

基于高频数据的套利研究

——对中国股市弱式有效的一个检验

仪垂林 王家琪

(南京财经大学金融学院)

(一) 引言

证券市场效率，一直是证券市场发展中的核心问题，也是金融研究的重要课题。在金融经济理论中，关于证券市场效率问题最有影响力的理论应首推 Fama 在 1970 年提出的“效率市场假说”(EMH)，该假说认为，若证券市场在价格形成中充分而准确及时地反映全部相关信息，则称该市场为有效率的。在有效市场中，过去的信息对预测股票将来的收益是无用的，因为通过许许多多的竞争的、利润最大化的投资者的努力，所有过去信息都已反映到股价上。根据 EMH，股票市场只对新信息有反映，但是新信息是不可预测的，因此股价的变动或股票的收益也是不可预测的。

国内有很多研究曾对上海股票市场的有效性进行了检验(俞乔, 1994; 宋颂兴, 金伟根, 1995; 沈艺峰, 吴世农, 1999; 陈小悦, 1997; 胡朝霞, 1998; 汤光华, 2000; 胡畏, 范龙振, 2000; 张亦春, 周颖刚, 2001; 李琦, 郭菊娥, 2003; 等等)。但至今为止并没有针对股票日内交易的高频数据的研究, 这篇文章通过研究上海股票市场有关指数的日内交易价格模式来检验 EMH。本文的结构如下: 第一部分是有关理论的回顾、方法以及实证研究的情况。第二部分是这篇文章采用的模型及数据的描述。第三部分是实证研究的结果。最后是小结及有关解释。

(二) 有关理论及国内外研究文献综述

根据 EMH, 当现在的价格充分反映价格历史序列数据中所包含的一切信息, 即

由过去股价构成的信息集, 投资者不可能通过股价的历史变动来预测未来股价的变动, 此时的市场为弱式有效。即

$$E[P_t - (P_{t-1}^*/I_{t-1})] = 0 \text{ 或者 } E(r_t/I_{t-1}) = 0 \quad (1)$$

这里 I_{t-1} 是时期 $t-1$ 的可得的信息集, P_t 是时期 t 的股票的真实价格, P_t^* 是基于信息集 I_{t-1} 的预测价格, $P_t - P_t^*$ 是预测误差, 其与信息集 I_{t-1} 中的变量是无关的。同样股票收益 r_t 与信息集 I_{t-1} 中的变量也是无关的。

Fama (1970) 又提出了效率市场的如下定义, 这个定义使效率市场假设成为一个联合假设:

$$Z_t = r_t - E(r_t/I_{t-1}) \quad (2)$$

$$\text{和} \quad E(z_t) = E[r_t - E(r_t/I_{t-1})] = 0 \quad (3)$$

这里, 经济项 z_t 是时期 t 基于信息集 I_{t-1} 的超过均衡期望收益的收益, 加上另外的关于均衡收益是不随时间变化的常数的假设, 因此收益自身与过去信息集里的变量是不相关的。

除了上述模型, Fama 还提出了对检验 EMH 有重要应用的另外的模型, 该模型认为如果证券下一个时期的价格是基于信息集的映射, 那么证券的价格经常比现在的价格要高, 因此在这种情形下, 期望收益经常是正的。即

$$E(P_{t+1}/I_t) > P_t \text{ 或者 } E(r_{t+1}/I_t) > 0 \quad (4)$$

这个模型的重要的实际应用在于收益非负的假设, 意味着任何基于信息集的专门的交易策略产生的收益都不会比单纯的“买入并持有”(Buy and hold)策略高。

在实证研究中, 通常区分三种市场来进行研究。弱式有效市场是指股价已经反映了全部能从市场交易数据中得到的信息, 这些信息包括过去的股价、交易量、空头的利益等。该假定以为市场的价格趋势分析是徒劳的, 过去的股价资料是公开的且几乎毫不费力就可以获得。半强式有效市场是指与公司前景有关的全部公开的已知信息一定已经在股价中反映出来了, 此外, 如果任何一个投资者能从公开已知资源获取这些信息, 可以认为它会被反映在股价中。强式有效市场中是指股价反映了全部与公司有关的信息, 甚至包括尚未公开的或者原本属于保密的内部信息。大多数检验市场效率的实证研究是检验通过已知信息是否能获得超常收益从而来判断市场效率的 (Muth, 1961; Cootner, 1962; Fama, 1965; Gowland 和 Baker, 1970; Cutler 等人, 1989; MacDonald 和 Taylor, 1988, 1989; Spiro, 1990; Cochrane,

