Department of Mathematics
Faculty of Science and Technology
City University of Hong Kong

"An Empirical Comparison of Alternative Models of the Short-term Interest Rate" (by Chan, Karolyi, Longstaff and Sanders (1992) published in Journal of Finance)

This excellent paper provides a comprehensive empirical comparison on a variety of well-known continuous-time short rate models in the finance literature in their abilities in explaining the actual behavior of short rate. The short rate models include Merton, Vasicek, CIR SR, Dothan, GBM, Brennan-Schwartz, CIR VR and CEV. First, a generalized form of continuous-time model is proposed in which the eight short rate models are nested. The generalized model possesses the mean-reverting property and the heteroskedastic property. The discrete-time econometric specification of the generalized model is an AR(1) model with time-varying volatility which depends on the current level of the short rate. The Generalized Moment of Moments (GMM) by Hansen (1982) technique is then employed for the estimation of the unknown parameters of the discrete time model. The test statistics inherent from the GMM are used for testing hypotheses on the unknown parameters and comparing the performance of various short-rate models in explaining the dynamic of the monthly T-Bill yields. Topics included in this presentation are listed as follows:

Prepared by Siu Tak Kuen, Ken

Section 1: Introduction

1. Background:

- Many short-rate models have been proposed in the finance literature
- A relatively little amount of work on comparing the short-rate models in their abilities to incorporate the actual behavior of short rates

2. Objectives:

- Estimate the unknown parameters of the general form of continuous-time short-rate (called unrestricted model) through its discrete-time specification using the Generalized Method of Moments (GMM)
- Perform students' t test on each of the unknown parameters of the unrestricted model. In particular, the t test for the beta parameter indicates whether the mean-reverting effect exists or not while the t test for the gamma parameter indicates whether the volatility of the short rate model is changing or remains constant (i.e. homoskedastic or heteroskedastic)
- Perform the chi-square test of goodness-of-fit for each of the eight nested models using the unrestricted one as benchmark using the GMM minimized criterion as a test statistic (i.e. the chi-square test statistic). This tells us the explanatory power of each nested models

- Provide pair-wise comparisons of nested models by constructing a R test statistic based on the difference between the GMM minimized criterion of each of the nested models and that of the unrestricted one. The test statistic R follows a chi-square distribution and the test resembles the likelihood ratio test

- Construct a test for whether there is a structural break in the one-month T-bill yield due to the shift in Federal Reserve Monetary Policy in October 1979. The structural break is incorporated by introducing a dummy variable. The test statistic is the GMM minimized criterion for the expanded model with the dummy variable

3. Implications from the empirical analysis:

- Identification of the most important parameter that differentiates short rate models

- The volatility of the short rate process (i.e. homoskedastic or heteroskedastic). The relationship between the short rate volatility and the level of the short rate

- Structural change in the short rate process in October 1979 due to the shift in the Monetary Policy in U.S.

4. Some remarks:

- This paper focuses on the comparison of different <u>*parametric*</u> short rate models. It seems to be analogues the problem of model selection among different parametric statistical models based on some criteria, for instance, Akaike Information Criteria (AIC) or Bayeisan Information Criteria (BIC) in parametric time series models and parametric regression models

- The estimation and related empirical issues of semi-parametric short-rate models and the non-parametric short-rate models has been studied by several researchers, for instance, Jiti Gao, Yacine Ait-Sahalia and Andrew Lo

- The short rate considered in the paper is the short-term nominal interest rate (i.e. without adjustment for inflation rate). The short-term interest rate that has been adjusted for inflation is called short-term real interest rate which is unpredictable due to the unpredictable nature of inflation rate

- For modelling short-term nominal interest rate, it is desirable to have a stochastic model which ensures that the nominal short rate is positive any time almost surely. However, for modelling short-term rate interest rate, a stochastic model which allows the real short rate taking negative values, for instance, the Vasicek model, the Merton model and the Ho-Lee model, is also suitable due to the fact that the inflation rate may be greater than the nominal short rate

