Estimating the benchmark Yield Curve
- A new approach using Stochastic Frontier Functions

Gangadhar Darbha, *
Consultant, National Stock Exchange, Mumbai.
e-mail: gdarbha@nse.co.in

September 2002

Abstract

Estimating a risk free Term Structure of Interest Rates or Zero Coupon Yield Curve from the observed bond prices would involve controlling for the effects of security specific non-interest rate factors that affect bond prices. In this paper we propose a new framework to estimate bench-mark default and liquidity risk free yield curve using the stochastic frontier functions. The methodology explicitly controls for the effects of security specific factors such as age, issue size, coupon and residual maturity on bond prices in estimating the yield curve. Using the daily secondary market data from NSE-WDM for the period Jan 1997 to July 2002, we find that the new methodlogy not only identifies frontier yield curve that is significantly different from the standard zero coupon yield curve and gives reasonable estimates of liquidity premia, but also performs better than the standard yield curve models in terms of predicting bond prices.

*Useful contributions from my colleagues Vardhana Pawaskar and Sudipta Dutta Roy and comments from Subrata Sarkar, Ajay Shah, Chitra Ramakrishna, Ravi Narain, Ganesh Kumar and Ashok Nag are greatfully acknowledged.

Estimating the benchmark Yield Curve
- A new approach using Stochastic Frontier Functions

Gangadhar Darbha

1 Introduction

The zero coupon yield curve (ZCYC) or the term structure of interest rates - that characterize the relationship between interest rates in the economy and the term to maturity - forms the basis for the valuation of all fixed income instruments. Modeled as a series of cashflows due at different points of time in the future, the underlying price of such instruments can be calculated as the sum of the present values of the cashflows, each discounted by the rate for the associated term to maturity. When the rates used in discounting are the risk-free rates of interest, the resulting value of the fixed income instrument would be its fundamental value devoid of either default or liquidity risks.

Risk free rate is the rate of return that is free of default and liquidity risk. It must reflect three components: a rental rate indicating the real return for lending out funds over a investment period thereby forgoing the consumption for which the funds otherwise could be used, inflation, and term risk indicating the risk that the principal's market value will increase or decrease during the term to maturity. The risk-free interest rate schedule or the bench-mark yield curve, therefore, represents the expectations of agents, on an average, about the movements of the above mentioned factors in the economy. Knowledge of the default risk free term structure of interest rates enables one to compute the fundamental value of default-risk free soveriegn bonds and their derivatives like STRIPS (Vasicek (1979) and Hull and White (1990)). In addition, it would also provide an essential input into the pricing of defaultable bonds and credit derivatives (Jarrow and Turnbull (1995) and Duffie and Singleton (1999)). Risk free yield curves are also required for risk management purposes , for example, in applying the historical simulation method to calculate the Value-at-Risk for deafult-risk free and defaultable bond portfolios (Saunders (1999) and Darbha (2001)).

An important obstacle in the above mentioned applications is that the term structure of risk-free interest rates is not directly observable in the market and has to be estimated from the market prices of traded instruments using statistical techniques. An obvious way to control for the default risk in estimating risk-free interest rates is to consider, at the first stage, only the default risk-free soveriegn bonds. The resulting interest rate schedule, while free from default risk, will still be affected by the liquidity risk if soveriegn bond prices contain liquidity premium. In other words, if the observed variation in soveriegn bond prices is due to variation in (expectations about) interest rates alone, one can estimate the yield curve from the bond prices using the present value relations in a conceptually straight forward manner. If, on the other hand, the soveriegn bond prices contain a significant liquidity premium, related to certain security specific factors such as the age of the bond, coupon, issue size etc., then estimating the yield curve from bond prices without controlling for these effects would lead to biases in the term structure. If these factors adversely affect the liquidity of bond and thereby lead to lower prices, any model estimating the yield curve ignoring this fact will over-estimate implied interest rates.

The empirical significance of these factors in affecting the liquidity premium and prices in bond markets is well recoginsed in the literatrure. Amihud and Mendelson (1991), analyzing the Treasury bill and note yield differential, find that liquidity is an important factor in affecting bond prices and it is inversely related to the term to maturity of bonds. Warga (1992) find that recently issued "on-the-run" treasury securities have lower returns than "off-the-run" issues, a result suggesting that greater liquidity of the former is positively priced by the market. Similarly, Sarig and Warga (1989) find that bond's liquidity tends to decrease with its age or time since issue. In a study of yield differences between Treasury notes and bills, Kamara (1994) finds that the liquidity premium increases with both interest rate volatility and turnover ratio in the notes market relative to that in the bill market. Crabbe and Turner (1995) analyze the relation between liquidity premia and the issue size in US corporate bond market and do not find a stable correlations between them. Elton and Green (1998), analysing the US bond price data, find that factors affecting liquidity, such as age of the bond, coupon, issue size, and volume are significantly correlated with bond prices over and above that explained by the interest rate factors. Similar evidence is found by Eom, Subramanyam and Uno (1998) in the context of Japanese bond market and Darbha, Dutta Roy and Pawaskar (2001) in the Indian context. This evidence highlighs the need for: controlling the effects of these variables on the estimated yield curve, and estimating the effects of these variables on bond prices simultaneously with the yield curve.

Not withstanding the empirical evidence on the importance of non-interest rate factors in affecting the liquidity premium and bond prices, the work on estimation of the yield curve, both in developed and emerging bond markets, has focussed primarily on identifying the functional form of the relationship between interest rates and the term to maturity using alternative parametric models that fit the present value model prices 'best' to the observed market prices. These methods include: polynomial splines (McCoullogh (1971, 75) and Mastronikola (1991)), exponential splines (Vasicek and Fong (1982) and Shea (1985)), B-splines (Shea (1985), Steely (1991) and Eom, Subramanyam and Uno (1998)), exponential polynomial forms for the forward rates (Nelson and Siegel (1987), Svensson (1994) and Bliss (1997)), and smoothing splines (Fisher, Nychka, and Zervos (1995), Bliss (1997) and Waggoner (1997)). In the Indian context, Thomas and Saple (2000), Subramanian (2001) and Darbha, Dutta Roy and Pawaskar (2001) estimate and compare the performance of alternative models of term structure of interest rates.