1991; Frennberg 和 Hansson, 1993; Jung 和 Boyd, 1996; Al-Loughani 和 Chappel, 1997; Niarchos 和 Alexakis, 2003; 等人)。这篇文章用同样的方式检验弱式有效市场, 主要是检验利用股市日内交易数据获利的可能性, 如果日内股票收益之间没有任何关系, 那么弱式有效市场便可以得到证实。

在最近的十几年里, 由于基于计算机交易的结果, 使得日内高频数据的提供成为可能, 因此推动了高频数据交易现象的研究。这样可以观察投资者在市场中的交易行为。一些高频数据的研究主要集中于日内股票收益是否依赖于时间的变化。经常用的计量经济学模型如下:

$$R_t = \alpha_1 D_1 + \alpha_2 D_2 + \alpha_3 D_3 + \cdots + \alpha_n D_n \quad (5)$$

这里 R_t 是一系列的股票的真实收益, D_1, D_2, \cdots, D_n 是虚拟变量, 分别与一天内的股票交易时刻 1, 2, \cdots, n 有关。按效率市场理论, 那么下式应成立。

$$\alpha_1 = \alpha_2 = \cdots = \alpha_n = 0$$

即股票收益应该与时间相互独立。

有关股票市场异象研究的文献认为日内股票收益显示出系统的特征。一些早期的研究文献认为日内平均收益显示出明显的模式——在开盘和收盘时往往收益比较高! Wood 等人 (1985) 利用 NYSE 的数据发现了与 *EMH* 相反的预测结果。Harris (1986) 利用 NYSE 的 1616 种股票构成的组合, 收集了 287 个交易日的数据, 发现在开盘的 45 分钟 (星期一除外) 和收盘的 15 分钟的收益明显为正, 且统计显著。另外 Terry (1986), Jain 和 Joh (1988) 利用每小时的数据也发现了类似的规律。许多研究表明“日末效应”(End of the day anomaly) 在几个国际股票市场是存在的。例如 Aitken 等人 (1995) 就证明了澳大利亚股票市场也存在这个现象。

既然从总体来看, 股票每天交易价格在收盘时通常比先前时段要高, 那么以收盘价格来代表股票价格有时就会引起很多问题。正如 Harris (1989) 指出的那样, 在所有的日交易价格中, 收盘价格是最重要的一个, 他们经常用来计算基金的净资产、某些交易合约的成本、研究人员计算收益的依据等等。之所以收盘价格被广泛应用, 一是因为他们很方便地得到, 另外假定他们可以代表股票的价值。

事实上, 不仅股票收益在日内呈现某种系统性的模式变化, 交易量也有类似的特征。Foster 和 Viswanathan (1993) 以及 Jain 和 Joh (1988) 发现交易量的系统变化, 并指出交易量在每天的交易中呈现 U 形分布的特点, 开盘时往往成交量最高, 期后减少, 在收盘时又重新增加。

许多解释股票日内交易价格模式的研究是基于股票市场的微观结构特点。在标准价值决定理论中暗含的市场结构对证券价格没有影响的假设现在受到了挑战。日益增长的有关市场微观结构的研究文献主要关注市场组织结构对价格形成过程的可能影响。这个过程的主要元素包括信息的产生和散布, 股票市场交易命令的传导以及交易规则等, 这些都会决定交易命令如何转换到交易中去。

另外市场涨跌幅限制以及竞价制度也经常是被微观结构理论解释日末收益较高的原因。Miller (1989) 指出某些做空者希望在交易日结束时结束他们的头寸并且努力在结算时取得一个正的收益, 因此会导致收盘价格较高。Amihud 和 Mendelson (1991) 还指出了交易命令的报价方式对收盘价格的影响, 他们区分了市价委托报单和限价委托报单的三种重要区别, 认为在最后的交易时刻, 投资者通常用市价委托而不是限价委托进行交易, 这样会导致收盘价格较高。

