应用概率统计 第 32 卷
第 6 期 2016 年 12 月

Chinese Journal of Applied Probability and Statistics
Dec., 2016, Vol. 32, No. 6, pp. 581-591
doi: 10.3969/j.issn.1001-4268.2016.06.002

噪音环境下跳–扩散模型中积分波动率的非参数估计 *

叶绪国[1,2]⋆ 林金官[1]

([1]东南大学数学系, 南京, 210096; [2]凯里学院数学科学学院, 凯里, 556011)

摘 要: 本文研究了噪音环境下跳–扩散过程的积分波动率非参数估计问题. 利用门限技术, 分别提出了积分波动率的门限两尺度与多尺度已实现波动率估计量, 并获得了相应的大样本性质, 推广了已有文献的结果.

关键词: 非参数估计; 门限技术; 微结构噪音; 跳–扩散过程; 积分波动率

中图分类号: O212.1

§1. 引 言

跳–扩散过程在概率论理论中是一类重要的随机过程, 同时也是描述金融资产收益的一个重要模型, 在金融和保险中得到广泛地应用. 模型的建立可追溯到1976年[1], 它是由伊藤随机微分方程所定义的跳–扩散过程

$$\mathrm{d}X_t = \mu_t\mathrm{d}t + \sigma_t\mathrm{d}B_t + \gamma_t\mathrm{d}J_t, \quad t \in [0, T], \tag{1}$$

其中μ_t, $\sigma_t\,(\sigma_t > 0)$分别是即时漂移项, 即时波动率项, 和B_t是一维标准布朗运动. J_t是一个与B_t独立, 且强度为λ的计数过程. $\gamma_t = X_{t,+} - X_t$是对数价格过程中对应跳时刻的跳跃尺度, 且J_t与γ_t相互独立. 假定J_t是一个与B_t独立的复合Poisson过程, 则过程(1)的半鞅形式为

$$X_t = X_0 + \int_0^t \mu_s\mathrm{d}s + \int_0^t \sigma_s\mathrm{d}B_s + \sum_{k=1}^{N_t} \gamma_k, \tag{2}$$

其中N_t表示在$[0, t]$时段内跳点的个数, 且服从强度为λ的Poisson过程. 跳跃尺度γ_k表示中心化的, 且相互独立的随机变量, 且N_t与γ_k相互独立.

我们定义对数价格过程X_t的二次变差为(Quadratic Variation, 简写为QV)

$$[X, X]_t = \int_0^t \sigma_s^2\mathrm{d}s + \sum_{k=1}^{N_t} \gamma_k^2. \tag{3}$$

---

*贵州省科技厅自然科学基金(LKK[2013]30)、凯里学院校级重点项目(Z1505)、江苏省自然科学基金项目(BK20141326)和江苏省普通高校研究生科研创新计划项目(KYLX15_0102)资助.

⋆通讯作者, E-mail: yexuguo522@126.com.

本文2015年1月4日收到, 2016年6月10日收到修改稿.

它由积分波动率与跳变差两部分组成, 我们分别定义为

$$\mathrm{IV}=\int_0^t \sigma_s^2 \mathrm{d}s \qquad 和 \qquad \mathrm{SJ}=\sum_{k=1}^{N_t} \gamma_k^2.$$

当跳不存在时, Andersen等[2,3]、Barndorff-Nielsen和Shephard[4,5]给出了二次变差QV的一个非参数估计量 — 已实现波动率(Realized Volatility, 简写RV)

$$\mathrm{RV}=\sum_{i=1}^{n}(X_{t_i}^*-X_{t_{i-1}}^*)^2, \tag{4}$$

其中X_t^*表示连续时间的对数资产价格过程, 且服从连续Itô过程

$$X_t^*=X_0^*+\int_0^t \mu_s \mathrm{d}s+\int_0^t \sigma_s \mathrm{d}B_s, \qquad t\in[0,T]. \tag{5}$$

当取样频率趋于无穷大($n\to+\infty$)时, 则已实现波动率依概率收敛积分波动率IV, 即

$$\mathrm{RV}\xrightarrow{\mathrm{p}}\int_0^T \sigma_s^2 \mathrm{d}s.$$

当跳出现时, Barndorff-Nielsen和Shephard[6]、Andersen等[7]证明了

$$\mathrm{RV}=\sum_{i=1}^{n}(X_{t_i}-X_{t_{i-1}})^2\xrightarrow{\mathrm{p}}\int_0^T \sigma_s^2 \mathrm{d}s+\sum_{k=1}^{N_T}\gamma_k^2, \tag{6}$$

从理论上说, 在$[0,T]$上资产收益取样频率越大, 已实现波动率越收敛于二次变差.

相对于跳变差部分, 归因于在对冲、期权定价、风险分析与投资组合方面波动率扮演着重要的角色, 积分波动率估计已成为现代计量经济学研究的热点. 为了消除跳对积分波动率估计的影响, Barndorff-Nielsen和Shephard[6,8]提出了已实现多次幂变差估计量. Mancini[9]利用门限技术实现了积分波动率的非参数估计. Corsi等[10]结合多次幂变差与门限技术, 研究了带有跳的门限多次幂变差估计策略. Veraart[11]比较了这两种方法的有限样本的表现. Andersen等[12]推广了多次幂变差技术, 提出最近邻截断估计量.

