A Locally Parametric Nonparametric Estimation of the Short Term Interest Rate Model

By Pouyan Mashayekh Ahangarani[1]

University of Southern California, Economics Department

January 2004

ABSTRACT

The nonparametric approach will be shown as a method for estimation of the continuous time interest rate model. Locally Parametric Nonparametric estimation will be used which is a method that has less bias relative to other standard nonparametric methods. The conditions will be derived in order to estimate the diffusion function nonparametrically. At the end, the results will be reported.

[1] I am grateful to the referees of my 2nd year paper at the Department of Economics at USC for giving me the hints to the recent papers in this area. Also I am grateful to Dr. Levon Goukosian for giving me the data that I used in this study.

Introduction

The short term interest rate is one of the most fundamental and important prices determined in financial markets. Most important than interest rate itself is its contingent claims on interest rate like caps, Floors, Swaps which are traded in large volumes in stock markets. Pricing these instruments has been a big challenge for people who are working in this industry.

The theory of derivative security pricing relies essentially on continuous-time arbitrage arguments since the pioneering Black and Sholes (1973) paper. As further demonstrated by the work of Merton (1973), pricing derivatives in the theoretical finance literature is generally much more tractable in a continuous-time framework than through binomial or other discrete approximations. This paper develops tools to estimate the actual model used in theoretical construction (which is continuous-time) by using the data available in discretized monthly Eurodollar interest rates.

The underlying process of interest is a diffusion represented by Ito stochastic differential equation:

$$dr_t = \mu(r_t)dt + \sigma(r_t)dw_t \qquad (1.1)$$

Where $\{w_t , t >= 0\}$, is a standard Brownian motion. The functions $\mu(.)$ and $\sigma(.)$ are respectively the drift (or instantaneous mean) and the diffusion (or instantaneous variance) functions of the process. It has long been recognized in finance literature that one of the most important features of (1.1) for derivative security pricing is the specification of $\sigma(.)$. Since, in Black-Scholes world, the existence of a risk free asset will eliminate the importance of deterministic part of the diffusion model (See Bjork 1998).

As a consequence every model is tried to specify $\sigma(.)$ correctly. To price interest rate derivatives, Vasicek (1977) specifies that the instantaneous volatility of the spot rate process is constant. The Cox-Ingersoll-Ross (1985) model of the term structure assumes that the instantaneous variance is a linear function of the level of the spot rate r, so the standard deviation is square root. There are many other models in the literature for the instantaneous variance of the short term interest rates (see chan et al 1992).

A commonly used method to estimate (1.1) consists in first parameterizing $\mu(.)$ and $\sigma(.)$, then discretizing the model in order to estimate the parameters. Because of specification error, discretizing the model will bring a bias to the estimation (Gourieroux 2001). Chan et al (1992) have used Generalized Method of Moments (GMM) to estimate the parameters and have compared different models. One method of eliminating the discretized bias is using Simulation based method techniques introduced by Gourieroux (1993). Mashayekh-Ahangarani(2003) has used this method for finding the consistent estimates of the parameters and has used the Voung (1989) test to choose the best model. Because of the importance of the diffusion functional form in pricing the derivatives and the nonobservability of instantaneous volatility of financial series, Ait-Sahalia (1996) has proposed the nonparametric estimation of diffusion functions. Bandi and Philips (2000) have proposed a fully Nonparametric method. Hjort and Jones(1996) have introduced a Locally Parametric Nonparametric (LPN) approach which has less bias in estimating the distribution function of a random variable relative to the standard Kernel density estimation (which is used in Ait-Sahalia (1996) paper). The objective of this paper is to implement the Locally Parametric Nonparametric estimator for estimation of the diffusion model.

This paper is organized as follows. Section 2 identifies nonparametrically the diffusion given a restriction in the drift function. Section 3 will discuss about the benefits of using LPNE relative to the standard Kernel estimation. Section 4 introduces the data and in section 5 the empirical results will be presented.

