On some problems of semi-parametric estimation related to stochastic volatility models

Yuri Golubev

CNRS, Université de Provence

Le Mans, December 19–20, 2002

Stochastic volatility model

Our goal is to estimate the unknown parameter θ based on the sample

$$X^n = \left\{ X\left(\frac{1}{n}\right), \dots, X\left(\frac{k}{n}\right), \dots, X(1) \right\}$$

of the random process $X(t)$ satisfying the SDE

$$dX(t) = c(t, v_\theta(t), X(t))dt + v_\theta(t)dw'(t), \quad t \in [0, 1],$$

where $w'(t)$ is a standard Wiener process, $c(\cdot, \cdot, \cdot)$ is a known fucntion.

It is assumed that the unobservable random process $v_\theta(t)$ is defined by

$$dv_\theta(t) = a(t, v_\theta(t))dt + b_\theta(v_\theta(t))dw(t), \quad t \in [0, 1],$$

where $a(\cdot, \cdot)$ and $b_\theta(\cdot)$ are known functions and $w(t)$ is a standard Wiener process independent of $w'(t)$.

Under natural conditions we have for large n

$$\sqrt{n}\left[X\left(\frac{k+1}{n}\right)-X\left(\frac{k}{n}\right)\right]\approx v_\theta\left(\frac{k}{n}\right)\xi_k,$$

where ξ_k are i.i.d. $\mathcal{N}(0,1)$.

Therefore it is assumed that we have in our disposal

$$Y_k=v_\theta\left(\frac{k}{n}\right)\xi_k+\varepsilon\xi'_k,$$

where ξ'_k are i.i.d. $\mathcal{N}(0,1)$ independent of ξ_s.

The small parameter $\varepsilon>0$ is assumed to be known. It plays a sufficiently important role regularizing the problem of recovering θ.

We consider the simplest case assuming

$$a(t,x) = 0, \quad b_\theta(x) = \theta.$$

So $v_\theta(t) = v(0) + \theta w'(t)$.

Then the observations Y_k can be rewritten in the following equivalent form

$$Y_k = \left[v(0) + \theta w\left(\frac{k}{n}\right) \right] \xi_k + \varepsilon \xi'_k, \quad k = 1, \ldots, n.$$

Some known facts:

- Gloter (2000) constructed an estimator $\widehat{\theta}$ such that
$$\mathbf{E}(\widehat{\theta} - \theta)^2 \le C n^{-1/2}.$$

- Hoffman (2002) has shown that the rate of convergence $n^{-1/4}$ can not be improved.

Roughly speaking, the rate of convergence $n^{-1/4}$ is due to the following reasons

- the trajectory of $w(t)$ is *not known*
- $w(t)$ is a *random process*
- the derivative of $w(t)$ is the *white Gaussian noise*

If the trajectory of $w(t)$ is known, then one can recover θ with the ordinary parametric rate $n^{-1/2}$.

If $w(t)$ is an unkown deterministic function, then the consistent estimator does not exist.

If $w(t)$ coincides with a Gaussian random variable η, then we can estimate the product $\theta\eta$ very well with the parametric rate $n^{-1/2}$. But evidently we can not recover θ.

The goal of my talk is to shed some light on computation of the rate of convergence up to a constant. For simplicity we will assume that $v(0)$ is known and strictly positive and $\theta > 0$.

The main advantage of the *semiparametric Bayesian approach* is that the optimal estimator is known

$$\widehat{\theta}_b = \arg\min_{\theta>0} \left\{ \mathbf{E}^w \mathrm{e}^{-L(\theta,w,Y)/2} \right\},$$

where $\mathbf{E}^w$ is the expectation with respect to the measure generated by the Wiener process, and

$$L(\theta, w, Y) = \sum_{i=1}^{n} \left\{ \frac{Y_i^2}{S^2[\theta w(i/n)]} + \log S^2[\theta w(i/n)] \right\}$$

with

$$S^2(x) = \varepsilon^2 + [v(0) + x]^2.$$

Note here that numerical computation of this estimator is difficult and interesting problem.

