ESTIMATION OF THE HEATH-JARROW-MORTON MODEL VIA THE KALMAN FILTER: A BOOTSTRAP ANALYSIS

Ramaprasad Bhar, Carl Chiarella

*School of Finance & Economics
University of Technology, Sydney
PO BOX123, Broadway
NSW – 2007
AUSTRALIA
Fax: +61 2 9514 7711

1#School of Banking and Finance
University of New South Wales

Abstract:

This paper considers the Heath-Jarrow-Morton (HJM) model of the term structure of interest rates for a fairly general specification of forward rate volatility, including stochastic variables. Estimation of this volatility function is at the heart of the identification of the HJM model. Reduction of the model to state space form is discussed and use of the Kalman filter as an estimation technique is proposed. Since typical data sets are small, a bootstrap procedure is used to determine the statistical significance of the estimates. A Monte-Carlo experiment is used to compare the bootstrap and “true” small-sample distributions of the estimates of the parameters of the volatility function.

Key words:  Heath-Jarrow-Morton model, Arbitrage-free, Forward rate volatility functions, Non-linear filtering, Bootstrap.

JEL Classification: G12, G13

Financial support from ARC grant A00103864 in the development of this work is gratefully acknowledged.

* corresponding author: e-mail: carl.chiarella@uts.edu.au
1. **INTRODUCTION**

A major advance in the modelling of the term structure of interest rates and of interest rate derivatives more generally was achieved by Heath, Jarrow and Morton (HJM, 1992). Their approach has three features which made it an advance on previous approaches of Merton (1973), Vasicek (1977) and Ball and Torous (1983) to the evaluation of derivative security prices in a stochastic interest rate environment. Firstly, they choose as the driving dynamic variable the instantaneous forward rate of interest instead of the instantaneous spot rate of interest or the bond price as in the alternative approaches referred to earlier. Since these various rates are related it should in principle make no difference which one chooses as the driving dynamics, however in practice the choice of the instantaneous forward rate proves to be far more convenient and encompasses more general specifications than the other two. Secondly, HJM make use of the concepts of change of probability measure and Girsanov's theorem in particular to express prices of interest rate derivative securities in a form which does not involve the empirically troublesome market price of interest-rate risk. Hence this approach avoids making assumptions about investor preferences. Thirdly, because of its choice of the instantaneous forward rate as the driving dynamics the term structure of interest rates that it implies is consistent with the currently observed yield curve.

Analogously to the role of the volatility parameter in the Black-Scholes stock option model, the key input variable to the HJM modelling framework is the function which describes the volatility of the instantaneous forward rate. The major task in implementations of the HJM model is an appropriate choice of this function. To date there has not been a great deal of work on the empirical estimation of the HJM term structure model apart from HJM (1990), Flesaker (1993), Amin and Morton (1994) and Bliss and
Ritchken (1996). Perhaps one reason for the paucity of empirical studies is the difficulty in expressing the equations of the HJM model in a form to which standard statistical estimation methodology can be applied. This difficulty stems in part from the fairly general specification of the instantaneous forward rate volatility that the HJM theory allows, which means that in general the system dynamics are non-Markovian. We make use of results developed by a number of authors including Carverhill (1994), Cheyette (1992), Ritchken and Sankarasubramanian (1995a), Bhar and Chiarella (1997a), Inui and Kijima (1998) and Chiarella and Kwon (2001) that show that by a specific (but nevertheless fairly general) choice of functional form for the volatility of the instantaneous forward rate it is possible to express the system dynamics in Markovian form. More importantly it is then possible to recast the system dynamics in state-space form, so that one may approach the estimation problem by use of the extended Kalman filtering methodology outlined for example by Harvey (1989). In Bhar and Chiarella (1997b) this technique has been applied to estimate the HJM model in the Australian market using futures contracts on 90-day bank bill and 3 year Treasury bond data. Since the futures contracts of sufficient liquidity trade only over the last three to four months of the contract life, one is typically dealing with short sample sizes so that it is difficult to rely for statistical inference upon the standard errors of the parameter estimates obtained by maximum likelihood estimation. Whilst these are asymptotically normally distributed it may be problematic to make this assumption when the samples are of the small length typical of most empirical studies, usually from 60 to 120. Hence the Monte-Carlo bootstrap methodology was employed to obtain a bootstrap sample distribution and to place statistical confidence intervals around the estimates.
The above approach to bootstrapping state space models has been proposed by Stoffer and Wall (1991). These authors have also suggested judging the reliability of the bootstrap sample distribution by comparing it with an approximation to the true small sample distribution obtained by a parametric Monte Carlo experiment. Ritchken and Sankarasubramanian (1995b) highlight the sensitivity of interest rate option prices in the HJM framework to the parameters which specify the forward rate volatility structure. It is, therefore, extremely important that when the parameters of the volatility structure are recovered from the available market prices the precision of the estimates are clearly established. Hence there is a need to compare the bootstrap sample distribution obtained from empirical estimations with the true small sample distribution.

