterministic jump times, the risk-neutral jump-size distributions satisfy

$$E_{t-}^*[\Delta x_t] = E_{t-} \left[ \frac{\Delta x_t \xi_t}{\xi_{t-}} \right] = E_{t-} \left[ \frac{\Delta x_t (\xi_{t-} + \Delta \xi_t)}{\xi_{t-}} \right] = E_{t-} [\Delta x_t (1 + J_\xi^D(\Delta x_t))].$$

At Poisson jump times, the risk-neutral distributions satisfy

$$E_{t-}^* [\Delta x_t] = E_{t-} \left[ \frac{\Delta x_t (1 + J_\xi^P (\Delta x_t))}{E_{t-} [1 + J_\xi^P (\Delta x_t)]} \right]$$

Since the expected jump in  $\xi$  at deterministic jump times is zero,  $E_{t-} [J_\xi^D] = 0$ , we can see that the last two equations are very similar.

**Example:** Suppose that the jump size  $\Delta x$  at Poisson jumps, the Poisson intensity  $\lambda$  and  $J_\xi^P$  are all constant. Then the risk-neutral jump size is unchanged  $\Delta x$ , only the jump intensity under the risk-neutral measure is different:  $\lambda^* = \lambda (1 + J_\xi^P)$ .

### 3.6 Negative short rates and jump intensities

Affine models do not constrain the short rate and jump intensities to be positive in general. Assumption I specifies the short rate to be affine in the state  $x$  which itself may take on negative values. Assumption IV.1 specifies jump intensities to be affine in  $x$ . Negative nominal short rates are undesirable, because they lead to arbitrage opportunities in economies with money. Negative intensities (just like negative probabilities) do not make sense by definition. Within affine models, there are only two ways out of this problem. The first way is to only include square-root processes in the state vector. Condition A allows these square-root processes to be positively correlated, but not negatively (see Section 3.1). Thus, the correlation structure in the model is severely restricted. For example, jump intensities of different Poisson processes can then only be positively correlated. But negative correlation in Poisson arrival rates is useful in various contexts. For example, up and down moves in a central bank's policy rate come with conditional probabilities that depend on the business cycle and are therefore negatively correlated.

The second way is to view the affine term structure model as a tool to approximate true bond prices. The true short rate and the true intensity are nonlinear,

$$\begin{aligned} r_t^{true} &= \max \{r_t, 0\} = \max \{ \delta_0 + \delta_1^\top x, 0 \} \\ \lambda_t^{true} &= \max \{\lambda_t, 0\} = \max \{ \lambda_0 + \lambda_1^\top x, 0 \} \end{aligned}$$

while  $r$  and  $\lambda$  are affine in the state  $x$  and enter the (now approximate) pricing equation (2). The approximating model for bond prices ignores the truncation induced by the max-operators and is therefore affine. To be clear, the approximating affine model may still allow arbitrage opportunities. For example, in states of the word where  $r$  takes on negative values, long-term bond yields from the approximating model may be negative as well, again giving rise to arbitrage strategies involving money. The approximation is good provided that the probability  $\Pr \{r < 0\}$  that  $r$  takes on negative values is small. Similarly,  $\Pr \{\lambda < 0\}$  needs to be small for the approximating model to work well. The accuracy of this approximation at a given parameter vector can be checked, for example,

by computing true bond yields based on  $r^{true}$  and  $\lambda^{true}$ . This computation either involves Monte-Carlo methods applied to (2) or numerically solving the PDE (8). I am not aware of any study of nominal bond yields that performs such a check regarding short rates. Some estimated affine models imply negative short rates on average, which suggests that such a check would be useful. Piazzesi (2001) performs this check for negative intensities.

Leaving the affine setup is another alternative. This does not necessarily mean giving up on closed-form solutions for yields. For example, the short rate is quadratic in Constantinides (1992), El Karoui, Myneni, and Viswanathan (1993), Ahn, Dittmar, and Gallant (2002). Both the short rate and jump arrival intensities are quadratic in Piazzesi (2001).

### 3.7 Expected returns

Expected returns in excess of the riskfree rate in affine models can be computed as follows. Ito's lemma implies that the volatility of bond returns is

$$\sigma_F(x_t, \tau) = b(\tau)^\top \sigma_x(x_t).$$

I insert the definition of the Brownian motion  $dz_t^* = dz_t - \sigma_\xi^\top(x_t) dt$  and compensated Poisson process

$$dM_t^* = dM_t - \lambda_t E_{t-}(J_\xi^P) dt$$

under  $Q^*$  into the SDE for the bond price (23). This leads to a CAPM-type equation linking expected bond returns under the two measures,  $\mu_F(x, \tau)$  and  $\mu_F^*(x, \tau) = R(x)$ , between any two deterministic jump dates:

$$\mu_F(x, \tau) - R(x) = -b(\tau)^\top \sigma_x(x) \sigma_\xi(x)^\top - \lambda(x) E[J_\xi^P(\Delta x) J_F^P(\Delta x, \tau)]. \quad (25)$$

Without Poisson jumps ( $\lambda(x) \equiv 0$ ), expected excess returns are determined by their covariance with the density  $\xi$ , which in continuous time is just the product of the volatilities. The volatility of bond returns is the factor loading  $b(\tau)$  times the volatility of the factor  $\sigma_x(x)$ . The volatility  $\sigma_\xi(x)$  of the density contains the *market prices of risk for each Brownian motion*. These risk prices have the usual mean-variance tradeoff interpretation: The  $i$ -th market price of risk measures the percentage change in expected return that compensates an investor for a 1% increase in return volatility attributable to the  $i$ -th Brownian motion. Typically,  $b(\tau)$  contains negative numbers (at least in one-factor models), so that  $b(\tau)^\top \sigma_x(x)$  is negative. Investors thus want more expected excess returns in compensation for holding extra risk, measured as  $-b(\tau)^\top \sigma_x(x)$ . The next section will show that Lucas models with a representative agent imply that  $\xi$  is high in recessions (when aggregate consumption growth is low). A high covariance between bond returns and the density means that bonds pay out in recessions, which makes bonds valuable. Low excess returns are therefore required to compensate the agent to hold the bond.

With Poisson jumps, expected excess returns also compensate for jump risk. The compensation is the probability of a jump times the expected jump in returns  $J_F^P$  weighted with the *market price of Poisson jump risk*  $J_\xi^P$ . Again, since  $b(\tau)$  usually contains negative numbers,  $J_F^P$  is negative.

At deterministic jump times, expected returns under the risk-neutral measure are zero

$$E_{t-}^* [J_F^D(\Delta x_t, \tau)] = 0.$$

To understand why, remember that expected excess returns under the risk-neutral measure are equal to the short rate. In the instant of a jump, the short rate is zero, and expected excess returns under  $Q^*$  must be therefore zero. Intuitively, the instant of a jump is too short for there to be a positive short rate. This implies that expected returns under  $Q$  satisfy an analogous condition to the one for Poisson jumps:

$$E_{t-} [J_F^D(\Delta x_t, \tau)] = -E_{t-} [J_\xi^D(\Delta x_t, x) J_F^D(\Delta x_t, \tau)].$$

Intuitively, expected returns are again the probability of a jump, which is equal to 1 for deterministic jump times, multiplied with the expected return weighted with the *market price of jump risk for deterministic jump arrival times*.

## 4 Affine general equilibrium models

The pricing equations derived so far did not link fundamentals to the yield curve. Moreover, the transition from the data-generating measure  $Q$  to the risk-neutral measure  $Q^*$  was specified exogenously and was not tied to preference parameters. For real bonds, this link to fundamentals can be achieved within a representative agent endowment economy along the lines of Exercise 10.3 in Duffie (2001). Suppose the representative agent has a time-separable utility function

$$U(c) = \int_0^\infty e^{-\delta t} u(c_t, \eta_t) dt \quad \text{with } u(c_t, \eta_t) = \frac{(c_t - \eta_t)^{1-\alpha}}{1-\alpha}$$

where  $\delta$  is the rate of time preference,  $\alpha$  is some power and  $\eta$  is an exogenous preference shock process. The agent eats an endowment process and receives preference shocks such that

$$c_t - \eta_t = \exp(\gamma^\top x_t),$$

where the state vector  $x$  is a diffusion. The coefficient of relative risk aversion

$$-\frac{c_t u_{cc}(c_t, \eta_t)}{u_c(c_t, \eta_t)} = \frac{\alpha c_t}{c_t - \eta_t}$$

is time-varying. In the absence of preference shocks ( $\eta \equiv 0$ ), the coefficient of relative risk aversion is the constant  $\alpha$ , the aggregate endowment is  $c_t = \exp(\gamma^\top x_t)$ , the instantaneous

expected endowment growth rate equals  $\gamma^\top \mu_x(x) + \frac{1}{2} \gamma^\top \sigma_x(x) \sigma_x(x)^\top \gamma$  and the volatility of endowment growth is  $\gamma^\top \sigma_x(x)$ .

The time- $t$  price of a bond paying one unit of consumption at time  $t + \tau$  is the conditional expected value of the marginal rate of substitution between  $t$  and  $t + \tau$ ,

$$P_t^{(\tau)} = E_t \left[ \frac{m_{t+\tau}}{m_t} \right],$$

where marginal utility is given by

$$m_t = \exp(-\delta t) u_c(c_t, \eta_t) = \exp(-\delta t - \alpha \gamma^\top x_t).$$

Ito's Lemma now implies that  $m$  is given by

$$\frac{dm_t}{m_t} = \mu_m(x) dt + \sigma_m(x) dz_t$$

with drift and volatility

$$\begin{aligned} \mu_m(x) &= -\delta - \alpha \gamma^\top \mu_x(x) + \frac{1}{2} \alpha^2 \gamma^\top \sigma_x(x) \sigma_x(x)^\top \gamma, \\ \sigma_m(x) &= -\alpha \gamma^\top \sigma_x(x). \end{aligned} \tag{26}$$

No-arbitrage is a necessary condition for an equilibrium to exist, and from Section 2.1, no-arbitrage is also equivalent to risk-neutral pricing. The marginal utility process  $m$  thus provides the link between the data-generating probability  $Q$  and the risk-neutral probability  $Q^*$ . The following equations hold:

$$P_t^{(\tau)} = E_t \left[ \frac{m_{t+\tau}}{m_t} \right] = E_t \left[ \frac{\xi_{t+\tau}}{\xi_t} \exp \left( - \int_t^{t+\tau} r(u) du \right) \right] = E_t^* \left[ \exp \left( - \int_t^{t+\tau} r(u) du \right) \right],$$

where

$$\xi_t = \frac{m_t}{m_0} \exp \left( \int_0^t r(u) du \right) \tag{27}$$

is the density of  $Q^*$  with respect to  $Q$  (a concept defined in Section 2.4).

Since the process  $\xi$  is a martingale, an application of Ito's Lemma to equation (27) implies that

$$\mu_m(x) = -r = -R(x).$$

Equation (26) therefore describes minus the short rate. The usual comparative statics arguments apply to this short-rate equation, at least in the case without preference

shocks. A higher rate of time preference  $\delta$  makes the agent want to save less, so that the real rate must be higher to compensate the agent for saving as much as before. Higher future expected endowment growth makes the agent want to consume more today. The real rate must therefore be higher to prevent him from borrowing. Higher endowment volatility activates a precautionary savings motive, so that the real rate must be lower to prevent the agent from saving.

The short-rate map  $R(x)$  is affine if the drift  $\mu_x(x)$  and variance-covariance matrix  $\sigma_x(x)\sigma_x(x)^\top$  are affine. In other words, the data-generating process for  $x$  has to be an affine diffusion for Assumption I to be satisfied. Equations (26) and (27) imply that market prices of risk are given by  $\sigma_m(x) = -\alpha\gamma^\top\sigma_x(x)$ . In the absence of preference shocks, market prices of risk are thus given by minus the coefficient of relative risk aversion times the volatility of consumption growth  $\gamma^\top\sigma_x(x)$ . A higher volatility of consumption growth makes recessions worse, and therefore makes bonds which pay out in these bad times even more attractive. The drift of  $x$  under  $Q^*$  satisfies

$$\mu_x^*(x) = \mu_x(x) - \alpha\sigma_x(x)\sigma_x(x)^\top\gamma$$

which is automatically affine.

To summarize, the real yield-curve is affine if the state  $x$  is an affine diffusion under  $Q$  in this model.<sup>6</sup> Campbell (1996) computes bond and stock prices in a discrete-time version of this economy in which consumption growth is a univariate ARMA process of any order. CIR (1981) discuss the specification of higher-order autoregressive processes in continuous time. Bekaert and Grenadier (2000) relax the homoskedasticity assumption on the state vector in a discrete time setting. Campbell (1996) and Bekaert and Grenadier (2000) allow for preference shocks to increase risk premia.<sup>7</sup> Campbell, Lo, and MacKinlay (1997) specify consumption growth as an AR(1) plus noise, which amounts to an ARMA(1,1). This specification differs from the one in Campbell because the number of shocks matters, for example, for determining the spanning number of assets. Wachter (2001) combines an ARMA(1,1) for consumption growth with a ‘surplus ratio’  $(c - \eta)/c = \exp(x_1)$  where  $x_1$  is a square-root process in discrete time as in Campbell and Cochrane (1999). The continuous time analogue of the aggregate endowment in her economy is  $c = \exp(x_2)$ , where  $x_2(t) = x_2(0) + \int_0^t x_3(s) ds + z_2(t)$  and  $x_3$  is a Gaussian autoregressive process. The expected instantaneous endowment growth rate is  $x_3$  plus a constant. In terms of the general specification outlined above, this amounts to choosing  $\gamma^\top x = x_1 + x_2$ . Expected excess returns in the data switch signs over time. Dai (2001) generates this stylized fact with a model of habit formation.<sup>8</sup>

<sup>6</sup>For stock pricing in affine economies, see Bakshi and Chen (1997), Bekaert and Grenadier (2001), Mamaysky (2002), and Longstaff and Piazzesi (2002).

<sup>7</sup>Alternatively, Telmer and Zin (1996) investigate the real term structure in an incomplete (non-affine) setting which also implies higher premia for long-term bonds.

<sup>8</sup>Seppala (2002) generates the fact using an economy with limited commitment. However, equilibrium bond prices in his economy do not have tractable solutions.

Even though there is no role for money in this economy, nominal bonds can still be priced by specifying an *exogenous* price process  $p_t$ . CIR (1985) do this in their Section 7. To be concrete, the dollar-price  $PN_t^{(\tau)}$  of a bond that pays out one dollar at  $\tau$  periods from now is

$$PN_t^{(\tau)} = E_t \left[ \frac{m_{t+\tau}}{m_t} \frac{p_t}{p_{t+\tau}} \right].$$

CIR (1985), Gibbons and Ramaswamy (1993), Pearson and Sun (1994), and Heston (1991) assume that  $m_{t+\tau}$  and  $1/p_{t+\tau}$  are independent. This assumption leads to

$$PN_t^{(\tau)} = P_t^{(\tau)} E_t \left[ \frac{p_t}{p_{t+\tau}} \right],$$

which is the nominal price  $P_t^{(\tau)} p_t$  of a bond that pays one consumption good at time  $t + \tau$  multiplied with how much one dollar at  $t + \tau$  is expected to be worth in terms of the consumption good  $E_t [1/p_{t+\tau}]$ . In this setting, the nominal yield is equal to the real yield plus expected inflation (plus a Jensen's inequality term). This is not true in general

$$PN_t^{(\tau)} = P_t^{(\tau)} E_t \left[ \frac{p_t}{p_{t+\tau}} \right] + cov_t \left( \frac{m_{t+\tau}}{m_t}, \frac{p_t}{p_{t+\tau}} \right)$$

due to the covariance of the pricing kernel and the inverted inflation rate. Pennacchi (1991) and Sun (1992) allow their exogenous inflation process to be correlated with real variables.

The real value of a dollar at  $t + \tau$  can be computed conveniently if the price level  $p$  and expected inflation  $\pi$  are specified as in CIR (1985, Section 7)

$$\begin{aligned} \frac{dp_t}{p_t} &= \pi_t dt + \sigma_p \sqrt{\pi_t} dz_t^p, \\ d\pi_t &= \kappa_\pi (\bar{\pi} - \pi_t) dt + \sigma_\pi \sqrt{\pi_t} dz_t^\pi. \end{aligned}$$

for constants  $\sigma_p, \kappa_\pi, \bar{\pi}, \sigma_\pi$  and independent Brownian motions  $z^p, z^\pi$ . Here, expected inflation is always positive. This specification boils down to evaluating

$$E_t \left[ \frac{p_t}{p_{t+\tau}} \right] = E_t \left[ \exp \left( - \int_t^{t+\tau} \left( 1 - \frac{1}{2} \sigma_p^2 \right) \pi_u du - \int_t^{t+\tau} \sigma_p \sqrt{\pi_u} dz_u^p \right) \right].$$

Since the conditional expected value of the second integral is zero, this expression has a closed form solution. More generally, any price process  $p_t = \exp(\rho^\top x_t)$  would work.

An *endogenous* price process can be derived, for example, in models with a cash-in-advance constraint. The motive for holding money in these models is that good purchases need to be made with money, so that the agent maximizes utility subject to a budget constraint and the cash-in-advance constraint

$$p_t c_t \leq M^d.$$

This constraint binds as long as the nominal interest rate is positive. This is a serious restriction within the class of affine yield-curve models, because many affine specifications allow nominal rates to become negative (Section 3.6 discusses how to maybe deal with this problem). The model also specifies an exogenous money-supply process  $M^s$  and equilibrium requires the money market to clear, so that  $M^d = M^s$ . Together with good-market clearing, the price process in this economy is implied by the quantity equation  $p_t c_t = M^s$ . Rebelo and Xie (1999) and Bakshi and Chen (1996) include money in the utility function. Money lowers the costs of consumption transactions in Wu (2001a). Taxation of nominal capital makes money nonneutral in the money-in-the-utility setup of Buraschi and Jiltsov (2001). Wu (2001b) computes an affine model by linearizing a model with sticky prices.

## 5 Some famous affine models

First generation affine models were based on one of the two basic diffusions.

