On Estimability of Parsimonious Term Structure Models: An Experiment with Three Popular Specifications

Vineet Virmani

W.P. No. 2006-06-01

June 2006

INDIAN INSTITUTE OF MANAGEMENT
AHMEDABAD-380 015
INDIA

On Estimability of Parsimonious Term Structure Models: An Experiment with Three Popular Specifications

Vineet Virmani
Post-Doctoral Fellow
Finance Area
Wing 11 – J
Indian Institute of Management (IIM)
Vastrapur
Ahmedabad - 380015
Ph: 91 79 2632 4909
Fax: 91 79 2630 6896
Email: vineet@iimahd.ernet.in

Abstract

This study addresses operational issues in estimation of parsimonious term structure models. When using price errors, objective function in term structure estimation is a highly non-linear function of the parameters. This necessary entails using numerical optimization techniques for estimation, which brings to fore the issue of (sensitivity of results to) the choice of initialization of the optimization routine.

This study assesses the sensitivity of the final objective function value and the final parameter vector to the choice of the initial vector for three popular specifications, namely, Nelson-Siegel (1987), Svensson (1994), and Cox-Ingersoll-Ross (1985). It turns out that given the nature of the objective function, the choice of the starting vector is far from obvious in all three cases. There exist regions in the shape of the objective function in all three where a slight change in (seemingly reasonable) initial vector takes one far from optimum. Choice of the (range of) 'best' starting vector turns out to be an empirical matter. Grid search is recommended. One must first get to a subset of initial values which results in the objective function value near the minimum and then assess the sensitivity of the final parameter vector to those (subset of) initial values. The study illustrates the process using a typical trading day's data.

[Preliminary Draft. Please Don't Quote]

I. Introduction

Both single factor equilibrium models (e.g. Vasicek, 1977, Brennan and Schwartz 1979, Cox-Ingersoll-Ross, 1985 etc.) and models of the Nelson-Siegel (1987; NS) family are quite popular in the term structure literature for modelling interest rates. Their attractiveness lies not only in their parsimony but also in their ability to give economically sensible estimates of the term structure.

Despite their popularity (or perhaps because of it), operational issues are rarely, if ever, discussed while reporting the results[1]. For anyone familiar with the functional form of these models, it is easy to see that the objective function (when using price errors) is a highly non-linear function of the parameters. This necessary entails using numerical optimization techniques for estimation. Though, depending on the exact nature of the objective function, estimability may or may not be an issue, unless there is some idea about the behaviour of the objective function and sensitivity of the final parameter vector to initialization, *a priori*, well, we just don't know.

This study is an experiment to see how the objective function value (*fval* hereafter) and the final parameter vector ($\tilde{b}$ hereafter) vary as the initial parameter vector (b_0 hereafter) is changed around its neighbourhood for a typical trading day's data. Three popular parsimonious models of the term structure, namely, NS, Svensson (1994; SV), and Cox-Ingersoll-Ross (1985; CIR) are selected for the purpose.

Since the literature is replete with the description/applications of term structure methods studied here (see Bliss, 1997, Ioannides, 2003 and others for a survey and original papers for the details), after briefly describing the methodology of estimation we move to study the behaviour of the objective function

II. The Objective Function

Estimating a term structure, Bliss (1997) notes, requires decision on the following three aspects:

[1] A notable exception is Bolder and Streliski (1999)

1. A Pricing Function
2. A Discount/Rate Function
3. Estimation Technique (the objective function)

2.1 The Pricing Function

In the literature (see, for example, Bolder and Streliski, 1999, Bliss, 1997, Darbha, Roy and Pawaskar, 2003 etc.) it is standard to specify the price of a default-risk free bond, in absence of arbitrage, as:

$$P = \sum_{m=1}^{M} c_m \delta_m \qquad \textbf{[1]}$$

where M is the time to maturity of the bond, c_m is the cash flow received at time m, and δ_m is what is called the 'discount function' in the term structure literature. The above equation relates the discounted cash flows from the bond in discrete time periods to the price of the bond. It is a rather straight forward matter to convert 'discount function' to a 'rate function' using the following equation:

$$r(m) = -\frac{\ln(\delta_m)}{m} \qquad \textbf{[2]}$$

Since, conditions for perfect markets don't exist in reality, and cash flows are received only at discrete times, in practice one needs to give a stochastic form to equation **[1]**, such as:

$$P = f\ [c_m, r(m)] + \varepsilon \qquad \textbf{[3]}$$

where ε is the 'error' term and accounts for whatever is not captured in the function f about how bonds are priced. Bliss (1997) uses the term "omitted pricing factors" for

"...factors which have been omitted from the bond pricing equation which nonetheless impact the pricing of bonds[2]"

2.2 The Discount / Rate Function

The next decision in the exercise of term structure estimation involves the selection of a form for the discount rate function. As stated already, the forms studied here are:

1. Nelson-Siegel (1987)
2. Svensson (1994)
3. Cox-Ingersoll-Ross (1985)

While the first two in the above list are empirical curve fitting exercise, the last belongs to a set of models arising from intertemporal general equilibrium description of a competitive economy with utility maximizing agents. A brief description of the three models follows. Details can be found in the references cited earlier.

2.2.1 NS

NS assume that the instantaneous forward rate is the solution to a second order differential equation with two equal roots. The forward rate function used by NS is:

$$f(m\,;b) = \beta_0 + \beta_1 \exp(-m/\tau_1) + \beta_2 \frac{m}{\tau_1} \exp(-m/\tau_1) \qquad \textbf{[4]}$$

where $b \equiv (\beta_0\,, \beta_1\,, \beta_2\,, \tau)$ is the vector of parameters to be estimated. The spot rate function can in turn be derived by integrating the above equation. This gives:

$$s(m\,;b) = \beta_0 + \left(\beta_1 + \beta_2\right)\frac{1 - \exp(-m/\tau_1)}{m/\tau_1} - \beta_2 \exp(-m/\tau_1) \qquad \textbf{[5]}$$

[2] R. R. Bliss (1997), "Testing Term Structure Estimation Methods", Advances in Futures and Options Research, 9

The spot rate function has four parameters. While β_0 and $\beta_0 + \beta_1$ are implied long-rate and short-rate respectively, β_2 gives the medium term component of the yield curve, and along with τ defines the shape of the curve. The possible shapes of the term structure that result as parameters vary can be found in NS, SV and Bolder and Streliski (1999) and won't be discussed here.

2.2.2 SV

SV adds a fourth term to the forward rate function given by NS, with two additional parameters, (β_3, τ_2), thereby adding to the flexibility of the shape of the term structure (possibility of a second 'hump' – or what is often referred to as an *S*-shaped curve in the literature – with β_3 and the other time decay parameter, τ_2). The corresponding functions are then given as:

$$f(m\,;b) = \beta_0 + \beta_1 \exp(-m/\tau_1) + \beta_2 \frac{m}{\tau_1}\exp(-m/\tau_1) + \beta_3 \frac{m}{\tau_2}\exp(-m/\tau_2) \qquad [6]$$

$$s(m\,;b) = \beta_0 + \left(\beta_1 + \beta_2\right)\frac{1-\exp(-m/\tau_1)}{m/\tau_1} - \beta_2 \exp(-m/\tau_1) + \beta_3 \frac{1-\exp(-m/\tau_2)}{m/\tau_2} - \beta_3 \exp(-m/\tau_2) \qquad [7]$$

2.2.3 Empirical Implications of the Cox-Ingersoll-Ross Model

The dynamics of the interest rate process in CIR is given as[3]:

[3] κ is the mean reversion coefficient, θ is the mean of the process, *r* is the instantaneous short rate, σ^2 is the scale factor for variance of *r*, and λ is the price of risk associated with *r*.

