INDIAN INSTITUTE OF MANAGEMENT
AHMEDABAD • INDIA

On the Choice of Optimization Routine in Estimation of Parsimonious Term Structure Models: Results from the Svensson Model

Vineet Virmani

W.P. No. 2013-01-02
January 2013

INDIAN INSTITUTE OF MANAGEMENT
AHMEDABAD-380 015
INDIA

On the Choice of Optimization Routine in Estimation of Parsimonious Term Structure Models: Results from the Svensson Model

Vineet Virmani
Assistant Professor, Indian Institute of Management Ahmedabad
Email: vineetv@iimahd.ernet.in

Abstract

Objective function in term structure estimation with price errors is not only non-linear but also non-convex in parameters. This makes the final results sensitive to both the choice of the optimization routine as well as to the starting guess. This study looks at the impact of the choice of the optimization routine to final parameter estimates for the Svensson model. While results are expected to differ numerically across routines, what is of interest is the economic impact. Using eleven different routines over a range of starting parameter values, it is found while there is significant variation in the final objective function value across routines, for the most part, implied short-rates and long-rates have low standard deviation. Also, while grid-search seems unavoidable, popular quasi-Newton methods allowing for linear constraints seem quite adequate for the task at hand.

On the Choice of Optimization Routine in Estimation of Parsimonious Term Structure Models: Results from the Svensson Model

I. Introduction

Non-convex optimization problems are notoriously hard to estimate. Finding global optimum, even if it is known to exist, is rare and final estimates are highly sensitive to the 'initial guess' provided to the optimization routine. With multiple parameters, the problem is even harder because *a priori* one has very little idea of the shape of the objective function – which means there is little help in available in selecting the 'right' optimization routine for the problem.

Term structure estimation using price errors with parsimonious models like Nelson-Siegel (1987) and Svensson (1994) is an important example of non-convex optimization in finance. Despite their popularity (or perhaps because of it), however, very few studies have explicitly studied estimation issues involved (e.g. Gilli, Grobe and Schumann, 2010; GGS hereafter) and the impact of the choice of optimization routine.

This study takes a look at the sensitivity of the final parameter estimates to the choice of the optimization routine for the Svensson model. In particular, it is studied how the objective function value (*fval* hereafter) and the final parameter vector ($\tilde{b}$ hereafter) vary across different optimization routines as the initial parameter vector (b_0 hereafter) is changed. While results are expected to differ quantitatively across optimization routines, what is of interest rate is whether the impact is economically significant.

II. The Objective Function

2.1 The Svensson spot rate

Svensson is an extension of the Nelson-Siegel specification to allow for more flexibility in the shape of the term structure. The resulting specification for the spot rate is:

$$s(m;b) = \beta_0 + \beta_1 \left[\frac{1- e^{-m/\tau_1}}{m/\tau_1}\right] + \beta_2 \left[\frac{1- e^{-m/\tau_1}}{m/\tau_1} - e^{-m/\tau_1}\right] + \beta_3 \left[\frac{1- e^{-m/\tau_2}}{m/\tau_2} - e^{-m/\tau_2}\right] \quad [1]$$

In the above equation β_0 and $\beta_0 + \beta_1$ are implied long-rate and short-rate respectively, β_2 and β_3 describe the medium-term component, and along with τ_1 and τ_2 (which determine the location of 'humps') define the shape of the curve.

2.2 The Objective Function

The optimization problem involved is minimizing the weighted sum of square of price errors, i.e. the objective function is:

$$min \sum_{i=1}^{N} (\omega_i \varepsilon_i)^2$$

[2]

subject to non-negativity constraints imposed on the short-rate, the long-rate $(m \to \infty)$ and on the τ *s*, and $\tau_1 - \tau_2 > 0.2$ (to enable unique identification of τ s); $\varepsilon_i = P_i - \hat{P_i}$ is the difference between the model price and the traded price and the weights ω_i are given as:

$$\omega_i = \frac{\frac{1}{d_i}}{\sum_{j=1}^{N} \frac{1}{d_j}}$$

[3]

where d_i is the Macaulay duration of the i^{th} bond. This weighing scheme corrects for heteroskedasticity as well as proxies for minimizing yield errors.