Noticeably none of these studies either address the issue of consequences of mis-specification to the estimation of term-structure or do not provide a framework to estimate the term structure that takes into account the effect of non-interest rate factors on bond prices. Of course, if the observed price variation across securities over and above that implied by the yield curve is random and not systematically related to the security specific factors, then the biases caused by the omission of these effects would be minimal. But some of these studies find that such a systematic relationship between security specific factors and bond prices is empirically significant. See McCoullogh (1971, 75), Carleton and Cooper (1976) and Shafer (1981) for tax timing and clientele effects in the US context, Elton and Greene (1998) for the liquidity and volume effects in the US context, and Eom, Subramanyam and Uno (1998) for the effects of coupon, age of the bond and term to maturity in the context of Japanese bond market. In the Indian context, Darbha, Dutta Roy and Pawaskar (2001) find that age, liquidity ratio and coupon have a significant effect on bond prices after controlling for interest rate effects. These studies typically estimate the yield curve at the first stage, and study the correlations between pricing errors (the difference between the prices predicted by the estimated yield curve and the observed prices) and the security specific factors at the second stage. As a result, while their findings are useful in highlighting the importance of these non-interest rate factors, they do not control for these effects in estimating the yield curve in the first stage itself. This makes the estimated yield curve sensitive to the biases caused by the omission of relevant variables from the bond pricing equation.

The empirical literature aimed at controlling for the effects of 'non-interest rate factors' has taken either of the following three approaches:

- select a sub-sample of 'bonds'from the full data set using some criteria based on 'liquidity' and estimate the yield curve based on the selected sub-sample of observations (Bolder and Sterilski (1997)). The idea being that the selected sample of bonds being the most liquid would be devoid of illiquidity premia, and hence their prices reflect primarily the interest rate factors.

- identify a single or a set of proxies for 'liquidity', generate a weight variable as a function of these proxies and estimate the yield curve by minimizing the weighted distance between observed and model prices (Subramanian (2001)). Liquidity based weighting by down weighting the illiquid bonds is expected to minimze the impact of the latter in estimating the yield curve.

- estimate the yield curve jointly with a liquidty function relating the bond prices to the other non-nterest rate factors (Elton and Greene (1998), Alonso, Blanco, del Rio and Sanchis (2000) and Darbha, Dutta Roy and Pawaskar (2001)).

These approahces while suggest many heuristic ways to identify a bench mark yield curve are not without problems. While the first approach may enable to estimate the 'liquid yield curve', its success hinges in two assumptions: that we can select the sub-smaple in a correct way in the sense that we do not throw out some relevant information wrongly; and that we have sufficient number of observations in the sub-sample spread across the maturity spectrum to enable us to fit a curve an in an efficient way. The latter will be more serious in the context of emerging debt markets where a substantial portion of secondary market trading is concentrated in handful of bonds that the market percieves as liquid.

The second approach does not leave out any observations from the data set, but it assumes that we can identify a liquidity function - both the proxy variables and the parametric form in which it enters into the pricing equation. Related to the latter point is that this approach does not take into account the one-sided nature of the liquidity / illiquidity premium on bond prices. That is, an illiquid bond would, generally, command a price (yield) lower (higher) than a liquid bond, leading to a negative illiquidity premium in prices. It is not clear how a simple liquidity based weighting scheme takes this aspect into account. Naturally, credibility of the yield curve, estimated from this approach, to act as a benchmark would depend upon the extent to which such wieghting schemes fully control for the other factors on bond prices. It is found that the pricing errors of the yield curve model, estimated using weights, are systematically related to security specific factors thereby suggesting that the weighting scheme alone is not sufficient to account for the effects of non-interest rate factors.

Finally, the regression based approach that suggests a joint estimation of a yield curve along with a liquidity function is consistent but does not explicitly recognize the one-sided nature of the liquidity effects on bond prices. Darbha, Dutta Roy and Pawaskar (2001) recognize this aspect when they relate the security specific factors to positive pricing errors, but the underlying framework of their analysis is not very clear in the sense that we do not know whether the estimated yield curve represents a 'most' liquid curve or an 'average' liquid curve.

The primary objective of this paper is to suggest a new framework for estimating a bench-mark yield curve that explicitly controls for and models the effect of non-interest rate factors on bond prices. The framework is based on the econometric models of stochastic frontier functions that, as we explain in detail in the next section, addresses the criticisms of the previous approaches. The second goal of this paper is to provide a methodology to estimate the liquidity premium for each bond and analyse its relation with various security specific factors. It is, in our view, the first comprehensive attempt to jointly estimate the benchmark yield curve and security-specific liquidity premium within a consistent statistical framework. While the empirical analysis is based on Government of India bond market, the proposed framework would be of relevance to all emerging bond markets where the problem of fragmentation and illiquidity is accute.

The rest of the paper is organized as follows. Section II provides a brief description of the Indian Government bond market, with a focus on the segmentation and illquidity in the bond markets. Section III outlines the econometric methodology and the empirical specification of the estimated model. An account of the data and related estimation issues is presented in Section IV. Results and analysis are presented in Section V. Section VI concludes.

2 Indian Government Bond Market

The Indian Government bond market comprises securities issued by the Government of India and the State Governments. Government of India securities include Treasury Bills (T-Bills) with maturity less than a year and dated Government securities (G'secs) with maturities exceeding a year. As on March 31, 2001, there were 116 G'secs outstanding with maturity dates ranging from 1 to 20 years. The total outstanding amount was Rs.38,78,540 million. There were 54 T-Bills outstanding for an aggregate amount of Rs.1,69,800 million. State Governments had an outstanding of 295 securities comprising Rs.4,31,760 million.

