July 1995

Modeling the term structure

A. R. Pagan
A. D. Hall
V. Martin

Follow this and additional works at: http://epublications.bond.edu.au/discussion_papers

Recommended Citation
http://epublications.bond.edu.au/discussion_papers/60

This Discussion Paper is brought to you by the Bond Business School at ePublications@bond. It has been accepted for inclusion in School of Business Discussion Papers by an authorized administrator of ePublications@bond. For more information, please contact Bond University's Repository Coordinator.
DISCUSSION PAPERS

"Modeling the Term Structure"

A.R. Pagan
Australian National University

A.D. Hall
Bond University

V. Martin
University of Melbourne

DISCUSSION PAPER NO 60

July 1995

University Drive, Gold Coast, QLD, 4229
AUSTRALIA
Bond University was established by Act of Parliament in 1987 as an independent, private University. The first student intake occurred in May 1989. The School of Business offers degrees in the undergraduate (BCom, BHospMgt and Diploma) and the graduate (MBA, MAcc, MCom, MTM and PhD) levels.

The School teaches and sponsors research in accounting, economics, econometrics, finance, marketing, management, organisational behaviour and related disciplines in hospitality and real estate fields.

The Discussion Paper series is intended to foster research and comments are invited. The views expressed in the papers are the opinion of the authors and do not necessarily reflect the views of the School or the University.

Lists of available Discussion Papers and copies of the papers (which are free of charge) may be obtained from:

The Senior School Administrator
School of Business
Bond University
GOLD COAST QLD 4229

Telephone: (07) 5595 2244
Fax: (07) 5595 1160
Modeling the Term Structure

A.R. Pagan
Australian National University

A.D. Hall
Bond University

V. Martin
University of Melbourne

March 1994, revised May 1995

\footnotesize{\textsuperscript{1}We are grateful for comments on previous versions of this paper by John Robertson, Peter Phillips and Ken Singleton. All computations were performed with a beta version of MICROFIT 4 and GAUSS 3.2}
Contents

1 Introduction .......................................................... 2
2 Characteristics of Term Structure Data ................................. 2
   2.1 Univariate Properties ............................................ 2
   2.2 Multivariate Properties ......................................... 10
      2.2.1 The Level of the Yield Curve ............................. 10
      2.2.2 The Shape of the Yield Curve ................... 11
3 Models of the Term Structure .......................................... 15
   3.1 Solutions from the Consumer's Euler Equations .......... 15
   3.2 One Factor Models From Finance ............................ 19
   3.3 Two Factor Models From Finance ............................ 21
   3.4 Multiple Non-Independent Factor Models in Finance .... 24
   3.5 Forward Rate Models ........................................... 25
4 Conclusion .............................................................. 26
5 References .............................................................. 26
1 Introduction

Models of the term structure of interest rates have assumed increasing importance in recent years in line with the need to value interest rate derivative assets. Economists and econometricians have long held an interest in the subject, as an understanding of the determinants of the term structure has always been viewed as crucial to an understanding of the impact of monetary policy and its transmission mechanism. Most of the approaches taken to the question in the finance literature have revolved around the search for common factors that are thought to underlie the term structure and little has been borrowed from the economic or econometrics literature on the subject. The converse can also be said about the small amount of attention paid in econometric research to the finance literature models. The aim of the present chapter is to look at the connections between the two literatures with the aim of showing that a synthesis of the two may well provide some useful information for both camps.

The paper begins with a description of a standard set of data on the term structure. This results in a set of stylized facts pertaining to the nature of the stochastic processes generating yields as well as their spreads. Such a set of facts is useful in forming an opinion of the likelihood of various approaches to term structure modelling being capable of replicating the data. Section 3 outlines the various models used in both the economics and finance literature, and assesses how well these models perform in matching the stylized facts. Section 4 presents a conclusion.

2 Characteristics of Term Structure Data

2.1 Univariate Properties

The data set examined involves monthly observations on 1, 3, 6 and 9 month and 10 year zero coupon bond yields over the period December 1946 to February 1991, constructed by McCulloch and Kwon (1993); this is an updated version of McCulloch (1989).

Table 1 records the autocorrelation characteristics of the series, with $\hat{\rho}_j$ being the $j^{th}$ autocorrelation coefficient, DF the Dickey-Fuller test, ADF(12) the Augmented Dickey-Fuller test with 12 lags, $r_t(\tau)$ the yield on zero-coupon bonds with maturity of $\tau$ months and $sp_t(\tau)$ is the spread $r_t(\tau) - r_t(1)$. It shows that there is strong evidence of a unit root in all interest rate series, because this would imply the possibility of negative interest rates, finance modellers have generally maintained that either there is no unit root and the series feature mean reversion or, in continuous time, that an appropriate model is given by the stochastic differential equation

$$dr_t = \alpha dt + \sigma r_t d\eta_t,$$

where, throughout the paper, $d\eta_t$ is a Wiener process. Because of the "levels effect" of $r_t$ upon the volatility of interest rate changes, we can think of this as an equation in $d\log r_t$ with constant volatility, and the logarithmic transformation ensures that $r_t$ remains positive.\(^1\) In any case, the important point to be made here is that interest rates seem to behave as integrated processes, certainly over the samples of data we possess. It may be that the autoregressive root is close to

\(^1\)It is known that, if $r_t$ is replaced by $r_t^\gamma$, the restriction $\gamma > 0$ ensures a positive interest rate while, if $\gamma = 0$, $\sigma < 2\alpha$ is needed.
<table>
<thead>
<tr>
<th></th>
<th>DF</th>
<th>ADF(12)</th>
<th>( \hat{\rho}_1 )</th>
<th>( \hat{\rho}_2 )</th>
<th>( \hat{\rho}_6 )</th>
<th>( \hat{\rho}_{12} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( r(1) )</td>
<td>-2.41</td>
<td>-2.02</td>
<td>.98</td>
<td>.98</td>
<td>.98</td>
<td>.98</td>
</tr>
<tr>
<td>( r(3) )</td>
<td>-2.15</td>
<td>-1.89</td>
<td>.98</td>
<td>.98</td>
<td>.98</td>
<td>.98</td>
</tr>
<tr>
<td>( r(6) )</td>
<td>-2.12</td>
<td>-1.91</td>
<td>.99</td>
<td>.99</td>
<td>.99</td>
<td>.99</td>
</tr>
<tr>
<td>( r(9) )</td>
<td>-2.12</td>
<td>-1.89</td>
<td>.99</td>
<td>.99</td>
<td>.99</td>
<td>.99</td>
</tr>
<tr>
<td>( r(120) )</td>
<td>-1.41</td>
<td>-1.53</td>
<td>.99</td>
<td>.99</td>
<td>.99</td>
<td>.99</td>
</tr>
<tr>
<td>( sp(3) )</td>
<td>-15.32</td>
<td>-3.37</td>
<td>.38</td>
<td>.33</td>
<td>.21</td>
<td>.38</td>
</tr>
<tr>
<td>( sp(6) )</td>
<td>-11.67</td>
<td>-4.21</td>
<td>.59</td>
<td>.51</td>
<td>.26</td>
<td>.30</td>
</tr>
<tr>
<td>( sp(9) )</td>
<td>-10.38</td>
<td>-4.38</td>
<td>.66</td>
<td>.55</td>
<td>.27</td>
<td>.26</td>
</tr>
<tr>
<td>( sp(120) )</td>
<td>-5.60</td>
<td>-4.15</td>
<td>.89</td>
<td>.80</td>
<td>.55</td>
<td>.32</td>
</tr>
</tbody>
</table>

The 5% critical value for the DF and ADF tests is -2.87.
Table 2: Autocorrelation Features, Forward Rates, Full Sample

<table>
<thead>
<tr>
<th>Feature</th>
<th>DF</th>
<th>ADF(12)</th>
<th>$\hat{\rho}_1$</th>
<th>$\hat{\rho}_2$</th>
<th>$\hat{\rho}_6$</th>
<th>$\hat{\rho}_{12}$</th>
<th>$\hat{\rho}_1(\Delta t)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$F(1)$</td>
<td>-2.28</td>
<td>-1.92</td>
<td>.98</td>
<td>.07</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$F(3)$</td>
<td>-2.18</td>
<td>-1.97</td>
<td>.98</td>
<td>.04</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$F(6)$</td>
<td>-2.39</td>
<td>-1.91</td>
<td>.98</td>
<td>.09</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$F(9)$</td>
<td>-2.14</td>
<td>-1.77</td>
<td>.98</td>
<td>.07</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$Fp(0,1)$</td>
<td>-17.08</td>
<td>-4.07</td>
<td>.29</td>
<td>.18</td>
<td>.11</td>
<td>.18</td>
<td></td>
</tr>
<tr>
<td>$Fp(2,3)$</td>
<td>-19.52</td>
<td>-5.17</td>
<td>.16</td>
<td>.06</td>
<td>.01</td>
<td>-.02</td>
<td></td>
</tr>
<tr>
<td>$Fp(5,6)$</td>
<td>-20.61</td>
<td>-5.77</td>
<td>.11</td>
<td>-.12</td>
<td>-.05</td>
<td>-.03</td>
<td></td>
</tr>
<tr>
<td>$Fp(8,9)$</td>
<td>-19.73</td>
<td>-4.69</td>
<td>.15</td>
<td>-.00</td>
<td>.08</td>
<td>.02</td>
<td></td>
</tr>
</tbody>
</table>

The 5% critical value for the DF and ADF tests is -2.87.

unity, rather than identical to it, but such “near integrated” processes are best handled with the integrated process technology rather than that for stationary processes.

