



市道轮换下的高频数据参数估计

柳向东,靳晓洁

暨南大学经济学院,广东广州 510632

在高频金融数据研究分析中引入市道轮换模型,结合自回归模型和波动率替代模型,对高频数据进行建模分析,运用滤波算法以及Kim平滑算法等进行参数估计和预测.利用2017- 01- 03至2017- 08- 02上证综指每5 min的收盘价格,用Matlab实现马氏市道轮换自回归模型的预测和推测,并构建波动率替代模型.结果表明,马氏市道轮换高频数据模型是一种具有模型创新且理论性强的分析方式.

关键词:应用统计学; 高频数据; 市道轮换模型; 滤波算法; Kim平滑算法; 参数估计

Parameter estimation via regime switching model for high frequency data

LIU Xiangdong and JIN Xiaojie

College of Economics, Jinan University, Guangzhou 510632, Guangdong Province, P.R.China

To analyze high-frequency financial data, we use the regime switching method combined with the autoregressive model and volatility replacement model. We perform modelling analysis for high frequency data and use filtering algorithm and Kim smoothing algorithm to perform parameter estimation and prediction. By using the closing prices of each 5 minutes of the 2017- 01- 03 to 2017- 08- 02 Shanghai Composite Index, we achieve regime-switching autoregressive model of prediction and speculation, and construct the volatility replacement model. The results show that the Markov high frequency data model is an innovation model and an analysis method with strong theoretical background.

Keywords: application of statistical mathematics; high frequency data; regime switching; filtering algorithm; Kim smoothing algorithm; parameter estimation

DOI: 10.3724/SP.J.1249.2018.04432

备注:

基金项目:国家自然科学基金资助项目(71471075); 教育部人文社会科学研究资助项目(14YJAZH052)

作者简介:柳向东(1973—)男,暨南大学教授、博士生导师,研究方向:概率统计在经济金融领域的应用研究, E-mail: tliuxd@jnu.edu.cn

中文责编:方 圆; 英文责编:淡 紫

引言

20世纪90年代以前,金融市场的研究多以低频数据为依据,低频数据是指日、周、月、季度及年的频率数据.近年来,随着计算机与通信技术的迅猛发展,记录高频数据日趋便捷,从而受到金融界的广泛关注.高频数据是指金融市场运行过程中以小时、分钟、秒或实际交易间隔为采集频率的日内交易数据.金融市场往往是连续运行的,交易频繁,市场情况瞬息万变,基于低频数据的金融市场研究无疑会损失大量有效市场信息,无法准确理解市场微观结构.而高频数据所包含的资产价格信息接近于理论上连续时间的资产价格信息,不仅包含金融资产价格更丰富的信息,还包括像交易时间间隔、交易量、买卖价差等众多其他维度的信息,能够对金融市场进行更精细的分析,对理解市场微观结构具有重要作用.结合金融学、计量经济学、时间序列分析及统计学等理论,不少学者从不同角度对高频数据进行了大量的金融计量研究.朱建平^[1]对国内外金融高频数据的研究现状进行剖析,澄清金融高频数据的概念与特征,从统计学视角审视金融高频数据.雷平生等^[2]利用改进的统计套利策略对高频数据进行研究,满足了机构投资者套利需求,同时引进一种更有效的全新投资方式. HANSEN等^[3]提出一种分析高频数据的新模型,即已实现测度与传统自回归条件异方差(ge neralized autoregressive conditional heteroskedasticity, GARCH)模型相结合的Realized GARCH模型,且通过实证表明,与基于低频的数据日收益率的一般GARCH模型相比,其有实质性的改进.王天一等^[4]考察Realized GARCH模型用于高频数据的研究,着重考察了沪深300指数1 min高频数据的收益率分布及波动率预测,指出不同抽样频率的已实现方差对模型预测有显著影响. ZHANG^[5]提出利用幂律函数的霍克斯过程(Hawkes processes with power-law kernels)对高频数据进行建模,且通过与其他模型对比,表明此模型有实际应用价值.李胜歌等^[6-11]也对高频数据的波动率进行研究,利用一阶偏差修正方法、ARFIMAX模型、GARCH模型、SV模型、ACD模型、ACH模型、扩展的ACH模型以及跳扩散模型等对高频数据进行了分析.近年来,对于高频数据的实证研究结果表明,金融市场存在资产价格跳跃行为.沈根祥^[12]构造了跳跃行为的Hausman检验统计量,并对沪深300指数高频数据进行分析,发现样本区间内有1/3以上的交易日存在跳跃行为.唐勇等^[13]基于非参数方法,结合A-J跳跃检验统计量,构建新的跳跃方差和连续样本路径方差对上证综指高频数据进行建模分析,包括高频数据中跳跃方差统计特征、跳跃方差贡献、跳跃幅度以及跳跃与经济信息关系,有助于投资者优化投资策略,为监管部门提供监管依据. BAJGROWICZ等^[14]则在对高频数据分析中,利用伪跳跃(spurious jump)检验法研究跳跃,认为跳跃发生与流动性压力有关.汪先珍^[15]基于BN-S理论框架对上证综指高频数据进行研究,证实跳跃的存在,进而剥离出股价行为中的跳跃成分,并对跳跃产生的时刻、幅度以及分布特征加以分析,着重探讨了几个跳跃个例产生的原因.