The maturity distribution of outstanding G'secs as on March 31, 2001 reveals that over 50 percent of the outstanding issues have a residual maturity less than or equal to 5 years. About 30 percent of the securities lie in the maturity range of 5 to 10 years and the balance 20 percent have maturity beyond 10 years. There are, in fact, only 3 securities with maturity dates beyond 2016. During the financial year 2001-02, the Government of India issued, for the first time, securities beyond 20-year maturity. The secondary market in Government securities is largely a telephone-based market, with trades subsequently reported on the Wholesale Debt Market segment of the National Stock Exchange (NSE-WDM) and the Subsidiary General Ledger of the RBI (RBI-SGL). Secondary market activity in Government securities witnessed an average growth of 91 per cent per annum during the period 1994-95 to 1999-2000. However, the size of the market is small compared to the amounts outstanding; the total turnover in 2000-01 was Rs.45,65,150 million implying a turnover ratio of about 1.5. Trading in Government securities on the NSE-WDM was thin prior to 1998, but has grown significantly over the last 3 years. Trading is usually spread over the entire maturity spectrum, which, for the purpose of estimating a term structure, has the advantage that there are no gaps in the data at any maturity bracket.

Like in most other markets, secondary market activity is concentrated in a few securities, also referred to as benchmark securities. The identity of these benchmark securities changes over time, and it is difficult to identify the 'optimal' mix of characteristics that distinguish benchmark (liquid) securities from other illiquid Government bonds. Investor interest in these papers is partly driven by the Reserve Bank of India (RBI) policy of reissuing certain securities at various maturities, which on the one hand increases the notional amount outstanding in these securities and on the other signals RBI preference for emergence of the benchmark. Barring few exceptions, other features in addition to high outstanding amounts - common to actively traded bonds are a residual time to maturity that lies between 4-8 years and time since issuance not exceeding 3 years (Table 1).

A comparison of the attributes and secondary market activity in 2 securities with residual maturity of 10 years is illustrative of the role of non-present-value factors in influencing investor preferences and, in turn, volumes and prices (yields) in the secondary market (Table 2). The 11.50and has a maturity date of August 5, 2011. The 11.5023, 2000 and will be redeemed on November 23, 2011. The cashflow amounts of the 2 bonds are the same and the cashflow structures (times to coupon and redemption) almost similar. Secondary market activity in these papers on NSE-WDM reveals a concentration of activity in the more recent issue, resulting in turn in significant (price and corresponding) yield differentials between the two securities. Similar anomalies in pricing of liquid (benchmark) vis--vis illiquid securities exist in almost every maturity segment. This increases the need for explicitly modelling the liquidity factor in estimating the bencgmark zero coupon yield curve. Econometric

3 Econometric Methodology

3.1 Standard Framework

The price function for the (default-risk free) Government bonds, without any optionalities, is given by:

$$P_i^d = (P_i^c + AI_i) = \Sigma_{j=1}^{T} \frac{CF_j}{(1+r_j)^j} \tag{1}$$

where P^d is the dirty price, P^c is the clean price, AI is the accrued interest rate, CF_j is the cash-flow due at time j, and r_j is the spot rate associated with the maturity j.

To reduce the dimensionality of estimating the 'T' number of spot rates associated with 'T' cashflows, the literature specifies a unique functional relation that relates spot rates with the term to maturity, i.e.

$$r_m = g(m; \beta)$$

A great deal of empirical literature focussed on identifying the function

$g(..)$ and estimating β, with the constraint that it should be sufficiently flexible to capture most patterns in the observed bond price data and at the same time should not excessively parametrized. Once such a functional relation is identified, one could routinely derive the spot rates for any given maturity.

In the above formulation, the (dirty) price of a bond is expressed as a determistic function of interest rates alone. In order to take into the fact that observed prices may deviate from the present value model prices in a random manner, most of the empirical studies add an error term to equation (1) implying,

$$P_d = P^* + \varepsilon \tag{2}$$

where

$$P^* = \Sigma_{j=1}^{T} \frac{CF_j}{(1+r_j)^j}$$

.

Assuming that the error term ε are independent and identically distributed (typically normal with zero mean) one could estimate the parameter vector β by minimizing the error sum of squares.

As explained earlier, if the security specific non-interest rate factors are systematically related to liquidity premium, the errors ε will not in general have a zero mean and also be non i.i.d. In addition, since liquidity premium will have a one-sided effect on bond prices, i.e. illiquid bonds will typically have a lower price than a liquid one, the error term capturing the effect of all omitted variables will not even be symmetric. Imposing a zero mean normal distribution on the data would, therefore, lead to biases in the estimated term structure.

3.2 Proposed framework

To control for and model the effects of non-interest rates factors on bond prices, we, following the extensive literature on frontier production functions, modify the equation (2) as:

$$P_d = P^* - U + \upsilon \tag{3}$$

where $P^* = \Sigma_{j=1}^{T} \frac{CF_j}{(1+r_j)^j}$, $\upsilon \sim N(0, \sigma_\upsilon^2)$, and $U \geq 0.0$ and $U \sim N^+(\mu_U, \sigma_U^2)$.

Notice that in the place of a single random error in equation (2), we have two random error terms in equation (3) which are assumed to be independent of each other and the regressors. The error term U, being a non-neagtive (turncated) normal random variable, is expected to capture the effect of liquidity premium on bond prices. The second error term υ, assumed to be normal random variate, is expected to capture noise due to the effect of all other factors that are omitted from the bond pricing equation.

According to equation (3),

$$E[P|.] = P^* - E[U|.] + E[\upsilon] = P^* - E[U|.],$$

since $E[\upsilon|.] = 0$. This would imply that a predicted price of a bond is given by the present value model price determined by the interest rate factors (P^*) minus the premium ($E[U|.]$) that it commands depending upon its degree of illiquidity. Note that U is a non-neagtive variable but enters the pricing equation with a minus sign. The higher is degree of illiquidity the greater will be the premium and lower will be the price. For the most liquid bonds the premium, $E[U|.]$, is expected to be 'zero' suggesting the interpretation that P^* would correspond to the price of bonds on the most 'liquid frontier'. The implied term structure of interest rates, g(m,β), correspond to the benchmark yield curve that is devoid of the effects of liquidity premium.