$$dr = \kappa\,(\theta - r)\,dt + \sigma\,\sqrt{r}\,dz \qquad \textbf{[8]}$$

CIR, just like other affine models, in absence of arbitrage, results in the following pricing equation:

$$P\,[r, t, T] = A\,[t, T]\,e^{-B\,[t, T]\,r} \qquad \textbf{[9]}$$

where for $\tau = T - t$

$$A\,[t,T] = \left[\frac{\phi_1 \exp\,(\phi_2\,\tau)}{\phi_2\left(\exp\,(\phi_1\,\tau) - 1\right) + \phi_1}\right]^{\phi_3} \qquad \textbf{[10]}$$

$$B\,[t,T] = \frac{\exp(\phi_1\,\tau) - 1}{\phi_2\left(\exp(\phi_1\,\tau) - 1\right) + \phi_1} \qquad \textbf{[11]}$$

where

$$\phi_1 = \left((\kappa + \lambda)^2 + 2\sigma^2\right)^{1/2} \qquad \textbf{[12]}$$

$$\phi_2 = \left(\kappa + \lambda + \phi_1\right) / \, 2 \qquad \textbf{[13]}$$

$$\phi_3 = 2\kappa\theta \,/\, \sigma^2 \qquad \textbf{[14]}$$

Value of a coupon bond can then be written as:

$$V\;[t, c, d] = \sum c_i\,P(r, t, d_i) \qquad \textbf{[15]}$$

where *d* is the vector of coupon payment dates.

Then, given the prices of the traded bonds, one can estimate the parameters ϕ_1, ϕ_2, ϕ_3 and *r* (though for actual dynamics it is not possible[4] to separately identify the parameters,

[4] for risk-neutral dynamics with $\lambda = 0$, the parameters of the process can be uniquely identified from equations 12 – 14

θ, κ and λ). The long-rate and volatility of the short-rate are given as a function of the parameters ϕ_1, ϕ_2, ϕ_3 as follows:

$$r_\infty = \phi_3 \left(\phi_1 - \phi_2\right) \quad [16]$$

$$\sigma^2 = 2\,\phi_2\left(\phi_1 - \phi_2\right) \quad [17]$$

Before moving further, it must be acknowledged that the theoretical CIR model describes the process for *real* rates, as opposed to nominal rates. However, that said, it is still attractive for modelling nominal rates because the structure of the model precludes negative interest rates.

It is also intuitive because, like NS and SV, the model implies that the long rate $(m \rightarrow \infty)$ converges to a constant. Now, although, volatility of the yield of the longest maturity bond traded in the money market is clearly not zero, the fact that it converges to a constant makes it appealing.

2.3 The Objective Function and the Estimation Strategy

The optimization problem is to minimize the weighted sum of square of (price) errors, i.e. the objective function is:

$$min \sum_i^N \left(\omega_i \varepsilon_i\right)^2 \quad [18]$$

subject to non-negativity constraints imposed on the short-rate, the long rate $(m \rightarrow \infty)$ and on the τ *s;* where $\varepsilon_i = P_i - \hat{P}_i$, and

$$\omega_i = \frac{1/d_i}{\sum_j^N 1/d_j} \quad [19]$$

where d_i is the Macaulay duration of the i^{th} bond[5].

The loss function above has been specified as a function of price errors. An alternative exists in taking the yield errors. However, since it is the bond *prices* that are traded in the market, and not the yields, it makes sense to specify a loss function in terms of the variable which is directly observed / traded in the market. Further, the weighting scheme used – other than taking care of heteroskedasticity – also takes care of minimizing yield errors indirectly. Recall that duration is a function of first derivative of price w.r.t yield, and the weighting scheme is inverse of duration.

The basic process of determining the minimum involves selecting initial vector of parameters, finding pricing errors based on starting vector and using a suitable routine for selecting the optimum.

III. Sensitivity of *fval* and $\tilde{b}$ to b_0: The Methodology

Before discussing the methodology used to assess the sensitivity of *fval* and $\tilde{b}$ to b_0, it is must be clarified that the experiment is *not* an evaluation of the optimization routine used (Sequential Quadratic Programming; SQP; as implemented in Optimization Toolbox of MATLABv6.5 R13; procedure: *fmincon*) to estimate the parameters. To the extent that the same technique is used to estimate the parameters across a range of initial values, loosely speaking, it does provide evidence on the routine too, but experiment is not designed for that purpose. The results, then, should be taken to be conditional on the routine used to estimate the parameters.

SQP methods are standard general purpose algorithms for solving not too big, smooth nonlinear optimization problems. It uses the popular BFGS (Broyden-Fletcher-Goldfarb-Shanno, 1970) solver in optimization literature to estimate the parameters. Quoting from MATLAB's manual for the Optimization Toolbox,

[5] This weighing scheme corrects for the heteroskedasticity problem in the error terms which occurs if the price errors are used instead of yield error. See Coleman, Fisher and Ibbotson (1995), Bliss (1997) and Bolder and Streliski (1999) for a discussion. Using duration weighted loss function is also a proxy to minimize yield errors when price errors are used in the loss function. Subramanian (2001) uses a liquidity (instead of duration) weighted loss function

"SQP methods represent the state of the art in nonlinear programming methods. Schittkowski (1985), for example, has implemented and tested a version that outperforms every other tested method in terms of efficiency, accuracy, and percentage of successful solutions, over a large number of test problems.[6]*"*

The problem dealt with here is clearly not big (4 parameters in two of them and 6 parameters in one, with trivial linear constraints). Given the nature of the functional forms used here, calculation of the objective function and its derivative are also uncomplicated. The fact that the spot rate (and in turn, bond price) in NS and SV (in CIR, it is one of the parameters) is a function of exponentials, in the least the objective function would be continuous. Weighted by inverse of duration (a function of the bond price itself), the objective function is suitably scaled too.

3.1 Data

To illustrate the behaviour of the objective function, trades data for first day of Jun, 2003 has been used. Though, to an extent arbitrary, the choice of the day was dictated by the number of bonds traded and the value of trades during the day. As it happens, in the history of the Wholesale Debt Market (WDM) segment of the National Stock Exchange (NSE), trading activity in WDM was highest around the middle of 2003.

Roughly 15% of bonds selected at random were kept as out-of-sample bonds[7]. Only bonds with $T + 0$ and $T + 1$ settlement dates have been taken for the purpose of estimation[8]. Value weighted prices are used while calculating pricing errors, and, as mentioned earlier, errors have been weighted by inverse of duration.

[6] T. Coleman, M.A. Branch, and A. Grace , "Optimization Toolbox for use with MATLAB", *MATLAB Optimization Toolbox User's Guide Version 2*, Section 2-24, Math Works Inc, 1999

[7] This being a companion paper to a detailed *time series* analysis of the three term structure models, the out-of-sample portion was 'retained' here too. Otherwise, this is not necessarily required.

[8] accounting for more than 90% of the number of trades on most days; dates where none of the bonds settled on $T + 0$ or $T + 1$ dates, bonds settling on $[T + 2]^{th}$ date is also included

3.2 The Methodology

We proceed in three steps. First each model is estimated using a 'reasonable' first choice b_0 and *fval* is noted. Choice of software apart, assuming that the choice of initial vector is not too off mark, this is, of course, how one goes about getting to $\tilde{b}$. Only this is not known if that is the best/safest thing to do.

Next, we change b_0 around its neighbourhood and study the sensitivity of *fval* and $\tilde{b}$ to b_0. Choice of neighbourhood is to a large extent arbitrary, and depending on the model varies. So, while $b_0 \pm 0.02$ may be too small a step in case of NS and SV, for the fourth parameter (*r,* the short rate) in CIR, this amounts to a step size of 2% interest rate, clearly quite large. The neighbourhood was selected after some trial and error, and the choice, to a degree, is subjective. But since this study is only illustrative in nature, it is alright.

