

# 沪深股票市场间的时变波动和相关性

## --基于五分钟和十五分钟数据的检验

陈守东, 韩广哲

(吉林大学数量经济研究中心, 吉林 长春 130012)

**摘要:** 本文以上海和深圳股票市场的高频数据为样本, 对中国股市的相关性和波动性进行时变分析。基于上证综指和深证成指的日内五分钟和十五分钟数据建立二元GARCH (Bivariate GARCH) 模型, 模拟计算出沪、深股票市场时变的条件方差、条件协方差和条件相关系数。研究结果表明两个股票市场高频收益率序列的条件波动之间存在着较强的正相关性。

**关键词:** 二元GARCH; 高频数据; 时变

**中文分类号:** F 224.0      **文献标识码:** A

### 1 引言

近年来, 对金融高频数据的研究与分析已成为学术界和实务界的热点问题和难点问题, 很多学者致力于研究用高频数据来解释关于市场行为、市场微观结构的问题。高频数据分析与市场微观结构理论是紧密联系在一起的。证券市场微观结构即证券市场的交易机制, 是指证券交易价格形成与发现的过程与运作机制。市场微观结构的核心是价格发现功能, 后者也是整个证券交易市场最核心的环节。证券流通市场的微观结构将影响市场价格波动、流动性以及潜在的投资者数量和交易量, 这正是市场微观结构的意义所在, 而高频数据分析是解释市场微观结构极为有效的手段。金融高频数据相关的研究及其进展, 为更好地解释金融市场的微观结构开辟了新的途径, 对中国的金融市场无论是市场交易及其效率还是市场监管都具有非常重要的理论意义和实际意义。

Robert Wood最早创建研究市场微观机构(金融高频)数据库。Wood(2000)从金融市场微观结构研究的初衷、结构数据的基础检验、TAQ数据库的组织形式和特征等角度对金融高频数据库的发展历程做了介绍, 并且讨论了金融高频数据量(如NASDAQ数据等)的快速增长趋势以及这种数据量的增长趋势在市场结构研究中的应用问题。Harris (1990)等人最早利用金融高频数据来分析市场微观结构, 他们研究了NYSE高频交易数据日内(intraday)收益与波动性时间序列的模式; 而Goodhart(1991,1997)等人是对外汇市场的高频交易数据进行研究。近几年, 关于对市场微观结构的实证研究在深度和广度方面又有了新的进展, 其中尤其以对股票市场高频数据的分析最具代表性。Hasbrouk(1999)、Zhang、Russell和Tsay(2001)利用高频交易数据对某一个特殊股票的报价与询价的动态性进行研究; Bollerslev、Zhang (2003)将股票市场的高频交易数据应用于对因素定价模型(factor pricing models)中系统风险因素的计量和建模等一系列

的相关研究。

与低频金融数据的统计特征不同，金融高频数据在较短期间内有着增长性的拖尾趋势（heavy tailed），并且数值具有离散性非正态性、非线性、均值回复和周期性等特点。Polson (2000)的研究发现S&P500指数的日收益数据具有非正态性；Jobson和Korkie (1980)的研究表明在决定最优证券组合的输入变量的均值—方差模型中，方差—协方差/期望收益与最优组合的权重之间的映射是高度非线性的；Andersen、Bollerslev(1997)引入可变傅立叶形式回归（Flexible Fourier Form Regression，简称FFF回归）方法能够很有效的确定高频数据的周期成分；Martens(2002)等说明在波动率模型中加入周期成分能够有效的提高其预测效果。

国内对于高频数据的研究起步于近三年，而且用于实证研究的样本数据绝大多数是五分钟频率数据，主要研究工作主要分为三个方面：一是对国外高频数据相关理论的引进与介绍（常宁等（2004）、周建等（2004）、郭兴义等（2002））；二是分析我国证券市场高频数据的基本统计特征，并与国际成熟市场作比较；三是利用我国证券市场的历史高频数据，对国外已有相关理论模型进行检验，研究日内价格模式的异同。