1- Nonparametric Identification of the Diffusion Function

Because of the Gaussian distribution of the Brownian increments, the continuous time process r in (1.1), in general, is not normally distributed. But, it turns out that under regularity conditions, an analogues property will hold for stochastic differential equations. The distributions of the process (marginal or transitional densities) are entirely characterized by the first two moments of the process, which are here the drift and diffusion functions. The joint parameterizations of μ(.) and σ(.) adopted in the literature imply specific forms for the marginal transitional densities of the process. For example, an Ornestien-Uhlenbeck process $dr_t = \beta(\alpha - r_t)dt + \gamma dw_t$ generates Gaussian transitional and marginal densities. The stochastic differential equation $dr_t = \beta(\alpha - r_t)dt + \gamma r_t^{0.5} dw_t$ yields a noncentral chi-squared transitional distribution, while the marginal density is a Gamma distribution.

Ait-Sahalia (1996) has used the equivalence between (μ ,σ) and transition density, to derive (μ ,σ) of the diffusion model. The densities of the processes can be straightforwardly estimated from the data on the short term rate. Therefore, instead of specifying parametrically both the drift and the diffusion functions and then accepting whatever marginal and transitional densities are implied by these choices, we start with nonparametric estimates of the densities. In this paper, we will assume the drift $\mu(.,\theta)$ depends on an unknown parameter vector θ, while the diffusion function σ(.) depends on an unknown function. The available data consists of realizations of the process sampled at equally spaced discrete dates 1,…, n. Let π (.) be the marginal density of the spot rate and $P(\Delta, r_{t+\Delta} | r_t)$ be the transition density function between two successive observations. From the Kolmogrov forward equations (Karlin and Taylor (p219, 1981)) we can obtain the following ordinary differential equation:

$$(d^2/dr^2)[\sigma^2(r)\,\pi(r)] = 2\,(d/dr)[\mu(r,\theta)\pi(r)] \qquad (2.1)$$

Where $r = r_{t+\Delta}$

Integrating (2.1) twice yields the diffusion function

$$\sigma^2(r) = \frac{2}{\pi(r)}\int_0^r \mu(u,\theta)\pi(u)du + C_1 r + C_2 \qquad (2.2)$$

The above expression needs to some initial value in order to be solved otherwise it just shows the diffusion difference between different interest rate values. Ait-Sahalia (1996) has assumed $C_1 = C_2 = 0$. But in this paper $C_1 = 0$, $C_2 = 100$ has been implemented.

Thus, if the drift parameters have been identified, the diffusion function can be found from marginal distribution of π (.). An identifying restriction on the drift is the linear mean –reverting specification $\mu(r_t,\theta) = \beta(\alpha - r_t)$.

By Dynkin formula (Karlyn and Taylor (1981, p310)) the following formula will be derived:

$$E(r_{t+\Delta} \mid r_t) = \alpha + e^{-\beta\Delta}(r_t - \alpha)$$

With using OLS we can easily find the drift parameters.

2- Locally Parametric Nonparametric approach

Let $X_1,\dots,X_n$ be i.i.d with density function f, The traditional Kernel estimator of f is $\tilde{f}(x) = n^{-1}\sum_{i=1}^{n} K_h(x_i - x)$ where $K_h(z) = h^{-1}K(h^{-1}z)$ and K(.) is some chosen unimodel density, symmetric about zero. The basic properties of $\tilde{f}$ are well known and under smoothness assumptions these include

$$E\tilde{f}(x) = f(x) + 0.5\,\sigma^2 h^2 f''(x) + O(h^4) \qquad (3.1)$$

$$Var\,\tilde{f}(x) = R(k)(nh)^{-1} f(x) - n^{-1}[f(x)]^2 + O(h/n) \qquad (3.2)$$

Where $\sigma^2 = \int z^2 k(z)dz$ and $R(k) = \int [k(z)]^2 dz$, See scott (1992, chapter6).

