Abstract

Models of interest rates often consider high frequency interest rate movements, either continuous or daily, over relatively short periods of time. Relatively few attempts have been made to model low frequency interest rate dynamics over extended periods of time. Normally there is little statistical analysis of the interest rate series being modelled, or validation of the postulated interest rate dynamics. In this paper I demonstrate that Australian interest rates have a distinctive statistical structure involving nonlinear dependence and time varying volatility. I then go on to present some results of fitting various stochastic models to Australian interest rate series, including examples of probability density estimates from simulated series. In the process I introduce a new class of nonlinear stochastic time series model, the Exponential Regressive Conditional Heteroscedasticity, or ERCH, model. ERCH models capture the observed nonlinear dependence and time varying volatility by explicitly modelling the logarithm of the conditional error standard deviation (and therefore variance). Allowance is also made for contemporaneous correlation between the shocks to the various series involved. The impact of data frequency is also discussed. Such research has implications for resilience reserving, capital adequacy, solvency, asset-liability modelling, matching, and medium and long term interest rate risk management.

Keywords: nonlinear, heteroscedasticity, ERCH models, model validation, probability density estimates, simulation, data frequency.
1. **INTRODUCTION**

This paper considers the statistical structure and modelling of interest rate series at monthly, quarterly and annual frequencies. The paper is divided into the following sections:

1. Introduction;
2. Statistical Analysis of the Interest Rate Series;
3. Rate Models;
4. Model Fitting Results.

The data used was Australian 13 week Treasury Note (T-note) issue yields, \( \{ T_t \} \), and Australian Commonwealth Government Bond 10 year indicator yields to maturity, \( \{ B_t \} \). When examining the statistical structure of interest rates the series investigated were the monthly, quarterly and annual changes in the natural logarithm of the yields, i.e. the differenced logarithmic series \( \Delta \ln T_t \), \( \Delta_3 \ln T_t \), \( \Delta_{12} \ln T_t \), \( \Delta \ln B_t \), \( \Delta_3 \ln B_t \), and \( \Delta_{12} \ln B_t \), where \( \Delta_m y_t = y_t - y_{t-m} \) and \( t \) is a monthly index. Monthly and quarterly data over the periods 1980-93 and 1970-93 were used, respectively. Annual data was examined over the period June 1950 to June 1993. The tender system for selling T-notes was introduced in December 1979. Since then the T-note yields used were the weighted average issue yields at the last available tender. The T-note yields used prior to then were the Treasury set issue yields. T-note yields were not available prior to June 1963, where a proxy was used.

The statistical analyses were based on the transformed yields for the following reasons:

i. it imposes a suitable positivity constraint on the yield (by taking logarithms);

ii. it produces a superficially stationary series, which would be covariance stationary for simple underlying processes (differencing);

iii. the differenced series has a much lower variance than the undifferenced series;

---

1 based on the weighted average bank overdraft interest rate (June 1956 to June 1962) and the housing loan variable interest rate (June 1950 to June 1955), via linear regression over the period where the series overlapped.
iv. statistical analysis of the undifferenced series would not be meaningful due to lack of stationarity and large serial correlation; and

v. it can be interpreted as the continuously compounded rate of change in interest rates (and can therefore be expressed as a percentage).

2. **Statistical Analysis of the Interest Rate Series**

Short and long term rates are highly correlated, with the short term rates displaying greater volatility.

![Monthly Australian Interest Rates](image)
Interest rates are clearly not covariance stationary. The change in log-yields appears much more stationary, though the level of volatility appears time dependent, in the case of short term rates at least.

The annual and quarterly changes represent sums of monthly changes, hence the Central Limit Theorem implies that such aggregated series can be expected to be more Normally distributed than the monthly series (i.e. non-Normality should tend to "average out" with increasing temporal aggregation).

2.1 Frequency Statistics

The sample skewness, $\gamma_1$, and excess kurtosis, $\gamma_2$, were calculated as

$$
\gamma_1 = \frac{n}{(n-1)(n-2)} \sum \left( \frac{x_i - \bar{x}}{s} \right)^3,
$$

and

$$
\gamma_2 = \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum \left( \frac{x_i - \bar{x}}{s} \right)^4 - \frac{3(n-1)^2}{(n-2)(n-3)},
$$

respectively, where $x_i$ is the $i$-th sample value, $s$ is the sample standard deviation and $n$ is the sample size. Under the null hypothesis that the returns are Normally distributed, Lomnicki (1961) shows that approximately
\[ \gamma_1 \sim \mathcal{N}
left(0, \frac{3}{n} \sum r(\tau) \right) \] and \[ \gamma_2 \sim \mathcal{N}
left(0, \frac{2}{\sqrt{n}} \sum r(\tau)^4 \right), \]
where \( r(\tau) \) is the sample autocorrelation at lag \( \tau \). It is therefore not difficult to determine the probability of a sample value at least as extreme as the observed skewness and excess kurtosis arising under the null hypotheses, which I refer to as the p-value, and also the critical values under the null hypotheses. Two-sided tests were generally used for skewness and one-sided tests for excess kurtosis.

The observed frequency distribution of changes in log T-note yields is summarised in the following table.
An anomaly occurs in the annual log T-note changes in that the annual changes appear more leptokurtic than the quarterly and monthly changes, contrary to the implications of the Central Limit Theorem. Inspection of the annual changes reveals an outlier at June 1974, where T-note yields rose to 10.75% from 4.91% a year earlier, resulting in a continuously compounded rate of increase of 78.4%. The next largest rates of change (i.e. log-changes) were -49.4% in June 1992 and 41.7% in June 1989. The average of the June 1973 and June 1975 T-note rates produces a rate of increase of 25.9%. Substituting 25.9% for the June 1974 rate of increase of 78.4% reduces the annual skewness and excess kurtosis statistics to only -0.47 and 1.03 respectively, neither of which is significant.

The frequency distribution of monthly and quarterly changes in log T-note yields and 10 year Bond yields are significantly leptokurtic, i.e. fat tailed. Compared to the Normal distribution, small deviations are more likely, as are extreme deviations, with moderate deviations being less likely. Becker (1991) similarly found excess kurtosis in monthly changes in log U.S. Treasury security yields.