事实上, 在高频金融数据环境下, 除了跳对积分波动率估计有一定的影响外, 还受到微结构噪音的影响(市场微观结构噪音的存在是归因于市场交易过程不完善, 包括竟要价跃动、不同步交易、闭市影响等现象). 正如Zhou[13]所指出, 当取样频率增加, 已实现波动率偏离了积分波动率, 并且市场微观结构噪音的影响随取样频率的增加而增强, 对积分波动率估计的影响是相当显著. 为了消除市场微观结构噪音, Aït-Sahalia等[14]、Bandi和Russell[15,16]研究了稀疏取样消噪方法. Zhang等[17]、Zhang[18]分别提出了双尺度已实现波动率(Two-Scale Realized Volatility, 简写TSRV)和多尺度已实现波动率(Multi-Scale Realized Volatility, 简写MSRV). Hansen和Lunde[19]、Barndorff-Nielsen等[20,21]利用已实现核估计方法消除噪音, 提高积分波动率的精度. Jacod等[22]、Podolskij和Vetter[23,24]应

用前平均方法(pre-averaging)消除噪音. Xiu[25]提出了拟似然方法. 这些方法能很好地消去噪音影响, 但他们的分析是基于连续时间的对数资产价格过程X_t^*被噪音污染的情况.

因此, 在噪音环境下, 怎样通过离散样本$\{X_{t_i}\}_{i=0}^n\,(t_i=Ti/n,\, t_i\in[0,T])$去估计对数价格过程$X_t$的积分波动率(IV)是我们非常关注的问题, 即噪音环境下跳–扩散模型中积分波动率估计问题. 这个问题已有部分研究者所关注, 例如Fan和Wang[26]先利用小波方法除去跳, 然后使用文献[18]的多尺度技术消噪, 提出了积分波动率估计方法. Podolskij和Vetter[24]、Bos等[27]先应用前平均方法消噪, 然后使用多次幂过滤跳, 提出了积分波动率的另一种方法. Jing等[28]引入文献[9]的门限技术, 研究了积分波动率估计的门限前平均方法. 本文结合前人的结论, 研究了门限技术下多尺度方法的积分波动率的非参数估计, 并获得了收敛速度为$n^{-1/4}$的结果.

§2.　基本定义、假设及符号

假设 1　为了方便起见, 假定实际对数交易价格过程X_t为

$$X_t=X_t^c+X_t^d=\int_0^t\sigma_s\mathrm{d}B_s+\sum_{k=1}^{N_t}\gamma_k, \tag{7}$$

其中X_t^c是Itô过程, X_t^d是复合Poisson过程, 且相互独立. 进一步要求, N_t服从强度为$\lambda>0$的Poisson过程, 且$N_t<\infty$. 跳跃尺度γ_k是相互独立的中心化随机变量序列, 并与N_t相互独立, 且$\mathsf{E}(\gamma_k)=0$, $\mathsf{E}(\gamma_k^2)=\mathsf{E}(\gamma^2)<\infty$, $\mathsf{E}(\gamma_k^4)=\mathsf{E}(\gamma^4)<\infty$.

注记 2　(7)式与(2)式相比, (7)式中没有漂移项(即$\int_0^t\mu_s\mathrm{d}s$部分), 这样的假设仅仅归因于方便本文的理论研究, 对方法本身不失一般性.

假设 3　设Y_t为所观察到的数据, X_t为潜在的真实数据. 由于市场微观结构噪音的存在, 所以观察采集到的高频金融数据Y_t与潜在的真实数据X_t之间存在着一定程度的偏差, 则记t_i时刻对数价格的观测值为

$$Y_{t_i}=X_{t_i}+\varepsilon_{t_i}=X_{t_i}^c+X_{t_i}^d+\varepsilon_{t_i}=Y_{t_i}^*+X_{t_i}^d,\qquad t_i\in[0,T],\ i=0,1,\ldots,n,$$

其中ε_{t_i}是 i.i.d. 微观结构噪音, 满足$\mathsf{E}\varepsilon_{t_i}=\mathsf{E}\varepsilon=0$, $\mathsf{E}\varepsilon_{t_i}^2=\mathsf{E}\varepsilon^2$, $\mathsf{E}(\varepsilon_{t_i}^4)=\mathsf{E}(\varepsilon^4)<\infty$, 且$\varepsilon_{t_i}$与对数价格$X_{t_i}$相互独立.

