Statistical Methods for High Frequency Data
Oslo January 2017

Per Mykland and Lan Zhang

Exercise Set 2

Make the assumptions in Section 3.2 of the notes. Your data is $X_0, X_{t_{n,1}}, ..., X_{t_{n,i}}, ..., X_{t_{n,n}} = X_T$. For simplicity we write t_i to mean $t_{n,i}$.

1. (Simple subsampling.) In addition to the usual grid $\mathcal{G}_n = \{0 = t_{n,0}, t_{n,1}, ..., t_{n,n} = T\}$, also consider
$\mathcal{G}_{n,1} = \{0, t_{n,1}, t_{n,3}, ..., t_{n,2i+1}, ..., T\}$ and $\mathcal{G}_{n,2} = \{0, t_{n,2}, t_{n,2}, ..., t_{n,2i}, ..., T\}$, so that, in particular, $\mathcal{G}_{n,1} \cup \mathcal{G}_{n,2} = \mathcal{G}_n$, while $\mathcal{G}_{n,1} \cap \mathcal{G}_{n,2} = \{0, T\}$. Consider the following estimator of volatility:

$$\widehat{[X, X]}_t = \frac{1}{2} \left([X, X]_t^{\mathcal{G}_{n,1}} + [X, X]_t^{\mathcal{G}_{n,2}} \right). \tag{1}$$

In addition to t_* from (2.27) in the notes, also define $t_{**} = \max\{t_i \in \mathcal{G} : t_{i+1} \leq t\}$.

   (a) Show that
$$\widehat{[X, X]}_t - [X, X]_t^{\mathcal{G}_n} = L_t^{(n,d)} + o_p(n^{-1/2}) \tag{2}$$
(where "d" stands for discrete), where
$$L_t^{(n,d)} = \sum_{t_{i+1} \leq t} \Delta X_{t_{i-1}} \Delta X_{t_i}. \tag{3}$$

   (b) Provide a continuous interpolation $L_t = L_t^{(n)}$ of $L_t^{(n,d)}$. (In analogy with the interpolation used to create a continuous M_t on p. 136 and a continuous $[X, X, X, X]_t^{\mathcal{G}}$ on p. 137. The interpolated process should be a local martingale.)

   (c) Explain why the difference between these two martingales ($L_t^{(n,d)}$ and $L_t^{(n)}$) is asymptotically negligible.

   (d) Provide a (data based and consistent) estimate of the quadratic variation $\widehat{\text{QV}}$ of L_t.

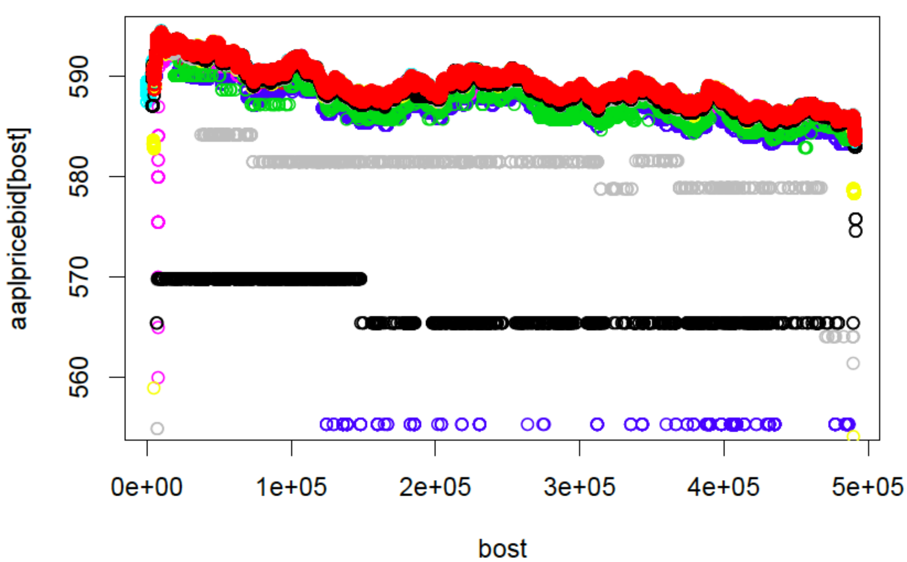
   (e) Derive the asymptotic value of $n[L^{(n)}, L^{(n)}]_t$ under the assumptions of, say, Proposition 2.21 (p. 143) in the notes.

2. (A first stab at data. With mystery.)

   (a) Retrieve the bid and offer prices for Apple (AAPL) and Google (GOOG) for Dec 3, 2012. The data are from the TAQ database in WRDS.