Hjort and Jones (1996) have proposed and investigated a new class of semiparametric competitors which have precision comparable to that of $\tilde{f}$ but sometimes better. For any given parametric family, $f(.,\theta) = f(.,\theta_1,\ldots,\theta_p)$ and for each given x, they have presented ways of estimating the locally best approximation to f and then use

$$\hat{f}(x) = f(x, \hat{\theta}_1(x),\ldots, \hat{\theta}_p(x)) \quad (3.3)$$

Thus the estimated density of x employs a parametric value which depends on x and whose choice is to be tailored to good estimation at x. In other words, the method amounts to a version of nonparametric parameter smoothing within the given parametric class.

We approximate the unknown distribution f of x by the pseudo family $F=\{f(x,\theta),\theta$ varying $\}$ on an interval A=[c-h, c+h]. The resulting approximation of the θ parameter is :

$$\tilde{\theta}_{c,h} = \quad (3.4)$$

$$\text{ArgMax } E[(1/h)K((x-c)/h)\log f(x;\theta)]-E[(1/h)K((x-c)/h)]\log\int (1/h)K((x-c)/h)f(x;\theta)dx$$

Which corresponds to the optimization of the Kullback-Leibler criterion. The above definition is valid for any Kernel K and x. The Local Parameter Function (LPF) is defined as the limit of $\tilde{\theta}_{c,h}$ when h tends to zero.

The Locally Parametric Nonparametric estimation shown in (3.3) has these characteristics:

$$E\hat{f}(x) = f(x) + 0.5\,\sigma^2h^2b(x) + O(h^4+(nh)^{-1})$$

$$\text{Var}\,\hat{f}(x) = R(k)\,(nh)^{-1}f(x) - n^{-1}[f(x)]^2 + O(h/n)$$

Which is just like (3.2) and (3.3) but with a bias factor function b(x) related to but different from $f''(x)$, with characteristics inherited from the parametric class. To the order of approximation used, the variance is simply the same, regardless of parametric family, The statistical advantage will be that for many “f"s , typically those lying in a broad nonparametric neighborhood of the parametric $f(.,\theta)$, b(x) will be smaller in size than $f''(x)$ for most x (Hjort and Jones (1996)).

In this paper the standard normal assumption for Kernel function and distribution function f, has been used, If we plug the normal distribution function in (3.4) we will come up with these conditions for unknown parameters μ_c and σ_c (see Appendix A).

$$\sum_{t=1}^{T} (1/h)K((x_t - c)/h)\, [\sigma_c^{-2}(x_t - \mu_c) + (\sigma_c^{2} + h^{2})^{-1} (\mu_c - c)] = 0 \qquad (3.5)$$

$$\sum_{t=1}^{T} (1/h)K((y_t - c)/h)[-(\mu_c - c)^{2} + (\sigma_c^{2} + h^{2})^{-1} + \sigma_c^{-4} (x_t - \mu)^{2} - \sigma_c^{-2}] = 0 \qquad (3.6)$$

3- DATA

The monthly Eurodollar interest rates in the period 1970-2002 are used in this study. Eurodollar deposits are dollar deposits held in banks outside United States and therefore exempt from Federal Reserve regulations. Appendix B shows the pattern of these market rates over the 32 years period which is 377 observations. Each observation is the average of the interest rate in each month.

4- Empirical results

In this section, the empirical results will be presented. The first step consists of estimating the drift parameters α and β by OLS. The results are reported in Appendix C.

In the next step the Parametric Nonparametric Estimator of marginal density has been estimated. The nonlinear system of equations defined in (3.5) and (3.6) has been solved for a grid of data points of available interest rate values. The MATLAB programs in Appendix D are written for the calculations of this part. The histogram of interest values is shown in Appendix D (diagram D.1). The graph of the obtained marginal density has been shown in diagram D.2. The distribution of diffusion function is constructed by plugging the obtained marginal density in (2.2). The trapezoid method has been used for computation of the integral. The diffusion function distribution has been reported in diagram D.3.

5- Conclusion

In this paper, I tried to find a model for short term Eurodollar interest rate. I used the method developed by Ait_Sahalia (1996) which assumes a parametric model for drift and tries to find a nonparametric estimation of diffusion function. I used the idea of Locally Parametric Nonparametric estimation for finding the diffusion function. The theory proposed by Hjort and Jones (1996) shows that this method has less bias relative to the standard Kernel estimators. By implementing the normal distribution assumption for functions f and K, the necessary conditions derived from the abstract theory. Then the continuous time model for Eurodollars has been found. The results show the increasing diffusion function for interest rates up to 8 and decreasing for upper values.