- The use of T-bill yields as a proxy for short rate can induce significant measurement error (see Pearson and Sun (1994) published in Journal of Finance). The presence of measurement error can make the time series analysis of short-term nominal interest rate useless since the underlying short rate cannot be measured or observed (see Campbell, Lo and MacKinlay (1997)). In this case, one needs to consider the Kalman filtering state-space representation for the formulation of the short rate model with measurement error (i.e. the observation process is the observable time series of the T-bill yields while the state is the unobservable time series of short nominal rate)

Section 2: The Generalized short rate model and its discrete-time specification

1. The general form of the short rate model (or the unrestricted model):

$$d\, r_t = (\alpha + \beta\, r_t)\, dt + \sigma\, r_t^{\gamma}\, d\, W_t$$

$\{r_t\}$ – the short rate process or instantaneous interest rate process

$\{W_t\}$ – a standard Brownian Motion

β – the mean-reverting parameter (i.e. non-zero if the effect exists)

γ – the elasticity parameter that describes how the current level of the interest rate affects the volatility of the change in the short rate

If $\gamma = 0$, the short rate process is homoskedastic; otherwise, it is heteroskedastic

The general form include most of the short-rate models as nested models (see Table 1 in Page 1212)

The nested short-rate models may be single-factor or multi-factor

It seems that this paper only consider the single factor model

Question: Can the general set up still apply for the multifactor model?

The general form cannot incorporate some short rate models, for instance, the Ho-Lee model, the Hull-White model, and the Pearson-Sun model.

Ho-Lee: $d\, r_t = \theta(t)\, dt + \sigma\, d\, W_t$ (i.e. $\theta(t) = \alpha$ => Merton)
Hull White: $d\, r_t = (\theta(t) - \beta\, r_t)\, dt + \sigma\, d\, W_t$
Pearson-Sun: $d\, r_t = (\alpha + \beta\, r_t)\, dt + \sigma\, (\theta_0 + \theta_1\, r_t)^{1/2}\, d\, W_t$

Pearson-Sun model is an extension of the square-root process (i.e. the CIR SR short rate model). If $\theta_0 = 0$ and $\theta_1 \neq 0$, Pearson-Sun model reduces to the square-root process. If $\theta_1 = 0$ and $\theta_0 \neq 0$, Pearson-Sun model becomes homoskedastic Vasicek short rate model

Some ideas:

To develop a new general form of short rate model that can incorporate or nest a wider class of short rate models in the literature

By using an extended version of the model, it is possible to obtain a comprehensive and representative empirical comparison

Since it is a continuous-time model, it is difficult to estimate its unknown parameters directly via data from discrete sampling

Quick solution: Consider the discrete specification of the general model

2. The discrete-time heteroskedastic AR(1) specification:

$$r_{t+1} - r_t = \alpha + \beta r_t + e_{t+1}$$

$$E(e_{t+1}) = \alpha + \beta r_t \text{ and } SD(e_{t+1}) = \sigma r_t^{\gamma}$$

The volatility of the change in the short rate = $SD(e_{t+1}) = \sigma r_t^{\gamma}$

The heteroskedastic property of the discrete-time specification is in line with that of the continuous-time counterpart
The discrete time specification is just an approximation and the approximation error has been shown by Campbell (1986) to be of second order important if the discrete-time sampling window is small

Very heuristically, by Taylor's expansion,

$$r_{t+dt} - r_t = (dr_t/dt)\, dt + (d^2r_t/dt^2)\, dt^2 + \ldots$$

As $dt \to 0$, the second order and the higher order terms has been neglected

It seems that a more rigorous justification can be found by employing Ito's lemma

Four unknown parameters α, β, γ and σ are estimated by GMM

Section 3: Review on Generalized Method of Moments and its use for comparing short rate models