定义 4　设观察数据Y_t分别在$0=t_0,t_1,\ldots,t_n=T$时刻取观测值, 则定义观测点全集

$$\mathscr{G}=\{t_0,t_1,\ldots,t_n\}, \tag{8}$$

其中$t_i=Ti/n$, $i=0,1,\ldots,n$ (等距离观测). 于是, 定义基于观测点全集$\mathscr{G}$的已实现波动率为

$$[Y,Y]^{(n,1)}=\sum_{i=1}^n(Y_{t_i}-Y_{t_{i-1}})^2=\sum_{t_i\in\mathscr{G},i\geqslant1}(Y_{t_i}-Y_{t_{i-1}})^2. \tag{9}$$

定义 5 设观察数据Y_t的观测点全集为$\mathscr{G}$, 将其划分为K个不相交子集$\mathscr{G}^{(k)}$, $k=1,2,\ldots,K$, 即$\mathscr{G}=\bigcup\limits_{k=1}^{K}\mathscr{G}^{(k)}$, 且当$k\neq l$时, $\mathscr{G}^{(k)}\cap\mathscr{G}^{(l)}=\emptyset$. 则第$k$个观测点子集是从第$k-1$个观测点开始, 每隔$K$个观测点取一次, 直到时刻$T$, 即

$$\mathscr{G}^{(k)}=\{t_{k-1},t_{k-1+K},t_{k-1+2K},\ldots,t_{k-1+n_kK}\},\tag{10}$$

其中$n_k=|\mathscr{G}^{(k)}|=(|\mathscr{G}|-K+1)/K=(n-K+1)/K$, 且$n_k$是一个使$t_{k-1+n_kK}$在子集$\mathscr{G}^{(k)}$内表示最后一个元素的正整数, 另外$\overline{n}_K=\sum\limits_{k=1}^{K}n_k/K=(n-K+1)/K$.

于是, 定义基于观测点子集$\mathscr{G}^{(k)}$的已实现波动率为

$$[Y,Y]^{(k)}=\sum_{j=1}^{n_k}(Y_{t_{k-1+jK}}-Y_{t_{k-1+(j-1)K}})^2=\sum_{t_{j,-},t_j\in\mathscr{G}^{(k)}}(Y_{t_j}-Y_{t_{j,-}})^2,\tag{11}$$

其中$k=1,2,\ldots,K$, 且如果$t_j\in\mathscr{G}^{(k)}$, 那么$t_{j,-}$表示在$\mathscr{G}^{(k)}$中t_j之前的那个元素. 进一步, 我们定义基于观测点子集$\mathscr{G}^{(k)}$ $(k=1,2,\ldots,K)$的平均已实现波动率

$$\begin{aligned}[Y,Y]^{(n,K)}&=\frac{1}{K}\sum_{k=1}^{K}[Y,Y]^{(k)}=\frac{1}{K}\sum_{k=1}^{K}\sum_{j=1}^{n_k}(Y_{t_{k-1+jK}}-Y_{t_{k-1+(j-1)K}})^2\\&=\frac{1}{K}\sum_{i=0}^{n-K}(Y_{t_{i+K}}-Y_{t_i})^2=\frac{1}{K}\sum_{t_i\in\mathscr{G},i\geqslant K}(Y_{t_i}-Y_{t_{i-K}})^2.\end{aligned}\tag{12}$$

在假设3下, $[Y,Y]^{(n,K)}$分解为

$$[Y,Y]^{(n,K)}=[X,X]^{(n,K)}+2[X,\varepsilon]^{(n,K)}+[\varepsilon,\varepsilon]^{(n,K)},\tag{13}$$

其中$[X,X]^{(n,K)}=K^{-1}\sum\limits_{t_i\in\mathscr{G},i\geqslant K}(X_{t_i}-X_{t_{i-K}})^2$, $[\varepsilon,\varepsilon]^{(n,K)}=K^{-1}\sum\limits_{t_i\in\mathscr{G},i\geqslant K}(\varepsilon_{t_i}-\varepsilon_{t_{i-K}})^2$和$[X,\varepsilon]^{(n,K)}=K^{-1}\sum\limits_{t_i\in\mathscr{G},i\geqslant K}(X_{t_i}-X_{t_{i-K}})(\varepsilon_{t_i}-\varepsilon_{t_{i-K}})$.

现在假定跳不存在, 则

$$Y_{t_{i+K}}-Y_{t_i}=Y^*_{t_{i+K}}-Y^*_{t_i}=X^c_{t_{i+K}}-X^c_{t_i}+\varepsilon_{t_{i+K}}-\varepsilon_{t_i}.$$

在布朗半秩模型下, 可知

$$(Y_{t_{i+K}}-Y_{t_i})\Big/\Big[\int_{t_i}^{t_{i+K}}\sigma_s^2\mathrm{d}s+2\mathsf{E}\varepsilon^2\Big]^{1/2}\sim\mathrm{N}(0,1).$$

于是, 我们可定义一个示性函数$I_Y^K(i,\eta)$:

$$I_Y^K(i,\eta)=\begin{cases}1, & (Y_{t_{i+K}}-Y_{t_i})^2\Big/\Big(\int_{t_i}^{t_{i+K}}\sigma_s^2\mathrm{d}s+2\mathsf{E}\varepsilon^2\Big)\leqslant\eta;\\ 0, & 其他.\end{cases}\tag{14}$$

借助文献[9]的门限技术思想, 利用示性函数$I_Y^K(i;\eta)$来截断带有潜在跳的收益, 从而构造噪音环境下跳–扩散模型中积分波动率的门限两尺度非参数估计量与门限多尺度非参数估计量. 下面我们介绍门限两尺度已实现波动率(TTSRV)和门限多尺度已实现波动率(TMSRV).