Much works remain to be done. The theory needs to more development for finding the functional form out of the Kolmogrov forward equations since the diffusion function relies on initial conditions. Also we can change the normal distribution assumption and use other distribution for deriving conditions for estimating the parameters. Also more can be done for finding the optimal bandwidth h. At the end, we can extend this paper for pricing the contingent claims of interest rate which is the ultimate goal of this research.

Appendix A – Deriving the locally parameters of Nonparametric estimation

In this section the procedure that the conditions (3.5) and (3.6) are derived from (3.4) will be shown. The assumptions we have used in this papers are:

$f(x; \theta) = (2\pi\sigma^2)^{-1} \exp[-½\sigma^{-2}(x-\mu)^2]$

$K((x-c)/h) = (2\pi)^{-1} \exp[-½h^{-2}(x-c)^2]$

Now if we use the above assumptions, we will have:

$\int (1/h)K((x-c)/h)f(x; \theta)dx =$

$$(1/(2\pi\sigma h)) \int \exp[-½(h^{-2}(x-c)^2 + \sigma^{-2}(x-c)^2)]dx \qquad (A.1)$$

Now in order to simplify (A.1), the expression in the integral should be simplified. So, we have:

$h^{-2}(x-c)^2 + \sigma^{-2}(x-\mu)^2 = \sigma^{-2} h^{-2}[h^2(x-c)^2 + \sigma^2(x-\mu)^2] =$

$= (h^{-2}+\sigma^{-2})x^2 - 2(ch^{-2} + \mu\sigma^{-2})x + (c^2 h^{-2} + \mu^{-2}\sigma^{-2}) =$

$= (h^{-2}+\sigma^{-2})[x^2 - 2(ch^{-2} + \mu\sigma^{-2})(h^{-2}+\sigma^{-2})^{-1} + (c^2 h^{-2} + \mu^{-2}\sigma^{-2})(h^{-2}+\sigma^{-2})^{-1}] =$

$= (h^{-2}+\sigma^{-2})[x^2 - 2(c\sigma^2 + \mu h^2)(h^2+\sigma^2)^{-1} + (c^2\sigma^2 + \mu^2 h^2)(h^2+\sigma^2)^{-1}] =$

$= (h^{-2}+\sigma^{-2})[(x-(c\sigma^2 + \mu h^2)(h^2+\sigma^2)^{-1})^2 + (c^2\sigma^2 + \mu^2 h^2)(h^2+\sigma^2)^{-1} - (c\sigma^2 + \mu h^2)^2(h^2+\sigma^2)^{-2}]$

After some algebraic simplification we will have:

$$= \sigma^{-2} h^{-2}(h^{-2}+\sigma^{-2})(x-(c\sigma^2 + \mu h^2)(h^2+\sigma^2)^{-1})^2 + (\mu-c)^2(h^2+\sigma^2)^{-1} \qquad (A.2)$$

$\int (1/h)K((x-c)/h)f(x; \theta)dx =$

$(2\pi(h^2+\sigma^2))^{-½} \exp(-½(\mu-c)^2(h^2+\sigma^2)^{-1}) \left\{ (h^2+\sigma^2)^{½}(2\pi h^2\sigma^2)^{-½} \int \exp[-½(h^2+\sigma^2)h^{-2}\sigma^{-2}(x-(c\sigma^2 + \mu h^2)(h^2+\sigma^2)^{-1})^2 dx \right\}$

Since the expression in the bracket is equal to 1, we will have:

$$\log \int (1/h)K((x-c)/h)f(x; \theta)dx = -.5(h^2+\sigma^2)^{-1}(\mu-c)^2 - .5 \log[2\pi(h^2+\sigma^2)] \qquad (A.3)$$

Also we have :

$$\log f(x; \theta) = -.5\,\sigma^{-2}(x-\mu)^2 - .5 \log[2\pi\sigma^2] \qquad (A.4)$$

