

文章编号:1003-207(2015)10-0019-11

DOI:10.16381/j.cnki.issn1003-207x.2015.10.003

高频波动率矩阵估计的比较分析

——基于有噪非同步的金融数据

赵树然¹, 姜亚萍¹, 任培民²

(1. 中国海洋大学经济学院金融系, 山东青岛 266100; 2. 青岛大学经济学院金融系, 山东青岛 266071)

摘 要:基于高频数据的波动率矩阵估计可有效解决传统低频估计面临的种种瓶颈问题。然而,由于受非同步和微观结构噪声等的影响,传统的高频波动率矩阵估计会产生艾普斯效应,并偏离其理论值。本文主要考虑非同步逐笔高频数据的三种同步化方法和五种传统已实现波动率矩阵的纠偏降噪方法,并从数值模拟和沪深股市的实证分析两个角度对两类方法分别展开了全面深入的比较研究。结果表明:更新时间同步化法最大程度地保留了数据信息,传统未纠偏的已实现波动率矩阵具有艾普斯效应,其偏差较大,多变量已实现核估计、双频已实现波动率矩阵估计、调整的已实现波动率矩阵估计的纠偏降噪效果较好,事先平均 HY 估计和 HY 估计相对表现较差。研究结果可为相关领域工作者进一步的研究与应用提供方法上的参考与指导。

关键词:波动率矩阵;高频数据;微观结构噪声;非同步交易

中图分类号:F830.9

文献标识码:A

1 引言

波动率矩阵作为金融资产价格变化的一种统计度量,是资产配置、现代资产定价以及金融风险管理等重要金融理论和活动的基础与核心要素。早在 1952 年马克维茨的资产组合选择模型中,波动率矩阵就作为二维分析框架中的一维扮演了十分重要的角色;当前,随着计算机技术快速发展及相关理论的不完善,波动率矩阵的估计逐渐成为现代金融研究的一个极其重要的研究领域。

概括讲,常用的波动率矩阵估计方法根据数据类型可分为三类:一是基于日度等低频收益数据实现,如常见的样本协方差阵^[21],二是由期权价格和 BS 期权定价公式推算出的隐含波动率矩阵,三是基于日内等高频数据实现的高频波动率矩阵,如传统的已实现波动率矩阵。第一种方法考虑到样本协方

差计算的准确性,其计算往往需基于较长的时间跨度,通常是资产维度越高,时间跨度应越长;然而,当今的投资者们往往需要的是能够准确反映其资产组合在相对短期内的风险测度,如一天的风险;显然,该估计方法在体现风险的动态性方面表现较弱,这种估计通常视为静态估计。第二种方法适用于波动率矩阵中的各资产均存在相应的期权产品方可,这对于尚未推出期权产品的我国,该方法是不适应的。相比于前两种方法,高频波动率矩阵具有如下显著优势:首先,高频波动率矩阵为关注短期动态风险的投资者们提供了一个有力工具;其次,高频波动率矩阵估计方法视波动率为显性变量,不需要具体的模型结构和参数估计就可直接对当期原本不可观测的波动率进行相合估计,从而实现波动率矩阵的可观测性;第三,从高频波动率矩阵出发建立的波动率矩阵模型由于其以“可观测”的高频波动率矩阵,直接分析、建模和预测波动率矩阵,而具有比较灵活且易于实现的明显优势;第四,相比与隐性的 GARCH 类模型^[20],从高频波动率矩阵出发建立的波动率矩阵模型能够较好地解决高维波动率矩阵模型中的维数灾难问题。

为得到有效、正定的高频波动率矩阵,高频数据的抽样频率应随资产维度的增加而增加。然而,随

收稿日期:2014-03-03; 修订日期:2015-01-06

基金项目:国家自然科学基金项目(71201147, 71103165);教育部人文社会科学研究青年基金(12YJC630161);全国统计科学研究项目(2014511)

作者简介:赵树然(1978-),女(汉族),湖南人,中国海洋大学经济学院副教授,博士,研究方向:金融计量与金融风险

管理。

着抽样频率的增加,高频波动率矩阵估计会面临两个重要难题:一是数据的非同步性问题,即不同资产的交易发生在不同时间,这是由不同资产具有不同的交易频率造成的。二是微观结构噪声问题,如离散抽样、买卖报价价差等。非同步性易引起资产统计意义上的虚假现象,典型的如 Epps^[8]在股票市场上发现的艾普斯效应(Epps Effect),即基于股市非同步交易数据计算的样本相关系数随着抽样频率的增加,而趋向于 0。Guillaume 等人^[10]和 Muthuswamy 等人^[14]在汇率市场也发现了相同的效应。微观结构噪声易导致市场价格偏离资产的有效价格,并且频率越高,这种偏离造成的影响越大。如何对多维非同步高频数据进行同步化,以及降低微观结构噪声的影响是实现波动率矩阵有效估计的关键。本文对近几年刚刚兴起的高频非同步金融数据的同步化技术、以及基于此的旨在纠偏降噪的高频波动率矩阵估计展开深入、全面的比较分析,具体包括理论分析、数值模拟比较、实证比较分析等,为国内相关学者和业界将各种方法运用于我国金融市场的实证分析提供有益参考。

2 重要概念及基础假设

考虑 p 维对数价格过程 $X = (X^{(1)}, X^{(2)}, \dots, X^{(p)})$, 其观测时间介于 $[0, T]$ 。为便于展开, 后文记 $T = 1$ 。对于第 i 个资产, 记其观测时间为 $t_1^{(i)}, t_2^{(i)}, \dots, t_{N(i)}^{(i)}$, 并且假设该时间序列严格递增, 同时, 记 $X^{(i)}$ 在该观测时间序列的各实现为 $X^{(i)}(t_j^{(i)})$ 。假设观测价格过程 X 由如下有效对数价格过程 Y 驱动:

$$Y(t) = \int_0^t a(u) du + \int_0^t \sigma(u) dW(u),$$

其中, a 是可测的局部有界漂移向量过程, σ 是 $p \times p$ 的即时波动率矩阵过程, $W(u)$ 为 p 维布朗运动。上述过程称为布朗半鞅过程或带漂移的随机波动率过程。通常, 假设观测价格是在有效对数价格 Y 的基础上叠加市场微观结构噪声形成, 即:

$$X^{(i)}(t_j^{(i)}) = Y^{(i)}(t_j^{(i)}) + U_j^{(i)}, j = 0, 1, \dots, N^{(i)},$$

其中, $U_j^{(i)}$ 代表噪声对观测价格的影响, 通常假设其期望为 0, 且协方差平稳。

称 $\sum(t) = \sigma_t \sigma_t'$ 为瞬时波动率矩阵, 并称其积分形式 $[Y] = \int_0^1 \sum(u) du$ 为积分波动率矩阵, 它反映了价格过程 Y 在 $[0, 1]$ 时间段内的总体变化情况, 在金融计量分析中占据重要位置^[9], 也是本文待研究的各估计量的估计对象。

如果观测价格是同步交易时, 易证明随最大抽样时间间隔 $\max\{\tau_{i+1} - \tau_i, i = 1, \dots, n-1\}$ 趋于 0, 基于有效价格的样本协变差矩阵和 $[Y, Y] = \sum(Y(\tau_j) - Y(\tau_{j-1}))(Y(\tau_j) - Y(\tau_{j-1}))'$ 收敛于 $[Y]$ ^[15] 和 Jacod 和 Shiryaev^[12], 这里, τ_j 为同步时间。因此, 许多研究者根据该性质利用已实现波动率矩阵 RC :

$$[X, X] = \sum (X(\tau_j) - X(\tau_{j-1}))(X(\tau_j) - X(\tau_{j-1}))'$$

来估计积分波动率矩阵。但是, 在实际运用中, 最大抽样时间间隔的收敛, 意味着要使用高频数据, 而实际高频数据通常面临非同步性和微观结构噪声问题。因此, 如何对数据进行同步化, 以及对已实现协方差矩阵进行纠偏降噪就成为一个非常现实的问题。