Instead of the yields one might examine the time series characteristics of the forward rates. The forward rate $F^k_i(\tau)$ contracted at time $t$ for a $\tau$ period bond to be bought at $t + k$ is $F^k_i(\tau) = \left[ \frac{1}{\tau + k} \right] \left[ (\tau + k)r_i(\tau + k) - kr_i(k) \right]$. For a forward contract one period ahead this becomes $F^1_i(\tau) = F_i(\tau) = \frac{1}{\tau + 1} \left[ (\tau + 1)r_i(\tau + 1) - r_i(1) \right]$. For reasons that become apparent later it is also of interest to examine the properties of the forward “spreads” $Fp_i(\tau, \tau - 1) = F_i(\tau - 1) - F_{i-1}(\tau)$. These results are to be found in Table 2. Generally, the conclusions would be the same as for yields, except that the persistence in forward rate spreads is not as marked, particularly as the maturity gets longer.

As Table 1 also shows, there is a lot of persistence in spreads between short-dated maturities; after fitting an AR(2) to $sp(3)$ the LM test for serial correlation over 12 lags is 80.71. This persistence shows up in other transformations of the yield series, e.g. the realized excess holding yield $h_{t+1}(\tau) = \tau r_i(\tau) - (\tau - 1)r_{t+1}(\tau - 1) - r_i(1)$, when $\tau = 3$, has serial correlation coefficients of .188 (lag 1), .144 (lag 8), and .111 (lag 10). Such processes are persistent, but not integrated, as the ADF(12) for $h_{t+1}(3)$ clearly shows with its value of -5.27. Papers have appeared concluding that the excess holding yield is a non-stationary process—Evans and Lewis (1994) and Hejazi (1994). That conclusion was reached by the authors performing a Phillips-Hansen (1990) regression of $h_{t+1}(\tau)$ on $F_{t-1}(\tau)$. Applying the same test to our data, with McCulloch’s forward rate series, produces an estimated coefficient on $F_{t-1}(\tau)$ of .11 with a t ratio of 10, quite consistent with both Evans and Lewis’ and Hejazi’s results. However, it does not seem reasonable to
Table 3: Autocorrelation Features, Pre October 1979

<table>
<thead>
<tr>
<th></th>
<th>DF</th>
<th>ADF(12)</th>
<th>$\hat{\rho}_1$</th>
<th>$\hat{\rho}_2$</th>
<th>$\hat{\rho}_4$</th>
<th>$\hat{\rho}_{12}$</th>
<th>$\hat{\rho}_1(\Delta r)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r(1)$</td>
<td>-.76</td>
<td>-.79</td>
<td>.97</td>
<td></td>
<td></td>
<td></td>
<td>-.14</td>
</tr>
<tr>
<td>$r(3)$</td>
<td>-.64</td>
<td>-1.03</td>
<td>.97</td>
<td></td>
<td></td>
<td></td>
<td>-.07</td>
</tr>
<tr>
<td>$r(6)$</td>
<td>-.52</td>
<td>-1.00</td>
<td>.98</td>
<td></td>
<td></td>
<td></td>
<td>.08</td>
</tr>
<tr>
<td>$r(9)$</td>
<td>-.55</td>
<td>-.89</td>
<td>.98</td>
<td></td>
<td></td>
<td></td>
<td>.11</td>
</tr>
<tr>
<td>$r(120)$</td>
<td>-.14</td>
<td>.33</td>
<td>.99</td>
<td></td>
<td></td>
<td></td>
<td>-.04</td>
</tr>
<tr>
<td>sp(3)</td>
<td>4.05</td>
<td>-4.30</td>
<td>.90</td>
<td>.84</td>
<td>.59</td>
<td>.23</td>
<td></td>
</tr>
<tr>
<td>sp(6)</td>
<td>11.99</td>
<td>-2.64</td>
<td>.46</td>
<td>.34</td>
<td>.22</td>
<td>.34</td>
<td></td>
</tr>
<tr>
<td>sp(9)</td>
<td>8.80</td>
<td>-2.89</td>
<td>.70</td>
<td>.56</td>
<td>.39</td>
<td>.38</td>
<td></td>
</tr>
<tr>
<td>sp(120)</td>
<td>8.20</td>
<td>-3.10</td>
<td>.71</td>
<td>.60</td>
<td>.38</td>
<td>.36</td>
<td></td>
</tr>
</tbody>
</table>

The 5% critical value for the DF and ADF tests is -2.87.
interpret this as evidence of non-stationarity. Certainly the series is persistent, and an I(1) series like the forward rate exhibits extreme persistence, so that regressing one upon the other can be expected to lead to some "correlation", but to conclude, therefore, that the excess holding yield is non-stationary is quite incorrect. A fractionally integrated process that is stationary would also show such a relationship with an I(1) process. Indeed, the autocorrelation functions of the spreads and excess yields are reminiscent of those for the squares of yield changes, which have been modelled by fractionally integrated processes—see Baillie et al. (1993). Nevertheless, the strong persistence in spreads is a characteristic which is a substantial challenge to term structure models.2

As is well known, there was a switch in monetary policy in the US in October 1979 away from targeting interest rates, and this fact generally means that any analyses have to be re-done to ensure that the results do not simply reflect outcomes from 1979 to 1982. Table 3 therefore presents the same statistics as in Table 1 but using only pre October 1979 data. It is apparent that the conclusions drawn above are quite robust.

It is also well known that there is a substantial dependence of the conditional volatility of \( \Delta r_t(\tau) \) upon the past, but the exact nature of this dependence has been subject to much less analysis. As will become clear, the most important issue is whether the conditional variance, \( \sigma^2_t \), exhibits a levels effect and, if so, exactly what relationship is likely to hold. Here we examine the evidence for a "levels effect" in volatility, i.e. \( \sigma^2_t \) depends on \( r_t(\tau) \), concentrating upon the five yields mentioned earlier. Evidence of the effect can be marshalled in a number of ways. By far the simplest approach is to plot \( (\Delta r_t(\tau) - \mu)^2 \) against \( r_{t-1}(\tau) \), (and this is done in Fig. 1 for \( r_1(1) \)).3 The evidence of a levels effect looks very strong. A more structured approach is to estimate the parameters of a diffusion process for yields of the form

\[
dr_t = (\alpha_t - \beta_t r_t)dt + \tau_t^n d\eta_t,
\]

and to examine the estimate of \( \gamma_t \). To estimate this requires some approximation scheme. Chan et al. (1992) consider a discretization based on the Euler scheme with \( h = 1 \) (\( ht \) being the discretized steps) producing

\[
\Delta r_t = \alpha_1 - \beta_1 r_{t-1} + \sigma_{\tau^n_t} \epsilon_t,
\]

where here and in the remainder of the paper \( \epsilon_t \) is n.i.d.(0,1).

Equation (2) can be estimated by OLS simply by defining the dependent variable as \( \Delta r_t \tau^n_{t-1} \), while the regressors become \( x_t = [r^n_{t-1} - r^n_{t-1}(\tau)] \), as the error term is then \( \sigma \epsilon_t \), which is n.i.d \((0, \sigma^2)\). Because the conditional mean for \( r_t \) depends only on \( \alpha_1, \beta_1 \), while the conditional variance of \( u_t = r_t - E_{t-1}(r_t) \) is \( \sigma^2 \tau^n_{t-1} \), which does not involve these parameters, we could estimate the parameters in the following way.

1. Regress \( \Delta r_t \) on 1 and \( r_{t-1} \) to get \( \hat{\alpha}_1 \) and \( \hat{\beta}_1 \).

---

2Throughout the paper we will take the term structure data as corresponding to actual observations. In practice this may not be so, as a complete term structure is interpolated from observations on parts of the curve. This may introduce some biases of unknown magnitude into relationships between yields. McCulloch and Kwon (1993) interpolate with spline functions. Others e.g. Gourieroux and Scaillet (1994) actually utilize some of the factor models discussed later to suggest forms for the yield curve that may be used for interpolation.

3Marsh and Rosenfeld (1983) also did this and commented on the relation.
2. Since

\[ E_{t-1}[u_t^2] = \sigma^2 r_{t-1}^{2\gamma_1}, \]  

then

\[ u_t^2 = \sigma^2 r_{t-1}^{2\gamma_1} + \nu_t, \]  

where \( E_{t-1}(\nu_t) = E[u_t^2 - E_{t-1}(u_t^2)] = 0 \). Hence we can estimate \( \gamma_1 \) by using a non-linear regression program.

3. We can re-estimate \( \alpha_1, \beta_1 \) by then doing a weighted regression of \( \Delta r_t r_{t-1}^{-\gamma_1} \) against \( r_{t-1}^{-\gamma_1} \) and \( r_{t-1}^{\gamma_1} \).

The above steps would produce a maximum likelihood estimator if \( \epsilon_t \) was taken to be \( N(0,1) \) and the estimation of \( \gamma_1 \) was done by a weighted non-linear regression on (3) using the conditional standard deviation of \( \nu_t \) as weights.\(^4\) Chan et al. (1992) use a GMM estimator, which jointly estimates \( \alpha_1, \beta_1, \gamma_1 \) and \( \sigma \) from the set of moments

\[ E(\epsilon_t) = 0, E(r_{t-1}\epsilon_t) = 0, E(\nu_t) = 0, E(r_{t-1}\nu_t) = 0. \]

Their estimator would coincide with the one described above if the last moment condition was replaced by \( E(r_{t-1}\nu_t) = 0 \). A potential problem with all the estimators is that, if \( \beta_1 \) is likely to be close to zero, the regressors in (2) and (4) will be close to \( I(1) \), and so non-standard distribution theory almost certainly applies to the GMM estimator.