Clearly, sensitivity of both *fval* and b_0 can be best appreciated in a partial derivative sense. At the same time, changing only one parameter at a time is not enough. One has to see how *fval* (and then $\tilde{b}$) varies as different combinations of initial values vary. Given the limitation on three dimensions that we can deal with (graphically) at a time, we judge the sensitivity of *fval* to changes in b_0 taken two at a time. Reporting such results in the form of tables would not only take up a lot of space, but would also make it mighty hard to interpret. So, in the last step, this is how we proceed.

We change two initial values (say, β_0 and β_1 in the case of NS) at a time and plot the variation in *fval* as both β_0 and β_1 vary. This is we do for all combinations of parameters taken two at a time (${}^4C_2 = 6$ in the case of NS and CIR, and ${}^6C_2 = 15$ in case of SV). Once it is known around which (sub) set of b_0 (say, b^*_0) *fval* reaches minimum (or minima, as the case may be), we study the behaviour of $\tilde{b}$ (only) around b^*_0. The idea is that once it is known how *fval* behaves as b_0 is changed locally, for final parameter estimation we select only that subset of b_0 (i.e. b^*_0), for which *fval* is the lower. This takes care of the problem of ending up with initial values for which *fval* is highly sensitive to slight changes around it.

IV. Results

Next we report results for each specification separately.

4.1 Nelson Siegel

For NS, zero coupon yield curve estimates provided by NSE helped guess parameters for first day of Jun, 2003. For NS, b_0 is given in *Table 1.*

With these values, roughly the short-rate for the day becomes 4.58 % and the long-rate 6.72 %. For the chosen b_0, *fval* is 0.00305. For sake of readability, in what follows we report 1000 times *fval*, i.e. the optimum value then is 3.05. Taking the short-rate and the long-rate as the cue, the range for initial values chosen for the four parameters has been taken to be as given in *Table 1b*.

Table 1a

Initial Parameter Vector (b_0) for NS

NS Parameter	Initial Value
β_0	6.7147
β_1	-2.1333
β_2	-0.0735
τ	4.0370

This implies a range (in initialization) for short-rate to be roughly 0.6 % to 8.6 %, and that for long rate to be 4.7 % to 8.7 %. This is the sense in which the term 'reasonable set of initial starting values' was used in the beginning of the study. Short-rate in money market[9] as on 2nd Jun, 2003 was 4.91 %.

[9] Corresponding to then mean Mumbai Inter-bank Offer Bid/Offer Rate (MIBID/MIBOR) as on 2nd Jun, 2003

Table 1b

Selected Neighbourhood of b_0 for NS

NS Parameter	Initial Value Minimum	Initial Value Maximum	Step Size / No. of Steps
β_0	4.7147	8.7147	0.2 / 20
β_1	-4.1333	-0.1333	0.2 / 20
β_2	-2.0735	1.9265	0.2 / 20
τ	2.0370	6.0370	0.2 / 20

During estimation, as required by construction, τ was constrained to be positive. Further, the long rate was constrained to lie between 0 and 20%, short-rate to be greater than -4%[10]. No other constraints were imposed *a priori* during estimation. During the experiment it was found that the upper bound for τ could be reduced from infinity to a large finite value (taken to be 100 here). This helped increase the estimation time. At no point, did τ come near the (imposed) upper bound.

In what follows shape of the objective function is shown as contour plots taking two parameters at a time, this implies in all ${}^4C_2$ (= 6) contour plots for *fval*.

4.1.1 Contour Plots

Figures 1a – 1f represent contour plots for *fval* - *NS* taking two parameters at a time.

For illustration/explanation sake, let's first look at **Figure 1a**. The center of the plot is the region where $fval \leq 3.1$ (with 2 'packets' representing the region where $fval = 3.1$). As we move outward towards the ends, *fval is* generally seen to increase, i.e. we go further away from the minimum. Note that the region between the isolines $fval = 3.1$ and $fval = 3.2$ represents the set of initial values for which $fval \in [3.1, 3.2]$. Similarly, region between the isolines of $fval = 3.1$ represents the region for which $fval \leq 3.1$. Note that, accordingly the region represented by isolines $fval = 3.5$ represents the region for which

[10] exactly = initial estimate of long rate for 2nd Jun, 2003 - 10% ~ - 4%

fval $\geq$ *3.5* (in particular note the 'rhombi' near $\beta_1 \in$ *[-2.25, -2]* and $\beta_0 <$ *5.1*, where *fval* shifts from a very low to a very high value with a only a slight change in β_1).

Although **Figure 1a** in itself says little about the shape of the objective function, it tells us that there exists a set of initial values for which we *fval* is the least/lower. Now, if in this range/subset of b_0 (using our earlier used notation, say, b_0^*), $\tilde{b}$ changes only marginally we can safely say that (in a limited sense) *fval* is well-behaved and that we can rely on our software to give reasonable estimates of $\tilde{b}$ once b_0 is chosen judiciously (i.e. whenever $b_0 = b_0^*$).

Figure 1a

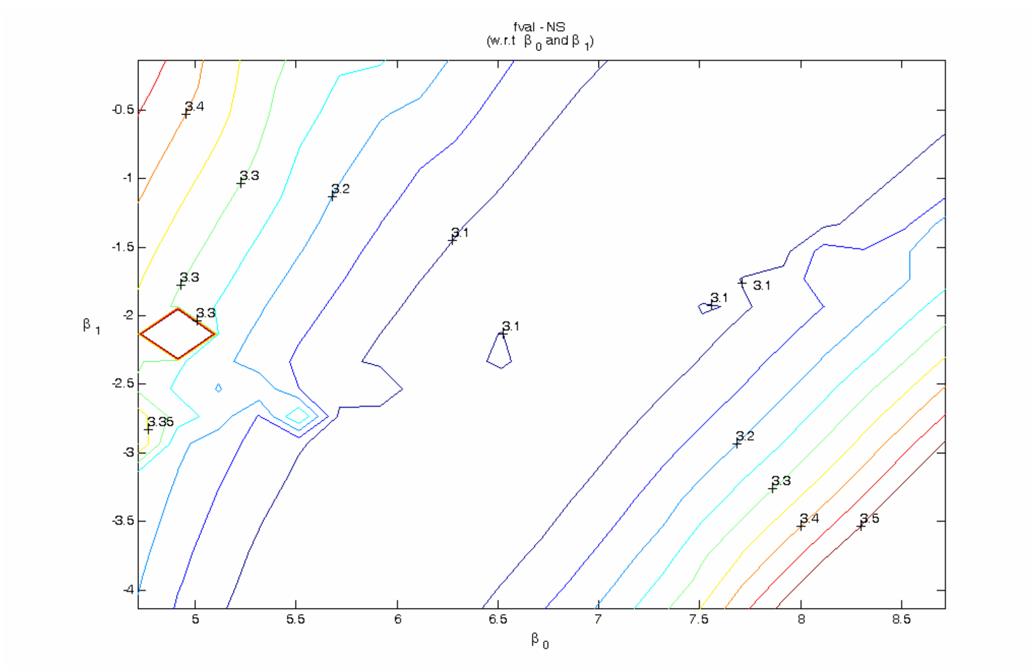
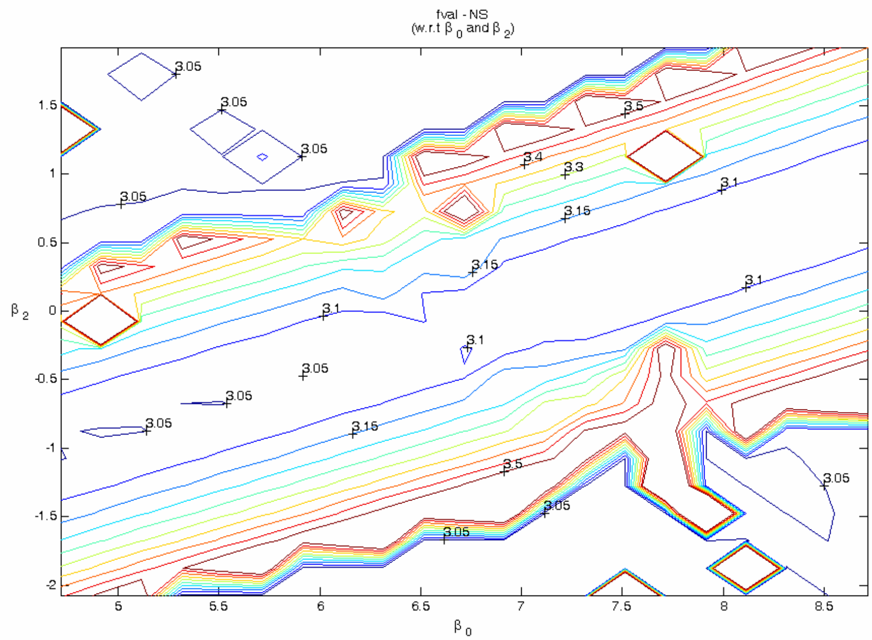