§3.　门限两尺度已实现波动率(TTSRV)

当跳不存在时, Zhang等[17]提出了积分波动率的双尺度已实现波动率估计量(简写TSRV*), 即

$$\mathrm{TSRV}^* = \Big(1-\frac{\overline{n}_K}{n}\Big)^{-1}\Big([Y^*,Y^*]^{(n,K)}-\frac{\overline{n}_K}{n}[Y^*,Y^*]^{(n,1)}\Big), \tag{15}$$

且在一定的条件下, 可获得收敛速度为$n^{-1/6}$, 即

$$\mathrm{TSRV}^*-\int_0^T\sigma_s^2\mathrm{d}s = O_{\mathrm{p}}(n^{-1/6}),$$

噪音方差$\mathsf{E}\varepsilon^2$的一个相合估计量: $\widehat{\mathsf{E}\varepsilon}_*^2=[Y^*,Y^*]^{(n,1)}/(2n)$.

当跳存在时, Dorfard[29]利用两尺度已实现波动率方法(记为TSRV)去估计二次变差$[X,X]_T$, 也可获得了收敛速度为$n^{-1/6}$的结果, 即

$$\mathrm{TSRV}-[X,X]_T = O_{\mathrm{p}}(n^{-1/6}),$$

其中

$$\mathrm{TSRV} = \Big(1-\frac{\overline{n}_K}{n}\Big)^{-1}\Big([Y,Y]^{(n,K)}-\frac{\overline{n}_K}{n}[Y,Y]^{(n,1)}\Big).$$

同时, 归因于估计量TSRV一致收敛于二次变差$[X,X]_T$, 估计量TSRV是高估积分波动率的. 考虑到$[Y,Y]^{(n,1)}$是二次变差$[X,X]_T$与$2n\mathsf{E}\varepsilon^2$之和的一个相合估计, 我们可得噪音方差$\mathsf{E}\varepsilon^2$的另一个相合估计量:

$$\widehat{\mathsf{E}\varepsilon}^2=\frac{1}{2n}([Y,Y]^{(n,1)}-\mathrm{TSRV}). \tag{16}$$

在研究二次变差估计时, 我们更关注积分波动率部分. 因此, 当跳存在时, 为了得到积分波动率的一致估计, 我们提出了门限两尺度已实现波动率(TTSRV). 重新定义$[Y,Y]^{(n,K)}$, 并表示为$\{Y,Y\}^{(n,K)}$:

$$\{Y,Y\}^{(n,K)}=\frac{\gamma_\eta}{K}\frac{\sum\limits_{i=0}^{n-K}(Y_{t_{i+K}}-Y_{t_i})^2\cdot I_Y^K(i,\eta)}{\dfrac{1}{n-K+1}\sum\limits_{i=0}^{n-K}I_Y^K(i,\eta)}, \tag{17}$$

其中$\gamma_\eta=F_{\chi_1^2}(\eta)/F_{\chi_3^2}(\eta)$是纠偏系数, 且$F_{\chi_m^2}$表示自由度为$m$的卡方分布.

注记 6 γ_η的产生归因于对门限截断的一个修正, 且定义为示性函数$I_Y^K(i;\eta)$的数学期望与一个自由度为1的卡方分布随机变差$\zeta\,(\zeta\in[0,\eta])$的数学期望之比. 在模拟与实证中, 我们可以假设$\eta=9$, 见文献[10].

于是, 我们定义门限两尺度已实现波动率(TTSRV):

$$\mathrm{TTSRV}=\Big(1-\frac{\overline{n}_K}{n}\Big)^{-1}\Big(\{Y,Y\}^{(n,K)}-\frac{\overline{n}_K}{n}\{Y,Y\}^{(n,1)}\Big). \tag{18}$$

注记 7 由于$\int_{t_i}^{t_{i+K}}\sigma_s^2\mathrm{d}s+2\mathsf{E}\varepsilon^2$在现实中是观测不到的(因为我们只能观测到过程$Y_t$), 所以在构造估计量(17)–(18)时存在让人误解的地方. 事实上, 当计算示性函数与估计量时, 我们使用(16)式去估计噪音方差. 对于$\int_{t_i}^{t_{i+K}}\sigma_s^2\mathrm{d}s$, 我们作近似计算:

$$\int_{t_i}^{t_{i+K}}\sigma_s^2\mathrm{d}s\approx\frac{t_{i+K}-t_i}{T}\cdot\widehat{\int_0^T}\sigma_s^2\mathrm{d}s. \tag{19}$$

如果日间波动率是强相依的(highly persistent), 此近似计算是有效的. 我们也可以采用其他方法来改善这个近似计算, 具体可参看文献[30].

对于$\widehat{\int_0^T\sigma_s^2}\mathrm{d}s$的计算, 我们采用一个迭代方法: 首先, 我们可利用文献[12]的估计量MedRV来初次估计$\int_0^T\sigma_s^2\mathrm{d}s$, 并代入(19)式得$\int_{t_i}^{t_{i+K}}\sigma_s^2\mathrm{d}s$的近似, 然后计算示性函数(14)与(17)式, 于是, 我们计算得到了门限两尺度已实现波动率(TTSRV)的初值. 然后, 利用初次得到的TTSRV来计算新的示性函数(14)与(17)式的值, 进而得到新的TTSRV值. 最后, 我们迭代这个过程, 获得新的门限值与TTSRV值, 直到没有更大的收益被截断为止.