Now if we plug (A.3) and (A.4) in (3.4) we will have :

$$(\theta_c , \mu_c) = \text{ArgMax} \; -.5 \left\{ \sum_{t=1}^{T} (1/h)K((x_t - c)/h) \, [\sigma^{-2} (x- \mu)^2 + \log(2\pi\sigma^2)] - \sum_{t=1}^{T} (1/h)K((x_t - c)/h) \, [(h^2+\sigma^2)^{-1} (\mu-c)^2 + \log (2\pi) + \log(h^2+\sigma^2)] \right\} =$$

$$= \text{ArgMax} \sum_{t=1}^{T} (1/h)K((x_t - c)/h) \, [(h^2+\sigma^2)^{-1} (\mu-c)^2 - \sigma^{-2} (x- \mu)^2 + \log (1+h^2\sigma^{-2})]$$

So if we take the derivative relative to μ we will have:

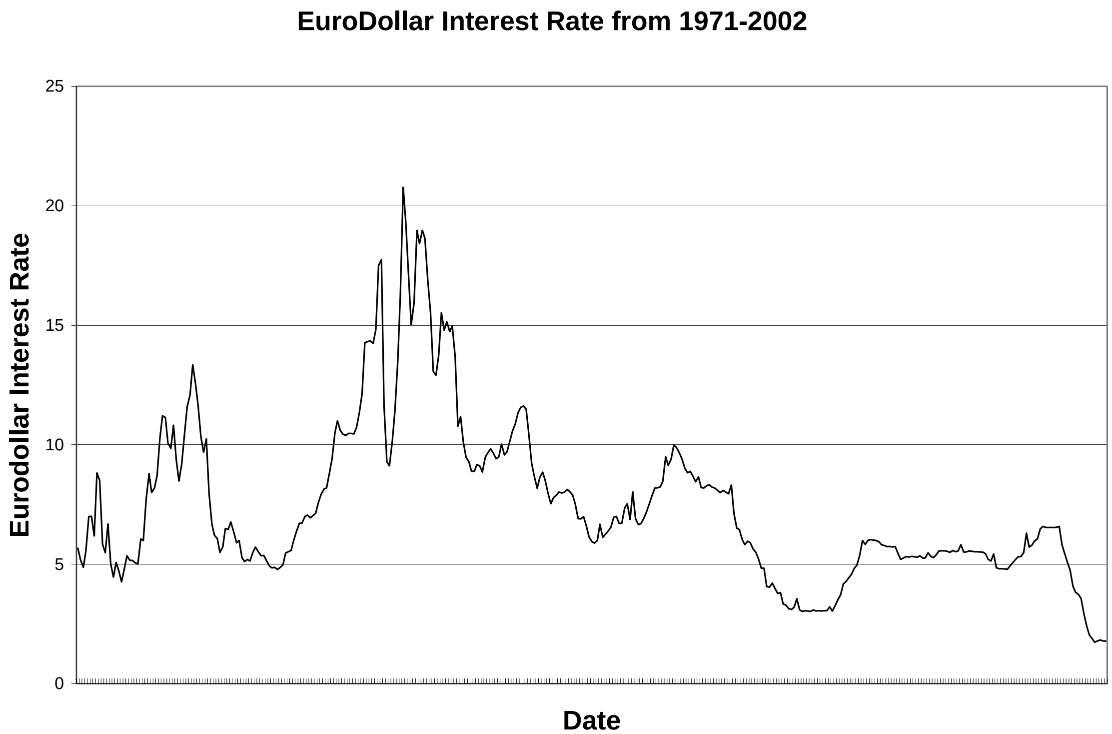
$$\frac{\partial}{\partial \mu}(.) = \sum_{t=1}^{T} (1/h)K((x_t - c)/h) \, [\sigma_c^{-2}(x_t - \mu_c) + (\sigma_c^{2} + h^2)^{-1} (\mu_c - c)] = 0$$

Which is conditioned defined as (3.5) and if we take the derivative relative to σ^2 we will have:

$$\frac{\partial}{\partial \sigma^2}(.) = \sum_{t=1}^{T} (1/h)K((y_t - c)/h)[-(\mu_c - c)^2 + (\sigma_c^{2} + h^2)^{-1} + \sigma_c^{-4} (x_t-\mu)^2 - \sigma_c^{-2}] = 0$$

Which is the expression shown as (3.6).