1. Vasicek-type models:  $x$  is Gaussian.
2. CIR-type models:  $x$  consists of independent square-root processes.
3. Mixture models:  $x$  consists of possibly correlated affine processes.

These early models were 1-factor models. The factor was called ‘short rate.’ The key features of the Vasicek model are:

$$\begin{aligned} R(x) &= x \\ \sigma_x(x) &= \Sigma \\ \sigma_\xi(x) &= q \end{aligned} \tag{28}$$

for constants  $\Sigma$  and  $q$ . Inserting these coefficients into equation (12) shows that the speed of mean reversion  $\kappa = \kappa^*$  in  $x$  (and therefore the short rate) is the same under both probability measures, only the long run mean differs, because  $\bar{x}^* = x - \kappa^{-1}\Sigma q$ . The market price of risk  $q$  is usually estimated to be negative. Intuitively, this means that yields are expected values of average future short rates (apart from a Jensen’s inequality term) which are on average higher  $\bar{r}^* > \bar{r}$  than their historical average. This is therefore an implicit form of risk adjustment.

The CIR model sets

$$\begin{aligned} R(x) &= x \\ \sigma_x(x) &= \Sigma\sqrt{x} \\ \sigma_\xi(x) &= q\sqrt{x} \end{aligned} \tag{29}$$

for constants  $\Sigma$  and  $q$ . Here, the change of measure not only affects the long-run mean but also the speed of mean reversion. A negative  $q$  implies that under the risk-neutral measure  $x$  mean-reverts more slowly ( $\kappa > \kappa^*$ ) to a higher mean ( $\bar{x} > \bar{x}^*$ ). The Vasicek and CIR model share the feature that the state is an affine diffusion under both the risk-neutral and the data-generating probability measure. Vasicek (1977) only contains the one-factor version of the model, which was later extended to the multifactor case by Langetieg (1980). CIR (1985) already contains the multifactor case in Section 6.

Duffie and Kan (1996) paved the way for a second generation of mixture models. Mixture models are built from the two basic building blocks. Duffie and Kan completely characterize the general class of multifactor affine models. To classify these mixture models, Dai and Singleton (2000) count the number  $m$  of processes that enter the volatility  $s(x)$ . More precisely,  $m = \text{rank}(s_1)$  where  $s_1 = [s_{11} \cdots s_{1N}]$ . In their notation,  $A_m(N)$  denotes a model with a total of  $N$  state variables, of which  $m$  enter the volatility. For example, the 1-factor Vasicek model is  $A_0(1)$ , the  $N$ -factor Vasicek model is  $A_0(N)$  and the multifactor CIR model is  $A_N(N)$ . The classification of models does not depend on how the risk-adjustment is specified (because of diffusion invariance).

Factor models need to specify what their factors stand for. Duffie and Kan (1996) propose to explain yields with *latent* factors. This means that the econometrician does not get to observe  $x$  directly, but may be able to infer  $x$  from yields. In other words, the state  $x$  can in this case be thought of as consisting of yields. Most papers with latent factors still try to give their variables intuitive labels. There are two broad types of labels. The first type refers to statistical properties of the short rate, while the second type refers to fundamentals of an underlying general equilibrium model.

## 5.1 Labels based on moments of the short rate

To be able to identify latent variables as moments of the short rate, these models feature one state variable which is called the short rate  $r$ . The linear map  $R(x)$  in these models thus picks just one component of the state vector, say the first, by setting  $\delta_0 = 0$  and  $\delta_1 = [1, 0, \dots]^\top$ . The one-factor Vasicek and CIR models are special cases with  $\delta_1 = 1$ . Multifactor models with this feature have a short rate which is not Markov under the risk-neutral probability measure, so that other variables (in addition to  $r_t$ ) help in forecasting the short rate and thus to compute bond yields.

*Stochastic mean models* take  $x = (r, \theta)$ , where the short rate  $r$  reverts quickly to a time-varying mean  $\theta$ , which reverts slowly to its long-run (unconditional) mean  $\bar{\theta}$ . The relevant SDEs are

$$\begin{aligned} dr_t &= \kappa_r (\theta_t - r_t) dt + \sigma_r dz_t^r \\ d\theta_t &= \kappa_\theta (\bar{\theta} - \theta_t) dt + \sigma_\theta (\theta_t) dz_t^\theta \end{aligned} \tag{30}$$

where  $\kappa_r, \kappa_\theta, \sigma_r$ , and  $\bar{\theta}$  are scalars, with  $\kappa_r > \kappa_\theta$  for  $\theta$  to be interpreted as stochastic mean. The Brownian motions  $z_r$  and  $z_\theta$  are independent. Balduzzi, Das, and Foresi (1998) assume that  $\sigma_\theta(\theta_t)$  does not depend on  $\theta_t$ , which makes the stochastic mean normally distributed. This model is a  $A_0(2)$ -model. Market prices of risk  $\sigma_\xi$  are constant in these models. Chen (1996) assumes that  $\theta$  is a square-root process, so that  $\sigma_\theta(\theta_t) = v\sqrt{\theta_t}$  for some constant  $v$ . This leads to  $z = (z_r, z_\theta)^\top$ , and

$$\begin{aligned}\sigma_x(x) = s(x) &= \begin{pmatrix} \sigma_r & 0 \\ 0 & v\sqrt{\theta} \end{pmatrix}, \\ \sigma_\xi(x) &= q^\top s(x),\end{aligned}$$

which constitutes an  $A_1(2)$ -model for some  $q \in \mathbb{R}^2$ . Here, the matrix  $\Sigma$  in the volatility  $\sigma_x(x) = \Sigma s(x)$  is a  $2 \times 2$  identity matrix  $I$ .

*Stochastic volatility models* take  $x = (r, v)^\top$  where  $v$  is interpreted as the volatility  $v$  of the short rate. To keep volatility positive, it is specified to be a square-root process:

$$\begin{aligned}dr_t &= \kappa_r(\bar{r} - r_t)dt + \sqrt{v_t}dz_t^r \\ dv_t &= \kappa_v(\bar{v} - v_t)dt + \sigma_v\sqrt{v_t}dz_t^v.\end{aligned}\tag{31}$$

for constants  $\kappa_r, \bar{r}, \kappa_v, \bar{v}, \sigma_v$  and independent Brownian motions  $z^r, z^v$ . This leads to  $z = (z^r, z^v)^\top$  and

$$\begin{aligned}\sigma_x(x) = s(x) &= \begin{pmatrix} \sqrt{v} & 0 \\ 0 & \sigma_v\sqrt{v} \end{pmatrix}, \\ \sigma_\xi(x) &= q^\top s(x).\end{aligned}$$

Again,  $q \in \mathbb{R}^2$  and  $\Sigma = I$ . Longstaff and Schwartz (1992) interpret their  $A_1(2)$ -model in this way.

Combinations of these labels can be found in many 3-factor models where the state  $x = (r, \theta, v)$  consists of the short rate, a stochastic mean and stochastic volatility. Examples are the  $A_1(3)$  model of Balduzzi, Das, Foresi, and Sundaram (1996) and the  $A_2(3)$  model of Chen (1996). Dai and Singleton (2000) write down the most flexible  $A_1(3)$  and  $A_2(3)$  models in which all parameters are just identified. Their model has the general form:

$$\begin{aligned}\sigma_x(x) &= \Sigma s(x) \\ \sigma_\xi(x) &= q^\top s(x)\end{aligned}\tag{32}$$

for a constant vector  $q \in \mathbb{R}^N$ . Under this assumption, the state is an *affine diffusion under both measures*. The drift parameters  $\kappa$  and  $\bar{x}$  under  $Q$  are related to those under  $Q^*$  in the following way:

$$\begin{aligned}\kappa &= \kappa^* - \Sigma\Phi \\ \bar{x} &= \kappa^{-1}(\kappa^*\bar{x}^* + \Sigma\psi),\end{aligned}\tag{33}$$

where the  $i$ -th row of  $\Phi \in \mathbb{R}^{N \times N}$  is given by  $q_i s_{1i}^\top$  and the  $i$ -th row of  $\psi \in \mathbb{R}^N$  is given by  $q_i s_{0i}$ . This shows that both the speed of mean reversion  $\kappa$  and the long-run mean  $\bar{x}$  may be different under the data-generating measure than under the risk-neutral measure, where  $\kappa^*$  and  $\bar{x}^*$  are responsible for determining the drift of the process  $x$ .

## 5.2 Labels based on fundamentals

Yield curves in general equilibrium models depend on state variables that have natural interpretations in terms of fundamentals. In principle, the model can then be estimated using observations on both macro variables and yields. This is, however, not what is usually done. The reason is that for ‘reasonable’ coefficients of relative risk aversion, representative agent models can neither match average excess returns on long bonds nor their time series properties when calibrated to aggregate quantities such as consumption. This ‘bond premium puzzle’ is documented for real bonds in Backus, Gregory, and Zin (1989) and Chapman (1997). Den Haan (1995) documents the puzzle for nominal bonds.

When the same models are estimated using asset prices alone, the model implies dynamics for the macro variables that have little to do with their historical behavior. In this sense, labels from fundamentals are often empty labels. For example, Pearson and Sun (1994) use the model of Section 7 in CIR (1985) with exogenously specified ‘expected inflation.’ Their estimation does not use any data on inflation, however, only data on yields. Similarly, ‘consumption growth’ in Buraschi and Jilsov (2001), ‘expected aggregate consumption growth’ in Wachter (2001) and ‘labor income’ in Dai (2001) are labels for latent variables. Sometimes data from outside the bond market are combined with many, often more than 5, yields. The key in these applications is that yields far outnumber the macro series, and Kalman filtering tends to match only moments of yields. In this case, again, the filtered variables usually have little to do with their names.

## 6 Estimation methods for affine models

To estimate affine models, various choices have to be made regarding measurement errors and estimation methods. This section is long, because these choices are not obvious. In thinking about these choices, it is useful to view affine models as *state space systems* with an observation equation which links observable yields to the state vector and a state equation which describes the dynamics of the state:

$$\begin{aligned} y_t^{(\tau)} &= A(\tau) + B(\tau)^\top x_t + \varepsilon_t^{(\tau)} \\ dx_t &= \mu(x_t) dt + \sigma(x_t) dz_t. \end{aligned} \tag{34}$$

The system (19) of ODEs provides the cross-equation restrictions for this system. Empirical applications start with a choice of how to add ‘measurement errors’  $\varepsilon_t^{(\tau)}$ , which I discuss in Section 6.1. I explain moment-based estimation methods in Section 6.3 and

likelihood-based methods in Section 6.2. Identification of parameters is tricky as in any state space model, and is discussed in Section 6.4.

## 6.1 Stochastic singularity

Affine models rely on a low-dimensional state vector to describe what drives the yield curve. Data on  $N$  different yields can therefore be used to back out  $N$  state variables. The  $N$  yields  $y^{(\tau_1)}, \dots, y^{(\tau_N)}$  can be used to invert equations (21) for  $\tau = \tau_1, \dots, \tau_N$  to obtain the model-implied state vector  $x$ . Any additional yield is predicted by the model with an  $R^2$  of 1. The model can therefore be rejected with a single observation on  $y^{(\tau_{N+1})}$ . Put differently, the variance-covariance matrix of  $N + 1$  yields in the model is singular, a feature called stochastic singularity.

Stochastic singularity is a problem, because we have lots of cross-sectional yield data (many different  $\tau$ 's) and want to use models with few state variables. Adding measurement error  $\varepsilon^{(\tau)}$  to the yield equation, as done in (34), breaks this singularity. Now different assumptions can be made on the properties of these measurement errors. Either all of the yields are observed with error or only a subset of yields are observed with error. The variance of the measurement error  $\varepsilon^{(\tau)}$  is nonzero for all  $\tau$  according to the first assumption, while some of the  $\text{var}(\varepsilon^{(\tau)})$  may be zero according to the second assumption.

The assumption that all yields are observed with error seems plausible. Data-entry mistakes and interpolation methods for constructing zero-coupon yields are among the obvious sources for such errors. When all yields have errors, we cannot invert the yield coefficients in (34) to compute the state vector. Kalman filtering is useful here, especially when the state vector is normally distributed (Gong and Remolona (1996), Pennacchi (1991), Campbell and Viceira (2001)), but also in more general setups (Collin-Dufresne, Goldstein, and Jones (2002)).

The alternative assumption is that data on  $N$  yields is flawless where  $N$  also happens to be the number of factors in the model. This assumption is clearly arbitrary. The econometrician is even supposed to know which  $N$  yields in his dataset are flawless. All other yields are observed with error, so that the model cannot be easily rejected. Some estimations include the yields with error in the estimation to exploit all available data (Chen and Scott (1993)). Other applications leave the contaminated yields out and then use them for an out-of-sample check of the model. To be clear, the check is 'out-of-sample' only in the cross-section, because these are yields not included in the estimation, not in the time-series sense (for example, Pearson and Sun (1994), Dai and Singleton (2000), Piazzesi (2001)).

The measurement errors recovered using any of these approaches are typically highly autocorrelated. This autocorrelation may be due to the interpolation method used to construct zero-coupon yields. This does not seem very plausible, however, because one

would expect data construction methods to generate measurement errors that are correlated with each other in the cross section and not necessarily over time. Moreover, swap yields are not interpolated and their measurement errors are still highly autocorrelated (Duffie and Singleton (1997), Collin-Dufresne, Goldstein, and Jones (2002)). Autocorrelation in measurement errors is worrisome, because it suggests that these errors might have in fact nothing to do with measurement issues but with omitted state variables or functional form assumptions. For example, Table IV in Dai and Singleton (2000) computes average measurement errors for their 3-factor affine model over periods with upward or downward sloping swap curve. Their preferred  $A_1$  (3) model makes larger errors when the yield curve is upward sloping. Nonlinearities may account for such a pattern. Model misspecification is not handled by the estimation methods and the computation of standard errors. Much more research is needed in this direction.

## 6.2 Likelihood-based methods

Maximizing the likelihood function relies on being able to compute the density  $f(x_{t+1}|x_t)$  of the state vector  $x_{t+1}$  given  $x_t$ . The conditional density of an  $N$ -dimensional vector of observed yields  $Y$  can be obtained by a change of variable. The density of  $Y$  is the product of the conditional density of  $x$  and the determinant of the Jacobian

$$f_y(Y_{t+1}|Y_t) = f(x_{t+1}|x_t) \left| \frac{dx_{t+1}}{dY_{t+1}} \right|.$$

The log-likelihood function of observed yields  $\{Y_t\}_{t=1}^T$  is then constructed as the usual sum of log densities  $\log f_y(Y_{t+1}|Y_t)$  over the sample. To maximize the log-likelihood, the state  $x_{t+1}$  is backed out from  $Y_{t+1}$  for any given parameter vector. This method works both with linear zero-coupon yields or invertible nonlinear functions of the state  $x$ , such as coupon-bond prices, because the nonlinearity is absorbed by the Jacobian term.

### *Closed form densities*

The density  $f(x_{t+1}|x_t)$  is known in closed form for only a few affine processes. For Gaussian processes,  $f$  is multivariate normal. Zero-coupon yields are affine in  $x$  and therefore also Gaussian. Their likelihood function is therefore particularly easy to compute. To implement the procedure, we only need the conditional expected value (14) and variance (15). For independent Gaussians, the conditional variance is (16).

For independent square-root processes,  $f$  is the product of noncentral chi-square densities. The formula for the densities is based on the modified Bessel function of the first kind of order  $q$  (see CIR (1985), p. 391-2). The command “besseli” computes the function in MATLAB. The conditional mean is given again by (14), but now the conditional variance is (17).

### *Quasi-maximum Likelihood*

For general affine diffusions,  $f$  cannot be computed in closed form. The temptation is then to discretize the SDE and apply maximum likelihood to the density of the discretized process. The discretization assumes that the data is generated from the stochastic difference equation

$$\Delta x_{t+h} = \mu(x_t) h + \sigma(x_t) \varepsilon_{t+h} \sqrt{h}, \quad (35)$$

where  $\varepsilon_{t+h}$  has an  $N$ -dimensional standard normal distribution and  $h$  is the length of the observation interval. The density of the discretized process  $x_{t+h}$  conditional on  $x_t$  is normal with mean  $\mu(x_t) h$  and conditional variance  $\sigma(x_t) \sigma(x_t)^\top h$ . The conditional distribution of the discretization (35) converges to the one of the SDE (6) when  $h$  tends to 0. The estimator that maximizes the likelihood function of the discretization is, however, not consistent for any given  $h$ . Lo (1988) shows this for explicit examples. The reason for this inconsistency is that the discretized process has conditional moments  $\mu(x_t) h$  and  $\sigma(x_t) \sigma(x_t)^\top h$ , while the true process has moments (14) and (15). Quasi-maximum likelihood estimation based on the right moments and a normal density, however, is consistent (Fisher and Gilles (1996)).

#### *Fourier inversion of the characteristic function*

The density  $f$  of affine diffusions can be computed by Fourier inversion of the characteristic function. The characteristic function  $\phi_t(u)$  is defined as the Fourier transform of the density of  $x_{t+1} \in D$  conditional on  $x_t$ ,

$$\phi_t(u) = E_t [\exp(iu^\top x_{t+1})] = \int_D f(x_{t+1}|x_t) \exp(iu^\top x_{t+1}) dx_{t+1}$$

for some  $u \in \mathbb{R}^N$  and the imaginary number  $i = \sqrt{-1}$ . Duffie, Pan, and Singleton (2000) show that the characteristic function  $\phi_t(u)$  can be computed in closed form for affine diffusions. The idea is to apply the Feynman-Kac approach to the conditional expected value

$$\phi_t(u) = \exp\left(\alpha(1) + \beta(1)^\top x_t\right) \quad (36)$$

with coefficients  $\alpha(\tau)$  and  $\beta(\tau)$  that start at  $\alpha(0) = 0$  and  $\beta(0) = iu$  and solve the complex-valued ODEs

$$\begin{aligned} \alpha'(\tau) &= \beta(\tau)^\top \kappa \bar{x} + \frac{1}{2} \sum_{i=1}^N \left[ \beta(\tau)^\top \Sigma \right]_i^2 s_{0i} \\ b'(\tau) &= \kappa^\top \beta(\tau) + \frac{1}{2} \sum_{i=1}^N \left[ \beta(\tau)^\top \Sigma \right]_i^2 s_{1i} \end{aligned}$$

For more details, see Duffie (2001), Appendix H.

