A nonparametric method of multi-step ahead forecasting in diffusion processes

Mariko Yamamura[a], Isao Shoji[b]

[a]School of Pharmacy, Kitasato University, Minato-ku, Tokyo, 108-8641, Japan.
[b]Graduate School of Systems and Information Engineering,
University of Tsukuba, Tsukuba Ibaraki, 305-8573, Japan.

May 9, 2008

Abstract

This paper provides a nonparametric model of multi-step ahead forecasting in diffusion processes. The model is constructed from the first order local polynomial model with the Gaussian kernel. The paper also provides some simulation studies to see how well the model can estimate drift coefficients of nonlinear stochastic differential equations and simultaneously to evaluate its performance of multi-step ahead forecasting comparing with the global linear model. The results show the model produces their reasonable estimates and better forecasting performance than the global linear model.
Keywords: Multi-step ahead forecasting; Nonparametric estimation; Local linear fitting; Diffusion process; Drift coefficients

1 Introduction

Recent rapid progresses in information technology enable us to get very high frequency data such as sampled at every minute or second. So, when modeling time series with such high frequencies, we often use continuous time stochastic processes formulated by stochastic differential equations. Actually, we can find wide varieties of their applications to real fields such as economics, finance, medicine, and electric engineering. Particularly, their theoretical and empirical applications to finance are quite extensive; see Campbell et al (1997) for example. Like regression models in discrete time, we have to specify two coefficients in a stochastic differential equation, called the drift coefficient μ and the diffusion coefficient σ as follows:

$$dX_t = \mu(X_t)dt + \sigma(X_t)dB_t, \tag{1}$$

where X_t stands for the process of interest and B_t stands for the standard Brownian motion. These coefficients are formulated by specific functions just like linear functions in regression models. The functions, however, are not always linear but supposedly nonlinear though we have no exact knowledge about their functional form beforehand. The uncertainty of their functional forms might cause a serious misspecification in the modeling; an incorrect model leads to entire lack of fit in data so that any statistical inference based on the incorrect model becomes quite suspicious. Instead of using parametric models, we might specify the drift and diffusion coefficients nonparametrically, which of course is free from the risk of misspecification.

Most nonparametric models are derived from the method of kernel estimation which has two approaches broadly. On one hand, target functions are computed from the transition probability density function which is estimated by kernel density estimators; for example, Aït-Sahalia (1996). On the other hand, target functions are directly estimated by the method of the kernel regression which locally approximates a target function by constant or linear functions; see Fan and Gijbels (1996) and Fan and Yao (2003) for example. The local constant kernel regression, so-called the Nadaraya-Watson kernel estimation method, was originally applied to estimation of diffusion coefficients by Florens-Zmirow (1993), and since then its related theoretical and empirical studies have been carried out until recently; for example Arapis and Gao (2006), Bnadi and Phillips (2003), Jiang and Knight (1997), Nicolau (2003), Stanton (1997).

Additionally the local linear kernel regression, or simply called the local linear model or fitting, is considered as an extended method of the local constant kernel regression because a family of linear functions includes that of constants. Fan and Yao (1998) proposed a method of bias reduction in the local liner fitting and Fan and Zhang (2003) fully discussed the method and its asymptotic properties particularly related to the test of nonlinearity in the function that is motivated by Stanton (1997).

All the studies cited above fully discuss both theoretical and empirical aspects of the nonparametric estimation, but don't mention much about forecasting of the nonparametric models, much less about a multi-step ahead forecasting. The multi-step ahead forecasting is useful particularly when considering a situation of handling high frequency data or data with very short sampling interval, since a single step forecasting based on such data leads to forecast simply a state of a process in the extremely near future. For example, suppose we exchange financial assets timely. We need to know their expected prices somewhat before the exchange, and so the single step forecasting might be meaningless if the execution time for the exchanging is more distant future than sampling interval. The aim of the paper is to propose a method for multi-step ahead forecasting based on the local linear fitting and evaluate its performance. A model used for the forecasting is estimated by the local linear regression without estimating the diffusion coefficient of a stochastic differential equation, and so the model basically has no risk of misspecification in the diffusion coefficient. Next, we evaluate the predictive power of the model by conducting numerical experiments using several stochastic differential equations with nonlinear drift coefficients.