Note that if spot rates can be bootstrapped or forward rates are known and one has some idea of location of the 'humps' in the yield curve (i.e. one can guesstimate τ s), estimation of the remaining parameters boils down to an exercise in simple least square regression. This means that even if one did not know the value of τ s, one could still potentially get away with a simpler exercise in least squares estimation over a set of τ s and then select the best among those. Unfortunately, illiquid debt markets like those in India neither allow for a sufficient number of bootstrapped spot/forward rates, nor can one take the liberty of assuming the location of the 'humps'.

III. Methodology

3.1 Estimation strategy and the choice of the optimization routine

Normally the estimation strategy for such problems would be to select a starting vector of parameters for the given day's data (this study uses the same dataset used in Virmani, 2012), find errors based on the starting vector and then use a suitable optimization routine for finding the optimum. One can then either write a computer program for optimization or select from among many software available. Given that the focus of this study is comparison of different optimization routines, choice of software is important.

This paper studies the optimization routines available in open source R software and contributed packages. The preference for R is driven by the availability of the source code (its open-source nature) and its popularity in the scientific community. The eleven routines used are given in Table 1 and cover the broad class of Newton-based, simplex-based and heuristics-based methods.

Table 1

Optimization Routines Used

Sr.	Routine	R Package	Developer
1	Augmented Lagrangian (AugLag; Sequential Quadratic Programming; constrained)	Rsolnp	Stefan Theussl, Institute for Statistics and Mathematics, University of Vienna
2	Broyden-Fletcher-Goldfarb-Shanno (BFGS; quasi-Newton; unconstrained)	optim	Part of R base (R Development)
3	Box-constrained BFGS (L-BFGS-B; quasi-Newton; bound constraints)	optim	- do -
4	BFGS with Constraints (BFGSConst.; quasi-Newton; constrained)	coptim	- do -
5	Conjugate Gradient (CG; unconstrained)	optim	- do -
6	Differential Evolution (DE; Heuristics-based; bound constraints)	RcppDE	Katharine Mullen, National Institute of Standards and Technology, US Department of Commerce.

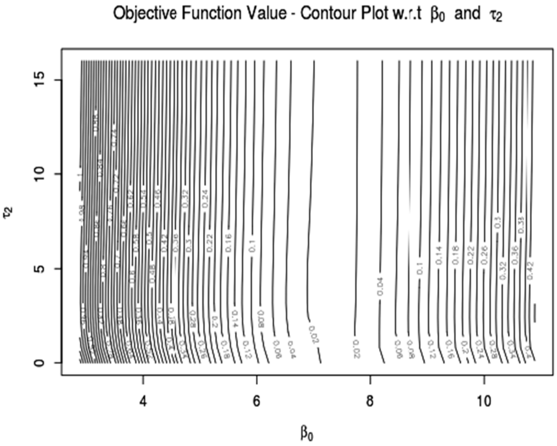
7	Nelder Mead (NM; Simplex; unconstrained)	optim	Part of R base (R Development)
8	Nelder Mead with Constraints (NMConst; Simplex; constrained)	coptim	- do -
9	Non-linear Minimization (NLM; Newton; unconstrained)	nlm	Part of R stats (R Development)
10	Trust Region (TR; Line-search; constrained)	minqa	Katharine Mullen, National Institute of Standards and Technology, US Dept. of Commerce.
11	Unconstrained Non-linear Optimization (UCMINF; quasi-Newton; unconstrained)	ucminf	R port to FORTRAN routine MINF by Hans Brunn Nielsen, Technical University of Denmark