从上述文献来看,对高频数据的研究并不少见,但利用马氏市道轮换研究高频数据的文献并不多.马氏轮换模型可以刻画金融市场的各种状态,对股票价格波动的描述更加切合实际,因此,在马氏轮换基础上推导各类具体模型对于高频数据的分析预测具有重要现实意义.近年来,国内外许多学者将马氏轮换运用到金融市场模型中,并得到较好的结论. HAMILTON^[16]首先提出马氏市道轮换,并利用该模型对时间序列的对数差分数据进行回归模拟,较好地刻画了实际产出增长的非线性动态和非对称特性. CAI等^[17-18]将马氏市道轮换引入ARCH模型,为存在结构转换的金融资产序列的波动性建模提供新方法. 佟杰^[19]在预测模型中引入马氏轮换,并用Baum Welch算法和弗罗贝尼乌斯范数最小法进行参数估计,并预测上证指数收益率,得到很好的预测效果. 刘金全等^[20]在CKLS中加入马氏轮换,将传统CKLS模型推广至更为一般的状态相依模型,并通过Hamilton滤波和Kim平滑概率得到比传统CKLS模型更好的估计和分析结果. 杨宝臣等^[21]针对SHIBOR动态特性中存在的状态转换和波动聚类现象,分别将马氏轮换和GARCH效应引入CIR模型,基于Kim滤波的极大似然估计法进行估计,并与CIR模型对比,发现改进的模型对风险溢价动态特性的刻画能力更强.本研究尝试在马氏市道轮换下对高频数据进行研究,高频数据选用上证综指每5 min收盘价作为分析,并用市道轮换下的自回归模型以及波动率替代模型对高频数据进行分析,其目的是由于在不同经济体制下会产生不同收益率,而马氏轮换能对这些不同经济体制进行明确研究.在估计技术方面,本研究采用极大似然法,并结合Hamilton滤波概率和Kim平滑概率等来估计马氏市道轮换模型,取得较精准的预测效果.本文还将市道轮换模型加入到波动率替代模型对高频数据进行分析,并给出了相应的估计.

1 高频数据模型

1.1 基本知识

定义1 一个离散的随机过程{S_t, t∈N}被称为马尔可夫链,若对所有t∈N,满足

$$P(S_{t+1}=s_{t+1}|S^t_1=s^t_1)=P(S_{t+1}=s_{t+1}|S_t=s_t)$$

这里利用S^t₁指代{S₁, S₂, ..., S_t}, 上式称为马尔可夫性.马尔可夫性是马尔可夫链最重要的性质,形象地说,马尔可夫性意味着“未来”仅与“现在”有关,而与“过去”无关.

定义2 在t时刻的n步转移概率定义为a_{ij}(n, t)=P(S_{t+n}=j|S_t=i), 即过程在t时刻位于状态i, 在t+n时刻转移到状态j的概率.

值得注意的是,若马尔可夫链在t时刻的n步转移概率a_{ij}(n, t)与t无关,则称为齐次马尔可夫链.此时n步转移概率记为a_{ij}(n), 而1步转移概率简称转移概率,记为a_{ij}, 下文中马尔可夫链均指齐次马尔可夫链,转移概率a_{ij}均指1步转移概率.

定义3 双离散随机过程{S_t, t∈N₁}, {Y_t, t∈N₂}被称为隐马尔可夫链,若满足

$$1)P(S_t=s_t|S^{t-1}_1=s^{t-1}_1)=$$

$$P(S_t=s_t|S_{t-1}=s_{t-1});$$

$$2)P(S_t=s_t|Y^{t-1}_1=y^{t-1}_1, S^{t-1}_1=s^{t-1}_1)=$$

$$P(S_t=s_t|S_{t-1}=s_{t-1});$$

$$3)P(Y_t=y_t|S^t_1=s^t_1, Y^{t-1}_1=y^{t-1}_1)=$$

$$P(Y_t=y_t|S_t=s_t);$$

$$4)P(Y_t=y_t|S^t_1=s^t_1, Y^{t-1}_1=y^{t-1}_1, Y^T_{t+1}=y^T_{t+1})=$$

$$P(Y_t=y_t|S_t=s_t).$$

其中, {S_t}称为状态过程,不可观测,实际上是一个齐次马尔可夫链; {Y_t}是观测值过程.以上性质可见,给定S_{t-1}, S_t与其他所有变量(过去的观测值和过去的状态)均无关,给定S_t, Y_t与其他所有变量(之前和之后的观测值、过去的状态)均无关.这些性质是下文中经常用到的性质.

引理1 P(AB|C)=P(A|C)P(B|AC)

引理1虽然是一个很简单公式,但在下文许多公式的推导中却频繁使用.

1.2 马氏市道轮转自回归模型

近年来,各种经济事件序列已作为市道轮换事件序列进行建模.建模中,该变量的分布被假设在一个特定的市道或状态条件下发生.会产生一系列重大变化.经典的自回归模型要求时间序列平稳,但实际上平稳性往往难以满足.金融模型的不稳定性可视为时间序列出现结构性的变化,即不同数据后面隐藏着不同状态,而引入市道轮换的自回归模型MS-AR(Markov switching model autoregressive model)可以解决这一难题^[22].下面以一阶市道轮换自回归模型MS-AR(1)为例进行介绍.

$$y_t-\mu_{S_t}=\beta(y_{t-1}-\mu_{S_{t-1}})+\varepsilon_t(1)$$

其中, S_t 是 t 时刻未知市道或者状态,本研究中只考虑两种市道的情况,即

$$S_t=\{1,2\}, \varepsilon_t\sim N(0, \sigma^2_{S_t})$$

$$y_t\sim N(\mu_{S_t}+\beta(y_{t-1}-\mu_{S_{t-1}}),\sigma^2_{S_t})(2)$$

$$y_t|S_{t-1}=i, S_t=j,$$

$$F_{t-1}\sim N(\mu_j+\beta(y_{t-1}-\mu_{S_j}),\sigma^2_{S_j})(3)$$

其中, $F_t=\{y_0, y_1, y_t\}$, F_t 指截止到时刻 t 所拥有的观测值 y_t 的所有信息,称为信息集或域流. $\{y_t\}$ 和 $\{S_t\}$ 即为隐马尔可夫链,可见其满足1.1节所述性质.