王春峰（2004）和方兆本（2004）等研究了我国证券市场高频数据的基本形态特征，实证得出上海股票市场波动性的日内周期特征和长记忆特征；吴世农（2002）等运用高频数据对我国深圳股票市场的买卖报价价差的变动模式进行实证分析，同时研究股票买卖报价价差的影响因素和成因，揭示了我国股票市场的微观结构特征；针对中国股票市场波动率特征，陈浪南（2003）等进行了有关高频估计、特性（分布、不对称、长期记忆）以及模拟与预测等各方面的研究；吴冲锋（2004）等应用高频数据分析股指与股指期货日内互动关系，研究发现标准普尔500指数现货市场与其期货市场收益率之间存在即时互动关系，股指期货先行时间明显比股指先行时间要长，股指与股指期货对不同类型的信息反映速度不一致。

Engle在1982年首先提出了ARCH模型对方差进行建模，1986年Bollerslev将ARCH模型推广，发展成为广义的ARCH模型，即GARCH模型。随后的十多年中，计量经济学家们对基本的GARCH模型进行了许多变形，现在已发展成为一个包含众多方法的模型类别。大量实证研究表明(Mills,T. C.,1999)，GARCH类模型特别适合于对金融时间序列数据的波动性和相关性进行建模，估计和预测波动性和相关性。Engle和Kroner（1993）提出了多元GARCH模型，并讨论多元情形下的参数估计方法。Engle（2000）对高频数据的计量经济学进行了综述。

本文以上海和深圳股票市场的综合指数——上证综指和深证成指为研究对象，着眼于建立反映上海和深圳两个股票市场相互影响关系的动态模型。我们使用二元GARCH模型对上海和深圳股票市场的收益率进行建模，样本数据是五分钟和十五分钟频率数据，计算得到条件相关性的时间序列。从二元和高频角度刻画两个市场间时变的波动和相关性。提出当金融资产之间存在着共同特征时，就会存在着Granger因果关系，也就是说资产收益之间存在着领先—滞后关系，相互影响的程度和方向存在着区别。无条件相关性和条件相关性之间的区别是显而易见的。条件相关性是一个时变的参数，不象很多经济模型所假设的那样存在着参数的真实值。当然，不管模型是基于常量相关的还是时变相关的，所得到的相关性估计量都会随着时间而变化，只不过常量相关模型中的相关性估计量的变化是因为取样的不同，而时变参数模型中相关性估计量的变化来源于参数真实值的变化。本文结构如下：第二部分介绍二元GARCH模型，第三部分对数据与模型进行检验和估计，第四部分是结论。

## 2 二元GARCH模型

首先介绍一元GARCH(p,q)模型。数据生成过程（也称均值方程）为：

$$R_t = \alpha + \sum_{i=1}^m \theta_i R_{t-i} + \sum_{j=1}^n \eta_j \varepsilon_{t-j} + \varepsilon_t$$

这里假设收益率  $R_t$  服从ARMA(m,n)过程，m、n随每个具体收益率序列而确定，其中残差序列  $\varepsilon_t$  是条件异方差过程，表示收益率波动的爆发性、聚类性和持续性，一般假设条件异方差序列满足以下的方差过程（也称条件方差方程，Bollerslev,1986）：

$$\sigma_t^2 = \omega + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^p \beta_j \sigma_{t-j}^2, \quad \omega > 0, \quad \alpha_i \geq 0, i=1, \dots, q, \beta_j \geq 0, j=1, \dots, p$$

这里，p是GARCH项的阶数，q是ARCH项的阶数。

当估计两个收益率序列之间的波动性和相关性时，可以建立二元GARCH（Bivariate GARCH）模型(Engle and Kroner, 1995)。二元GARCH模型有两个条件均值方程、三个条件方差方程。条件均值方程一般都采用简单的形式，假定收益率等于一个常数加上误差项：

$$R_{1t} = \mu_1 + \varepsilon_{1t}$$

$$R_{2t} = \mu_2 + \varepsilon_{2t}$$

三个条件方差方程中，一个是两收益率间的条件协方差方程，另外两个分别是两个收益率的条件方差方程。对二元GARCH模型使用Vech参数化方法，则二元GARCH(1,1)模型的条件方差方程为：

$$\sigma_{1,t}^2 = \omega_1 + \alpha_1 \varepsilon_{1,t-1}^2 + \beta_1 \sigma_{1,t-1}^2$$

$$\sigma_{2,t}^2 = \omega_2 + \alpha_2 \varepsilon_{2,t-1}^2 + \beta_2 \sigma_{2,t-1}^2$$

$$\sigma_{12,t}^2 = \omega_3 + \alpha_3 \varepsilon_{1,t-1} \varepsilon_{2,t-1} + \beta_3 \sigma_{12,t-1}^2$$