The corresponding 10 year Bond statistics are summarised in the following table.
<table>
<thead>
<tr>
<th>Frequency Statistics Change in Log 10 year Bond Yields</th>
</tr>
</thead>
<tbody>
<tr>
<td>---</td>
</tr>
<tr>
<td>Standard Dev.</td>
</tr>
<tr>
<td>Mean Abs. Dev.</td>
</tr>
<tr>
<td>Range</td>
</tr>
<tr>
<td>Skewness</td>
</tr>
<tr>
<td>P-value</td>
</tr>
<tr>
<td>5% Crit. Value</td>
</tr>
<tr>
<td>Excess Kurtosis</td>
</tr>
<tr>
<td>P-value</td>
</tr>
<tr>
<td>1% [5%] Crit. Value</td>
</tr>
</tbody>
</table>

The corresponding frequency statistics for \( \{T_i\}, \{\ln T_i\}, \{B_i\} \) and \( \{\ln B_i\} \) would be less meaningful due to their obvious lack of stationarity and large serial correlation.

### 2.2 Time Varying Volatility (Heteroscedasticity)

The observed leptokurtic frequency distribution of the changes in log-yields is consistent with a time varying volatility, i.e. heteroscedasticity. The proposition that interest rate volatility is time dependent was formally tested. The test used can be derived using the likelihood ratio method, and is described in Kendall & Stuart (1979) and Johnson & Tetley (1966). The test statistic is here expressed in terms of the relationship between the arithmetic and geometric average sample variances. The test statistic used assumes large sample Normal distributions, and is defined as

\[
J = N \ln \left( \frac{\frac{1}{k} \sum s_i^2}{\prod s_i^{2/k}} \right) \sim \chi^2_{k-1},
\]

where \( N \) is the total number of yields in the sample and \( k \) is the number of groups.
The change in log-yields at monthly intervals over the period 1980-93 were divided into calendar years, and the mean and variance of monthly values calculated for each year. The null hypothesis was that the variance of the monthly values was constant from year to year. Similarly quarterly and annual changes in log-yields over the periods 1970-93 and June 1950 to June 1989 were grouped by trienniums and decades, respectively. The results for changes in log T-note yields are set out in the following table.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Arithmetic Av.</td>
<td>0.006241</td>
<td>0.02128</td>
<td>0.03714</td>
</tr>
<tr>
<td>Geometric Av.</td>
<td>0.003264</td>
<td>0.01045</td>
<td>0.02394</td>
</tr>
<tr>
<td>J Statistic</td>
<td>108.9</td>
<td>68.25</td>
<td>17.57</td>
</tr>
<tr>
<td>P-value</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>1% [5%] Crit.</td>
<td>27.7</td>
<td>18.5</td>
<td>[7.81]</td>
</tr>
</tbody>
</table>

Non-Normality means that the significance of the test is less than it appears to be. An approximate adjustment would be to divide \((1 + \frac{1}{2} \gamma_2)^{\frac{1}{2}}\) into the J statistic. Application of the adjustment does not alter the above conclusions at the 1% [5%] level. Adjusting for the June 1974 outlier as done in section 2.1 reduces the J statistic for annual changes from 17.57 to 8.68, which is still significant at the 5% level. Further adjustment for excess kurtosis then reduces the J statistic to approximately 7.1, which is no longer significant at the 5% level. **There is therefore evidence that changes in log T-note yields are heteroscedastic** (certainly for monthly and quarterly intervals, and even for annual intervals the data is more consistent with heteroscedasticity than homoscedasticity).

The corresponding results for changes in log 10 year Bond yields are set out in the following table.
### Time Dependence of the Volatility of Bond Yields

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Arithmetic Av.</strong></td>
<td>0.001358</td>
<td>0.00343</td>
<td>0.01328</td>
</tr>
<tr>
<td><strong>Geometric Av.</strong></td>
<td>0.001156</td>
<td>0.00287</td>
<td>0.01158</td>
</tr>
<tr>
<td><strong>J Statistic</strong></td>
<td>27.1</td>
<td>17.08</td>
<td>5.49</td>
</tr>
<tr>
<td><strong>P-value</strong></td>
<td>0.01</td>
<td>0.02</td>
<td>0.14</td>
</tr>
<tr>
<td><strong>1% [5%] Crit.</strong></td>
<td>27.7</td>
<td>[14.1]</td>
<td>[7.81]</td>
</tr>
</tbody>
</table>

Approximately adjusting for excess kurtosis reduces the monthly statistic to 21.8, which has a p-value of 5.9%, and is therefore only marginally significant. The quarterly statistic is not significant at the 5% level after adjustment for excess kurtosis. Nevertheless the data is still more consistent with heteroscedasticity than homoscedasticity.

Becker (1991) concluded that the change in log U.S. Treasury security yields was heteroscedastic.

Time varying volatility, i.e. heteroscedasticity, can explain the observed leptokurtic frequency distributions.

#### 2.3 Serial Correlation in the Interest Rate Series

Significant autocorrelations (at the 5% level) are observed in the $\{\ln T_t\}$ series over the period 1980-93 at lags of 7 and 12 months, being 0.21 in both cases. In the case of the quarterly 1970-93 series significant autocorrelations (at the 5% level) were observed at lags of 4 and 14 quarters, with the largest being 0.24 at lag 4 quarters, consistent with the lag 12 month autocorrelation in the monthly series.