下面我们给出估计量TTSRV的相合性.

定理 8 当$n\to\infty$, $K\to\infty$时, $n/K\to\infty$. 假定观察时刻点$\{t_i\}_{i=0}^n$是非随机的, 时间渐进二次变差$H(t)$存在, 且连续可微, 并满足$\min\limits_{0\leqslant t\leqslant T}H'(t)>0$. 设$K/n^{2/3}\to c\,(n\to\infty)$, 在假设1和3下, 则

$$\mathrm{TTSRV}\xrightarrow{\mathrm{p}}\int_0^T\sigma_s^2\mathrm{d}s,\qquad \text{并且 }\mathrm{TTSRV}-\int_0^T\sigma_s^2\mathrm{d}s=O_{\mathrm{p}}(n^{-1/6}).$$

证明: 重新定义Y真实过程为$Y=X^c+\varepsilon+J=Y^*+J$, 其中$X^c=\int_0^t\sigma_s\mathrm{d}B_s$和$Y^*=X^c+\varepsilon$.

如果$J=0$, 在假设1和3下, 根据文献[17]的定理4, 我们有

$$\mathrm{TSRV}^*\xrightarrow{\mathrm{p}}\int_0^T\sigma_s^2\mathrm{d}s,\qquad \text{并且 }\mathrm{TSRV}^*-\int_0^T\sigma_s^2\mathrm{d}s=O_{\mathrm{p}}(n^{-1/6}).$$

如果$J\neq0$, 因为J是一个有限跳过程(finite activity jump process), 并且X^c是一个独立于噪音ε的布朗半秩. 基于文献[9]的定理1和文献[10]的定理1, 对于足够小的δ, 我们有

$$(Y_{t_{i+K}}-Y_{t_i})^2\cdot I_Y^K(i,\eta)=(Y^*_{t_{i+K}}-Y^*_{t_i})^2-I^*_{i,K}\cdot(Y^*_{t_{i+K}}-Y^*_{t_i})^2,$$

其中$I_{i,K}^*$是一个示性函数, 且当时间区间$[t_i, t_{i+K}]$内存在跳时, $I_{i,K}^*=1$, 当其他时, $I_{i,K}^*=0$. 不考虑计算式中偏差校正系数下, 我们可得

$$\{Y,Y\}^{(n,K)}=\frac{1}{K}\sum_{i=0}^{n-K}(Y_{t_{i+K}}^*-Y_{t_i}^*)^2-\frac{1}{K}\sum_{i=0}^{n-K}I_{i,K}^*\cdot(Y_{t_{i+K}}^*-Y_{t_i}^*)^2,$$

因此, 利用布朗运动的连续模性质(the modulus of continuity of the Brownian motion), 且N_T表示区间$[0,T]$内出现跳的个数, 且有限($N_T<+\infty$), 于是, 我们有

$$\{Y,Y\}^{(n,K)}-[Y^*,Y^*]^{(n,K)}=O_{\mathrm{p}}\Big(\frac{N_t}{K}(\delta\ln|\delta|+1)\Big).$$

依据上等式, 我们可知

$$\begin{aligned}\mathrm{TTSRV}-\mathrm{TSRV}^* &=\Big(1-\frac{\overline{n}_K}{n}\Big)^{-1}\Big(\{Y,Y\}^{(n,K)}-\frac{\overline{n}_K}{n}\{Y,Y\}^{(n,1)}\Big)\\ &\quad-\Big(1-\frac{\overline{n}_K}{n}\Big)^{-1}\Big([Y^*,Y^*]^{(n,K)}-\frac{\overline{n}_K}{n}[Y^*,Y^*]^{(n,1)}\Big)\\ &=\Big(1-\frac{\overline{n}_K}{n}\Big)^{-1}\Big((\{Y,Y\}^{(n,K)}-[Y^*,Y^*]^{(n,K)})\\ &\quad-\frac{\overline{n}_K}{n}\cdot(\{Y,Y\}^{(n,1)}-[Y^*,Y^*]^{(n,1)})\Big)\\ &=o_{\mathrm{p}}(1).\end{aligned}$$

因此, 我们证明了下式成立

$$\mathrm{TTSRV}-\mathrm{TSRV}^*=o_{\mathrm{p}}(1),$$

即TTSRV与TSRV*具有相同的极限性质.　□

注记 9　如果噪音序列是相依的, Aït-Sahalia等[31]证明了TSRV*是有偏的. 于是他们提出了一个修正估计方法, 即在估计量TSRV*中, 用$[Y,Y]^{(n,J)}$取代$[Y,Y]^{(n,1)}$, 并表示为

$$\mathrm{TSRV}^{**}=\Big(1-\frac{\overline{n}_K}{\overline{n}_J}\Big)^{-1}\Big([Y,Y]^{(n,K)}-\frac{\overline{n}_K}{\overline{n}_J}[Y,Y]^{(n,J)}\Big),$$