Knowing the characteristic function  $\phi_t(u)$  of an affine diffusion means that its conditional density  $f$  can be computed by Fourier inversion

$$f(x_{t+1}|x_t) = \frac{1}{\pi^N} \int_{\mathbb{R}^N} \operatorname{Re} \left\{ \exp(-iu^\top x_{t+1}) \phi_t(u) \right\} du, \quad (37)$$

where  $\text{Re}$  denotes the real part of complex numbers. Maximum likelihood by Fourier inversion has been implemented in the univariate case. Singleton (2001) estimates a one-factor CIR model by maximizing the likelihood function obtained with this method. The conditional density (37) is computed using Gauss-Legendre quadrature. For higher-dimensional state spaces, this computation becomes costly. The number of grid points used for the quadrature grows from  $d$  for  $N = 1$  to  $d^N$  for general  $N$ . The MATLAB command “quad” performs these computations.

For general diffusions, not necessarily affine, the density can be computed by numerically solving a PDE, simulation or Hermite expansions. Of these three methods, only simulation has been applied to the case of many factors so far.

### *Solving the PDE for the density*

The PDE for the conditional density  $f$  is given by the usual forward Kolmogorov equation (see, for example, Lo (1988)). The PDE can be solved numerically. The curse of dimensionality applies here as well, see Lo (1988) and Jensen and Poulsen (1999).

### *Simulated maximum likelihood*

Pedersen (1995) and Santa-Clara (1995) propose to simulate the likelihood function. Simulations of general diffusions cannot be based on their true density  $f$ , which is unknown. Instead, the simulations use the Euler scheme (35). Starting with the observed value  $x_t$  at time  $t$ , the  $s$ -th simulated path of the state vector  $\hat{x}^{x_t}[s]$  is taken from (35) using independent draws  $\hat{\varepsilon}[s]$  from an  $N$ -dimensional standard normal distribution. The MATLAB command “randn” takes these draws. The idea is to write the density of  $x_{t+1}$  conditional on the last observation  $x_t$ , using Bayes’ Rule and the Markov property of  $x$ , as

$$f(x_{t+1}|x_t) = \int_D f(x_{t+1}|x_{t+1-h}) f(x_{t+1-h}|x_t) dx_{t+1-h} \quad (38)$$

for any time interval  $h$ . The density  $f(x_{t+1}|x_{t+1-h})$  is now approximated with the density  $\hat{f}$  of the discretized process (35) for small  $h$ . This density is normal with mean  $x_{t+1-h} + \mu(x_{t+1-h})h$  and standard deviation  $\sigma(x_{t+1-h})\sqrt{h}$ . The integral in (38) can then be computed using Monte Carlo

$$f(x_{t+1}|x_t) \approx \frac{1}{S} \sum_{s=1}^S \hat{f}(x_{t+1}|\hat{x}_{t+1-h}^{x_t}[s]),$$

where the summation is over a total of  $S$  simulated paths of the state that start at the last observation  $x_t$  at time  $t$ . The computer only needs to store the terminal simulated value  $\hat{x}_{t+1-h}^{x_t}[s]$  for each simulation  $s$ , not the entire simulated path. Standard variance reduction techniques, such as antithetic sampling, can be used to improve the efficiency of Monte Carlo integration (for a survey, see Geweke (1996)). Brandt and Santa-Clara (2002) use this simulated maximum likelihood (SML) method to estimate a multifactor

diffusion model. Piazzesi (2001) extends SML to the case of jumps with time-varying jump intensities. Honoré (1998) conducts a Monte Carlo to compare the accuracy of the SML estimator for different discretization intervals  $h$  and numbers of simulations  $S$ . His findings suggest that even coarse discretizations and small simulated samples improve considerably over quasi-maximum likelihood (which amounts to  $h = 1$ ). Durham and Gallant (2002) investigate importance-sampling techniques to improve the accuracy of this method.

### *Hermite expansions*

Aït-Sahalia (2001) approximates  $f(x_{t+1}|x_t)$  for univariate diffusions  $x$  by constructing a standardized version  $\tilde{x}$  of the process  $x$  and then approximating the density of  $\tilde{x}$  by Hermite expansions. The reason for the standardization is that convergence results for Hermite expansions only apply to densities that are ‘close to normal’, not for densities of general diffusions. The standardized version is a diffusion with unity volatility:  $\tilde{x} = \int^x 1/\sigma(w) dw$ . For every  $x \in \mathbb{R}^N$ , Hermite polynomials are given by

$$H_j(x) = \exp(0.5x^2) \frac{\partial^j}{\partial x^j} \exp(-0.5x^2), \quad j = 0, 1, \dots, J$$

For large  $J$ , the conditional density  $f$  of  $x$  can be written in terms of the density  $f_{\tilde{x}}$  of  $\tilde{x}$  which can be approximated with Hermite expansions

$$\begin{aligned} f(x_{t+1}|x_t) &\approx \frac{1}{\sigma(\tilde{x}_{t+1})} f_{\tilde{x}}(\tilde{x}_{t+1}|\tilde{x}_t) \\ &\approx \frac{1}{\sigma(\tilde{x}_{t+1})} \exp(-0.5(\tilde{x}_{t+1} - \tilde{x}_t)^2) \sum_{j=0}^J \eta^{(j)}(\tilde{x}_t) H_j(\tilde{x}_{t+1} - \tilde{x}_t) \end{aligned}$$

with coefficients

$$\eta^{(j)}(\tilde{x}_t) = \frac{1}{j!} E[H_j(\tilde{x}_{t+1} - \tilde{x}_t) | \tilde{x}_t]$$

which are conditional moments of functions of  $\tilde{x}$ . Aït-Sahalia (2002) computes closed-form expressions for these coefficients using Taylor approximations. Jensen and Poulsen (1999) compare the accuracy of Hermite expansions with other methods for the case of a univariate square-root process. Aït-Sahalia (2002) and Aït-Sahalia and Kimmel (2002) extend the method to multivariate diffusions.

## **6.3 Matching moments**

The computation of moments for Hansen (1982)’s GMM depends on whether the yield equation is affine and on whether the data-generating process of the state vector is an affine diffusion. Moments of affine diffusions can be computed in closed form using the characteristic function. But this result is only useful for matching moments of zero-coupon yields, not for nonlinear yield formulas (which arise with coupon bonds and swaps,

for example.) To avoid nonlinear yield formulas, zero-coupon yields can be constructed by interpolating swap data or other ‘nonlinear’ yield data. For non-affine dynamics under the data-generating measure and nonlinear yield formulas, moments can either be simulated using the methods explained in Gallant and Tauchen (2002) or computed using operator-methods explained in Ait-Sahalia, Hansen, and Scheinkman (2002).

Higher order moments of affine diffusions can be conveniently computed from the characteristic function. First and second moments were already computed in Section 3.2. Conditional cross-moments of the  $i$ -th and  $j$ -th component of  $x$  are given by

$$\begin{aligned} E_t (x_{i,t+1}^m x_{j,t+1}^n) &= i^{m+n} \frac{\partial^m}{\partial u_i^m} \frac{\partial^n}{\partial u_j^n} \phi_t(u) \Big|_{u=0} \\ &= i^{m+n} \frac{\partial^m}{\partial u_i^m} \frac{\partial^n}{\partial u_j^n} \exp \left( \alpha(1) + \beta(1)^\top x_t \right) \Big|_{u=0} \end{aligned}$$

for  $1 \leq i, j \leq N$ . This computation is particularly convenient if the coefficients  $\alpha(1)$  and  $\beta(1)$  can be computed with paper and pencil. For an early estimation of a CIR model with GMM, see Gibbons and Ramaswamy (1993).

Another set of moments is computed in Singleton (2001) and Chacko and Viceira (2001). In these papers, the characteristic function is used to set up moment conditions

$$E_t \left[ \exp(iu^\top x_{t+s}) - \phi_t(u) \right] = 0.$$

Each such complex-valued moment condition implies two real-valued moment conditions based on the real and the imaginary part of the expression. Singleton (2001) shows that GMM is efficient in this case when the number of grid points  $u$  goes to infinity. Carrasco, Chernov, Florens, and Ghysels (2001) demonstrate how to actually implement this efficiency result.

## 6.4 Identification

Just like in a state-space system with latent state dynamics, the conditions ensuring identification of parameters are tricky. Ideally, we would not have to care about these conditions, because identification should be an invertibility condition on the information matrix. In practice, this invertibility condition is not useful for checking whether parameters are identified. The likelihood function would have to be maximized for many different trial-parametrization to find out which of the parameters are not identified. Moreover, numerical gradient-methods for computing the information matrix are imprecise, especially when the likelihood function is computed numerically or simulated. The information matrix may turn out to be numerically invertible, even in cases where parameters are not identified. Theoretical results are therefore important. Dai and Singleton (2000) provide such results.

## 7 Empirical evidence on affine models

Empirical studies of term structure models usually pick a set of stylized facts about yields and tailor their model to match these. There is no benchmark for evaluating the performance of different models, because different sets of facts are being matched, and there is no consensus about the relative importance of these facts. I therefore organize the discussion of the empirical findings of affine models around these stylized facts and, in particular, around the moments of yields that are being matched. After discussing data issues in Section 7.1, I focus on factor interpretation in Section 7.2, cross-sectional fitting errors in Section 7.3, unconditional and conditional first moments in Sections 7.4 and 7.5, unconditional and conditional second moments in Section 7.6 and 7.7, higher order moments in Section 7.8, seasonalities in Section 7.9 and zoom back in on fitting the short end in Section 7.10. Then I discuss joint systems of yields with macroeconomic variables in Section 8.

### 7.1 Data issues

The choice of suitable data to estimate yield-curve models needs to balance concerns about measurement errors, sample length, observation frequency, nonlinearities in pricing formulas and even the documentation quality of different data sets. In this chapter, I use monthly Fama CRSP zero-coupon bond tapes from 1964:1 to 2001:12. CRSP provides detailed documentation for this data set. The dataset is problematic, because these yields are interpolated from traded Treasuries which introduces measurement error. Moreover, the 1-month T-bill from the dataset looks strange when compared to the other short rates in the same dataset. For example, the persistence of the 1-month rate decreases after 1985 while the persistence of all other short rates increases. Watson (1999) documents an increase in persistence in the (overnight) Fed funds rate after 1985, which means that the behavior of the 1-month rate does not seem to be a special feature of very short rates. I also omit data from 1952:1 - 1963:12 like Fama and Bliss (1997). The data over this early period behaves much different from the rest in terms of, for example, predictability regressions. Then there are data entry errors: September 1987 shows a 0% yield for the 6-month T-bill in the CRSP file for short maturity T-bills. I therefore interpolate that datapoint. Like ants, errors usually come in company, and this company may be less obvious.<sup>9</sup> Figure 2 plots some of the Fama-Bliss yields used in this chapter.

#### *Short yields*

Short-maturity yields are often used as proxies for the short rate. Seasonality in measurement error is a worry in this context. For example, overnight rates (like fed funds in Hamilton (1996), repo in Piazzesi (2001)), other short-term rates (like term fed funds in Balduzzi, Bertola and Foresi (1996), 7-day Eurodollar in Durham (2001)) and

---

<sup>9</sup>An example for a less obvious data entry error is the Federal Reserve target-rate change that happened on February 4, 1994. Datastream assigns this Fed move to February 3.

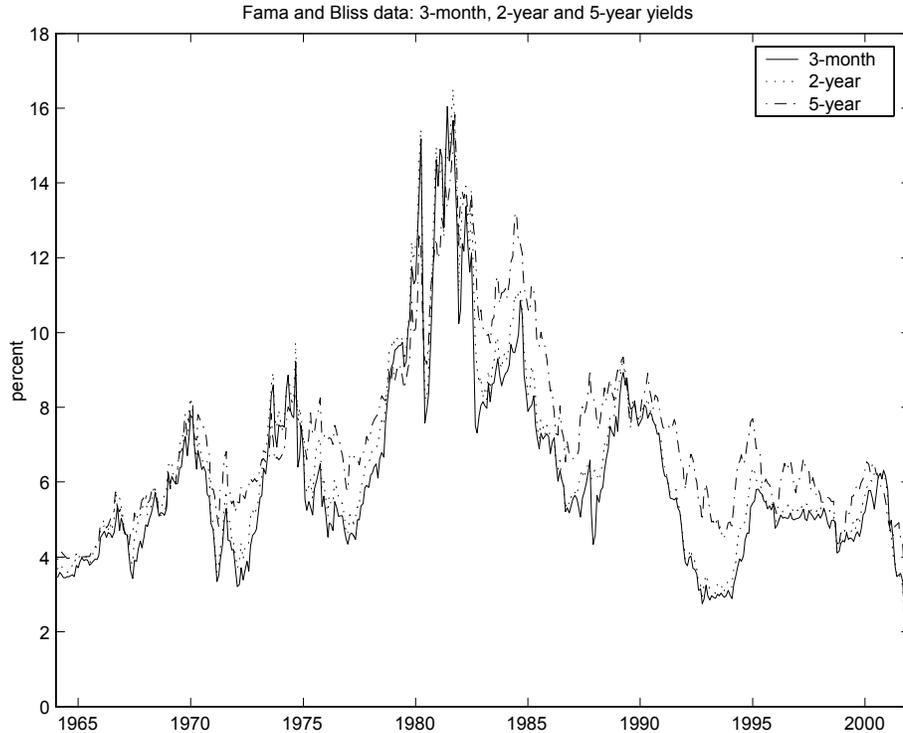


Figure 2: Monthly Fama-Bliss data for 3-month, 2-year and 5-year yields, 1964:1-2001:12.

even yields with maturities of a few months (like the 3-month T-bill in Durham (2001), 6-month LIBOR in Piazzesi (2001)) have been shown to be affected by the two-week reserve maintenance period of banks. These two-week periods start on a Thursday and ending on the so-called ‘settlement Wednesday.’ During this period, banks must hold required reserves in accounts at the Federal Reserve. These reserves are associated with opportunity costs for banks, because the Fed does not pay any interest on these accounts. Until July 30, 1998, the Fed used a contemporaneous reserve maintenance system. In this system, the reserve computation period, the period over which required reserves are actually computed, overlapped with the reserve maintenance period. This overlap implied that the exact amount of reserves that banks were required to hold was not known until the very end of the maintenance period. To avoid the opportunity costs of excess reserve holdings, banks used to hold few reserves until they knew the required amount and then started borrowing on the day before the settlement Wednesday. The increased demands for funds at the end of reserve maintenance periods lead to huge seasonal spikes in interest rates in this contemporaneous reserve system. This seasonality has weakened since 1998, when the Fed adopted a lagged reserve maintenance system. According to the new rules, the reserve computation period ends 30 days before the maintenance period, so that banks know the required amount before they start holding reserves. Settlement Wednesdays and other day-of-the-week effects, such as FOMC meetings, introduce seasonalities in interest rates which may bias, for example, estimates of mean-reversion parameters (more on this in Section 7.9). Piazzesi (2001) and He (2001) therefore argue to use the target rate

set by the Federal Reserve as ‘cleaner’ measure of the short rate. Chapman, Long, and Pearson (1999) argue that short T-bill rates are good short-rate proxies, at least when used in 1-factor affine models. Duffee (1996), however, points out that ‘very short’ T-bill rates behave differently from other short rates. More concretely, T-bills with maturities less than 3 months do not share much variation with other short-term yields such as Eurodollar rates or Fed funds rates.

### *Long yields of zero-coupon bonds*

Zero-coupon bonds have the advantage that the yield equation is easy to invert for  $x$ . True zero-coupon bonds are, however, not easy to come by. These bonds are supposed to be default-free, which may apply to government securities in the United States but certainly not to those in many other countries such as Italy and Spain (Favero, Giavazzi, and Spaventa (1997)), or Russia (Duffee, Pedersen, and Singleton (2002)). U.S. Treasury bills are zero-coupon bonds with maturities up to one year. Duffee (1996) documents that T-bills with maturities less than three months seem to be disconnected from longer term Treasuries. Treasury notes have longer maturities (from 2 to 30 years), but they do pay semiannual coupons. The principal and coupons of these notes can be stripped and traded as separate securities since 1985. This means that data on prices of some long zero-coupon bonds exists (Jordan, Jorgensen, and Kuipers (2000), Grinblatt and Longstaff (2000), Sack (2000)). Various authors have developed interpolation methods to construct long time series of zero-coupon bond yields. Of course, these data-construction methods introduce measurement error. The “McCulloch-Kwon” data until 1991 are available on the website of J. Huston McCulloch at Ohio. Bliss (1999) updates this dataset until the end of 1998. The “Fama-Bliss” dataset is updated each year and available from the Fama CRSP zero-coupon bond tapes. Both datasets consist of monthly observations over the whole postwar period. Estimations of affine models with zero-coupon bonds include Balduzzi, Das, Foresi, and Sundaram (1996), Duffee (2002), Ang and Piazzesi (2002), Brandt and Yaron (2001), Wu (2001a), Wu (2001b), and Buraschi and Jiltsov (2001).

### *Long yields of Treasuries*

The U.S. Treasury interpolates the yields of traded securities when computing constant-maturity Treasury yield data which is released by the Federal Reserve Board in its H.15 release. Daily data since 1962 on these yields are posted on the Federal Reserve’s website (which also has the short-term Treasury bill data). Treasuries pay semiannual coupons. Their yields can be computed as par bond rates  $yc_t^{(\tau)}$  from

$$1 = \sum_{j=1}^{2\tau} P_t^{(0.5j)} \frac{yc_t^{(\tau)}}{2} + P_t^{(\tau)}.$$

Solving this equation for  $yc_t^{(\tau)}$  gives

$$yc_t^{(\tau)} = \frac{2 \left(1 - P_t^{(\tau)}\right)}{\sum_{j=1}^{2\tau} P_t^{(0.5j)}}. \quad (39)$$

The inversion of observed yields for the unobserved state  $x$  can no longer be accomplished by hand with coupon-yields, because the pricing map (39) is nonlinear. Instead, the map needs to be inverted numerically for each observation  $t$  in the sample. The speed of this loop can be increased considerably by supplying the analytical gradient  $\partial y c_t^{(\tau)} / \partial x_t$  to the gradient-based method that inverts the pricing map. For example, Pearson and Sun (1994) use prices of traded Treasuries in their estimation.