The $\{\ln B_t\}$ series exhibits less serial correlation than the short term yield series. The largest and most consistent autocorrelation in the monthly series over the period 1980-93 is observed at lag 9 months, being 0.18. The only significant autocorrelation in the quarterly 1970-93 series was 0.25 at lag 3 quarters, which is consistent with the monthly series.
2.4 Serial Correlation in the Nonlinear Deviations

Significant serial correlation is observed in the nonlinear deviations of monthly, quarterly and annual changes in log T-note yields, implying a high degree of nonlinear dependence in short term interest rates. The largest and most persistent autocorrelations were generally observed in the log-absolute deviations, e.g. $\ln(\nabla \ln T_t - \mu_t)$. Significant autocorrelations (at the 5% level) in the log-absolute deviations of the monthly T-note series (1980-93) are observed at lags 4, 7, 8 and 16 months.$^2$

$^2$ the 5% critical value for the monthly autocorrelations is 0.15.
In the case of the quarterly T-note log-absolute deviations (1970-93), eight of the first eleven quarters autocorrelations are significant at the 5% level\(^3\). In the case of the annual T-note log-absolute deviations (6/50-6/93), the first three years' autocorrelations are significant at the 5% level, as are the autocorrelations at lags 8 to 11 years\(^4\). Adjusting for the June 1974 outlier as done in section 2.1 does not alter the level of the large autocorrelations observed in the annual nonlinear deviations.

\(^3\) the 5% critical value for the quarterly autocorrelations is 0.20.

\(^4\) the 5% critical value for the yearly autocorrelations is 0.30.
Note that the autocorrelations in the nonlinear deviations are generally positive, which implies persistence in the volatility of yield changes, i.e. clustering of large and small yield changes (of either sign).

Similar, though somewhat less pronounced, features were observed in the log 10 year Bond nonlinear deviations.

3. Rate Models

Models of interest rates often consider high frequency interest rate movements, either continuous or daily, over relatively short periods of time. Relatively few attempts have been made to model low frequency interest rate dynamics over extended periods of time. Normally there is little statistical analysis of the interest rate series being modelled, or validation of the postulated interest rate dynamics.

Low frequency models of untransformed nominal interest rates over extended time periods have problems constraining the modelled rates to positive values, a problem easily overcome by modelling the logarithmic transformed rates instead. I will therefore only consider models of the logarithm of the rates. While the statistical analysis of the interest rate series was based on the change in the logarithmic transformed rates, models should preferably be expressed in terms of the logarithmic transformed rates themselves rather than the differences. This is because there is important
information in the level of the rates and their interaction with other series, information which may be lost in the differenced series. This is certainly the case with annual interest rate data. Interestingly however, when monthly rates are considered, models of the level of rates collapse to become models of the differences, which has important implications for the modelling of interest rate dynamics. This point is discussed in the final section, section 4.4, where models are fitted to monthly T-note yields over the period 1980-93.

3.1 Random Walk, Autoregressive and Related Models

Common interest rate models include the Random Walk and first order Autoregressive, AR(1), models of the logarithm of interest rates.

The Random Walk model, $\ln T_t = e_t$, with $e_t \sim i.i.d. N(0, \sigma^2)$, is also referred to as the Log-normal model of interest rates. The AR(1) model, used by Tilley (1992) and the Australian Capital Adequacy Sub-Committee (1992) for example, is of the form

\[
\ln T_t = \mu + \alpha \ln T_{t-1}^* + e_t,
\]
i.e.

\[
\nabla \ln T_t = -(1-\alpha) \ln T_{t-1}^* + e_t,
\]

where $\mu$ is the unconditional mean of $\{\ln T_t\}$ and a superscript asterix refers to mean adjustment, i.e $\ln T_t^* = \ln T_t - \mu$ and $E(\ln T_t^*) = 0$. When modelling yields at various maturities, e.g. $\ln T_t$ and $\ln B_t$, the contemporaneous shocks to the series can be assumed to be correlated, i.e.

\[
E(\ln T_t e_t \ln B_s e_s) = \rho \cdot \sigma_{\ln T} \cdot \sigma_{\ln B}
\]
when $t = s$, and zero when $t \neq s$, where $\rho$ is the contemporaneous correlation between the shocks to the two series.

It follows that if the monthly log-yield process is AR(1), as above, then the quarterly ($m = 3$) and annual ($m = 12$) log-yield processes are also AR(1), with dynamics given by

\[
\ln T_t = \mu + \alpha^m \ln T_{t-m}^* + \xi_t,
\]
i.e.

\[
\nabla_m \ln T_t = -(1-\alpha^m) \ln T_{t-m}^* + \xi_t,
\]
the validity of which can be tested. Note that interest rates are highly serially correlated, hence $\alpha$ is close to 1. Note also that the AR(1) rate parameter and
process serial correlation can be expected to reduce with reducing data frequency (i.e. increasing \( m \)).

The commonly used continuous time Ornstein-Uhlenbeck process is equivalent to the above AR(1) model. Another commonly used continuous time model is the Cox-Ingersol-Ross model, which adds a particular form of past rate dependent volatility to the Ornstein-Uhlenbeck process. The discrete time model of Mobbs (1985) is along similar lines. These models can be seen to be consistent with highly constrained ERCH models, defined in section 3.3.

### 3.2 Wilkie Bond Model

A commonly used annual stochastic asset model, at least in the U.K., is the model introduced by Wilkie (1986). Price inflation, modelled as an AR(1) process, drives the model. Irredeemable bond, or Consul, yields were modelled in terms of inflation and the residuals of his share dividend yield model, which itself was driven by inflation. A somewhat simpler version of the bond model was also given which excluded three parameters, including the dividend yield residual term (which was very small in any case). Wilkie (1986) reported that his investigations showed that very similar long term forecasts were obtained from the simpler model. The simpler bond model can be represented as follows:

\[
F_t = M_F + \alpha_F F_{t-1} + \zeta_t \\
Q_t = \beta F_t + (1 - \beta) Q_{t-1} \\
\ln(B_t - Q_t) = M_B + \alpha_B \left( \ln(B_{t-1} - Q_{t-1}) - M_B \right) + \beta \xi_t
\]

where \( F_t \) is the continuously compounded annual rate of price inflation, \( M_F \) is the corresponding unconditional mean rate of inflation, \( Q_t \) is an inflation carry forward, \( B_t \) is the long bond yield, and \( t \) is a yearly index. Note that \( 0 \leq \beta < 1 \) and \( B_t > Q_t \) \( \forall t \). The Wilkie (1986) model can be seen to represent the logarithm of the “real” bond yield, \( B_t - Q_t > 0 \), as an Autoregressive process\(^5\). Anticipated future inflation is modelled by the inflation carry forward, \( Q_t \), which is an exponentially weighted moving average of past inflation rates. Note that the above bond model collapses to an AR(1) model