2 Local polynomial fitting

The local polynomial fitting is a method to fit polynomial functions locally into an unknown function. Each polynomial function may be different from one after another for every local domain, or local neighborhood, specified by the so-called kernel functions. So, the method is not simply a global polynomial fitting. Here, we give a brief explanation about the local polynomial fitting though its detail discussion can be seen in Fan and Gijbels (1996) and Fan and Yao (2003) for example.

In the first place, suppose Y is expressed as a regression model of X defined by,

$$Y = f(X) + \varepsilon_t. \tag{2}$$

where $f(x)$ is assumed to be a smooth, possibly nonlinear, function but unknown. Using the Taylor expansion around x_0, $f(x)$ could be well approximated by a polynomial function of degree p as long as x lies in some local neighborhood of x_0. That is,

$$f(x) \approx f(x_0) + f'(x_0)(x - x_0) + \frac{f''(x_0)}{2!}(x - x_0)^2 + \cdots + \frac{f^{(p)}(x_0)}{p!}(x - x_0)^p. \qquad (3)$$

For simplicity, replacing every coefficient in the right-hand side with parameter β_j, we get,

$$f(x) \approx \sum_{j=0}^{p} \beta_j (x - x_0)^j. \qquad (4)$$

Here, if we specify all the parameters β_j, we can get local polynomial approximation of $f(x)$. Practically, we could use time series of X and Y and so we can relatively easily estimate these parameters by the least square estimation for example. But, we note the approximation should work properly only when the independent variable moves around a suitable local neighborhood of x_0. So, a kernel function is used for specifying the neighborhood. Let $\{x_t\}_{1 \le t \le n}$ and $\{y_t\}_{1 \le t \le n}$ be time series of X and Y, respectively. To estimate the parameters, we conduct the following weighted least square estimation instead of the conventional least square one:

$$\min_{\beta_j} \sum_{t=1}^{n} \left\{ y_t - \sum_{j=0}^{p} \beta_j (x_t - x_0)^j \right\}^2 K\left(\frac{x_t - x_0}{h} \right), \qquad (5)$$

where $K(\cdot)$ is a kernel function, and h is the size of local neighborhood, called bandwidth. A kernel function typically has such characteristics that values in the neighborhood of x_0 have contribute more to getting estimates of β_j while remote values from x_0 have less contribution in the estimation. In the paper we use the Gaussian kernel defined by

$$K\left(\frac{x_t - x_0}{h} \right) = \left(\frac{1}{\sqrt{2\pi h^2}} \right) \cdot \exp\left\{ -\frac{1}{2h^2}(x_t - x_0)^2 \right\}. \qquad (6)$$

The weighted least square estimates of β_j can be easily obtained from $\{x_t\}_{1 \le t \le n}$ and $\{y_t\}_{1 \le t \le n}$. Now, let Y and β be vectors and X be a matrix for simplicity, which are defined by

$$\boldsymbol{Y} = \begin{pmatrix} y_1 \\ \vdots \\ y_T \end{pmatrix}, \ \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_p \end{pmatrix}, \ \boldsymbol{X} = \begin{pmatrix} 1 & (x_1 - x_0) & \cdots & (x_1 - x_0)^p \\ \vdots & \vdots & & \vdots \\ 1 & (x_n - x_0) & \cdots & (x_n - x_0)^p \end{pmatrix}.$$

Then, the objective function of (5) can be rewritten as $(\boldsymbol{Y}-\boldsymbol{X\beta})'\boldsymbol{W}(\boldsymbol{Y}-\boldsymbol{X\beta})$, where $\boldsymbol{W}$ is an $n \times n$ diagonal matrix whose diagonal elements are $K((x_t - x_0)/h)$ $(1 \leq t \leq n)$. The weighted least square estimate of β is given as,

$$\begin{aligned} \hat{\beta} &= \underset{\beta}{\operatorname{argmin}} (\boldsymbol{Y}-\boldsymbol{X\beta})'\boldsymbol{W}(\boldsymbol{Y}-\boldsymbol{X\beta}) \\ &= (\boldsymbol{X'WX})^{-1}\boldsymbol{X'WY}. \end{aligned} \tag{7}$$