下文若未特别说明, 记 τ_i 为第 i 个同步时间, n 为同步时间数, $x^{(i)}(t_j^{(i)}) = X^{(i)}(t_j^{(i)}) - X^{(i)}(t_{j-1}^{(i)})$ 为第 i 个资产在 $(t_{j-1}^{(i)}, t_j^{(i)})$ 内的对数收益率, $x(\tau_j) = (x^{(1)}(\tau_j), \dots, x^{(p)}(\tau_j))$ 为 $(\tau_{j-1}, \tau_j]$ 内的对数收益率向量。

3 非同步高频数据的同步化方法及其比较

理论上, 在有效的证券市场上, 每只证券的价格都应迅速反应影响整体市场的信息。但现实中, 由于交易摩擦的存在、信息到达不同投资者时间的差异等因素, 使得证券对信息的反应不一, 从而导致同一证券市场上不同证券具有不同的交易频率, 甚至是单一证券, 其交易强度也会随时间和日期的变化而变化。

对于日内高频数据, 非同步交易会导致典型的艾普斯效应。该效应导致组合的风险(协方差)被低估。直观地, 随着抽样频率的增加, 非同步交易会导致越来越多的零回报, 这些零回报在已实现协方差及其相关统计量中逐渐占主导地位, 并致使这些估计量逐渐趋向于零。因此, 利用非同步高频数据估计波动率矩阵时, 必须事先对数据进行同步化处理。

目前较常用的多元高频数据同步化方法主要是删除法, 即通过删除部分数据, 并借助某种价格替代技术实现同步目标。应用比较广泛的主要有如下三种, 其中, 以后两种的使用为最常见。

(1) 时间区间同步法。Hayashi 和 Yoshida^[11]最早对标准形式的已实现协方差, 即 $[X, X]$ 的非对角元进行了修正, 提出了 HY 估计量(以 (i, j) 非对角元为例):

$$\langle X^{(i)}, X^{(j)} \rangle_{HY} = \sum_{k,s} (X^{(i)}(t_k^{(i)}) - X^{(i)}(t_{k-1}^{(i)})) \\ (X^{(j)}(t_s^{(j)}) - X^{(j)}(t_{s-1}^{(j)})) \times I((t_{k-1}^{(i)}, t_k^{(i)}) \cap (t_{s-1}^{(j)}, t_s^{(j)})),$$

其中, I 是示性函数。显然, HY 的计算过程, 隐含着数据的同步化方法, 其同步化过程实质是时间区间的同步化, 即将原来非完全重叠的时间区间视为完全重叠。但是, 当一个大区间同时嵌套多个小区间时(这些小区间必然相连), 同步化过程则是这个大区间与几个小区间合并而成的大区间进行同步。因而, 该种同步化技术造成合并区间内交易数据携带信息的损失。

(2) 前值法。Zhang Lan^[17] 的前值法重在构建抽样时间上各资产价格的选取原则。其以各资产上一最近时间的价格作为抽样时间的价格, 而抽样时间的间隔可以是等间隔, 如常见的 5 分钟间隔, 也可能是依赖于数据观测时间的时变间隔, 如由下文的更新时间法确定的时间间隔。除上一最近价格可选外, AïT-SAHALIA 等人^[22] 的广义抽样时间法进一步将其推广到上一抽样时间段内的任意一点。

(3) 更新时间法。在 Barndorff-Nielsen 等人^[4] 的更新时间法中, 同步化后的样本数及同步时间(即抽样时间)通常是由流动性相对最差资产的价格序列决定, 其中, 同步时间公式为:

$$\tau_1 = \max(t_1^{(1)}, \dots, t_1^{(p)}), \quad \tau_{j+1} = \max(t_{N^{(1)}(\tau_j)+1}^{(1)}, \dots, t_{N^{(p)}(\tau_j)+1}^{(p)})$$

这里, $N^{(i)}(\tau_j) = \#\{t_k^{(i)} \leq \tau_j, 1 \leq k \leq N(i)\}$ 为截止到时间 τ_j 第 i 个资产的观测样本数。图 1 以三资产为例, 说明更新时间法中各同步时间的确定。在图中, 实点为实际交易时间, 虚线为基于更新时间法的同步时间。

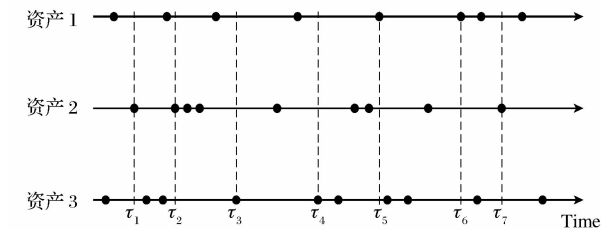


图 1 三资产的更新时间图(Refresh Time)

各同步时间点对应的各资产价格可依据前值法确定。上述同步化过程保证了相邻两个抽样时间点之间各资产至少发生一次交易, 且流动性最差资产最多发生一次。显然, 该确保原则导致其他资产在各抽样时间上的实际交易总是领先于流动性相对较

差资产实际交易的发生。而这种隐含的领先效应有可能导致波动率矩阵估计产生偏差。更新时间法是前值法框架下保留数据最多的一类同步化技术。

4 非同步、微观结构噪声下高频波动率矩阵估计的降噪方法及其比较

标准形式的已实现波动率矩阵 RC 是多元波动率矩阵估计中最为常用的估计。当观测价格为同步的有效价格时, 该估计是无偏和一致的。然而, 受交易成本和交易摩擦等市场微观结构噪声的影响, 观测价格通常偏离有效价格, 从而使得 RC 估计随抽样频率的增加而收敛于零矩阵, 产生严重的偏差。因此, 在利用上述高频数据同步化技术对原始高维高频数据进行同步处理的基础上, 探讨高频波动率矩阵的纠偏降噪技术是解决非同步和微观结构噪声条件下高频波动率矩阵估计偏差和低效的主要途径。

目前, 纠偏降噪技术主要有两种。一是在频率选择与噪声影响之间进行某种平衡, 其主要思想是通过合适的频率选择一方面尽可能降低噪声干扰, 另一方面尽可能多使用数据, 如 Zhang Lan^[17] 以均方误差为优化目标, 给出最优同步抽样频率的封闭解。二是对由于噪声影响而导致有偏的估计进行纠偏, 从而可以在即使同步数据频率很高的情况下, 也能得到性质良好的高频波动率矩阵估计。目前, 绝大部分的高频数据纠偏降噪技术属于这一种。下文就第二类纠偏方法展开讨论。

(1) 多变量已实现核估计 MRK。Barndorff-Nielsen 等人^[4] 利用更新时间法对原始高频数据同步化后, 将 Barndorff-Nielsen 等人^[2] 的一维已实现核估计的纠偏技术扩展到多维情形:

$$MRK = \sum_{h=-\tilde{n}}^{\tilde{n}-1} k(h/H) \Gamma_h$$

其中, 当 $h \geq 0$ 时, $\Gamma_h = \sum_{j=h+1}^{\tilde{n}} x_j x_{j-h}$, 当 $h < 0$ 时, $\Gamma_h = \Gamma_{-h}$; $k(\cdot)$ 为 $[0, 1]$ 上的核函数, 且 $k(0) = 1, k'(0) = 0; x_j = X_j - X_{j-1}, j = 1, 2, \dots, \tilde{n}, \tilde{n} = n - 2m + 1, X_j = X(\tau_j), j = 1, 2, \dots, \tilde{n} - 1, X_0 = \sum_{i=1}^m X(\tau_i)/m, X_{\tilde{n}} = \sum_{i=1}^m X(\tau_{n-m+i})/m$ 。