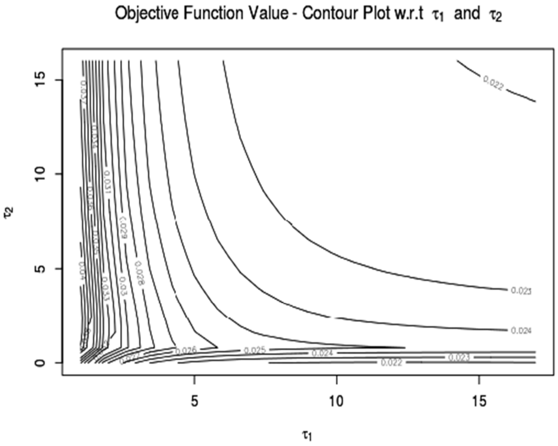
While comparing the implementation details of the selected routines in R is beyond the scope of this short note (R package repository describes them in some detail), it may be mentioned that most of the routines selected in this study are ported versions of their C++ or FORTRAN counterparts which have been in use for decades. Also, only those routines/packages have been used which have been developed by the R development team or at a research institution.

3.2 The Methodology: Sensitivity of fval and $\tilde{b}$ to b_0 for a given optimization routine

Since for non-convex optimization problems final results are sensitive to the choice of starting guess, for each optimization routine the final results are evaluated over a set of initial guesses. Starting with a reasonable choice of the starting guess b_0 (such that the resulting short and long rates are meaningful) a grid search is done around it to see how *fval* varies with local changes in b_0.

Figures 1 shows two contour plots (out of a total of ${}^6C_2 = 15$) for the shape of the objective function. The two are indicative of the flat portions in the shape of the objective function, multiple local minima and high sensitivity of *fval* to initial guess – typical of non-convex optimization.





IV. Results and Discussion

Given the variation of *fval* for changing first stage starting guess, the second stage initial parameter vector was taken to be:

$b_0^* = \{\beta_0 = 7.9,\ \beta_1 = -1.0,\ \beta_2 = -0.5,\ \beta_3 = 0.5,\ \tau_1 = 5.0,\ \tau_2 = 4.0\}$. The range of starting values for the first four parameters then was $\beta_0 \in [2.9, 10.9]$, $\beta_1 \in [-5.0, 3.0]$, $\beta_2 \in [-6.5, 1.5]$, $\beta_3 \in [-5.5, 2.5]$ with a step-size for each of *0.4*, and range for the next two was $\tau_1 \in [1.0, 17.0]$ with a step-size of *1.0* and $\tau_2 \in [0.00+, 16.0]$ with a step-size of *0.8*.

For DE the parameters to specify include the “initial population” (*NP*), “mutation” (*F*), “crossover probability” (*CR*) and “strategy”. NP is selected to be an array of size 60 (recommended to be ten times the number of parameters) from the subset of the initial parameter vector space above. Recommended “strategy” for such problems, and default in package *RcppDE*, is local-to-best/1/bin (for more on choice of DE parameters see Storn and Price, 1997).

Selecting *F* and CR is a matter of some trial and error, and they are varied with a step-size of *0.1* between 0.1 and 1.9 for *F*, and 0.1 and 0.9 for *CR*. Other parameter settings across optimization routines are available from author on request.

Rather than estimating parameters for all possible starting vectors (which would mean 21^6 > 80 million iterations), parameters are varied one at a time keeping others fixed at their value in b_0^*. This results in 21 * 6 = 126 different starting guesses for each optimization routine other than DE and 19 * 9 = 171 for DE over the range of *F* and *CR*.

Table 2 presents results on the variation in short-rate, long-rate, final objective function value and number of outliers (optimization routine failing to converge). Entries in the table represent the range with the associated standard deviation in parentheses below.