估计模型的参数采用极大似然方法,因此,求得似然函数的表达式是最重要的一步.考虑 $\{y_1, y_2, \dots, y_T\}$ 的似然函数 $f(y_1, y_2, \dots, y_T)$, 反复利用引理1可得

$$L=f(y_1, y_2, \dots, y_T)=\prod_{t=1}^T f(y_t|F_{t-1})$$

具体证明不再赘述,由此式可知求得 $f(y_t|F_{t-1})$ 的表达式即求得了似然函数的表达式,因此,以下主要围绕求 $f(y_t|F_{t-1})$ 的表达式展开.

$$f(y_t|F_{t-1})=\sum_i\sum_j f(y_t, S_{t-1}=i, S_t=j|F_{t-1})=$$

$$\sum_i\sum_j [f(y_t|S_{t-1}=i, S_t=j, F_{t-1})\times$$

$$P(S_{t-1}=i, S_t=j|F_{t-1})]$$

为方便起见定义

$$\eta_{i,j}(t)=f(y_t|S_{t-1}=i, S_t=j, F_{t-1});$$

$$\gamma_{i,j}(t)=P(S_{t-1}=i, S_t=j|F_{t-1}).$$

则下式成立

$$f(y_t|F_{t-1})=\sum_i\sum_j \eta_{i,j}(t)\gamma_{i,j}(t)(4)$$

由于市道 S_t 不可观测,只能通过计算 S_t 在不同时期的取值概率来推测 S_t 的取值,推断概率方法包括:事前概率、Hamilton滤波及Kim平滑概率.

事前概率指基于 $t-1$ 时刻信息集 F_{t-1} 对 t 时刻的市道进行推断的条件概率,记为 P_{it} .

$$P_{it}=P(S_t=i|F_{t-1})$$

Hamilton滤波概率指基于 t 时刻信息集 F_t 对 t 时刻的市道进行推断的条件概率,记为 $c_i(t)$,

$$c_i(t)=P(S_t=i|F_t)$$

这里 $c_i(0)=P(S_0=i)=\pi_i$, 其中, π_i 为 $\{S_t\}$ 位于状态 i 的初始分布, $i=1,2$.

Kim平滑概率是基于所有信息对 t 时刻的状态进行推断的条件概率,记为 $p_j(t)$,

$$p_j(t)=P(S_t=j|F_T)$$

$\gamma_{i,j}(t)=P(S_{t-1}=i, S_t=j|F_{t-1})$ 的计算过程分为预测和更新两个过程,设初始为 $c_i(0)=\pi_i$, 具体过程为:

1)预测.

$$\gamma_{i,j}(t)=P(S_{t-1}=i|F_{t-1})P(S_t=j|S_{t-1}=i, F_{t-1})=P(S_{t-1}=i|F_{t-1})P(S_t=$$

$$j|S_{t-1}=i)=c_i(t-1)a_{ij}(5)$$

2)更新.

$$P(S_{t-1}=i, S_t=j|F_t)=P(S_{t-1}=i, S_t=$$

$$j|F_{t-1}, y_t)=(f(y_t, S_{t-1}=i, S_t=j|F_{t-1}))/(f(y_t|F_{t-1}))=$$

$$(\eta_{i,j}(t)\gamma_{i,j}(t))/(\sum_m\sum_n \eta_{lm}, n(t)\gamma_{lm}, n(t))(6)$$

记 $\kappa_{i,j}(t)=P(S_{t-1}=i, S_t=j|F_t)$,在 t 时刻 y_t 被观测到后,信息集 F_{t-1} 被更新到 $F_t=\{F_{t-1}, y_t\}$,此时, $\kappa_{i,j}(t)$ 相当于 $\gamma_{i,j}(t)$ 的更新,即 $\gamma_{i,j}(t)$ 更新为

$$\gamma_{i,j}(t)=(\eta_{i,j}(t)c_i(t-1)a_{ij})/(\sum_m\sum_n \eta_{lm}, n(t)e_{mn}(t-1)a_{mn})(7)$$

当完成一轮预测和更新后,即完成式(5)至式(7)的计算后,可计算滤波概率,即式(8). 然后继续开始新一轮计算.

$$c_j(t)=\sum_i P(S_{t-1}=i, S_t=j|F_t)=$$

$$\sum_i \gamma_{i,j}(t)(8)$$

当上述全部完成时,可计算有对数似然函数.用 $\kappa_{i,j}(t)$ 代替 $\gamma_{i,j}(t)$, 有对数似然函数为

$$\ln L=\sum_{t=1}^T \ln f(y_t|F_{t-1})=$$

$$\sum_{t=1}^T \ln [\sum_i\sum_j \eta_{i,j}(t)\kappa_{i,j}(t)](9)$$

1.3 马氏市道轮换波动率替代模型

由于日内数据可以记录到很高频率的交易数据,日均大约几千个数据,此时记录之间的时间间隔已非常小,所以,对于高频数据的建模,自然的想法是利用连续时间模型,连续时间金融理论主要是考虑使用半鞅过程来刻画价格或者收益率的变动过程.考虑金融资产的连续时间对数价格过程(或者对数收益率过程) $\{Y(u)\}$, 称其是一个半鞅,若有如下分解:

$$Y(u)=A(u)+M(u)$$

其中,过程 $\{A(u)\}$ 是一个有界变差的过程; $\{M(u)\}$ 是一个局部鞅.连续时间金融理论中使用半鞅的原因在于,只有在半鞅上才能定义随机积分^[22].它被广泛用于衍生品(期权,期货及债券)的定价中,对于半鞅而言,一般用它的二次变差过程衡量其波动性.二次变差 QV 可被定义为已实现方差随抽样频率趋于无穷的概率极限,

$$QV=P\text{-}\lim_{\Delta t\downarrow 0} RQV$$

其中,定义 RQV 为已实现方差.而波动率定义为 $(QV)^{1/2}$,故波动率可以用已实现方差的平方根估计,称 $(RQV)^{1/2}$ 为已实现波动率,记为 RV .

BOLLERSLEV^[23]提出的广义GARCH模型,因为能很好的解释波动率汇聚和收益率厚尾等现象,已成为研究和预测时变波动率的标准工具. GARCH模型通常用于低频数据,如日、周、月数据,高频数据包含更丰富的信息,而与高频数据相结合成为一个重要的方向.为使GARCH模型与日内的高频数据结合发挥模型优势,VISSER^[24]提出基于GARCH模型的尺度模型和波动率替代模型,将日内的高频数据嵌入到模型的框架中,且指出通过构造合适的波动率替代,可以改进模型的伪极大似然估计的渐近方差.考虑日内对数收益率 $R_t(u)$, 日内波动率可以被已实现波动率 RV_t 代替.下面介绍尺度模型和波动率替代.

$$R_t(u)=(v_t\tau)^{1/2}\psi_t(u), 0\leq u\leq 1(10)$$

$$v_t=1+\alpha\tau^2_{t-1}+\beta v_{t-1}(11)$$

其中, $\psi_t(u)$ 服从均值为0、方差为1的标准分布; $\{\psi_0(u), \psi_1(u), \dots\}$ 独立同分布.两者为尺度模型. $R_t(0)$ 为隔夜收益率,而 $R_t(1)$ 即 r_t , 故当 $t=1$ 时,尺度模型即转代为GARCH(1,1).

VISSER^[24]定义了替代波动率的统计量,称为波动率替代.一般地,称随机变量 $H_t=H(R_t(\cdot))$ 为波动率替代,如果 H 为正且满足正齐次性,则 $H(sR_t(\cdot))=sH(R_t(\cdot))$, $s\geq 0$

波动率替代一般有 $|r_t|$, 日内最高-最低价差、已实现波动率等.

$$H(R_t(\cdot))=H((v_t\tau)^{1/2}\cdot\psi_t(\cdot))=$$

$$(v_t\tau)^{1/2}H(\psi_t(\cdot))(12)$$

记 $\mu^H=E(H(\psi_t(\cdot)))$, $0\leq \mu^H<\infty$, $\tau_H=\tau(\mu^H)^2$, $Z_{H,t}=H(\psi_t(\cdot))/\mu^H$, 则

$$H_t=(v_t\tau_H)^{1/2}Z_{H,t}(13)$$

$$v_t=1+\alpha r_{t-1}^2+\beta v_{t-1} \quad (14)$$

即波动率替代模型. 令 $F_t=\{R_t, R_{t-1}, \dots\}$, 有 $E(H_t^2|F_{t-1})=v_t \tau_H$.

伴随经济政策的出台和金融市场监管制度的变迁自身完善,在一个不成熟的金融市场上,马氏轮换是一个普遍存在的现象,对这种结构性变换的描述是必要的^[25]. 而GRAY等^[26-27]在GARCH方程中引入马氏轮换,建立MRS-GARCH模型,利用滞后的条件方差取期望值,从而避免了路径依赖问题,通过检验表明MRS-GARCH模型较传统的GARCH模型在拟合和预测能力上均有提高. 由于尺度模型和波动率替代模型是在GARCH模型基础上提出,因此,很自然考虑到在这两个模型上加入马氏市道轮换. 这里首先对尺度模型(10)和(11)加入市道轮换,即

$$R_t(u)=(v_{s_t, t} \tau_{s_t})^{1/2} \psi_t(u) 0 \leq u \leq 1 \quad (15)$$

$$v_{s_t, t}=1+\alpha_{s_t} r_{t-1}^2+\beta_{s_t} v_{s_t, t-1} \quad (16)$$

由式(12)知H具有齐次性,用式(12)所述H作用于式(15),由齐次性即可得到马氏市道轮换波动率替代模型

$$H_t=(v_{s_t, t} \tau_{s_t})^{1/2} Z_{H,t} \quad (17)$$

$$v_{s_t, t}=1+\alpha_{s_t} r_{t-1}^2+\beta_{s_t} v_{s_t, t-1} \quad (18)$$

其中, $F_t=\{H_t, H_{t-1}, \dots\}$; $E(H_t^2|F_{t-1}, s_t=i)=v_{i,t} \tau_{Hi}$. 假设 $Z_{H,t}$ 服从标准正态分布,则 $H_t|F_{t-1}, s_t=i \sim N(0, v_{i,t} \tau_{Hi})$.