为得到相合的 MRK, 上式就 $X(\tau_j)$ 的首尾两端进行了 m 长度的平均。MRK 具有渐近混合正态性、一致性和噪声稳健性, 并适用于具有内生性且序列相关的噪声。其最优收敛速度为 $\tilde{n}^{-1/5}$, 略慢于一维情形时的 $\tilde{n}^{-1/4}$ 。该估计的一致性受核函数选取的影响, 并且, 面临敏感参数窗宽 H 的选取。一般地,

在均方误差最小原则下,窗宽取为 $H \propto \tilde{n}^{3/5}$ 。

关于核函数的选取,Barndorff-Nielsen 等人^[4]在综合考虑核函数的相对渐近效及已实现交叉协方差阵计算量的基础上,最终选取了如下 Parzen 型核函数:当 $x \in [0, 0.5]$ 时, $k(x) = 1 - 6x^2 + 6x^3$; 当 $x \in [0.5, 1]$ 时, $k(x) = 2(1 - x)^3$; 但 $x \in [1, +\infty]$ 时, $k(x) = 0$ 。显然, Parzen 型估计只需计算 H 个已实现交叉协方差阵, 计算量大大缩减。因此, 在后文的数值模拟和实证分析中, 我们将选用 Parzen 核函数。

(2) 双频已实现波动率矩阵估计 TSCV。Zhang Lan^[17]将 Zhang Lan 等人^[18]提出的一维双频已实现波动率估计推广到多元情形, 得到相合的 TSCV 估计(以矩阵 (i, j) 非对角元为例):

$$\text{TSCV}(i, j) = [X^{(i)}, X^{(j)}]_T^{(K)} - \frac{\tilde{n}_K}{n_j} [X^{(i)}, X^{(j)}]_T^{(J)}$$

其中, $[X^{(i)}, X^{(j)}]_T^{(K)} = \sum_{s=K+1}^n (X^{(i)}(\tau_s) - X^{(i)}(\tau_{s-K}))(X^{(j)}(\tau_s) - X^{(j)}(\tau_{s-K}))/K, \tilde{n}_K = (n - K)/K, 1 \leq J \leq K$ 。 $[X^{(i)}, X^{(j)}]_T^{(K)}$ 实质是基于子抽样法的平均 RC(具体见原文)。相比于全样本的 RC, 其方差仅是原来的 $1/K$ 。但由于是有偏估计, 因此需要利用更高频率(J)来估计噪声方差, 由此得到上面的 TSCV 估计。在无内生性的白噪声条件下, 该估计可同时消除非同步和噪声引起的偏差。但由于 TSCV 是基于两种不同频率下的 RC 做差实现, 因而无法保证其正定性。实际计算 TSCV 时, 一般取 $J=1, K=O(n^{2/3})$ 。

(3) 事先平均估计(Pre-averaging Estimators)。独立于上述研究工作, Christensen 等人^[6]通过对多维高频数据序列进行等步长的移动加权平均来消除无内生性的独立同分布噪声影响, 使求得的平均序列更接近有效价格, 同时传承原始序列的同步或异步属性:

$$\bar{x}^{(i)}(t_k^{(i)}) = \sum_{j=1}^{k_n-1} g(j/k_n) x^{(i)}(t_{k+j}^{(i)}), \quad k = 1, \dots, N(i) - k_n + 1$$

其中, g 为 $[0, 1]$ 上的连续函数, 且满足 $g(0) = g(1) = 0, k_n$ 为正整数。一方面, 若平均序列非同步, 则将其嵌入到 HY 估计中, 得到渐近无偏的事先平均 HY 估计(以非对角元为例):

$$\text{PAHY}(i, j) = \sum_{k=0}^{N(i)-k_n+1} \sum_{s=0}^{N(j)-k_n+1} \bar{x}^{(i)}(t_k^{(i)}) \bar{x}^{(j)}(t_s^{(j)}) I((t_k^{(i)}, t_{k+k_n}^{(i)}) \cap (t_s^{(j)}, t_{s+k_n}^{(j)})) / (\psi_{HY} k_n)^2$$

其中, $\psi_{HY} = \int_0^1 g(u) du$ 。PAHY 较好地结合了事先平均法降噪技术与 HY 估计的区间同步化技术, 解决了标准 HY 估计面临的微观结构噪声问题。

另一方面, 若平均序列同步, 作者则进一步构建如下尺度调整的 RC 矩阵以实现波动率矩阵的估计:

$$\text{MRC} = \frac{n}{n - k_n + 2} \frac{1}{\psi_2 k_n} \sum_{k=0}^{n-k_n+1} \bar{x}(\tau_k) \bar{x}(\tau_k)'$$

其中, $\bar{x} = (\bar{x}^{(1)}, \dots, \bar{x}^{(p)})', \psi_2 = \int_0^1 g(u)^2 du$ 。

当平均步长 $k_n / n^{1/2} = \theta + o(n^{-1/4})$ ($\theta > 0$) 时, MRC 具有正定性和渐近混合正态性, 且可达到最优收敛速度 $n^{-1/4}$, 但是有偏估计。因此, 作者提出两种改进途径。一是从 MRC 中减去偏差矩阵的估计量 $(\int_0^1 (g'(u))^2 du)^{k_n} \hat{\Psi}_n / (\theta^2 \psi_2^{k_n})$, 其中, $\hat{\Psi}_n = \frac{1}{2n} \sum_{i=1}^n \bar{x}(\tau_i) (\bar{x}(\tau_i))'$ 为噪声的方差估计。由于是两项相减, 因而无法保证减后估计的正定性。但该纠偏估计(记为 MRC5)保留了 MRC 的最优收敛速度。二是提高移动平均的步长使其满足 $k_n / (n^{0.5+\delta}) = \theta + o(n^{-1/4+\delta/2})$ ($\delta > 0$)。由此所得的 MRC(记为 MRC6)具有正定性和渐近无偏性, 但其收敛速度却降为 $n^{-1/5}$, 并且稳定收敛于非中心的混合正态分布。

关于平均窗宽 k_n 或者进一步参数 θ 的选取, Christensen 等^[6]利用数值模拟研究样本容量和噪声对最优 θ 的影响时, 发现 MSE 曲线关于 k_n 高度不对称, 他们更倾向于选取略高于最优 k_n 的窗宽, 而其实证研究表明较保守的 k_n 有助于减缓价格离散取值带来的不利影响。在后文的数值模拟和实证分析中, 我们建议选取保守的 k_n , 即取 $\theta=1$ 。

综上所述, 事先平均法的去噪思想不同于其他三类方法: 其他方法是先计算 RC, 再对 RC 去噪, 而事先平均法则是先对序列去噪, 然后再计算 RC, 这也是其名称“事先”的由来。在对白噪声序列的要求方面, 已实现核估计的要求最宽泛, 即同时允许噪声序列具有内生性和序列相关性, 而其他方法均是基于噪声独立无内生性实现。在收敛速度方面, MRC5 的收敛速度最快, 可达到目前的最快速度 $n^{-1/4}$, 其他估计小于等于 $n^{-1/5}$ 。在协方差矩阵的内属性半正定性要求方面, MRK 和 MRC6 的构造保证他们的半正定性, 但 TSCV 和 PAHY 不一定满足该要求, 甚至可能出现方差为负的情形。