\(^5\) 1st order, i.e. AR(1), in the simpler model, 3rd order, i.e. AR(3), in the more complicated model.
for the logarithm of the bond yield when $\beta = 0$, which is exactly what happened when the above model was fitted to Australian annual price inflation and 10 year bond rates over the period June 1950 to June 1993. The constrained maximum likelihood estimate of $\beta$ is zero, and there is no increase in the AR(1) log-likelihood. Periods of high inflation and relatively low bond yields, such as occurred in the early 1950s and mid-1970s, are poorly represented by the model, and contributed to the model's poor showing. The addition of a second autoregressive term in the real yield equation was also insignificant at any reasonable level.

In addition the AR(1)/Wilkie (1986) inflation model has been found to be inconsistent with observed inflation experience, as noted by, for example, Daykin & Hey (1989), Kitts (1990), Carter (1991) and Harris (1995).

The Wilkie (1986) bond model can therefore be considered to be an unnecessary complication of the common AR(1) model, at least in the case of Australian annual long bond yields.

### 3.3 ERCH Models

In response to the observed interdependence between financial series conditional volatilities and means, Harris (1994) introduced a class of model where the logarithm of the conditional error standard deviation (and therefore variance) was expressed as a linear combination of lagged exogenous and dependent explanatory variables. He referred to the class as Exponential Regressive Conditional Heteroscedasticity, or ERCH, models.

The ERCH regression model, denoted by ERCH$(\psi_i ; \phi_i)$, is defined as follows:

\[
\ln T_t = \mu + \theta^T \cdot \psi_t^* + \xi_t,
\]

\[
\xi_t = S_t \cdot Z_t \quad Z_t \sim \text{i.i.d. } N(0,1)
\]

\[
\ln S_t = \omega_0 + \omega^T \cdot \phi_t
\]

where $\theta$ and $\omega$ are parameter vectors, and $\psi_i$ and $\phi_i$ are vectors of lagged explanatory variables. A superscript asterix again refers to mean adjustment. Because volatility is modelled in terms of the logarithm of the error standard

---

6 the model was fitted by maximum likelihood (refer to Harris (1995)), with both $\beta$ and the initial inflation carry forward entering as additional variables.
deviation, the modelled variance is appropriately constrained to be positive (being an exponential function). Note that an ERCH process has a multiplicative variance structure which may involve exogenous variables, while a GARCH\textsuperscript{7} process has an additive volatility structure involving only past dependent variance and errors.

An ERCH model requires the simultaneous modelling of the exogenous variables, resulting in an ERCH system, where each series is modelled in terms of past values of itself and the other series. The contemporaneous shocks to the series within the ERCH system may be assumed to be correlated, which is likely to be the case between short and long term interest rates as a minimum. An $m$ series ERCH system may be expressed in multivariate form as

$$X_t = M + \Theta \Psi_t^* + \xi_t$$

$$\xi_t = \Lambda_t Z_t$$

$$\ln \Lambda_t = \text{diag}\{\omega_0 + \Omega \Phi_t\}$$

$$Z_t \sim N(0, \Sigma_z)$$

$$E(Z_t^T Z_s) = \begin{cases} 0 & t \neq s \\ \Sigma_z & t = s \end{cases}$$

where,

$X_t$ is an $m \times 1$ column vector of series values at time $t$;

$M = E(X_t)$, an $m \times 1$ column vector of unconditional series means;

$\Theta$ is an $m \times p$ conditional mean parameter matrix;

$\Psi_t$ is a $p \times 1$ column vector of lagged explanatory variable values at time $t$;

$\xi_t$ is an $m \times 1$ column vector of conditionally multivariate Normal random errors or shocks to the series at time $t$;

$\Lambda_t$ is an $m \times m$ diagonal matrix of error standard deviations at time $t$;

$Z_t$ is an $m \times 1$ column vector of multivariate standard Normal standardised errors or shocks to the series at time $t$;

\textsuperscript{7} Generalised Autoregressive Conditional Heteroscedasticity, a common model of time varying and persistent volatility, introduced by Bollerslev (1986).
\[ \ln \Lambda_t \text{ is an } m \times m \text{ diagonal matrix of the logarithms of the error standard deviations at time } t; \]
\[ \text{diag}(...) \text{ is a diagonal matrix whose } i\text{-th non-zero element is equal to the } i\text{-th element of its vector argument}; \]
\[ \omega_0 \text{ is an } m \times 1 \text{ column vector of parameters;} \]
\[ \Omega \text{ is an } m \times q \text{ conditional volatility parameter matrix;} \]
\[ \Phi_t \text{ is a } q \times 1 \text{ column vector of lagged explanatory variable values at time } t; \]
\[ \Sigma_z \text{ is an } m \times m \text{ contemporaneous correlation matrix, the } i,j\text{-th element of which is equal to the contemporaneous correlation between the } i\text{-th and } j\text{-th components of the } Z_t. \]

4. MODEL FITTING RESULTS

Models were fitted to Australian data by maximising the log-likelihood function. The parameter estimates are therefore the maximum likelihood estimates. When comparing different models it should be noted that the addition of \( p \) independent (unrelated) variables can be expected to increase the maximum log-likelihood by \( p/2 \) on average. The likelihood ratio test given by Akgiray (1989) was used to determine whether any particular increase was statistically significant, and hence whether the null hypothesis that the additional variables have no explanatory ability can be rejected. The test is

\[ 2 \times \Delta \ln L = 2 \times [\max \ln L(\theta_I) - \max \ln L(\theta_0)] \sim \chi^2_p, \]

where \( p \) is the number of additional parameters in \( \theta_I \) (strictly the models should be nested, i.e. \( \theta_I \supset \theta_0 \)). At the 1% level, the required difference in maximum log-likelihoods is therefore 3.32 for 1 additional parameter, 4.61 for two, and 5.67 for three. At the 5% level the required increases are 1.92, 3.00 and 3.91 respectively. Model fitting and validation are discussed further in Harris (1995).