We will look for the minimal functional $I(\theta, w)$ such that

$$\inf_{\widehat{\theta}} \sup_{\theta} \mathbf{E}(\widehat{\theta} - \theta)^2 \sqrt{n} I(\theta, w) \geq 1 + o(1), \quad n \to \infty.$$

Why do we need the lower bound in this form? The reason is that we want to control *precisely* the quality of the optimal estimator. If we suppose that there exists an oracle which provides us with the hole trajectory of the unobservable Wiener process $w(t)$ and θ, then we could say that

$$(\widehat{\theta}_b - \theta)^2 \gtrsim \frac{1}{\sqrt{n} I(\theta, w)}.$$

Of course such oracle does not really exist, but usually we can estimate the functional $I(\theta, w)$ based on the data $Y_k,\ k = 1, \ldots, n$.

$$Y_k = \left[v(0) + \theta w\left(\frac{k}{n}\right)\right]\xi_k + \varepsilon\xi'_k,\ \ k = 1, \ldots, n.$$

In statistical analysis of this model there exist three principal difficulties:

- the data Y_k are non-Gaussian
- $w(t)$ is an irregular function of t
- $w(t)$ is a long memory process

Our approach is based on the idea to replace the Wiener process $w(t)$ by an approximation $\tilde{w}^n(t)$.

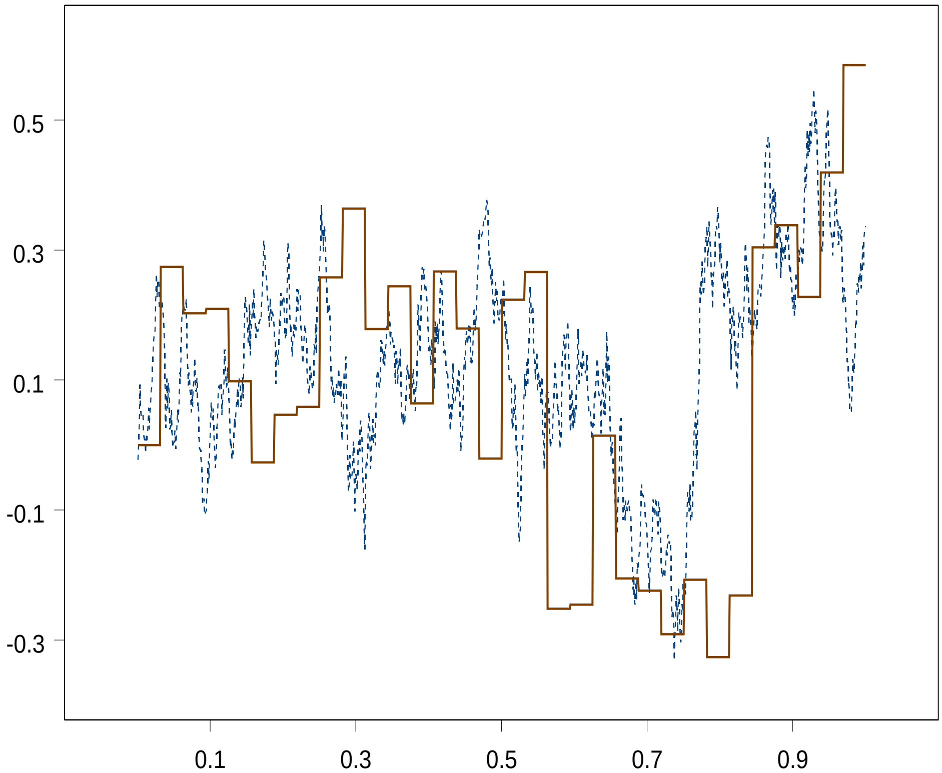
Two scale method:

- $\tilde{w}^n(t)$ is a step function with $\Delta_n \ll n^{-1/2}$

- $\tilde{w}^n(kh_n) = w(kh_n), \ k = 1, \ldots, h_n^{-1}$, where $h_n \gg n^{-1/2}$

- given $w(kh_n)$ the random variables $\tilde{w}^n(t)$ and $\tilde{w}^n(s)$ are independent for $|t - s| \geq h_n$

- the risk of recovering θ based on Y_k is asymptotically greater then the risk of estimation θ based on
$$\tilde{Y}_k = \left[v(0) + \theta \tilde{w}^n \left(\frac{k}{n} \right) \right] \xi_k + \varepsilon \xi_k'.$$



An approximation of the Wiener process.