除上述主流方法外,还有一些非常规的方法。如刘丽萍^[19]考虑到市场微观结构噪声和跳跃的同时存在,提出修正的门限预平均已实现协方差阵估计。该估计通过事先平均法减少市场微观结构噪声的影响,并通过门限来剔除跳跃的影响。Liu Cheng 等^[13]利用伪极大似然方法对高频波动率矩阵进行估计。由于这些方法的应用不是很广泛。因此,本文不再做详细介绍。

下文从数值模拟和实证分析角度对上述各估计展开比较分析。两者区别在于前者的对数价格数据来源于已知数学模型的模拟,后者则来源于现实的金融市场,其数据所遵循的数学模型是未知的。因此,在数值模拟中,所要做的是通过与已知真实波动率矩阵的各种比较,来考察波动率矩阵估计的优劣性。而在实证分析中,由于真实波动率矩阵未知,故只能对各估计的性质、其对现实的解释情况,及相互间的横向比较等进行分析。综合而言,前者属于数学上的绝对比较,后者是结合现实的相对比较。

5 数值模拟

本节利用数值模拟来比较各估计的优劣。我们所采用的时间区间为 $t \in [0, 1]$, 并采用如下多变量因子随机波动率模型来模拟对数价格:

$$dY^{(i)} = \mu^{(i)} dt + dV^{(i)} + dF^{(i)}, dV^{(i)} = \rho^{(i)} \sigma^{(i)} dB^{(i)}, dF^{(i)} = \sqrt{1 - (\rho^{(i)})^2} \sigma^{(i)} dW \quad (1)$$

其中, B 中各元素和 W 均是独立的标准布朗, $F^{(i)}$ 是公因子, $V^{(i)}$ 是奇异成分 (Idiosyncratic)。

进一步地,假定瞬时波动率服从独立过程 $\sigma^{(i)} = \exp(\beta_0^{(i)} + \beta_1^{(i)} \zeta^{(i)})$, 其中, $\zeta^{(i)}$ 服从 Ornstein-Uhlenbeck 过程 $d\zeta^{(i)} = \alpha^{(i)} \zeta^{(i)} dt + dB^{(i)}$ 。这意味着 $V^{(i)}$ 和 $\sigma^{(i)}$ 的增量具有较好的相关性, $Y^{(i)}$ 和 $\zeta^{(i)}$ 的增量的相关性为 $\rho^{(i)}$, $Y^{(1)}(t)$ 和 $Y^{(2)}(t)$ 的相关系数为 $[(1 - (\rho^{(1)})^2)(1 - (\rho^{(2)})^2)]^{1/2}$ 。

各随机过程通过离散化来模拟各时点的价格。由于我国股市每天的交易时间为四个小时,故将 $[0, 1]$ 时间段离散化为 $N = 14400 (= 60 \times 60 \times 4)$ 个等长度区间,每个区间对应 1 秒,即每隔 $1/N$ 时间产生一个对数有效价格,同时,共模拟了 1000 条价格路径。模型中各参数值为 $(\mu^{(i)}, \beta_0^{(i)}, \beta_1^{(i)}, \alpha^{(i)}, \rho^{(i)}) = (0.03, -5/16, 1/8, -1/40, -0.3)$, 这些特殊取值保证了波动率过程的标准化,即 $E(\int_0^1 [\sigma_s^{(i)}]^2 ds) = 1$ 。我们通过从 $\zeta^{(i)}$ 的平稳分布 $N(0, (-2\sigma^{(i)})^{-1})$ 中抽样来确定每天的初始值 $\zeta^{(i)}(0)$ 。通过上面的设计,我们

有 σ 的方差为 $\exp(-2\beta_1^{(i)})^2/\alpha^{(i)} - 1 \approx 2.5$ 。

进一步地,模拟噪声过程为: $U_j^{(i)} | \sigma, Y \sim N(0, \omega^2)$, iid, $\omega^2 = \xi^2 (N^{-1} \sum_{j=1}^N \sigma^{(i)4} (j/N))^{0.5}$, 其中,噪声信号比 ξ^2 取 0, 0.001, 0.01, 分别对应无噪声、低噪声和高噪声三种情形。如 Bandi 等^[1]所述,上述噪声的设定形式表明噪声会随着有效价格方差的增加而增加。

由上述过程产生的是同步的高频数据。为使数据非同步,还需从泊松过程中抽样产生实际观测时间 $\{t_j^{(i)}\}$ 。设两个泊松过程独立,并且具有强度参数 (λ_1, λ_2) , 这里, λ_i 代表从完全数据 $Y^{(i)}$ 中抽取 1 个实际观测数据所需等待的平均时间。因此,平均每天有 N/λ_i 个 $Y^{(i)}$ 的观测值。设定 λ_1 分别取 (3, 10, 60) 以反映流动性对各估计的影响,同时,设定 $\lambda_2 = 2\lambda_1$ 使得 $Y^{(2)}$ 的更新速度平均为 $Y^{(1)}$ 的 1/2。

我们分别研究了基于各高频波动率矩阵的方差、协方差、相关系数。模拟结果见表 2 至表 4。在表 2, 两个 TSCV 估计中取 $J=1$, $K=cn^{2/3}$, 其中, $TSCV_{0.8}$ 中 $c=0.8$, $TSCV_{OPT}$ 中 c 取使协方差 MSE 最小的 c 。MRC5 和 MRC6 分别为 $k_n = n^{1/2}$, $n^{3/5}$ 的 MRC 估计, PAHY 中 $k_n = n^{1/2}$, 三个估计中均取 $g(x) = \min(x, 1-x)$ 。MRK 中取 Parzen 核函数, $m=2$, H 取 $H_i = 3.51\tilde{n}^{3/5} (\tilde{A}^{(i)2}/RV_{15m}^{(i)})^{2/5}$ 的均值, 其中, $\tilde{A}^{(i)2} = RV_{sy}^{(i)2}/2\tilde{n}$ 为第 i 个资产基于同步数据的噪声方差估计量, $RV_{sy}^{(i)}$ 为第 i 个资产基于同步数据的 RV , $RV_{15m}^{(i)}$ 为基于 15 分钟高频数据的 RV 。由于 HY 估计只适合估计协方差,不适于方差和相关系数的估计,因此在表 2 至表 4 中未列出 HY 估计结果。同时,受篇幅所限,本文只给出了资产 2 各已实现方差估计的偏差和均方误差,但下文关于方差的结论也适用于资产 1。从表 2 至表 4 我们可得出如下综合结论:

(1) 从纠偏角度分析: 两个 RV 估计表现均远不如所有其他纠偏估计,这说明了纠偏降噪技术确实起了较大作用,同时, HY 估计也明显优于两个 RV 估计,这可能得益于 HY 估计在同步化时保留数据的能力强于 1 分钟和 15 分钟间隔的前值法,且这种优势在流动性较强情形表现更明显,这进一步验证了我们的结论;此外,虽然 RV_{1m} 利用的数据量远大于 RV_{15m} ,但是其受到的微观结构噪声的影响也较大,以至于前者带来的信息增加等有利作用难敌后者的不利影响,使得 RV_{1m} 估计表现较差。

(2) 从流动性角度分析: $TSCV_{0.8}$ 、 $TSCV_{OPT}$ 、

MRC5、MRC6、MRK、PAHY 和 HY 估计在协方差、相关系数和方差三种估计情况下 rmse 均随流动性的减弱(λ 增加)而增加,该结论符合常规认识,即如果估计合理,则流动性越强,进入估计的样本数就越多,自然估计与真值的差异就越小。