Table 2

Range of Short-rate, Long-rate and Objective Function Value across Optimization Routines

Sr.	Routine	Range of Short-Rate ($\beta_0 + \beta_1$)	Range of Long-Rate (β_0)	Range of Objective Function Value (*fval* * 10^4)	Number of Outliers
1	AugLag	4.84 – 4.98 (0.02)	6.61 – 9.19 (0.45)	2.63 – 3.09 (0.09)	0
2	BFGS	4.92 – 5.10 (0.02)	6.74 – 8.04 (0.24)	2.12 – 2.65 (0.06)	0
3	L-BFGS-B	4.89 – 5.10 (0.02)	6.63 – 11.51 (0.48)	1.93 – 2.69 (0.08)	1
4	BFGSConst.	4.69 – 5.09 (0.03)	3.89 – 8.26 (0.82)	2.14 – 2.85 (0.08)	0
5	CG	4.86 – 5.01 (0.03)	6.65 – 8.89 (0.42)	2.56 – 4.17 (0.32)	0
6	DE	4.80 – 6.20 (0.28)	0.05 – 15.00 (2.78)	1.83 – 106.30 (17.25)	NA
7	NM	4.90 – 5.13 (0.05)	3.16 – 12.90 (1.05)	2.19 – 2.67 (0.10)	9
8	NMConst.	4.91 – 5.09 (0.03)	4.57 – 8.90 (0.04)	2.12 – 2.79 (0.09)	0
9	NLM	4.92 – 4.93	6.80 – 7.21	2.63 – 2.65	29

		(0.001)	(0.07)	(0.002)	
10	TR	4.85 – 5.97 (0.11)	4.72 – 8.65 (0.66)	2.09 – 4.47 (0.21)	0
11	UCM	4.92 – 4.94 (0.002)	6.74 – 7.84 (0.19)	2.62 – 2.65 (0.01)	2

It is not unexpected that results would vary across different optimization routines. However, for the range of reasonable starting guess studied in this note, Table 2 suggests that other than DE, NLM and NM (all unconstrained procedures), there is clearly little to choose among the rest.

The estimate of the short-rate ($\beta_0 + \beta_1$) and long-rate across the remaining eight routines (1008 different estimations) is between 4.69 – 5.97 % and 4.57 – 9.19% respectively (ignoring two extreme values of long-rate of 3.89% and 11.51%). While long-rate is more variable than the short-rate, range in both is not alarming. For a given a routine, over the set of reasonable starting guesses, results are even more encouraging in that the range of short-rate and long-rate is very reasonable. This suggests that while grid search over a range of initial guesses is unavoidable, both quasi-Newton and Trust-region based methods work well.

It must be added, however, that for specific values *F* and *CR*, DE results in the lowest value of *fval*, which is what GGS base their recommendation of DE on. *A priori*, however, case for DE is slightly weak because while it may give the a better optimum (lower *fval*), it does not promise an economically meaningful result as it does not naturally allow for constraints.

V. Conclusion

Estimation of parameters based on non-convex optimization is susceptible to both the choice of the initial guess as well as to the choice of the optimization routine. This note has looked at the sensitivity of the final parameter to the choice of the optimization routine for the Svensson model. While final results were expected to vary numerically, economically results are not very variable across popular optimization routines. Given that parsimonious term structure models of the Svensson-type are more popular with the central bankers for studying monetary policy, this is good news – as robustness is more appealing than goodness-of-fit for such purposes.

As an aside, while an R-user is clearly spoilt for choice when it comes to implementing optimization in practice, better documentation and a suite of test dataset and programs for validation would definitely go a long way in helping select the 'right' routine.

References

Gilli, M., Grobe, S. and Schumann, E., 2010. Calibrating the Nelson-Siegel-Svensson model. *COMISEF Working Paper WPS-031*.

Nelson, C. R. and Siegel, A. F., 1987. Parsimonious modeling of yield curves. *Journal of Business*, 60, 4, pp. 473-489

Storn, R. and Price, K., 1997, Differential evolution – A simple and efficient heuristic for global optimization over continuous space. *Journal of Optimization*, 11, pp. 341-359.

Svensson, L. O., 1994. Estimating and interpreting forward interest rates: Sweden 1992-1994. *NBER Working Paper 4871*.

Virmani, V., 2012. On estimability of parsimonious term structure models: an experiment with the Nelson-Siegel specification. *Applied Economics Letters*, 19, 17, 1 November, pp. 1703-1706