Statistical Methods for High Frequency Data
Oslo January 2017

Per Mykland and Lan Zhang

Exercise Set 1

1. The purpose of this problem is to start thinking about inference when the semimartingale is contaminated by microstructure noise. We shall do this in the context of preaveraging, and for a parametric specification. Note that the parametric problem can also be tackled with an MLE, but this is computationally more complicated.

   We shall suppose that we observe data in the time interval $[0, T]$. We have n equidistant observations. In other words, set $\Delta t_n = T/n$, and our observations take place at times $t_{n,i} = i\Delta t_n$. We observe

   $$Y_i = X_{t_{n,i}} + \epsilon_{t_{n,i}},$$

   where $X_t = \sigma W_t$, where (W_t) is a Brownian motion, and the $\epsilon_{t_{n,i}}$ are i.i.d. (independent and identically distributed), with distribution $N(0, \nu^2)$. We assume that the $\epsilon_{t_{n,i}}$'s are independent of the Brownian motion (W_t). Use notation $\Delta X_{t_{n,i}} = X_{t_{n,i}} - X_{t_{n,i-1}}$, etc. Note that

   $$\Delta Y_i = \Delta X_{t_{n,i}} + \Delta \epsilon_{t_{n,i}}. \tag{1}$$

   (a) Find $\text{Cov}(\Delta Y_i, \Delta Y_{i-k})$ for $k = 0, 1, 2, \cdots$.

   —

   We shall consider averages of $X_{t_{n,i}}$ and Y_i in blocks of size M. Specifically, let $I = 1, \cdots, n_M = \lfloor n/M \rfloor$ (the largest integer $\leq n/M$). In block $\#I$, set

   $$\bar{X}_I = \frac{1}{M} \sum_{i=M(I-1)+1}^{MI} X_{t_{n,i}} \text{ and } \bar{Y}_I = \frac{1}{M} \sum_{i=M(I-1)+1}^{MI} Y_i.$$

   (For simpicity, we do not consider overlapping blocks. We also throw away the observations from $M\lfloor n/M \rfloor + 1$ to n.)

   (b) Set $\tau_{n,I} = IM\Delta t_n = t_{n,MI}$. Note that we can alternatively write $\bar{X}_I = \frac{1}{M} \sum_{\tau_{n,I-1} < t_{n,i} \leq \tau_{n,I}} X_{t_{n,i}}$, and so on. Find the variance $\sigma^2 V_1$ of $\bar{X}_1 - X_0$ and the covariance $\sigma^2 V_2$ of $\bar{X}_1 - X_0$ and $X_{\tau_{n,1}} - X_0$. [Hints: (1) Use telescope sums: for example, $\bar{X}_1 - X_0 = \frac{1}{M} \sum_{i=1}^{M} \sum_{j=1}^{i} \Delta X_{t_{n,j}} = \frac{1}{M} \sum_{j=1}^{M} \sum_{i=j}^{M} \Delta X_{t_{n,j}}$. (2) $\sum_{k=1}^{M} k^2 = (2M+1)(M+1)M/6$.]

(c) Find the joint distribution of $X_{\tau_{n,I}} - \bar{X}_I$ and $\bar{X}_I - X_{\tau_{n,I-1}}$, $I = 1, \cdots, \lfloor n/M \rfloor$. You can use σ^2, V_1 and V_2 to characterize the process. [Hints: (1) Independent normal sums are normal. (2) Normal random variables are characterized by their means, variances and covariances.]

(d) Consider the process $\Delta\bar{X}_I = \bar{X}_I - \bar{X}_{I-1}$. Determine $\mathrm{Var}(\Delta\bar{X}_I)$ and $\mathrm{Cov}(\Delta\bar{X}_I, \Delta\bar{X}_{I-K})$ for $K = 1, 2. \cdots$, in terms of σ^2, V_1 and V_2.

(e) Similarly, consider the process $\Delta\bar{Y}_I = \bar{Y}_I - \bar{Y}_{I-1}$. Determine $C_K = \mathrm{Cov}(\Delta\bar{Y}_I, \Delta\bar{Y}_{I-K})$ for $K = 0, 1, 2. \cdots$ (including $C_0 = \mathrm{Var}(\Delta\bar{Y}_I)$) in terms of σ^2, V_1, V_2, ν^2, and M.

(f) For what values of σ^2, V_1, V_2, M, and ν^2 can you write

$$\Delta\bar{Y}_I = \zeta\Delta B_{\tau_{n,I}} + \Delta\eta_{\tau_{n,I}} \ , \tag{2}$$

where (B_t) is a Brownian motion, and the $\eta_{\tau_{n,I}}$ are i.i.d. normal and independent of (B_t)? What is the value of ζ^2 and the variance of $\eta_{\tau_{n,I}}$?

(g) Any comments on the relationship between (2) and (1)?

(h) Recall that $n_M = \lfloor n/M \rfloor$. Consider the estimators $K = 1, 2, 3, \cdots$:

$$\hat{C}_K = \frac{1}{n_M - K} \sum_{I=1}^{n_M - K} \Delta\bar{Y}_I \Delta\bar{Y}_{I+K} \tag{3}$$

Find an unbiased estimator $\hat{\sigma}_n^2$ of σ^2 based on $\hat{C}_0$ and $\hat{C}_1$.

—

The following extra questions are hard, and optional:

(i) Is the estimator in the previous question consistent? Does the answer change if you allow M to depend on n (*i.e.*, $M = M_n$)?

(j) What is the asymptotic distribution of $\hat{\sigma}_n^2$? Is there an optimal relationship between M_n and n?

2. Stochastic Calculus Roundup

(a) Show Lévy's Theorem (Thm 2.14 in the notes) using a charactertistic function argument and Itô's formula.

(b) Argue Itô's formula (Thm 2.15 in the notes) using a Taylor expansion argument. Also try to argue the more general version for discontinuous semimartingales.

(c) Go though the leverage effect argument in Section 2.2.5.

(d) Derive a version of the options trading result (in Section 2.2.6 in the notes) where we assume that there is also a traded zero coupon bond Λ_t which terminates in value $\Lambda_T = 1$ (cf. p. 21/p. 129 in the notes). The constraints (2.21) in the notes are replaced by the constraint

$$\Xi^+ \geq \int_0^T (\sigma_t^*)^2 dt \geq \Xi^-$$