Appendix B-Data



Appendix C- Regression Output

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.973519052
R Square	0.947739344
Adjusted R Square	0.94759961
Standard Error	0.772250884
Observations	376

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	0.162279988	0.10	1.67	0.10
X Variable 1	0.976838804	0.01	82.36	0.00

delta	0.083333333
alfa	7.006546084
beta	0.281203581

Appendix D – Programs and their outputs

Mainprog.m

```
% This is the Main program
% for finding the diffusion function of a
% continous time interest rate model
clear;
clc;
i=0;
stp=0.1; % this is the step of interest rate grids
tic
for c=1.8:stp:18
    i=i+1
    wk1write('c:\user\c',c); % Reading the data
    yt=[1,2]; % Initial values
    [v,FVAL,EXITFLAG]=fsolve(@fprog,yt,optimset('MaxFunEvals',1220));
   mu(i)=v(1,1)
    sigma2(i)=(v(1,2))^2;
     f(i)=(1/(sqrt(sigma2(i)*2*pi)))*exp(-0.5*((c-mu(i))^2)/sigma2(i));
     x(i)=c;
 end
figure
plot(x,f) % plotting the distribution
alfa=7.006546 % Value found from OLS
beta=0.281203
miu=beta*(alfa-x);
integ=miu.*f;
intsum=0;
k=1;
% In this loop the integral will be calculated
for j=2:i
    k=k+1;
    intsum=intsum+0.5*stp*(integ(j)+integ(j-1));
    sig(k)=(2/f(k))*intsum+100;
end
figure
plot(x,sig) % the graph of the diffusion function
toc
```

fprog.m

% The condintions (3.4) and (3.5) are defined in this function

```
function x=fprog(teta)
mu=teta(1,1);
s=teta(1,2);
data=wk1read('c:\user\datahist');
nn=size(data);
n=nn(1,1);
c=wk1read('c:\user\c');
h=4;
for i=1:n
   f(i)=exp(-0.5*((data(i)-c)/h)^2)*((data(i)-c)/(s^2+h^2)+s^(-2)*(data(i)-mu));
   g(i)=exp(-0.5*((data(i)-c)/h)^2)*(-((mu-c)/(s^2+h^2))^2+1/(s^2+h^2)+s^(-4)*(data(i)-mu)^2-s^(-2));
  end
x(1,1)=sum(f);
x(2,1)=sum(g);
```