(3)从噪声角度分析:除 MRK、HY 和 RV_1m 估计外,其余估计的 bias 和 rmse 略受噪声的影响,但影响不大,说明他们对噪声稳健;RV_1m 在估计方差时,其 bias 和 rmse 随噪声的增加而增加较快,MRK 在 4 种特征参数情况下,均受到较大的噪声影响。

(4)两两关联估计比较:①相比于 HY 估计,纠偏的 PAHY 在强噪声时的优势显著强于低噪声和无噪声时的优势,恰验证其纠偏的功能。②无论是 bias 还是 rmse, TSCV_{OPT} 的优势都强于 TSCV_{0.8}, 这是很自然的一个结果,但同时也说明 TSCV 估计受其频率选择的影响,但我们在 c 取其他值的数值模拟结果和后文的实证分析表明, TSCV 估计还是比较稳定的。受篇幅所限,这些结果未列出。③就协方差和两个方差而言, MRC5 估计的偏差均大于 MRC6,但其 rmse 值均较小,说明 MRC5 估计稳定性较好,但易低估风险;但是,相比于 TSCV、MRK 等其他类型的估计, MRC5 的偏差还算较小;在相关

系数估计方面, MRC5 表现出了双重优势,说明其不但较接近真实相关系数,并且稳定性也较好。

综合而言, TSCV、MRC 和 MRK 三个估计优于 PAHY 和 HY 估计,而后两个又优于 RV 估计;下面进一步比较 TSCV、MRC 和 MRK:①从协方差角度, MRK 最优,其次是 MRC5 和 TSCV_{OPT},最后依次是 MRC6 和 TSCV_{0.8},其中, MRC5 的偏差均小于 TSCV_{OPT},同时,在流动性较差时, MRC5 的 rmse 值小于 TSCV_{OPT},这可能源于此时 TSCV 估计所采用的子抽样法降低了样本容量,使得 TSCV 估计相对不稳定。②在方差方面, MRK 的偏差较大,尤其是高噪声情形,同时,其 rmse 在高噪声时的表现不如低噪声或无噪声情形,说明 MRK 受噪声影响大; MRC5 在偏差和 rmse 两方面的表现均优于 TSCV_{OPT}; MRC6 在偏差方面优于 MRK,而低噪声时, MRK 的 rmse 优于 MRC6。③在相关系数方面,由于其不是我们估计和优化的目标,因此,有些估计表现不佳,其中, MRK 表现最差,同时,低噪声时, TSCV_{OPT} 表现最好,而在高噪声时, MRC5 表现最好;综上分析,在协方差估计方面,建议选用 MRK 估计,在低噪声的方差估计方面,建议选用 MRK,而在高噪声情形,建议选取 MRC5 估计。

表 2 各已实现协方差估计的偏差和均方误差开方的比较

		TSCV _{0.8}		TSCV _{OPT}		MRC5		MRC6		MRK		PAHY		HY		RV_15m		RV_1m	
		bias	mse	bias	mse	bias	mse	bias	mse	bias	mse	bias	mse	bias	mse	bias	mse	bias	mse
$\xi^2=0$	(3,6)	-0.078	0.332	-0.047	0.101	-0.006	0.125	-0.004	0.173	-0.007	0.065	-0.009	0.521	0.012	0.263	1.021	2.335	0.812	2.388
	(10,20)	-0.111	0.413	-0.070	0.144	-0.010	0.157	-0.004	0.222	-0.011	0.107	0.033	0.655	-0.066	0.611	0.609	1.610	0.341	1.769
	(60,120)	-0.211	0.521	-0.149	0.249	-0.041	0.242	-0.010	0.336	-0.012	0.173	0.078	1.210	0.028	1.480	0.391	1.554	-0.216	1.818
$\xi^2=0.001$	(3,6)	-0.078	0.332	-0.046	0.102	-0.006	0.125	-0.005	0.172	0.002	0.093	-0.011	0.513	0.015	0.397	0.975	2.295	0.777	2.390
	(10,20)	-0.111	0.413	-0.070	0.144	-0.010	0.157	-0.004	0.222	-0.009	0.122	0.033	0.655	-0.054	0.643	0.609	1.606	0.341	1.792
	(60,120)	-0.210	0.521	-0.147	0.248	-0.039	0.242	-0.009	0.337	-0.023	0.179	0.081	1.209	0.027	1.499	0.396	1.537	-0.213	1.806
$\xi^2=0.01$	(3,6)	-0.078	0.331	-0.035	0.117	-0.006	0.126	-0.004	0.170	0.003	0.135	-0.007	0.540	0.167	1.413	1.074	2.431	0.868	2.422
	(10,20)	-0.110	0.417	-0.063	0.162	-0.012	0.168	-0.003	0.226	-0.011	0.174	0.031	0.653	-0.132	1.105	0.625	1.668	0.340	1.837
	(60,120)	-0.211	0.522	-0.129	0.258	-0.037	0.258	-0.006	0.345	-0.049	0.240	0.049	1.302	0.124	1.610	0.371	1.579	-0.224	1.874

表 3 资产 2 各已实现方差估计的偏差和均方误差开方的比较

		TSCV _{0.8}		TSCV _{OPT}		MRC5		MRC6		MRK		PAHY		RV_15M		RV_1m	
		bias	mse	bias	mse	bias	mse	bias	mse	bias	mse	bias	mse	bias	mse	bias	mse
$\xi^2=0$	(3,6)	-0.111	0.574	-0.072	0.192	-0.015	0.219	-0.010	0.301	0.009	0.103	0.483	1.188	1.801	3.506	3.167	6.065
	(10,20)	-0.162	0.731	-0.110	0.283	-0.020	0.277	-0.002	0.395	0.005	0.178	0.644	1.788	0.919	2.729	2.532	5.438
	(60,120)	-0.311	0.916	-0.220	0.477	-0.053	0.426	0.000	0.563	0.064	0.303	2.165	5.375	1.224	4.355	2.186	5.502
$\xi^2=0.001$	(3,6)	-0.111	0.574	-0.071	0.189	-0.015	0.215	-0.008	0.298	0.122	0.235	0.488	1.204	1.784	3.498	3.557	6.396
	(10,20)	-0.162	0.728	-0.108	0.278	-0.018	0.276	0.002	0.396	0.077	0.252	0.651	1.785	0.954	2.752	2.999	5.839
	(60,120)	-0.311	0.915	-0.220	0.486	-0.050	0.439	0.005	0.568	0.022	0.302	2.405	5.539	1.247	4.423	2.410	5.736
$\xi^2=0.01$	(3,6)	-0.111	0.573	-0.062	0.261	-0.014	0.236	0.012	0.296	0.249	0.429	0.533	1.239	2.131	3.942	7.763	11.718
	(10,20)	-0.163	0.738	-0.109	0.337	-0.023	0.296	0.026	0.404	0.289	0.468	0.728	1.856	1.200	3.081	7.055	11.114
	(60,120)	-0.311	0.918	-0.191	0.505	-0.027	0.487	0.047	0.608	0.241	0.622	4.498	7.250	1.476	4.441	4.455	7.875