To elaborate, if U is specified corrcetly to capture the liquidity premium in bond prices, the expected price of a bond should always be less than or equal to P^*. Of course, the observed price of a bond can be greater than the liquid frontier price due to the other unaccounted for factors that affect the price of bonds captured by υ.

The formulation in (3) readily addresses the criticisms that we have raised with respect to other approaches. It doesnt require any pre-sampling of data; it doesnt require to specify, without estimating, a liquidity based weighting scheme; and it explicitly recognizes the one-sided nature of the effect of liquidity premium by characterizing with a truncated random variable.

In order to estimate the yield curve fom the bond pricing equation (3), we need to specify: the exact functional form for g(.;β), and the form of the density function for the truncated random variable, U. Given a density function for U, and υ, we can write the likelihood function based on the joint density function of (U,υ). The expressions for log-likelihood function are derived for cases in which the one-sided random variable has either a truncated normal, exponential or gamma densities (Greene (1993)). The parameter vector β characterizing the yield curve, along with the mean / variance parameters of U and υ, are estimated by maximizing the log-likelihood function using the standard numerical optimization techniques. Given the estimates of these parameters, we could derive the estimates of liquidity premia by computing $E[U|.]$ using the formulae given in Jandrow et all (1982).

3.2.1 Model Specification and Estimation

In terms of model specification, we need to specify a functional form for the term structure or ZCYC, g(m;β), as well as the density for the error terms U and υ. For the yield curve, we use the well-known Nelson-Siegel's exponential polynomials linking interest rates and the residual maturity, and truncated normal normal and normal distributions for the error terms U and υ respectively. This would imply:

$$r(m,\beta) = \beta_0 + \beta_1 * \left(\frac{1-\exp\left(-\frac{m}{\tau}\right)}{\frac{m}{\tau}} \right) + \beta_2 * \left(\frac{1-\exp\left(-\frac{m}{\tau}\right)}{\frac{m}{\tau}} - \exp\left(-\frac{m}{\tau}\right) \right)$$

$$f(U) = \frac{2}{\sqrt{2\pi}\sigma_U \Phi(-\frac{\mu_U}{\sigma_U})} \exp\left[-\frac{(U-\mu_U)^2}{\sigma_U^2} \right]$$

$$f(\upsilon) = \frac{1}{\sqrt{2\pi}\sigma_\upsilon} \exp(-\frac{\upsilon^2}{\sigma_\upsilon^2})$$

where r(m;β) is the interest rate function or the ZCYC, f(.) is the density function of error terms, μ is the mode of the normal distribution truncated from below at zero, and Φ is the standard normal cumulative distribution function. As derived in the Appendix I, the likelihood function associated with this specification for a sample of N observations is given by:

$$lnL = -Nln\sigma - Nln\Phi\left(-\frac{\mu}{\sigma_U}\right) + \sum_i \left\{ ln\Phi\left(\frac{\mu}{\sigma_U\lambda} - \frac{\varepsilon_i\lambda}{\sigma}\right) - 0.5\left(\frac{\varepsilon_i+\mu}{\sigma}\right)^2 \right\}$$

The log-likelihood function can be maximized with respect to the parameter vector to obtain the ML estimates of the parameter vector ($\beta_0;\beta_1;\beta_2;$ $\tau;\mu;\sigma_U;\sigma_\upsilon$). Given the degree of non-linearity of both the bond pricing model and truncated distribution model, the maximization of the likelihood function will have to be done using numerical nonlinear optimization methods.

In order to allow for the possibility that the liquidity premia is typically non-i.i.d, and more importantly, would be systematically related to security specific factors, we, following Kumbhakar, Ghosh and McGuikin (1991) in the context of frontier functions, propose to generalize the model (3) by allowing the mode of the one-sided error term U to be depedent on security specific factors such as issue size, age and coupon. i.e.

$$\mu_i = \gamma_0 + \gamma_1 \ issz + \gamma_2 \ agex + \gamma_3 \ cpn$$

where issz, agex and cpn are some affine transforms of variables issue size, age and coupon of each security respectively. The intution for introducing these variables is as follows: the larger the issue size, the greater will be the stock in circulation, the lower will be the search costs involved in matching an order imbalance, resulting in lower transaction costs and greater liquidity, thereby implying a negative relation between issue size and illquidity. Similarly, because fixed income instruments have the characterstic of maturity date being known at finite date besides certain pay-offs and the value of the bond approaches a cretain redemption value as the date of maturity approaches, the incentive to trade that security in secondary market will decrease and hold it till maturity will increase as the date of maturity is approached. This, in turn, would imply that a greater proportion of bonds issued long ago would go into a hold till maturity stock than recently issued ones, thereby causing their illquidity to go up. Therefore, one would expect a positive relation between age and illiquidity premia. Finally, due to differentail taxation of capital gains and interest income, and accounting norms, low coupon bonds are favoured than high coupon bonds at the margin, leading to a negative relation between coupon and illiquidity. While we keep these explanations in mind in specifying an illiquidity function, we do not impose any constraints on the coefficients to match our intution. In fact, as we discuss later, we consider this as a test for validdating the specification for the illiquidity function. If the estimated γ coefficients have the expected signs, then its an indication that the illiquidity function is correctly specified, and hence the resulting yield curve is a risk free yield curve. Similarly, we would also extend the model (3) by allowing for heteroskedasticity in error terms U and υ, by making the error variances a function of residual maturity and duration respectively. This is to control for the effect of heterogeneity in cross-section regression errors on the parameter estimates and illiquidity premia.

As for the derivation of likelihood function, one needs to just replace the μ value with the μ_i function given above. As shown by Kumbhakar, Ghosh and McGuikin (1991), this will not pose any additional conceptual difficulties, though the computational complexity will increase due to increase in the number of parameters. Since the objective function is a highly non-linear function in parameters, standard maximization methods such DFP or BFGS may have difficulty in finding the global optimum (Goffe (1998)). Following the suggestion of Goffe(1998), we use the Simulated Annealing algorithm, which is found to be robust in finding the parameter vector associated with the global optimum. All computations are undertaken in GAUSS package using the Constrained Optimization (CO) module[1].