4.1 Annual Treasury Note Yield Models

The first set of results presented is for the series of Australian annual 13 week Treasury Note yields over the period June 1950 to June 1993. The rate parameter of the fitted AR(1) model was \( \alpha = 0.92 \). A number of regression models were tried based on the cross-correlation between the model residuals and lagged exogenous variables, particularly past price inflation, real economic growth, bond yields and share price index returns. The observed
relationships were best captured by a regression model which included the two year lagged continuously compounded share price index return, $R_t$, as a further explanatory variable, i.e.

$$\ln T_t = -2.87 + 0.915 \ln T_{t-1}^* + 0.42 R_{t-2}^* + \ln T \xi_t,$$

$$\ln T \xi_t \sim \text{i.i.d. } \mathcal{N}(0, 0.184^2).$$

The following ERCH model was found to capture the observed T-note yield structure (estimated parameter standard errors are shown in brackets below the parameters). The statistical significance of the ERCH model is consistent with the observed nonlinear dependence and heteroscedasticity, which are modelled explicitly.

$$\ln T_t = -2.87 + 0.87 \ln T_{t-1}^* + 0.44 R_{t-2}^* + \ln T \xi_t,$$

(0.004) (0.004) (0.002)

$$\ln S_t = -1.02 + 0.301 \ln T_{t-1}^*,$$

(0.235) (0.090)

i.e. $\ln T \xi_t = S_t Z_t = 0.361 \ln T \xi_{t-1}^* \xi_{t-1} \sim \mathcal{N}(0, 0.361 \ln T \xi_{t-1}^* \xi_{t-1}^* )^2$.

The volatility structure can also be expressed in the form

$$\ln |\xi_t| = \ln S_t + \ln |Z_t|$$

$$= -1.02 + 0.3 \ln |\xi_{t-1}| + \ln |Z_t|$$

$$= \text{constant} + \ln |Z_t| + 0.3 \ln |Z_{t-1}| + 0.3^2 \ln |Z_{t-2}| + .. ,$$

which can be seen to be akin to an AR(1) process in the logarithm of the magnitude of the model residuals, or an exponentially weighted moving average of past shocks.\(^8\)

\(^8\) the $\ln |Z_t|$ form a sequence of i.i.d. random variables, each with a $\frac{1}{2} \ln \chi^2_i$ distribution.
A consequence of the nonlinear dependence in the T-note series is that large magnitude shocks to the rates (of either sign) tend to occur in clumps. Similarly small magnitude shocks tend to cluster together. For instance, based on the ERCH model, if the shock in the previous year was $\xi_{t-1} = \pm 0.1$, then the shock in the current year $\xi_t$ would be Normally distributed with standard deviation 0.18. The likelihood that the magnitude of the current year’s shock would exceed 0.3 would be only about one in ten. However, if the shock in the previous year was $\pm 0.4$, then the shock in the current year would be Normally distributed with standard deviation 0.27, and the likelihood that the magnitude of the current year’s shock would exceed 0.3 would be about one in three-and-a-half. Hence the occurrence of a “large” shock (magnitude $> 0.3$) would be about three times as likely in the second scenario as in the first. Contrast the above with the AR(1) model in which the standard deviation of the shock would always be 0.20, and the likelihood that the magnitude of the shock in any year would exceed 0.3 would always be about one in seven-and-a-half, regardless of the past history.

The following table compares the fit of the three models described above. Each model is significantly better than the preceding model. $s.e.$ refers to the model residual standard error (deviation).

<table>
<thead>
<tr>
<th></th>
<th>AR(1)</th>
<th>Regression</th>
<th>ERCH</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\ln L$</td>
<td>8.09</td>
<td>12.06</td>
<td>18.63</td>
</tr>
<tr>
<td>$\Delta \ln L$</td>
<td>-</td>
<td>3.97</td>
<td>10.54</td>
</tr>
<tr>
<td>$s.e.$</td>
<td>0.203</td>
<td>0.185</td>
<td>0.188</td>
</tr>
<tr>
<td>$\gamma_1(z)$</td>
<td>1.14</td>
<td>1.01</td>
<td>0.88</td>
</tr>
<tr>
<td>$\gamma_2(z)$</td>
<td>3.90</td>
<td>3.46</td>
<td>1.76</td>
</tr>
<tr>
<td>$\gamma_1(z)^*$</td>
<td>0.06</td>
<td>-0.15</td>
<td>0.12</td>
</tr>
<tr>
<td>$\gamma_2(z)^*$</td>
<td>1.07</td>
<td>0.16</td>
<td>-0.57</td>
</tr>
</tbody>
</table>

$\gamma_1(z)$ and $\gamma_2(z)$ are the sample residual skewness and excess kurtosis, respectively. Examination of the residuals indicates that June 1974 is again

---

9 $z$ represents the sample standardised residuals, i.e. the model errors divided by the model standard deviation (which is a function of time in the ERCH case).
an outlier. \(\gamma_1(z)^*\) and \(\gamma_2(z)^*\) have been based on the data adjusted for June 1974, as described in section 2.1\(^\text{10}\).

The difference in the maximum log-likelihoods of the ERCH and AR(1) models (10.54) is distributed as \(\frac{1}{2} \chi^2_2\) under the null hypothesis of no difference between the two models. The ERCH model therefore significantly better fits the data at much less than the 0.1\% level. Notably, the ERCH model explicitly models the nonlinear dependence and conditional heteroscedasticity observed in the interest rate process.

Note that the share price return series, \(\{R_t\}\), is itself very much subject to nonlinear dependence and time varying volatility\(^\text{11}\), hence the degree of nonlinear dependence and conditional heteroscedasticity captured in the above ERCH model is larger than it appears to be from the above model equation.

