Estimating Prices Transition Rate with Ultra-High-Frequency Data

Jian-ming Sun
Huazhong University of Science and Technology
Hubei Polytechnic University (430064 Wuhan China)

Abstract: Ultra-high-frequency data is defined to be a full record of transactions and their associated characteristics. In this paper marked point processes are applied to describe ultra-high-frequency data. By producing general marked point process sample function density, inserting the Markov process, which describes prices transition into the marked point process, prices transition rate is estimated by ML.

Key words: Prices transition rate   marked point processes   ultra-high-frequency data   ML estimation

1. Introduction

Ultra-high-frequency data is defined to be a full record of transactions and their associated characteristics (Robert F. Engle 2000). Transaction data inherently arrive in irregular time intervals. Of course, one can aggregate this data up to some fixed time interval (Tim Bollerslev 2001). There is naturally a loss of information in such aggregate. If the intervals are small the large number of zeros makes econometrics analysis very complex.

The times and prices of transactions are random. As in science and engineering, stochastic processes can be used to solve these econometrics problems. Engle and Russell (1997, 1998) developed an autoregressive conditional duration model, which is based on Poisson process. Robert F.Engle (2000) introduced a framework to estimate models for ultra-high-frequency data. The latter paper pay eyes on the transaction intensity. However the transaction prices are the best important, the changes of prices should be the center being paid attention to. This paper will develop a marked point process model, deduce sample function density, estimate parameters by ML, and insert the Markov process which describes the price changes of transactions into the marked point process, and then deduce the estimation formulation of prices transition rate.

2. Formulating The Economic Questions Statistically

Transactions data can be described two types of random variables. The first is the time of the transaction, and the second is a vector observed at the time of the transaction. In the literature of point processes, these latter variables are called marks as they identify or further describe the event that occurred.

At to any arrival time w_i of point process there is a mark u_i. u_i randomly take a mark in countable set $\{U_1, U_2 \cdots\}$. Assume T be a interval in $[t,\infty)$, A be a subset of mark space U. we denote the number of the points by N $(T \times A)$ which occur in time interval T and marks in A. Similarly N $([s,t) \times \{U\})$ is the total number in $[t_0,t)$. As to $t_o \le s < t$, $N_{s.t} = N_t - N_s$, Let N_t=N $([t_0,t) \times \{U\})$。

Assume that: a. $\{N_t; t \ge t_0\}$ is conditional orderliness.

b. Defining function

$$a(\Delta t, N_t, U_i) = \begin{cases} (\Delta t)^{-1} \Pr[N([t,t+\Delta t)\times\{U_i\}) = 1 \mid N_t], & N_t = 0 \\ (\Delta t)^{-1} \Pr[N([t,t+\Delta t)\times\{U_i\}) = 1 \mid N_t, w_1 \cdots w_{N_t}, u_1 \cdots u_{N_t}], & N_t \ge 1 \end{cases} \quad (1)$$

as $\Delta t \to o$, to almost all realizations of the marked point process, the limit of the function $a(\Delta t, N_t, U_i)$ exists.

c. $\Pr[N_{t_0} = 0] = 1$

As to N_t=0, let

$$m_{t,U_i}(0) = \lim_{\Delta t \to 0} (\Delta t)^{-1} \Pr[N([t, t+\Delta t) \times \{U_i\}) = 1 \mid N_t] \quad (2a)$$

As to $N_t \geq 1$, let

$$m_{t,U_i}(N_t; w_1 \cdots w_{N_t}, u_1 \cdots u_{N_t}) = \lim_{\Delta t \to 0} (\Delta t)^{-1} \Pr[N([t, t+\Delta t) \times \{U_i\}) = 1 \mid N_t, w_1 \cdots w_{N_t}, u_1 \cdots u_{N_t}] \quad (2b)$$

Define stochastic process $\{ l_{t,U_i}; t \geq t_0 \}$

$$l_{t,U_i} = \begin{cases} m_{t,U_i}(0), & t_0 \leq t \leq w_1 \\ m_{t,U_i}(N_t; w_1 \cdots w_{N_t}, u_1 \cdots u_{N_t}), & w_{N_t} < t \leq w_{N_t+1} \end{cases} \quad (3)$$