如同一元GARCH模型，二元GARCH模型中的系数也应有非负约束，从而保证协方差矩阵的正定性。在任一时点上，将协方差估计除以两序列标准差估计之积就可得到相关性的估计，即

$$\rho_{12,t} = \frac{\text{COV}_{12,t}}{\sqrt{\sigma_{1,t}^2 \sigma_{2,t}^2}}$$

### 3 数据、模型检验及估计

#### 3.1 数据描述

本文使用的上证综指（HU）和深证成指（SHEN）的五分钟、十五分钟数据来源于分析家软件，这些数据已经考虑了还权的问题，所以得到的指数值能够直接用于收益的计算。样本区

间为2004年12月13日到2005年7月1日，132个交易日，五分钟数据的样本观察值个数是6336，十五分钟数据的样本观察值个数是2112。图1—图4给出了上证综指、深证成指以及它们五分钟收益率的图形。我们以RHU和RSHEN上证综指和深证成指的对数收益率序列，计算公式如下：

$$RHU = 100 \times DLOG(HU) \quad RSHEN = 100 \times DLOG(SHEN)$$

从图1—图4中我们可以看出，上海和深圳股票市场五分钟数据的走势及收益率图形非常相近，波动增大、减小的幅度以及持续的时间相近，描述统计量也表明两市收益率序列的均值、标准差、峰度和偏度很接近，尖峰特征显著。检验收益率序列和收益率平方序列自相关和偏相关的Ljung-Box Q统计量（滞后阶数：1—200）在1%水平下显著，说明收益率序列存在着波动聚类现象，而且两个股票市场的波动聚类的表现（持续时间、波动幅度）非常相似。GARCH模型能够很好的刻画波动聚类现象，本文同时考虑上证综指和深证成指的收益率序列，把两个市场联系到一起，建立二元GARCH（1，1）（Bivariate GARCH）模型。（十五分钟数据的表现与五分钟数据相同。限于篇幅，检验结果略。）

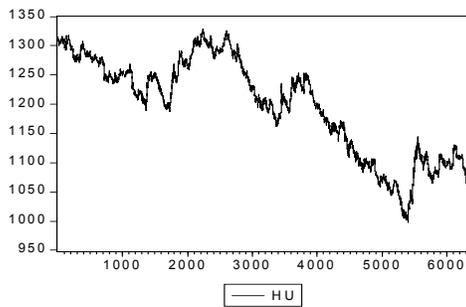


图1 上证综指

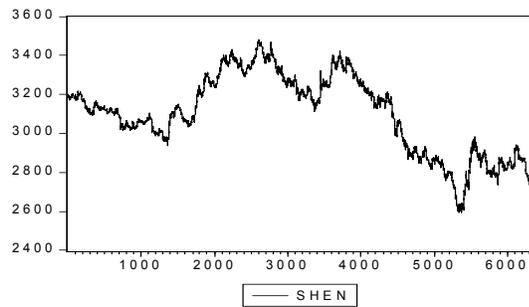


图2 深证成指

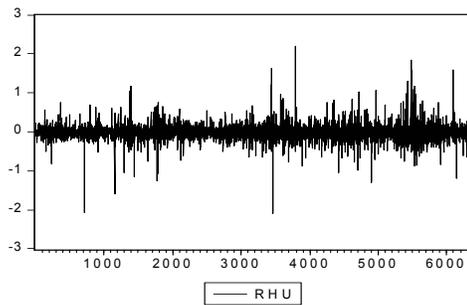


图3 上证综指五分钟收益率

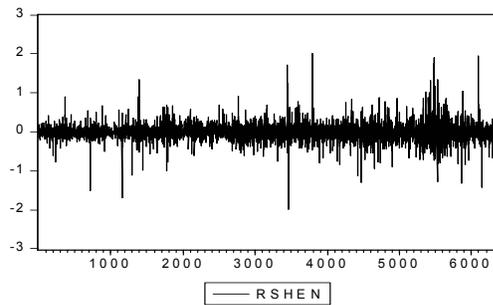


图4 深证成指五分钟收益率

### 3.2 沪深股市收益率间的无条件相关系数

在全部样本区间内，上证综指和深证成指五分钟收益率序列之间的无条件相关系数  $\rho = 0.898$ ，上证综指和深证成指十五分钟收益率序列之间的无条件相关系数  $\rho = 0.919$ ，表明沪深股市高频收益率序列之间存在很强的正相关性。