将马氏轮换引入GARCH模型后存在严重的路径依赖问题,即t时刻的条件方差将依赖于所有变量t时刻之前全部时刻的取值及其状态. 当样本较大时,这使模型异常复杂难以估计. 同样,将马氏轮换引入波动率替代模型也存在路径相依问题,因此,可模仿MRS-GARCH的估计方法. GRAY^[26]利用一期条件方差的期望值代替条件方差本身,从而解决了一模型中存在的路径依赖问题. 此后,KLAASSEN^[27]充分利用信息,进一步改善模型. 本研究所使用的方法,正是基于KLAASSEN所提出的方法. 具体来说,式(14)中等号右边条件方差项的表达式为

$$v_{s_t, t-1}=E(v_{S_{t-1}, t-1}|S_t, F_{t-1})=$$

$$P_{1s_t, t-1} v_{1, t-1} + P_{2s_t, t-1} v_{2, t-1} \quad (19)$$

$$P_{ji, t-1}=P(S_{t-1}=j|S_t=i, F_{t-1})=$$

$$(P(S_t=i|F_{t-1}, s_{t-1}=j)P(S_{t-1}=j|F_{t-1})) / (P(S_t=i|F_{t-1}))=$$

$$(a_{ji} e_j(t-1)) / (p_{it})(20)$$

为方便计算 $e_j(t)$, 引入符号 η_{jt} , 即

$$\eta_{jt}=P(H_t|S_t=j, F_{t-1})=1 / ((2\pi v_{i,t} \tau_{Hi})^{1/2}) e^{-(H_t^2 / (2v_{i,t} \tau_{Hi}))} \quad (21)$$

$$e_j(t)=P(S_t=j|F_t)=P(S_t=j|F_{t-1}, H_t)=$$

$$(P(H_t, S_t=j|F_{t-1})) / (P(H_t|F_{t-1}))=$$

$$(P(H_t|S_t=j, F_{t-1})P(S_t=j|F_{t-1})) / (\sum_i P(H_t|S_t=i, F_{t-1})P(S_t=i|F_{t-1}))=$$

$$(\eta_{jt} p_{jt}) / (\sum_{i=1}^2 \eta_{it} p_{it}) \quad (22)$$

$$p_{it}=P(S_t=i|F_{t-1})=$$

$$\sum_{j=1}^2 P(S_t=i, S_{t-1}=j|F_{t-1})=$$

$$\sum_{j=1}^2 a_{ji} e_j(t-1) \quad (23)$$

$$\hat{f}(H_t|F_{t-1})=\sum_{i=1}^2 \hat{f}(H_t, S_t=i|F_{t-1})=$$

$$p_{1t} \eta_{1t} + p_{2t} \eta_{2t} \quad (24)$$

由上式迭代可得出似然函数L

$$L=\sum_{t=1}^T \ln(p_{1t} \eta_{1t} + (1-p_{1t}) \eta_{2t}) \quad (25)$$

2 模型实证研究

选用2017- 01- 03 09:35至2017- 08- 02 14:55上证综指每5 min的收盘价格数据作为研究对象. 为方便观察和分析,将收盘价缩小100倍,调整后的收盘价时序和直方图见图1和图2.

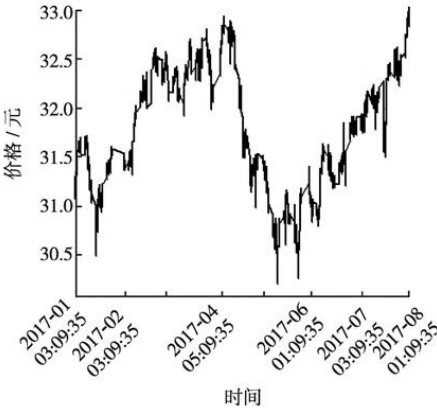


图 1 收盘价时序图

Fig. 1 Closing price sequence diagram

图1 收盘价时序图

Fig.1 Closing price sequence diagram

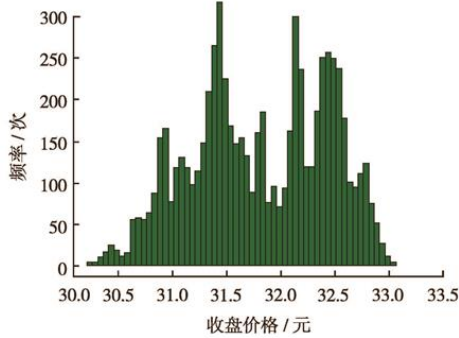


图 2 收盘价直方图
Fig. 2 Closing price histogram

图2 收盘价直方图
Fig.2 Closing price histogram

图3为收益率的时序图. 可见,收益率在固定范围内变化,同时波动率存在聚集性,即大的波动率后面常常伴随较大的波动率,小的波动率后面也常常伴随小的波动率. 统计分析得出上证综指每5 min 收盘价的基本分析结果如表1. 可见,收盘价在30 元. 在22~33.05 元(收盘价缩小100倍之后的价格)内波动,平均价格为31.80 元.

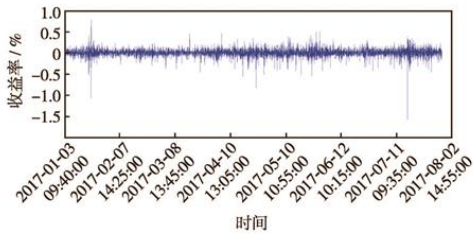


图 3 收益率时序图
Fig. 3 Return sequence diagram

图3 收益率时序图
Fig.3 Return sequence diagram

表 1 基本统计分析
Table 1 Basic statistical analysis

指标	日期时刻	收盘价/元
最小值	2017-01-03 09: 35: 00	30. 22
上四分位数	2017-02-28 12: 41: 15	31. 34
中位数	2017-04-21 00: 17: 30	31. 80
平均数	2017-04-21 02: 09: 02	31. 80
下四分位数	2017-06-14 11: 53: 45	32. 36
最大值	2017-08-02 15: 00: 00	33. 05

表1 基本统计分析
Table 1 Basic statistical analysis

另外, LB检验结果显示 $p\text{-value} < 2.2 \times 10^{-16}$,拒绝原假设,收盘价序列不是白噪声,因此,有分析价值; ADF检验结果表明,检验统计量值为-1.848 2,大于显著性水平0.05下的分位数-2.86,则收盘价序列非平稳,而马氏市道轮换转换模型可解决这一问题.