Figure 1a alone, however, gives only a partial picture. Let's look at other plots. Here we can pick either of **Figures 1b** to **1f**, as they are all quite revealing about the characteristics of the (sensitivity) of *fval* to b_0. While, like in **Figure 1a**, in all plots there is a 'central' region for which *fval* is seen to be minimum, *unlike* **Figure 1a**, there appears to be a region even in the periphery where *fval* reaches a minimum, and that too in the vicinity of the points for which it is far away from the optimum. For example, in **Figure 1b**, around region represented by $\beta_0 \in$ *[5.3, 8]* and $\beta_2 \in$ *[-1, 0.2]*, *fval* varies from as low as *3.05*

to as high as *3.5*. Even, within that region there are ‘packets’ where *fval* shifts from a very low to very high value. Note the presence of ‘rhombi’ as pointed out in **Figure 1a** in all the plots.

Figure 1b



It must be reminded at this stage, that these plots *do not* represent the shape of the objective function. They represent sensitivity of *fval* to change in b_0. They show that there exists a set of b_0 for which *fval* is highly sensitive to even slight changes in b_0. *A priori,* thus, one can’t be sure if any given b_0 will result in minimum *fval,* or even a near value near the optimum.

Figure 1c

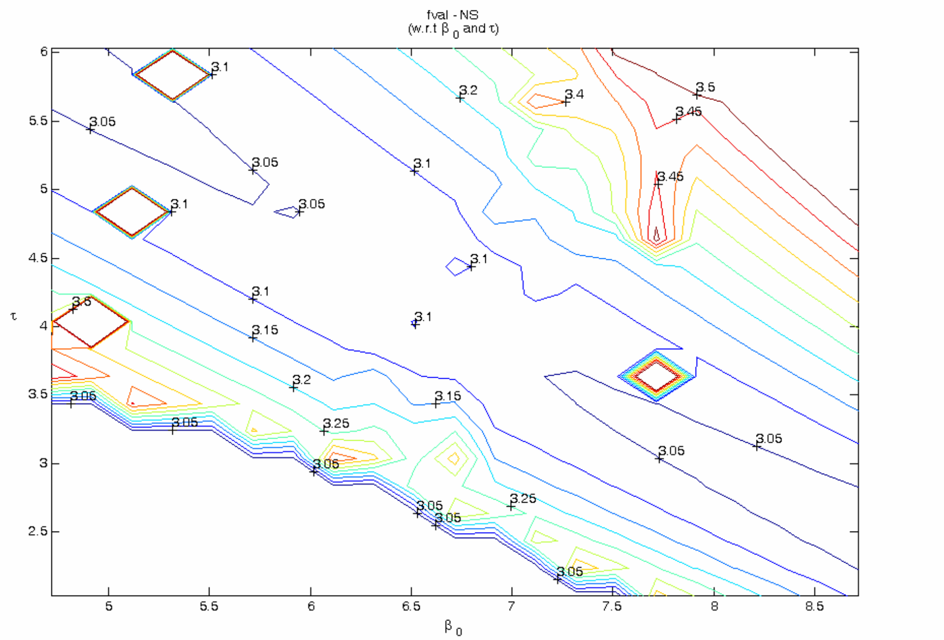


Figure 1d

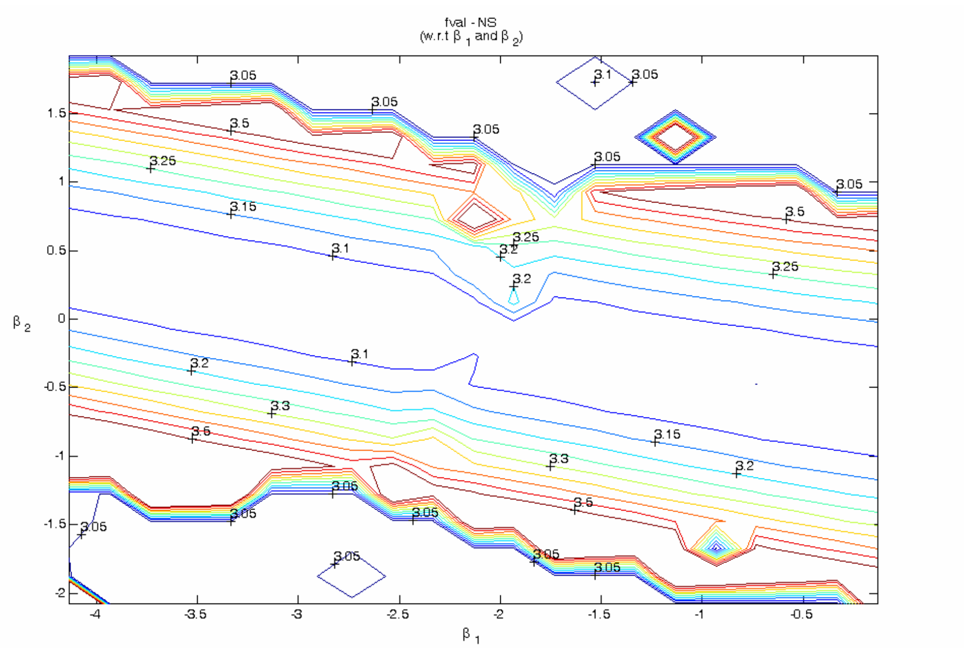


Figure 1e

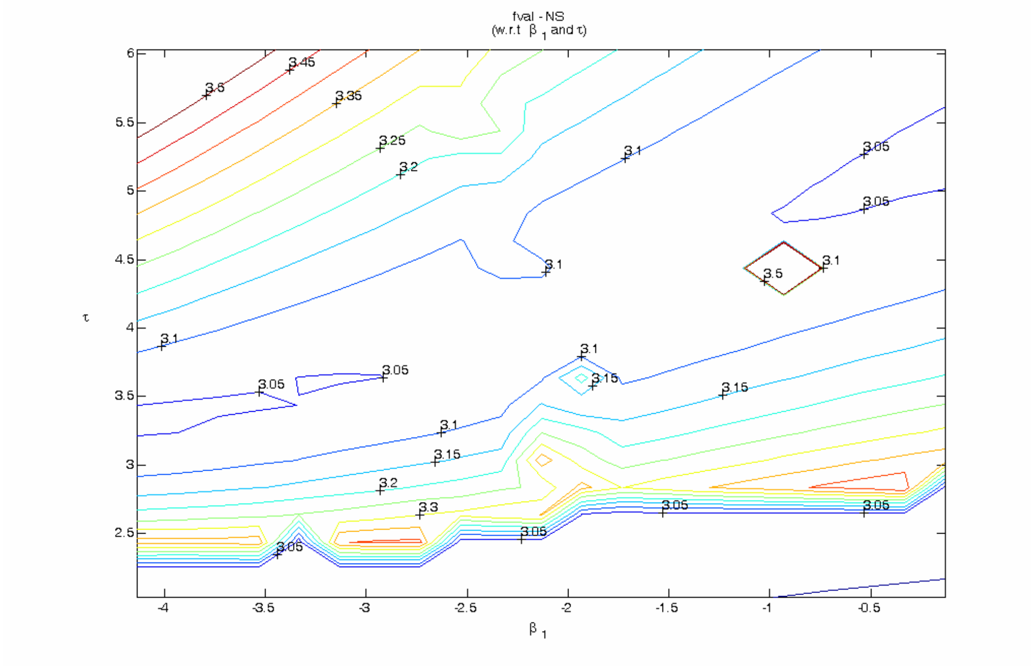
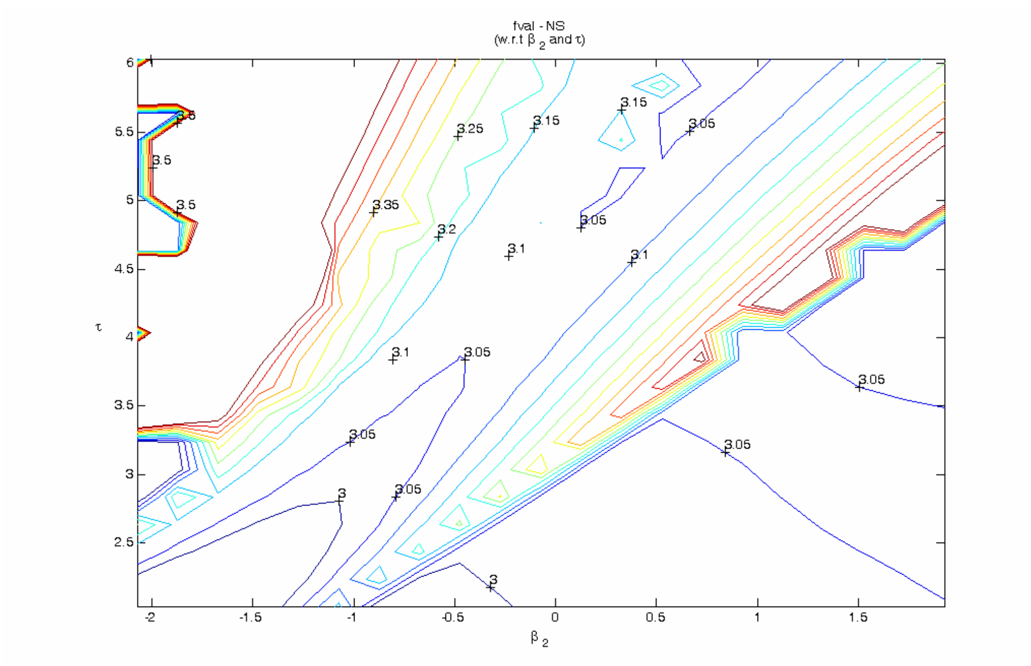


Figure 1f



Finally, look at **Figure 1f**. It is in this plot we see a region with $fval \leq 3$ [the region $\beta_2 \leq -1$ and $\tau \leq 3$]. Thus, given our choice of b_0 'space' there does seem to exist a unique minimum, though not identifiable very easily.

Given the initial value that we selected in *Table 1a*, optimization routine stopped when *fval* reached *3.05*. This says as much about the shape of the objective function, as it does about the SQP method.

Our choice of b_0 for NS was based on actual estimates for the day provided by NSE. Even though NSE's exact objective function/estimation strategy is not known, this is the best one could have done in terms of the selection of the *starting* vector. However, as the contour plots here show, *a priori,* we just can't know which set of initial values results in a minimum *fval*. Thus, selection of a 'good' b_0 is strictly an empirical matter and is not only far from obvious, one cannot realise it unless one studies the sensitivity of *fval* to a range of values around the first choice b_0.