### 3.3 单位根检验

对上证综指和深证成指五分钟和十五分钟的对数序列及收益率序列进行单位根检验，确定

各序列的平稳性特征。计算单位根检验的ADF统计量(Augmented Dickey-Fuller统计量)和PP统计量(Phillips-Perron统计量), 在1%的显著性水平下, 指数对数序列均接受存在单位根的原假设, 收益率序列均显著拒绝存在单位根的原假设, 这说明指数对数序列都是一阶单整(非平稳)的, 收益率序列是平稳过程。(检验结果略)

### 3.4 沪深股市收益率的Granger因果关系检验

在1%置信水平下, 上证综指和深证成指五分钟收益率序列在在滞后1阶至240阶的情况下均存在显著的互为因果关系, 这说明沪深股市五分钟收益率序列互相影响关系显著。

上证综指和深证成指十五分钟收益率序列的Granger因果关系检验结果与五分钟数据存在着一定的差异。在1%显著性水平和滞后1-80阶的情况下, 原假设H均被拒绝, 认为深市收益率对沪市收益率有着显著的影响。而原假设H'在滞后2阶时, 在1%水平上被显著拒绝; 在滞后3、4、48、64和80阶时, 在5%水平上被显著拒绝; 在滞后1、8、16和32阶时, 在5%水平上接受了原假设H'。表明沪市收益率对深市收益率的影响关系没有一致的结果, 需要进一步分析。

表1: 沪深股市十五分钟收益率的Granger因果关系检验

量 原假设	F 统计									
	滞后阶数									
	1	2	3	4	8	16	32	48	64	80
H	16.7**	10.6**	8.1**	6.1**	3.6**	2.3**	1.9**	1.9**	1.6**	1.5**
H'	3.7	4.6**	3.7*	2.8*	1.8	1.6	1.4	1.5*	1.4*	1.3*

(H: RSHEN does not Granger Cause RHU, H': RHU does not Granger Cause RSHEN; \*和\*\*分别表示在1%和 5%的水平上显著)

### 3.5 二元GARCH(1,1)模型估计结果

使用极大似然估计方法估计上证综指和深证成指五分钟数据, 结果如下:

$$RHU_t = -0.006 + \varepsilon_{RHU,t}$$

$$RSHEN_t = -0.004 + \varepsilon_{RSHEN,t}$$

$$\sigma_{RHU,t}^2 = 0.0005 + 0.02\varepsilon_{RHU,t-1}^2 + 0.965\sigma_{RHU,t-1}^2$$

$$\sigma_{RSHEN,t}^2 = 0.006 + 0.017\varepsilon_{RSHEN,t-1}^2 + 0.973\sigma_{RSHEN,t-1}^2$$

$$\sigma_{RHU,RSHEN,t} = 0.0004 + 0.019\varepsilon_{RHU,t-1}\varepsilon_{RSHEN,t-1} + 0.969\sigma_{RHU,RSHEN,t-1}$$

采用相同的方法, 可以得到上证综指和深证成指十五分钟数据的二元GARCH(1,1)模型:

$$RHU_t = -0.00001 + \varepsilon_{RHU,t}$$

$$RSHEN_t = -0.00001 + \varepsilon_{RSHEN,t}$$

$$\sigma_{RHU,t}^2 = 0.0000002 + 0.025\varepsilon_{RHU,t-1}^2 + 0.959\sigma_{RHU,t-1}^2$$

$$\sigma_{RSHEN,t}^2 = 0.024\varepsilon_{RSHEN,t-1}^2 + 0.967\sigma_{RSHEN,t-1}^2$$

$$\sigma_{RHU,RSHEN,t} = 0.024\varepsilon_{RHU,t-1}\varepsilon_{RSHEN,t-1} + 0.963\sigma_{RHU,RSHEN,t-1}$$