This process is called the ith mark density process. The probability that there is a point marked U_i is nearly $l_{t,U_i}\Delta t$。

$$\begin{aligned} \Pr[N_{t,t+\Delta t} = 1 \mid N_t = 0] &= \Pr[N([t, t+\Delta t) \times \mathbf{U}) = 1 \mid N_t = 0] \\ &= \sum_i \Pr[N([t, t+\Delta t) \times \{U_i\}) = 1 \mid N_t = 0] \\ &= m_t(0)\Delta t + o(\Delta t) \end{aligned} \quad (4)$$

Note $m_t(0) = \sum_i m_{t,U_i}(0)$

similarly: as to $N_t \geq 1$

$$\begin{aligned} \Pr[N_{t,t+\Delta t} = 1 \mid N_t; w_1 \cdots w_{N_t}, u_1 \cdots u_{N_t}] \\ = m_t(N_t; w_1 \cdots w_{N_t}, u_1 \cdots u_{N_t})\Delta t + o(\Delta t) \end{aligned} \quad (5)$$

Note

$$m_t(N_t; w_1 \cdots w_{N_t}, u_1 \cdots u_{N_t}) = \sum_i m_{t,U_i}(N_t; w_1 \cdots w_{N_t}, u_1 \cdots u_{N_t})$$

When given point numbers the arrival time and the marks of the points, the conditional probability that there is a point over $[t, t+\Delta t)$ is nearly $l_t \Delta t$

$$l_t = \Sigma l_{t,U_i} = \begin{cases} \sum_i m_{t,U_i}(0) & t_0 \leq t \leq w_1 \\ \sum_i m_{t,U_i}(N_t; w_1 \cdots w_{N_t}, u_1 \cdots u_{N_t}), & w_{N_t} < t \leq w_{N_t+1} \end{cases} \quad (6)$$

According to the assumption (a), the conditional probability that there is not a point is nearly $1 - l_t \Delta t$。

As to N_s=0

$$\Pr[N_{s.t} = 0 \mid N_s = 0] = \exp[-\int_s^t m_s(0) ds] \quad (7)$$

As to $N_s \geq 1$

$$\Pr[N_{s.t} = 0 \mid N_s; w_1 \cdots w_{N_s}, u_1 \cdots u_{N_t}] = \exp[-\int_s^t m_s(N_s; w_1 \cdots w_{N_s}, u_1 \cdots u_{N_t}) ds] \quad (8)$$

If giver a marked point process, an event occurs, then the conditional probability that there is a point marked U_i is nearly $l_{t,u_i} l_t^{-1}$.

3. Sample Function Density

The sample function density of marked point process is defined as:

$$f[\{Z_s; t_0 \le s < t\}] = \begin{cases} \Pr(N_t = 0) & N_t = 0 \\ f_w(w, u_1 = x_1, \cdots u_n = x_n, N_t = n) & N_t = n \ge 1 \end{cases} \quad (9)$$

As to $i(1 \le i \le n)$, $x_i \in \mathbf{U}$

$$f_w(w, u_1 = x_1 \cdots u_n = x_n, N_t = n) = \lim_{\substack{\max(\Delta w_i) \to 0 \\ 1 \le i \le n}} (\prod_{i=1}^{n} \Delta w_i)^{-1} \Pr[E_n, N_t = n]$$

Where as n=1. E_n denotes the events

$$\{N([t_0, w_1) \times \mathbf{U}) = 0, N([w_1, w_1 + \Delta w_1) \times \{x_1\}) = 1\}$$

as n≥2 denotes

$$\{N([t_0, w_1) \times \mathbf{U}) = 0, N([w_1, w_1 + \Delta w_1) \times \{x_1\}) = 1, \cdots,$$
$$N([w_{n-1} + \Delta w_{n-1}, w_n) \times \mathbf{U}) = 0, N([w_n, w_n + \Delta w_n) \times \{x_n\}) = 1\}$$

$f[\{Z_s; t_0 \le s < t\}]\Delta w_1 \Delta w_2 \cdots \Delta w_n$ can be understood as a particular realization probability which there is a marked point process over $[t_0, t)$, the realization has N_t=n points, which occur at $w_1 = W_1 \cdots w_n = W_n$ and have marks $u_1=\xi_1 \cdots u_n=\xi_n$