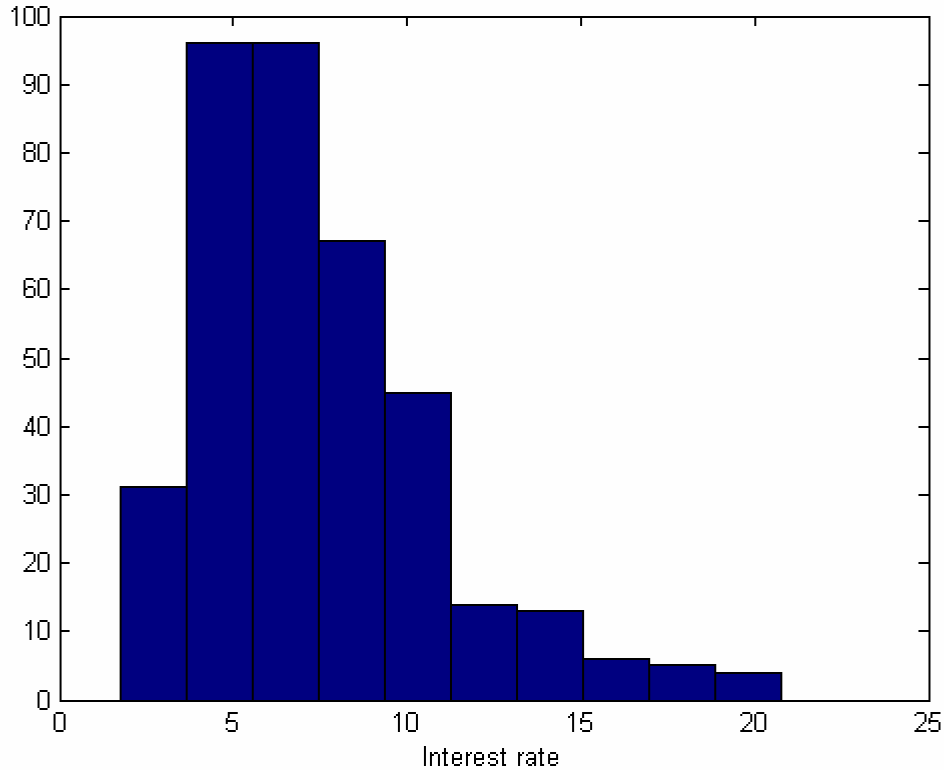


Diagram D.1 – The histogram of interest rate values.

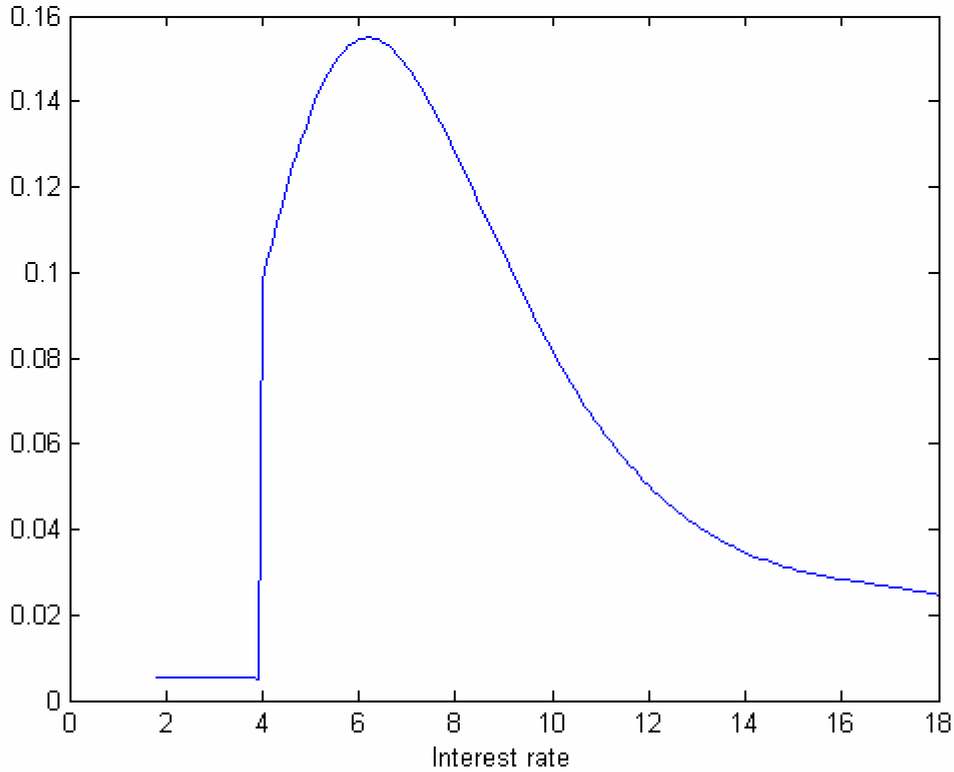


Diagram D.2– The pdf of interest rate values found from Locally Parametric Nonparametric method.