以上两个模型的参数估计量均在1%水平上显著（统计量略）。由沪深股市收益率的二元GARCH(1,1)模型的估计结果我们可以得到：对于五分钟和十五分钟数据均有  $\hat{\beta}_2 > \hat{\beta}_1$  和  $\hat{\alpha}_2 + \hat{\beta}_2 > \hat{\alpha}_1 + \hat{\beta}_1$ 。这说明：对于收益率波动的自相关特性和持续性来讲，深圳股票市场要略大于上海股票市场，也即对于同样的一个波动率冲击（例如：新出台的宏观经济、市场管理政策，国际商品、股票市场的波动等），深圳股票市场吸收、反应信息的周期即波动持续期要长于上海股票市场。

由估计得到的条件方差序列和条件协方差序列可以计算条件相关系数。图5和图6分别给出了上证综指和深证成指五分钟和十五分钟收益率序列的条件相关系数序列。

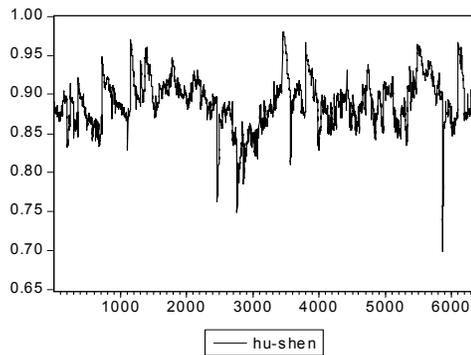


图5 沪深股市五分钟数据的相关性

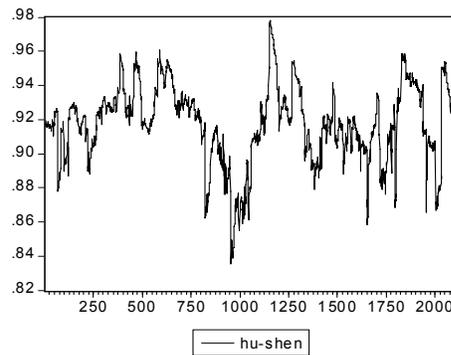


图6 沪深股市十五分钟数据的相关性

我们在图7中给出了沪、深股市五分钟成交量（HUVOLUME、SHENVOLUME）与五分钟收益率的条件方差（VAR\_HU、VAR\_SHEN）的图形。成交量是股票市场重要的信息指标之一，Granger因果关系检验结果表明成交量与收益率的条件方差之间存在着显著的相互影响关系，但是沪、深股市五分钟成交量与五分钟收益率的条件方差的无条件相关系数分别是0.43和0.46，这说明成交量对股票市场收益率条件波动的解释能力有限。