3.2.2 Model Validation

In order to check the empirical validity of our model specification, we propose to examine the results by asking the following questions:

- does the model pass the statistical diagnostics? As can be seen from the model specification, under the null that σ_U and all γ coefficents are equal to zero, the proposed model will be equal to the standard yield curve estimation problem. This would give us the possibility to conduct an LR test comparing the restrcted model with an unrestricted frontier function model. Rejection of this null would mean an indicate

[1]We have combined and modified GAUSS codes of Professors Lars svenson and Bill Goffe for this purpose.

the validity of the frontier function framework to model the illiquidty phenomenon. This in our view would constitute a first step in checking the model adequacy.

- do the estimates of the liquidity premium meaningful, and whether the relation between liquidity premium and security specific factors such as age, issue size etc.. in tune with the literature?

- how does the yield curve associated with the frontier price function compared with that of the average price function? If our argument about the effect of liquidity premium on the average yield curve is indeed correct, one should find that the frontier yield curve genrally lies below the average yield curve.

- how does the new framework perform vis-a-vis the earlier approaches in terms of fit to the observed bond prices? If our model specification is correct, it should predict the bond prices more closer to the observed prices than the average yield curve, since the ability to estimate liquidity premia must ultimately improve the ability to predict the bond prices. The test would relate to the error statistics such as MAE associated with the fronteir yield cuvre vis-a-vis the average yield curve.

In conducting the model validation tests, we are confining to the in-sample tests as we are not aware of procedures to do that on an out-of-smaple basis within a well specified statistical model. In any case, since the objective of the study is more to do with providing a framework to estimate benchmark yield curves and not to forecast bond prices, the above validation methodology should suffice. We leave the out-of-sample comparisons for future work.

4 Data details

The data for the study are compiled from secondary market trades in Government securities reported on every day on the Wholesale Debt Market (WDM) segment of the National Stock Exchange (NSE). The estimation framework takes into account various institutional details related to secondary market trading and is adapted from our earlier study [Darbha, Dutta Roy and Pawaskar (2001)]. The proposed period of analysis is from 1^{st} January 1997 to 31^{st} July 2002.

The NSE-WDM data [2] constitutes, on an average, about 60-70 per cent of the total trades negotiated and comprise those trades that are negotiated through member-brokers. The price information relates to 'traded prices' rather than 'quotes', and is not time-stamped. On every trade date and for each individual trade, we have information on the security traded, traded price, traded volume and settlement date. Security details viz. date of issue, date of maturity and details of cashflows for the bond, are available from a masterfile of securities available with NSE.

Bulk of the trading is in securities issued by the Central Government, ie. GoI securities; state government securities (SGS) account for a very small number of the trades conducted on any given day. It is useful to mention at this point that, state Governments being perceived as less credit-worthy than the Centre, SGS are issued and traded at a credit spread over GoI securities of same maturity. There are, in addition, differences in perceived credit-worthiness across states that is reflected in inter-state coupon (yield) differentials. To purge the estimated sovereign term structure of any of these effects, the dataset we use comprises only GoI securities. A widely held perception in the Indian markets is that instruments with maturity less than a year, being traded as money market instruments, reflect pricing

[2] Information on trades reported on RBI-SGL is publicly disseminated only on the day of settlement and could have trades conducted on different trade dates, which renders it difficult to use it for the exercise at hand.

considerations different from that of longer-maturity securities. Further, pricing differences are observed between T-Bills and G'secs of the same residual maturity. Subramanian [2001] cites these as reasons for excluding such observations from the sample. Inasmuch as the objective of the current exercise is to analyse the nature and extent of such influences, in addition to providing daily estimates of the term structure, we do not apply any such prior filter on the dataset.

Volume weighted average prices are used, where the average is computed over trades with the same settlement date. This means that for each security, we have as many observations as the range of settlement horizons for trades negotiated on a given trade date.

Present value computations require information on time to coupon payments and redemption. These are calculated with reference to the settlement date. Market conventions require computation of accrued interest on a 30/360 basis for instruments with residual maturity exceeding a year and on actual/365 basis otherwise (this includes T-Bills), and these are adhered to in the computation of coupon accrual and time to cashflows.

There are various factors to which intra-security variation in prices can be attributed. First, the scope for price discovery in negotiated deals is limited, and even the dissemination of the transacted price is available to the market after a considerable lag, an outcome of the current state of the market where reporting rules are not very stringent. This may be an important factor contributing to the observed dispersion in prices across different trades in the same security. Further, within the T+5 settlement system, trades negotiated on a given day can have settlement dates varying from current date to 5 days hence. There are two mechanisms through which this exerts an impact on the price. First, expectations about the likely directionality of interest rates would be built into the contract if the term structure is expected to undergo a significant change by the time the deal is settled. To discount the cashflows for deals that do not settle on the current day, therefore, the appropriate rates to be used are those that are expected to prevail on the settlement date. We use implied forward rates - the best predictors of expected future spot rates - to discount these cashflows.

Secondly, the negotiated price for a transaction that does not settle on the same day would need to incorporate the net cost of carry. From the point of view of the seller, the opportunity cost involved in settling a deal T days into the future is approximated by the foregone return in the call money market (say), while the return is given by the coupon that accrues for these days. If the net cost of carry is positive (negative), the negotiated futures price will be higher (lower) than the spot price. To compute the net cost of carry for the purpose of the empirical exercise, we proxy the overnight rate by the short-term rate ($\beta_0 + \beta_1$) derived from the estimated term structure. The cost of carry is added to the estimated spot price to arrive at the estimated futures price.

5 Results

The proposed frotier function model (3) is estimated on the daily cross-section of observations on bond prices. Since it is a cross-section regression, we have all in all about 1600 regressions. We present the summary statistics of these regressions to give the synoptic view of the results. Table 3 presents the number of times the LR statistic for the null hypothesis that σ_U and all γ coefficents are individually and collectively equal to zero. The results indicate that except for the constant term for the year 1997, all other coeeficents are statistically different from zero. The LR test in the last column also indicates that the standard specification (valid under the null) can not be accepted for more than 90 percent of the days for all the years, thereby supporting the evidence if favor of the frontier function specification. Thus, the bond price data does seem to support the model specification presented above [3].