In this paper we compare the bootstrap sample distribution of the Kalman filter estimators of the HJM model with the true small-sample distribution. We do this by simulating the model for known sets of parameters which generate typical market yield curves, and thereby producing time-series of bond prices of length usually dealt with in actual market studies. The Kalman filter estimates of the known parameters are obtained and a bootstrap sample distribution for them is generated. The approximation to the true small-sample distribution is also generated by a parametric Monte Carlo experiment. We are thereby able to draw some conclusions about the statistical reliability of the extended Kalman filter and associated bootstrapping methodology as applied to the HJM model.

Our major conclusion is that the bootstrap resampling technique is a viable one for precise estimation of the parameters of the volatility function that characterises the HJM model.
The plan of the paper is as follows:- Section 2 reviews the relevant details of the HJM model under a class of volatility functions important for applications. Section 3 expresses the stochastic differential equation in state space form and sets up the filtering algorithm. Section 4 describes the bootstrap experiment and procedures. Section 5 describes the selection of parameters for the volatility function and initial yield curve. Section 6 discusses the empirical results. Section 7 concludes.

2. MARKOVIAN REPRESENTATION OF THE HEATH-JARROW-MORTON MODEL

In the HJM framework assume that the forward rate process is driven by the one factor stochastic integral equation

\[ f(t, T) - f(0, T) = \alpha(v, T, \cdot) \, dv + \sigma(v, T, \cdot) \, dW(v). \] (1)

Here \( \alpha(t, T, \cdot) \), \( \sigma(t, T, \cdot) \) are respectively the drift and diffusion coefficients at time \( t \) for an instrument maturing at time \( T \) (the blank third argument indicates possible dependence on the stochastic variable \( r(t) \), the instantaneous spot rate of interest)\(^1\), and \( dW(v) \) are the increments of the single Wiener process driving the stochastic market fluctuations of the forward rate curve.

The critical result of HJM is that if the economy is arbitrage free then the drift cannot be chosen independently of the volatility function but rather according to

\(^1\)In HJM, \( \alpha \) and \( \sigma \) can also depend on the forward rate itself or even more generally on the path histories. We do not allow such dependence here as the state-space representation would then assume a far more
\[ \alpha(v, T, \cdot) = \sigma(v, T, \cdot) \left[ \int_0^t \sigma(v, y, \cdot) dy - \psi(v) \right], \]

where \( \psi(v) \) is the market price of risk. As a consequence the stochastic process for the forward rate becomes

\[ f(t, T) = f(0, T) + \int_0^1 \sigma(v, T, \cdot) \int_0^T \sigma(v, y, \cdot) dy \, dv + \int_0^1 \sigma(v, T, \cdot) d\tilde{W}(v). \quad (2) \]

Here \( d\tilde{W}(t) := d\left[ W(t) \int_0^t \psi(s) \, ds \right] \) are the increments of a Wiener process from the equivalent martingale probability measure, which arises in the HJM theory by application of Girsanov's theorem.

From equation (2) it is possible to derive the stochastic differential equation for the instantaneous spot rate \( r(t) = f(t, t) \) as

\[ dr(t) = \left[ f_2(0, t) + \frac{\partial}{\partial t} \int_0^1 \int_0^T \sigma(v, y, \cdot) dy \, dv + \int_0^1 \sigma(v, t, \cdot) d\tilde{W}(v) \right] dt \]

\[ + \sigma(t, t, \cdot) d\tilde{W}(t). \quad (3) \]

---

HJM show that with the spot interest rate process (3) under the equivalent measure, the market observed bond price at $t$ of a pure discount bond maturing at $T$ is given by

$$P(t, T) = E_t \left[ \exp \left( -\int_t^T r(s) ds \right) \right],$$

where $E_t$ is the expectation at time $t$ under the equivalent measure. An application of Ito's lemma readily shows that $P(t, T)$ and its logarithm $B(t, T) = \ln P(t, T)$ are driven by the stochastic differential equations

$$\frac{dP(t, T)}{P(t, T)} = r(t) dt + \left[ -\int_t^T \sigma(t, v, \cdot) dv \right] d\bar{W}(t)$$

$$\equiv r(t) dt + \sigma_B(t, T, \cdot) d\bar{W}(t),$$

and

$$dB(t, T) = \left[ r(t) - \frac{1}{2} \sigma_B^2(t, T, \cdot) \right] dt + \sigma_B(t, T, \cdot) d\bar{W}(t)$$

$$\equiv \mu_B(t, T, \cdot) dt + \sigma_B(t, T, \cdot) d\bar{W}(t).$$

The principle difficulty in application of the HJM model lies in the second and third terms in the expression for the drift in the stochastic differential equation (3) for the instantaneous spot rate. These expressions involve integrals over the history of the process and the shock process up to time $t$ and so make the stochastic process driving bond prices non-Markovian. Most applications of the HJM approach essentially seek some simplification, that renders the driving dynamics Markovian.