首先,用MS-AR(1)模型对高频数据进行分析. 这里高频数据的需要估计参数向量为 $\theta=[\mu_1, \mu_2, \sigma_1, \sigma_2, \beta, p, q]$.

使用Hamilton滤波算法计算似然函数,进行极大似然估计,得出参数的极大似然估计值如表2. 分析可见,熊市为市道1,牛市为市道2,熊市均值为30.499 0 元,标准差为0.049 0; 牛市的均值为30.527 0,标准差为0.008 4. 均值均在收盘价区间内,且熊市的标准差比牛市的大,这说明熊市价格震荡幅度更大.

转移概率矩阵 $A=[0.107\ 7\ 0.892\ 3$

$0.214\ 6\ 0.785\ 4]$.

由此可见,牛市转换成牛市的概率高,即牛市的第2天还是牛市的概率更大,熊市转换成牛市的概率也很大.

表 2 MS-AR (1) 参数估计¹⁾
Table 2 MS-AR (1) parameter estimation

参数	初始估计	优化参数
μ_1	31.0	30.499 0
μ_2	33.0	30.527 0
σ_1	1.8	0.049 0
σ_2	2.0	0.008 4
β	0.3	0.995 1
p	0.6	0.107 7
q	0.8	0.785 4
$\ln L$	-11 467.0	18 485.000 0

¹⁾ 该统计日期同表 1

表2 MS-AR(1)参数估计 1)
Table 2 MS-AR(1)parameter estimation

下面进行市道解码,即把2017- 01- 03至2017- 08- 02上证综指每5 min收盘价的市道推断出来,采用Kim平滑概率,市道即对应平滑概率较大的状态;滤波概率和平滑概率见图4和图5.

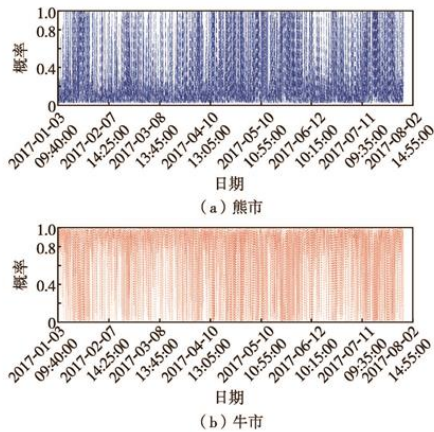


图 4 滤波概率图
Fig. 4 (Color online) Filter probability figure

图4 滤波概率图
Fig.4 (Color online)Filter probability figure

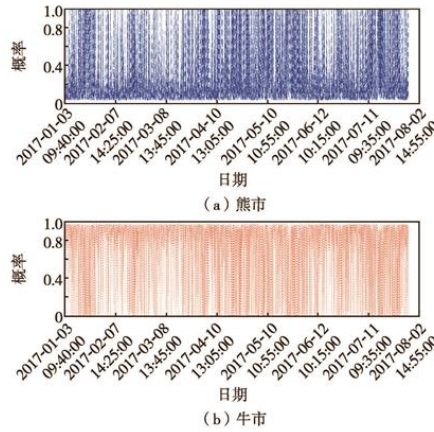


图 5 平滑概率图
Fig. 5 (Color online) Smoothing probability figure

图5 平滑概率图
Fig.5 (Color online)Smoothing probability figure

为看清楚高频数据状态预测之间的变化,选取了2017- 04- 24的数据,如图6. 可见,高频数据之间存在明显的市道转换.

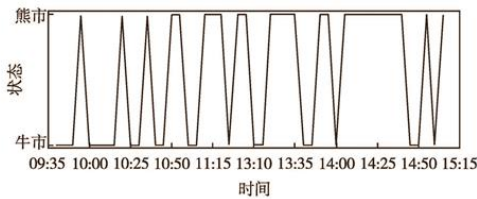


图 6 状态转换图
Fig. 6 Regime switching figure

图6 状态转换图
Fig.6 Regime switching figure

为评价模型的预测效果,引入平均相对误差ARPE为

$$ARPE=(\sum_{t=2}^T (|y_t-(\hat{y}_t)|)/(|y_t|))/(T-1)$$

其中, y_t 为真实值, (\hat{y}_t) 为预测值; $t=1,2,\dots,T$, T 为总时间.

通过评估可知,预测价格与真实价格之间的平均相对误差为 $7.493\ 2\times10^{-4}$,平均相对误差较小,说明预测效果相对较好,为清楚地看到预测效果,只选取2017- 08- 01至2017- 08- 02内收盘价作预测图,见图7.

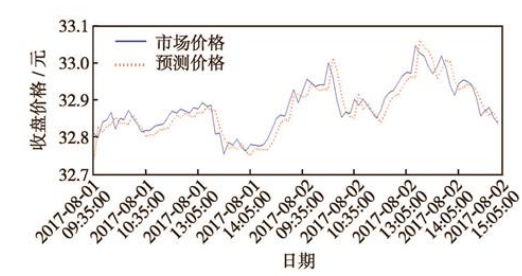


图 7 收盘价预测图

Fig.7 (Color online) Closing price forecast figure

图7 收盘价预测图

Fig.7 (Color online)Closing price forecast figure

这里给出2017- 08- 02 15:00的点预测,此时真实价格为32.850 6元,模型预测价格为32.818 3元,预测的标准差为0.016 9元,效果比较理想.