### *Long yields for swaps*

Swap rates are truly constant maturity yields which makes interpolation unnecessary. Swaps are agreements to exchange fixed and floating rates semiannually for a time of  $\tau$  years. The  $\tau$ -year swap rate is the fixed coupon rate in this contract, while the floating side is usually specified to be the 3-month or 6-month London Interbank offered rate (LIBOR). Under the assumption that swap rates can be valued as par bond rates, the formula (39) also applies to swap rates. Without default risk, the formula follows from the absence of arbitrage. With default risk, the formula applies if the credit quality in LIBOR and swap markets is the same. The assumption is somewhat problematic because of the institutional features of swap markets. For example, netting features imply that swap rates are minimally affected by credit risk apart from being tied to LIBOR rates (Duffie and Huang (1996), Collin-Dufresne and Solnik (2001)). Swaps have only started trading at the end of the 1980s, which means that swap data are silent about periods of high volatility such as the monetary experiment in the early 1980s in Figure 2. Daily data on both swap rates and LIBOR can be obtained from Datastream, which only supplies poor documentation of this data. Moreover, the data are asynchronous, because LIBOR data are recorded at 11 a.m. London time, while swap data are recorded at the end of the business day in London. Estimations of affine models usually ignore this issue. See, for example, Duffie and Singleton (1997), Dai and Singleton (2000), Piazzesi (2001), He (2001), Collin-Dufresne, Goldstein, and Jones (2002) and Liu, Longstaff, and Mandell (2002).

### *Other data*

Term structure models can, of course, also be estimated with data on futures, caps, floors and other derivative securities. For example, Jegadeesh and Pennacchi (1996) use Eurodollar futures. Attari (2001) looks at options on Eurodollar futures. Jagannathan, Kaplin, and Sun (2001) include data on caps and swaptions. Data on all these contracts can be found in Datastream.

## **7.2 Level, slope, and curvature**

Traditional factor analysis already delivers much of the intuition for what drives yields. Principal components can be computed from levels and changes in yields, I will do both. Suppose the econometrician has data on  $K$  different yields that are contained in the

vector  $Y_t$  at time  $t$ . The variance-covariance matrix of  $Y_t$  can be written as

$$\text{var}(Y_t) = \Omega \Lambda \Omega^\top$$

where  $\Lambda$  is a diagonal matrix of eigenvalues of the matrix  $\text{var}(Y_t)$  and  $\Omega$  is an orthogonal matrix (which means it satisfies  $\Omega^\top = \Omega^{-1}$ ) whose columns are standardized eigenvectors. Principal components  $pc$  are then defined by

$$pc_t = \Omega^\top (Y_t - \bar{Y}), \quad (40)$$

where  $\bar{Y} \in \mathbb{R}^K$  is the sample mean of the yields. The variance of the  $k$ th principal component is just equal to  $\Lambda_k$ , the  $k$ th eigenvalue of  $\text{var}(Y_t)$ . It is also true that the total variation in yields  $\text{tr}(\text{var}(Y_t))$  is equal to the total variation of principal components  $\text{tr}(\Lambda)$ , where  $\text{tr}$  denotes trace. Details can be found, for example, in Mardia, Kent, and Bibby (1979). The same procedure can be repeated for yield changes by replacing  $Y_t$  with  $\Delta Y_t$  and  $\bar{Y}$  with 0 in the above formulas.

Looking at principal components of yield changes reveals that much of the variance in yield changes is explained by the first few principal components. I use  $K = 11$  Fama and Bliss yields from CRSP with maturities 1, 2, 3, 4, 5, 6 months and 1, 2, 3, 4, 5 years for the months 1964:1-2001:12. Table 1 computes the cumulative percentage in the variation of yield changes and levels explained by the first  $k$  principal components. The table shows that for the postwar period the first  $k = 3$  principal components already explain over 96% of the total variation in yield changes. This number is similar for weekly (Litterman and Scheinkman (1991), Chapman and Pearson (2001)) and even daily yields (Hull (2000), Chapter 14.10). In the case of yield levels, the numbers are higher.

To use only  $k \leq K$  principal components, I define the  $K \times k$  matrix  $\tilde{\Omega}$  by

$$\tilde{\Omega}_{i,j} = \begin{cases} \Omega_{i,j} & \text{for } j \leq k \\ 0 & \text{otherwise} \end{cases}$$

and compute the  $k$  principal components of yield levels as

$$pc_t = \tilde{\Omega}^\top (Y_t - \bar{Y}).$$

Table 1: % Variation in yield changes and levels explained by the first  $k$  principal components

$k$	1	2	3	4	5
% explained in $\Delta Y_t$	79.7	91.7	96.5	97.5	98.3
% explained in $Y_t$	96.5	99.5	99.8	99.9	100

NOTE: The total variation in yields is given by  $tr(\Lambda)$ , where  $\Lambda$  is the diagonal matrix of eigenvalues of  $\text{var}(\Delta Y_t) = \Omega\Lambda\Omega^\top$  in the first row and  $\text{var}(Y_t) = \Omega\Lambda\Omega^\top$  in the second row. The numbers in the table are the percentage variation in yield changes (yield levels) explained by the first  $k$  principal components computed as

$$100 \times \frac{\sum_{i=1}^k \Lambda_i}{tr(\Lambda)}$$

The yields are from the Fama tapes of CRSP. The maturities are 1, 2, 3, 4, 5, 6, 12, 24, 36, 48 and 60 months. The sample is 1964:1-2001:12.

The  $k$  principal components are linear combinations of  $K = 11$  yields. Figure 3 plots the coefficients of these linear combinations (or loadings), which are the  $k = 3$  columns of  $\tilde{\Omega}$ , as function of the maturity of the yields in months. Figure 3 looks very similar for the loadings of principal components of yield changes, so I do not include them here. The loadings of the first principal component are horizontal. This pattern means that changes in the first principal component correspond to parallel shifts in the yield curve. This principal component is therefore called the *level factor*. The loading of the second principal component is downward sloping. Changes in the second principal component thus rotate the yield curve. This means the second component is a *slope factor*. The loading of the third principal component is hump shaped. The hump occurs at intermediate maturities. The third principal component therefore affects the curvature of the yield curve, which is why it is called the *curvature factor*. These three principal components can be ordered according to their persistence. The level factor is very persistent with a monthly autocorrelation of 0.98. The slope factor is less persistent with an autocorrelation of 0.92. The curvature factor is the least persistent with an autocorrelation of 0.47.

The interpretation of these principal components in terms of level, slope and curvature goes back to Litterman and Scheinkman (1991). These labels have turned out to be extremely useful in thinking about the driving forces of the yield curve until today. The *latent factors* implied by estimated affine models typically behave like principal components. This empirical finding applies to different sample periods, datasets and model specifications. More concretely, the coefficients from estimated yield equations (21) show the same general patterns as in Figure 3 for the case of  $N = 3$  state variables. This applies to models with only square-root processes as in Chen and Scott (1993), only Gaussian processes as in Gong and Remolona (1996) and Ang and Piazzesi (2002), or mixture models like Balduzzi, Das, Foresi, and Sundaram (1996) and Dai and Singleton (2000). There is no one-to-one mapping between labels such as ‘stochastic mean’ and ‘stochastic volatility’ and the Litterman-Scheinkman labels. For example, ‘stochastic volatility’ behaves like a curvature factor in some estimated models, but it turns out to be so persistent that it becomes the level factor in others. Lower dimensional models

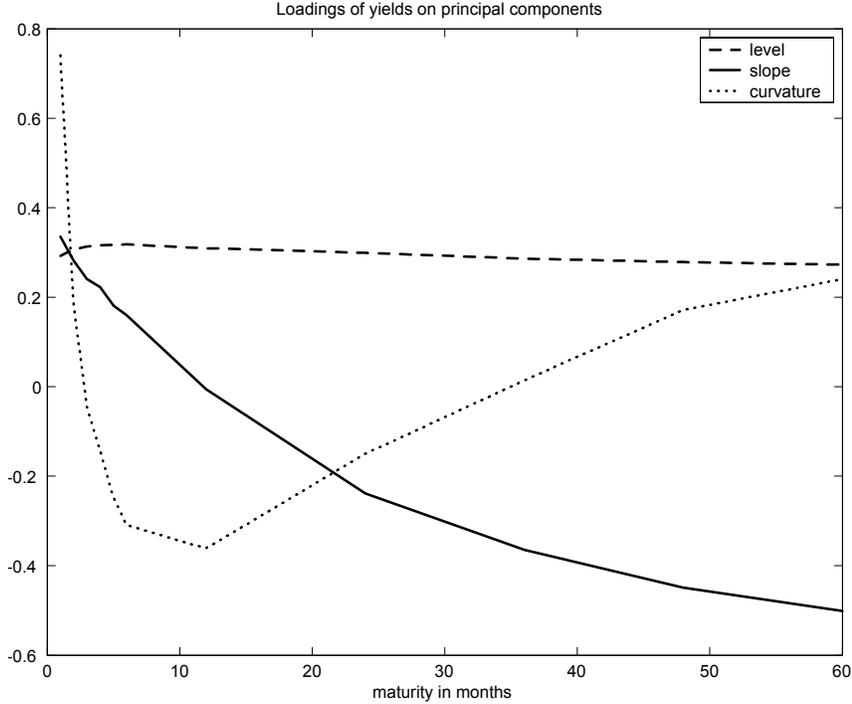


Figure 3: Principal components are linear combinations of yields in the dataset. The coefficients of these linear combinations are the columns of  $\tilde{\Omega}$ . The coefficients of each of the three principal components are plotted as a function of the maturity of the yields.

with  $N < 3$  feature state variables with yield coefficients that correspond to the first  $N$  principal components. In other words, models with only two state variables find a level and a slope factor. Again, this empirical finding is robust across specifications. In particular, it holds for stochastic mean models (30) and stochastic volatility models (31) alike. The square-root case for  $N = 2$  is in Chen and Scott (1993) and the Gaussian case is in Balduzzi, Das, and Foresi (1998). Models with only one state variable (namely the short rate  $r$ ) have one persistent level factor.

### 7.3 Cross-sectional performance

Affine models predict yields of any maturity  $\tau$  with an  $R^2$  of 1. Once we fix a time series of factors, the yields are just linear functions of these factors. Traditional factor models provide a natural benchmark for the cross-sectional fit. Factor models based on  $k$  principal components predict all  $K$  yields in the cross-section as

$$\hat{Y}_t = \bar{Y} + \tilde{\Omega}pc_t, \quad (41)$$

where  $pc_t$  is given by (40). The yield coefficients in this prediction do not impose the cross-equation restrictions from no arbitrage. Unlike in a term structure model, there is no link between the data-generating process of the factors  $pc$  and the way yields depend on

pc. The model implies *fitting errors* for yields which are defined as the difference between actual yields  $Y_t$  and model-predicted yields  $\hat{Y}_t$ . Table 2 computes the mean, standard deviation and maximum of the absolute value of these fitting errors for  $k = 3$  principal components. The absolute fitting errors are less than 11 basis points for all yields in the dataset. This means that this low-dimensional factor model not only explains much of the variance in yields by construction, but also performs extremely well according to this additional metric.

Table 2: Absolute value of fitting errors for yields

maturity	1	3	6	12	24	36	48	60
mean	0.09	0.09	0.06	0.11	0.09	0.05	0.05	0.08
std dev	0.09	0.09	0.06	0.11	0.07	0.05	0.05	0.06
max	0.60	0.80	0.72	1.09	0.50	0.47	0.40	0.48

NOTE: The Table shows the mean, standard deviation, and maximum of absolute fitting errors  $|Y_t - \hat{Y}_t|$  where  $\hat{Y}_t$  is computed as in equation (41) with  $k = 3$  principal components of yield levels. The maturity of yields is in months. The yield data are from the Fama tapes of CRSP for 1964:1-2001:12.

The fitting errors in Table 2 turn out to be hard to beat in practice with an affine model. In other words, the difference between the yields predicted by an affine model at the estimated parameter values and the actual yield data can be substantial. Affine models do not deal with measurement errors explicitly. Such errors are usually tagged onto yields by the econometrician. But they are usually larger in absolute value than those in Table 2. Moreover, these errors are usually highly autocorrelated. Affine models with many Gaussian factors tend to do relatively better in the cross-section than models with many square-root factors. For example, the fitting errors from the  $A_2(3)$  model in Table IV of Dai and Singleton (2000) are larger than the errors from the  $A_1(3)$  model. The fitting errors from the 3-factor jump model without the Gaussian ‘inertia factor’ in Table 2 of Piazzesi (2001) are larger than the errors from the 3-factor jump model with stochastic volatility.

## 7.4 Unconditional first moments (positive slope)

Yields of bonds with longer maturities are on average higher than those of bonds with shorter maturities. This means that the yield curve is on average upward sloping. Figure 4 shows this stylized fact by plotting the sample average of the Fama and Bliss yields as a function of maturity. The solid line of point estimates is shown together with dotted approximate 95% confidence bounds (two times Newey-West standard errors using 6 lags). The plot shows that the shortest yield is significantly lower than the longest yield on average.

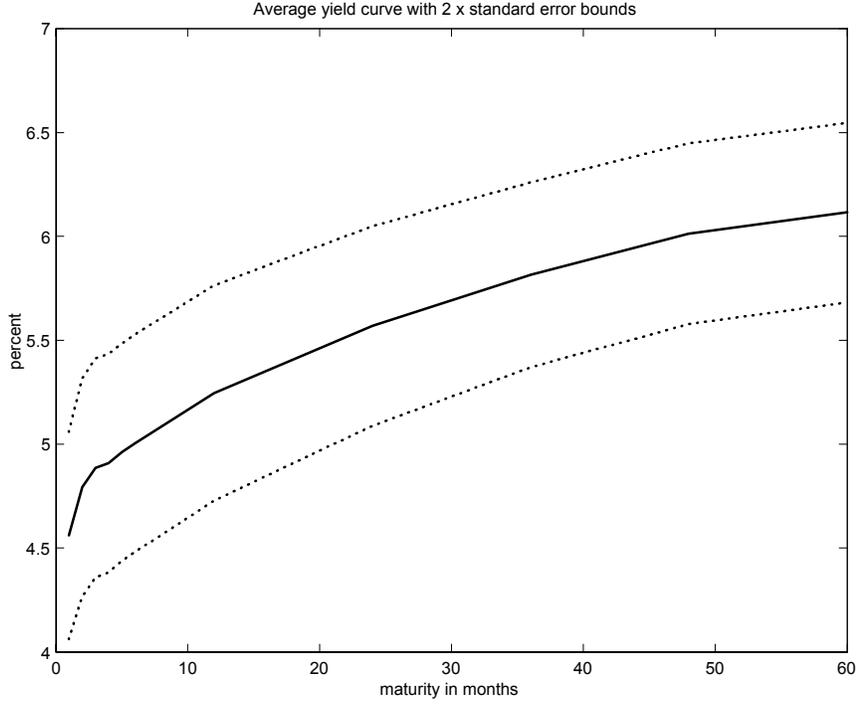


Figure 4: The average yield curve is computed using the Fama and Bliss yields with maturities 1, 2, 3, 4, 5, 6, 12, 24, 36, 48 and 60 months. The sample period is 1964:1-2001:12. The dotted lines are two times Newey-West standard error bounds computed using 6 lags.

An upward sloping yield curve is easy to generate with an affine model. To do this, the risk-neutral long-run mean of the short rate must be higher than its true long-run mean,

$$\bar{r}^* > \bar{r}.$$

From the short-rate equation in assumption I, these parameters are linked to the long-run mean of the state vector  $x$  under  $Q^*$  and  $Q$  as follows:

$$\delta_0 + \delta_1^\top \bar{x}^* > \delta_0 + \delta_1^\top \bar{x}.$$

The parameters  $\bar{x}^*$  and  $\bar{x}$  differ only when market prices of risk differ from zero. For example, constant market prices of risk together with constant factor volatility do the job. From equation (12), this assumption gives:

$$\begin{aligned} \sigma_\xi(x) &= q^\top \\ \sigma_x(x) &= \Sigma \\ \bar{x}^* &= \bar{x} - \kappa^{-1} \Sigma q \end{aligned}$$

When  $R(x) = x$ , the risk-neutral mean  $\bar{x}^*$  is larger than  $\bar{x}$  as long as the market price of risk  $q < 0$ . The CAPM-type equation (25) shows that expected excess returns on bonds

are *constant and positive* in this example. The Vasicek model makes these assumptions (see equation (28)).

## 7.5 Conditional first moments (persistence, predictability, non-linearity)

### *Persistence*

Yields are highly autocorrelated. Table 3 shows the autocorrelations of Fama-Bliss yields together with standard errors around the autocorrelation estimates (which are not corrected for small-sample bias). The monthly autocorrelation coefficient of the 5-year yield implies that shocks to this yield have a half-life of  $\log(0.5)/\log(0.985) \approx 46$  months, almost 4 years. This persistence in yield levels is behind the large standard errors around the mean estimates in Figure 4. Longer yields tend to be more persistent than short yields, at least judging from the point estimates.

There is some evidence that persistence in short rates has increased over time, again according to the point estimates. For example, the autocorrelation of the 3-month yield in the Fama-Bliss file goes from 0.969 before 1985 to 0.993 after 1985. The evidence is not strong, through. Watson (1999) is unable to detect a change in persistence using a Chow test for structural break based on distributions that take into account small sample bias. There is strong evidence that persistence in short rates has increased since the creation of the Federal Reserve in 1914. Mankiw and Miron (1986) document higher predictability of short-rate changes for quarterly data from 1880 to 1914 than after 1914.