1. Classical Method of Moments (MM):

- Form a moment equation by matching a moment with its empirical estimate
- Obtain a moment estimate of the unknown parameter by solving the moment equation
- The number of moment equations is equal to the number of unknown parameters
- A moment estimate of an unknown parameter is not unique

2. A simple example of MM:

Suppose $X_1, X_2, \ldots, X_n$ iid ~ $t(\upsilon)$

Estimate υ by the MM

$E(X_i^2) = \upsilon / (\upsilon - 2)$ if $\upsilon > 2$

By matching $E(X_i^2)$ with its sample version $(\sum_{i=1}^{n} X_i^2) / n$, we get the following moment equation:

$$(\sum_{i=1}^{n} X_i^2) / n = \upsilon / (\upsilon - 2)$$

The moment estimate υ_e for υ is given by:

$$\upsilon_e = (2 \sum_{i=1}^{n} X_i^2) / (\sum_{i=1}^{n} X_i^2 - n)$$

3. Consistency: Weak consistency and Strongly consistency

- Weakly consistent estimate: Converge to the ‘true’ value of the unknown parameter in probability as the sample size tends to infinity

- Strongly consistent estimate: Converge to the ‘true’ value of the unknown parameter with probability one as the sample size tends to infinity

- Consistency is the minimum requirement for an estimate

- Other requirements include, for instance, sufficiency and efficiency

4. The MM estimate is a consistent estimate

5. It is a weaker method compared with maximum likelihood estimation (or Quasi maximum likelihood estimation)and least-square estimation

6. The MM estimate may not be an asymptotically efficient estimator. Recall that an unbiased estimator is said to be more efficient than the other if the variance of the former is smaller than the variance of the latter. An unbiased estimator is said to be asymptotically efficient if its variance tends to the inverse of the Fisher information evaluated at the value of the estimator which is the lowest variance bound of an unbiased estimator, namely the Cramer-Rao lower bound (i.e. The variance of an unbiased estimator must be greater than or equal to the Cramer-Rao lower bound). Note that the Fisher information is defined as the negative of the expectation of the second derivative of the log-likelihood function with respect to the unknown parameter.

7. Both MLE and GMM estimates are efficient

8. Since the moment estimate of an unknown parameter is not unique, we can find another moment estimate based on another moment equation (e.g. We can find another moment estimate for the degree of freedom υ based on another moment equation obtained from matching the fourth moment by its sample version)

9. One cannot choose the single parameter υ so that it can satisfy two different moment equations simultaneously

10. One possible solution is to select υ so that it can satisfy both of the moment equations as close as possible

11. Generalized Method of Moments (GMM):

    Suppose $X_1, X_2, \ldots, X_n$ iid ~ $t(\upsilon)$

    Write $X(n)$ for the n-dimensional vector $(X_1, X_2, \ldots, X_n)$

    Estimate υ by the GMM using the second and the fourth moments:

    $E(X_i^2) = \upsilon / (\upsilon - 2)$ if $\upsilon > 2$

    $E(X_i^4) = 3\upsilon^2 / [(\upsilon - 2)(\upsilon - 4)]$ if $\upsilon > 4$

    Let $h(\upsilon, X_i) := (h_1(\upsilon, X_i), h_2(\upsilon, X_i))^T$,

    where $h_1(\upsilon, X_i) := X_i^2 - \upsilon / (\upsilon - 2)$ and

    $h_2(\upsilon, X_i) := X_i^4 - 3\upsilon^2 / [(\upsilon - 2)(\upsilon - 4)]$

Prior sampling, X_i is random, and hence, so is $h(\upsilon, X_i)$

In this case, we can define the ensemble average $E(h(\upsilon, X_i))$ of $h(\upsilon, X_i)$, where the expectation is taken over all possible values for X_i