### 4.2 Annual 10 year Bond Yield Models

The second set of results presented is for the series of Australian annual 10 year Commonwealth Government Bond yields over the period June 1950 to June 1993. The rate parameter of the fitted AR(1) model was \(\alpha = 0.93\). The corresponding \(\ln L\) was 31.33. As discussed in section 3.2, the Wilkie (1986) bond model was found to collapse to an AR(1) model when fitted to the Australian data using maximum likelihood.

The following ERCH model was found to best capture the observed 10 year Bond yield structure:

\[
\ln B_t = -2.62 + 0.92 \ln B_{t-1}^* + 0.26 F_{t-1}^* + 0.23 R_{t-2}^* + \ln B \xi_t,
\]

\[
\ln S_t = -1.48 + 0.27 \ln |\ln B \xi_{t-1}|.
\]

The corresponding \(\ln L\) was 44.62, representing an increase of 13.29 over the AR(1) model, which is extremely significant (being distributed as \(\frac{1}{2} \chi^2_3\) under the null hypothesis of no difference between the two models). The ERCH model can again be seen to provide a significantly better fit to the

\(^{10}\) note that the models were fitted to the unadjusted data however, the adjustment was only made for the purposes of calculating more meaningful residual skewness and excess kurtosis measures.

\(^{11}\) see Harris (1994).
observed yields than the commonly used first order Autoregressive model. The 10 year Bond ERCH model is very similar in form to the 13 week T-note ERCH model. In the case of long bond yields however, past inflation was found to significantly capture structure in addition to that captured by past share returns.

Again note that the ERCH model captures the observed nonlinear dependence and conditional heteroscedasticity, both explicitly via the conditional volatility equation, and implicitly via conditionally heteroscedastic explanatory variables, i.e. \( \{F_t\} \) and \( \{R_t\} \) in this case\(^{12}\).

The contemporaneous residuals from the two ERCH annual interest rate models were found to be highly correlated, as expected, i.e. random shocks to short and long rates tended to be correlated. This feature can be modelled by assuming that \( \mathbb{E}(\ln T_t \cdot \ln B_s) = 0.6 \) when \( t = s \), and zero when \( t \neq s \).

### 4.3 Annual ERCH System

The ERCH models presented earlier include past share price returns and inflation as explanatory variables. To enable annual interest rates to be generated from the ERCH models, one needs to also model share price returns and inflation, as a minimum. I therefore present an ERCH system that has been fitted to Australian data for the years ending June 1950 to June 1993. The series modelled are real GDP growth, \( \{G_t\} \), price inflation, \( \{F_t\} \), share price index returns, \( \{R_t\} \), 13 week Treasury Note yields, \( \{T_t\} \), and 10 year Government Bond yields, \( \{B_t\} \). The first three are continuously compounded annual rates, while the last two are nominal annual rates. Validation and discussion of the ERCH share return and inflation models can be found in Harris (1994) and Harris (1995), respectively.

\(^{12}\) refer to Harris (1995) and Harris (1994), respectively.
\[
\begin{pmatrix}
G_t \\
F_t \\
R_t \\
\ln T_t \\
\ln B_t
\end{pmatrix} = \begin{pmatrix} 0.038 \\ 0.062 \\ 0.068 \\ -2.87 \\ -2.62 \end{pmatrix} + \begin{pmatrix} 0 & -0.305 & 0.066 & 0 & 0.40 & 0 & 0 \\ 0.42 & 0.85 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -3.15 \\ 0 & 0 & 0.87 & 0 & 0 & 0 & 0.44 \\ 0 & 0 & 0.26 & 0 & 0.92 & 0 & 0.23 \\ -4.01 & -5.65 & -3.88 & -1.02 & -1.48 & 35 & 0 & 9 & 0 & 0 & 0.30 & 0 & 0.27 \end{pmatrix} \begin{pmatrix}
G_{t-1} \\
F_{t-1} \\
R_{t-1} \\
\ln T_{t-1} \\
\ln B_{t-1} \\
(G \cdot R)_{t-2} \\
\end{pmatrix} + \xi_t
\]

Examination of the model residuals when the above ERCH system is fitted to Australian annual data over the period June 1950 to June 1993 reveals no significant serial correlation in the five series of residuals, nor any significant cross-correlation between the residuals and the previous year's residuals of any other series. As expected however, there is significant contemporaneous cross-correlation between the series of standardised residuals, particularly between the two interest rate series, as noted earlier (0.6). The Bond yield standardised residuals were found to be significantly correlated to the contemporaneous standardised residuals of each of the other four series. The other significant contemporaneous cross-correlations in the standardised residuals were between real GDP growth and inflation (0.3), and between inflation and Treasury Note yields (0.3).

To generate the contemporaneously correlated standardised residuals, \{Z_t\}, one decomposes the contemporaneous correlation matrix, \(\Sigma_z\), to obtain a lower triangular "square root" matrix, \(L\), which is multiplied by a vector of i.i.d. standard Normal random variates. The approach is summarised below.
and $\mathbf{Z}_t = \mathbf{L} \mathbf{Z}_t$, where $\mathbf{Z}_t$ is a $5 \times 1$ vector of i.i.d. $N(0,1)$ variates.

The system shown was derived by fitting ERCH models to the individual data series. To generate future scenarios one would probably wish to adjust the unconditional mean vector to reflect one's expectations of the long run differentials between the series. The experienced mean share return of 6.8% p.a. in particular seems relatively low compared to the other series and most commentators expectations. To use the above ERCH system to generate future scenarios one would need to restrict the values of some of the elements of $\mathbf{\Phi}_t$ to improve the stability of the nonlinear system. Further refinements such as the use of reflecting and absorbing barriers may also be employed. Refer to Harris (1995, section 6.) for a discussion of these aspects.

Simulation was used to generate 5,000 Treasury Note and 10 year Bond yields from each of the ERCH system\textsuperscript{13} and AR(1) models. The characteristics of the changes in the yields (the log-yields, i.e. the continuously compounded rates of change, e.g. $\Delta \ln T_t$), and their magnitudes (e.g. $|\Delta \ln T_t|$), were of most interest. The characteristics of the simulated series are summarised below, including the percentiles of the unconditional distributions.