Table 3: Autocorrelations of yields

maturity	1	3	6	12	24	36	48	60
autocorrelation	0.954	0.975	0.979	0.978	0.982	0.983	0.983	0.985
standard errors	0.019	0.016	0.014	0.014	0.013	0.013	0.012	0.011

NOTE: The maturity of yields is in months. Standard errors (in brackets) are computed with 6 Newey West lags. The yield data are from the Fama tapes of CRSP for 1964:1-2001:12.

Affine models describe yields as affine in the factors. This implies that persistence of yields must come from persistent factors. In fact, all estimations of affine models find a level factor which is very persistent. This fact was already mentioned in the context of principal component models (Section 7.2).

Persistence contributes to the practical problems associated with the estimation of affine models. Consider the simple example of estimating the parameters of an AR(1),

$x_t = \mu + \rho x_{t-1} + \varepsilon_t$ , with Gaussian errors  $\varepsilon_t$  using maximum likelihood. When the autoregressive coefficient  $\rho$  is close to one, gradient-based optimization methods typically converge very slowly. They take tiny steps around reasonable values of  $\rho$  while still being far from reasonable values of  $\mu$ . The reason is that the likelihood function is essentially flat in  $\mu$  but very steep in  $\rho$ . There is a simple solution to this problem in the univariate case. Fixing a value for  $\rho$  close to one while optimizing over  $\mu$  typically delivers a great starting value for  $\mu$  that can be used in the optimization over both parameters. With multidimensional AR(1) systems, the same strategy can be applied but often becomes more tricky.

### *Predictability of excess returns*

Fama and Bliss (1987) investigate whether excess holding period returns on bonds are predictable using the forward-spot spread. The regressions involve holding periods of one year. For this horizon, the 1-year rate  $y_t^{(1)}$  is the riskless rate. Holding period returns are the difference in log prices  $hpr_{t \rightarrow t+1}^{(\tau)} = p_{t+1}^{(\tau-1)} - p_t^{(\tau)}$ , where  $t \leq t+1 \leq t+\tau$ . Fama and Bliss regress excess holding period returns on the forward-spot spread  $f_t^{(n-1 \rightarrow n)} - y_t^{(1)}$ , where the forward rate is  $f_t^{(n-1 \rightarrow n)} = p_t^{(n-1)} - p_t^{(n)}$ . Table 4 reports the  $R^2$  from these predictability regressions.  $R^2$ s are substantial, 14% for bonds with maturities from 2 to 4 years. Excess returns on 5-year bonds are less predictable, with an  $R^2$  of 6%. Expected excess returns are thus not constant over time.

Cochrane and Piazzesi (2002) show that this  $R^2$  more than doubles when all forward rates  $f_t^{(1 \rightarrow 2)}, f_t^{(2 \rightarrow 3)}, f_t^{(3 \rightarrow 4)}, f_t^{(4 \rightarrow 5)}$  are included on the right hand side of this regression. Moreover, they find that the predictability is mostly due to a single factor. This ‘return-predicting’ factor turns out to be only weakly related to level, slope, and curvature. Also, monthly VARs with 1 lag in the yields with maturities 1-5 years do not find the return-predicting factor.

Table 4:  $R^2$  from predictability regressions

$n$	2	3	4	5
FB (1987)	0.14	0.14	0.15	0.06
CP (2001)	0.34	0.34	0.37	0.34

NOTE: The Table reports  $R^2$  from two predictability regressions. Fama-Bliss (1987) regress  $hpr_{t \rightarrow t+1}^{(n)} - y_t^{(1)}$  on a constant and  $f_t^{(n-1 \rightarrow n)} - y_t^{(1)}$ . Cochrane-Piazzesi (2002) regress the same variable on a constant and  $y_t^{(1)}, f_t^{(1 \rightarrow 2)}, f_t^{(2 \rightarrow 3)}, f_t^{(3 \rightarrow 4)}, f_t^{(4 \rightarrow 5)}$ . The sample is 1964:1-2001:12.

This evidence against the expectations hypothesis is of course based on a rather small sample. Bekaert, Hodrick, and Marshall (1997), Valkanov (2001) and Bekaert and Hodrick (2001) stress the importance of taking into account the small sample distributions

of these  $R^2$ 's and other standard tests of the expectations hypothesis. Tests based on small sample distributions tend to make a weaker case against the expectations hypothesis. Cochrane and Piazzesi (2002) construct 95% confidence intervals for  $R^2$  for the predictability regressions. Confidence intervals for  $R^2$  in the Fama-Bliss regressions indeed include 0. Confidence intervals for  $R^2$  from the regression on all forward rates are far away from 0, however, even if the expectations hypothesis is imposed on the bootstrap. The cross-country evidence about the expectations hypothesis is more mixed. In some countries like Germany, the expectations hypothesis seems to hold up better than in the U.S. (Hardouvelis (1994), Bekaert, Hodrick, and Marshall (1997), Gerlach and Smets (1997), Bekaert, Wei, and Xing (2002)). Finally, the very short end of the yield curve seems to conform better with the expectations hypothesis (Longstaff (2000b)).

Figure 5 shows the fitted values from the predictability regressions in Table 4 for  $n = 3$ . Excess returns on long-term bonds are small and on average positive. For Fama-Bliss yields, average excess returns are 0.42%, 0.65%, 0.79%, 0.72% for  $n = 2, 3, 4, 5$ . This stylized fact can be generated even in the simple 1-factor Vasicek model. Inserting coefficients (28) into the CAPM equation (25) gives expected excess returns  $-b(\tau)^\top \Sigma q$ , where  $b(\tau) > 0$  because the short-rate coefficient  $\delta_1$  is equal to 1. Expected excess returns are positive provided that  $q < 0$ .

However, expected excess returns in the Vasicek model are constant. According to Figure 5 expected excess returns seem to vary through time. In particular, they switch signs over time. In other words, expected excess returns are not always positive, but also sometimes negative. They tend to be positive when the term structure is upward sloping and negative when the term structure is downward sloping. Time-variation alone is easy to generate within an affine model, but many popular affine models are unable to generate the switching signs. The modelling key to this stylized fact is either in the market prices of risk  $\sigma_\xi(x)$  or in the factors loadings  $b(\tau)$ . I will discuss this in detail next.

#### *Affine diffusions under both measures*

Examples of models with time-varying expected excess returns that are not able to switch signs are the CIR and Dai and Singleton (2000) models. From the CAPM equation (25), expected excess returns in the CIR model are  $-b(\tau)^\top \Sigma r_t q$ , where  $b(\tau) > 0$  because  $\delta_1 = 1$ . The time variation in expected excess returns thus comes through the short rate  $r_t$  which is a square-root process and therefore always positive. This implies that expected excess returns are either always positive or always negative - they cannot switch signs. The same mechanism is at work in Dai and Singleton (2000). The coefficients (32) imply expected excess returns  $-b(\tau)^\top \Sigma s(x) s(x)^\top q$ . Condition A allows only square-root processes to enter the volatility  $s(x)$ . Together with the assumption that  $\delta_1 = (1, 0, \dots, 0)$ , the model cannot generate expected excess returns that switch signs.

Recent affine models have attacked this problem in different ways. Backus, Foresi,

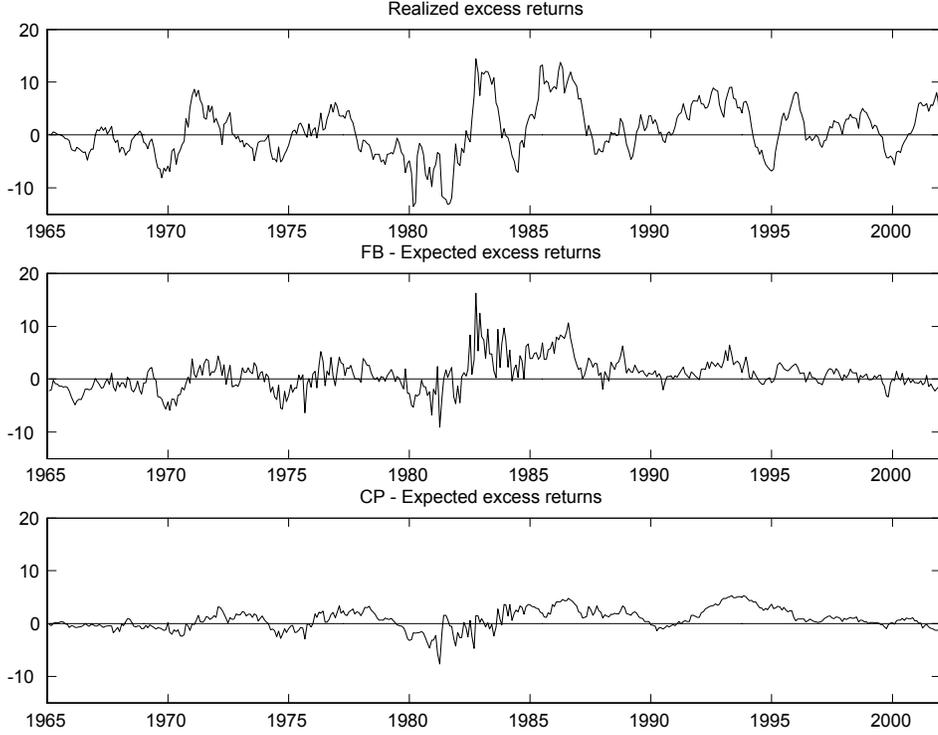


Figure 5: Realized excess returns are  $hpr_t^{(3)} - y_t^{(1)}$ . FB-Expected excess returns are the fitted values of the Fama and Bliss (1987) regression in Table 3 for  $n = 3$ . CP-Expected excess returns are the fitted values of the Cochrane and Piazzesi (2001) regression for  $n = 3$ . Returns expected for  $t$  are plotted together with returns at  $t$ .

Mozumdar, and Wu (1998) make the following assumptions:

$$\begin{aligned}
 R(x) &= x_1 - x_2 \\
 \sigma_x(x) &= \begin{pmatrix} \sqrt{x_1} & 0 \\ 0 & \sqrt{x_2} \end{pmatrix} \\
 \sigma_\xi(x) &= q^\top \sigma_x(x)
 \end{aligned}$$

Both factors  $x_1$  and  $x_2$  are square-root processes, and the short rate is the difference between the two. The CAPM-equation (25) shows that the coefficients  $b(\tau)$  generate the switch in expected excess returns in this ‘negative CIR model.’ This can be seen from the fact that yield coefficients  $B(\tau) = -b(\tau)/\tau$  go to  $\begin{bmatrix} 1 & -1 \end{bmatrix}$  as maturity  $\tau$  goes to 0.

Market prices of risk had so far always the form  $\sigma_\xi(x) = q^\top \sigma_x(x)$ . A number of papers obtain switching signs in expected excess returns with other functional forms for  $\sigma_\xi(x)$ . El Karoui, Myneni, and Viswanathan (1992) and Ahn, Dittmar, and Gallant (2002) propose a setting with Gaussian  $x$ , affine market prices of risk and a linear-quadratic short rate  $r$ . When the short rate is constrained to be affine in  $x$  (no quadratic terms),

this framework collapses to the affine model<sup>10</sup>

$$\begin{aligned} R(x) &= \delta_0 + \delta_1^\top x \\ \sigma_x(x) &= \Sigma \\ \sigma_\xi(x) &= q_0^\top + x^\top q_1 \end{aligned}$$

where  $q_0 \in \mathbb{R}^N$  and  $q_1 \in \mathbb{R}^{N \times N}$ . This is the model considered in Fisher (1998) and Dai and Singleton (2002). Ang and Piazzesi (2002) specify a discrete-time version of the model. In this setup, again, the process  $x$  is Gaussian under both measures as in Vasicek (1977), but now both the long-run mean and the speed of mean reversion differ under the two measures (unlike in the Vasicek model before, where only the long-run mean was different). In particular, expected excess returns switch signs because of  $q_1$ . The setup is a special case of Duffee (2002) who considers

$$\begin{aligned} \sigma_x(x) &= \Sigma s(x) \\ \sigma_\xi(x) &= q_1^\top s(x) + x^\top q_2 s(x)^- \end{aligned}$$

with  $q_1 \in \mathbb{R}^N$ ,  $q_2 \in \mathbb{R}^{N \times N}$  and

$$s_{ii}(x_t)^- = \begin{cases} (s_{0i} + s_{1i}^\top x_t)^{-1/2} & \text{if } \inf (s_{0i} + s_{1i}^\top x_t) > 0; \\ 0 & \text{otherwise.} \end{cases}$$

This definition ensures that  $s(x)^-$  does not explode as diagonal elements in  $s(x)$  go to zero. Suppose  $x$  is Gaussian so that  $s_{0i} = 1$  and  $s_{1i} = 0$ , then  $\sigma_\xi(x)$  is affine in  $x$ .

#### *Risk-neutral affine diffusions with nonlinear data-generating process*

In all setups considered so far,  $x$  is an affine diffusion under both probability measures  $Q$  and  $Q^*$ . As noted before, there is no reason for  $x$  to be an affine diffusion under  $Q$ , at least not for pricing bonds. Duarte (2000) considers this case by specifying

$$\begin{aligned} \sigma_\xi(x) &= q_0 + q_1^\top s(x) \\ \sigma_x(x) &= s(x) = \text{diag}(\sqrt{x}) \end{aligned}$$

$q_0 \in \mathbb{R}$  and  $q_1 \in \mathbb{R}^N$ , and a vector  $x$  of square-root processes under  $Q^*$ . The data-generating process for  $x$  is no longer affine, because the drift under  $Q$  depends on  $\sqrt{x}$  and  $x$ . Again, switching signs in expected returns are generated through the matrix  $q_1$ .

#### *More on nonlinearities*

The statistical discussion about nonlinearities in the term structure literature has much to do with the macroeconomic discussion about the policy rule of the Federal

---

<sup>10</sup>Constantinides (1992) is an early solution to the problem of switching signs within a quadratic term structure model. In his model, the short rate is quadratic in Gaussian state variables  $x$ . The model does not collapse to an affine model because the drift of  $r = x^2$  depends on both  $\sqrt{r}$  and  $r$ .

Reserve. I postpone policy rules to Section 8.1 and regime switching models to Section 7.8, and discuss the purely statistical evidence here. Aït-Sahalia (1996), Stanton (1997) and Conley, Hansen, Luttmer, and Scheinkman (1997) present considerable evidence against affine conditional first moments of short-term interest rates for univariate settings. Ghysels and Ng (1998) and Boudoukh, Richardson, Stanton, and Whitelaw (1999) also find such evidence for 2-factor models. The main empirical pattern found in these papers is that the speed of mean reversion seems to be higher when the short rate is far away from its mean. The pattern may, however, be due to small sample biases. Chapman and Pearson (2000) argue that empirical evidence about what happens in the tails of the distribution, far away from the mean, is necessarily based on few data points. Moreover, they simulate short-rate data under the null of an affine conditional mean and find nonlinearities in the mean using the nonparametric estimators of Aït-Sahalia (1996) and Stanton (1997). Some of these observations may also be due to seasonal measurement error. Large spikes in short rates occur on certain calendar days, such as at the end of calendar years. The verdict is still out on whether there are nonlinearities in conditional means and, more importantly, whether they matter for long-term bond yields.

## 7.6 Unconditional second moments (vol curve)

The volatility curve or ‘vol curve’ is the standard deviation of yield changes  $\Delta Y_t$ . Figure 6 plots the volatility curve for the Fama-Bliss data during the Greenspan era (1987:8 - now). During this time period, the curve is ‘snake-shaped’: high for short maturities (<6 months), low at 6-months, then increasing with a peak at 2-3 years, and then again decreasing. The ‘back’ of the snake, or hump, in volatility around 2-3 years can also be found in swap data or Treasury yield data over this period. The ‘head’ of the snake in Figure 6 comes from the 1 and 2-month yields from the Fama tapes which may not be reliable. Data on the federal funds rate, short term repo and LIBOR rates, however, confirm the overall picture. The ‘snake’ is documented in Piazzesi (2001). Over different time periods, the volatility curve looks different. While the volatility of short-term interest rates is always high, the hump at 2-3 years disappears, for example, during the monetary experiment of the early 1980s.

The modelling key to the ‘back’ of the snake, the hump in volatility, is correlation between factors. For example, the multifactor CIR model based on independent square-root processes is unable to generate the hump in volatility. A stochastic mean model can generate the hump. Intuitively, the shocks to the stochastic mean do not affect the short rate directly, and therefore only affect the volatility of longer-term yields. To match the hump quantitatively, negative correlation between the Brownian motion of the short-rate and the stochastic mean seems to be needed. Dai and Singleton (2000) document this finding within their 3-factor setup. More evidence on the importance of negative correlation between factors is given in Duffie and Singleton (1997). In their two-factor CIR model, the two model-implied factors turn out to be negatively correlated. The correlation coefficient of the factors is  $-0.5$  when they are inverted from yield data. The correlation is thus far from zero, which is what the theoretical model assumes.