Table 4 presents estimates of the parameters \( \alpha_1, \beta_1 \) and \( \gamma_1 \) found by using three estimation methods. The first one is based on estimating the diffusion with an Euler approximation,

\[ \Delta r_{th} = \alpha_1 h - \beta_1 h r_{(t-1)h} + \sigma h^{1/2} r_{(t-1)h}^{\gamma_1} \epsilon_t, \]  

with \( h = 1 \). It is the estimator described above as GMM. The others stem from the modern approach of indirect estimation proposed by Gouriéroux et. al. (1993) and Gallant and Tauchen (1992). In these methods one simulates \( K \) multiple sets of observations \( r_{th}^k \) \((k = 1, ..., K)\) from (5), with given values of \( h \) (we use 1/100) and \( \theta' = (\alpha_1, \beta_1, \gamma_1, \sigma^2) \), and then finds the estimates of \( \theta \) that set \( \sum_{t=1}^T \{K^{-1} \sum_{k=1}^K d_\phi(r_{th}^k; \hat{\phi})\} \) to zero, where \( \hat{\phi} \) is an estimator of the parameters of some auxiliary model found by solving \( \sum_{t=1}^T d_\phi(r_t; \phi) = 0 \).\(^5\) The logic of the estimator is that, if the model (5) is true, then \( \hat{\phi} \rightarrow \phi^* \), where \( E[d_\phi(r_t; \phi^*)] = 0 \), and the term in curly brackets estimates this expectation by simulation. Consistency and asymptotic normality of the indirect estimator follows from the properties of \( \hat{\phi} \) under mis-specification. It is important to note that the auxiliary model need not be correct, but it should be a good representation of the data, otherwise the indirect estimator will be very inefficient. We use two auxiliary models and, in each instance, \( d_\phi \) are the scores for \( \phi \) from those models. The first is (5) with \( h = 1 \) and \( \epsilon_t \) being assumed \( n.i.d.(0,1)(MLE) \), while the second has \( r_t \) being an AR(1) with EGARCH(1,1) errors. The visual evidence of Figure 1 is strongly supported by the estimated parametric models,

---

\(^4\)Frydman (1994) argues that the distribution of the MLE of \( \beta_1 \) is non-standard when \( \gamma_1 = 1/2 \) and there is no drift.

\(^5\)A Mihlestein (1974) rather than Euler approximation of (5) was also tried, but there were very minor differences in the results.
Table 4: Estimates of Diffusion Process Parameters

<table>
<thead>
<tr>
<th></th>
<th>( r_t(1) )</th>
<th>( r_t(3) )</th>
<th>( r_t(6) )</th>
<th>( r_t(9) )</th>
<th>( r_t(120) )</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>GMM</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \alpha_1 )</td>
<td>.106</td>
<td>.090</td>
<td>.089</td>
<td>.091</td>
<td>.046</td>
</tr>
<tr>
<td></td>
<td>(2.19)</td>
<td>(1.82)</td>
<td>(1.80)</td>
<td>(1.87)</td>
<td>(1.77)</td>
</tr>
<tr>
<td>( \beta_1 )</td>
<td>.020</td>
<td>.015</td>
<td>.015</td>
<td>.015</td>
<td>.006</td>
</tr>
<tr>
<td></td>
<td>(1.52)</td>
<td>(1.24)</td>
<td>(1.25)</td>
<td>(1.31)</td>
<td>(.98)</td>
</tr>
<tr>
<td>( \gamma_1 )</td>
<td>1.351</td>
<td>1.424</td>
<td>1.532</td>
<td>1.516</td>
<td>1.178</td>
</tr>
<tr>
<td></td>
<td>(6.73)</td>
<td>(5.61)</td>
<td>(4.99)</td>
<td>(5.12)</td>
<td>(9.80)</td>
</tr>
<tr>
<td><strong>MLE</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \alpha_1 )</td>
<td>.071</td>
<td>.047</td>
<td>.041</td>
<td>.037</td>
<td>.015</td>
</tr>
<tr>
<td></td>
<td>(2.17)</td>
<td>(2.48)</td>
<td>(3.00)</td>
<td>(3.82)</td>
<td>(3.74)</td>
</tr>
<tr>
<td>( \beta_1 )</td>
<td>.012</td>
<td>.007</td>
<td>.005</td>
<td>.004</td>
<td>-.001</td>
</tr>
<tr>
<td></td>
<td>(.74)</td>
<td>(7.89)</td>
<td>(2.08)</td>
<td>(1.08)</td>
<td>(2.35)</td>
</tr>
<tr>
<td>( \gamma_1 )</td>
<td>.583</td>
<td>.648</td>
<td>.694</td>
<td>.753</td>
<td>1.136</td>
</tr>
<tr>
<td></td>
<td>(2.39)</td>
<td>(1.92)</td>
<td>(2.31)</td>
<td>(3.34)</td>
<td>(19.30)</td>
</tr>
<tr>
<td><strong>EGARCH</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \alpha_1 )</td>
<td>.107</td>
<td>.044</td>
<td>.045</td>
<td>.043</td>
<td>.009</td>
</tr>
<tr>
<td></td>
<td>(1.63)</td>
<td>(1.89)</td>
<td>(4.34)</td>
<td>(1.67)</td>
<td>(3.30)</td>
</tr>
<tr>
<td>( \beta_1 )</td>
<td>-.004</td>
<td>-.010</td>
<td>-.008</td>
<td>-.008</td>
<td>-.009</td>
</tr>
<tr>
<td></td>
<td>(2.15)</td>
<td>(1.57)</td>
<td>(2.24)</td>
<td>(2.04)</td>
<td>(2.36)</td>
</tr>
<tr>
<td>( \gamma_1 )</td>
<td>.838</td>
<td>.974</td>
<td>.947</td>
<td>.941</td>
<td>1.104</td>
</tr>
<tr>
<td></td>
<td>(2.67)</td>
<td>(5.73)</td>
<td>(7.21)</td>
<td>(3.09)</td>
<td>(4.88)</td>
</tr>
</tbody>
</table>

Asymptotic t-ratios in parentheses

although there is considerable diversity in the estimates obtained. Perhaps the most interesting aspect of the table is the fact that \( \gamma_1 \) tends to increase with maturity. Based on the evidence from the indirect estimators, \( \gamma_1 = 1/2 \) seems a reasonable choice for the shortest maturity, which would correspond to the diffusion process used by Cox et al. (1985).

A problem in simply fitting a model with a "levels" effect is that the observed conditional heteroskedasticity in the data might be better accounted for by a GARCH process, and so the appropriate questions should either be whether there is evidence of a levels effect after removing a GARCH process, or, whether a levels representation fits the data better than a GARCH model does. To shed some light on these questions, our strategy was to fit augmented EGARCH(1,1) models to \( \Delta r_t(\tau) = \mu + \sigma_{r_t} \epsilon_{r_t} \), \( \epsilon_{r_t} \sim \mathcal{N}(0,1) \), of the form

\[
\log \sigma_{r_t}^2 = a_0 + a_1 \log \sigma_{r_{t-1}}^2 + a_2 \epsilon_{r_{t-1}} + a_3 \left( \epsilon_{r_{t-1}} - \mathcal{N} \right) + \delta r_{t-1}(\tau). \tag{6}
\]

This specification is used to generate a diagnostic test for the presence of a levels effect, and is
Table 5: δ and t Ratios for Levels Effect

<table>
<thead>
<tr>
<th>δ</th>
<th>0.050</th>
<th>0.025</th>
<th>0.023</th>
<th>0.021</th>
<th>0.019</th>
</tr>
</thead>
<tbody>
<tr>
<td>t</td>
<td>3.73</td>
<td>3.51</td>
<td>3.42</td>
<td>3.04</td>
<td>2.42</td>
</tr>
</tbody>
</table>

not intended to be a good representation of the actual volatility. Hence the t-statistic for testing if δ is zero can be regarded as a valid test for more general specifications, e.g. \( \delta g(\tau_{t-1}(\tau)) \), where \( g(\cdot) \) is some function, provided \( \tau_{t-1}(\tau) \) is correlated with \( g(\tau_{t-1}(\tau)) \). Table 5 gives the estimates of δ and the associated t ratios. Every yield displays a levels effect, although with the 10 year maturity it seems weaker. The same conclusion applies to the spreads between forward rates, \( Fp_t(\tau, \tau-1) \). Fitting EGARCH(1,1) models to these series for \( \tau = 1, 3, 6 \) and 9 months maturity, and allowing the levels effect to be a function of \( F_{t-1}(\tau) \), the t-ratios that this coefficient was zero were 3.85, 3.72, 17.25 and 12.07 respectively.

A number of studies have appeared that look at this phenomenon. Apart from our own work, Chan et al. (1992), Broze et al. (1993), Koedijk et al. (1994), and Brenner et al. (1994) have all considered the question, while Vetzal (1992) and Kearns (1993) have tried to allow for stochastic volatility, i.e. \( \sigma_t^2 \) is not only a function of the past history of yields. To date no formal comparison of the different models is available, unlike the situation for stock returns e.g. Gallant et al. (1994). All studies find strong evidence for a levels effect on volatility. Brenner et al. provide ML estimates of the parameters of a discretized joint GARCH/levels model in which the volatility function, \( \sigma_t^2 \), is the product of a GARCH(1,1) process and a levels effect i.e. \( \sigma_t^2 = (a_0 + a_1\sigma_{t-1}^2 + a_2\tau_{t-1})\tau_{t-1} \). The estimated value of \( \gamma \) falls to around 0.5, but remains highly significant. Koedijk et al. (1993) have a similar formulation except that \( \sigma_t^2 \) is driven by \( \tau_{t-1} \) rather than \( \sigma_{t-1}^2 \). Again \( \gamma \) is reduced but remains highly significant.

One might question the use of conventional significance levels for the "raw" t ratios, owing to the fact that one of the regressors is a near-integrated process. To examine the effects of this we simulated data from an estimated model, equation (6) for \( \tau_{t}(1) \), treating the estimates obtained by MLE estimation as the true parameter values, and then found the distribution of the t ratio for the hypothesis that \( \delta = 0 \) using the MLE, constructed by taking one step from the true values of the coefficients (this would be a simulation of the asymptotic distribution). The results indicate that the distribution of the t-ratio has fatter tails than the normal with critical values for two tailed tests of 2.90 (5%) and 2.41 (10%), but use of these would not change the decisions.