Using this estimate of β, we finaly get the estimate of $f(x)$ for every x, denoted by $\hat{f}(x)$:

$$\hat{f}(x) = \sum_{j=0}^{p} \hat{\beta}_j (x - x_0)^j. \tag{8}$$

Here, note that every different local neighborhood, which can be specified by its center and bandwidth (x_0, h), produces a different estimate of β, and that the Gaussian kernel implies any distant values from the center contribute to the estimation no matter how small the contributions may be; the Gaussian function never reaches zero. Hence, for any x, multiple approximations of $f(x)$ are obtained so that we need to unify them by evaluating contributions of those approximations. Since the contributions are measured by values of the Gaussian kernel at x, we claim $\hat{f}(x)$ is defined by the weighted average of $\hat{f}_i(x)$ for all local neighborhoods, $(x_{0,i}, h)_{1 \leq i \leq w}$:

$$\hat{f}(x) = \sum_{i=1}^{w} \left(\frac{k_i \cdot \hat{f}_i(x)}{\sum_{i=1}^{w} k_i} \right), \tag{9}$$

where,

$$k_i(x) = K\left(\frac{x - x_{0,i}}{h}\right).$$

Applying the method above to estimate the drift function of (1), a nonparametric estimate of $\mu(x)$ is given as,

$$\hat{\mu}(x) = \sum_{i=1}^{w} \left(\frac{k_i \cdot \hat{\mu}_i(x)}{\sum_{i=1}^{w} k_i} \right). \tag{10}$$

In the application, we need to know what kind of data should be used as the dependent variable Y. Note the estimate of β is given by (7). In the weighted least square, we take $(x_t - x_{t-1})/\Delta t$ as dependent variables and regress it with $\sum_{j=0}^{p} \beta_j (x_{t-1} - x_0)^j$. The theoretical background of this way of the regression is explained by Fang and Zhang (2003) for example, but intuitively it can be understood by the following discretization. Applying discretization by the Euler method to (1), we get,

$$X_t - X_{t-1} = \mu(X_{t-1})\Delta t + \sigma(X_{t-1})(B_t - B_{t-1}).$$

Dividing the both sides of the above equation by Δt, the basic formula of the regression is obtained. Since $\mu(x)$ is approximated by $\hat{\mu}(x)$ in the local linear fitting, the above way of the regression makes sense.

3 Multi-step forecasting by local linear fitting

An estimate of the drift function is given as local polynomial of degree p in (8) mathematically. It is often said that at most the first order approximation is enough for practical purposes, so we use the local linear fitting in this paper. The linearity makes simple forecasting formula of multi-step ahead as well as one-step ahead. In the first place, we focus on a local neighborhood with its center x_0 and its bandwidth h. In the local neighborhood, $\mu(x)$ is approximated by $\hat{\mu}(x) = \hat{\beta}_0 + \hat{\beta}_1(x - x_0)$, and so the target stochastic differential equation can be thought locally as a linear stochastic differential equation in drift:

$$dX_t = \hat{\mu}(X_t)dt + \sigma(X_t)dB_t. \tag{11}$$

The above stochastic differential equation can be solved as,

$$X_t = X_s + \frac{\hat{\mu}(X_s)}{\hat{\mu}'(X_s)}(\exp(\hat{\mu}'(X_s)(t-s)) - 1) + \int_s^t \exp(\hat{\mu}'(X_s)(t-u))\sigma(X_u)dB_u. \tag{12}$$

Here note the first derivative $\hat{\mu}'$ is fixed at $\hat{\beta}_1$ because of the linearity in drift. Using as forecast the conditional expectation of X_t at s, denoted by $E_s[X_t]$, we get,

$$E_s[X_t] = X_s + \frac{\hat{\mu}(X_s)}{\hat{\mu}'(X_s)}(\exp(\hat{\mu}'(X_s)(t-s)) - 1). \tag{13}$$

When the sampling interval is Δt, m-step ahead forecast is given by,

$$E_s[X_{s+m\Delta t}] = X_s + \frac{\hat{\mu}(X_s)}{\hat{\mu}'(X_s)}(\exp(\hat{\mu}'(X_s)m\Delta t) - 1). \tag{14}$$