国内目前尚无有关股票日内交易价格模式变化的研究文献。已有的文献主要集中于对效率市场的检验上, 没有利用高频数据进行研究。主要的研究如下: 1994 年俞乔通过对深、沪两市 1990 年 12 月 29 日至 1994 年 4 月 28 日综合指数的变动趋势通过随机游走模型进行分析, 得到的实证结果表明, 深、沪市场均不具有弱式有效。1995 年宋颂兴和金伟根认为由于上海股市的发展存在明显的阶段性, 因此检验时将中国股市分为两个阶段, 对沪市 1993 年初至 1994 年 10 月 29 种股票的周收益率进行分析后结果表明, 29 种股票价格具有随机游动的特征, 据此认为上海股市达到弱式有效市场。2000 年, 汤光华对沪市 1992 年 6 月 1 日至 1999 年 3 月 31 日的日收益率时间序列进行序列相关检验和流动检验。其检验结果表明, 基本肯定上海股市具有弱式有效, 并且认为近年信息有效程度在不断地提高。从上述对目前中国股市有效性进行的实证检验结果来看, 中国股市在是否达到弱式有效市场还存在着分歧, 但 1993 年后大多数学者的实证研究都证明中国股市正趋于“近似弱式有效”, 但离强式有效和半强式有效还相去甚远。

(三) 数据及模型

本文利用指数的日内交易数据, 采集了上证综合指数、综合类指数、制造业指数、批发零售业指数、建筑业指数。以每 30 分钟为一时间间隔, 这样从开盘 (9:30) 到收盘 (15:00) 便有 9 个观测值。之所以选间隔 30 分钟的数据是考虑到衡量股票价格变动的敏感性以及市场对新信息的反映程度, 过短和过长的间隔都不是很合适。时间段

自 2004 年 9 月 1 日到 2004 年 12 月 31 日，数据来源于天相投资公司。

我们对指数价格 P 作对数变换，记为 $\log P$ ，收益用下式计算

$$R_t = \log P_t - \log P_{t-30}$$

检验日内收益是否独立于时间的模型如下：（字母含义同前）

$$R_t = \alpha_1 D_1 + \alpha_2 D_2 + \alpha_3 D_3 + \cdots + \alpha_9 D_9 \quad (6)$$

另外考虑到收益的序列相关及市场对信息的非同步性等特点，需要其他解释变量如滞后项对模型进行修正，那式（6）可以调整如下：

$$R_t = \alpha_1 D_1 + \alpha_2 D_2 + \alpha_3 D_3 + \cdots + \alpha_9 D_9 + \sum_{i=1}^k \beta_i R_i \quad (7)$$

R_i 为滞后的收益项。

在 EMH 的框架下，下式应成立。

$$\alpha_1 = \alpha_2 = \cdots = \alpha_9 = 0 \quad (8)$$

（四）实证研究

表 1 列出了序列的基本统计特征，显示五个指数都不符合正态分布的特点。

表 1 指数收益统计量特征描述

指 数	平均收益	标准差	斜度	峰度
上证指数	-0.0000336	0.001999	0.076581	6.460288
综合类指数	-0.0000140	0.002257	-0.154174	6.623682
制造业指数	-0.0000397	0.002101	-0.266611	7.569290
批发零售业指数	-0.0000192	0.002015	-0.377564	7.798475
建筑业指数	-0.00000493	0.002226	-0.114105	5.464670

指数回报率序列是否平稳，是我们采用何种回归方法的基础。为此，我们采用单位根（ADF）来检验 R_t 的平稳性，检验结果如表 2，显示所有的序列是平稳序列。

表 2 指数收益单位根检验 (ADF)

指 数	统计量 (99% 置信水平)	ADF 检验统计量
上证指数	-3.438877	-27.74482
综合类指数	-3.438889	-27.74482
制造业指数	-3.438889	-27.42326
批发零售业指数	-3.438889	-27.36864
建筑业指数	-3.438889	-27.11926

另外对序列相关及异方差的检验表明，所有序列均具有 ARCH 效应的特点。因此本文回归的方法采用 GARCH 模型。

回归的结果如下（表 3 到表 7）：

表 3 上证指数收益回归结果

	Coefficient	Std. Error	z-Statistic	Prob.
D1	-9.74E-05	0.000191	-0.509815	0.6102
D2	-0.000650	0.000139	-4.666444	0.0000
D3	-0.000228	0.000184	-1.241980	0.2142
D5	0.000168	0.000178	0.948805	0.3427
D4	-8.91E-05	0.000177	-0.502099	0.6156
D6	-6.60E-05	0.000240	-0.275573	0.7829
D7	4.69E-05	0.000182	0.257553	0.7968
D8	-0.000322	0.000165	-1.953563	0.0508
D9	0.000641	0.000173	3.696375	0.0002
Variance Equation				
C	1.79E-07	6.18E-08	2.899986	0.0037
ARCH(1)	0.171637	0.034353	4.996229	0.0000
GARCH(1)	0.795850	0.036889	21.57438	0.0000