表 4 各已实现相关系数估计的偏差和均方误差开方的比较

		TSCV _{0.8}		TSCV _{OPT}		MRC5		MRC6		MRK		PAHY		RV_15M		RV_1m	
		bias	mse	bias	mse	bias	mse	bias	mse	bias	mse	bias	mse	bias	mse	bias	mse
$\xi^2=0$	(3,6)	-0.010	0.062	-0.001	0.012	-0.002	0.019	-0.003	0.029	-0.027	0.037	-0.272	0.711	-0.276	0.495	-0.463	0.627
	(10,20)	-0.013	0.082	-0.002	0.020	-0.003	0.026	-0.005	0.038	-0.033	0.046	-0.325	0.739	-0.068	0.199	-0.451	0.604
	(60,120)	-0.027	0.143	-0.004	0.042	-0.008	0.046	-0.014	0.062	-0.100	0.125	-0.347	0.839	-0.250	0.469	-0.688	0.787
$\xi^2=0.001$	(3,6)	-0.010	0.061	0.000	0.018	-0.002	0.019	-0.006	0.030	-0.119	0.138	-0.274	0.603	-0.302	0.515	-0.536	0.659
	(10,20)	-0.013	0.082	-0.002	0.024	-0.004	0.028	-0.009	0.039	-0.099	0.115	-0.325	0.639	-0.081	0.200	-0.539	0.641
	(60,120)	-0.027	0.143	-0.003	0.046	-0.009	0.048	-0.018	0.065	-0.112	0.134	-0.465	0.741	-0.257	0.465	-0.719	0.800
$\xi^2=0.01$	(3,6)	-0.009	0.062	0.007	0.070	-0.003	0.028	-0.027	0.043	-0.215	0.243	-0.326	0.558	-0.339	0.523	-0.703	0.746
	(10,20)	-0.012	0.082	0.005	0.076	-0.009	0.043	-0.036	0.058	-0.264	0.290	-0.467	0.688	-0.205	0.298	-0.752	0.780
	(60,120)	-0.027	0.147	0.009	0.111	-0.023	0.081	-0.058	0.097	-0.283	0.313	-0.738	0.788	-0.363	0.519	-0.821	0.854

表 5 各同步化方法的平均每日数据保留比率

	GHYX	DFMZ	ZGZT	GSYH	JLAD		GHYX	DFMZ	ZGZT	GSYH	JLAD
更新时间法	0.5072	0.5140	0.3975	0.3439	0.3632	15 秒	0.3683	0.3519	0.3204	0.2715	0.2930
5 秒	1.1049	1.0556	0.9613	0.8145	0.8790	1 分钟	0.0921	0.0880	0.0801	0.0679	0.0732

6 实证分析

我们选取沪深 A 股 2012 年 10 月至 2013 年 9 月的 level2 逐笔高频交易数据进行实证分析。数据来源于国泰安 CSMAR 数据库。共选取 5 支来自不同行业的代表性股票:歌华有线(GHYX)、东方明珠(DFMZ)、中国中铁(ZGZT)、工商银行(GSYH)、吉林敖东(JLAD)。首先对数据进行如下预处理:①删除不在 9:30—11:30 和 13:00—15:00 间的记录,这说明集合竞价的数据被删除;②删除交易价格、交易量为负的记录;③对记录不完整的交易日,删除该日内所有股票记录。④若一个时间点上有多 个记录,则对该点上的各交易价格按交易量进行加权平均。前 3 步通常对数据的删除比率较小,第 4 步一般较大。经该过滤方法之后,四支股票数据的保留比例依次为 47.11%、43.14%、43.26%、46.06%和 44.44%。

表 5 进一步给出了较常用的更新时间法和 3 种等间隔抽样前值法的平均每日数据保留比率。从表 5 可看出,虽然 5 秒等间隔前值法保留数据比率最大,但对本身交易相对不频繁的歌华有线和东方明珠两支股票而言,该方法的抽样频率过大,从而大量重复原始信息,使得其数据保留比率大于 1。与 15 秒和 1 分钟相比,更新时间法保留数据比率最大。另一方面,注意到 1 分钟数据的保留比率小于 1%,可见其损失信息较多。除特别说明,下文仅利用经更新时间法处理后的同步化数据展开各协方差估计。

首先,为说明艾普斯效应,我们利用同步化数据计算传统的已实现波动率矩阵 RC。考虑到数据的

损失,我们采用子抽样法估计 RC,即 TSCV 估计式的第一项 $[X^{(i)}, X^{(j)}]_T^{(K)}$ 。受篇幅所限,以下以歌华有线和吉林敖东为例,其他情形具有相似结论。图 3 给出两者不同抽样频率下的已实现协方差估计曲线图。图 2 证实了艾普斯效应,在非常高的抽样频率下,已实现协方差迅速从 0.18 降至 0.035,而在较高的抽样频率下,由于微观结构噪声的影响,已实现协方差具有一定的向上偏差。

图 3 为歌华有线和吉林敖东不同频率组合(即不同 K 和 J 的组合)的 TSCV 估计的特征图,图 4 和图 5 分别为其对应的截面图,其中,图 4 为高频 J 固定的五种情况下的截面图,图 6 为低频 K 固定的五种情况下的截面图。与图 2 相比,图 3 的 TSCV 估计基本平稳在 1.7×10^{-4} 附件,如果不考虑 K 和 J 的选取,TSCV 估计基本在 0.5×10^{-4} 和 1.8×10^{-4} 之间做窄幅波动,显示了良好的降噪效果。图 4 和图 5 展示了固定一个频率,TSCV 随另一个频率的变化情况,从图 4 可看出,随着高频 J 的增加,TSCV 估计递减,但随着低频 K 的增加,这种差异越来越小,说明噪声对抽样间隔较长数据的 RV 影响较小,与实际市场相符。从图 5 可看出,对较小的低频 K (如 $K=50,100$), J 的变化对 TSCV 的影响较大,而对较大的 K (如 400),TSCV 基本不受 J 的影响。这是因为 K 所在项估计的是价格波动率矩阵, J 所在项估计的是噪声波动率矩阵,而 K 和 J 若过于接近,两项所代表的波动率较难区分,故 TSCV 对 J 的变化异常敏感;而若 K 和 J 差距较大,即使 J 做适当的变化,两者所在项估计的对象也很分明。

表 6 给出了各已实现协方差矩阵估计,其中,

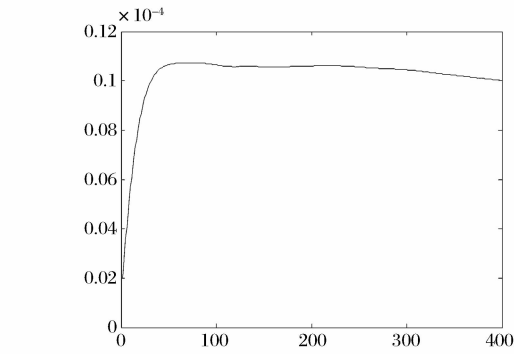


图 2 歌华有线和吉林敖东的已实现协方差函数

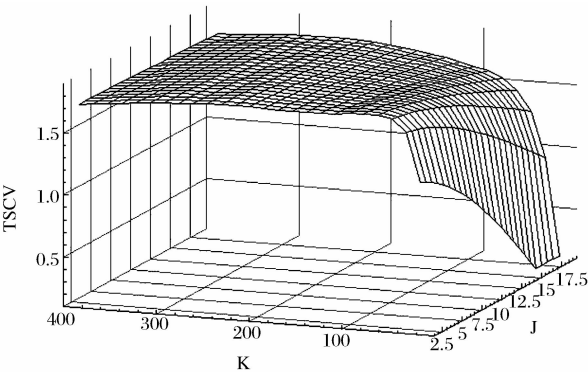


图 3 不同频率组合的 TSCV 估计的特征图

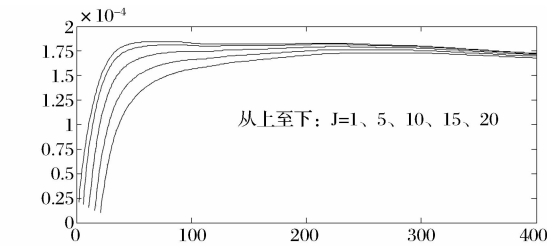


图 4 固定高频 J 下 TSCV 估计的截面图

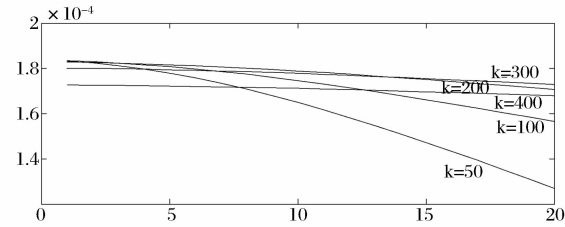


图 5 固定低频 K 下 TSCV 估计的截面图

RV_ave(5m,15m)为分别基于 5 分钟和 15 分钟高频数据的 RV 估计的均值, RV_15s 为 15 秒高频数据的 RV,其他符号意义同表 1。从表 6 可得出如下结论:(1)虽然 RV_15s 各协方差项都位于正值区域,但各项严重趋向于 0,说明随着抽样频率的迅速