其中$\overline{n}_J=(n-J+1)/J$, J充分大于1, 且$K>J$. 借助文献[31]的方法, 我们可以提出相依噪音下门限两尺度已实现波动率(TTSRV*), 即

$$\mathrm{TTSRV}^*=\Big(1-\frac{\overline{n}_K}{\overline{n}_J}\Big)^{-1}\Big(\{Y,Y\}^{(n,K)}-\frac{\overline{n}_K}{\overline{n}_J}\{Y,Y\}^{(n,J)}\Big). \tag{20}$$

§4.　门限多尺度已实现波动率(TMSRV)

当跳不存在时, 为了提高收敛速度与精度, Zhang[18]推广了两尺度已实现波动率估计方法, 对积分波动率提出了多尺度已实现波动率估计量(简写MSRV*). 多尺度已实现波动率估计量(MSRV*)的定义如下:

$$\mathrm{MSRV}^*=\sum_{i=1}^{M}\alpha_i[Y^*,Y^*]^{(n,K_i)}, \tag{21}$$

其中$M > 2$. 为了保证MSRV*无偏估计$\int_0^T \sigma_s^2 \mathrm{d}s$, 要求

$$\sum_{i=1}^{M} \alpha_i = 1 \qquad 和 \qquad \sum_{i=1}^{M} \alpha_i \frac{n+1-K_i}{K_i} = 0. \tag{22}$$

若令$a_1 = \alpha_1 - [(n+1)(1/K_1 - 1/K_2)]^{-1}$, $a_2 = \alpha_2 - (a_1 - \alpha_1)$, $a_i = \alpha_i$ $(i \geqslant 3)$, 则条件(22)等价于

$$\sum_{i=1}^{M} a_i = 1 \qquad 和 \qquad \sum_{i=1}^{M} \frac{a_i}{K_i} = 0, \tag{23}$$

那么多尺度已实现波动率(MSRV*)可改写为

$$\mathrm{MSRV}^* = \sum_{i=1}^{M} a_i [Y^*, Y^*]^{(n,K_i)}. \tag{24}$$

这里需要说明的是a_i, M, K_i依靠于n, 因此它们也可以表示为$a_{n,i}$, M_n, $K_{n,i}$.

当跳存在时, 多尺度已实现波动率估计量(缩写为MSRV)可作为二次变差$[X,X]_T$的一个相合估计, 并且在一定的假设下, 可知

$$\mathrm{MSRV} - [X,X]_T = O_{\mathrm{p}}(n^{-1/4}),$$

其中

$$\mathrm{MSRV} = \sum_{i=1}^{M} a_i [Y,Y]^{(n,K_i)}.$$

注记 10 此结论的证明将另文给出, 此处忽略.

为了消去跳对积分波动率估计的影响与提高收敛速度, 我们引入门限技术来改进多尺度已实现波动率估计量MSRV, 从而构造新的积分波动率估计量. 我们把它称为门限多尺度已实现波动率, 并用TMSRV来表示:

$$\mathrm{TMSRV} = \sum_{i=1}^{M} a_i \cdot \{Y,Y\}^{(n,K_i)}, \tag{25}$$

其中$\{Y,Y\}^{(n,K_i)}$由公式(17)定义. 进一步, 我们有下列结论.

首先, 我们介绍引理11.

引理 11 设$K_{n,1} = O(1)$, $K_{n,2} = O(1)$ (当$n \to \infty$). 在假设1和3下, 则

$$\mathrm{MSRV} = \sum_{i=1}^{M} a_i [Y,Y]^{(n,K_i)} + 2\mathsf{E}\varepsilon^2 + O_{\mathrm{p}}(n^{-1/2}).$$

证明:

$$\mathrm{MSRV} = \sum_{i=1}^{M} a_i [Y,Y]^{(n,K_i)} + (\alpha_1 - a_1)([Y,Y]^{(n,K_1)} - [Y,Y]^{(n,K_2)}),$$

其中$[Y,Y]^{(n,K_i)}=[X,X]^{(n,K_i)}+2[X,\varepsilon]^{(n,K_i)}+[\varepsilon,\varepsilon]^{(n,K_i)}$, $i=1,2$.

在假设1和3下, $[X,X]^{(n,K_i)}=O_{\mathrm{p}}(1)$, $i=1,2$. 根据文献[17]的引理A.2和定理A.1, 我们可得

$$[X,\varepsilon]^{(n,K_i)}=O_{\mathrm{p}}\Big(\frac{1}{\sqrt{K_i}}\Big) \qquad 和 \qquad [\varepsilon,\varepsilon]^{(n,K_i)}=2\overline{n}_i\mathsf{E}\varepsilon^2+O_{\mathrm{p}}\Big(\frac{\sqrt{n}}{K_i}\Big),$$

其中$\overline{n}_i=(n-K_i+1)/K_i$, $i=1,2$. 于是

$$\begin{aligned}\mathrm{MSRV}&=\sum_{i=1}^{M}a_i[Y,Y]^{(n,K_i)}+\Big[(n+1)\Big(\frac{1}{K_1}-\frac{1}{K_2}\Big)\Big]^{-1}[2(\overline{n}_1-\overline{n}_2)\mathsf{E}\varepsilon^2]+O_{\mathrm{p}}(n^{-1/2})\\&=\sum_{i=1}^{M}a_i[Y,Y]^{(n,K_i)}+2\mathsf{E}\varepsilon^2+O_{\mathrm{p}}(n^{-1/2}).\end{aligned}$$