Let υ_0 denote the "true" value of the unknown parameter υ

By definition of the function h, we require that υ_0 satisfies the following vector equation:

$$E(h(\upsilon_0, X_i)) = (0, 0)^T$$

Let
$$g_{n,1}(\upsilon, X(n)) := \left(\sum_{i=1}^{n} h_1(\upsilon, X_i)\right) / n = \left(\sum_{i=1}^{n} X_i^2\right) / n - \upsilon / (\upsilon - 2)$$

and

$$g_{n,2}(\upsilon, X(n)) := \left(\sum_{i=1}^{n} h_2(\upsilon, X_i)\right) / n = \left(\sum_{i=1}^{n} X_i^4\right) / n - 3\upsilon^2 / [(\upsilon - 2)(\upsilon - 4)]$$

Write $g_n(\upsilon, X(n)) := [g_{n,1}(\upsilon, X(n)), g_{n,2}(\upsilon, X(n))]^T$

$g_n(\upsilon, X(n))$ is the sample average or the time average of $h(\upsilon, X_i)$

Under the assumptions of strict stationarity and ergodicity for the time series X(n), it is reasonable to use $g_n(\upsilon, X(n))$ to approximate $E(h(\upsilon, X_i))$

$$E(h(\upsilon, X_i)) \approx g_n(\upsilon, X(n))$$

Recall that

- A time series $\{X_t\}$ is said to be covariance-stationary or weakly stationary if its unconditional mean and unconditional auto-covariance are invariant over time

- A time series $\{X_t\}$ is said to be strictly stationary if any of its finite-dimensional distributions are invariant over time

- A covariance-stationary time series $\{X_t\}$ is said to be ergodic for the mean if its time averages $(\sum_{t=1}^{n} X_t)/n$ will eventually converge in probability to the ensemble average $E(X_t)$ as n goes large

Since $E(h(\upsilon_0, X_i)) = (0, 0)^T$, we expect that $g_n(\upsilon, X(n))$ should be as close as possible to $(0, 0)^T$ when $\upsilon = \upsilon_0$

In fact, the main idea is to find the estimate υ of the "true" value υ_0 so that $g_n(\upsilon, X(n))$ is as close as $(0, 0)^T$
(i.e. We try to minimize the distance of each component $g_{n,i}(\upsilon, X(n))$ $(i = 1, 2)$ from 0)

Let $W := (w_{ij})$ denote a (2×2) positive definite symmetric weighting matrix which indicates the importance given to satisfying each of the moment equations

For instance, the larger is w_{11}, the greater is the importance of being as close possible to satisfying the moment equation given by the second moment

Define the criterion function $J_n(\upsilon, X(n))$ in the following quadratic form:

$$J_n(\upsilon, X(n)) = g_n(\upsilon, X(n))^T W g_n(\upsilon, X(n))$$

The GMM estimate υ_g is given by:

$$\upsilon_g = \arg\min_{\upsilon > 4} J_n(\upsilon, X(n))$$

Different names for this GMM estimate are "Minimium Chi-Square" Estimator by Cramer (1946), "Minimum Distance" Estimator, "GMM" estimator by Hansen (1982) in Econometrica

Hansen (1982) was the first to provide the most general characterization of the GMM approach and derive the asymptotic properties for serially dependent process

A recent paper by Hansen (2001) investigated the estimation of dynamic economic system using GMM from the time series perspective

The general formulation for the problem of GMM estimation by Hansen (1986) is given as follows:

For illustration, we consider the case that observations are scalars. The case of vector observations can be generalized easily.