   (b) Use R or excel to construct signature plots for the realized volatility of the bid and the offer, the midquote ((bid+offer)/2). Also provide a signature plot for the realized co-volatility (covariance). [Hints: You may need substantially larger values of K to get comparable results. For standardization, please plot $\sqrt{(250 * r)}$, where r is any of the realized quantities. If $r < 0$, use sign of r $\times \sqrt{(250 * |r|)}$.] Explain as well as you can the effects that you observe. [Hint: there is a mystery here, with a clue (see the plot below, courtesy of CH). You are encouraged to try to solve it.]

   (c) For the same data (possibly after any adjustments what you may with to make based on the solution to the mystery), compute the statistic $L_T^{(n,d)}$ from the previous problem, as well as $\widehat{\mathrm{QV}}$. We shall later show that $L_T^{(n,d)}/\sqrt{\widehat{\mathrm{QV}}}$ is asymptotically N(0,1) if the X process is an Itô process. On this basis, test the null hypothesis H_0 that the X process is an Itô process.



3. (Mathematical subsampling tricks.)

   (a) Let $p > 1$. Define $\mathcal{R}_p(\mathcal{G}_n) = \sum_{i=1}^n (\Delta t_i)^p$, where $\Delta t_i = t_i - t_{i-1}$. Now form a new grid $\mathcal{G}_{n-1}$ by deleting one of the observation points (except t_0 or t_n). For example, if t_1 is deleted, then $\mathcal{G}_{n-1} = \{0 = t_0, t_2, t_3 ..., t_n = T\}$. Determine when $\mathcal{R}_p(\mathcal{G}_{n-1}) \geq \mathcal{R}_p(\mathcal{G}_n)$.

   (b) Consult Example 2.19 (ii), and set more generally $N_t = \#\{i : t_i \in (0, t]\}$. We define that the arrival of points follow an inhomogenlus Poisson process with intensity process (λ_t) if $N_t - \int_0^t \lambda_s ds$ is a martingale.

   Important properties are as follows: suppose that the intensity process is bounded from above and below by constants, say

$$0 < \lambda^- \leq \lambda_t \leq \lambda^+ < \infty. \tag{4}$$

   Then the points t_i can be generated as follows: Generate a set of points from a Poisson process with intensity λ^+, call these u_i. For each i flip a coin with probability of heads λ_{u_i}/λ^+. If heads, u_i is added to the grid of t_i's, otherwise it is discarded. Also note that given the points t_i generated from (λ_t), one can further subsample (keep t_i with probability λ^-/λ_{t_i}) to obtain time points from a Poisson process with intensity λ^-.

   Question: Determine whether the conditions of Proposition 2.17 are satisfied when observation points come from a inhomogenous Poisson process with intensity process (λ_t), assuming that (4) is satisfied.

4. (Picking every second Poisson arrival.) Consult Example 2.24 (ii), and note that when observations arrive according to a Poisson process, then, in probability,

$$n[M, M]_t \;\rightarrow\; 4T \int_0^t \sigma_s^4 ds \tag{5}$$

   Suppose instead that observations arrive according to a Poisson process, but for technical reasons we can only use every second observation. What is the corresponding limit in (5)? (Use n as the actual number of observations used.).

   [Notes: given the previous problem, we take for granted that the regularity conditions of Propositions 2.17 and 2.21 are satisfied.]

5. (Another way of generating irregular times.) Suppose that F is an increasing and continuously differentiable function which maps $[0, T]$ to $[0, T]$. Let G be the inverse function of F, so that $F(G(x)) = G(F(x)) = x$. We suppose that for each t, $F(t)$ is a stopping time. Set $u_i = u_{n,i} = Ti/n$ (equidistant sampling points) and define $t_i = t_{n,i} = F(u_i)$. As before, $dX_t = \sigma_t dW_t$. Define $Y_t = X_{F(t)}$. Let $\mathcal{G}_n$ be the grid based on the t_i's, while $\mathcal{H}_n$ is the grid based on the u_i's.

(a) Show that $[X,X]^{\mathcal{G}_n}_{F(t)} = [Y,Y]^{\mathcal{H}_n}_t$.

(b) Find the volatility of the process Y_t.

(c) Let M^X and M^Y be the error martingales for $[X,X]^{\mathcal{G}_n}_t - [X,X]_t$ and $[Y,Y]^{\mathcal{H}_n}_t - [Y,Y]_t$. Use the limit of $n[M^Y, M^Y]_T$ to find the limit of $n[M^X, M^X]_T$.

(d) Can this model for the generation of observation times be used to generate Poisson observation times?