---

\textsuperscript{13} $\mathbf{\Phi}_t$ was “data limited”, a reflecting barrier was applied to the inflation series at 30%, and absorbing barriers were applied to the share price return series at $\pm 85\%$, as described in Harris (1995, section 6.). The impact of the barriers was minimal.
Characteristics of the Simulated Yield Series

<table>
<thead>
<tr>
<th></th>
<th>ERCH System</th>
<th>AR(1) Models</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\gamma_1$</td>
<td>$\gamma_2$</td>
</tr>
<tr>
<td>$\gamma_1$</td>
<td>-0.02</td>
<td>0.57</td>
</tr>
<tr>
<td>$\gamma_2$</td>
<td>0.05</td>
<td>0.65</td>
</tr>
<tr>
<td>$\rho(1)$</td>
<td>1.29</td>
<td>1.32</td>
</tr>
<tr>
<td>$\frac{1}{5}$</td>
<td>-0.02</td>
<td>-0.02</td>
</tr>
<tr>
<td>$\frac{1}{10}$</td>
<td>0.99</td>
<td>0.93</td>
</tr>
<tr>
<td>$\frac{1}{5}$</td>
<td>-38.6%</td>
<td>-29.3%</td>
</tr>
<tr>
<td>$\frac{1}{10}$</td>
<td>-21.3%</td>
<td>-15.7%</td>
</tr>
<tr>
<td>$\frac{9}{10}$</td>
<td>1.4%</td>
<td>2.7%</td>
</tr>
<tr>
<td>$\frac{9}{10}$</td>
<td>0.7%</td>
<td>1.5%</td>
</tr>
</tbody>
</table>

Note the differences in the excess kurtosis of the yield changes and the serial correlation in the magnitude of the changes between the ERCH system and the AR(1) models. The differences are more apparent when one considers conditional distributions. To illustrate the point, consider the magnitude of yield changes given that the previous year's change was "large". For this purpose I defined "large" to be $|\ln(T_{i+1})| > 25\%$ and $|\ln(B_{i+1})| > 15\%$, which corresponds to the largest 20-25% of the 5,000 simulated changes. In terms of nominal T-note yield changes, "large" would imply that yields increased by more than 28% (i.e. by a factor of $>1.28$) or fell by more than 22% in a year (i.e. to <78% of the yield a year earlier), e.g. from 7% to 9%, or from 11% to 14%. Similarly for nominal 10 year Bond yield changes, "large" would imply that yields increased by more than 16% (i.e. by a factor of $>1.16$) or fell by more than 14% in a year, e.g. from 9.5% to 11%, or from 12% to 14%. Nonlinear dependence in the ERCH system can be expected to show up as clustering of large changes, and therefore a higher expected conditional magnitude. The characteristics of the conditional simulated distributions are summarised in the following table.
Characteristics of the Conditional Magnitudes of the Yield Changes

<table>
<thead>
<tr>
<th></th>
<th>ERCH System</th>
<th>AR(1) Models</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(</td>
<td>\ln T</td>
</tr>
<tr>
<td>50%'ile</td>
<td>18.0%</td>
<td>9.5%</td>
</tr>
<tr>
<td>90%'ile</td>
<td>46.3%</td>
<td>25.8%</td>
</tr>
<tr>
<td>95%'ile</td>
<td>56.1%</td>
<td>31.2%</td>
</tr>
<tr>
<td>P(&gt;15%)</td>
<td>0.58</td>
<td>0.31</td>
</tr>
<tr>
<td>P(&gt;25%)</td>
<td>0.34</td>
<td>0.11</td>
</tr>
<tr>
<td>P(&gt;40%)</td>
<td>0.15</td>
<td>0.02</td>
</tr>
</tbody>
</table>

The distribution of conditional changes under the ERCH system is shifted to the right and fatter tailed than under the AR(1) models, as expected. This result is consistent with the discussion in section 4.1. Note that under the ERCH system the likelihood of a large change in yields, given that the change in the previous year was large, is significantly higher than the case with the AR(1) model, and the unconditional likelihood in the ERCH case (shown in the previous table).

The above characteristics can be represented graphically in the form of conditional probability density functions. The conditional probability density functions were estimated using kernel density estimates. The method involves conceptually centring a kernel function, a Normal density in this case, at each observed data point, \(x_i\). The kernel density estimate of the probability density function at point \(x\) is calculated as the average of the contributions of the Normal densities from each observed data point. In formulae:

\[
\tilde{f}(x) = \frac{1}{N} \sum_{i=1}^{N} K_b(x - x_i)
\]

where \(b\) is the “band-width” of the kernel function, and represents a smoothing parameter (0.075 and 0.05 were used for the T-note and Bond estimates, respectively). The resulting conditional density functions are shown on the next page.
Conditional Density of the Magnitude of Changes in Log T-note Yields

Magnitude of Change in Log Yields
Conditional Density of the Magnitude of Changes in Log 10 year Bond Yields
4.4 *Monthly Treasury Note Yield Models*

In this, the final section, models are fitted to Australian 13 week Treasury Note yields at monthly intervals over the period 1980-93. The AR(1) model was found to be insignificantly different to the simple Random Walk model. The maximum likelihood estimate of the monthly rate parameter, \( \alpha \), is 0.991, which is consistent with the annual rate parameter of 0.92 from section 4.1, since \( 0.991^{12} = 0.90 \). Given the magnitude of the annual rate parameter, it is not surprising that the AR(1) process is indistinguishable from a Random Walk model at monthly or higher frequencies.

ERCH models that exploit the serial correlation and nonlinear dependence observed in the monthly series in sections 2.3 and 2.4 were found to be highly significant. Some model fitting results are shown in the following table. The Normality of the standardised residuals is examined for excess kurtosis, \( \gamma_2(z) \), and nonlinear dependence, \( r_{lny}(1) \), both of which should ideally be zero. The 2½% critical values are 0.74 and 0.15 respectively (one-sided in both cases).