Recall the forecast constructed as above is derived from only one contribution of all local neighborhoods, $(x_{0,i}, h)_{1\leq i\leq w}$. Hence, taking all the contributions into account, the forecast, denoted by $\hat{X}_{s+m\Delta t|s}$, should be given as,

$$\hat{X}_{s+m\Delta t|s} = \sum_{i=1}^{w}\left(\frac{k_i \cdot E_s[X_{s+m\Delta t}]_i}{\sum_{i=1}^{w} k_i}\right), \tag{15}$$

where $E_s[X_{s+m\Delta t}]_i$ stands for the conditional expectation derived from i-th local neighborhood $(x_{0,i}, h)$.

4 Numerical experiment

This section presents simulation studies to evaluate how accurately the local linear fitting can produce estimates of the drift function and what different performance the proposed forecast show in multi-step forecasting as compared with a misspecified model such as a global linear model.

4.1 Estimation of the drift function

We consider the following examples of stochastic differential equations with nonlinear drift coefficients:

$$\begin{aligned}
\textbf{model1}: \quad dX_t &= \frac{-X_t(X_t^2-5)}{1+0.1X_t^2}dt + \sigma dB_t \\
\textbf{model2}: \quad dX_t &= \frac{-X_t(X_t^2-1)(X_t^2-5)}{1+0.1X_t^4}dt + \sigma dB_t \\
\textbf{model3}: \quad dX_t &= (\exp(-2X_t)-1)dt + \sigma dB_t
\end{aligned}$$

where σ is a constant and fixed at 2 through the following experiments. All the models have strong solutions; model1 and model2 satisfy the Lipschitz condition in drift coefficients, and model3 is known to have an explicit solution as explained in Kloeden and Platen (1995). Hence the simulation study is meaningful from a mathematical viewpoint.

Using the local linear fitting, we try to estimate the drift functions of model1-model3 from ten thousand discrete observations with sampling interval $\Delta t = 1/100$. The sample path is generated by the Euler method with time interval of one-tenth of Δt. This means data for estimation are sampled out of every tenth data generated by the Euler method. For local linear fitting, we chose the bandwidth from visual inspection and set up ninety four local neighborhoods in total.

The results of estimation are displayed by figure 1. 1st-3rd graphs correspond to model1-model3, where the solid line stands for the true curve and quarry shaped signs for the estimates. The local linear fitting produced reasonable estimates basically though the difference is somewhat large in the end regions of observations; this is supposedly due to sparse data around the regions.

4.2 Performance comparison in multi-step forecasting

Using model1-model3 as in the previous section, we compare the forecasting performance with the global linear model. Since model1-model3 have nonlinear drift coefficients, the global linear model is of course a misspecified one. As far as the estimation of drift coefficients is concerned, it is not surprising that the local linear fitting as a nonparametric model shows much better performance than the global linear model. But, the performance of forecasting is different from that of estimation; intuitively the former concerns out-of sample properties, but the latter concerns in sample ones. It is sometimes reported that nonparametric models show good performance in sample, but not out-of sample. Hence, it is questionable whether the proposed forecasting model shows really better performance than the global linear model.

The global linear model as a counterpart is formulated as follows:

$$dX_t = (\alpha_0 + \alpha_1 X_t)dt + \sigma dB_t.$$

The conditional expectation at time s is given as,

$$\hat{X}^L_{s+m\Delta t|s} = E_s[X_{s+m\Delta t}] = X_s + (\alpha_0/\alpha_1 + X_s)(\exp(\alpha_1 m\Delta t) - 1). \tag{16}$$

And, it is known that $X_{s+m\Delta t}$ conditionally follows the normal distribution with mean $X^L_{s+m\Delta t|s}$ and variance $\sigma^2(\exp(2\alpha_1 m\Delta t) - 1)/(2\alpha_1)$. This fact is used for estimation of the parameters by the maximum likelihood method.