Remark. Note that the Hellinger distance between the measures generated by Y_k and $\tilde{Y}_k$ tends to infinity as $n \to \infty$. It means in particular that the statistical experiments of recovering vectors with the components $\theta w(k/n)$ and $\theta\tilde{w}(k/n)$ are not asymptotically equivalent. But nevertheless one could show that the statistical experiments related to the estimation θ are asymptotically equivalent in the Le Cam sense.

Two scale approach leads to the following result. Let

$$I(\theta, w) = \frac{1}{\sqrt{2}\theta} \int_0^1 \frac{|v(0) + \theta w(t)|}{|v(0) + \theta w(t)|^2 + \varepsilon^2} \, dt$$

Theorem 1 *As $n \to \infty$ we have*

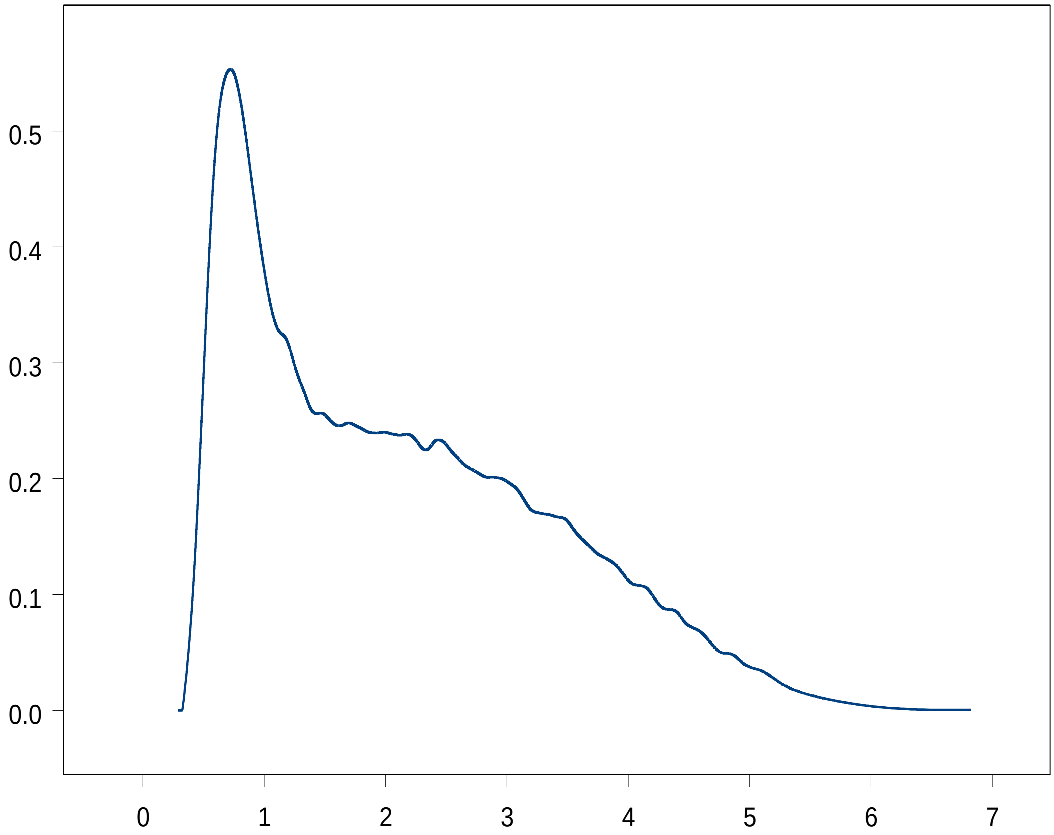
$$\inf_{\widehat{\theta}} \sup_{\theta : |\theta - \theta_0| \leq \varepsilon_n} \mathbf{E}_\theta [\widehat{\theta}(Y) - \theta]^2 \sqrt{n} I(\theta, w) \geq (1 + o(1)),$$

where inf is taken over all estimators based on the observations

$$Y_k = \left[v(0) + \theta w \left(\frac{k}{n} \right) \right] \xi_k + \varepsilon \xi'_k, \;\; k = 1, \dots, n$$

and the sequence ε_n is such that

$$\lim_{n \to \infty} \varepsilon_n = 0, \quad \lim_{n \to \infty} \varepsilon_n n^{1/4} = \infty.$$