In Tables 4 we report the averages of the estimates of the coefficients relating the security specific factors and illiquidity premia taken over the number of days in each year. The coefficients on issue size and age have expected signs across all years, i.e 'issue size' ('age') is negatively (positively) correlated with illiquidity premia. The coefficient on 'coupon' has also been consistently positive indicating a preference for low coupon bonds by the market, except for the year 1998[4]. All these coefficients are statistically significant, supporting the specification for the liquidiy function. In fact, we would like to interpret this result to indicate that the one-sided error term that we specify indeed captures the illiquidity premia in bond prices. The results do not indicate, however, that the postulated relation is sufficient characterization of illquidity factor as we may have ommitted the effect of some other variables.

In table 5, we report the averages of the estimated illquidity premia according to volume and maturity classes for the years 2001 and 2002. The results clearly indicate that there is no monotonic relation between volume / maturity and illquidity premia though the illiquidity may be negatively (positively) correlated with volume (residual maturity). This would imply that usage of volume as a proxy for liquidity is at best imperfect and studies such as Subramanian (2000) that regularly use volume weights to control for illquidity effects in bond pricing may introduce some biases into the estimation. The size of the estimated illquidity premia are also in tune with the subjective estimates obtained from some of the market participants, reinforcing the validity of the specified model that we established through stastical tests earlier.

[3]It should be kept in mind that the standard significance tests are not strictly valid for the coeffients in variance function as the later have inequality restrictions of zero on them

[4]The averages are taken over all days and not just those that are statistically significant.

Having found evidence supporting the model specification, we now present in Table 6 the estimates of the term structure of interest rates associated with frontier price function vis-a-vis standard (average) price function. As mentioned above, if our argument about the effect of liquidity premium on the average yield curve is indeed correct and the frontier function approach controls for it appropriately, then we would expect the terms structure associated with the later should lie below that of the former. This is because illiquidity has a depressing effect on prices and if one attributes all the variations in bond prices to the interest rates, then the later will be over estimated. In Table 6 we present the averages of the difference between the frontier and average yield curve taken over the number of days across years. The results clearly indicate that, for most part, the fronteir term structure lies below the average term structure, and the differences are significant. For the years 2001 and 2002, the average difference ranges from 15 to 30 basis points over various tenors implying that omission of liquidity factors bias not only the level but slope of the yield curve as well. Typically the overestimation is more at the short end. This could be due to the fact as the residual maturity falls below one year, the trading activity in government securities falls and the illiquidity rises due to which the standard approach overestimates the interest rates at the short end. To the extent the term structure is expected to act as a guide for monetary policy in so far as it captures inflation expectations, it is important not to use the average term structure as the later is signifcantly affected by the omitted illiquidity factors. The frontier yield curve on the otherhand will provide a better alternative since it explicitly controls for the effect of illiquidity factor.

From the point of view of pricing bonds and other bond derivatives, the biases caused by using the average yield curve could be substantial. The frontier yield curve, however, by itself does not provide a better alternative since it can generate only a shadow liquid price for every security in the portfolio and not a price at which the secuirty could potentially be liquidated. For this purpose, we need to deduct the estimated illiquidity premia for a security, given its attributes, from the liquid or frontier price that it could have commanded had it been most liquid. Note, therefore, that we have two models that together generate the price of a bond: frontier yield curve giving a liquid price and an illiquidity function giving an illiquidity premia. Whether these models predict the price of a bond in a finer manner than an average yield curve is purely an empirical question that can be answered by analysing how these predicted prices fare vis-a-vis observed bond prices. It is to this end that we now turn. Table 7 presents the mean absolute errors (MAE) of our specification vis-a-vis standard approach across volume and maturity classes for the years 2001 and 2002. The numbers in each cell are the averages over the securities and number of days within each year. The results clearly show that the MAE statistics for the frontier model are signifcantly smaller for all classes and across years, indicating that the pricing based on the proposed framework tracks the observed prices much more closely than the standard framework. Also to be noted is the fact that MAEs associated with the standard approach clearly increase with maturity questioning the assumption that the errors in pricing equation are random. No such patterns can be found in the MAEs associated with the frontier function model validating the basic specification. The superior 'in-sample' fit provided by the proposed model, therefore, provides a strong case for the later as a benchmark for pricing bonds and their derivatives. It should, however, be noted that there is an increase in the MAEs across classes over years 2001 and 2002 for both the proposed and standard models. This could be due to the increase in the volatlity of interest rates and bond markets in the recent past due to various policy changes. To the extent some of these factors havent been taken into account explicitly in modelling the illiquidity premia, the frontier function model will also not capture the increased variation in bond prices.

6 Conclusion

Estimating a risk free Term Structure of Interest Rates or Zero Coupon Yield Curve from the observed bond prices would involve controlling for the effects of security specific non-interest rate factors that affect bond prices. In this paper we propose a new framework to estimate bench-mark default and liquidity risk free yield curve using the stochastic frontier functions. The methodology explicitly controls for the effects of security specific factors such as age, issue size, coupon and residual maturity on bond prices in estimating the yield curve. The proposed model is estimated using the daily secondary market data from NSE-WDM for the period Jan 1997 to July 2002. We find that the new methodlogy not only identifies frontier yield curve that is significantly different from the standard zero coupon yield curve and gives reasonable estimates of liquidity premia, but also performs better than the standard yield curve models in predicting bond prices.