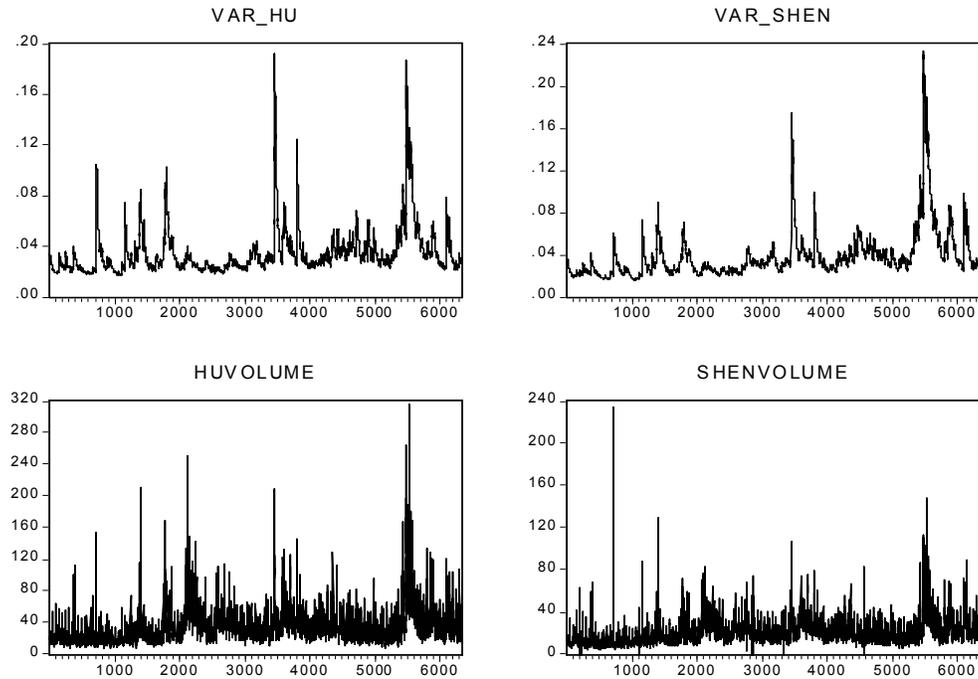


图7 沪深股市五分钟成交量与五分钟收益率的条件方差（成交量单位：万手）

#### 4 结 论

我们以上证综指和深证成指五分钟和十五分钟数据为样本，建立二元GARCH(1,1)模型，对上海和深圳两个股票市场股指收益率之间的相关性进行了量化估计。值得注意的是，这里的相关性反映的是沪深股市高频数据的短期波动之间的相关程度。估计得到的五分钟和十五分钟数据的条件相关性序列的均值分别是0.89和0.91，考虑到样本数据的高频特征，说明我国上海和深圳股票市场的短期波动之间存在着较强的正相关性，且持续性也较强。从而，从高频角度说明了两个市场存在（强）正相关性，也说明了两个市场对共同信息的反应速度及收益率变化程度是相近的。原因是：两市受到国家经济政策、市场管理制度等因素的共同影响，两市的上市公司种类和结构也类似，投资资金在两市之间流动性较好，随着两个股票市场的不断发展和完善以及投资者的逐步成熟，更重要的是收益率的短期波动在两个股票市场之间相互迅速且充分传递，使得两股票市场收益率的波动正相关程度很大。

对于同样的一个信息冲击，深圳股票市场消化信息的时间要略长于上海股票市场，也即深圳股票市场的波动持续期要大于上海股票市场。成交量是股票市场重要的信息指标之一，通常将其视为信息到达时刻的代理变量，将成交量作为股票市场条件波动的重要解释变量。在本文中，Granger因果关系检验结果表明五分钟成交量与收益率的条件方差之间存在着显著的相互影响关系，但二元GARCH模型估计结果表明沪、深股市五分钟成交量与五分钟收益率的条件方差的无条件相关系数分别是0.43和0.46，这说明成交量对股票市场收益率条件波动的解释能力有限。

#### 参考文献

- [1] 常宁，徐国祥。金融高频数据分析的现状与问题研究[J]。财经研究，2004，(3)：31-39。

- [2] 房振明, 王春峰. 上海股票市场收益日内效应的研究[J]. 北京理工大学学报(社会科学版), 2004, (4): 38-41.
- [3] 房振明, 王春峰, 蒋祥林. 中国股市回报波动性分析——高频数据揭示股市的特征[J]. 系统工程, 2004(2): 13-19.
- [4] 郭兴义, 杜本峰, 何龙灿. (超)高频数据分析与建模郭兴义[J]. 统计研究, 2002, (11): 28-31.
- [5] 黄后川, 陈浪南. 中国股票市场波动率的高频估计与特性分析[J]. 经济研究, 2003, (2): 75-82.
- [6] 屈文洲, 吴世农. 中国股票市场微观结构的特征分析——买卖报价价差模式及影响因素的实证研究[J]. 经济研究, 2002, (1): 56-63.
- [7] 陶利斌, 方兆本, 潘婉彬. 中国股市高频数据中的周期性和长记忆性[J]. 系统工程理论与实践, 2004, (6): 26-32.
- [8] 肖辉, 吴冲锋. 股指与股指期货日内互动关系研究[J]. 系统工程理论与实践, 2004, (5): 15-21.
- [9] 周建, 潘慧峰. 金融计量学的理论体系及其新进展[J]. 哈尔滨工业大学学报(社会科学版), 2004, (3): 93-99.
- [10] Andersen TG. Some reflections on analysis of high frequency data[J]. Journal of Business & Economic Statistic, 2000, 4, 146-153.
- [11] Andersen TG, Bollerslev T.. Intra day periodicity and volatility persistence in financial markets [J]. Journal of Empirical Finance, 1997, 4(2-3): 115-158.
- [12] Bollerslev Tim.. Generalized Autoregressive Conditional Heteroscedasticity[J]. Journal of Econometrics, 1986, 31: 307-327.
- [13] Bollerslev Tim, Zhang Benjamin Y.B.. Measuring and modeling systematic risk in factor pricing models using high-frequency data[J]. Journal of Empirical Finance, 2003, (10), 533-558.
- [14] Charles A. E., Goodhart, Maureen O' Hara. High frequency data in financial markets: Issues and applications[J]. Journal of Empirical Finance, 1997, 4, 73-114.
- [15] Engle R. F.. Autoregressive Conditional Heteroscedasticity With Estimate of the Variance of United Kingdom Inflation[J]. Econometrica, 1982, (50): 987-1008.
- [16] Engle R. F.. The econometrics of ultra - high frequency data[J]. Econometrica, 2000, 68(1).
- [17] Engle R.F., and K.Kroner (1993). Multivariate Simultaneous Generalized ARCH Econometric Theory, 11, 122-150.
- [18] Engle R. F. and C.W.J. Granger. 1987. Co-Integration and Error Correction: Representation, Estimation and Testing[J]. Econometrica, 55: 251-76.
- [19] Goodhart, C.A.E, Hall, S.G., Henry, S.G.B and Pesaran, B. News effects in a high frequency model of the sterling-dollar exchange rate. LSE Financial Markets Group Discussion Paper NO.119, 1991.
- [20] Harris, A. Statistical properties of the roll serial covariance Bid/ Ask spread estimator [J]. Journal of