$R^2 = 0.02 \quad Q_1 = 0.1246 \quad Q_{36} = 34.825$

$Q_{(sq)1} = 0.8614 \quad Q_{(sq)36} = 19.215$

表 4 综合类指数收益回归结果

	Coefficient	Std. Error	z-Statistic	Prob.
D1	7.65E-05	0.000232	0.330115	0.7413
D2	-0.000535	0.000166	-3.232220	0.0012
D3	-0.000278	0.000219	-1.269602	0.2042
D4	-0.000248	0.000210	-1.180219	0.2379
D5	0.000170	0.000209	0.817343	0.4137
D6	5.56E-05	0.000298	0.186528	0.8520
D7	8.03E-05	0.000227	0.353059	0.7240
D8	-0.000207	0.000200	-1.035116	0.3006
D9	0.000632	0.000196	3.220070	0.0013
Variance Equation				
C	2.18E-07	7.11E-08	3.066006	0.0022
ARCH(1)	0.133710	0.031775	4.207980	0.0000
GARCH(1)	0.832067	0.035783	23.25336	0.0000

$$R^2 = 0.012 \quad Q_1 = 0.7760 \quad Q_{36} = 53.396$$

$$Q_{(sq)1} = 0.1746 \quad Q_{(sq)36} = 16.740$$

表 5 制造业指数收益回归结果

	Coefficient	Std. Error	z-Statistic	Prob.
D1	-0.000116	0.000229	-0.506005	0.6129
D2	-0.000725	0.000144	-5.054506	0.0000
D3	-0.000223	0.000188	-1.189486	0.2342
D4	-0.000134	0.000183	-0.731102	0.4647
D5	0.000177	0.000186	0.954721	0.3397
D6	7.45E-06	0.000242	0.030769	0.9755
D7	5.03E-05	0.000192	0.261759	0.7935
D8	-0.000200	0.000176	-1.132611	0.2574
D9	0.000650	0.000167	3.893987	0.0001
Variance Equation				
C	1.69E-07	5.94E-08	2.841879	0.0045
ARCH(1)	0.164483	0.034377	4.784674	0.0000
GARCH(1)	0.809875	0.037459	21.62028	0.0000

$$R^2 = 0.015 \quad Q_1 = 0.8959 \quad Q_{36} = 42.973$$

$$Q_{(sq)1} = 0.6985 \quad Q_{(sq)36} = 17.468$$

表 6 批发零售业指数收益回归结果

	Coefficient	Std. Error	z - Statistic	Prob.
D1	-0.000102	0.000225	-0.451358	0.6517
D2	-0.000709	0.000146	-4.848244	0.0000
D3	-0.000242	0.000198	-1.220396	0.2223
D4	-0.000150	0.000203	-0.737566	0.4608
D5	0.000233	0.000187	1.246463	0.2126
D6	8.67E -05	0.000234	0.370322	0.7111
D7	0.000134	0.000189	0.711108	0.4770
D8	-0.000215	0.000170	-1.267888	0.2048
D9	0.000633	0.000151	4.188336	0.0000
Variance Equation				
C	2.56E -07	7.79E -08	3.292500	0.0010
ARCH(1)	0.176401	0.035213	5.009491	0.0000
GARCH(1)	0.775565	0.041987	18.47170	0.0000

$$R^2 = 0.018 \quad Q_1 = 1.1451 \quad Q_{36} = 41.334$$

$$Q_{(sq)1} = 0.1657 \quad Q_{(sq)36} = 11.046$$

表 7 建筑业指数收益回归结果

	Coefficient	Std. Error	z - Statistic	Prob.
D1	2.93E -05	0.000263	0.111408	0.9113
D2	-0.000780	0.000151	-5.162653	0.0000
D3	-0.000164	0.000208	-0.788100	0.4306
D4	-0.000114	0.000215	-0.531312	0.5952
D5	-1.22E -05	0.000230	-0.052949	0.9578
D6	6.91E -05	0.000276	0.250911	0.8019
D7	7.88E -05	0.000235	0.335197	0.7375
D8	-0.000192	0.000214	-0.896841	0.3698
D9	0.000311	0.000203	1.531277	0.1257
Variance Equation				
C	1.32E -07	4.88E -08	2.709684	0.0067
ARCH(1)	0.079429	0.017513	4.535499	0.0000
GARCH(1)	0.894764	0.023581	37.94363	0.0000