According to (7) as to N_t=0

$$f[\{z_s; t_0 \le s < t\}] = \exp[-\int_{t_0}^{t} m_s(0) ds] \quad (10)$$

As to N_t=1

$$\Pr(E_1, N_t = 1) = \Pr(N_{w_1} = 0) \times \Pr[N([w_1, w_1 + \Delta w_1) \times \{x_1\})$$
$$= 1 \mid N_{w_1} = 0] \times \Pr[N_{w_1 + \Delta w_1, t} = 0 \mid E_1]$$

combine (9) (8) with (2)

$$f[\{z_s, t_0 \le s < t\}] = m_{w_1, x_1}(0) \exp[-\int_{t_0}^{w_1} m_s(0) ds - \int_{w_1}^{t} m_s(1, w_1 x_1) ds] \quad (11)$$

Note $m_t(1; w_1, x_1) = \sum_i m_{t, U_i}(1, w_1 x_1)$

as to N_t=n≥2

$$\Pr(E_n, N_t = n) = \Pr(N_{w_1} = 0) \times \Pr[N([w_1, w_1 + \Delta w_1) \times \{x_1\}) = 1 \mid N_{w_1} = 0]$$
$$\times \{\prod_{i=2}^{n} \Pr[N([w_i, w_i + \Delta w_i) \times \{x_1\}) = 1 \mid E_{i-1}, N_{w_i} = i - 1]$$
$$\times \Pr(N_{w_{i-1} + \Delta w_{i-1}, w_i} = 0 \mid E_{i-1})\} \Pr(N_{w_n + \Delta w_n, t} = 0 \mid E_n)$$

$$f[\{z_s, t_0 \le s < t\}] = m_{w_1, x_1}(0) \left\{ \prod_{i=2}^{n} m_{w_i, x_i}(i-1; w_1 \cdots w_{i-1}, x_1 \cdots x_{i-1}) \right\} \times \exp[-\int_{t_0}^{w_1} m_s(0) ds -$$
$$\sum_{i=2}^{n} \int_{w_{i-1}}^{w_i} m_s(i-1, w_1 \cdots w_{i-1}, x_1 \cdots x_{i-1}) ds - \int_{w_n}^{t} m_s(n; w_1 \cdots w_n, x_1 \cdots x_n) ds] \quad (12)$$

4. Estimating Model Parameters

Take a realization $\{N_t, w_i \cdots w_{N_t}, u_1 \cdots u_{N_t}\}$ of the marked point process over $[t_0, t)$ as observation data, the Likelihood function is

$$\ln f\left[\{z_s ; t_0 \le s < t\}\right]$$

$$= \begin{cases} \int_{t_o}^{t} m_o(0)ds, & N_t = 0 \\ \int_{t_o}^{w_1} m_s(0)ds + \int_{w_1}^{t} m_s(1; w_1, m_1)ds + \ln m_{w_1,u_1}(0) & N_t = 1 \\ \int_{t_o}^{w_1} m_s(0)ds + \sum_{i=2}^{N_t}\int_{w_{i-1}}^{w_i} m_s(i-1; w_1 \cdots w_{i-1}, m_1 \cdots m_{i-1})ds \\ + \int_{w_{N_t}}^{t} m_s(N_t; w_1 \cdots w_{N_t}, u_1 \cdots u_{N_t})ds + \ln m_{w_1,u_1}(0) \\ + \sum_{i=2}^{N_t} \ln m_{w_i,u_i}(i-1, w_1 \cdots w_{i-1}, u_1 \cdots u_{i-1}) & N_t \ge 2 \end{cases} \quad (13)$$

The likelihood estimator is the values which make likelihood functions (13) take maximum, where it is needed to be represented by data.

$$N_t = n \quad w_1 = W_1 \cdots w_n = W_n, u_1 = x_1 \cdots u_n = x_n$$

5. Estimation Prices Transition Rate

Let x_t be prices, $\{x_t, t \ge t_0, x_{t_0} = X_0\}$ be a jump Markov process, define

$$a_{x_i,x_k}(t) = \begin{cases} \lim_{\Delta t \downarrow 0}(\Delta t)^{-1} \Pr[x_{t+\Delta t} = x_k] \mid x_t = x_i], & i \ne k \\ -\sum_{j(j\ne i)} a_{x_i},_{x_j}(t) & i = k \end{cases} \quad (14)$$