Finance , 1990 , 45 (2) .

[21]Hasbrouck, J. The dynamics of discrete bid and ask quotes[J] . Journal of Empirical Finance , 1999, 54, 2109-2142.

[22] Jobson, J. D. and Korkie, B. M.. Estimation for Markowitz Efficient Portfolios[J]. Journal of the American Statistical Association, 1980( 75), 544-554.

[23] Martens M, Chang YC, Taylor SJ. A comparison of seasonal adjustment methods when forecasting intraday volatility [J]. Journal of Financial Research,2002,25(2):283-299.

[24] Nelson ,c.. Conditional Heteroskedasticity in Asset Return: A New Approach[J].Econometrica,59,347-370.

[25] Nicholas G . Polson, Bernard V . Tew . Bayseian portfolio selection: an empirical analysis of the S&P500 index 1970~1996. [J] . Journal of Business & Economic Statistics, 2000,4, 164~176.

[26] Robert. Wood. Market microstructure research databases: history and projections [J]. Journal of Business & Economic Statistics, 2000, 4, 140~145.

[27] Terence C.Mills. The Econometric Modelling of Finacial Time Series[M]. Original english Language Edition Published By Cambridge University Preess,1999.

[28] Zhang, M. Y., Russell, J. R., and Tsay, R. S. Determinants of bid and ask quotes and implications for the cost of trading, Working paper, Graduate School of Business, University of Chicago, 2001.

## **Study of Time-varying Volatility and Correlation between Shanghai and Shenzhen Stock Market ——A test based on five-minute and fifteen-minute's data**

CHEN Shou-dong, HAN Guang-zhe

(Business School of JiLin University, JiLin ChangChun,130012)

**Abstract:** In this paper, we use high-frequency data as our sample and study the time-varying volatility and correlation between Shanghai and Shenzhen Stock market. With the intraday five-minute and fifteen-minute's data, we build Bivariate GARCH(1,1) model to simulate and compute return series' variance, covariance and correlative coefficient of that are changing with time. The results show that the conditional volatilities between the Stock markets have strongly positive correlation.

**Key words:** Bivariate GARCH; High-frequency Data; Time-varying

**收稿日期:** 2005-08-10

**基金项目:** 04年教育部重大项目(05JJD790005)、05年国家社会科学基金项目(05BJY100)、05年国家自然科学基金项目(70573040)、01年国家自然科学基金项目(70173043)、02年教育部重大项目(02JAZJD790007)、“吉林大学‘985工程’项目”。

**作者简介:** 陈守东: 男, 1955年生, 汉族, 天津市蓟县人。吉林大学数量经济研究中心、商学院财务系主任、教授, 博士生导师, 博士。研究方向: 数量经济学。韩广哲: 男, 1978年生, 朝鲜族, 吉林省磐石市人。吉林大学商学院, 数量经济学专业博士研究生。