This study, being illustrative in nature, focused only on the local neighbourhood of b_0. However, given the time (< 30 seconds) it takes to estimate the parameters for a day on a PC and the cheap availability of computing resources, unless one must provide real time estimates of the term structure, such sensitivity analysis for an even larger neighbourhood, before narrowing down on the 'best' set of starting values, is very much do-able.

For example, then, given the six contour plots for NS above, the best b_0^* can be roughly identified as $\beta_0 \in [4.7, 7.4]$, $\beta_2 \leq -1$, $\tau \leq 2.25$. As it happens, in the selected neighbourhood, the choice of β_1 doesn't seem to matter. Here, in the specific case of NS, however, a useful starting point is available can be obtained using the money market short-rate. Another alternative exists in constraining the short-rate to the money-market rate.

4.1.2 Sensitivity of $\tilde{b}$ to change in b_0^* ($b_0 : fval \le 3.05$)

If for the region in which $fval \le 3.05$, variation in $\tilde{b}$ turns out to be only marginal or economically insignificant, we needn't worry much. If, however, there is a large variation in $\tilde{b}$ for b_0^* for which $fval \le 3.05$, it is a sign that the objective function is flat for a range of $\tilde{b}$. If, that is indeed the case, then for a given set of b_0, not only the region of minimum *not known*, it is *also unidentifiable* for the problem at hand[11]. This would be bad news.

In what follows we report results on variation in $\tilde{b}$ for which $fval \le 3.05$. For each combination of initial values (as in 6 contour plots above), variation in each component of $\tilde{b}$ is reported as density plots and summary statistics. To preserve space, density plots and summary statistics tables are shown only for those parameters that have direct economic interpretation, i.e. β_0 $(= r_\infty)$ and $\beta_0 + \beta_1$ $(= r_0)$ for NS. Density plots and summary statistics for sensitivity of other parameters are available on request.

- $\beta_0 = r_\infty$ (the Long Rate)

Table 2a

	Min	Max	Mean	Median	Std. Dev.
1 - 2	6.55	6.87	6.73	6.76	0.09
1 - 3	6.69	6.77	6.73	6.73	0.01
1 - 4	6.65	6.77	6.71	6.71	0.02
2 - 3	6.69	6.77	6.73	6.73	0.01
2 - 4	6.68	6.76	6.71	6.71	0.02
3 - 4	6.54	6.77	6.71	6.72	0.04

[11] Note that this is *not* conditional on the optimization routine used. A contour plot is just a plot of variation in *fval* as b_0 changes