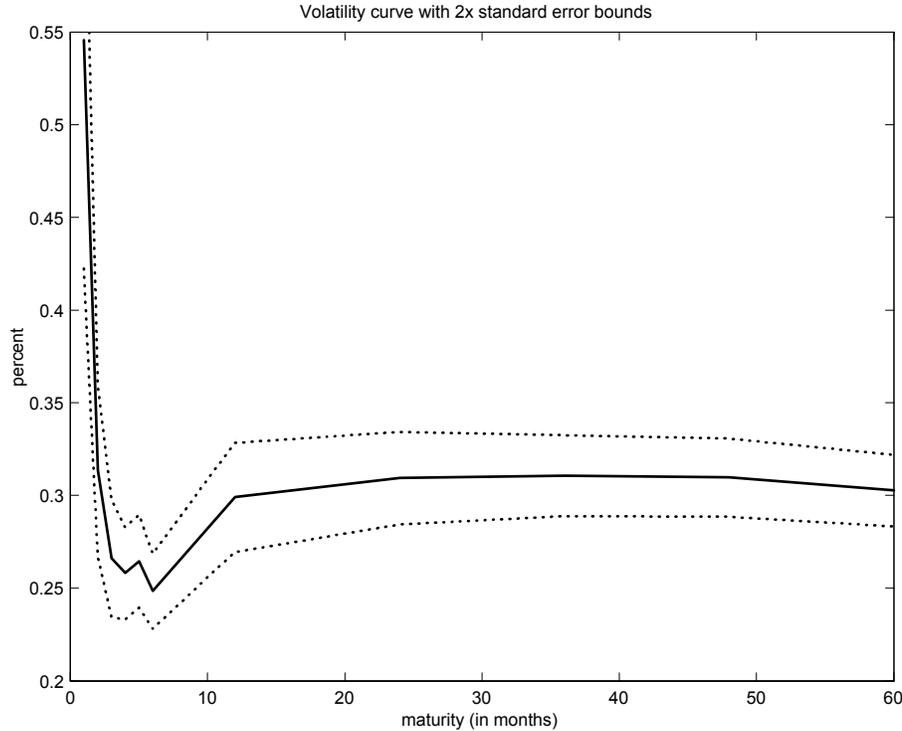


Figure 6: The volatility curve is the standard deviation of yield changes. Yields are from the Fama tapes, 1964:1-2001:12.

The ‘back’ of the snake can be linked to *policy inertia* by the Federal Reserve, defined as positive autocorrelation in target-rate changes. This positive autocorrelation is induced by the Fed’s tendency to moves its policy rate, the Fed funds target rate, in a series of small steps. Piazzesi (2001) builds an affine model with interest-rate targeting by the Fed in which policy inertia generates the ‘back’ of the snake for the period after 1994. The different look of the snake over different subperiods may be explained by the varying degree of policy inertia under different Fed chairmen. The ‘head’ of the snake is money market noise: short-lived deviations of the short rate from the target rate.

## 7.7 Conditional second moments (stochastic vol)

To gather some evidence about conditional second moments of yields, I estimate a vector autoregression of all 11 Fama-Bliss yields (with maturities of 1-6 months and 1-5 years) and compute the squared residuals from this VAR. Figure 7 plots the time series of these squared VAR-residuals for the 1-year yield (together with the 1-year yield itself). Several stylized facts about volatility become clear from this Figure. First, volatility varies over time. In particular, time-varying volatility is really about two episodes: the oil price shock in 1974 and the monetary experiment in 1979-1982. Any volatility study therefore has to decide first on how to treat these two episodes. The choice already starts with

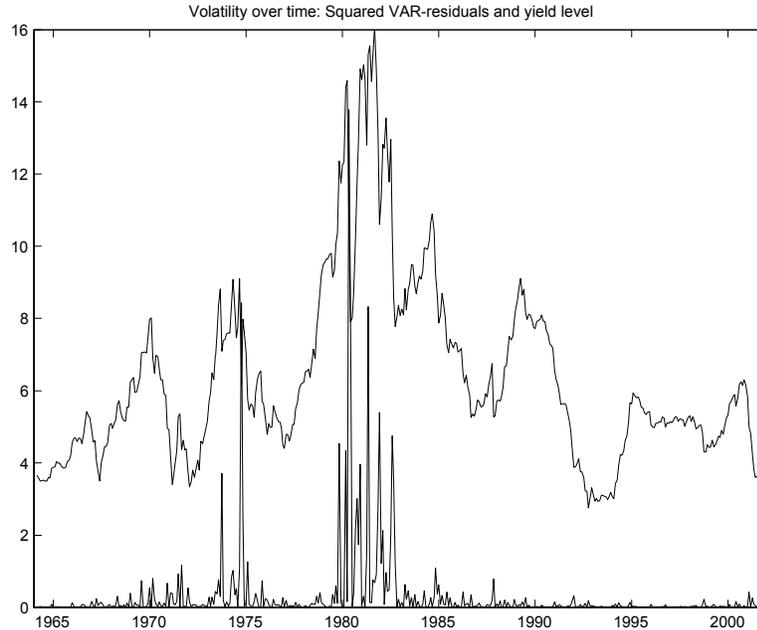


Figure 7: The Figure plots the 1-year Fama-Bliss yield together with the squared residuals for the 1-year yield estimated from a VAR with all 11 Fama-Bliss yields (with maturities 1-6 months and 1-5 years). The squared residuals are on the bottom of the graph.

the dataset. For example, studies with swap yields are completely silent about these episodes because swaps only started trading at the end of the 1980s. Stationary  $A_0(N)$ -models of the Vasicek-type are obviously unable to match the volatility experience of the entire sample, but Figure 7 suggests that a model with 2 regimes - high volatility and low volatility - may be enough. Another possibility is that the world is not stationary, and these ‘regimes’ are really structural breaks. To just describe the experience of the most recent years, constant volatility models may then be enough. More evidence on this ‘return to normality’ is given in the next section in terms of higher order moments.

Second, volatility is positively correlated with the level of interest rates. This becomes clear from eyeballing the years 1974 and 1979-1982 in Figure 7. More precise evidence is given in the ‘slope’-row of Table 5 which computes the slope coefficient of regressing the VAR-squared residuals of any given yield on the level of the same yield. The slope coefficient is positive for all squared residuals and significant for most (at least judging from OLS-standard errors which are not adjusted for the 2-step estimation procedure). The positive correlation between volatility and yield levels motivated the square-root specification for the short rate in the CIR model and later multifactor models such as the model by Longstaff and Schwartz (1992) or Chen (1996), which feature volatility as one of their factors. The residuals from the regression of the squared residuals in Figure 7 on all Fama-Bliss yields still show spikes in 1974 and 1979-82. In other words, yield levels only explain some of the time-variation in volatility. This can also be seen from the  $R^2$  in Table 5, which range from 10 to 15%.

Third, volatility is autocorrelated. Table 5 computes the autocorrelation of the squared VAR-residuals for all maturities. The autocorrelation is positive and significant for all maturities but the 5-year yield (again, standard errors ignore the VAR-step of the estimation).

Table 5: Properties of squared VAR-residuals

maturity	1	3	6	12	24	36	48	60
Regression of squared residuals on yield level								
slope	0.17	0.15	0.15	0.13	0.09	0.07	0.05	0.05
standard errors	0.03	0.02	0.02	0.02	0.01	0.01	0.01	0.01
$R^2$	0.10	0.12	0.15	0.14	0.12	0.11	0.11	0.11
Autocorrelations of squared residuals								
autocorrelation	0.27	0.37	0.40	0.27	0.25	0.15	0.16	0.12
standard errors	0.05	0.10	0.12	0.10	0.10	0.07	0.07	0.07

NOTE: Squared residuals come from a VAR with 1 lag using all 11 Fama-Bliss yields (1-6 months, 1-5 years). ‘slope’ is the slope coefficient from the regression of squared VAR-residuals for yield  $i$  on a constant and the level of yield  $i$ . Standard errors for the slope coefficients are standard OLS standard errors. ‘ $R^2$ ’ is the  $R^2$  from this regression. ‘autocorrelation’ is the first-order autocorrelation of squared VAR-residuals together with Newey-West standard errors computed with 6 lags. All these standard errors ignore sampling noise from the VAR. The yield data are from the Fama tapes of CRSP for 1964:1-2001:12.

Stochastic volatility is a feature of the data that standard affine models may have problems to match. The problem arises because volatility plays two roles in affine models. One role is to match the time-series properties of the short rate. Volatility can in fact be computed from the second moment of some short-rate proxy. For example, Chan, Karolyi, Longstaff, and Sanders (1992) do this. Another role of volatility is to match the cross-section of yields. Volatility can be inverted from affine yield equations. Brown and Dybvig (1986) and Brown and Schaefer (1994) take this route. There is a natural tension between these two roles in any affine model, and panel data studies of affine models need to deal with this tension.

In practice, the tension arises when the inversion of affine yield-equations leads to negative values for volatility. These negative values make some estimation procedures choke, such as maximum-likelihood, and tricks have to be used to avoid them. For example, Duffie and Singleton (1997) find that it helps to add a constant to the short rate equation in a 2-factor CIR model. Other estimation methods are more robust to negative volatility. Efficient method of moments is an example, because volatility is

only simulated to evaluate the EMM objective function, not inverted from yields. The problem remains, however, of how to interpret parameter estimates that lead to negative model-implied volatility in-sample.

More evidence on this tension is documented by Collin-Dufresne, Goldstein, and Jones (2002) who estimate  $A_1$  (3) models with swap data. Ideally, the stochastic volatility factor in these models should be highly correlated with conditional second moments of the short rate implied by the model. Collin-Dufresne et al. proxy the conditional second moment with a Garch model estimated with the time series of the model-implied short rate. Garch-volatility turns out to be *negatively* correlated with the volatility factor from the affine model.

A way out of this tension is to construct models in which bond markets are incomplete in a way that volatility cannot be inverted from the cross section of bond yields. This is the idea behind the ‘string models’ proposed by Santa-Clara and Sornette (2001), ‘random field models’ by Goldstein (2000) and Kimmel (2001) and ‘unspanned volatility models’ by Collin-Dufresne and Goldstein (2001). Useful results on how affine models are related to such more flexible models of volatility are in De Jong and Santa-Clara (1999) and Collin-Dufresne and Goldstein (2001). Empirical support for these models is given in Longstaff, Santa-Clara, and Schwartz (2001) who compare option prices computed from standard affine models with those from string models and find that affine models underprice options. Also, Collin-Dufresne and Goldstein (2001) regress returns on straddles on swap yield changes for different countries and find low  $R^2$ s, while affine models would predict  $R^2$ s of 1. Straddles are portfolios of caps and floors which are particularly sensitive to volatility.

When volatility is not invertible from the cross section of bond yields, it is truly latent when only information about bond yields is used in the estimation. Collin-Dufresne, Goldstein, and Jones (2002) therefore use Kalman filtering techniques to estimate their model. Brandt and Santa-Clara (1999) make volatility an observable factor by using data on at-the-money options.

## 7.8 Higher order moments (jumps and regimes)

Yields are not normally distributed over the sample 1964:1 - 2001:12. If they were, then yield changes would have to be normally distributed as well. Table 6 computes the skewness and kurtosis of yield changes. Benchmark normal distributions are symmetric around the mean, so that their skewness is 0. Their kurtosis is 3, anything beyond that is called excess kurtosis. The distribution of yield changes shows negative skewness. This means that the distribution of yield changes is skewed to the left. (The distribution has a long thin left tail, while most of the probability mass is around and above the mean.) The evidence for skewness is weak, however, because GMM-standard errors around the esti-

mates are large. The distribution of yield changes shows clear excess kurtosis, however, which means that its tails are heavier compared to the normal distribution.

Recently, yields seem to have become ‘more Gaussian.’ Table 6 also computes the skewness and kurtosis of yield changes for the subsample 1990:1-2001:12. The difference is striking. There seems to be only weak evidence against normality during the last 10 years. The squared residuals in Figure 7 during this period are, in fact, barely visible to the eye. Have yields ‘returned to normal’?

Table 6: Higher order moments of yield changes

maturity	1	3	6	12	24	36	48	60
Skewness								
full sample	-1.00	-1.35	-1.51	-1.01	-0.68	-0.11	-0.16	-0.26
std err	(0.80)	(0.61)	(0.80)	(0.69)	(0.57)	(0.39)	(0.31)	(0.30)
1990s	-0.18	-0.88	-0.60	-0.05	0.05	0.19	0.19	0.13
std err	(0.39)	(0.32)	(0.29)	(0.26)	(0.18)	(0.16)	(0.16)	(0.17)
Kurtosis								
full sample	14.34	13.92	16.93	15.38	11.90	9.16	7.07	6.84
std err	(4.48)	(3.10)	(5.24)	(4.78)	(3.15)	(2.71)	(1.22)	(1.25)
1990s	5.31	5.47	4.56	3.59	2.73	2.64	2.68	2.72
std err	(0.72)	(1.08)	(0.74)	(0.37)	(0.24)	(0.23)	(0.26)	(0.26)

NOTE: Skewness is  $m_3/m_2^{3/2}$  and kurtosis is  $m_4/m_2^2$ , where  $m_i$  is the  $i$ -th central moment of yield changes. ‘Full sample’ compute the statistic over the entire sample 1964:1-2001:12, while ‘1990s’ uses the subsample 1990:1-2001:12. Standard errors (in brackets) are computed using GMM with 6 Newey West lags. The yield data are from the Fama tapes of CRSP for 1964:1-2001:12. Maturities are in months.

Affine models offer two ways to deviate from normality: stochastic volatility and jumps. The two ways are difficult to distinguish just from first and second moments. Das and Sundaram (1997) try to distinguish the two with higher order moments. Das (2002) and Johannes (2001) estimate short-rate models with jumps. The nonparametric method by Johannes implies jump arrival rates in-sample that place high probability on jumps occurring at scheduled macroeconomic news releases (which are discussed in the next subsection).

Regime switching models also generate non-normal distributions and are consistent with evidence on nonlinearities in conditional first moments. In general, these models are outside the affine class (Hamilton (1994, Chapter 22), Gray (1996), Ang and Bekaert (1998), Veronesi and Yared (2000), Bansal and Zhou (2002)). For some special cases,

affine solutions can be still be obtained (Landen (2000), Dai and Singleton (2003), Ang and Bekaert (2003)). Estimated regime switching models tend to find two regimes: a high-persistence low-volatility regime, and a low-persistence high-volatility regime. This finding is intuitive from Figure 7 which suggests that we are in the high-persistence low-volatility regime most of the time, with the exception of the oil price shock and the monetary policy experiment.

## **7.9 Seasonalities (settlement Wednesdays & macro releases)**

Fleming and Remolona (1997), Furfine (2001) and Johannes (2001) go back to see whether the largest yield-movements over a given time period coincide with certain events. Fleming and Remolona (1999) and Furfine (2001) use 5-minute price changes in the 5-year Treasury note, while Johannes (2001) uses daily data on the 3-month T-Bill rate. Table 7 summarizes the findings of these three studies. The results show that most large yield-movements happen around employment releases and Federal Reserve policy rate moves. The bulk of these events are scheduled announcements. Only few happen at random times, such as the outbreak of the Gulf war which coincides with one of the 10 largest yield movements in the Johannes sample. More evidence on seasonalities around macroeconomic news releases is documented in Jones, Lamont and Lumsdaine (1996), Fleming and Remolona (1997), Balduzzi, Elton, and Green (2001), Li and Engle (2000).

Table 7: Largest moves in bond yields

Daily data on 3-month T-Bill, Jan 1991 - Dec 1993		
1.	Jan 2, 1992	Fed policy rate
2.	Dec 20, 1991	Fed policy rate
3.	Sep 4, 1992	Employment
4.	Apr 9, 1992	Large Japanese market decline
5.	Feb 1, 1991	Employment, Fed policy rate
5-Minute data on 5-year T-note, Aug 1993 - Aug 1994		
1.	Aug 5, 1994	Employment
2.	May 6, 1994	Employment
3.	Jul 8, 1994	Employment
4.	Apr 1, 1994	Employment
5.	Jul 29, 1994	GDP
5-Minute data on 5-year T-note, Jan 1999 - Dec 1999		
1.	Jun 30, 1999	Fed policy rate
2.	May 5, 1999	Employment
3.	Sep 3, 1999	Employment
4.	May 14, 1999	Consumer price index
5.	Aug 6, 1999	Employment

NOTE: The three panels show the 5 largest bond-yield moves in their subsample. The first column indicates the rank of the move, the second column gives the date of the move, and the third column indicates what happened during the move. The amount of the move is not available for the top panel, so it is not included here. The top panel combines information from Table 4 and Figure 5 in Johannes (2001). The middle panel is from Table 3 in Fleming and Remolona (1999). The lower panel is from Table 1 in Furfine (2001).

When large yield movements at macroeconomic news releases are modelled as jumps, the timing of these jumps is deterministic. Piazzesi (2001) builds an affine model with deterministic jump times and state-dependent jump size distributions to be able to impose the release calendar on the estimation.

Another type of seasonality is documented in Hamilton (1996). Large spikes in the federal funds rate occur on so-called ‘settlement Wednesdays’, which mark the end of the bi-weekly reserve maintenance period. Less pronounced spikes on these days can also be found in other short-term interest rates, such as overnight LIBOR or repos (Piazzesi (2001)). An estimation with data on these very short rates therefore needs to carefully take into account these seasonalities. High-frequency studies of the effects of monetary policy are especially affected by these seasonalities, because most target moves happen around settlement Wednesdays.

## 7.10 Fitting errors at the short end

When affine models are estimated with panel data, the fit at the very short end of the yield curve often turns out to be bad. One of the reasons is that estimations typically do not use any data on extremely short yields. Indeed, seasonalities and other microstructure noise would make such an estimation difficult, as mentioned in the previous section. Piazzesi (2001) and He (2001) argue that the Federal Reserve target rate can serve as a ‘clean’ short rate that helps in fitting the short end. These papers use the target rate to pin down the short end of the swap curve. Another reason for the poor fit is that it seems like more than three factors are needed to capture the short end of the yield curve. Evidence of the need of a fourth factor is in Knez, Litterman, and Scheinkman (1994), Longstaff, Santa-Clara, and Schwartz (2001) and Piazzesi (2001).

## 8 Joint system with other macroeconomic variables

Macroeconomics often views the Federal Reserve as setting the short end of the nominal yield curve. Many issues that are being debated in the term structure literature - such as persistence, predictability, nonlinearities, and structural breaks in short yields - are therefore also being debated in monetary economics. Section 8.1 links some of these discussion points. Yields have much to do with other macroeconomic variables. Inflation in Section 8.2 and the other macro variables in Section 8.3 may teach us something about yields.