---

2 It is important to realise that even though these SDE’s are under the equivalent measure, the bond prices are the market observed ones, unlike the $r(t)$ in equation (3).
Bhar and Chiarella (1997a) show that by assuming for the forward rate volatility the specific functional form.

\[ \sigma(t, T, \cdot) = [a_0 + a_1(T-t) + \ldots + a_n(T-t)^n]e^{-\lambda(T-t)}G(r(t)) , \]

where \( G \) is a sufficiently well behaved function, the dynamics driving \( r(t) \) and \( B(t, T) \) turn out to be Markovian. However the dimension of this Markovian system increases rapidly with \( n \). In this paper we consider the special case

\[ \sigma(t, T, \cdot) = a_0 e^{-\lambda(T-t)}r(t)^\gamma , \]  \hspace{1cm} (6)

so that

\[ \sigma_b(t, T, \cdot) = \frac{a_0}{\lambda}[e^{-\lambda(T-t)} - 1]r(t)^\gamma , \] \hspace{1cm} (7)

where \( a_0, \lambda \) are parameters to be determined and \( \gamma \) will be set at some pre-determined values in the empirical analysis below. The specification (6) captures the essential features of the time dependence of the forward rate volatility and also allows, via the \( r(t)^\gamma \) term, for some dependence on stochastic variables.

3. STATE SPACE FORM AND THE FILTERING EQUATIONS

Bhar and Chiarella (1997a) show that the non-Markovian stochastic dynamics system (3), (5) for \( r(t) \), \( B(t, T) \), with \( \sigma(t, T, \cdot) \) given by (6) can be expressed as the Markovian stochastic differential system
\[
\begin{align*}
\text{d}S(t) &= \left[ J(t) + H(S(t), t)S(t) \right] \text{d}t + V(S(t), t) \text{d}\tilde{W}(t), \quad \text{(8)}
\end{align*}
\]

where

\[
S(t) \equiv [B(t, T), r(t), \phi(t)]', \quad \text{(9a)}
\]

\[
J(t) \equiv [0, f_2(0, t) + \lambda f(0, t), 0]', \quad \text{(9b)}
\]

\[
V(S(t), t) \equiv [a_0 r(t)^\gamma (e^{-\lambda(T-t)} - 1)/\lambda, a_0 r(t)^\gamma, 0]', \quad \text{(9c)}
\]

\[
H(t) = \begin{bmatrix}
0 & h_{12} & 0 \\
0 & -\lambda & a_0^2 \\
0 & h_{32} & -2\lambda
\end{bmatrix}, \quad \text{(9d)}
\]

with

\[
h_{12} = 1 - \frac{1}{2} r(t)^{2\gamma-1} \left[ e^{-\lambda(T-t)} - 1 \right]^2/\lambda^2, \quad h_{32} = r(t)^{2\gamma-1},
\]

and

\[
\phi(t) \equiv \int_0^t r(u)^{2\gamma} e^{-2\lambda(t-u)} \text{d}u. \quad \text{(9e)}
\]

The state variable \(\phi(t)\), which summarises characteristics of the path history of the instantaneous spot rate process, is not directly observable. Since the dynamics of the instantaneous spot rate \(r(t)\) in equation (3) are under the equivalent measure, this is also
not directly observable. So the only element of the state vector \( S(t) \) which is considered observable is \( B(t, T) \), hence we have the observation vector (in this case a scalar)

\[
Y(t) = CS(t) + \epsilon(t),
\]

where, \( \epsilon(t) \), the observation error is introduced to account for the spread in open-close or high-low quotes available from the market, and

\[
C = [1, 0, 0].
\]

It should be pointed out that, similar reductions to a Markovian system have been found in the earlier cited works of Cheyette (1992), Carverhill (1994), Ritchken and Sankarasubramanian (1995a), Inui and Kijima (1998), de Jong and Santa-Clara (1999), Bhar, Chiarella, El-Hassan and Zheng (2000) and Chiarella and Kwon (2001), though without the specific emphasis on state space form and non-linear Kalman filtering.

We will also assume that the continuous-discrete system (in the sense used in the filtering literature eg. Jazwinski (1970), page 195) defined by the equations (8) and (10) have error sequences that are not correlated. Furthermore, the distributional characteristic of \( \epsilon(t) \) is defined by \( \epsilon(t) \sim N(0, \sigma_\epsilon), \sigma_\epsilon > 0 \). The observation noise may also be present due to thin trading. Since the distributional characteristic of such noise is likely to be non-Gaussian, the Kalman filter implementation would be quite different. However, we prefer to leave incorporation of such refinements for future research.

---

3 Even if we transformed to the historical measure, \( r(t) \) should preferably be treated as unobservable since the shortest rate that could be used in Australia would be a 30-day rate or a weekly rate. In the U.S. data on overnight rates exist but are considered too "noisy" for reliable empirical work.
In equation (8) the specification of the initial forward rate term structure \( f(t,T) \) is required. Implementation details of this forward rate curve is discussed in section 5 on empirical results. The initial value of the spot rate is obtained from \( f(t,t) \) at \( t = 0 \).

For ease of exposition, the equation (8) is expressed as,

\[
dS(t) = F(S(t); \theta) \, dt + V(S(t); \theta) \, dW(t),
\]

where \( \theta \equiv [a, \lambda] \). In general, \( F(.) \) and \( V(.) \) will be non-linear in both the state variables as well as the parameters \( \theta \). Estimation of the parameter \( \theta \) in equation (11) from discretely observed data will involve some form of discretisation from which the conditional moments over successive time intervals can be calculated, and for this purpose the Euler-Maruyama scheme has been adopted; see Kloeden and Platen (1992).