---

6. It is interesting to observe that the distribution of the Dickey-Fuller test is very sensitive to whether there is a levels effect or not. To see this we simulated a model in which \( \Delta \tau_t = .001 + .01\tau_{t-1}^2\varepsilon_t \), where \( \varepsilon_t \sim iid(0,1) \) and \( \gamma \) either took the value of zero or unity. A small drift was added, although its influence upon the distribution is likely to be small. The simulated critical values for 1%, 2.5% and 5% significance levels when \( \gamma = 0,1 \) are (-3.14, -6.41), (-2.71, -4.97) and (-2.39, -4.03) respectively. Clearly, the presence of a levels effect in volatility means that the critical values are much larger (in absolute terms), strengthening the claim that Table 1 suggests a unit root in yields.
2.2 Multivariate Properties

2.2.1 The Level of the Yield Curve

As was mentioned in the introduction a great deal of work on the term structure views yields as being driven by a set of $M$ factors

$$
\tau_t(\tau) = \sum_{j=1}^{M} \beta_{j\tau} \xi_{j,t},
$$

and it is important to investigate whether this is a reasonable characterization of the data. It is useful here to recognise that the modern econometrics literature on multivariate relations admits just such a parameterization. Suppose the yields are collected into an $(n \times 1)$ vector $y_t$ and that it is assumed that $y_t$ can be represented as a VAR. Then, if $y_t$ are $I(1)$ and, in the $n$ yields there are $k$ co-integrating vectors, Stock and Watson (1988) showed this to mean that the yields can be described in the format

$$
y_t = J\xi_t + u_t
$$

where $\xi_t$ are the $n - k$ common trends to the system, and $E_{t-1}u_t = 0$. The format (8) is commonly referred to as the Beveridge-Nelson-Stock-Watson (BNSW) representation. If there are $(n - 1)$ co-integrating vectors, there will be a single common factor, $\xi_{1t}$, that determines the level of the yields. How the yields relate to one another is governed by $y_t - J\xi_{1t} = u_t$ i.e. the yield curve is
Table 6: Tests for Cointegration Amongst Yields

<table>
<thead>
<tr>
<th></th>
<th>Max</th>
<th>Crit. Value (.05)</th>
<th>Tr.</th>
<th>Crit. Value (.05)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 vs 4 trends</td>
<td>273.4</td>
<td>33.5</td>
<td>586.9</td>
<td>68.5</td>
</tr>
<tr>
<td>4 vs 3 trends</td>
<td>184.7</td>
<td>27.1</td>
<td>313.5</td>
<td>47.2</td>
</tr>
<tr>
<td>3 vs 2 trends</td>
<td>95.6</td>
<td>21.0</td>
<td>128.8</td>
<td>29.6</td>
</tr>
<tr>
<td>2 vs 1 trends</td>
<td>30.9</td>
<td>14.1</td>
<td>33.3</td>
<td>15.4</td>
</tr>
<tr>
<td>1 vs 0 trends</td>
<td>2.4</td>
<td>3.8</td>
<td>2.4</td>
<td>3.8</td>
</tr>
</tbody>
</table>

a function of \( u_t \).

Johansen's (1988) tests for the number of co-integrating vectors may be applied to the data described earlier. Table 6 provides the two most commonly used — the maximum eigenvalue (Max) and trace (Tr) tests — for the five yields under investigation, and assuming a VAR of order one.\(^7\) From this table there appears to be four co-integrating vectors, i.e. a single common trend. Johnson (1994), Engsted and Tanggaard (1994) and Hall et al. (1992) reach the same conclusion. Zhang (1993) argues that there are three common trends but Johnson shows that this is due to Zhang's use of a mixture of yields from zero and non-zero coupon bonds.

What is the common trend? There is no unique answer to this. One solution is to find a yield that is determined outside of the system, as that will be the driving force. For a small country, that rate is likely to be the "world interest rate", which in practice either means a Euro-Dollar rate or some combination of the US, German and Japanese interest rates. Another candidate for the common trend is the simple average of the rates.\(^8\) In any case we will take this to be the first factor \( \xi_1 \) in (7).

2.2.2 The Shape of the Yield Curve

The existence of \( k \) co-integrating vectors \( \alpha \) (\( \alpha \) is an \( (n \times k) \) matrix), such that \( \zeta_t = \alpha' y_t \) is \( i(0) \), means that any VAR in \( y_t \) has the ECM format

\[
\Delta y_t = \gamma \zeta_{t-1} + D(L)\Delta y_{t-1} + e_t, \tag{9}
\]

where \( E_{t-1}(e_t) = 0 \) and \( D(L) \) is a polynomial in the lag operator. It is also possible to show that \( u_t \) in (8) can be written as a function of the \( k \) EC (error correction) terms \( \zeta_t \) and this suggests that we might take these to be the remaining factors \( \xi_{j} \) (\( j = 2, \ldots, K \)) in (7). To make the following discussion more concrete assume that the expectations theory of the term structure

\(^7\)Changing this order to four does not affect any conclusions, but restricting it to unity fits in better with the theoretical discussion.

\(^8\)See Gonzalo and Granger (1991) for other alternatives.
holds i.e. a \( \tau \) period yield is the weighted average of the expected one period yields into the future. In the case of discount bonds the weights are equal to \( \frac{1}{\tau} \) so that the theory states

\[
\tau_t(\tau) = \frac{1}{\tau} \sum_{k=0}^{\tau-1} E_t(\tau_{t+k}(1)).
\]

Of course this is an hypothesis, albeit one that seems quite sensible. It implies that

\[
\tau_t(\tau) - \tau_t(1) = \left\{ \frac{1}{\tau} \sum_{k=0}^{\tau-1} E_t\tau_{t+k}(1) - E_t\tau_t(1) \right\}
\]

\[
= \left\{ \frac{1}{\tau} \sum_{i=1}^{\tau-1} \sum_{j=1}^{\tau-i} E_t\Delta \tau_{t+j}(1) \right\}.
\]

Now, if the yields are \( I(1) \) processes, the yield spread \( \tau_t(\tau) - \tau_t(1) \) should be \( I(0) \), i.e. \( \tau_t(\tau) \) and \( \tau_t(1) \) should be co-integrated with co-integrating vector \([1 - 1]\), and these spreads are the EC terms. Therefore, to test the expectations hypothesis for the five yields we need to test if the matrix of co-integrating vectors has the form

\[
\alpha' = \begin{bmatrix}
-1 & 1 & 0 & 0 \\
-1 & 0 & 1 & 0 \\
-1 & 0 & 0 & 0 \\
-1 & 0 & 0 & 1
\end{bmatrix}
\]

Johansen's (1988) test for this gives a \( \chi^2(4) \) of 36.8, leading to a very strong rejection of the hypothesis. Such an outcome has also been observed by Hall et al. (1992), Johnson (1994) and Engsted and Tanggaard (1994).

A number of possible explanations for the rejection were canvassed in those papers, involving the size of the test statistic being incorrect etc. One's inclination is to examine the estimated matrix of co-integrating vectors given by Johansen's procedure, \( \hat{\alpha} \), and to see how closely these correspond to the hypothesized values but, unfortunately, the vectors are not unique and the estimated quantities will always be linear combinations of the true values. Some structural information is needed to recover the latter, and to this end we write \( \alpha' = A\hat{\alpha} \), where

\[
\hat{\alpha} = \begin{bmatrix}
-\beta_3 & 1 & 0 & 0 \\
-\beta_6 & 0 & 1 & 0 \\
-\beta_9 & 0 & 0 & 0 \\
-\beta_{120} & 0 & 0 & 1
\end{bmatrix}
\]

and then proceed to solve the equations \( \hat{\alpha} = \hat{A}\hat{\alpha} \), where \( \hat{A} \) is some non-singular matrix. This produces \( \beta_3 = 1.038, \beta_6 = 1.063, \beta_9 = 1.075 \) and \( \beta_{120} = 1.076 \), which indicates that the point estimates are quite close to those predicted by the expectations theory. It is also possible to estimate the \( \beta \) by "limited information" rather than "full-information" methods. To that end the Phillips-Hansen (1990) estimator was adopted with a Parzen kernel and eight lags being used to form the long-run covariance matrices, producing \( \hat{\beta}_3 = 1.021, \hat{\beta}_6 = 1.034, \hat{\beta}_9 = 1.034 \) and \( \hat{\beta}_{120} = 0.91 \). With the exception of the 10 year rate, neither set of estimates seems to be greatly divergent from that predicted.
Some insight into why the rejection occurs may be had from (9). Given that co-integration has been established, and working with a VAR(1), i.e. $D(L) = 0$ in (9), the change in each yield should be governed by

$$\Delta r_t(\tau) = \sum_{j=2}^{5} \gamma_{j\tau}(r_{t-1}(j) - \beta_j r_{t-1}(1)) + e_{rt},$$  \hspace{1cm} (11)

where $j = 2, \ldots, 5$ maps one to one into the elements $\tau = 3, 6, 9, 120$. If the expectations theory is valid $\beta_j = 1$ and the system becomes

$$\Delta r_t(\tau) = \sum_{j=2}^{5} \gamma_{j\tau}(r_{t-1}(j) - r_{t-1}(1)) + e_{rt},$$

and the hypothesis $H_0 : \beta_j = 1$ can be tested by computing the likelihood ratio statistic. It is well known that such a test will be distributed as a $\chi^2(4)$ under the null hypothesis.