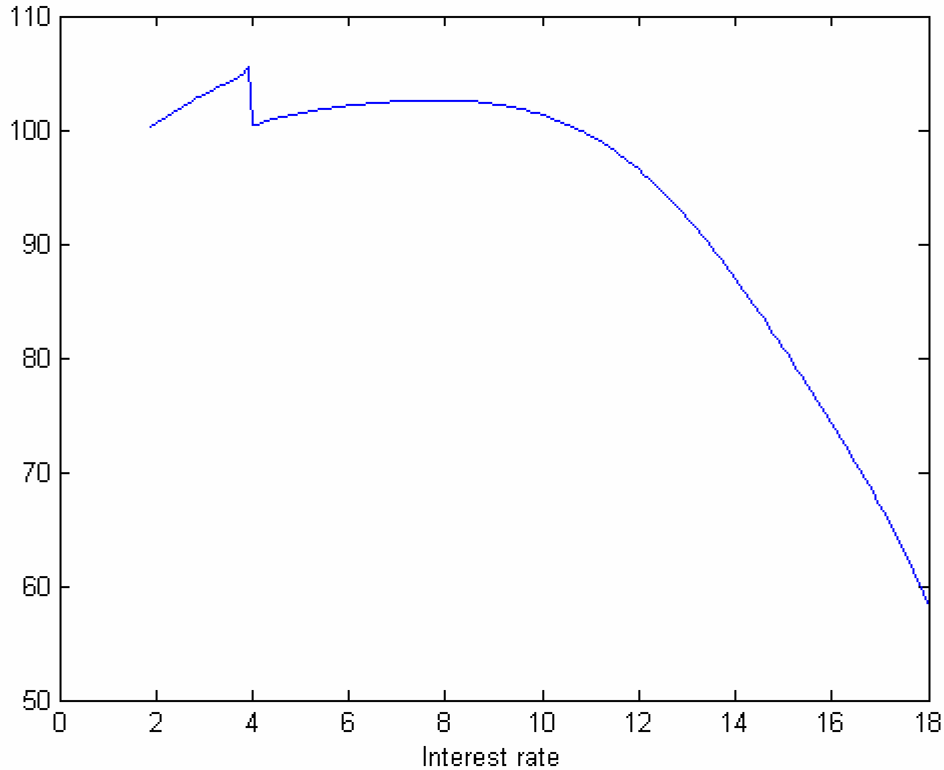


Diagram D.3– The diffusion function.

References

Ait-Sahalia, Yacine, 1996, Nonparametric pricing of interest rate derivative securities, *Econometrica* 64, 527-560.

Bjork, Tomas , 1998. Arbitrage Theory in Continuous Time. *Oxford University Press*.

Black, F. and M. Sholes, The pricing of options and corporate liabilities, *Journal of Political Economy*, 81 (May/June 1973) 637-59.

Chan K.C.; G. Andrew Karolyi; Francis A. Longstaff; Anthony B. Sanders, 1992 , An empirical comparison of alternative models of the short term interest rate, *Journal of Finance*, 1209-1227.

Cox, John C., Jonathan E. Ingersoll, and Stephan A. Ross,1985, A theory of the term structure of interest rates, *Econometrica* 53, 385-407.

Gourieroux, C., Joan Jasiak , 2001, Financial Econometrics. *Princeton University Press*.

Gourieroux, C., A. Monfort and E. Renault (1993), Indirect Inference, *Journal of Applied Econometrics*, 8, 85-118.

Hjort, N., M. C. Jones (1996), Locally Parametric Nonparametric Density Estimation, *The annals of Statistics*, Vol 24, No 4, 1619-1647.

Karlin, S., and H. M. Taylor (1981): *A Second Course in Stochastic Processes*, New York: Academic Press.

Mashayekh-Ahangarani, Pouyan , 2003 **,** An Empirical Estimation and Model Selection of the Short-Term Interest Rates, Working Paper at University of Southern California.

Merton, Robert C., 1973, Theory of rational option pricing, *Bell Journal of Economics and Management Science* 4, 141-183.

Scott, D. W. (1992). *Multivariate Density Estimation: Theory,Practice, and Visualization*. Wiley, New York.

Vasicek, Oldrich, 1977, An equilibrium characterization of the term structure, *Journal of Financial Economics* 5, 177-188.

Voung, H. Quang, 1989, Likelihood ratio tests for model selection and non-nested hypothesis, *Econometrica*, *57*, 307-333.