With \( \zeta \sim N(0,1) \) and \( \delta_k \) being the length of the time interval \( (t_k, t_{k+1}) \), the Euler-Maruyama scheme applied to equation (11) results in,

\[
S_{k+1} = S_k + F(S_k; \theta) \, \delta_k + V(S_k; \theta) \, \sqrt{\delta_k} \, \zeta_k,
\]

Where \( S_k = S(t_k) \). The equations (12) and (10) now constitute the state transition and the measurement or observation equations respectively in the context of the discrete Kalman filter. Kalman filter methodology is derived from the maximum likelihood approach. Essentially, we enter each period with a prior belief about the distribution of the state vector and update this prior belief after confronting the observation vector. The
observation vector is, however, noisy. This leads to an adjustment of the estimate of the state vector, which is a function of the error in the previous estimate to predict the observation vector. It is this prediction error that is used to generate the likelihood function that is to be maximised.

From equation (12) the first two conditional moments of the evolution of the state vector $S(t)$ are given by,

$$E(S_{k+1}|S_k) = S_k + F(S_k; \theta)\delta_k,$$  \hfill (13)

and,

$$Q_{k+1} = \text{Cov}(S_{k+1}|S_k) = V(S_k; \theta)V^T(S_k; \theta)\delta_k,$$  \hfill (14)

These last two quantities feed into the standard Kalman filter mean and variance update equations, which we set out here for completeness of exposition. From equation (13) the best forecast of $S$ at $t_{k+1}$ made at $t_k$ (knowing the observation $Y_k$ at $t_k$) is,

$$\hat{S}_{k+1|k} = \hat{S}_{k|k} + F(\hat{S}_{k|k}; \theta)\delta_k,$$  \hfill (15)

and the best forecast of variance of $S_{k+1}$ is,

$$P_{k+1|k} = P_{k|k} + Q_{k+1},$$  \hfill (16)

where, $Q_{k+1}$ is given by (14). The estimation error is, therefore, given by

$$Y_{k+1} - C\hat{S}_{k+1|k},$$  \hfill (17)

and the variance of the estimation error is,
\[ v_{k+1} = C P_{k+1|k} C' + R. \]  

(18)

It is also assumed that the prior value of \( S \) i.e. \( S_0 \sim N(\hat{S}_0, P_0) \). The updating equation for the state vector is

\[ \hat{S}_{k+1|k+1} = \hat{S}_{k+1|k} + K_{k+1} \left( Y_{k+1} - C \hat{S}_{k+1|k} \right), \]  

(19)

where \( K_{k+1} \), the Kalman gain matrix, is given by

\[ K_{k+1} = P_{k+1|k} C' v_{k+1}^{-1}. \]  

(20)

(Note that \( v_{k+1} \) in this case is a scalar).

The recursion for the error covariance completes the specification of the Kalman filter updating equations,

\[ P_{k+1|k+1} = \left[ I - K_{k+1} C \right] P_{k+1|k} \left[ I - K_{k+1} C \right]' + K_{k+1} \sigma_{e}^2 K_{k+1}'. \]  

(21)

Under the assumption of normality as incorporated in the equations (13) and (14), the transition probability density function for the state vector \( S_k \) to \( S_{k+1} \) can be written for a given set of observations \( T \) with the help of the updating equations (19) - (21). Following the argument in Harvey (1989) the prediction error decomposition form of the quasi likelihood function is given by,

\[ \log L(\theta) = -\frac{T}{2} \log 2 \pi - \frac{1}{2} \sum_{k=1}^{T} \log |v_k| - \frac{1}{2} \sum_{k=1}^{T} e_k^2 v_k^{-1}, \]  

(22)
where,

\[ e_k = e_k(\theta) = Y_k - \hat{Y}_{k|k-1} \]  \hspace{1cm} (23)

The quasi likelihood function in (22) when maximised by a suitable numerical optimisation algorithm will yield a Gaussian maximum likelihood estimate \( \hat{\theta} \) of the parameter vector \( \theta \). The standard errors for the estimators are obtained from the inverse of the information matrix at the point of convergence.

For the convenience of the reader not acquainted with the Kalman filter literature, we set out in table 1 the structure and temporal evolution of the Kalman filter recession procedure.
Table 1: Kalman Filter Recursion

<table>
<thead>
<tr>
<th>Step</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>k</td>
<td>Predict state: $\hat{S}_{k</td>
</tr>
<tr>
<td>k+1</td>
<td>Update state: $\hat{S}_{k+1</td>
</tr>
<tr>
<td></td>
<td>Predict measurement: $\hat{Y}_{k+1</td>
</tr>
<tr>
<td></td>
<td>Prediction error: $e_{k+1} = Y_{k+1} - \hat{Y}_{k+1</td>
</tr>
</tbody>
</table>

S: State vector  
P: Covariance matrix  
Y: Measurement vector  
e: Prediction error  
$\nu$: Variance of prediction error  

The (k+1)st term of the prediction error decomposition form of the likelihood function:  
Equation (23)
4. THE BOOTSTRAP EXPERIMENT AND PROCEDURES

In order to obtain bootstrap estimates of the parameters and their standard errors, bootstrap samples are generated by resampling the errors with replacement from the above Gaussian maximum likelihood estimation process. The parameter vector $\theta$ and the covariance parameter $\sigma^2$ were chosen as described in section 5 and used to generate a representative sample. Using this representative sample the extended Kalman filter was employed to generate a parameter estimate $\hat{\theta}$. The procedure to generate bootstrap samples from $\hat{\theta}$ then proceeds as follows:

1. Standardised innovations are obtained as $e_k^*(\hat{\theta}) = e_k(\hat{\theta}) \sqrt{v_k}$, so as to ensure that the innovations have the same first two moments. From these the innovations for the state transition equation are generated as $\zeta_k^*(\theta) = K_k e_k^*(\hat{\theta})$. (See Anderson and Moore (1979), page 231 for discussions on the innovations form of state space representations).