### 8.1 Monetary policy

Most central banks in industrialized countries target some overnight interbank lending rate. The Federal Reserve does so by fixing a target rate for the funds rate which is implemented by the New York Fed using open market operations. The way the Fed sets the target is usually described with policy rules, which are maps from macro variables to the target. According to the ‘Taylor rule,’ for example, the Fed sets the target in response to inflation and the output gap. Policy rules are structural equations which can be identified in several ways (for a survey, see Christiano, Eichenbaum, and Evans (1999)). The identification scheme proposed by Christiano, Eichenbaum, and Evans boils down to taking conditional expectations. The literature on structural breaks in policy rules and interest-rate smoothing by the Fed is thus intimately related to the statistical discussion about short-rate dynamics. Monetary policy regimes are usually associated with Fed chairmen. An example is the 1979-1982 monetary experiment under Paul Volcker. During this time period, the Fed stopped targeting short-term interest rates and started targeting nonborrowed reserves instead. The economy underwent two recessions during this three year experiment, but eventually inflation was under control. When policy rules are estimated over different regimes, coefficient estimates are very different. Cogley and Sargent (2001, 2002) address this issue with a random-coefficients

model. Sims (1999) and Sims and Zha (2002) argue that what looks like nonlinearities and structural breaks in policy rules may be due to time-varying second moments.

Knowledge about the operations procedures of the Fed can be used for yield-curve modeling. The model proposed by Piazzesi (2001) explicitly uses the meeting calendar of the Fed to determine the short end of the yield curve. The Fed meets 8 times per year and changes its target-short rate mostly at meetings since 1994. The market short rate fluctuates around the target. Between meetings, the short rate is thus likely to stay close to the old target level. The target is constant for long periods of time, which can be captured with counting processes for up and down moves. The yield-curve model predicts, for example, the reaction of yields to monetary policy surprises, defined as the difference between the actual target and the Fed's policy rule. The reaction of yields to these surprises turns out to be large. Cochrane and Piazzesi (2002) confirm this finding in a setting that does not impose no arbitrage. Traditional studies in which the Fed only reacts to macroeconomic variables tend to find small reactions of yields to policy shocks (Evans and Marshall (1998, 2001)). In fact, the policy rule estimated with the yield-curve model captures Fed behavior better compared to traditional rules based exclusively on macro variables. The estimated rule features both interest rate smoothing (autocorrelation in levels) and policy inertia (autocorrelation in changes).

Mankiw and Miron (1986) find that short rate movements have become much less predictable since the creation of the Fed in 1914. The reason is that the Fed smooths short rates (increases their autocorrelation in levels), which makes changes in the short rate unpredictable. Mankiw and Miron conclude that rejections of the expectations hypothesis may be due to the Fed-induced random walk character of interest rates. This idea can be used to construct a term structure model by estimating the short rate and then computing long yields with the expectations hypothesis (Balduzzi, Bertola, and Foresi (1996)). This explanation may be consistent with the fact that the persistence of short rates seems to have increased in the 1990s (Watson (1999)), a decade during which the expectations hypothesis has failed spectacularly.

Yield-curve models that incorporate interest-rate targeting by the Fed can be used to learn about policy rules. For example, the reaction of yields to macroeconomic surprises, such as nonfarm payroll numbers, seems to be hump-shaped with peaks around 2-3 years. This evidence is not consistent with a yield-curve model in which the Federal Reserve reacts to current macro variables (Piazzesi (2001)). The reason is that macroeconomic release surprises, measured as the difference between actual released number and analyst forecasts, do not seem to forecast future macroeconomic variables. Since future employment numbers are unaffected by the surprise, the Fed is likely not move the short rate in the future. This implies that long yields should not be reacting to release surprises - but they do. The Fed thus seems to react to some moving average of past release surprises instead of current macro variables.

Central banks in other countries use different operational procedures. For example,

the Bundesbank used to make its policy decisions at bi-weekly meetings without announcing an official target. Yield-curve models can then be used to learn about the latent target (Piazzesi (2002)).

## 8.2 Inflation

Central banks need to decompose nominal yields into expected inflation, risk premia and real yields (which are yields on real bonds, see Section 4 for definitions). The policy response to high expected inflation is much different from the response to high real yields, while both situations are characterized by high nominal yields. Holders of nominal bonds worry about future inflation because that is what determines the real value of the principal payment at the maturity date. This implies that expected inflation should matter for at least the determination of nominal yields. Expected inflation may also matter for real yields in a world in which money is not neutral.

Fama (1990) documents that the correlation between expected inflation and state variables that drive the real yield curve is negative, at least for horizons up to one year. More precisely, he defines the real rate as the difference between the nominal one-year yield and actual inflation over one year. The negative correlation is between the expected real short rate and expected inflation. Expected values are measured by the fitted values from regressions of actual inflation and the real short rate on a particular yield spread, the difference between the 5-year and the 1-year yield. The finding is that regression coefficients have opposite signs so that high yield spreads forecast higher inflation and lower real rates. For horizons up to one year, the signs of these forecasts cancel each other such that the yield spread does not forecast changes in the nominal rate. The  $R^2$  from these forecasts are around 20% for inflation and about half that for the real rate. For horizons beyond one year, the yield spread only predicts inflation and not the real short rate. The  $R^2$  from the real-rate forecasts decrease fast with horizon, while the  $R^2$  from the inflation forecasts stay high for several years. This means that yield spreads are able to predict changes in the nominal short rate for longer horizons.

Barr and Campbell (1996) and Campbell and Viceira (2001) also find negative correlation using affine 2-factor models. Expected inflation and the actual real short rate are Gaussian and market prices of risk are constant. Expected excess returns on all bonds, real and nominal, are therefore constant. The model for the real yield-curve is a 1-factor Vasicek model, while the nominal yield-curve is a 2-factor Vasicek model. Nominal yields in this model can thus become negative with positive probability. Expected inflation and the real rate are inherently latent variables, which are estimated using Kalman filtering with McCulloch-Kwon zero-coupon yield data and CPI inflation. All yields are assumed to be observed with error. Buraschi and Jiltsov (2001) also find negative correlation with a 3-factor model. Finally, Barr and Campbell (1996) and Campbell and Shiller (1996) find the same result with UK data on indexed bonds.

During the monetary experiment, inflation was high and the Fed under Volcker increased nominal short rates dramatically. This fight against inflation was successful in the sense that the estimated yield-curve model by Campbell and Viceira (2001) shows that expected inflation has been much less variable since 1983. Real rates, however, have become more variable. This is also reflected in the persistence of these variables since 1983. While expected inflation appears to have a unit root over the whole postwar period and shocks to the real short rate have a half-life of 5 quarters, these results are almost reversed for the subperiod after 1983. Expected inflation has become much less persistent, with shocks to expected inflation having half-lives of only 5 quarters. Shocks to the real short rate now have a half-life of 12 years!

### 8.3 Other macroeconomic variables

Ang and Piazzesi (2002) address whether macro variables add to our understanding of yields by looking at out-of-sample forecasts of yields. The forecasts are computed using a discrete-time Gaussian yield-curve model with macro variables as observable factors. Discrete time makes it easy to incorporate higher order autoregressive lags that are often needed to capture the dynamics of macroeconomic variables. Longer lags can be incorporated simply by expanding the state space. Hansen and Sargent (1991) discuss how to do this in continuous time. Market prices of risk in the model are affine in the Gaussian state variables. The estimation uses data on various inflation measures and real activity measures. The authors find that yield curve models with macro variables turn out to do better in out-of-sample forecasting. Ang, Piazzesi, and Wei (2002) estimate a 3-factor model based on a short rate, term spread and GDP growth. Their model completely characterizes the predictive regressions of GDP growth over different horizons and different term spreads on the right-hand side. Their model recommends the use of the short rate instead of any term spread for forecasting growth. This finding is in contrast to unrestricted OLS regressions and is confirmed in out-of-sample forecasts. Intuitively, arbitrage-free pricing imposes restrictions that improve the efficiency of parameter estimates. This gain in efficiency leads to vastly different point estimates in the case of forecasting GDP growth, even though the yield-curve parameters are well within confidence bounds of the corresponding OLS regression. A lot more research is needed in this area.

## References

- [1] Ahn, Dong-Hyun, Robert Dittmar, and A. Ronald Gallant (2002). “Quadratic term structure models: Theory and evidence” *Review of Financial Studies* 15, 243-88.
- [2] Aït-Sahalia, Yacine (1996). “Testing continuous-time models of the spot interest rate.” *Review of Financial Studies* 9, 385-426.
- [3] Aït-Sahalia, Yacine (2001). “Maximum likelihood estimation of discretely sampled diffusions: a closed form approximation approach.” *Econometrica* 70, 223-62.
- [4] Aït-Sahalia, Yacine (2002). “Closed-form likelihood expansions for multivariate diffusions.” Working paper, Princeton University.
- [5] Aït-Sahalia, Yacine and Robert Kimmel (2002). “Estimating affine multifactor term structure models using closed-form likelihood expansions.” Working paper, Princeton University.
- [6] Aït-Sahalia, Yacine, Lars Peter Hansen, and Jose A. Scheinkman (2002). “Discretely sampled diffusions.” forthcoming in *Handbook of Financial Econometrics*.
- [7] Alvarez, Fernando and Pablo Andres Neumeyer (1999). “Constructing historical time and maturity dependent yield spreads for emerging country sovereign debt.” Working paper, University of Chicago.
- [8] Ang, Andrew and Geert Bekaert (1998). “Regime switches in interest rates.” *NBER working paper* 6508.
- [9] Ang, Andrew and Geert Bekaert (2003). “The term structure of real rates and expected returns.” Working paper, Columbia University.
- [10] Ang, Andrew and Monika Piazzesi (2002). “A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables.” forthcoming *Journal of Monetary Economics*.
- [11] Ang, Andrew, Monika Piazzesi and Min Wei (2002). “What does the yield curve tell us about GDP growth?” Working paper, UCLA.
- [12] Angeletos, George-Marios (2001). “Fiscal policy with non-contingent debt and the optimal maturity structure.” forthcoming *Quarterly Journal of Economics*.
- [13] Attari, Mukarram (2001). “Testing interest-rate models: What do futures and options data tell us?” Working paper, University of Wisconsin.
- [14] Babbs, Simon H. and Nick J. Webber (1993). “A theory of the term structure with an official interest rate.” Working Paper, Warwick University.
- [15] Backus, David, Silverio Foresi, Abon Mozumdar, and Liuren Wu (1998). “Predictable changes in yields and forward rates.” *NBER Working paper* no 6379.

- [16] Backus, David, Silverio Foresi, and Chris Telmer (1998). "Discrete-time models of bond pricing." *NBER Working paper* no 6736.
- [17] Backus, David, Allan W. Gregory, and Stan E. Zin (1989). "Risk premiums in the term structure: Evidence from artificial economies." *Journal of Monetary Economics* 24, 371-99.
- [18] Backus, David and Stan E. Zin (1994). "Reverse engineering the yield curve." *NBER working paper* no 4676.
- [19] Balduzzi, Pierluigi, Giuseppe Bertola and Silverio Foresi (1996). "A model of target changes and the term structure of interest rates." *Journal of Monetary Economics* 39, 223-49.
- [20] Balduzzi, Pierluigi, Sanjiv R. Das, and Silverio Foresi (1998). "The central tendency: A second factor in bond yields." *Review of Economics and Statistics* 80, 62-72.
- [21] Balduzzi, Pierluigi, Sanjiv R. Das, Silverio Foresi, and Rangarajan K. Sundaram (1996). "A simple approach to three factor affine term structure models." *Journal of Fixed Income* 6, 43-53.
- [22] Balduzzi, Pierluigi, Edwin J. Elton, and T. Clifton Green (2001). "Economic news and the yield curve: Evidence from the U.S. Treasury market." *Journal of Financial and Quantitative Analysis* 36, 523-43.
- [23] Bakshi, Gurdip S. and Zhiwu Chen (1996). "Inflation, asset prices, and the term structure of interest rates in monetary economies." *Review of Financial Studies* 9, 241-75.
- [24] Bakshi, Gurdip S. and Zhiwu Chen (1997). "An alternative valuation model for contingent claims." *Journal of Financial Economics* 44, 123-65.
- [25] Bansal, Ravi and Hui Zhou (2002). "Term structure of interest rates with regime shifts." forthcoming, *Journal of Finance*.
- [26] Barr, David G. and John Y. Campbell (1996). "Inflation, real interest rates, and the bond market: a study of UK nominal and index-linked government bond prices." *NBER Working paper* no 5821.
- [27] Bekaert, Geert and Steven R. Grenadier (2000). "Stock and bond pricing in an affine economy." Working paper, Columbia Business School.
- [28] Bekaert, Geert and Robert Hodrick (2001). "Expectations hypothesis tests." *Journal of Finance* 56, 115-38.
- [29] Bekaert, Geert, Robert Hodrick, and David Marshall (1997). "On biases in tests of the expectations hypothesis of the term structure of interest rates." *Journal of Financial Economics* 44, 309-48.

- [30] Bekaert, Geert, Min Wei, and Yuhang Xing (2002). “Uncovered Interest Rate Parity and the Term Structure” Working paper, Columbia Business School, 2002.
- [31] Bliss, Robert (1999). “Fitting term structures to bond prices.” Working paper, Chicago Fed.
- [32] Boudoukh, Jacob, Matthew Richardson, Richard Stanton, and Richard Whitelaw (1999). “The stochastic behavior of interest rates: Implications from a nonlinear continuous-time model.” Working Paper, NYU Stern and UC Berkeley.
- [33] Brandt, Michael W. and Pedro Santa-Clara (2002). “Simulated likelihood estimation of diffusions with an application to exchange rates dynamics in incomplete markets.” *Journal of Financial Economics* 63, 161-210.
- [34] Brandt, Michael W. and Amir Yaron (2001). “Time-consistent no-arbitrage models of the term structure.” Working paper, Wharton School.
- [35] Brown, Stephen J. and Philip H. Dybvig (1986). “The empirical implications of the Cox, Ingersoll, and Ross theory of the term structure of interest rates.” *Journal of Finance* 41, 617-32.
- [36] Brown, Roger H. and Stephen Schaefer (1994). “The term structure of real interest rates and the Cox, Ingersoll, and Ross Model.” *Journal of Financial Economics* 35, 3-42.
- [37] Buraschi, Andrea and Alexei Jiltsov (2001). “How large is the inflation risk premium? A monetary model of the yield curve.” Working paper, London Business School.
- [38] Brémaud, Pierre (1981). “*Point processes and queues: Martingale dynamics.*” New York: Springer.
- [39] Campbell, John Y. (1986). “A defense of the traditional hypotheses about the term structure of interest rates.” *Journal of Finance* 41, 183-93.
- [40] Campbell, John Y. and John H. Cochrane (1999). “By force of habit: A consumption-based explanation of aggregate stock market behavior.” *Journal of Political Economy* 107, 205-51.
- [41] Campbell, John Y., Andrew W. Lo, and A. Craig MacKinlay (1997). “*The econometrics of financial markets.*” Princeton University Press: Princeton.
- [42] Campbell, John Y. and Robert Shiller (1991). “Yield spreads and interest rates: A bird’s eye view.” *Review of Economic Studies* 58, 495-514.
- [43] Campbell, John Y. and Luis Viceira (2001). “Who should buy long term bonds?” *American Economic Review* 91, 99-127.

- [44] Carrasco, Marine, Mikhail Chernov, Jean-Pierre Florens, and Eric Ghysels (2001). “Estimation of jump-diffusions with a continuum of moment conditions.” Working paper, University of Rochester.
- [45] Chacko, George and Luis Viceira (2001). “Spectral GMM estimation of continuous-time processes.” Working paper, Harvard University.
- [46] Chan, K, Andrew Karolyi, Francis Longstaff, and Anthony Sanders (1992). “The volatility of the short-term interest rates: An empirical comparison of alternative models of the term structure of interest rates.” *Journal of Finance* 68, 1209-27.
- [47] Chapman, David, John Long, and Neil D. Pearson (1999). “Using proxies for the short rate: When are three months like an instant?” *Review of Financial Studies*, 12, 763-806.
- [48] Chapman, David (1997). “The cyclical properties of consumption growth and the real term structure.” *Journal of Monetary Economics*, 39, 145-72.
- [49] Chapman, David and Neil D. Pearson (2000). “Is the short rate drift actually nonlinear?” *Journal of Finance* 55, 355-88.
- [50] Chapman, David and Neil D. Pearson (2001). “Recent advances in estimating term-structure models.” *Financial Analysts Journal*, July/August, 77-95.
- [51] Chen, Lin (1996). “*Stochastic mean and stochastic volatility - A three-factor model of the term structure of interest rates and its application to the pricing of interest rate derivatives.*” Blackwell Publishers, Oxford, UK.
- [52] Chen, Ren-Row and Louis Scott (1992). “Pricing interest-rate options in a two-factor Cox-Ingersoll-Ross model of the term structure.” *Review of Financial Studies* 5, 613-36.
- [53] Chen, Ren-Row and Louis Scott (1993). “Maximum likelihood estimation for a multifactor equilibrium model of the term structure of interest rates.” *Journal of Fixed Income* 3, 14-31.
- [54] Christiano, Lawrence, Martin Eichenbaum, and Charles Evans (1999). “Monetary policy shocks: What have we learned and to what end?” *Handbook of Macroeconomics*, Michael Woodford and John B. Taylor, Eds., North Holland: Amsterdam.
- [55] Cochrane, John H. (2001). “Long term debt and optimal policy in the fiscal theory of the price level” *Econometrica* 69, 69-116.
- [56] Cochrane, John H. and Monika Piazzesi (2002). “Bond risk premia.” Working paper, Chicago GSB and UCLA.
- [57] Cochrane, John H. and Monika Piazzesi (2002). “The Fed and Interest Rates: A High-frequency Identification.” *American Economic Review* 92, 90-95.