下面对高频数据进行波动率研究,从2017- 01- 04上证综指的日内数据开始. 5 min高频数据对于进行波动率研究是合理的. 唐勇^[28]以上证综指高频数据为例,实证了5 min抽样频率数据的合理性. LIU等^[29]研究了超过400种不同的波动率估计量,实证发现,至少在统计意义上很难显著地击败以5 min 抽样数据所估计的波动率.

极大似然估计参数如表3. 可见,状态2时波动率较大,状态1时波动率较小,且高波动率之后依旧是高波动率的概率较大.

表 3 波动率的参数估计¹⁾

Table 3 Volatility parameter estimation

参数	初始估计	优化参数
α_1	0. 360 0	0. 359 9
α_2	0. 810 0	0. 810 1
β_1	1. 690 0	1. 300 3
β_2	0. 360 0	0. 947 4
τ	9. 000 0	$8. 341\ 6\times10^{-6}$
p	0. 700 0	0. 457 0
q	0. 400 0	0. 708 0
$\ln L$	-475. 562 2	511. 559 6

¹⁾与表1、表2 统计日期一致

表3 波动率的参数估计 1)

Table 3 Volatility parameter estimation

3 结语

高频数据不仅丰富了金融资产价格信息,而且还包含众多其他维度的信息,如交易时间间隔、交易量及买卖价差等,这些可以对金融市场进行更精细的分析,对于理解市场价格形成和信息的传递机制等市场微观结构方面的特征,具有相当重要的作用. 本研究通过市道轮换对高频数据进行分析研究,包括市场价格和波动率两部分,发现日内数据中存在市道轮换,并对市场价格未来做了比较准确的预测,贴近事实,吻合市场. 同时模型参数少,简洁实用,可操作性强,可为市场投资策略提供参考和指导. 在本研究的后半部分,尝试用市道轮换下的波动率替代模型研究高频数据的波动率,并给出分析模型和方法. 未来研究可尝试通过改进市道轮换下的波动率替代模型对高频数据的波动率进行进一步的估计和预测,并且尝试研究超高频数据. 另外,半马氏能够克服马氏轮换过程一些缺点,如无记忆性^[30-32],未来可以把马氏市道轮换下的高频数据模型推广到半马氏的情形,同时讨论带跳的估计问题.

参考文献

[1] 朱建平,魏 瑾,谢邦昌. 金融高频数据挖掘研究评述与展望[J]. 经济动态,2011(6):59- 62.

ZHU Jianping, WEI Jin, XIE Bangchang. Review and prospect of research on financial high frequency data mining[J]. Economic Perspectives, 2011(6): 59- 62.(in Chinese)

[2] 雷井生,林 莎. 基于高频数据的统计套利策略及实证研究[J]. 科研管理,2013,34(6):139-145.

LEI Jingsheng, LIN Sha. Statistical arbitrage strategies based on high frequency data[J]. Science Research Management, 2013, 34(6): 139-145.(in Chinese)

[3] HANSEN P R, HUANG Z, SHEK H. Realized GARCH: a joint model of returns and realized measures of volatility[J]. Journal of Applied Econometrics, 2012, 27(6): 877-906.

[4] 王天一,赵晓军,黄 卓. 利用高频数据预测沪深300指数波动率-基于Realized GARCH模型的实证研究[J]. 世界经济文汇,2014(5):18-30.

WANG Tianyi, ZHAO Xiaojun, HUANG Zhuo. Forecasting the volatility of Shanghai-Shenzhen 300 index by using high frequency data: an empirical study based on realized GARCH model[J]. World Economic Papers, 2014(5): 18-30.(in Chinese)

[5] ZHANG Changyong. Modeling high frequency data using Hawkes processes with power-law kernels[J]. Procedia Computer Science, 2016, 80: 762-771.

[6] 李胜歌,张世英. 高频金融数据的两种波动率计算方法比较[J]. 系统管理学报,2007,16(4):426- 431.

LI Shengge, ZHANG Shiyong. Comparative study of two volatility estimation methods of high frequency financial data[J]. Journal of Systems & Management, 2007, 16(4): 426- 431.(in Chinese)

[7] 唐 勇,张世英. 高频金融数据的加权已实现极差波动及其实证分析[J]. 系统工程,2006, 24(8):52-57.

TANG Yong, ZHANG Shiyong. Weighted realized range-based volatility based on high-frequency data and its empirical analysis[J]. Systems Engineering, 2006, 24(8): 52-57.(in Chinese)

[8] 周泽宇. 基于高频数据的跳扩散模型的参数估计[D]. 北京:北方工业大学,2015.

ZHOU Zeyu. The jump diffusion models parameter estimation method of high-frequency data[D]. Beijing: North China University of Technology, 2015.(in Chinese)

[9] 徐正国,张世英. 调整“已实现”波动率与GARCH及SV模型对波动的预测能力的比较研究[J]. 系统工程,2004,22(8):60- 63.

XU Zhengguo, ZHANG Shiyong. The comparative research on volatility prediction ability of adjusted realized volatility, GARCH model and SV model[J]. Systems Engineering, 2004, 22(8):60- 63.(in Chinese)

[10] 杨 科,陈浪南. 基于C_TMPV的中国股市高频波动率的跳跃行为研究[J]. 管理科学,2011,24(2):103-112.