2. These standardised innovations are sampled with a uniform distribution with replacement to obtain one set of bootstrap innovations, denoted as $e_k^*(\hat{\theta}), \zeta_k^*(\hat{\theta}) (k = 1, \ldots, T)$.

3. The set of bootstrap innovations are employed in a complete recursion through the equations (10) and (12) using $e_k^*(\hat{\theta}), \zeta_k^*(\hat{\theta})$ respectively in place of the original
error sequences. In this way one set of bootstrap observations, \( Y^*_k (k=1,\ldots,T) \), is generated. In this process the initial conditions of the filter remain unchanged from the original Gaussian maximum likelihood set up that yielded the innovations in step 1.

4. Steps 2 and 3 are repeated \( N \) (typically 10,000) times to generate a pool of samples for each of which Gaussian maximum likelihood estimates are obtained ie. \( \hat{\theta}_i^* \), \( i \in \{1,\ldots,N\} \).

5. The bootstrap point estimator is computed as, \( \hat{\theta}_b^* = \frac{1}{N} \sum_{i=1}^{N} \hat{\theta}_i^* \), and the bootstrap variance of \( \hat{\theta} \) is computed as, \( \hat{\nu}_b = \frac{1}{N-1} \sum_{i=1}^{N} (\hat{\theta}_i^* - \hat{\theta}_b^*)^2 \).

In order to ensure that all bootstrap samples are representative, it is necessary that the innovations obtained from the Gaussian maximum likelihood estimates are well behaved. While one step ahead forecast errors are computed by the filter the first few iterations usually generate noisy covariance estimates. In other words, the filter requires the first few iterations to stabilise. In the current implementation it is found that the first two estimates of the error covariance should be ignored ie. the first two innovations are kept out of the resampling in step 2, using instead \( \hat{e}_k^*(\hat{\theta}) = \hat{e}_k^*(\hat{\theta}) \) for \( k = 1, 2 \).

An approximation to the true small sample distribution is required to compare the performances of the Gaussian maximum likelihood estimators and the bootstrap estimators. This is achieved by generating another 10,000 samples through recursion of the state space form given by (12) and (10) and using innovations \( \tilde{\xi} \sim N(0,1) \). The initial
values are held invariant the same way as in the bootstrap replications, but the parameter vector is set to known values chosen as explained in the next section. The rationale for the sample size selection is also described there. Gaussian maximum likelihood estimates are obtained for each of these 10,000 samples and this in turn provides the "true" small sample distribution of the parameter estimates. The choice of the number of samples as 10,000 has been found to be sufficient\(^\ddagger\). The sampling procedures are illustrated in tables 2 and 3.

\(^\ddagger\)We have also analysed the sampling distribution obtained from 20,000 samples and the results are not significantly different.
Step 1

- Generate a typical sample from the known model

Step 2

- Estimate the model using the sample
- Store the model standardised residual \( \varepsilon_k, k=1,2\ldots,T \) and the estimated parameter vector \( \hat{\Theta} \)

Step 3

- Using \( \varepsilon_k^*, k=1,2\ldots,T \) and the Kalman gain matrix generate the innovations for the state transition equation \( \zeta_k^*, k=1,2\ldots,T \)

Step 4

- Sample \( \varepsilon_k^*, \zeta_k^* \) with replacement using a uniform distribution to obtain one set of bootstrap innovations ie. \( \varepsilon_k^* (\hat{\Theta}), \zeta_k^* (\hat{\Theta}) \)

Step 5

- Use these bootstrap innovations ie. \( \varepsilon_k^* (\hat{\Theta}), \zeta_k^* (\hat{\Theta}) \) and recursion through the state space system to generate one bootstrap sample of observations.

Step 6

- Use the bootstrap sample to obtain the estimate \( \hat{\Theta}_i^* \)

Step 7

- Repeat steps 4, 5 and 6 to obtain the set of estimates \( \hat{\Theta}_i^*, i=1,2\ldots,N \)

Step 8

- Use the \( \hat{\Theta}_i^* (i=1,2\ldots,N) \) to generate the bootstrap small sample distribution.
Step 1

- Select the parameter vector

Step 2

- Use random normal variate and $\Theta$ to generate a sample of size $T$ through simulation.
- Repeat this $N$ times.
- This will produce $N$ samples each of size $T$.