定理 12　设a_i满足(23)式, 并且当$n\to\infty$, $M\to\infty$时, $M=o(n)$, $M^3/n\to\infty$, $\sum_{i=1}^{M}a_i=o(-(n/M)^{1/2})$, $\sum_{i=1}^{M}a_i/K_i=o(M^{-3})$, 且$\max\limits_{1\leqslant i\leqslant M}\|a_i/(i\cdot\gamma_n)\|\to 0$. 假定观察时刻点$\{t_i\}_{i=0}^{n}$是非随机的, 时间渐进二次变差$H(t)$存在, 且连续可微, 满足$\min\limits_{0\leqslant t\leqslant T}H'(t)>0$. 设$M/n^{1/2}\to c\,(n\to\infty)$, 在假设1和3下, 则

$$\mathrm{TMSRV}\xrightarrow{\mathrm{p}}\int_0^T\sigma_s^2\mathrm{d}s, \qquad 并且\ \mathrm{TMSRV}-\int_0^T\sigma_s^2\mathrm{d}s=O_{\mathrm{p}}(n^{-1/4}).$$

证明:　类似定理8的证明, 可知, 对于足够小的δ与适当选取序列$\{K_i\}_{i=1}^{M}$, 我们有

$$\big(Y_{t_{i+K_i}}-Y_{t_i}\big)^2\cdot I_Y^{K_i}(i,\eta)=\big(Y^*_{t_{i+K_i}}-Y^*_{t_i}\big)^2-I^*_{i,K_i}\cdot\big(Y^*_{t_{i+K_i}}-Y^*_{t_i}\big)^2,$$

其中I^*_{i,K_i}是一个示性函数, 且当时间区间$[t_i,t_{i+K_i}]$内存在跳时, $I^*_{i,K_i}=1$, 当其他时, $I^*_{i,K_i}=0$. 利用布朗运动的连续模性质, 且N_T表示区间$[0,T]$内出现跳的个数, 且有限($N_T<+\infty$), 于是, 我们有

$$\{Y,Y\}^{(n,K_i)}-[Y^*,Y^*]^{(n,K_i)}=O_{\mathrm{p}}\Big(\frac{N_t}{K_i}(\delta\ln|\delta|+1)\Big).$$

因此, 依据条件(23)与定理8的假设下, 我们可知

$$\begin{aligned}\mathrm{TMSRV}-\mathrm{MSRV}^*&=\sum_{i=1}^{M}a_i\cdot\{Y,Y\}^{(n,K_i)}-\sum_{i=1}^{M}a_i[Y^*,Y^*]^{(n,K_i)}\\&=\sum_{i=1}^{M}a_i\cdot\big(\{Y,Y\}^{(n,K_i)}-[Y^*,Y^*]^{(n,K_i)}\big)=o_{\mathrm{p}}(1).\end{aligned}$$

即TMSRV与MSRV*具有相同的极限性质. 证毕.　□

注记 13　鉴于篇幅限制, 本文未给出相关的模拟与实证研究, 但这方面的比较研究我们将另文给出.

参 考 文 献

[1] Merton R C. Option pricing when underlying stock returns are discontinuous [J]. *J. Financ. Econ.*, 1976, **3(1-2)**: 125–144.

[2] Andersen T G, Bollerslev T, Diebold F X, et al. The distribution of realized exchange rate volatility [J]. *J. Amer. Statist. Assoc.*, 2001, **96(453)**: 42–55.

[3] Andersen T G, Bollerslev T, Diebold F X, et al. Modeling and forecasting realized volatility [J]. *Econometrica*, 2003, **71(2)**: 579–625.

[4] Barndorff-Nielsen O E, Shephard N. Econometric analysis of realized volatility and its use in estimating stochastic volatility models [J]. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 2002, **64(2)**: 253–280.

[5] Barndorff-Nielsen O E, Shephard N. Estimating quadratic variation using realized variance [J]. *J. Appl. Econometrics*, 2002, **17(5)**: 457–477.

[6] Barndorff-Nielsen O E, Shephard N. Power and bipower variation with stochastic volatility and jumps [J]. *J. Financ. Economet.*, 2004, **2(1)**: 1–37.

[7] Andersen T G, Bollerslev T, Diebold F X. Roughing it up: including jump components in the measurement, modeling, and forecasting of return volatility [J]. *Rev. Econom. Statist.*, 2007, **89(4)**: 701–720.

[8] Barndorff-Nielsen O E, Shephard N. Econometrics of testing for jumps in financial economics using bipower variation [J]. *J. Financ. Economet.*, 2006, **4(1)**: 1–30.

[9] Mancini C. Non-parametric threshold estimation for models with stochastic diffusion coefficient and jumps [J]. *Scand. J. Stat.*, 2009, **36(2)**: 270–296.