After estimating the coefficients of (15) and (16) from n-observations, we compute up to m-step ahead forecasts for the two models and then measure their forecasting errors at q successive discrete times. The performance comparison is conducted on the base of the following statistics:

$$V_{NONPAR} = \sum_{s=n}^{n+q-1} \sum_{j=1}^{m} (\hat{X}_{s+j\Delta t|s} - X_{s+j\Delta t})^2 \tag{17}$$

$$V_{LIN} = \sum_{s=n}^{n+q-1} \sum_{j=1}^{m} (\hat{X}^L_{s+j\Delta t|s} - X_{s+j\Delta t})^2 \tag{18}$$

In the experiment, we take 1,000 for both n and q and 1, 2, 5 and 10 for m. Since V_{NONPAR} and V_{LIN} are thought to be variable depending on sample paths, we compute them from one thousand different sample paths and then take their means and standard deviations.

The results of the experiments are presented by table 1, where numbers stand for means of V_{NONPAR} and V_{LIN} while numbers in parenthesis stand for their standard deviations. In model1 and model2 the difference between the proposed model and the global linear model are all great in terms of mean and standard deviation. Smaller standard deviations for the proposed model imply its forecasting is more stable than the global linear model. The stability of forecasting in the proposed model is also observed in model3 while the difference is not so large as model1 and model2. Over all, the proposed model outperforms the global linear model in multi-step ahead forecasting.

5 Concluding remarks

This paper proposed a nonparametric model of multi-step ahead forecasting in diffusion processes when data are observed at discrete times. The proposed model is based on the local linear fitting, or the first order local polynomial model, which is often used for nonparametric estimation in continuous time processes as well as discrete time ones. Using the local linearity, we derived the discretized process on each local neighborhood, and then combining them with each other, the forecasting model was constructed.

Through the simulation studies by using stochastic differential equations with nonlinear drift coefficients, it was confirmed that the model can estimate well the drift coefficients. And besides, we also carried out a comparative study of multi-step ahead forecasting with the global linear model, and found the proposed model performs better than the global linear model.

References

Aït-Sahalia, Y. (1996). Nonparametric pricing of interest rate derivative securities. *Econometrica* **64**, 527-560.

Arapis, M. and Gao, J. (2006). Empirical Comparisons in Short-Term Interest Rate Models Using Nonparametric Methods. *Journal of Financial Econometrics* **4**, 310-345.

Bandi, F. M. and Phillips, P. C. B. (2003) Fully nonparametric estimation of scalar diffusion models. *Econometrica* **71**, 241-283.

Campbell, J.Y., Lo, A. W. and MacKinlay, A. C. (1997). *The Econometrics of Financial Markets.* Princeton University Press: Princeton, New Jersey.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications.* Chapman & Hall: London.

Fan, J. and Yao, Q. (1998). Efficient estimation of conditional variance functions in stochastic regression. *Biometrika* **85**, 645-660.

Fan, J. and Zhang,, C. (2003). A reexamination of diffusion estimators with applications to financial model validation. *Journal of the American Statistical Association* **98**, 118-134.

Florens-Zmirou, D. (1993). On estimating the diffusion coefficient from discrete observations. *Journal of Applied Probability* **30**, 790-804.

Jiang, G. and Knight, J. (1997). A nonparametric approach to the estimation of diffusion processes with an application to a short-term interest rate model. *Econometric Theory* **13**. 615-645.

Kloeden, P. E. and Platen, E. (1995). *Numerical Solution of Stochastic Differential Equations: 2nd ed.* Springer.

Nicolau, J. (2003). Bias reduction in nonparametric diffusion coefficient estimation. *Econometric Theory* **19**. 754-777.

Stanton, R. (1997). A nonparametric model of term structure dynamics and the market price of interest rate risk. *Journal of Finance* **52**, 1973-2002.

forecast step		1	2	5	10
model1					
	NONPAR	44.24	139.77	783.79	3206.43
		(14.41)	(68.34)	(652.07)	(3665.96)
	LIN	46.84	151.96	895.48	3795.30
		(19.05)	(89.94)	(841.78)	(4606.89)
model2					
	NONPAR	40.61	122.48	617.85	2274.21
		(8.81)	(40.43)	(361.35)	(1898.20)
	LIN	42.84	132.94	714.92	2799.38
		(14.34)	(66.23)	(591.02)	(3066.49)
model3					
	NONPAR	40.25	121.10	609.08	2242.46
		(4.05)	(18.84)	(195.53)	(1253.11)
	LIN	40.59	122.72	624.51	2331.50
		(4.92)	(22.80)	(222.95)	(1326.34)