YANG Ke, CHEN Langnan. Empirical research on jump behavior of high-frequency volatility based on C_TMPV in Chinese stock markets[J]. Journal of Management Science, 2011, 24(2):103-112.(in Chinese)

[11] 张 伟,李 平,曾 勇 中国股票市场个股已实现波动率估计[J]. 管理学报,2008, 5(2):269-273.

ZHANG Wei, LI Ping, ZENG Yong. Estimating realized volatility of individual stock in Chinese stocks market[J]. Chinese Journal of Management, 2008, 5(2):269-273.(in Chinese)

[12] 沈根祥. 沪深300指数日内跳的Hausman检验[J]. 数理统计与管理,2010,29(4):713-718.

SHEN Genxiang. Hausman test for the intraday jumps in Shanghai-Shenzhen 300 index[J]. Journal of Applied Statistics and Management, 2010, 29(4): 713-718.(in Chinese)

[13] 唐 勇,张伯新. 基于高频数据的中国股市跳跃特征实证分析[J]. 中国管理科学,2013, 21(5):30-39.

TANG Yong, ZHANG Boxin. The empirical analysis of jump characteristics of China's stock market based on high frequency data[J]. Chinese Journal of Management,2013, 21(5):30-39.(in Chinese)

[14] BAJGROWICZ P, SCAILLET O, TRECCANI A. Jumps in high-frequency data: spurious detections, dynamics and news[R].[S.l.]: Swiss Finance Institute, 2012.

[15] 汪先珍. 中国股市价格的跳跃行为——基于上证综指高频数据的分析[D]. 厦门:厦门大学,2009.

WANG Xianzhen. The jump behavior of China's stock market prices: an analysis with the high frequency data of the SSECI[D]. Xiamen: Xiamen University, 2009.(in Chinese)

[16] HAMILTON J D. A new approach to the economic analysis of nonstationary time series and the business cycle[J]. Econometrica, 1989, 57(2): 357-384.

[17] CAI J. A Markov model of switching-regime ARCH[J]. Journal of Business & Economic Statistics, 1994, 12(3): 309-316.

[18] HAMILTON J D, SUSMEI R. Autoregressive conditional heteroskedasticity and changes in regime[J]. Journal of Econometrics,1994, 64: 307-333.

[19] 佟 杰. 基于改进的隐马尔科夫体制转换ARMA-GARCH模型的高频数据预测[D]. 北京: 北京理工大学, 2015.

TONG Jie. Improved hidden Markov ARMA-GARCH model forecasting in high frequency sample[D]. Beijing: Beijing Institute of Technology, 2015.(in Chinese)

[20] 刘金全, 郑挺国. 利率期限结构的马尔科夫区制转移模型与实证分析[J]. 经济研究, 2006,41(11):82-91.

LIU Jinquan, ZHENG Tingguo. Markov regime switching model and empirical analysis of the term structure of interest rates[J]. Economic Research Journal, 2006, 41(11): 82-91.(in Chinese)

[21] 杨宝臣, 苏云鹏. SHIBOR市场风险溢价动态特性的机制转换[J]. 系统工程学报, 2012, 27(2): 202-207.

YANG Baochen, SU Yunpeng. Regime switching in risk premium dynamics of SHIBOR[J]. Journal of Systems Engineering, 2012, 27(2): 202-207.(in Chinese)

[22] 王 犇. 隐马尔可夫链、马尔可夫状态转换模型及在量化投资中的应用[M]. 北京: 清华大学出版社, 2016: 110-144.

WANG Ben. Hidden Markov chain, Markov switching model, and the applications in quantitative finance[M]. Beijing: Tsinghua University Press, 2016: 110-144.(in Chinese)

[23] BOLLERSLEV T. Generalized autoregressive conditional heteroskedasticity[J]. Journal of Econometrics, 1986, 31(3): 307-327.

[24] VISSER M P. Garch parameter estimation using high-frequency data[J]. Journal of Financial Econometrics, 2011, 9(1): 162-197.

[25] 孙金丽, 张世英. 具有结构转换的GARCH模型及其在 中国股市中的应用[J]. 系统工程, 2003, 21(6):87-91.

SUN Jinli, ZHANG Shiyng. Regime-swithing GARCH in China's stock market[J]. Systems Engineering, 2003, 21(6): 87-91.(in Chinese)

[26] GRAY S F. Modeling the conditional distribution of interest rates as regime-switching process[J]. Journal of Financial Economics, 1996, 42(1): 27- 62.

[27] KLASSEN F. Improving GARCH volatility forecasts with regime-switching GARCH[J]. Empirical Economics, 2002, 27(2): 363-394.

[28] 唐 勇. 金融资产跳跃检验方法实证比较[J]. 中国管理科学, 2012, 20: 290-299.

TANG Yong. The empirical comparison among jump tests of financial asset based on high frequency data[J]. Chinese Journal of Management Science, 2012, 20: 290-299.(in Chinese)

[29] LIU L, PATTON A J, Sheppard K. Does anything beat 5-minute RV? a comparison of realized measures across multiple asset classes[J]. Journal of Econometrics, 2015, 187(1): 293-311.

[30] SILVESTROV D, STENBERG F. A pricing process with stochastic volatility controlled by a semi-Markov process[J]. Communications in Statistics: Theory and Methods, 2004, 33(3): 591- 608.

[31] ZHU Quanxin. Razumikhin-type theorem for stochastic functional differential equations with Levy noise and Markov switching[J].International Journal of Control, 2017, 90(8): 1703-1712.

[32] ZHU Quanxin. Asymptotic stability in the pth moment for stochastic differential equations with Levy noise[J]. Journal of Mathematical Analysis and Applications, 2014, 416(1): 126-142.