Figure 2a

Variation in $\beta_0 = r_\infty$ *(NS)*

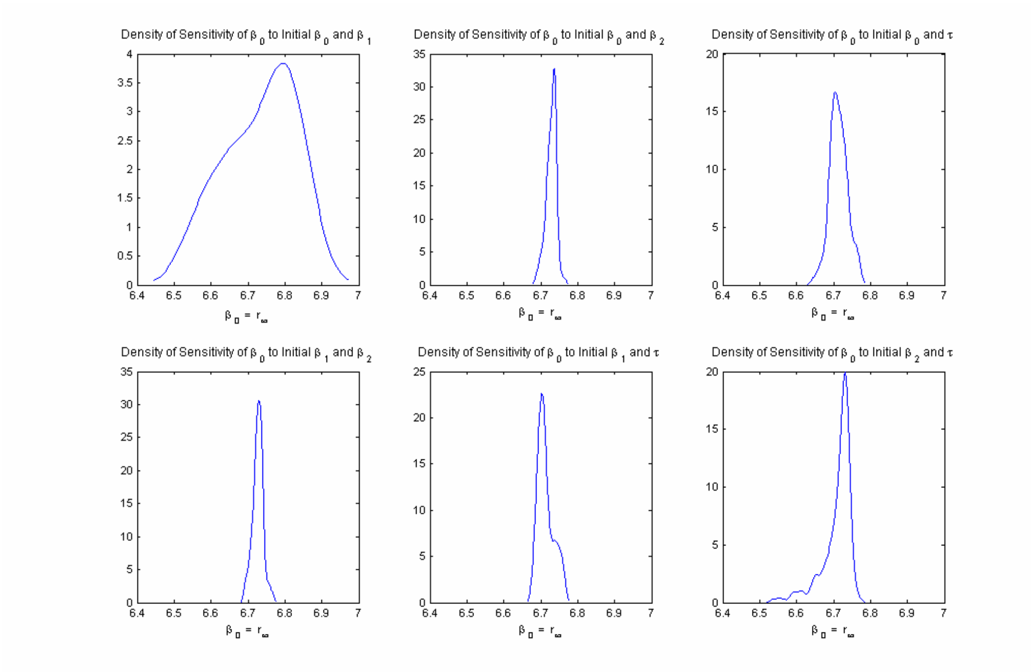
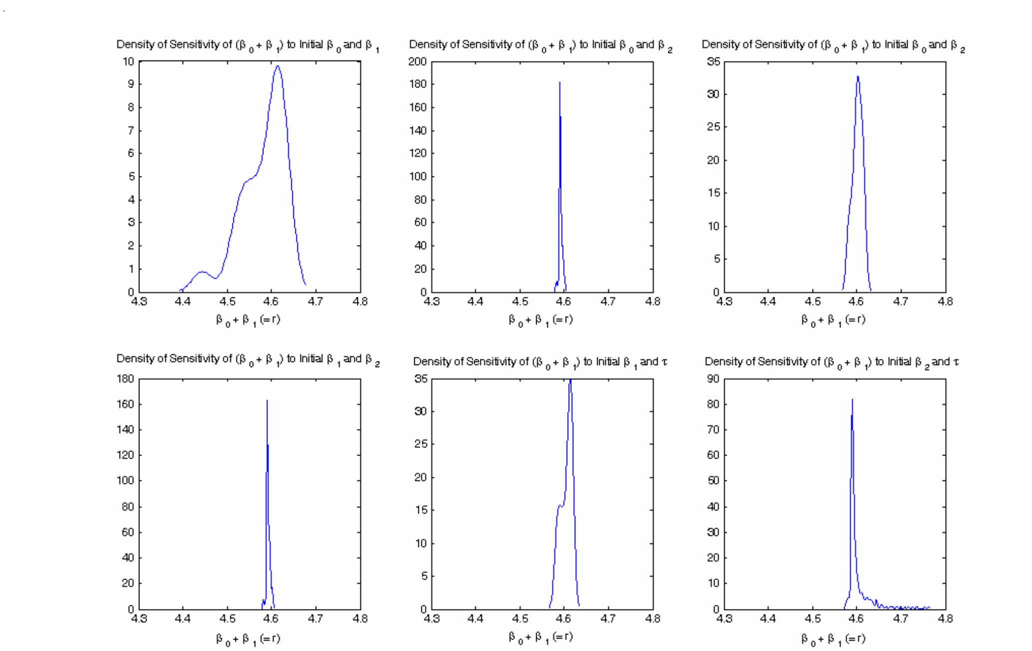


Figure 2b

Variation in $\beta_0 + \beta_1 = r_0$ *(NS)*



- $\beta_0 + \beta_1 = r_0$ (the Short Rate)

Table 2b

	Min	Max	Mean	Median	Std. Dev.
1 - 2	4.43	4.64	4.58	4.60	0.05
1 - 3	4.58	4.60	4.59	4.59	0.00
1 - 4	4.58	4.62	4.60	4.60	0.01
2 - 3	4.58	4.60	4.59	4.59	0.00
2 - 4	4.58	4.62	4.60	4.61	0.01
3 - 4	4.58	4.76	4.61	4.59	0.03

Clearly, in the sub-set of starting values considered here there is little variation in the final parameter values. The range of long rate comes out to be 6.5 % to 6.9 %, and that for short rate 4.4 % to 4.8 %, with most density plots centered around 6.7 % for long rate and 4.6 % for short rate. The summary statistics in *Tables 2a* and *2b* confirm that.

In the subsequent sub-sections, for CIR and SV we use the same strategy as followed in the case of NS. That is, after selecting a first choice b_0 we narrow down on a 'best' b_0^* and report the sensitivity of final parameter values to b_0^*. To preserve space, however, contour plots (in all ${}^4C_2 + {}^6C_2 = 21$) for these would are not shown here. They are available on request from the author.

4.2 CIR

For CIR, while selection of a range of values for r (the fourth parameter) was easy, trial and error had to used to come to a minimum and maximum values for ϕ_1, ϕ_2 and ϕ_3. Short rate (r) was constrained to be greater than 1% and less than 20%. Further, the initial values were chosen to be such that ϕ_1 was always greater than ϕ_2. All parameter values were constrained to be greater than zero during estimation.
The selected initial parameter vector and the neighbourhood then are given in *Tables 3a* and *3b* respectively.

Table 3a

Initial Parameter Vector for NS

NS Parameter	Initial Value
ϕ_1	0.2683
ϕ_2	0.2332
ϕ_3	1.8952
r	0.0436

Table 3b

Initial Parameter Vector Range for CIR

NS Parameter	Initial Value Minimum	Initial Value Maximum	Step Size / No. of Steps
ϕ_1	0.1933	0.2933	0.005 / 20
ϕ_2	0.0332	0.1732	0.007 / 20
ϕ_3	1.1452	2.1452	0.05 / 20
r	0.0436	0.0636	0.001 / 20

The process of identifying the 'best' b_0^* was same as followed in the case of NS. (1000 times) *fval* corresponding to the first choice b_0 (as in *Table 3a*) was 3.048. Then we proceeded to assess the sensitivity of $\tilde{b}$ to change in $b_0 : fval \leq 3.048$.

Now we report the sensitivity of $\tilde{b}$ to b_0^*; again only for those variables/values which have a direct economic interpretation, i.e. the short and the long rate. As in the case of NS, results are reported as summary statistics (*Tables 4a* and *4b*) and density plots (*Figures 3a* and *3b*). Density plots and summary statistics for sensitivity of other parameters are available on request.

➢ $r = r_0$

Table 4a

	Min	Max	Mean	Median	Std. Dev.
1 - 2	4.57	4.65	4.60	4.60	0.02
1 - 3	4.57	4.64	4.60	4.60	0.02
1 - 4	4.56	4.65	4.61	4.61	0.02
2 - 3	4.57	4.64	4.61	4.61	0.02
2 - 4	4.57	4.65	4.61	4.60	0.02
3 - 4	4.57	4.65	4.60	4.60	0.02

➢ $\phi_3 (\phi_1 - \phi_2) = r_\infty$

Table 4b

	Min	Max	Mean	Median	Std. Dev.
1 - 2	6.26	6.79	6.60	6.64	0.15
1 - 3	6.27	6.78	6.65	6.68	0.11
1 - 4	6.28	6.79	6.63	6.70	0.14
2 - 3	6.25	6.78	6.54	6.52	0.14
2 - 4	6.24	6.77	6.60	6.68	0.17
3 - 4	6.24	6.76	6.58	6.62	0.15

For CIR also, in the subset of starting values considered here (b_0^*), variation in the short-rate is limited to 4.55 % to 4.65 %, though in the long-rate is much higher with the range approximately 6.2 % to 6.8 %, though not too different from that obtained in NS, standard deviation in case of CIR is higher.

Following the same procedure as above, finally, we report results for SV.