If the yields were taken to be $I(0)$, the simplest way to test if $\beta_j = 1$ would be to re-write (11) as

$$\Delta r_t(\tau) = \sum_{j=2}^{5} \gamma_{j\tau}(r_{t-1}(j) - r_{t-1}(1)) + \left( \sum_{j=2}^{5} \gamma_{j\tau}(1 - \beta_j) \right) r_{t-1}(1) + e_{rt},$$  \hspace{1cm} (12)

and to test if the coefficient of $r_{t-1}(1)$ in each of the equations for $\Delta r_t(\tau)$ was zero. For a number of reasons this does not reproduce the $\chi^2(4)$ test cited above - there are five coefficients being tested and $r_{t-1}(1)$ will be $I(1)$, making the distribution non-standard. Nevertheless, the separate single equation tests might still be informative. In this case the $t$-values that $r_{t-1}(1)$ has a zero coefficient in each equation were -4.05, -1.77, -1.72, -0.24 and 0.55 respectively, suggesting that the rejection of (10) lies in the behaviour of the one month rate i.e. the spreads are not capable of fully accounting for its movement. Engsted and Tanggaard (1994) also reach this conclusion. It may be that $r_{t-1}(1)$ is proxying for some omitted variable, and the literature has in fact canvassed the possibility of non-linear effects upon the short-term rate. Anderson (1994) makes the influence of spreads upon $\Delta r_t(1)$ non-linear, while Pfann et al. (1994) take the process driving $r_t(1)$ to be a non-linear autoregression - in particular, the latter allow for a number of regimes according to the magnitude of $r_t(1)$, with some of these regimes featuring $I(1)$ behaviour of the rate while others do not. Another possibility, used in Conley et al. (1994) is that the “drift term” in a continuous time model has the form $\sum_{j=m}^{m} a_j r_t^j$ and this would induce a non-linearity into the relation between $\Delta r_t$ and $r_{t-1}$.

Instead of a mis-specification in the mean, rejection of (10) may be due to levels effects in $e_{rt}$. As noted earlier, the Dickey-Fuller test critical values are very sensitive to this effect, and the test that $r_{t-1}$ has a zero coefficient in the $\Delta r_t(1)$ equation in (12) is actually an ADF test, if the augmenting variables are taken to be the spreads. This led us to produce a small Monte Carlo simulation of Johansen’s test for (10) under different assumptions about levels effects in the errors of the VAR. The example is a simplified version of the system above featuring only two variables $y_{1t}$ and $y_{2t}$ with co-integrating vector $\begin{bmatrix} 1 & -1 \end{bmatrix}$ and being generated from the vector ECM.
\[ \Delta y_{1t} = -0.8(y_{1t-1} - y_{2t-1}) + 0.1y_{1t-1}^2 \epsilon_{1t} \]
\[ \Delta y_{2t} = -0.1(y_{1t-1} - y_{2t-1}) + 0.1y_{1t-1}^2 \epsilon_{2t}. \]

The 95\% critical value for Johansen’s test that the co-integrating vector is the true one varies according to the value of \( \gamma \): 3.95(\( \gamma = 0 \)), 4.86(\( \gamma = .5 \)), 5.87(\( \gamma = .6 \)), 11.20(\( \gamma = .8 \)) and 23.63(\( \gamma = 1 \)). Clearly, there is a major impact of the levels effect upon the sampling distribution of Johansen’s test, and the phenomenon needs much closer investigation, but it is conceivable that rejection of (10) may just be due to the use of critical values that are too small.

Even if one rejects the co-integrating vectors predicted by the expectations theory, the evidence is still that there are \( k = n - 1 \) error correction terms. It is natural to equate the remaining \( M - 1 \) factors in (7) (after elimination of the common trend) with these EC terms, but this is not very helpful, as it would mean that \( M = n \), i.e. the number of factors would equal the number of yields. Hall et al. (1992) provide an example of forecasting the term structure using the ECM relation (9), imposing the expectations theory co-integrating vectors to form \( \zeta_t \), and then regressing \( \Delta y_t \) on \( \zeta_{t-1} \) and any lags needed in \( \Delta y_t \). Hence their model is equivalent to using a single factor, the common trend, to forecast the level, and \( (n - 1) \) factors to forecast the slope (the EC or spread terms). In practice however they impose the feature that some of the coefficients in \( \gamma \) were zero, i.e. the number of factors determining the yield varies with the maturity being examined. It is interesting to note that their representation for \( \Delta y_t \) (4) has no EC terms i.e. it is effectively determined outside the system and plays the role of the “world interest rate” mentioned earlier.

In an attempt to reduce the number of non-trend factors below \( n - 1 \), it is tempting to assume that (say) only \( m = M - 1 \) of the \( (n - 1) \) terms in \( \zeta_t \) appear as determinants of \( \Delta y_t(\gamma) \) and that these constitute the requisite common factors, but such a restriction would necessitate \( m \) of the columns of \( \gamma \) being zero, thereby violating the rank condition, \( \rho(\gamma) = n - 1 \). Consequently the factors will need to be combinations of the EC terms. Now, pre-multiplying (7) by \( \alpha' \) gives

\[ \alpha'y_t = \alpha' \sum_{j=1}^{M} b_j \xi_{jt}, \quad (13) \]

where \( b_j = [\beta_{j1} \ldots \beta_{jn}] \) is a \( 1 \times n \) vector. If we designate the first factor as the common trend then it must be that \( \alpha'b_1 = 0 \) as the LHS is \( I(0) \) by construction, meaning that

\[ \zeta_t = \alpha' \sum_{j=2}^{K} b_j \xi_{jt} = \alpha'B \Xi_t, \quad (14) \]

where \( \Xi_t \) is the \( (K - 1) \times 1 \) vector containing \( \xi_{2t} \ldots \xi_{Kt} \), and \( B \) is an \( n \times (K - 1) \) matrix with \( \rho(B) = K - 1 \), where \( \rho(\cdot) \) designates rank.

Equation (14) enables us to draw a number of interesting conclusions. Firstly, \( \rho(cov(\zeta_t)) = \min(\rho(\alpha), \rho(B)) \), provided \( cov(\Xi_t) \) has rank \( K - 1 \). Since \( K \leq n \) implies \( K - 1 \leq n - 1 \), it must be that \( \rho(B) \leq \rho(\alpha) \), and therefore \( \rho(cov(\zeta_t)) = K - 1 \) i.e. the number of factors in the term structure (other than the common trend) may be found by examining the rank of the covariance matrix.
of the co-integrating errors. Secondly, since \( C = \alpha'B \) has \( \rho(C) = K - 1, F_t = (C'C)^{-1}C' \), and hence the factors will be linear combinations of the EC terms. Applying principal components to the data set composed of spreads \( s_p(t,3), s_p(t,6), s_p(t,9) \) and \( s_p(t,120) \), the eigenvalues of the covariance matrix are 4.1, .37, .02 and .002, pointing to the fact that these four spreads can be summarized very well by three components (at most).\(^9\) The three components are:

\[
\begin{align*}
\phi_{1t} &= .32 s_p(t,3) - .86 s_p(t,6) - .37 s_p(t,9) + .17 s_p(t,120) \\
\phi_{2t} &= -.78 s_p(t,3) + .00 s_p(t,6) - .55 s_p(t,9) + .29 s_p(t,120) \\
\phi_{3t} &= .54 s_p(t,3) + .52 s_p(t,6) - .58 s_p(t,9) + .37 s_p(t,120).
\end{align*}
\]

3 Models of the Term Structure

In this section we describe some popular ways of modelling the term structure. In order to assess whether these models are capable of replicating observed term structures, it is necessary to decide on some way to compare them to the data. There is a small literature wherein formal statistical tests have been performed on how well the models replicate the data in some designated dimension. Generally, however, the reasons for any rejection of the models remain unclear, as many characteristics are being tested at the one time. In contrast, this chapter uses the method of "stylized facts", i.e. it seeks to match up the predictions of the model with the nature of the data as summarized in section 2. Thus, we look at whether the models predict that yields are near-integrated, have levels effects in volatility, exhibit specific integrating vectors, produce persistence in spreads, and would be compatible with two or (at most) three factors in the term structure.\(^10\)

3.1 Solutions from the Consumer's Euler Equations

Consider a consumer maximising expected utility over a period subject to a budget constraint, i.e.

\[
\max_{C_t} E_t \left[ \sum_{s=t}^{\infty} U(C_s) \beta^s \right],
\]

where \( \beta \) is a discount factor, and \( C_s \) is consumption at time \( s \). It is well known that a first order condition for this is

\(^9\)The principal components approach, or variants of it, has been used in a number of papers — Litterman and Scheinkman (1991), Dybvig (1989) and Egginton and Hall (1993). This technique finds linear combinations of the yields such that the variance of each combination is as small as possible. Thus the \( i^{th} \) principal component of \( y_t \) will be \( b_i y_t \), where \( b_i \) is a set of weights. Because one could always multiply through by a scale factor the \( b_i \) are normalized, i.e. \( b_i b_i' = 1 \). With this restriction \( b \) becomes the eigenvectors of \( \text{var}(y_t) \). Since \( b_i \) is an eigenvector it is clear that \( b' \text{var}(y_t) b = \Lambda \), where \( \Lambda \) is a diagonal matrix with the eigenvalues \( (\Lambda_1, \ldots, \Lambda_n) \) on it, and that \( tr[b' \text{var}(y_t) b] = \sum_{i=1}^{n} \Lambda_i \). It is conventional to order the components according to the magnitude of \( \Lambda_i \); the first principal component having the largest \( \Lambda_1 \). There is a connection between principal components and common trends. Both seek linear combinations of \( y_t \) and, in many cases, one of the components can be interpreted as the common trend, e.g. in Egginton and Hall (1993) the first component is effectively the average of the interest rates, which we have mentioned as a possible common trend earlier.

\(^10\)There are many other characteristics of these yields that we ignore in this paper but which are challenging to explain e.g. the extreme leptokurtosis in the density of the change in yields and in the spreads.
where $V_t$ is the value of an asset (or portfolio) in terms of consumption goods. This can be re-arranged to give

$$E_t \left[ \frac{V_s}{V_t} \rho^{s-t} U'(C_s) / U'(C_t) \right] = 1. \quad (15)$$

Assuming that the asset is a discount bond, and the general price level is fixed, consider setting $s = t + \tau$ giving $V_t = f_t(\tau)$. The solution of this equation will then provide a complete set of discount bond prices for any maturity. It is useful to re-express (15) as

$$f_t(\tau) = E_t \left[ \rho^{t+\tau} U'(C_{t+\tau}) / U'(C_t) \right], \quad (16)$$

imposing the restriction that $f_t(t + \tau) = 1$, so as to find the price of a zero coupon bond paying $\$1$ at maturity. Hence the term structure would then be determined. If the price level is not fixed (16) needs to be modified to

$$f_t(\tau) = E_t \left[ \rho^{t+\tau} P_t U'(C_{t+\tau}) / U'(C_t) P_{t+\tau} \right], \quad (17)$$

where $P_t$ is the price level at time $t$.