Step 3

- Estimate the model using each of the $N$ samples
- Store the estimates $\hat{\Theta}_j$, $j=1,2,\ldots,N$

Step 4

- From $\hat{\Theta}_j$, $j=1,2,\ldots,N$, obtain the means of the parameters, which represents $\hat{\Theta}_{true}$.
- Experiment with $N$ and $T$ so that $\hat{\Theta}_{true}$ is as close as possible to $\Theta$.
- Use the final set of $\hat{\Theta}_j$ ($j=1,2,\ldots,N$) to generate the “true” small sample distribution.

Table 3: Generation of "True" Small Sample Distribution
5. PARAMETER SELECTION

5.1 Volatility Function

In order to generate the sample of bond price data for the simulation experiments outlined in the previous section, an appropriate volatility function has to be chosen for the HJM model.

The parameter $\gamma$ of the equation (6) is selected to be 0.5 as this value corresponds to the square-root model for volatility in Cox, Ingersoll and Ross (1985) model. The selection of the other two parameters, $a_0$ and $\lambda$, is guided by the observed shapes of the interest rate volatility reported by Kahn (1991). The humped shape and the convex shape observed in that paper are applied to the bond price volatility in the HJM context. The result of this investigation leads to the form of the volatility function shown in Figure 1. In generating these shapes a sample of overnight rates applicable to the local Australian market is used as a proxy for the instantaneous spot rate needed in (6). The values of $a_0$ and $\lambda$ are obtained by experimentation to arrive at representative values of the bond price volatility (corresponding to three year bonds during 1991 in the Australian market) as well as the shapes given in Figure 1. The focus on the particular year is due to the fact that $f(0,t)$ required in the equation (9b) also represents the same year. The method of obtaining $f(0,t)$ is described later in the section on the empirical results.

The other important parameter choice for the bootstrap tests is the sample size. The sample sizes of 60 and 120 are thought to be representative of typical empirical studies. The smaller number represents the approximate number of trading days during the last three months of the life of an interest rate futures contract trading on the Sydney Futures Exchange, the period when it is most actively traded. The larger number represents the
most actively traded part of the life of a three-year Treasury bond when prices are quoted weekly.

The filter also requires specification of the prior values of the state vector described in the equation (9a). This is set to \([-0.37223, 0.10, 0]\)’. The first element, the logarithm of the bond price, represents the discount bond price with three years to maturity at an yield to maturity of 12% per annum. The second element, the instantaneous spot interest rate, is representative of the overnight deposit rate and the third element given by the equation (9e) has, by definition, prior value zero. The parameter R in the covariance of the measurement error is equated to 0.00001. This value appears to represent the typical dispersion in the high-low or open-close quotes in the interest rate futures contracts trading on the Sydney Futures Exchange. The filtering algorithm also requires a prior covariance matrix of the state vector. In practical implementation of most state space systems this prior covariance matrix is not known precisely. In such situations it is usually specified as \(\kappa I\), where I is the identity matrix and \(\kappa\) is a large integer. For more information on diffuse prior specification see Harvey (1989, page 121). In this study \(\kappa\) is set to 1000.

### 5.2 Initial Yield Curve

The estimation process requires the specification of the initial forward rate curve \(f(0, t)\). The method used in this study is based upon fitting a polynomial of the form \(\beta_1 + \beta_2 x + \beta_3 x^2 + \beta_4 x^3 + \beta_5 x^4\) to the short-term bank bill futures data trading on the Sydney Futures Exchange. Here x represents the delivery time of the futures contracts. Each day nearly 11 or 12 different maturities are traded representing a span of nearly three
years. The parameters are estimated using a non-linear least squares algorithm and the estimated forward rate curve is used to compute the values of \( f(0, t) \) and \( f_2(0, t) \) required by the filter algorithm (see equation (9b)). The procedure is repeated for each observation interval implying that the state estimate \( S_{k+1} \) (given by the equation (12)) is obtained with reference to the forward rate curve at time \( k \). Ideally, the parameter vector \([\beta_1, \beta_2, \beta_3, \beta_4, \beta_5]'\) of the forward rate curve should be estimated at the beginning of each interval. However, it is assumed in this implementation that this parameter vector remains fairly constant over the period of interest. More information on estimation of these forward rate curves can be found in Bhar and Hunt (1993).

6. **EMPIRICAL RESULTS**

The simulation is carried out on a PentiumTM personal computer using the GAUSSSTM (version 3.1) programming language. The maximisation of the likelihood function given by the equation (23) is obtained by the secant method of Broyden, Fletcher, Goldfarb, and Shanno using the line search due to Brent (1972), which is a variation of the golden section method. All gradient calculations are carried out by the central finite difference method. The termination condition of the optimisation routine is when the gradients of the estimated parameters are less than or equal to 0.0001. On average convergence is achieved in around 10 iterations for both the "true" system samples and the bootstrap samples. Standard errors of the estimated parameters are obtained from the square root of the inverted Hessian matrix at the point of convergence.