Figure 3a

Variation in $r = r_0$ *(CIR)*

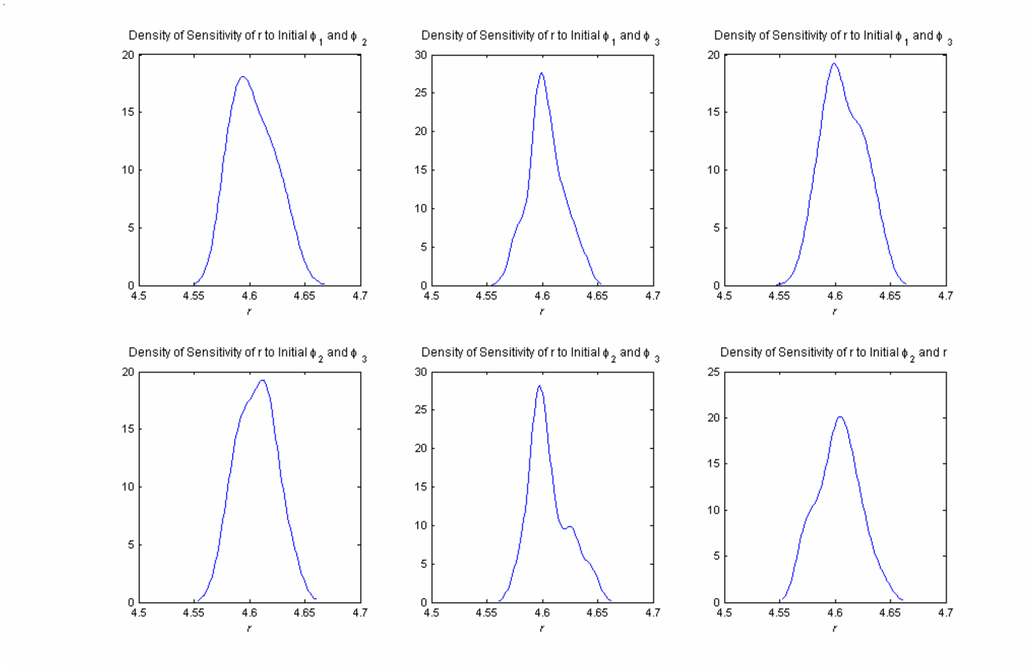
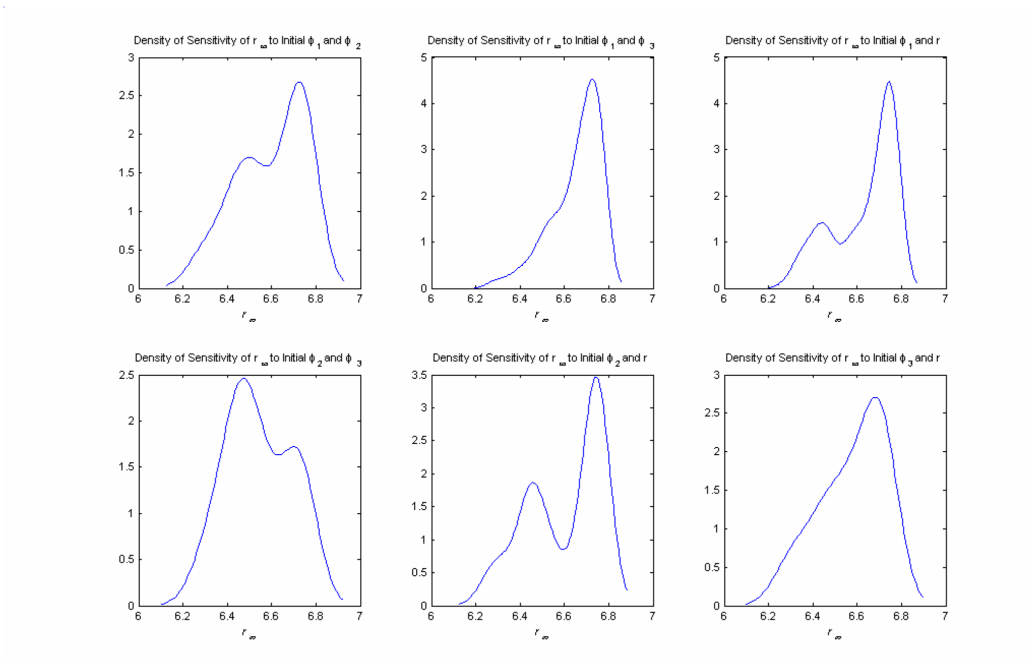


Figure 3b

Variation in $\phi_3\ (\phi_1 - \phi_2) = r_\infty$ *(CIR)*



4.3 SV

Same starting values as NS were used for SV with β_3 being initialized (approx.) same as β_2 and τ_2 initialized as (approx.) 1 more than τ_1 (some degree of trial and error was involved in the process). To enable identification, difference between τ_1 and τ_2 was constrained to be least 0.25. As in NS, both τ_1 and τ_2 were constrained to be greater than zero. Same constraints as in the case of NS were imposed as in SV for the short and the long rate. The selected initial parameter vector and the neighbourhood in the case of SV are given in *Tables 5a* and *5b* respectively.

Table 5a

Initial Parameter Vector for SV

NS Parameter	Initial Value
β_0	6.6578
β_1	-2.0255
β_2	-0.7043
β_3	-0.7133
τ_1	2.1088
τ_2	2.9280

Table 5b

Initial Parameter Vector Range for SV

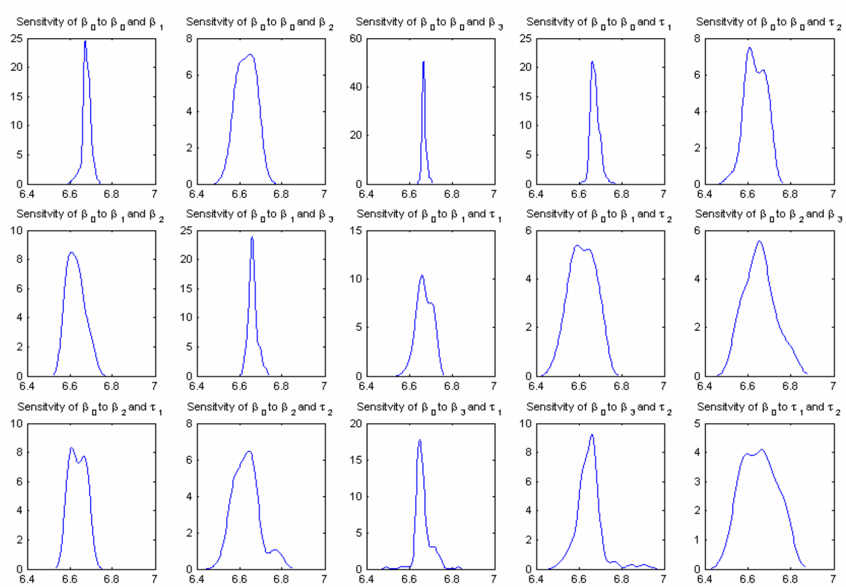
NS Parameter	Initial Value Minimum	Initial Value Maximum	Step Size / No. of Steps
β_0	4.6578	8.6578	0.2 / 20
β_1	-4.0225	-0.0225	0.2 / 20
β_2	-2.7043	1.2957	0.2 / 20
β_3	-2.7133	1.2867	0.2 / 20
τ_1	0.1088	4.1088	0.2 / 20
τ_2	0.9280	4.9280	0.2 / 20

- $\beta_0 = r_\infty$ (the long rate)