There have been a few attempts to price bonds from (16) or (17). Canova and Marrinan (1993) and Boudoukh (1993) do this by assuming that $C_t = \log(C_{t+1}/C_t) - 1$ and $P_t = \log(P_{t+1}/P_t) - 1$, follow a VAR process with some volatility in the errors, and that the utility function has the CRAA form, $U(C_t) = \frac{C_t^{1-\gamma}}{1-\gamma}$, where $\gamma$ is the coefficient of risk aversion.\(^\dagger\)

It is necessary to evaluate (17) for the yield $\tau_t(\tau) = -\tau^{-1} \log f_t(\tau)$.

$$\tau_t(\tau) = -\frac{1}{\tau} \log E_t \left[ \beta^{t+\tau} (C_{t+\tau}/C_t)^{-\gamma} (P_t/P_{t+\tau}) \right]$$

$$= -\frac{1}{\tau} \log E_t \left[ \beta^{t+\tau} (1 + c_{t+\tau})^{-\gamma} (1 + p_{t+\tau})^{-1} \right],$$

where

$$c_{t+\tau} = C_{t+\tau}/C_t - 1 \simeq \log C_{t+\tau} - \log C_t,$$

$$p_{t+\tau} = P_{t+\tau}/P_t - 1 \simeq \log P_{t+\tau} - \log P_t.$$
Equation (19) points to a four factor model of the term structure with the level being driven by the first two conditional moments of the inflation rate and consumption growth. However, the relation is not easily interpreted as a linear one, since the weights attached to volatilities are functions of the conditional means.

The problem remains to evaluate the conditional moments. To complete the model it is necessary to assume something about the evolution of $z_{1t}$ and $z_{2t}$. These are generally taken to be AR processes of the form

$$z_{1t} = \Phi_{0j} + \Phi_{1j} z_{1t-1} + \epsilon_{1t}.$$  

Canova and Marrinan (1993) take $\sigma_{z_{1t+1}}^2 = \text{var}_t(e_{1t+1})$ to be GARCH processes of the form

$$\sigma_{z_{1t+1}}^2 = a_0 + a_1 \sigma_{z_{1t}}^2 + a_2 \epsilon_{1t}^2,$$

whereby the formulae in Baillie and Bollerslev (1992) can be used to evaluate $E_t(z_{1t})$ and $\text{var}_t(z_{1t})$, while Boudoukh (1993) has $\sigma_{z_{1t}}^2$ as a stochastic volatility process. For GARCH models $\text{var}_t(z_{1t})$ is a linear function of $\sigma_{z_{1t+1}}^2$.

How well does this model perform in replicating the stylized facts of the term structure? To produce a near unit root in yields it is necessary that $\log(1 + E_t(p_{1t})) \sim E_t(p_{1t})$ be near integrated i.e. inflation must be a near integrated process, as it is the only one of the two series that has such persistence in either mean or variance — see Boudoukh (1993) for a description of the time series properties of the two series. Then the inflation rate becomes the common trend in the term structure, and the spreads will depend upon consumption growth and the two volatilities. As there is rather weak evidence for much dependence in either inflation or consumption volatility - see the test statistics in Boudoukh - it is difficult to see the persistence in spreads being explained.
by these models. Whether a levels effect in $\Delta r_t(\tau)$ can be produced is unclear; the GARCH structures used by Canova and Marrinan will not produce it, but Boudoukh's stochastic volatility formulation does allow for a levels effect in $\text{var}_t(p_t)$. Moreover, even if volatilities were constant, the conditional means enter the weights attached to them, and this dependence might be used to induce a levels effect into $\Delta r_t(\tau)$. Whilst $E_t(c_t)$ is likely to be close to a constant due to the weak autocorrelation in consumption growth, there is strong serial correlation in inflation rates, and, with inflation as the common trend, it is conceivable that the requisite effect could be found in that way, although the question was not addressed by the authors.

Another attempt at working within this framework is Constantinides (1992) who writes (17) as

$$f_t(\tau) = E_t[K_{t+\tau}/K_t],$$

where $K_t = \beta^tU'(C_t)/P_t$ is referred to as a "pricing kernel". He then makes assumptions about the evolution of $K_t$, in particular that

$$K_t = \exp \left\{ -(g + \frac{\sigma^2}{2})t + z_{0t} + \sum_{i=1}^{N}(z_{it} - \alpha_i)^2 \right\}.$$

He works in continuous time and makes $z_{0t}$ a Weiner process while the other $z_{it}$ are Ornstein-Uhlenbeck diffusion processes with parameters $\lambda_i$ and variances $\sigma_i^2$. Each of the $z_{it}$ are taken to be independent. Under these assumptions it turns out that

$$f_t(\tau) = \left\{ \prod_{i=1}^{N} H_i(\tau) \right\}^{-1/2} \exp \left\{ (-g + \sum_{i=1}^{N} \lambda_i)\tau + \sum_{i=1}^{N} H_i^{-1}(\tau)(z_{it} - \alpha_i e^{\lambda_i \tau})^2 - \sum_{i=1}^{N}(z_{it} - \alpha_i)^2 \right\},$$

where $H_i(\tau) = \sigma_i^2/\lambda_i + (1 - \sigma_i^2/\lambda_i)e^{2\lambda_i \tau}$. Consequently, $r_t(\tau)$ has the format

$$r_t(\tau) = \delta_0 + \sum_{i=1}^{N} \delta_i(\tau)(z_{it} - \alpha_i e^{\lambda_i \tau})^2 + \sum_{i=1}^{N} \tau^{-1}(z_{it} - \alpha_i)^2.$$
3.2 One Factor Models From Finance

Finance theory has developed by working with factor models to determine the term structure. Common to the material just discussed is the use of models of an economy in which there is inter-temporal optimization, but a notable difference is the introduction of a production sector and a concern with ensuring that the pricing formulae prohibit the possibility of arbitrage i.e. the solution tends to be closer to a general rather than partial equilibrium solution. The basic work horse of the literature is the model due to Cox, Ingersoll and Ross (1985) (CIR). Essentially they propose an economy driven by a number of processes that affect the rate of return to assets e.g. technological change and (possibly) an inflation factor. Dealing with the simplest case where there is just a single state vector, $\mu_t$, perhaps total factor productivity (TFP), it is assumed that this variable follows a diffusion process of the form:

$$d\mu_t = (b - \kappa \mu_t)dt + \phi \mu_t^{1/2}d\eta_t.$$  

General equilibrium in asset markets for such an economy results in an expression for the instantaneous rate of interest of the form:

$$dr_t = (\alpha - \beta r_t) dt + \sigma r_t^{1/2}d\eta_t. \tag{20}$$

Once one has the expression for the instantaneous rate the whole term structure $f_t(\tau)$ is priced according to a partial differential equation

$$\frac{1}{2} \sigma^2 f_{rr} + (\alpha - \beta \tau) f_r + f_t - \lambda r f_r - \tau f = 0, \tag{21}$$

where $f_{rr} = \partial^2 f / \partial \tau \partial \tau$, $f_r = \partial f / \partial \tau$, $f_t = \partial f / \partial t$ and the term $\lambda r f_r$, which depends upon the covariance of the change in the price of the factor with the percentage change in the optimal portfolio, is the "market price of risk" associated with that factor. This partial differential equation comes from the fact that a zero coupon riskless bond maturing at $t + \tau$ must be valued at

$$f_t(\tau) = E_t \left[ \exp \left( - \int_t^{t+\tau} r(\psi)d\psi \right) \right]. \tag{22}$$

Since the expected rate of change of the price of the bond is given by $r + \lambda r f_r/f$, it also can be interpreted as a liquidity premium. It is clear that we could group together the terms $(\alpha - \beta \tau)f_r$ and $-\lambda \tau f_r$ and treat the problem as one of pricing an asset using a "hypothetical " instantaneous rate that is generated by

$$dr_t = (\alpha - \beta r_t - \lambda r_t) dt + \sigma r_t^{1/2}d\eta_t, \tag{23}$$

$$= (\alpha - \gamma r_t) dt + \sigma r_t^{1/2}d\eta_t.$$ 

The distinction is between the true probability measure in (20) and the "equivalent martingale measure " in (23).

The analytic solution for the term structure in the CIR model is then (see Cox et. al. (p. 393))
\[ f_1(\tau) = A_1(\tau) \exp(-B_1(\tau)\tau_1), \]

where \( A_1(\tau) = \left[ \frac{2\delta \exp((\delta+\gamma)\tau/2)}{(\delta+\gamma)(\exp(\delta\tau)-1)+2\delta} \right]^{2\alpha/\sigma^2}, B_1(\tau) = \left[ \frac{2(\exp(\delta\tau)-1)}{(\delta+\gamma)(\exp(\delta\tau)-1)+2\delta} \right], \) and \( \delta = (\gamma + 2\sigma^2)^{1/2}. \)

Converting to a yield

\[ r_1(\tau) = \left\{ -\log(A_1(\tau)) + B_1(\tau)r_t \right\}/\tau. \]

(24)

This is a single factor model with the instantaneous rate or, more fundamentally, the "returns" factor, driving the whole term structure, i.e. the level of the term structure depends on the value of \( \tau \) at any point in time. The slope of the yield curve depends upon the parameters of the diffusion equation as well as the market price of risk.

Perhaps the biggest problem with this methodology is that it will never exactly reproduce an observed yield curve. This bothers practitioners a lot. One response has been to allow \( \alpha \) to change according to \( \tau \) and \( t \). This does is to add on "fudge factors" to the model based yield curve so that the modified curve equals the observed yield structure. Then, after forecasting \( \tau_{t+1} \) and finding the predicted term structure, the "fudge factors" from the previous period are added on. The need for "fudge factors" suggests that there is substantial mis-specification in the CIR model as a description of the term structure, just as "intercept corrections" in macro econometric models were given such an interpretation.