Suppose $X(T) := (X_1, X_2, \ldots, X_T)$ denote a T-dimensional random vector representing observations up to and including time T

Let θ denote an N-dimensional vector of unknown parameters

Define a vector-valued function $h: R^N \times R \rightarrow R^M$ (where $M \geq N$)

The function $h(\theta, X_i)$ is used to characterize the moment conditions or the moment equations

Write θ_0 for the "true" value of the unknown parameter θ

By definition of the function h, we require that θ_0 satisfies the following M-dimensional vector equation:

$$E(h(\theta_0, X_i)) = 0_M,$$

where the expectation $E(h(\theta_0, X_i))$ is the ensemble average of $h(\theta_0, X_i)$

We call the M rows of the above vector equation orthogonality conditions

Define the sample average or the time average $g_T(\theta, X(T))$ of $h(\theta, X_i)$ as follows:

$$g_T(\theta, X(T)) = \left(\sum_{t=1}^{T} h(\theta, X_t)\right)/T,$$

We call $g_T(\theta, X(T))$ the generalized sample moment

Use $g_T(\theta, X(T))$ to approximate the generalized population moment $E(h(\theta, X_i))$ under the assumptions of strict stationarity and ergodicity for $\{X_t\}$

When $\theta = \theta_0$, we expect that $g_T(\theta, X(T))$ should be as close as 0_M

The main idea of the GMM is to find an estimate θ_e for the "true" parameter θ_0 so as to make $g_T(\theta, X(T))$ as close as possible to 0_M

Let W_T denote the $(M \times M)$ positive definite symmetric weighting matrix, where W_T may be a function of $X(T)$

Define the criterion function $J_T(\theta, X(T))$ in the following quadratic form:

$$J_T(\theta, X(T)) = g_T(\theta, X(T))^T W_T \, g_T(\theta, X(T))$$

$J_T(\theta, X(T))$ can be interpreted as the weighted sum of the squared distances between each component of $g_T(\theta, X(T))$ and zero

The GMM estimator θ_e for θ_0 is given by:

$$\theta_e = \arg\min_{\theta} J_T(\theta, X(T))$$

The above non-linear minimization problem can be solved numerically

Consider the problem of estimating the unknown degree of freedom υ of the students' t distribution

By letting $W_T = 1$, $\theta = \upsilon$, $h(\theta, X_i) = X_i^2 - \upsilon / (\upsilon - 2)$, the GMM method reduces the MM method

Some of the commonly used estimation methods, for instance ordinary least squares (OLS) estimation, non-linear simultaneous equations estimators, instrumental variable estimation, two-stage least-squares, and in some cases MLE

When the number of unknown parameters to be estimated N is exactly the same as the number of orthogonality conditions M, we set the weighting matrix as the diagonal matrix with each diagonal element being the reciprocal of N.

In this case, the objective function becomes:

$$J_T(\theta, X(T)) = g_T(\theta, X(T))^T g_T(\theta, X(T))$$

Then, the GMM estimator is given by solving the following N-dimensional vector equation:

$$g_T(\theta_e, X(T)) = 0_N$$

If M > N, then objective function becomes the original one with the general weighting matrix W_T

In this case, the closeness between each of the components of $g_T(\theta_e, X(T))$ to zero depends on the amount of weight the corresponding orthogonalility condition (i.e. the moment condition or equation) is given by the weighting matrix W_T

Theorems:

- Suppose $\{X_t\}$ is strictly stationary and ergodic and that the function h is continuous. Then, for any value of θ, the sample average or the time average of $g_T(\theta, X(T))$ of $h(\theta, X_i)$ converge in probability to the population average or the ensemble average $E(h(\theta, X_i))$ as T tends to infinity. This is the weak law of large numbers. Recall that the strong law of large numbers is the one with convergence almost surely (i.e. with probability one)

- Suppose

  (i) $\{X_t\}$ is strictly stationary and ergodic
  (ii) $E(h(\theta, X_i))$ is a continuous function of θ
  (iii) θ_0 is the unique solution to $E(h(\theta, X_i)) = 0_M$
  (iv)The moment conditions are general enough (see the paper by Hansen (1982) in Econometrica)