增加,传统 RV 估计显示出了艾普斯效应,因此,其易低估风险。该现象对 RV_ave(5m,15m)表现较不明显,原因在于 RV_ave(5m,15m)未受到微观结构噪声和艾普斯效应的严重影响。因此, RV_ave(15m,15m)可较好捕捉协方差结构的平均水平。(2)HY 估计各项均偏离 RV_ave(5m,15m)较大,这是因为 HY 估计没有进行纠偏;纠偏的 PAHY 对 HY 则有了较大的改进,但离 RV_ave(5m,15m)还是有一定偏差。(3)从协方差角度, RV_ave(5m,15m)的各项协方差介于 MRC5 和 MRC6 之间,并普遍大于 MRK 和 TSCV。

下面从时序图角度比较歌华有线和吉林敖东各协方差估计。从图 6 可看出, MRC5、MRC6、MRK、PAHY 和 RV_ave(5m,15m)估计表现非常相似,而 RV_15s 和 HY 估计与他们相差较大。

最后研究各协方差矩阵估计的正定性问题。合理的协方差矩阵估计应该具有正定性,这是其应具有的内属性。由前一节的理论分析可知, MRC6 和 MRK 估计的构造结构已保证其非负定性,而其余估计单从结构上看,很难判断其正定与否。但在本实证分析中,经验证, MRC5、TSCV 和 HY 估计每日的协方差矩阵均为正定,而 PAHY 的非正定日期较多,共 38 天是非正定的,占比 15.97%。因此,在实际应用中,若采用 PAHY 法对协方差矩阵进行估计时,最好引入一些特征值清除技术,以对其非正定性问题进行校准。

7 结语

波动率矩阵的估计是现代金融经济学的重要课题,其对于资产配置、组合 VaR 计算等金融活动具有重要意义。基于高频数据的波动率矩阵估计可有效解决传统低频估计面临的维数危机等种种瓶颈问题。理论上,传统的已实现波动率矩阵 RV 收敛于积分波动率矩阵,是其理想的简洁估计。然而,现实中,由于非同步交易和微观结构噪声的影响, RV 估计会产生艾普斯效应而偏离其理论值。因此,需对高频数据进行同步化和纠偏处理。实证分析表明,更新时间法是保留数据最多的同步化方法,而常见的 1 分钟等数据损失了近 90% 的数据信息。在降噪技术方面,各波动率矩阵估计的数值模拟和实证分析结果表明:

(1)传统的 RV 估计表现均不如其他纠偏估计,这说明了纠偏降噪技术确实起了较大作用,在实际应用中降噪技术的运用是非常必要的。

表 6 各已实现协方差矩阵估计($\times 10^{-3}$)

MRC5					MRC6					PAHY $kn=n^{1/2}$				
GHYX	DFMZ	ZGZT	GSYH	JLAD	GHYX	DFMZ	ZGZT	GSYH	JLAD	GHYX	DFMZ	ZGZT	GSYH	JLAD
0.5349	0.1231	0.0695	0.0318	0.1742	0.5142	0.1647	0.0925	0.0482	0.2103	0.4699	0.1747	0.1018	0.0547	0.2081
0.1231	0.5168	0.0649	0.0401	0.1374	0.1647	0.4860	0.0836	0.0455	0.1677	0.1747	0.4840	0.0940	0.0480	0.1665
0.0695	0.0649	0.2284	0.0610	0.1050	0.0925	0.0836	0.2205	0.0552	0.1363	0.1018	0.0940	0.1980	0.0485	0.1494
0.0318	0.0401	0.0610	0.1522	0.0658	0.0482	0.0455	0.0552	0.1138	0.0862	0.0547	0.0480	0.0485	0.0840	0.0907
0.1742	0.1374	0.1050	0.0658	0.5233	0.2103	0.1677	0.1363	0.0862	0.5052	0.2081	0.1665	0.1494	0.0907	0.4715
MRK					TSCV0.8					HY				
GHYX	DFMZ	ZGZT	GSYH	JLAD	GHYX	DFMZ	ZGZT	GSYH	JLAD	GHYX	DFMZ	ZGZT	GSYH	JLAD
0.5772	0.1417	0.0594	0.0277	0.1436	0.3125	0.1441	0.0711	0.0335	0.1676	2.4132	0.0928	0.1601	0.1026	0.0317
0.1417	0.5557	0.0594	0.0386	0.1147	0.1441	0.2236	0.0593	0.0280	0.1227	0.0928	3.3559	0.2767	0.1930	0.0393
0.0594	0.0594	0.2773	0.0655	0.0922	0.0711	0.0593	0.1192	0.0359	0.1115	0.1601	0.2767	13.8957	0.9094	0.0630
0.0277	0.0386	0.0655	0.1970	0.0563	0.0335	0.0280	0.0359	0.0506	0.0673	0.1026	0.1930	0.9094	8.5718	0.0469
0.1436	0.1147	0.0922	0.0563	0.5188	0.1676	0.1227	0.1115	0.0673	0.3309	0.0317	0.0393	0.0630	0.0469	1.4406
RV_15s					RV_ave(5m,15m)									
GHYX	DFMZ	ZGZT	GSYH	JLAD	GHYX	DFMZ	ZGZT	GSYH	JLAD					
0.9818	0.0474	0.0535	0.0278	0.0290	0.4981	0.1401	0.0821	0.0462	0.1731					
0.0474	1.2534	0.0210	0.0125	0.0345	0.1401	0.4676	0.0781	0.0464	0.1419					
0.0535	0.0210	4.3589	0.0198	0.0475	0.0821	0.0781	0.3435	0.0504	0.1206					
0.0278	0.0125	0.0198	2.2408	0.0231	0.0462	0.0464	0.0504	0.1718	0.0815					
0.0290	0.0345	0.0475	0.0231	0.6536	0.1731	0.1419	0.1206	0.0815	0.4726					