Brown and Dybvig (1986) estimated the parameters of the CIR model by maximum likelihood and then computed the residuals defined by the gap between the observed bond prices \( f_t \) and the predictions of the model \( f_t^* \). Examination of the residuals pointed to specification errors in the model.\(^{15}\) Looking at the CIR model in the light of stylized facts, the data should possess the characteristic that interest rates are near-integrated processes and possibly co-integrated with co-integrating vectors between any pair of rates of \([1 -1]\) i.e. the spreads should be \( I(0) \). The question that arises is whether the CIR model would deliver such a prediction. One problem to be overcome is quantifying the market price of risk, \( \lambda \), in the CIR bond formulae. As CIR point out, \( \lambda = 0 \) if the factor had no effect on the real economy e.g. if it was some nominal quantity such as the inflation rate. Accordingly, we will adopt this interpretation, allowing us to set \( \lambda = 0 \).

To induce a unit root we set \( \beta = 0 \), and we also put the drift term \( \alpha = 0 \). This makes \( \delta = \sqrt{2}\sigma, A_1(\tau) = 1, B_1(\tau) = \frac{2(\exp(\delta\tau)-1)}{d(\exp(\delta\tau)-1)+2\delta}. \)

Now the spread \( SP_t(\tau) \) will be

\[ r_t(\tau) - r_t(1) = \left[ \tau^{-1}B_1(\tau) - B_1(1) \right]r_t, \]

so that we will not get spreads to be \( I(0) \) unless the term in square brackets is zero. Generally it will not be. Realistic values for \( \sigma, \beta \) and \( \alpha \) might be the GMM estimates for \( r_t(1) \) of .049, .02 and .106. These produce values of \( \tau^{-1}B_1(\tau) = .990, .967, .930, .890 \) and .181 for the five maturities. In the limit \( (\tau \to \infty)B_1(\tau) = 2/\delta, \) and so the spreads between adjoining yields tend to zero as

\(^{15}\)Since there are \( n \) yields but only one factor they needed to add on a vector of errors to the model to produce a non-singular covariance matrix for \( f_t \), in order to be able to form a likelihood. It may be that the mis-specification reflects the assumptions made in this step.
the maturity lengthens.

The source of the failure of the spreads to be \( I(0) \) is the fact that \( \delta \neq 0 \). If \( \delta = 0 \) then, using L'Hopital's rule, \( B_1(\tau) = \tau \), and so the spreads should be identically zero. By making \( \sigma \) very small we can always produce results in which the spreads will be very close to being \( I(0) \) i.e. even if \( \sigma \) is not exactly zero it can be regarded as sufficiently close to zero that the spreads are nearly non-integrated, although the longer the maturity which the spread is based on the less likely we are to see such an outcome.

Another way of understanding the problem is to look at a discrete form of the fundamental pricing equation (22), \( f_t(\tau) = E_t[\exp(-\sum_{j=i}^{t+T-1} \tau_j)] \). Suppose that \( \tau_t \) is \( \mathcal{I}(1) \) with martingale difference innovations that are normally distributed. Then \( f_t(\tau) = \exp(-\tau \pi_t \{ \prod_{j=1}^{i-1} [\frac{1}{2} (\tau - j)^2 \text{var}_t(\sum_{j=1}^{i-1+1} \Delta \tau_{t+j})] \} \). If the conditional variance is a constant the spreads will therefore be \( I(0) \). However, if it depends upon the level of the instantaneous rate, the spreads at any maturity would be equal to a non-linear function of \( \tau_t \). For example, substituting the "square-root" formulation of CIR gives \( \text{var}_t(\Delta \tau_{t+1}) = \sigma^2 \tau_t \), and \( sp_t(\tau) = \text{cnst} - (1 - \frac{1}{2}) \log \tau_t \). Thus, it is important to determine the nature of the conditional variances in the data. Most econometric models of the term structure make these conditional variances GARCH processes, which effectively means that they are functions of \( \Delta \tau_{t-j} \). But, as seen in the section examining the term structure data, there is prima facie evidence of a levels effect after allowing for a GARCH specification of the conditional variance.

Given the conflicting evidence in section 2, one might look at other co-integrating vectors when performing the comparison with CIR. In general, the CIR model points towards co-integrating vectors that are of the form

\[
\tau_t(\tau) = d(\tau)\tau_t(1),
\]

where \( d(\tau) < 1 \) and decreasing with \( \tau \). As seen in section 2, with one exception both the Johansen and Phillips-Hansen estimates of \( d(\tau) \) have \( \hat{d}(\tau) > 1 \) and increasing in \( \tau \). The predictions from CIR type models are therefore diametrically opposed to the data.\(^{16}\)

### 3.3 Two Factor Models From Finance

Another response to the discrepancy between the model based prediction of a yield curve and the observed one, is to seek to make the model more complex. It is not uncommon in this literature to see people "bypassing" the step between the instantaneous rate and the fundamental driving forces and simply postulating a process for the instantaneous rate, after which this is used to price all the bonds. An example of this is the paper by Chen and Scott (1992) who assume that the instantaneous rate is the sum of two factors

\[
\tau_t = \xi_{1t} + \xi_{2t}, \tag{25}
\]

where

\[
d\xi_{1t} = (\alpha_1 - \beta_1 \xi_{1t}) dt + \sigma_1 \xi_{1t}^{1/2} d\eta_{1t}
\]

\(^{16}\)Brown and Schaefer (1994) find that the CIR model closely fits the term structure of real yields, where these are computed from British government index-linked bonds. Note in constructing the Johansen and Phillips-Hansen estimators that an intercept was allowed into the relations in order to correspond to \( A(\tau) \).
\[ \frac{d \xi_1}{d t} = (\alpha_1 - \beta_1 \xi_1) \frac{d t}{d t} + \sigma_1 \xi_1^{1/2} d \eta_1, \]

where \( d \eta_1 \) are independent, thereby making each factor independent. Then the solution for the bond price is

\[ f_1(\tau) = A_1(\tau) A_2(\tau) \exp \{ -B_1(\tau) \xi_1 - B_2(\tau) \xi_2 \}, \]

where \( A_2 \) and \( B_2 \) are defined analogously to \( A_1 \) and \( B_1 \). Obviously this framework could be extended to encompass any number of factors, provided they are assumed to be independent.

Another method is that of Longstaff and Schwartz (1992) who also have two factors but these are related to the underlying rate of return process \( \mu_t \) rather than directly to the instantaneous rate. In particular they wish to have the two factors being linear combinations of the instantaneous rate and its conditional variance. The model is interesting because the second factor they use, \( \xi_2 \), affects only the conditional variance of the \( \mu_t \) process, whereas both factors affect the conditional mean. This is unlike Chen and Scott’s model which has \( \xi_1 \) and \( \xi_2 \) affecting both the mean and variance. Empirically, the two factors are regarded as the short term rate and its conditional volatility, where the latter is estimated by a GARCH process when assessing the quality of the model.

Tests of the model are limited to how well it replicates the unconditional standard deviations of yield changes.

There are a number of other two factor models. Brennan and Schwartz (1979) and Edmister and Madan (1993) begin with the long and short rates following a joint diffusion process. After imposing the “no arbitrage condition” and assuming that the long rate is a traded instrument, Brennan and Schwartz find that the price of the instantaneous risk associated with the long rate can be eliminated, and the two factors then effectively become the instantaneous rate and the yield spread between that rate and the long rate. Eliminating the price of risk for the long rate makes the model non-linear and they need to linearize to find a solution. Even then there is no analytical solution for the yield curve as with CIR. Another possibility for a two factor model might be to allow for stochastic volatility as a factor. Edmister and Madan find closed form solutions for the term structure in their formulation.

Suppose that the first factor in Chen and Scott’s model is a “near I(1)” process whereas the second factor is I(0). Then the instantaneous rate has the common trend format (compare (25) and (8) recognising that \( J \) can be regarded as the unit column vector). Using the same parameter values for the first factor as the polar case discussed in the preceding sub-section i.e. \( \beta_1 = 0, \lambda_1 = 0, \sigma_1 = 0 \), the first factor disappears from the spreads, which now equal

\[ \tau_1(\tau) - \tau_1(1) = \log(A_2(1)/A_2(\tau)) + [\tau^{-1}B_2(\tau) - B_2(1)] \xi_2. \]

Hence, they are now stochastic and inherit the properties of the second factor. For them to
be persistent, it is necessary that the second factor have that characteristic. Notice also that 
$r_t(\tau) - r_t(\tau - 1)$ will tend to zero as $\tau \to \infty$, and this may make it implausible to use this model with a large range of maturities.

Consequently, this two factor model can be made to reproduce the standard results of the co-integration approach in the sense that the EC terms are decomposed into a smaller number of factors. Of course the model would predict that the coefficients on the factors would be negative as $\tau^{-1}B_2(\tau) \leq B_2(1)$. The conclusion of negative weights extends to any number of factors, provided they are independent, so it is interesting to look at the evidence upon the signs of the coefficients of the factors in our data set, where the non-trend factors are equated with the principal components. Although one cannot uniquely move from the principal components/spreads relation to a spreads/principal components relation, a simple way to get some information on the relationship between spreads and factors is to regress each of the spreads against the principal components. Doing so the $R^2$ are .999, .999, .98 and .99 respectively, showing that the spreads are well explained by the three components. The results from the regressions are

$$sp_t(3) = .36\psi_{1t} - .83\psi_{2t} + .48\psi_{3t}$$

$$sp_t(6) = -.76\psi_{1t} - .09\psi_{2t} + .42\psi_{3t}$$

$$sp_t(9) = -1.28\psi_{1t} + .33\psi_{2t} + .44\psi_{3t}$$

$$sp_t(120) = -1.44\psi_{1t} + 1.84\psi_{2t} + 2.12\psi_{3t},$$

where $\psi_{jt}$ are the first three principal components. It is clear that independent factor models would not generate the requisite signs. Formal testing of two factor pricing models is in its infancy. Pearson and Sun(1994) and Chen and Scott(1993) estimate the parameters of the model by maximum likelihood and provide some evidence that at least two factors are needed to capture the term structure adequately.