Table 6a

	Min	Max	Mean	Median	Std. Dev.
1 - 2	6.60	6.73	6.68	6.68	0.02
1 - 3	6.52	6.73	6.63	6.63	0.04
1 - 4	6.64	6.71	6.67	6.67	0.01
1 - 5	6.62	6.76	6.68	6.67	0.02
1 - 6	6.51	6.73	6.64	6.63	0.05
2 - 3	6.56	6.73	6.63	6.62	0.04
2 - 4	6.62	6.73	6.66	6.66	0.02
2 - 5	6.57	6.73	6.67	6.67	0.04
2 - 6	6.48	6.74	6.62	6.61	0.06
3 - 4	6.51	6.83	6.66	6.65	0.07
3 - 5	6.57	6.72	6.64	6.64	0.04
3 - 6	6.49	6.81	6.63	6.63	0.06
4 - 5	6.49	6.83	6.66	6.66	0.04
4 - 6	6.49	6.94	6.65	6.65	0.06
5 - 6	6.50	6.80	6.65	6.66	0.08

Figure 4a
Variation in $\beta_0 = r_\infty$ (SV)



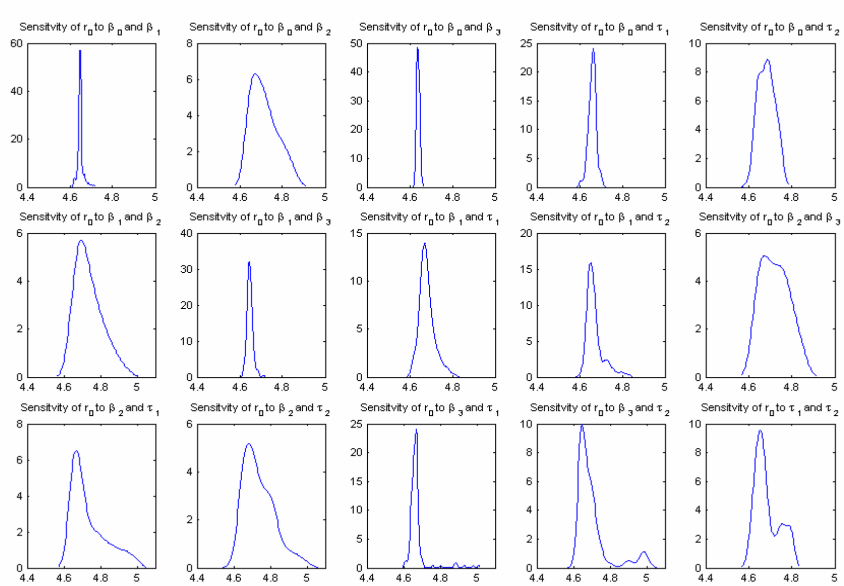
➤ $\beta_0 + \beta_1 = r_0$ (the short rate)

Table 6b

	Min	Max	Mean	Median	Std. Dev.
1 - 2	4.62	4.71	4.65	4.65	0.01
1 - 3	4.63	4.86	4.72	4.70	0.06
1 - 4	4.62	4.66	4.64	4.64	0.01
1 - 5	4.60	4.71	4.66	4.66	0.02
1 - 6	4.60	4.75	4.68	4.68	0.04
2 - 3	4.62	4.96	4.74	4.72	0.07
2 - 4	4.62	4.71	4.65	4.65	0.01
2 - 5	4.61	4.81	4.68	4.67	0.04
2 - 6	4.60	4.83	4.67	4.66	0.04
3 - 4	4.62	4.86	4.72	4.72	0.06
3 - 5	4.62	5.00	4.74	4.70	0.10
3 - 6	4.60	5.01	4.74	4.72	0.08
4 - 5	4.61	5.01	4.67	4.66	0.06
4 - 6	4.60	5.02	4.70	4.67	0.09
5 - 6	4.60	4.81	4.68	4.66	0.06

Figure 4b

Variation in $\beta_0 + \beta_1 = r_0$ (SV)



The process of identifying the 'best' b_0^* was again the same described above. (1000 times) *fval* corresponding to the first choice b_0 (as in *Table 4a*) was 2.98. We assessed the sensitivity of $\tilde{b}$ to change in b_0: $fval \leq 2.98$.

Tables 6a and *6b* and density plots *Figures 4a* and *4b* respectively report the sensitivity of $\tilde{b}$ to b_0^*; again only for those variables/values which have a direct economic interpretation, i.e. the short and the long rate. Density plots and summary statistics for sensitivity of other parameters are available on request.

The variation in long-rate in SV is around 6.4 % to 7 % and that in the short-rate is 4.6% to 7 %. The standard deviation in variation is of the same order as that in the case of NS.

V. Conclusion

As it turns out, given the nature of the objective function in term structure estimation using parsimonious models considered here, the selection of the starting vector during optimization is a highly non-trivial matter. There exist regions in the shape of the objective function in all three cases where a slight change in (seemingly reasonable) initial vector takes one far from optimum.

Thus, for each day for which the term structure has to be estimated, a grid of reasonable starting values must be chosen and shape of *fval* assessed as b_0 is varied before narrowing down on a subset for which *fval* turns out to be minimum. It is for this subset of b_0 that $\tilde{b}$ must be analysed.

References

Bliss, R.R. (1997), "Testing Term Structure Estimation Methods", *Advances in Futures and Options Research,* 9, pp. 197-232

Bolder, D. and D. Streliski (1999), "Yield Curve Modeling at the Bank of Canada", *Technical Reports* 84, Bank of Canada

Brennan, M. J. and E. S. Schwartz (1979), "A Continuous Time Approach to Pricing Bonds", *Journal of Banking and Finance*, 3, pp. 133-135

Broyden, C.G. (1970), "The Convergence of a Class of Double-rank Minimization Algorithms", *Journal of the Institute of Mathematics and its Applications*, Vol. 6, pp 76-90

Coleman, T., L. Fisher and R. Ibbotson (1995), "Estimating the Term Structure of Interest Rates from Data that Include the Prices of Coupon Bonds," *Journal of Fixed Income*, 2, pp. 85-116

Coleman, T., Branch, M.A. and Grace, A. (1999), "Optimization Toolbox for use with MATLAB", *MATLAB Optimization Toolbox User's Guide Version 2*, Section 2-24, Math Works Inc

Cox, J. C., Ingersoll, J. E. and Ross, S. A. (1985), "A Theory of the Term Structure of Interest Rates", *Econometrica*, 53, pp. 385-407

Darbha, G. (2003), "Estimating the Benchmark Yield Curve – A New Approach Using Stochastic Frontier Functions", *Indian School of Business* Working Paper

Darbha, G., Roy, S. D. and V. Pawaskar (2003a), "Term Structure of Interest in India: Issues in Estimation and Pricing", *Indian Economic Review,* 38, 1, 1-19

Darbha, G., Roy, S. D. and V. Pawaskar (2003b), "Idiosyncratic Factors in Pricing Sovereign Bonds: An Analysis Based on GoI Bond Market", *Journal of Emerging Market Finance*

Fletcher, R. (1970), "A New Approach to Variable Metric Algorithms", *Computer Journal*, Vol. 13, pp 317-322, 1970

Goldfarb, D. (1970), "A Family of Variable Metric Updates Derived by Variational Means", *Mathematics of Computing*, Vol. 24, pp 23-26

Nelson, C. R. and A. F. Siegel (1987), “Parsimonious Modeling of Yield Curves”, *Journal of Business* 60, 4, pp. 473-489

Schittkowski, K. (1985) "NLQPL: A FORTRAN-Subroutine Solving Constrained Nonlinear Programming Problems", *Annals of Operations Research*, Vol. 5, pp 485-500

Shanno, D.F. (1970), "Conditioning of Quasi-Newton Methods for Function Minimization," *Mathematics of Computing*, Vol. 24, pp 647-656

Svensson, L. E.O (1994), “Estimating and Interpreting Forward Interest Rates: Sweden 1992-1994”, *NBER* Working Paper 4871

Subramanian, V. (2001), “Term Structure Estimation in Illiquid Markets”, *Journal of Fixed Income,* June, pp. 77-86

Vasicek, O. A. (1977), “An Equilibrium Characterization of the Term Structure”, *Journal of Financial Economics* 5, November, pp. 177 – 88