- [58] Cogley, Timothy and Thomas J. Sargent (2001). "Evolving post-world war II U.S. inflation dynamics." forthcoming in *NBER Macroannual 2001*.
- [59] Cogley, Timothy and Thomas J. Sargent (2002). "Drifts and volatilities: Monetary policies and outcomes in the post WWII U.S." Working paper, Stanford University.
- [60] Collin-Dufresne, Pierre and Robert Goldstein (2001). "Do bonds span the fixed income markets? Theory and evidence for unspanned stochastic volatility." forthcoming, *Journal of Finance*.
- [61] Collin-Dufresne, Pierre, Robert Goldstein, and Christopher Jones (2002). "Identification and estimation of maximal affine term structure models: An application to stochastic volatility." Working paper, Carnegie-Mellon University.
- [62] Collin-Dufresne, Pierre and Bruno Solnik (2001). "On the term structure of default premia in the swap and LIBOR markets." *Journal of Finance* 56, 1095-15.
- [63] Conley, Tim G., Lars Peter Hansen, Erzo G.J. Luttmer, and Jose A. Scheinkman (1997). "Short-term interest rates as subordinated diffusions." *Review of Financial Studies* 10, 525-77.
- [64] Constantinides, George (1992). "A theory of the nominal term structure of interest rates." *Review of Financial Studies* 5, 531-52.
- [65] Cox, John C., Jonathan E. Ingersoll, and Stephen A. Ross (1981). "A reexamination of traditional hypotheses about the term structure of interest rates." *Journal of Finance* 36, 321-46.
- [66] Cox, John C., Jonathan E. Ingersoll, and Stephen A. Ross (1985). "A theory of the term structure of interest rates." *Econometrica* 53, 385-407.
- [67] Dai, Qiang (2001). "Asset pricing in a neoclassical model with limited participation." Working paper, NYU Stern.
- [68] Dai, Qiang and Kenneth Singleton (2000). "Specification analysis of affine term structure models." *Journal of Finance* 55, 1943-78.
- [69] Dai, Qiang and Kenneth Singleton (2002). "Expectation puzzles, time-varying risk premia, and affine models of the term structure." *Journal of Financial Economics* 63, 415-41.
- [70] Dai, Qiang and Kenneth Singleton (2003). "Term structure modeling in theory and reality." Forthcoming *Review of Financial Studies*.
- [71] Darolles, Serge, Christian Gouriéroux, and Joann Jasiak (2001). "Compound autoregressive models." Working paper, CREST, Paris.
- [72] Das, Sanjiv (2002). "The surprise element: Jumps in interest rates." *Journal of Econometrics* 106, 27-65.

- [73] Das, Sanjiv and Silverio Foresi (1996). “Exact solutions for bond and option prices with systematic jump risk.” *Review of Derivatives Research* 1, 7-24.
- [74] Das, Sanjiv and Rangarajan K. Sundaram (1997). “Taming the skew: Higher-order moments in modeling asset price processes in finance.” *NBER working paper* no 5976.
- [75] De Jong, Frank and Pedro Santa-Clara (1999). “The dynamics of the forward interest rate: A formulation with state variables.” *Journal of Financial and Quantitative Analysis* 34, 131-57.
- [76] Den Haan, Wouter (1995). “The term structure of interest rates in real and monetary economies.” *Journal of Economic Dynamics and Control*, 19, 909-40.
- [77] Duarte, Jefferson (2000). “The relevance of the price of risk in affine term structure models.” Working paper, University of Chicago.
- [78] Duffee, Gregory (1996). “Idiosyncratic variation of Treasury bill yields.” *Journal of Finance* 51, 527-52.
- [79] Duffee, Gregory (2002). “Term premia and interest rate forecasts in affine models.” *Journal of Finance*. 57, 405-43.
- [80] Duffie, Darrell (2001). “*Dynamic asset pricing theory*.” 3rd Edition. Princeton: Princeton University Press.
- [81] Duffie, Darrell, Damir Filipovic, and Walter Schachermayer (2001). “Affine processes and applications in finance.” Working paper, Stanford GSB.
- [82] Duffie, Darrell and Ming Huang (1996). “Swap rates and credit quality.” *Journal of Finance* 51, 921-49.
- [83] Duffie, Darrell and Rui Kan (1996). “A yield-factor model of interest rates.” *Mathematical Finance* 6, 379-406.
- [84] Duffie, Darrell, Jun Pan, and Kenneth Singleton (2000). “Transform analysis and asset pricing for affine jump-diffusions.” *Econometrica* 68, 1343-76.
- [85] Duffie, Darrell, Lasse Pedersen, and Kenneth Singleton (2002). “Modeling sovereign yield spreads: A case study of Russian debt,” forthcoming *Journal of Finance*.
- [86] Duffie, Darrell and Kenneth Singleton (1997). “An econometric model of the term structure of interest rate swap yields” *Journal of Finance* 52, 1287-1323.
- [87] Durham, Garland B. (2001). “Likelihood-based specification analysis of continuous-time models of the short-term interest rate.” Working paper, University of Iowa.
- [88] Durham, Garland B. and A. Ronald Gallant (2002). “Numerical techniques for maximum likelihood estimation of continuous time diffusion processes.” forthcoming *Journal of Business and Economic Statistics*.

- [89] El Karoui, Nicole, R. Myneni, and R. Viswanathan (1993). "Arbitrage pricing and hedging of interest rate claims with state variables." Working Paper, Université de Paris VI, Laboratoire de Probabilité.
- [90] Estrella, Arturo and Gikas A. Hardouvelis (1991). "The term structure as a predictor of real economic activity." *Journal of Finance* 46, 555-76.
- [91] Evans, Chales L. and David Marshall (1998). "Monetary policy and the term structure of nominal interest rates: Evidence and theory." *Carnegie-Rochester Conference Series on Public Policy* 49, 53-111.
- [92] Evans, Charles L. and David Marshall (2001). "Economic determinants of the term structure of nominal interest rates." Working paper, Chicago Fed.
- [93] Fama, Eugene F. (1990). "Term-structure forecasts of interest rates, inflation, and real returns." *Journal of Monetary Economics* 25, 59-76.
- [94] Fama, Eugene F. and Robert R. Bliss (1987). "The information in long-maturity forward rates." *American Economic Review* 77, 680-92.
- [95] Farnsworth, Heber and Richard Bass (2001). "The term structure with credible and semi-credible targeting." Working Paper, Washington University in St. Louis.
- [96] Favero, Carlo, Francesco Giavazzi, and Luigi Spaventa (1997). "High yields: The spread on german interest rates." *Economic Journal* 107, 663.
- [97] Fisher, Mark (1998). "A simple model of the failure of the expectations hypothesis." Working paper, Federal Reserve Atlanta.
- [98] Fisher, Mark and Christian Gilles (1996). "Estimating exponential-affine models of the term structure." Working paper, Federal Reserve Atlanta.
- [99] Fisher, Mark and Christian Gilles (1998). "Around and around: The expectations hypothesis." *Journal of Finance* 53, 365-83.
- [100] Fleming, J. Michael and Eli M. Remolona (1997). "What moves the bond market?" *FRBNY Economic Policy Review* December, 31-50.
- [101] Fleming, J. Michael and Eli M. Remolona (1999). "The term structure of announcement effects." Working Paper, Federal Reserve Bank of New York.
- [102] Furfine, Craig (2001). "Do macroeconomic announcements still drive the Treasury market?" *BIS Quarterly Review*, June, 49-57.
- [103] Gallant, A. Ronald and George Tauchen (2002). "Simulated score methods and indirect inference for continuous-time models." forthcoming in *Handbook of Financial Econometrics*.

- [104] Gerlach, Stefan and Frank Smets (1997). "The term structure of euro-rates: some evidence in support of the expectations hypothesis." *Journal of International Money and Finance* 16, 305-21.
- [105] Geweke, John (1996). "Monte Carlo simulation and numerical integration." in *Handbook of Computational Economics* 13, H.M. Amman, D.A. Kendrick and J. Rust eds., Elsevier Science, North-Holland, Amsterdam.
- [106] Ghysels, Eric and Serena Ng (1998). "A semi-parametric factor model of interest rates and tests of the affine term structure." *Review of Economics and Statistics* 80, 535-48.
- [107] Gibbons, Michael R. and Krishna Ramaswamy (1993). "A test of the Cox, Ingersoll and Ross model of the term structure." *Review of Financial Studies* 6, 619-58.
- [108] Goldstein, Robert (2000). "The term structure of interest rates as a random field." *Review of Financial Studies* 13, 365-84.
- [109] Gong, Frank F. and Eli M. Remolona (1996). "A three-factor econometric model of the U.S. term structure." Working paper, Federal Reserve Bank of New York.
- [110] Gouriéroux, Christian, Alain Monfort, and Vassilis Polimenis (2002). "Affine term structure models." Working paper, CREST, Paris.
- [111] Grinblatt, Mark and Francis Longstaff (2000). "Financial innovation and the role of derivative securities: An empirical analysis of the treasury STRIPS program." *Journal of Finance* 55, pp. 1415-36.
- [112] Hamilton, James D. (1996). "The daily market for federal funds." *Journal of Political Economy* 104, 26-56.
- [113] Hamilton, James D. (1994). "*Time series analysis*" Princeton University Press, Princeton, New Jersey.
- [114] Hamilton, James D. and Dong Kim (2002). "A re-examination of the predictability of the yield spread for real economic activity." *Journal of Money, Credit, and Banking* 34, 340-60.
- [115] Hansen, Lars Peter (1982). "Large sample properties of generalized method of moments estimators." *Econometrica* 50, 1029-54.
- [116] Hansen, Lars Peter and Thomas J. Sargent (1991). "Exact linear rational expectations models: Specification and estimation." in *Rational Expectations Econometrics*, Westview Press, Oxford.
- [117] Hardouvelis, Gikas (1994). "The term structure spread and future changes in long and short rates in the G7 countries: Is there a puzzle?" *Journal of Monetary Economics* 33, 255-83.

- [118] Harvey, Campbell R. (1988). “The real term structure and consumption growth.” *Journal of Financial Economics* 22, 305-33.
- [119] He, Hua (2001). “Modeling term structures of swap spreads.” Working paper, Yale School of Management.
- [120] Heston, Steven L. (1991). “Testing continuous time models of the term structure of interest rates.” Working paper, Yale University.
- [121] Honoré, Peter (1998). “Maximum likelihood estimation of non-linear continuous-time term-structure models.” Working paper, Aarhus School of Business, Denmark.
- [122] Hull, John (2000). “*Options, futures, and other derivatives.*” (4th edition). Englewood Cliffs, NJ: Prentice Hall
- [123] Jagannathan, Ravi, Andrew Kaplin, and Steve Guoqiang Sun (2001). “An evaluation of multi-factor CIR models using LIBOR, swap rates, and cap and swaption prices.” *NBER Working paper* 8682.
- [124] Jegadeesh, Narasimhan and George G. Pennacchi (1996). “The behavior of interest rates implied by the term structure of Eurodollar futures.” *Journal of Money, Credit and Banking* 28, 420-446.
- [125] Jensen, Bjarke and Rolf Poulsen (1999). “A comparison of approximation techniques for transition densities of diffusion processes.” Working paper, Aarhus University.
- [126] Johannes, Michael (2001). “The statistical and economic role of jumps in continuous-time interest-rate models.” Working paper, Columbia Business School.
- [127] Jones, Charles M., Owen Lamont, and Robin Lumsdaine (1996). “Macroeconomic news and bond market volatility.” *Journal of Financial Economics* 47, 315-37.
- [128] Jordan, Bradford D., Randy D. Jorgensen, and David R. Kuipers (2000). “The relative pricing of U.S. Treasury STRIPS: Empirical Evidence.” *Journal of Financial Economics* 56, 89-123.
- [129] Karatzas, Ioannis and Steven E. Shreve (1988). “Brownian motion and stochastic calculus.” Springer Verlag, Heidelberg.
- [130] Kimmel, Robert (2001). “Modeling the term structure of interest rates: A new approach.” Working paper, Princeton University.
- [131] Knez, Peter, Robert Litterman, and José Scheinkman (1994). “Explorations into factors explaining money market returns.” *Journal of Finance* 49, 1861-82.
- [132] Landen, Camilla (2000). “Bond pricing in a hidden Markov model of the short rate.” *Finance and Stochastics* 4, 371-89.

- [133] Langetieg, Terence C. (1980). "A multivariate model of the term structure." *Journal of Finance* 25, 71-97.
- [134] Li, Li and Robert F. Engle (1998). "Macroeconomic Announcements and Volatility of Treasury Futures." Working Paper No. 98-27, UC San Diego.
- [135] Litterman, Robert and José Scheinkman (1991). "Common factors affecting bond returns." *Journal of Fixed Income* 1, 54-61.
- [136] Liu, Jun, Francis Longstaff, and Ravit Mandell (2002). "The market price of credit risk: An empirical analysis of interest rate swap spreads." Working paper, UCLA.
- [137] Lo, Andrew W. (1988). "Maximum likelihood estimation of generalized Ito processes with discretely-sampled data." *Econometric Theory* 4, 231-47.
- [138] Longstaff, Francis (2000a). "Arbitrage and the expectations hypothesis." *Journal of Finance* 55, 989-94.
- [139] Longstaff, Francis (2000b). "The term structure of very short term rates: New evidence for the expectations hypothesis." *Journal of Financial Economics* 58, 397-96.
- [140] Longstaff, Francis, Pedro Santa-Clara, and Eduardo Schwartz (2001). "The relative valuation of caps and swaptions: Theory and empirical evidence" *Journal of Finance* 56, 2067-109.
- [141] Longstaff, Francis and Eduardo Schwartz (1992). "Interest rate volatility and the term structure: A two-factor general equilibrium model." *Journal of Finance* 47, 1259-82.
- [142] Longstaff, Francis and Monika Piazzesi (2002). "Corporate earnings and the equity premium." Working paper, UCLA.
- [143] Mamaysky, Harry (2002). "A model for pricing stock and bonds." Working paper, Yale University.
- [144] Mankiw, Gregory N. and Jeffrey A. Miron (1986). "The changing behavior of the term structure of interest rates." *Quarterly Journal of Economics*, CI(2), 211-28.
- [145] Mardia, Kani V., John T. Kent, and John M. Bibby (1979). "*Multivariate analysis*." Academic Press, San Diego.
- [146] McCulloch, J. Huston and Heon-Chul Kwon (1993). "U.S. Term structure data, 1947-1991." Working paper, Ohio State University.
- [147] Mishkin, Frederik S. (1990). "What does the term structure tell us about future inflation?" *Journal of Monetary Economics* 25, 77-95.
- [148] Missale, Alessandro (1997). "Managing the public debt: The optimal taxation approach." *Journal of Economic Surveys* 11, 235-265.

- [149] Naik, Vasant and Moon Hoe Lee (1997). "Yield curve dynamics with discrete shifts in economic regimes: Theory and Estimation." Working paper, University of British Columbia, Canada.
- [150] Nelson, Charles R. and Andrew F. Siegel (1987). "Parsimonious modelling of yield curves." *Journal of Business*, 60, 473-89.
- [151] Pearson, Neil D. and Tong-Sheng Sun (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.
- [152] Pedersen, Asger Roer (1995). "A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations." *Scandinavian Journal of Statistics* 22, 55-71.
- [153] Pennacchi, George G. (1991). "Identifying the dynamics of real interest rates and inflation: Evidence using survey data." *Review of Financial Studies* 4, 53-86.
- [154] Piazzesi, Monika (2001). "Macroeconomic jump effects and the yield curve." Working paper, UCLA.
- [155] Piazzesi, Monika (2002). "Bond yields and the Bundesbank." Working paper, UCLA.
- [156] Rebelo, Sergio and Danyang Xie (1999). "On the optimality of interest-rate smoothing." *Journal of Monetary Economics* 43, 263-82.
- [157] Sack, Brian (2000). "Using Treasury STRIPS to measure the yield curve." Working paper, Federal Reserve Board.
- [158] Santa-Clara, Pedro (1995). "Simulated likelihood estimation of diffusions with an application to the short term interest rate." Ph.D. Dissertation, Insead, France.
- [159] Santa-Clara, Pedro and Didier Sornette (2001). "The dynamics of the forward interest rate curve with stochastic string shocks." *Review of Financial Studies* 14, 149-85.
- [160] Sargent, Thomas J. (1979). "A Note on Maximum Likelihood Estimation of the Rational Expectations Model of the Term Structure." *Journal of Monetary Economics* 5, 133-43.
- [161] Seppala, Juha Ilmari (2002). "The term structure of real interest rates: Theory and evidence from the U.K. Index-linked bonds." Working paper, University of Illinois at Urbana-Champaign.
- [162] Sims, Christopher (1999). "Drifts and breaks in monetary policy." Working paper, Princeton University.
- [163] Sims, Christopher and Tao Zha (2002). "Macroeconomic switching." Working paper, Princeton University.

- [164] Stanton, Richard (1997). “A nonparametric model of term structure dynamics and the market price of interest rate risk.” *Journal of Finance* 52, 1973-2002.
- [165] Singleton, Kenneth (2001). “Estimation of affine asset pricing models using the empirical characteristic function.” *Journal of Econometrics* 102, 111-41.
- [166] Stock, James and Mark Watson (2001). “Forecasting output and inflation: The role of asset prices.” Working paper, Princeton University.
- [167] Sun, Tong-Sheng (1992). “Real and nominal interest rates: A discrete-time model and its continuous time limit.” *Review of Financial Studies* 5, 581-611.
- [168] Telmer Chris and Stan E. Zin (1996). “The yield curve: The terms of endearment or terms of endowment.” Working paper, Carnegie-Mellon University.
- [169] Wachter, Jessica (2001). “Habit formation and returns on bonds and stocks.” Working paper, NYU Stern.
- [170] Watson, Mark (1999). “Explaining the increased variability in long term interest rates.” Working Paper, Princeton University.
- [171] Wu, Shu (2001a). “Changes in monetary policy and the term structure of interest rates: Theory and evidence.” Working paper, University of Kansas.
- [172] Wu, Tao (2001b). “Macro factors and the affine term structure of interest rates.” Working paper, Yale University.
- [173] Valkanov, Ross (2001). “The term structure with highly persistent interest rates.” Working paper, UCLA.
- [174] Vasicek, Oldrich A. (1977). “An equilibrium characterization of the term structure.” *Journal of Financial Economics* 5, 177-88.
- [175] Veronesi, Pietro and Francis Yared (2000). “Short and long horizon term and inflation risk premia in the U.S. term structure: Evidence from an integrated model for nominal and real bond prices under regime shifts.” Working paper, Chicago GSB.