The estimated parameters for a sample size of 60 and for the four different values of \( \gamma \) are given in the Table 3 together with the standard error or standard deviations of these estimates. Although, the percentiles of the empirical distributions may be used to judge the
performance of the estimation procedure, the small sample performance of this simple
percentile method has been found unsatisfactory. Various improvements have been
suggested eg. bias-corrected percentile method. An excellent reference for this topic is
Maddala, Rao and Vinod (MRV, 1993, chapter 21 and 23). An alternative method for this
purpose is known as the pivotal method (also described in MRV (1993)) eg. bootstrap-t
also known as percentile-t. The main idea here is to use the empirical distribution of the
quantity \( \frac{\hat{\theta} - \theta}{SE(\hat{\theta})} \), where SE stands for standard error, instead of the distribution of
\( \hat{\theta} \). The critical values obtained from the empirical distribution of such a pivotal quantity
can be used for hypothesis tests.

Concentrating on the top panel of table 3, the "true" system data reveal that the mean \( a_0 \) is
0.05 standard deviation away from the value used to generate these samples. Similarly,
the mean \( \lambda \) is 0.45 standard deviation away from the corresponding value used to generate
these "true" samples. This discrepancy is attributable to the relatively small number of
resamplings used in this study. The objective now is to analyse whether there is any
significant difference in the estimates obtained by the Gaussian Maximum Likelihood
(GML) method and the estimates obtained from the empirical distribution of the bootstrap
samples.

The hypothesised "true" value against which both the GML estimates and the estimates
from the bootstrap sample distribution should be checked are given by the 50th percentile
(i.e. the median) of the "true" system distribution depicted in the Figures 2 and 3. The
GML estimate of \( a_0 = 0.05903 \) is closer to the 50th percentile position (i.e. median) of the
"true" distribution of \( a_0 \) as given in the Figure 2 compared to the mean \( a_0 (0.03247) \) of the
corresponding bootstrap distribution. Also, the standard error of the GML estimate is
much smaller than the bootstrap standard deviation. However, the opposite is true for $\lambda$ as can be seen from the Figure 3.

The sample statistics of the pivotal quantities for both the bootstrap replications and the "true" system are also given in Table 3. The empirical distributions of the pivotal quantities are depicted in the Figures 4 and 5 for the "true" system and in the Figures 8 and 9 for the bootstrap samples. Both pivotal quantities from the bootstrap distribution in the top panel in the Table 1 are very close to the 50th percentile position found in the Figures 4 and 5. It is thus clear that in terms of the pivotal quantities the bootstrap resampling technique performs well.

Still continuing with the data related to the top panel of the Table 1 and the Figures 2 and 3, another observation can be made. The "true" system distribution in the Figures 2 and 3 show that not all combinations of the parameters have high probability of occurrence. In particular for $\gamma=0.5$ and the given sample of observations, the most likely values of $a_0$ and $\lambda$ are given by the corresponding mean less 0.3 standard deviations. This is a significant finding and suggests that the HJM model in the state space form, as developed in this paper, is identifiable only within a limited range of values of the parameters of the specified volatility function of the forward rate. Furthermore, the bootstrap distribution of the pivotal quantities given in the Figures 8 and 9 suggest the suitability of the bootstrap scheme in locating the most likely combination of the two parameters $a_0$ and $\lambda$. 
7. CONCLUSIONS

The paper deals with the issue of the estimation of the HJM model in the state space framework under a fairly general specification of the forward rate volatility function. The precision of the Gaussian Maximum Likelihood estimates of the parameters is compared with those from the empirical distribution of the bootstrap samples. An algorithm is also described that generates the bootstrap samples for the state space model.

It is shown that the bootstrap samples offer a convenient way to obtain precise parameter estimates. The statistical significance is established with the help of the empirical distribution of the so-called pivotal quantities, which are known to have better small sample performance. The distribution of the estimated parameters clearly show that only a certain combination of the parameter pair \((a_0, \lambda)\) are highly likely. In some cases, the distribution of \(a_0\) appears to have a bimodal distribution.

The estimation of the HJM model in the state space framework produces estimates of the instantaneous spot rate of interest (a state variable), which are analysed to gain insight into the behaviour of the implied spot rate. Many other models of the term structure of interest rates model the spot rate of interest directly, where as, the HJM model starts with the forward rate of interest. It is, therefore, important to examine the characteristics of the implied spot rate to be able to link with alternative term structure of interest rate models.
REFERENCES


Figure 1
Different Shapes of Bond Price Volatility

Volatility is expressed in per annum basis. The graphs represent different values of the combination $(\alpha, \lambda, \gamma)$.
A : $(0.10, 0.52, 0.50)$
B : $(0.08, -0.52, 1.0)$
C : $(0.50, 0.52, 1.50)$
D : $(1.20, 0.10, 2.0)$
<table>
<thead>
<tr>
<th>Table 3</th>
<th>Comparison of the Estimated Parameters</th>
</tr>
</thead>
</table>

A: \((a_0 = 0.10, \lambda = 0.52, \gamma = 0.5), \text{ Sample Size} = 60\)

<table>
<thead>
<tr>
<th></th>
<th>Gaussian ML</th>
<th>Bootstrap</th>
<th>&quot;True&quot; System</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>Std. Error</td>
<td>Mean</td>
</tr>
<tr>
<td>(a_0)</td>
<td>0.05903</td>
<td>0.00029</td>
<td>0.03247</td>
</tr>
<tr>
<td>(\lambda)</td>
<td>0.50341</td>
<td>0.04845</td>
<td>0.79310</td>
</tr>
<tr>
<td>Pivot (a_0)</td>
<td>-134.68</td>
<td></td>
<td>5.54</td>
</tr>
<tr>
<td>Pivot (\lambda)</td>
<td>-94.87</td>
<td></td>
<td>279.64</td>
</tr>
</tbody>
</table>