Table 1: Comparison of performance of 1, 2, 5 and 10 step ahead forecasting by proposed and global linear models; NONPAR for the proposed model and LIN for the global linear model

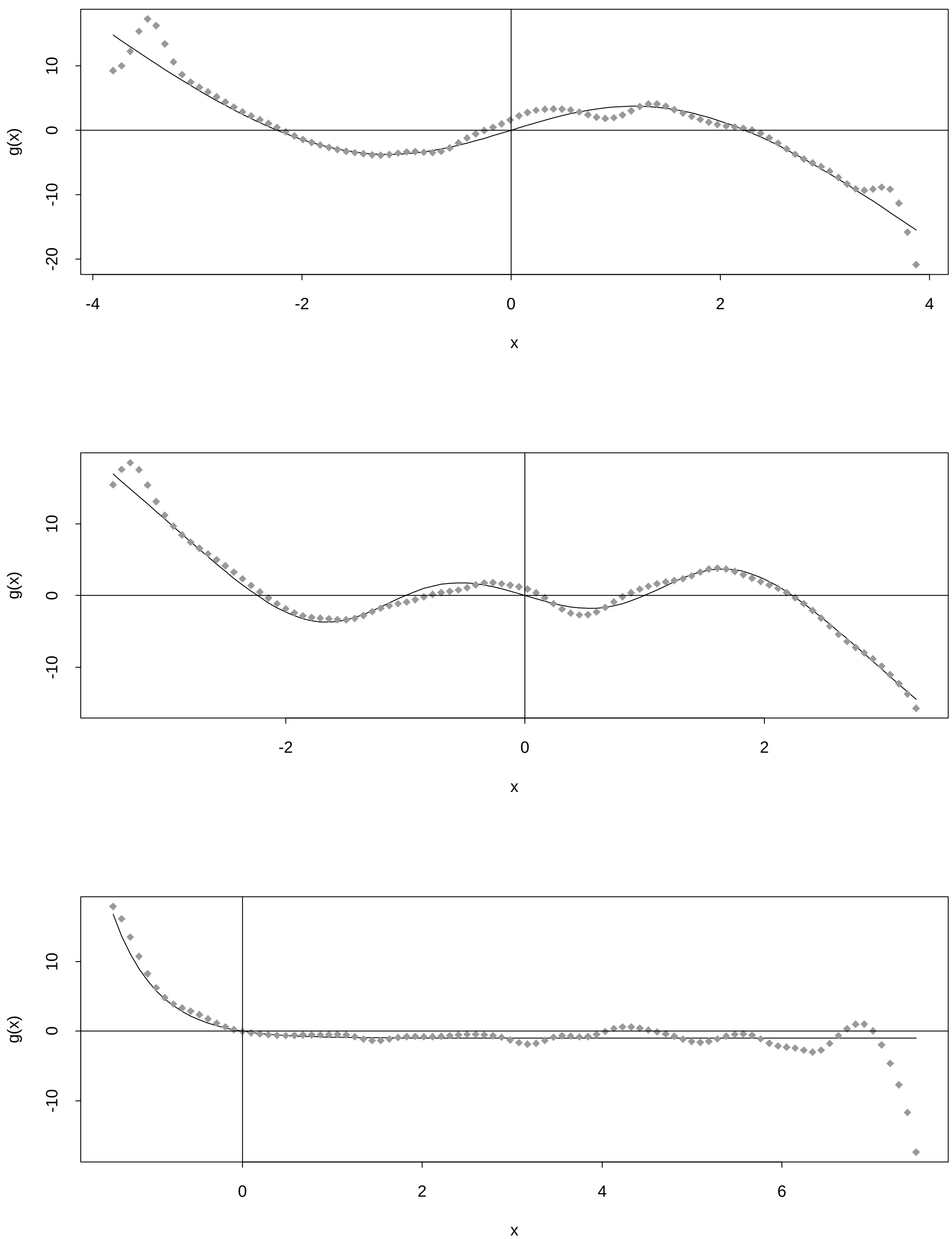

Figure 1: Estimation of drift functions: the upper to the lower for model1 to model3