In terms of future work, the present exercise can be extended in four related directions. First, to establish the superiority of the proposed model over the standard yield curve models, it is important to obtain the out-of-sample comparisons. One neds to design a clever sample separation methodology for this purpose as any simplistic schemes of splitting the sample into in-sample and out-of-sample data sets will lead to biases in the estimation and testing. Second, the proposed framework could be extended by considering alternate functional forms both for the term structure relation and the densities for error term U. For example, one could consider Cubic or B spline models for the yield curve and exponential or gamma ditribution functions for the one-sided error terms characterizing the illiquidity premia. Third, the model for illiquidity premia could be extended by considering alternative functional forms through which the security specific factors affect the illiquidity premia. Any improvement in the fit in this regard will improve the overall fit of the model. Finally, assuming that the illiquidity function is stable over some pre-specified period of time, one could consider formulating this problem in a panel frontier function framework by pooling the time series of cross sectional data. This would signifcantly increase the precision with which we can estimate the illquidity premia associated with each security. We propose to undertake some of these extensions in our future work.

Appendix I - Econometrics of Frontier functions[5]

The bond pricing equation that accounts for the effects of liquidity premia may be written as,

$$P_d = P^* - U + \upsilon \tag{4}$$

where

$$P^* = \Sigma_{i=1}^{T} \frac{CF_i}{(1+r_i)^i},$$

$$\upsilon \sim N(0, \sigma_\upsilon^2),$$

and

$$U \geq 0.0$$

and

$$U \sim N^+(\mu_U, \sigma_U^2).$$

The error term U, being a non-neagtive (turncated) normal random variable, is expected to capture the effect of liquidity premium on bond prices. The second error term υ, assumed to be normal random variate, is expected to capture noise due to the effect of all other factors that are omitted from the bond pricing equation.

The density functions of these random variables, U and υ, are, therefore, given by:

$$f(\upsilon) = \frac{1}{\sqrt{2\pi}\sigma_\upsilon} exp(-\frac{\upsilon^2}{\sigma_\upsilon^2})$$

$$f(U) = \frac{2}{\sqrt{2\pi}\sigma_U \Phi(-\frac{\mu_U}{\sigma_U})} exp\left[-\frac{(U-\mu_U)^2}{\sigma_U^2}\right]$$

[5]This section borrows from Kumbhakar (2000)

where μ is the mode of the normal distribution truncated from below at zero, and Φ is the standard normal cumulative distribution function.

Under the assumption that υ and U are independently distributed, we can write their joint fdensity as,

$$f(\upsilon) = \frac{1}{\sqrt{2\pi}\sigma_{\upsilon}} exp(-\frac{\upsilon^2}{\sigma_{\upsilon}^2})$$

$$f(U,\upsilon) = \frac{2}{\sqrt{2\pi}\sigma_U\sigma_{\upsilon}\Phi(\frac{-\mu_U}{\sigma_U})} exp\left[-\frac{(U-\mu_U)^2}{2\sigma_U^2} - \frac{\upsilon^2}{2\sigma_{\upsilon}^2}\right]$$

Since, $\varepsilon = \upsilon - U$, we can write the joint density of U and ε as,

$$f(U,\varepsilon) = \frac{2}{\sqrt{2\pi}\sigma_U\sigma_{\upsilon}\Phi(\frac{-\mu_U}{\sigma_U})} exp\left[-\frac{(U-\mu_U)^2}{2\sigma_U^2} - \frac{(\varepsilon+U)^2}{2\sigma_{\upsilon}^2}\right]$$

The marginal density of ε is

$$f(\varepsilon) = \int_0^{\infty} f(U,\varepsilon)dU = \frac{1}{\sigma}\phi\left(\frac{\varepsilon+\mu}{\sigma}\right)\Phi\left(\frac{\mu}{\sigma\lambda} - \frac{\varepsilon\lambda}{\sigma}\right)\left[\Phi\left(-\frac{\mu}{\sigma_U}\right)\right]^{-1}$$

where, $\sigma = \sqrt{(\sigma_U^2 + \sigma_{\upsilon}^2)}$ and $\lambda = \frac{\sigma_U}{\sigma_{\upsilon}}$, and ϕ is the standard normal density function.

The composite error, ε, is asymmetrically distributed with mean and variance given by,

$$E(\varepsilon) = -E(U) = -\frac{\mu\alpha}{2} - \frac{\sigma_U\alpha}{\sqrt{2\pi}} exp\left\{-\frac{1}{2}\left(\frac{\mu}{\sigma_U}\right)^2\right\},$$

$$V(\varepsilon) = \mu^2\frac{\alpha}{2}\left(1-\frac{\alpha}{2}\right) + \frac{\alpha}{2}\left(\frac{\pi-\alpha}{\pi}\right)\sigma_U^2 + \sigma_{\upsilon}^2$$

where, $\alpha = \left[\Phi\left(-\frac{\mu}{\sigma_U}\right)\right]^{-1}$.

The log-likelihood function for a sample of N observations can be written as

$$lnL = -Nln\sigma - Nln\Phi\left(-\frac{\mu}{\sigma_U}\right) + \sum_i \left\{ ln\Phi\left(\frac{\mu}{\sigma_U\lambda} - \frac{\varepsilon_i\lambda}{\sigma}\right) - 0.5\left(\frac{\varepsilon_i + \mu}{\sigma}\right)^2 \right\}$$

The log-likelihood function can be maximized with respect to the parameter vector to obtain the ML estimates of the parameter vector $(\beta; \mu; \sigma_U; \sigma_v)$. Given the degree of non-linearity of both the bond pricing model and truncated distribution model, the maximization of the likelihood function will have to be done using numerical nonlinear optimization methods. Since the objective function is a highly non-linear function in parameters, standard maximization methods such DFP or BFGS may have difficulty in finding the global optimum (Goffe (1998)). Following the suggestion of Goffe(1998), we use the Simulated Annealing algorithm, which is found to be robust in finding the parameter vector associated with the global optimum.

Given the joint density of U and ε and the marginal density of ε, we can derive the conditional density of U as:

$$f(U|\varepsilon) = \frac{f(U,\varepsilon)}{f(\varepsilon)} = \frac{1}{\sqrt{2\pi}\sigma_*[1 - \Phi(-\mu_i^*/\sigma_*)]} exp\left\{-\frac{(U-\mu_i^*)^2}{2\sigma_*^2}\right\},$$

where $\mu_i^* = \frac{(-\sigma_U^2\varepsilon + \mu\sigma_v^2)}{\sigma^2}$ and $\sigma_*^2 = \frac{\sigma_U^2\sigma_v^2}{\sigma^2}$.