Is transition rate from condition x_i to condition x_k, when given beginning condition X_0, a_{x_i,x_j} statistically denote $\{x_t, t \ge t_o, x_{t_o} = X_0\}$

if x has a transition from x_t to x_{t+} at time t, the marked point has a point marked $x_{t+} - x_t$. The mark density process l_{t,U_i} can be obtained by

$$l_{t,U_i} = a_{z_t+X_0, z_t+X_0+U_i}(t) \quad (15)$$

z_t is the accumulation of all marks over $[t_0, t)$, and $z_t + X_0 = x_t$

Replace (15) into (3)(13)

$$l = -\sum_i [(W_1 - t_0)a_{X_0,X_0+U_i} + (W_2 - W_1)a_{X_0+x_1,X_0+x_1+U_i]} + \cdots + (T - W_n)a_{X_0+x_1+\cdots+x_n, X_0+x_1+\cdots+x_n+U_i}]$$
$$+[\ln a_{X_0,X_0+x_1} + \ln a_{X_0+z_1,X_0+x_1+x_2} + \cdots + \ln a_{X_0+x_1,+\cdots x_{n-1}, X_0+x_1+\cdots+x_n}] \quad (16)$$

n is the observed transitions number of x over $[t_o, T)$, $w_1, w_2 \cdots w_n$ are the transition time. $x_1, x_2 \cdots x_n$ are observed transitions.

$$\text{Note} \quad I_U(x) = \begin{cases} 1 & x = U \\ 0 & x \ne U \end{cases} \quad (17)$$

Likelihood function can be rewritten as:

$$l = -\sum_i \sum_{j(\ne i)} [(W_1 - t_o)I_{X_0}(x_i) + (W_2 - W_1)I_{X_0+x_1}(x_i)] + \cdots + (T - W_n)I_{X_0+x_1+\cdots x_n}(x_i)]a_{x_i,x_j}$$

$$+\sum_{i}\sum_{j(\neq i)}[I_{X_0}(x_i)I_{X_0+x_1}(x_j)+I_{X_0+x_1}(x_i)I_{X_0+x_1+x_2}(x_j)+I_{X_0+x_1+\cdots+x_{n-2}}(x_j)I_{X_0+x_1+\cdots+x_{n-2}}(x_j)]\ln(a_{x_i,x_j})$$

As to $i \neq j$

$$\frac{\partial l}{\partial a_{x_i,x_j}} = -T(x_i) + \frac{N(x_i \to x_j)}{a_{x_i,x_j}} \tag{18}$$

Note:

$$T(x_i) = (W_1 - t_0)I_{x_0}(x_i) + (W_2 - W_1)I_{x_0+x_1}(x_i) + \cdots + (T - W_n)I_{x_0+x_1+\cdots+x_n}(x_i)$$

$$N(x_i \to x_j) = I_{X_0}(x_i)I_{X_0+x_1}(x_i) + I_{X_0+x_1}(x_i)I_{X_0+x_1+x_n}(x_j) + \cdots + I_{X_0+x_1+\cdots+x_{n-1}}(x_i)I_{X_0+x_1+\cdots+x_n}(x_j)$$

$T(x_i)$ is the total time of x in condition x_i, $N(x_i \to x_j)$ is the transition times from condition x_i to x_j. Then the ML estimator $\hat{a}_{x_i x_j}$ of $a_{x_i x_j}$ is

$$\hat{a}_{x_i,x_j} = \frac{N(x_i \to x_j)}{T(x_i)} \tag{19}$$

$T(x_i) \neq 0$

If $T(x_i) = 0$, let $\hat{a}_{x_i,x_j} = 0$

6. Conclusion

This paper proposed a new technique for modeling irregularly spaced time series data using marked point processes that is particularly well suited for financial transactions data. The formulation of prices transition rate estimation can be used to calculate the value at risk.

References

[1] T.Bollerslev, J.H.Wright, "High-Frequency Data, Frequency Domain Inference And Volatility Forecasting", the Review of Economics and Statistics, 2001.83(4): 596-602

[2] T.Bollerslev. Financial econometrics: Past developments and future challenges. Journal of econometrics 100(2001), 41-51

[3] R.F.Engle "The Econometrics of Ultra-High frequency Data", Econometric 68 (2000), 1-22

[4] R.F.Engle Russell, J.R "Autoregressive conditional duration: a new model for irregularly spaced transactions prices", Econometrica 66.(1998) 1127-1163

[5] R.F. Engle Financial econometrics-A new discipline with new methods. Journal of econometrics 100(2001), 53-36

[6] D.L.Snyder, M.I. Miller: Random point processes, (1974) New York: Springer –Verlag.