[10] Corsi F, Pirino D, Renò R. Threshold bipower variation and the impact of jumps on volatility forecasting [J]. *J. Econometrics*, 2010, **159(2)**: 276–288.

[11] Veraart A E D. How precise is the finite sample approximation of the asymptotic distribution of realised variation measures in the presence of jumps? [J]. *AStA Adv. Stat. Anal.*, 2011, **95(3)**: 253–291.

[12] Andersen T G, Dobrev D, Schaumburg E. Jump-robust volatility estimation using nearest neighbor truncation [J]. *J. Econometrics*, 2012, **169(1)**: 75–93.

[13] Zhou B. High-frequency data and volatility in foreign-exchange rates [J]. *J. Bus. Econom. Statist.*, 1996, **14(1)**: 45–52.

[14] Aït-Sahalia Y, Mykland P A, Zhang L. How often to sample a continuous-time process in the presence of market microstructure noise [J]. *Rev. Financ. Stud.*, 2005, **18(2)**: 351–416.

[15] Bandi F M, Russell J R. Separating microstructure noise from volatility [J]. *J. Financ. Econ.*, 2006, **79(3)**: 655–692.

[16] Bandi F M, Russell J R. Microstructure noise, realized variance, and optimal sampling [J]. *Rev. Econom. Stud.*, 2008, **75(2)**: 339–369.

[17] Zhang L, Mykland P A, Aït-Sahalia Y. A tale of two time scales: determining integrated volatility with noisy high-frequency data [J]. *J. Amer. Statist. Assoc.*, 2005, **100(472)**: 1394–1411.

[18] Zhang L. Efficient estimation of stochastic volatility using noisy observations: a multi-scale approach [J]. *Bernoulli*, 2006, **12(6)**: 1019–1043.

[19] Hansen P R, Lunde A. Realized variance and market microstructure noise [J]. *J. Bus. Econom. Statist.*, 2006, **24(2)**: 127–161.

[20] Barndorff-Nielsen O E, Hansen P R, Lunde A, et al. Realized kernels in practice: trades and quotes [J]. *Econom. J.*, 2009, **12(3)**: C1–C32.

[21] Barndorff-Nielsen O E, Hansen P R, Lunde A, et al. Subsampling realised kernels [J]. *J. Econometrics*, 2011, **160(1)**: 204–219.

[22] Jacod J, Li Y Y, Mykland P A, et al. Microstructure noise in the continuous case: the pre-averaging approach [J]. *Stochastic Process. Appl.*, 2009, **119(7)**: 2249–2276.

[23] Podolskij M, Vetter M. Bipower-type estimation in a noisy diffusion setting [J]. *Stochastic Process. Appl.*, 2009, **119(9)**: 2803–2831.

[24] Podolskij M, Vetter M. Estimation of volatility functionals in the simultaneous presence of microstructure noise and jumps [J]. *Bernoulli*, 2009, **15(3)**: 634–658.

[25] Xiu D C. Quasi-maximum likelihood estimation of volatility with high frequency data [J]. *J. Econometrics*, 2010, **159(1)**: 235–250.

[26] Fan J Q, Wang Y Z. Multi-scale jump and volatility analysis for high-frequency financial data [J]. *J. Amer. Statist. Assoc.*, 2007, **102(480)**: 1349–1362.

[27] Bos C S, Janus P, Koopman S J. Spot variance path estimation and its application to high-frequency jump testing [J]. *J. Financ. Economet.*, 2012, **10(2)**: 354–389.

[28] Jing B Y, Liu Z, Kong X B. On the estimation of integrated volatility with jumps and microstructure noise [J]. *J. Bus. Econom. Statist.*, 2014, **32(3)**: 457–467.

[29] Dorfard A. Estimating the quadratic variation of Poisson jump diffusion processes with noisy high-frequency data [D/OL]. Frankfurt: Goethe University Frankfurt, 2007. http://www.math.uni-frankfurt.de/~fmfi/diplom_dorfard.pdf.

[30] Boudt K, Croux C, Laurent S. Robust estimation of intraweek periodicity in volatility and jump detection [J]. *J. Empir. Financ.*, 2011, **18(2)**: 353–367.

[31] Aït-Sahalia Y, Mykland P A, Zhang L. Ultra high frequency volatility estimation with dependent microstructure noise [J]. *J. Econometrics*, 2011, **160(1)**: 160–175.

Nonparametric Estimation of the Integrated Volatility of Jump-Diffusion Processes with Noisy High-Frequency Data

YE Xuguo[1,2] LIN Jinguan[1]

([1]*Department of Mathematics, Southeast University, Nanjing, 210096, China*)

([2]*School of Mathematical Sciences, Kaili University, Kaili, 556011, China*)

Abstract: This paper studies nonparametric estimation of the integrated volatility of Poisson jump-diffusion processes with noisy high-frequency data. We propose jump-robust two-scale and multi-scale estimators. The estimators are based on a combination of the multi-scale method and threshold technique, which serves to remove microstructure noise and jumps, respectively. Furthermore, asymptotic properties of the proposed estimators, such as consistency, are established.

Keywords: nonparametric estimation; threshold technique; market microstructure noise; jump-diffusion processes; integrated volatility

2010 Mathematics Subject Classification: 62G05; 62M05