Following Jandrow et all (1982), we can derive the conditional mean of U is given by

$$E(U_i|\varepsilon_i) = \sigma_*\left[\frac{\mu_i^*}{\sigma_*} + \frac{\phi(\mu_i^*)}{1 - \Phi(-\mu_i^*)}\right].$$

The point estimates of the conditional and unconditional illquidity premia for any bond with observed price P_i is thus given by the estimators $E(U|\varepsilon)$ and $E(\varepsilon)$ respectively.

Selected References

Alonso, F., R. Blanco, A. del Rio, and A. Sanchis (2000); "Estimating Liquidity premia in Spanish Government securities market", BIS working paper, Bank of International Settlements, Switzerland.

Amihud, Y. and H. Mendelson (1986); "Liquidity, maturity and the yields on U.S. Treasury securities", Journal of Finance, 46, 1411-1425.

Anderson, N., F. Breedon, M. Deacon, A. Derry and G. Murphy (1996); 'Estimating and Interpreting the Yield Curve'; Wiley Series in Financial Economics and Quantitative Analysis.

Bliss, R.R. (1997); "Testing term structure estimation methods", Advances in Futures and Options Research, 9, 197-232.

Bolder, D. and D. Streliski (1999); "Yield Curve Modelling at the Bank of Canada", Bank of Canada Working Paper.

Carleton, Willard T. and Ian A. Cooper (1976); "Estimation and Uses of the Term Structure of Interest Rates", Journal of Finance XXXI (4), 1067-1083.

Crabbe, L.E. and C. Turner (1995); "Does the Liquidity of a Debt Issue Increase with its Size? Evidence from the corporate bond and medium term note markets", Journal of Finance, L(5), 1719-35.

Darbha, G., Dutta Roy S., and V. Pawaskar (2001); "Term structure of Interest Rates in India - issues in estimation", forthcoming in Indian Economic Review.

Darbha, G., Dutta Roy S., and V. Pawaskar (2002); "Idiosyncratic factors in pricing of sovereign bonds - An analysis based on GoI bond market", forthcoming in Journal of Emerging Markets Finance.

Elton, Edwin J. and T. Clifton Green (1998); "Tax and Liquidity Effects in Pricing Government Bonds", Journal of Finance, LIII(5), 1532-62.

Fisher, M., D. Nychka and D. Zervos (1995); "Fitting the term structure of interest rates with smoothing splines", Working paper 95-1, Finance and Economics discussion series, Federal Reserve Board.

Greene, W. (1993); 'Econometric Analysis', Prentice Hall.

Jandrow,J., Lovell, C.A.K, Materov, I. and P. Schmidt (1982); "On the estimation of inefficiency in the stochastic production function model", Journal of Econometrics, 19, 233-238.

Kamara, A. (1994); "Liquidity, taxes and short-term treasury yields", Journal of Financial and Quantitative Analysis, 29, 403-17.

Kumbhakar, S., S. Ghosh and J. McGuckin (1991); "A Generalzed Production Frontier approach for estimating determinants of inefficiency in US dairy farms", Journal of Business and Economic Statistics, 9, 279-286.

McCulloch, J.H. (1971); "Measuring the Term Structure of Interest Rates", Journal of Business, XLIV (January); 19-31.

McCulloch J.H. (1975); "The Tax Adjusted Yield Curve", Journal of

Finance XXX (3), 811- 830.

Mastronikola, K. (1991); "Yield Curves for Gilt-edged stocks: A new model", Technical series, 49, Bank of England Discussion paper.

Nelson, Charles R. and Andrew F. Siegel (1987); "Parsimonious Modeling of Yield Curves", Journal of Business 60(4), 473-89.

Nag, Ashok K. and Sudip K. Ghose (2000); "Yield Curve Analysis for Government Securities in India", Economic and Political Weekly, January 29, 339-48.

Robichek, Alexander A. and W. David Neibuhr (1970); "Tax-induced Bias in Reported Treasury Yields", Journal of Finance XLIII (3); 1081-1090.

Sarig, O. and A. Warga (1989); "Bond price data and bond market liquidity", Journal of Financial and Quantitative Analysis, 24, 367-378.

Schaefer, Stephen M. (1981); "Measuring a Tax-specific Term Structure of Interest Rates in the Market for British Government Securities", The Economic Journal, 91 (June), 415-438.

Schoenberg, R. (1998); 'Constrained Optimisation - Application Module', GAUSS Applications, Aptech Systems Inc.

Seber, G.A.F and C.J. Wild (1989); 'Nonlinear Regression', Wiley Series in Probability and Mathematical Statistics, John Wiley and Sons Inc.

Shea, G. (1985); "Interest Rate term structure estimation with exponential splines: A note", Journal of Finance, 11, 319-25.

Subramanian, K.V (2001); "Term Structure Estimation in Illiquid Government Bond Markets: An Empirical Analysis for India", Journal of Fixed Income, June, 77-86.

Steely, J. (1991); "Estimating the Gilt-edged term structure: basis splines and confidence intervals", Journal of Business, Finance and Accounting, 18, 512-529.

Svensson, Lars E.O (1994); "Estimating and Interpreting Forward Interest Rates: Sweden 1992-1994", NBER WP 4871.

Thomas, S. and V. Saple (2000); "Estimating the Term Structure of Interest Rates in India", unpublished paper, IGIDR.

Vasicek, O.A. and H.G. Fong (1982); "Term structure modelling using exponential splines", Journal of Finance, XXXVII, 339-348.

Waggoner, F.D. (1997); "Spline methods for extracting interest rate curves from coupon and bond prices", Working paper 97-10, Federal Reserve Bank of Atlanta.

Warga, A., (1992); "Bond returns, Liquidity and missing data", Journal of Financial and Quantitative Analysis, 27, 605-617.