The probability density of $I(\theta, w)$ computed by the Monte-Carlo method for $\theta = 1$ and $v(0) = 0.5$. Sample size is 400000.

Linear model

In the heart of our approach lies the following simple problem of recovering θ. Suppose we want to estimate θ based on the following noisy observations

$$dx(t) = \theta w(t)dt + \frac{\sigma}{\sqrt{n}} dw'(t), \ t \in [0, 1],$$

where $w(t)$, $w'(t)$ are independent standard Wiener process and $\sigma > 0$ is a known parameter.

Let $\varphi_k(t)$ and λ_k be eigenfunctions and eigenvalues of the correlation operator related to the Wiener process

$$\int_0^1 \min(t,s)\varphi_k(s)ds = \lambda_k \varphi_k(t).$$

Then the Wiener process can be represented as

$$w(t) = \sum_{k=1}^{\infty} \sqrt{\lambda_k}\varphi_k(t)\xi_k,$$

where ξ_k are i.i.d. $\mathcal{N}(0,1)$.

Denoting

$$x_k = \int_0^1 x(t)\varphi_k(t)\,dt, \quad \eta_k = \int_0^1 \varphi_k(t)\,dw(t)$$

we find the equivalent representation of $x(t)$

$$x_k = \theta\sqrt{n\lambda_k}\xi_k + \sigma\eta_k \quad k = 1, 2, \ldots,$$

where ξ_k are i.i.d. $\mathcal{N}(0,1)$ independent of η_s.

The probability density of $x_k,\ k = 1, \ldots, m$

$$p_\theta(x_1, \ldots, x_m) = \exp\left\{ -\frac{1}{2} \sum_{k=1}^{m} \left[\frac{x_k^2}{\sigma^2 + n\theta^2\lambda_k} \right.\right.$$
$$\left.\left. + \log[2\pi(\sigma^2 + n\theta^2\lambda_k)] \right] \right\}.$$

So the Fisher information can be computed as follows

$$\begin{aligned} I^m(\theta) &= \mathbf{E}_\theta \left[\frac{d}{d\theta} \log p_\theta(x_1, \ldots, x_m) \right]^2 \\ &= \frac{2}{\theta^2} \sum_{k=1}^{m} \left(\frac{1}{1 + \sigma^2\theta^{-2}n^{-1}\lambda_k^{-1}} \right)^2. \end{aligned}$$

In order to continue recall that

$$\lambda_k = \left(\pi k + \frac{\pi}{2}\right)^{-2} = (1 + o(1))(\pi k)^{-2}, \quad k \to \infty.$$

So we obtain as $n \to \infty$

$$\begin{aligned} I(\theta) &= \lim_{m\to\infty} I^m(\theta) \\ &= \frac{2}{\theta^2} \sum_{k=1}^{\infty} \left(\frac{1}{1 + \sigma^2\theta^{-2}n^{-1}\lambda_k^{-1}} \right)^2 \\ &= (1 + o(1)) \frac{2\sqrt{n}}{\theta\pi\sigma} \int_0^\infty \frac{dx}{(1 + x^2)^2} \\ &= (1 + o(1)) \frac{\sqrt{n}}{2\theta\sigma}. \end{aligned}$$

Next using the standard technique based on the Van Trees inequality we arrive at the following