B: \((a_0 = 0.08, \lambda = -0.52, \gamma = 1.0), \text{ Sample Size} = 60\)

|        | Estimate    | Std. Error | Mean          | Std. Dev. |
|--------|-------------|------------|---------------|
| \(a_0\)|             |            |               |           |
| \(\lambda\)|           |            |               |           |

C: \((a_0 = 0.50, \lambda = 0.52, \gamma = 1.5), \text{ Sample Size} = 60\)

|        | Estimate    | Std. Error | Mean          | Std. Dev. |
|--------|-------------|------------|---------------|
| \(a_0\)|             |            |               |           |
| \(\lambda\)|           |            |               |           |

D: \((a_0 = 1.20, \lambda = 0.10, \gamma = 2.0), \text{ Sample Size} = 60\)

|        | Estimate    | Std. Error | Mean          | Std. Dev. |
|--------|-------------|------------|---------------|
| \(a_0\)|             |            |               |           |
| \(\lambda\)|           |            |               |           |
Percentiles and Other Statistics Refer to the Figures Above

<table>
<thead>
<tr>
<th></th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percentile</td>
<td>Percentile</td>
<td>Percentile</td>
<td>Percentile</td>
<td>Percentile</td>
</tr>
<tr>
<td>0.025</td>
<td>-0.12298</td>
<td>0.025</td>
<td>-0.76106</td>
<td>0.025</td>
</tr>
<tr>
<td>0.050</td>
<td>-0.08860</td>
<td>0.050</td>
<td>-0.51444</td>
<td>0.050</td>
</tr>
<tr>
<td>0.100</td>
<td>-0.04398</td>
<td>0.100</td>
<td>-0.24563</td>
<td>0.100</td>
</tr>
<tr>
<td>0.500</td>
<td>0.08597</td>
<td>0.500</td>
<td>0.77652</td>
<td>0.500</td>
</tr>
<tr>
<td>0.900</td>
<td>0.29403</td>
<td>0.900</td>
<td>2.60629</td>
<td>0.900</td>
</tr>
<tr>
<td>0.950</td>
<td>0.36963</td>
<td>0.950</td>
<td>3.20728</td>
<td>0.950</td>
</tr>
<tr>
<td>0.975</td>
<td>0.44986</td>
<td>0.975</td>
<td>3.95605</td>
<td>0.975</td>
</tr>
</tbody>
</table>

\( \gamma = 0.5, JB = 166.71 \) \( \gamma = 0.5, JB = 299.09 \) \( \gamma = 0.5, JB = 24369.36 \) \( \gamma = 0.5, JB = 22132.36 \)
Percentiles and Other Statistics Refer to the Figures Above

<table>
<thead>
<tr>
<th></th>
<th>6</th>
<th></th>
<th>7</th>
<th></th>
<th>8</th>
<th></th>
<th>9</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>Percentile</strong></td>
<td></td>
<td><strong>Percentile</strong></td>
<td></td>
<td><strong>Percentile</strong></td>
<td></td>
<td><strong>Percentile</strong></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.025 -0.02548</td>
<td></td>
<td>0.025 -0.56038</td>
<td></td>
<td>0.025 -2563.68</td>
<td></td>
<td>0.025 -712.93</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0.050 -0.01727</td>
<td></td>
<td>0.050 -0.35801</td>
<td></td>
<td>0.050 -760.57</td>
<td></td>
<td>0.050 -229.15</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0.100 -0.00902</td>
<td></td>
<td>0.100 -0.19908</td>
<td></td>
<td>0.100 -316.87</td>
<td></td>
<td>0.100 -108.21</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>0.500 0.02288</td>
<td></td>
<td>0.500 0.50884</td>
<td></td>
<td>0.500 -7.10</td>
<td></td>
<td>0.500 -5.28</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.900 0.08359</td>
<td></td>
<td>0.900 2.01709</td>
<td></td>
<td>0.900 136.73</td>
<td></td>
<td>0.900 38.67</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.950 0.11625</td>
<td></td>
<td>0.950 2.98839</td>
<td></td>
<td>0.950 250.97</td>
<td></td>
<td>0.950 90.79</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.975 0.15240</td>
<td></td>
<td>0.975 3.58434</td>
<td></td>
<td>0.975 394.96</td>
<td></td>
<td>0.975 147.87</td>
<td></td>
</tr>
</tbody>
</table>

$\gamma=0.5$, $JB=17146.14$  $\gamma=0.5$, $JB=11975.05$  $\gamma=0.5$, $JB=22033.11$  $\gamma=0.5$, $JB=50793.31$
r10, r20 etc. denote spot rate distribution after 10, 20 periods respectively of the beginning of the bond price observation series in the "True" system. Sample size is 60 and $\gamma = 0.5$. Spot rate implies instantaneous spot rate in the context of HJM model.

Sample size is 60 and $\gamma = 0.5$ in the "True" system. Spot rate implies instantaneous spot rate in the context of HJM model. The dashed line represents mean.