The two factor model is also useful for examining some of the literature on the validity of the expectations hypothesis. Campbell and Shiller(1991) pointed out that the hypothesis implies that

$$\tau_{t+1}(\tau - 1) - \tau_t(\tau) = \alpha_0 + \frac{1}{\tau - 1}[\tau_t(\tau) - \tau_t(1)],$$

(26)

if the liquidity premium was a constant. They found that this restriction was strongly rejected by the data. With McCulloch and Kwon's data and $\tau = 3$, the regression of $\tau_{t+1}(2) - \tau_t(3)$ against $\tau_t(3) - \tau_t(1)$ yields an estimated coefficient of -.09, well away from the predicted value of .5. Of course, the assumption of a constant premium is incorrect. Bond prices are determined by (22) which, when discretized, would be,

$$f_t(\tau) = E_t[\exp(-\sum_{j=t}^{t+\tau-1} r_j)]$$
\[
\begin{align*}
t_{t+1} & = \exp(-E_t(\sum_{j=t}^{t+1} r_j)) \nu_t \\
& = f_t^E(\tau) \nu_t,
\end{align*}
\]

where \(f_t^E(\tau)\) is the bond price predicted by the expectations theory. Thus \(\tau_t(\tau)\) differs from that of the expectations theory by the term \(-\tau^{-1} \log \nu_t\), and this in turn will be a function of the conditional moments of \(\Delta \tau_t\). In the case where \(\Delta \tau_t\) is conditionally normal it depends upon the conditional variance, and the equation corresponding to (26) will now feature a time varying \(a_0\) that depends on this moment. If the conditional variance relates to the spreads with a negative coefficient, then that could cause there to be a negative bias in the coefficient of \(\tau_t(\tau) - \tau_t(1)\) in the Campbell and Shiller regressions. One scenario in which this happens is if the conditional variance depended upon \(\Delta \tau_t\), as happens with an EGARCH model. Then, due to cointegration amongst yields, \(\Delta \tau_t\) could also be replaced by the lagged spreads, and these will have negative coefficients. More generally, since we observed in section 2 that the factors influencing the term structure, such as volatility, could be written as linear combinations of the spreads, there is a possibility that term structure anomalies might be explained in this way.

3.4 Multiple Non-Independent Factor Models in Finance

Duffie and Kan (1993) present a multi-factor model of the term structure where the factors may not be independent. As for the two factor models it is assumed that the instantaneous rate is a linear function of \(M\) factors, collected in an \(M \times 1\) vector \(\xi_t\), which evolves according to the diffusion process

\[
d\xi_t = \mu(\xi_t) dt + \sigma(\xi_t) d\eta_t,
\]

where \(d\eta_t\) is a vector of standard Brownian motions and \(\mu(\xi_t), \sigma(\xi_t)\) are vectors and matrices corresponding to drift and volatility functions. They then ask what type of functions \(\mu(\cdot)\) and \(\sigma(\cdot)\) are capable of producing a solution for the \(n\) bond prices \(f_t(\tau), \tau = 1, \ldots, n\), of the exponential affine form

\[
f_t(\tau) = \exp[(A(\tau) + B(\tau)\xi_t)]
\]

\[
= \exp [(A(\tau) + \sum_{i=1}^M B_i(\tau)\xi_i)].
\]

It turns out that \(\mu(\xi_t)\) and \(\sigma(\xi_t)\) should be linear (affine) functions of \(\xi_t\). Thereupon the solution for \(B(\tau)\) can be found by solving an ordinary differential equation of the form

\[
\dot{B}(\tau) = B(B(\tau)), \quad B(0) = 0.
\]

In most cases only numerical solutions for \(B(\tau)\) are available. Duffie and Kan consider some special cases, differing according to the evolution of \(\xi_t\). When the \(\xi_t\) are joint diffusions driven by Brownian motion with covariance matrix \(\Omega\) that is not diagonal, there is the possibility that the weights attached to the factors can have different signs, and so the principal defect with the
two factor models of the preceding sub-section might be overcome. To date little empirical work seems to be available on these models, with the exception of El Karoui and Lacoste (1992) who make $\xi_t$ Gaussian with constant volatility.

3.5 Forward Rate Models

In recent years it has become popular to model the forward rate structure directly rather than the yields, e.g. in Ho and Lee (1986) and Heath, Jarrow and Morton (1992) (HJM). Since the forward rates are linear combinations of the yields, specifications based on the nature of the forward rate structure impose some restriction upon the nature of the yield curve, and conversely. In the light of what is known about the behavior of yields, this sub-section considers the likelihood that popular models of forward rates can replicate the term structure. In what follows, one step ahead forward rates are used along with the HJM framework. In the interest of space only a simple Euler discretization of the HJM stochastic differential equations describing the evolution of the forward rate curve is used.

Many variants of these equations have emerged, but they have the common format,

$$F_t(\tau - 1) - F_{t-1}(\tau) = c_{t,\tau-1} + \sigma_{t,\tau-1}\xi_{t,\tau-1},$$

where $\xi_{t,\tau-1}$ is n.i.d.$(0,1)$. Differences among the models reflect differences in the assumptions made about volatilities. Examples would be a constant volatility model in which $c_{t,\tau-1} = a_0 + \sigma^2\tau$ and $\sigma_{t,\tau-1} = \sigma$, or a proportional volatility model that has $c_{t,\tau-1} = -\sigma F_t(\tau) + \sigma F_t(\tau)(\sum_{k=1}^{n} F_t(k))$ and $\sigma_{t,\tau-1} = \sigma F_t(\tau)$. The nature of $c_{t,\tau-1}$ reflects the no-arbitrage assumption. After some manipulation it can be shown that

$$(28)$$

$$F_t(\tau - 1) - F_{t-1}(\tau) = \left[\frac{1}{\tau(\tau+1)}\right] sp_t(\tau) - \frac{1}{\tau^2} (\tau_t(\tau + 1) - \tau_t(\tau)) + \frac{\tau+1}{\tau} \Delta r_t(\tau+1) - \frac{1}{\tau} \Delta r_t(1),$$

so that the equation used by HJM for the evolution of the forward rate incorporates spreads and changes in yields. In turn, using co-integration ideas, $\Delta r_t(\tau + 1)$ depends upon spreads, and this shows quite clearly that the characteristics of $F_t(\tau - 1) - F_{t-1}(\tau)$ will be those of the spreads - see Table 2. Consequently, at least for small $\tau$, constant volatility models with martingale difference errors could not adequately describe the data. It is possible that proportional volatility models might do so due to the dependence of their $c_{t,\tau-1}$ upon $F_t(\tau)$, as the latter is near integrated. To check this out we regressed $F_t(2) - F_{t-1}(3)$ against $c_{t,2}$ and $sp_{t-1}(3)$ for $n = 9$ and a variety of values for the market price of risk $\lambda$. For $\lambda = 0$ the t ratio of the coefficient of $sp_{t-1}(3)$ was -4.37, while for very large $\lambda$ it was -4.70. Adopting other values for $\lambda$ resulted in t ratios between these extremes. Hence, the conditional mean for the forward rates is far more complex than that found in HJM models. Moreover, the rank of the covariance matrix of the errors $\xi_{t,\tau-1}$ must reflect the number of factors in the term structure, which appears to be two or three, so that the common assumption of a single error to drive all forward spreads seems inaccurate.

A number of formal investigations have been made into the compatibility of the HJM model with the data - Abken(1993) and Thurston(1994) fitted HJM models to forward rate data by GMM whilst Amin and Morton(1994) used options prices to recover implied volatilities whose
evolution was compared to those of the most popular variants of the HJM model. Abken and Thurston reach conflicting conclusions— the latter favours a constant volatility formulation and the former a proportional one, although his general conclusion was that all models were rejected by the data. Consequently, it seems interesting to look at the stylized facts regarding volatility and to compare them with model specifications. Equation (28) is useful for this task. As it has been shown that there is a levels effect in $\Delta \tau_t(k)$, in order to have constant volatility it would be necessary that there be some "co-levels" effect, analogous to the co-persistence phenomenon of the GARCH literature—Bollerslev and Engle (1993)— i.e. even though $\Delta \tau_t(k)$ displays a levels effect the linear combination $\tau_{t-1}(t) + \frac{1}{2} \Delta \tau_t(1)$ does not. This contention is easily rejected—a plot of that variable squared against $\tau_{t-1}(3)$ looks almost identical to figure 1, and such an observation points to the proportional volatility model as being the appropriate one.

4 Conclusion

This chapter has described methods of modeling the term structure that are to be found in the econometrics and finance literatures. By utilizing a factor representation we have been able to show that there are many similarities in the two approaches. However, there were also some differences. Within the econometrics literature it is common to assume that yields are integrated processes and that spreads constitute the co-integrating relations. Although the finance literature takes the stance that yields are near integrated but stationary, it emerges that the models used in that literature would not predict that the spreads are co-integrating errors if we actually replaced the stationarity assumption by one of a unit root. The reason for this outcome is found to lie in the assumption that the conditional volatility of yields is a function of the level of the yields. Empirical work tends to support such an hypothesis and we suggest that the consequences of such a relationship can be profound for testing propositions about the term structure. We also document a number of stylized facts about a set of data on yields that prove useful in assessing the likely adequacy of many of the models that are used in finance for capturing the term structure.

5 References


Canova, F. and J. Marrinan (1993), "Reconciling the Term Structure of Interest Rates with the Consumption Based ICAP Model", (mimeo, Brown University)


Hejazi, W. (1994), "Are Term Premia Stationary?", (mimeo, University of Toronto)


Pfann, G.A., P.C. Schotman and R. Tschernig (1994), "Nonlinear Interest Rate Dynamics and Implications for the Term Structure", (mimeo, University of Limburg)