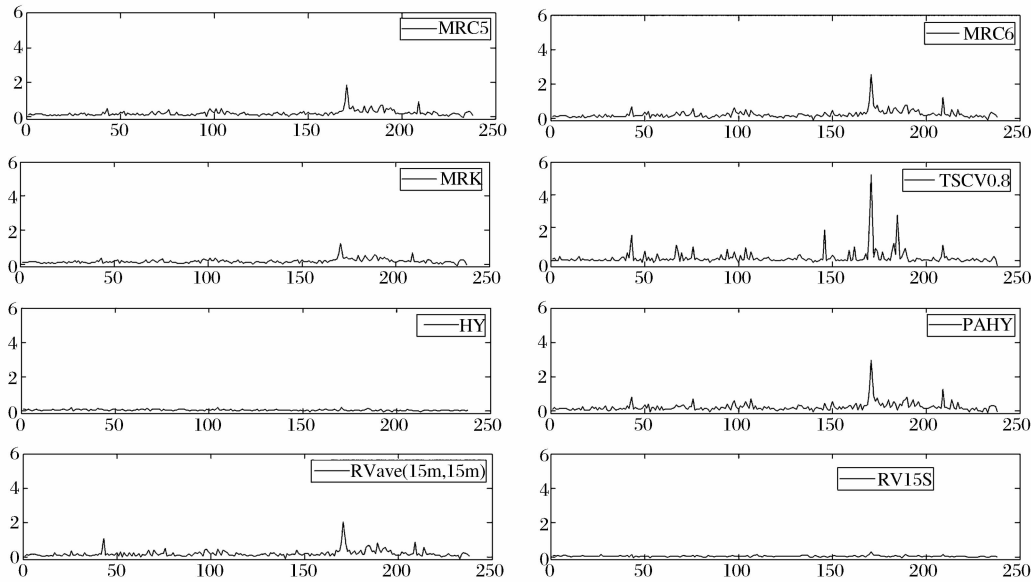


图 6 歌华有线和吉林敖东各协方差时间序列

(2)TSCV、MRC、MRK、PAHY 和 HY 估计受流动性的影响较大,进入估计的样本数越多,各估计的均方误差就越小。

(3)MRK 受噪声的影响较大,而 TSCV 和 MRC 则对噪声相对比较稳健。

(4)综合而言,TSCV、MRC 和 MRK 三个估计优于 PAHY 和 HY 估计,而后两个又优于未纠偏的 RV 估计;进一步地,在协方差估计和低噪声的方差估计方面,建议选用 MRK 估计,而在高噪声的方差

估计情形,建议选取 MRC5 估计。

(5)MRC6 和 MRK 估计的构造结构保证了他们的非负定性,而其余估计单从结构上看,很难判断正定与否。实证分析表明在所研的日期中,PAHY 估计近 15.97%的天数为非正定,说明了对其正定校准的必要性,而 MRC5、TSCV 和 HY 估计全部为正定。

波动率矩阵估计的优劣直接影响其建模好坏、以及最终的投资决策。从目前已有研究成果来看,

这方面的研究还仅处于起始阶段,比如,如果各资产流动性差异较大,既有同步化方法必然会损失样本区间内的大量价格信息,尤其是流动性较强的价格信息;再如,现有高频波动率矩阵估计的高维适用性有限问题,在许多高维高频波动率矩阵估计问题中,样本量的绝对数也许不算小,但相对于资产维数,就有可能相当小了。这些会是未来研究中很重要的课题,对于现有研究也会是很有意义的扩充。

参考文献:

- [1] Bandi F M, Russell J R. Separating microstructure noise from volatility [J]. *Journal of Financial Economics*, 2006, 79(3): 655—692.
- [2] Barndorff-Nielsen O E, Hansen P R, Lunde A, et al. Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise [J]. *Econometrica*, 2008, 76(6): 1481—1536.
- [3] Barndorff-Nielsen O E, Hansen P R, Lunde A, et al. Multivariate realised kernels: Consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading [J]. *Journal of Econometrics*, 2011, 162(2): 149—169.
- [4] Bollerslev T, Tauchen G, Zhou Hao, Expected stock returns and variance risk premia [J]. *Review of Financial Studies*, 2009, 22(11), 4463—4492.
- [5] Christensen K, Kinnebrock S, Podolskij M. Pre-averaging estimators of the ex-post covariance matrix in noisy diffusion models with non-synchronous data [J]. *Journal of Econometrics*, 2010, 159(1): 116—133.
- [6] Christensen K, Oomen R. Podolskij M. Realised quantile-based estimation of the integrated variance [J]. *Journal of Econometrics*, 2010, 159(1): 74—98.
- [7] Epps T W. Comovements in stock prices in the very short run[J]. *Journal of the American Statistical Association*, 1979, 74(366a): 291—296.
- [8] Ghysels E, Harvey A C, Renault E. Stochastic volatility[M]//Maddala G S, Rao C R. *Statistical methods in finance*. North-Holland, Amsterdam; Elsevier, 1996, 119—191.
- [9] Guillaume D M, Dacorogna M M, Dave R, et al. From the bird's eye to the microscope: A survey of new stylized facts of the intradaily foreign exchange markets[J]. *Finance and Stochastics*, 1997, 1(2): 95—129.
- [10] Hayashi T, Yoshida N. On covariance estimation of non-synchronously observed diffusion processes[J]. *Bernoulli*, 2005, 11(2): 359—379.
- [11] Jacod J, Shiryaev A N. *Limit theorems for stochastic processes* [M]. 2nd ed, Berlin: Springer, 2003.
- [12] Liu Cheng, Tang Chengyang. A quasi-maximum likelihood approach for integrated covariance matrix estimation with high-frequency data [J]. *Journal of Econometrics*, 2014, 180(2): 217—232.
- [13] Muthuswamy J, Sarkar S, Low A, et al. Time variation in the correlation structure of exchange rates: high-frequency analysis [J]. *Journal of Futures Markets*, 2001, 21(2): 127—144.
- [14] Protter P. *Stochastic integration and differential equations* [M]. New York: Springer-Verlag, 2004.
- [15] Tsay R S. *Analysis of financial time series* [M]. Hoboken, New Jersey: Wiley & Sons, Inc, 2005.
- [16] Zhang Lan. Estimating covariation: epps effect and microstructure noise [J]. *Journal of Econometrics*, 2011, 160(1): 33—47.
- [17] Zhang Lan, Mykland P A, Ait-Sahalia Y. A tale of two time scales: Determining integrated volatility with noisy high-frequency data [J]. *Journal of the American Statistical Association*, 2005, 100(472): 1394—1411.
- [19] 刘丽萍. 基于市场微观结构噪声和跳跃的金融高频协方差阵的估计及其应用研究[D]. 成都:西南财经大学, 2013.
- [20] 刘志东, 薛莉. 金融市场高维波动率的扩展广义正交 GARCH 模型与参数估计方法研究[J]. *中国管理科学*, 2010, 18(6): 33—41.
- [21] 熊和平. 投资组合协方差矩阵的性质与最优组合的选择[J]. *中国管理科学*, 2002, 10(2): 12—14.
- [22] Ait-Sahalia Y, Fan J, Xiu D. High-Frequency covariance estimate with noisy and asynchronous financial data [J]. *Journal of the American Statistical Association*, 2010, 105: 1504—1517.

Comparing Estimators of the High-Frequency Volatility Matrix in the Presence of Non-synchronous Trading and Market Microstructure Noise

ZHAO Shu-ran¹, JIANG Ya-ping¹, REN Pen-min²

(1. School of Economics, Ocean University of China, Qingdao 266100, China;

2. School of Economics, Qingdao University, Qingdao 266071, China)

Abstract: High-frequency volatility matrix estimator can effectively solve some bottleneck problems faced by traditional low-frequency estimators. However, because of the influence of non-synchronous trading and market microstructure noise, it has epps effect and some big bias. So mainly three kinds of synchronization methods for non-synchronous step-by-step high-frequency data and five types of the noised-reduction methods for the traditional realized volatility matrix are considered in this paper. The two kinds of methods are deeply compared separately, from data simulation and empirical analysis. The results suggest that refresh time method includes the largest amount of data among methods we considered, realized volatility matrix has epps effect and serious bias, multivariate realized kernels, two scales realized volatility matrix estimator and modulated realized volatility matrix estimator effectively reduce noise, but pre-averaging HY and HY estimators behave a little bad. The research results can provide a useful reference and guidance on methods for workers in related fields of further research and application.

Key words: volatility matrix; high-frequency data; market microstructure noisy; non-synchronous trading