Lemma 2 *As* $n \to \infty$ *we have*

$$\inf_{\widehat{\theta}} \sup_{\theta:|\theta-\theta_0|\le\varepsilon_n} \mathbf{E}_\theta(\widehat{\theta}-\theta)^2 \frac{\sqrt{n}}{2\theta\sigma} \ge (1+o(1)),$$

where inf is taken over all estimators based on the observations

$$dx(t) = \theta w(t)dt + \frac{\sigma}{\sqrt{n}} dw'(t),\ t \in [0,1],$$

and the sequence ε_n *is such that*

$$\lim_{n\to\infty} \varepsilon_n = 0, \quad \lim_{n\to\infty} \varepsilon_n n^{1/4} = \infty.$$

Remark. One can check that in the considered problem we have the LAN. Also it is not difficult to show that the estimator

$$\widehat{\theta}_b = \arg\min_{\theta} \sum_{k=1}^{\infty} \left[\frac{x_k^2}{\sigma^2 + n\theta^2 \lambda_k} + \log(\sigma^2 + n\theta^2 \lambda_k) \right]$$

is asymptotically normal and

$$\lim_{n\to\infty} \mathbf{E}_\theta (\widehat{\theta}_b - \theta)^2 \frac{\sqrt{n}}{2\theta\sigma} = 1.$$

Approximating the Wiener process

Consider the stepwise approximation

$$\bar{w}^n(t) = \sum_{k=0}^{1/\Delta_n} w(k\Delta_n)\mathbf{1}\left\{k\Delta_n \le t < (k+1)\Delta_n\right\}.$$

We want to estimate θ based on

$$dx^n(t) = \theta\bar{w}^n(t)dt + \frac{\sigma}{\sqrt{n}}dw'(t),\ \ t \in [0,1].$$

Note that this model is equivalent to

$$y_k^{\Delta} = \theta w(k\Delta_n) + \frac{\sigma}{\sqrt{\Delta_n n}}\xi_k,\ \ k = 1,\ldots,\Delta_n^{-1}.$$

Lemma 3 *If*

$$\lim_{n\to\infty} \sqrt{n}\Delta_n = 0,$$

then we have

$$\inf_{\widehat{\theta}} \sup_{\theta:|\theta-\theta_0|\le\varepsilon_n} \mathbf{E}_\theta\left[\widehat{\theta}(y^{\Delta}) - \theta\right]^2 \frac{\sqrt{n}}{2\theta\sigma} \ge (1+o(1)).$$

Remark. It is not difficult to check that the Hellinger distance H_n between the Gaussian measures generated by continuous and discreet observations is of the order

$$H_n \approx 1 - \exp(-Cn\Delta_n).$$

So with respect to recovering $\theta w(t),\ t \in [0,1]$ the corresponding statistical experiments are asymptotically equivalent in the Le Cam sense if

$$\lim_{n\to\infty} n\Delta_n = 0.$$

But in view of Lemma 3 one can expect that the statistical experiments of recovering θ become equivalent when

$$\lim_{n\to\infty} \sqrt{n}\Delta_n = 0.$$

Partitioning the Wiener process

Suppose we observe

$$dx(t) = \theta w(t)dt + \frac{\sigma}{\sqrt{n}} dw'(t), \ t \in [0,1],$$

and we have also in our disposal the vector $\mathbf{v}$ with the components

$$v_k = w(kh_n), \ k = 1, \dots, h_n^{-1}.$$

Using this additional information we can improve the quality of recovering θ.

Lemma 4 *Let*

$$\lim_{k\to\infty} h_n\sqrt{n} = \infty.$$

Then as $n \to \infty$

$$\inf_{\widehat{\theta}} \sup_{\theta:|\theta-\theta_0|\le\varepsilon_n} \mathbf{E}_\theta[\widehat{\theta}(x,\mathbf{v}) - \theta]^2 \frac{\sqrt{n}}{2\theta\sigma} \ge 1 + o(1)$$

and moreover

$$\inf_{\widehat{\theta}} \sup_{\theta:|\theta-\theta_0|\le\varepsilon_n} \mathbf{E}_\theta \left\{ [\widehat{\theta}(x,\mathbf{v}) - \theta]^2 \frac{\sqrt{n}}{2\theta\sigma} \middle| \mathbf{v} \right\} \ge 1 + o(1),$$

where inf is taken over all estimators of θ *based on the data* $x(t), \ t \in [0,1]$ *and* $\mathbf{v}$.