<table>
<thead>
<tr>
<th>Model</th>
<th>( \ln L )</th>
<th>( \gamma_2(z) )</th>
<th>( r_{lny}(1) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random Walk:</td>
<td>185.7</td>
<td>2.94</td>
<td>0.175</td>
</tr>
<tr>
<td>( \ln T_t = e_t )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \ln \sigma_t = -2.524 )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AR(1):</td>
<td>185.8</td>
<td>2.88</td>
<td>0.215</td>
</tr>
<tr>
<td>( \ln T_t = -2.172 + 0.99 \ln T_{t-1} + e_t )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \ln \sigma_t = -2.525 )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ERCH(( \ln L_e_t, l )):</td>
<td>193.9</td>
<td>2.32</td>
<td>-0.05</td>
</tr>
<tr>
<td>( \ln T_t = e_t )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \ln \sigma_t = -1.89 + 0.20 \ln</td>
<td>L_e_t, l</td>
<td>)</td>
<td></td>
</tr>
<tr>
<td>ERCH(( e_t, \ln L_e_t, l, \ln L_e_t, s, l )-GED):</td>
<td>214.5</td>
<td>1.57</td>
<td>-0.04</td>
</tr>
<tr>
<td>( \ln T_t = 0.19 e_{t-7} + e_t )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \ln \sigma_t = -1.494+0.177\ln L_e_t, l+0.16\ln L_e_t, s, l )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( z_t \sim \text{GED}(\nu = 1.19) )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ERCH(( e_t, l, \ln \sigma_t, l, \ln T_{t-1}, l, \ln T_{t-2}, l, \ln B_{t, l} )):</td>
<td>218.4</td>
<td>0.66</td>
<td>-0.01</td>
</tr>
<tr>
<td>( \ln T_t = 0.23 e_{t-1} + e_t )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \ln \sigma_t = 0.64+0.86 \ln \sigma_{t-1} - 1.365 \ln T_{t-1} \ + 1.084 \ln T_{t-2} + 0.78 \ln B_{t-1} )</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Where the residual excess kurtosis remained large the model could generally be improved by assuming that the errors were drawn from a Generalised Error Distribution (GED)\(^{14}\) or Student \(t\) distribution rather than from a Normal distribution. The GED distribution was found to give the higher log-likelihood in the case of the ERCH models fitted. The standard GED density function is given by

\[
f(z) = \frac{\nu \Gamma(\frac{3}{\nu})^\frac{1}{2}}{2 \Gamma(\frac{1}{\nu})^3} \cdot e^{-\left[ \frac{\nu}{\Gamma(\frac{1}{\nu})^3} \right] \cdot \frac{z^2}{2}},
\]

and the corresponding log-likelihood function, which needs to be maximised, is therefore

\[
\ln L = N \ln(\frac{\nu}{\sigma}) - \frac{N}{2} \ln \left[ \frac{\Gamma(\frac{3}{\nu})}{\Gamma(\frac{1}{\nu})^3} \right] - \sum_{i=1}^{N} \left( \ln \sigma_i + \left( \frac{c_i^2}{\sigma_i^2} \cdot \frac{\Gamma(\frac{3}{\nu})}{\Gamma(\frac{1}{\nu})} \right)^\frac{1}{2} \right).
\]

When the GED parameter, \(\nu\), equals 2 then the standard Normal distribution results. When \(0 < \nu < 2\) the distribution is leptokurtic.

Consider the ERCH\((e_{t,7}; \ln e_{t,1}, \ln e_{t,8})\)-GED model, and compare it with the observed statistical structure of the monthly change in log T-note yields over the period 1980-93. Recall that the series was found to be leptokurtic (\(\gamma_2 = 2.94\)), possessed significant autocorrelation at lag 7 months (\(r(7) = 0.21\)), and displayed significant autocorrelation in the log-absolute deviations (with the largest being \(r_{\ln \ln T-\mu}(8) = 0.36\)). Note that the model residual excess kurtosis of 1.57 is accounted for by assuming that the standardised residuals come from a Generalised Error Distribution with parameter 1.19, rather than a standard Normal distribution. The ERCH model is clearly superior to the Random Walk model. The increase in the maximum log-likelihood is distributed as \(\chi^2_4\) under the null hypothesis that the data was generated from the Random Walk model, thus the observed increase of 28.8 is significant at much less than the 1\% level. Further the Random Walk

\(^{14}\) also referred to as the Exponential Power Distribution.
LOW FREQUENCY STATISTICAL INTEREST RATE MODELS

model residuals display significant excess kurtosis and nonlinear dependence, features that are accounted for by the ERCH model.

As has already been observed, non-Normality and heteroscedasticity are more prominent in monthly data than annual data. With monthly data less can be said about the conditional mean than is the case with annual data, but more can be said about the conditional volatility. In fact it appears that the level of interest rates is virtually unpredictable in the short term (consistent with a Random Walk), hence the relative lack of structure in the monthly conditional mean equations. This is consistent with the results of Deaves (1993), who used various models to predict nominal quarterly Canadian interest rates. He found that none of the models were able to more successfully predict interest rate levels than a naive no change model (i.e. a Random Walk model) for periods out to a few years.

The monthly ERCH models are consistent with the annual ERCH models in a broad sense. The conditional volatility equations in particular are consistent. However the monthly models do not reflect the long term relationships between the series, which are better captured by the annual ERCH models. Statistically this is reflected in the inability to distinguish between models of the level of rates and the corresponding model of the change in the rates, which in the monthly models is manifested in a coefficient of $\ln T_{t,t}$ which is insignificantly different from one. Thus over an extended projection period the monthly model may “lose” its way in terms of the relativities between the various series levels. The fact that the data frequency has a major impact on the nature of the model fitted should not be surprising, since the information set available and the statistical structure of the series being modelled both alter with increasing temporal aggregation.

ACKNOWLEDGMENTS

I would like to acknowledge the support received from my employer, the Australian Mutual Provident Society, Finance Division, without whom my attendance at the 5th AFIR International Colloquium would not be possible. The views expressed in this paper, and any omissions made, are of course my own.
REFERENCES


LOW FREQUENCY STATISTICAL INTEREST RATE MODELS