  Then, the GMM θ_e is a consistent estimate of θ_0

11. The choice of the optimal weighting matrix W_T

- Assume that the process $\{h(\theta_0, X_t)\}$ is strictly stationary with mean 0_M and r-th lag auto-covariance matrix given by:

$$A_r = E(h(\theta_0, X_t)h(\theta_0, X_{t-r})^T)$$

  where

  (i) r ranges from the negative infinity to the positive infinity

  (ii) A_r is an $(M \times M)$ positive definite symmetric matrix. It depends on θ_0 and is not known in advance

- Assume that $\sum_{r=-\infty}^{\infty} \sum_{i=1}^{M} \sum_{i=1}^{M} | h_i(\theta_0, X_t)\, h_j(\theta_0, X_{t-r}) | < \infty$

- Define the following sum S of auto-covariance matrices:

$$S = \sum_{r=-\infty}^{\infty} A_r$$

- Theorem:

Suppose $\{Z_t\}$ is a weakly stationary M-dimensional time series with the unconditional (or marginal) mean vector and variance-covariance matrix given as follows:

$$E(Z_t) = m \text{ and } E((Z_t - m)(Z_{t-r} - m)^T) = A_r$$

Consider a sample $\{Z_1, Z_2, \ldots, Z_T\}$ of size T and define the following sample average:

$$M_Z = (\sum_{t=1}^{T} Z_t) / T$$

Then, the sample average M_Z satisfies:

(i) $M_Z \to m$ in probability as $T \to \infty$

(ii) $T\, E((M_Z - m)(M_Z - m)^T) \to \sum_{r=-\infty}^{\infty} A_r$ as $T \to \infty$

Note that $E((M_Z - m)(M_Z - m)^T)$ is the variance-covariance matrix of the sample average M_Z (i.e. It is an estimator for m). The statement (ii) provides the asymptotic (i.e. $T \to \infty$) variance-covariance matrix for the estimator M_Z

- Now, we consider a sample $\{h(\theta_0, X_1), h(\theta_0, X_1), \ldots, h(\theta_0, X_T)\}$ of size T. Then, the sample average is given by:

$$g_T(\theta_o, X(T)) = (\sum_{t=1}^{T} h(\theta_0, X_t))/T$$

$g_T(\theta_o, X(T))$ is an estimator for $h(\theta_0, X_i)$

By the above theorem, we have

(i) $g_T(\theta_o, X(T)) \rightarrow E(h(\theta_0, X_i)) = 0_M$ in probability as $T \rightarrow \infty$

(ii) $T\, E(g_T(\theta_o, X(T))\, g_T(\theta_o, X(T))^T) \rightarrow \sum_{r=-\infty}^{\infty} A_r = S$ as $T \rightarrow \infty$

Hence, by the statement (i), the variance-covariance matrix for the estimator $g_T(\theta_o, X(T))$ is given $E(g_T(\theta_o, X(T))\, g_T(\theta_o, X(T))^T)$.

From the statement (ii), we see that S is closely related to the asymptotic variance-covariance matrix of $g_T(\theta_o, X(T))$

- Intuitively, the asymptotic variance of the GMM estimator θ_e that is chosen to minimize depends on the choice of W_T since the objective function $J_T(\theta, X(T))$ depends on W_T

  The asymptotic efficiency of the GMM estimator θ_e can be maximized by making the asymptotic variance-covariance as small as possible via the choice of the weighting matrix W_T

  It has been shown in Hansen (1982) that the optimal choice of W_T for minimizing the asymptotic variance-covariance matrix for θ_e is given by the inverse S^{-1} of the matrix S

  In other words, the asymptotic variance of the GMM estimator θ_e minimized when θ_e is chosen to minimize the following objective function

$$J_T(\theta, X(T)) = g_T(\theta, X(T))^T S^{-1} g_T(\theta, X(T))$$

The intuition behind the optimal choice of weighting matrix can be obtained by considering the use of the generalized least-squares (GLS) method (i.e. the weighted least-squares method) for estimating linear models

Section 4: Empirical Results and Their Implications

~ End of the presentation~