$$R^2 = 0.018 \quad Q_1 = 0.8946 \quad Q_{36} = 50.833$$

$$Q_{(sq)1} = 3.0354 \quad Q_{(sq)36} = 32.557$$

另外，我们对每一时间间隔的收益进行了汇总统计，见图1。

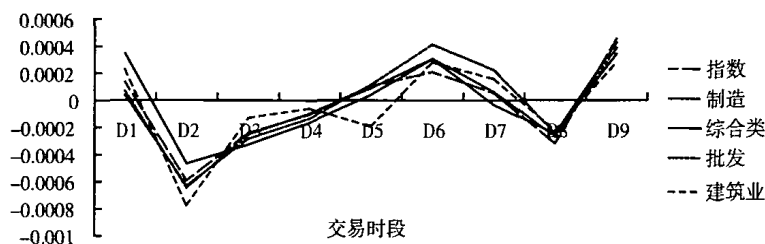


图1 日内指数收益特点

从图1可以看出，所有指数日内收益也呈现U形分布的特点。从表3到表7可以发现，从9:30到10:00收益明显是负的，且统计显著。而且除建筑业之外的所有指数收益呈现明显的“日末效应”，即收盘收益明显较高，统计上也是非常显著的。作为模型解释力的 R^2 从0.01到0.02变化，对残差的 Q 及 Q_{sq} 的检验显示模型的应用没有计量上的问题。

基于上述结果，我们可以设计如下的交易策略：在整个交易期间的每天上午的10:00买入股票，在收盘时（15:00）卖出股票，以这样的投资策略和“买入并持有”的投资策略进行比较，结果见表8。通过表8可以看出，所有的情形都表明交易策略比被动的“买入并持有”策略产生的收益要高，而且尤其要强调的是以标准差表度的风险明显都变小。虽然基于交易策略的综合类指数的收益为负，但依然比“买入并持有”的策略要好的多。也就是说基于交易策略的交易不但收益更高而且风险更小。这样看来，中国股票市场并未达到“弱式有效”。

表8 买入并持有策略（Buy and hold）与既定交易策略（Trading rule）收益比较

	平均收益	标准差
上证指数		
Buy and hold	-0.000285	0.006212
Trading rule	0.000158	0.005267
综合类指数		
Buy and hold	-0.000126	0.007707
Trading rule	-0.0000172	0.006584

续表

	平均收益	标准差
制造业指数		
Buy and hold	-0.000350	0.006850
Trading rule	0.000223	0.005819
批发零售业指数		
Buy and hold	-0.000167	0.006693
Trading rule	0.000462	0.005715
建筑业指数		
Buy and hold	-0.000435	0.007961
Trading rule	0.000107	0.006608

(五) 结 论

这篇文章的计量检验表明中国股票市场存在着系统性的日内交易价格模式。在整个交易期间股票收益遵循着近似 U 形分布的特点, 特别需要说明的是, 对所有的指数而言, 收益在 10:00 时统计上显著为负, 在收盘时显著为正。因此基于这些研究结果的交易策略可以产生比单纯的“买入并持有”策略产生更高的收益, 同时风险还大为降低。

对上述规律进行解释并不容易。大量的研究都证明在不同的市场临近收盘时的收益一般都比较低, 但是这种效应并不是很容易地能被微观市场组织理论所解释, 因为不同国家的市场结构并不相同。一个可能的解释是交易的末期往往有利好消息的发布或者一些投资者的交易习惯使得收盘价格较高。然而, 这个解释是很难在实证中被检验的。

另外, 解释这个现象的一个方向是基于有关市场操纵的研究文献。股票市场可以被操纵这一点在理论界没有争议。股票价格操纵有许多形式, 并且这也是监管当局进行监管的主要理由。以中国股票市场为例, 自 1996 年以来实行了 10% 的涨跌幅限制, 而涨跌幅计算的依据便是依据上一个交易日的收盘价。另外因为经济新闻、电视、广播一般报道收盘价格, 所以收盘价格通常被认为是股票表现的衡量依据。大多数技术分析人士通常利用收盘价格进行分析和预测, 共同基金计算净值也往往利用收盘价格。既然收盘价格被广泛应用为股票市场的指示器, 因此市场被操纵就会产生显著的利益。例如, 一部分投资者可以通过持续的买入来操纵收盘价格在投

投资者中间制造乐观情绪从而导致更多的买盘。这个问题容易引起争议的是投资者同样可以通过卖出的办法制造悲观情绪导致更多的卖盘而使收盘价格走低。这种形式的操纵可以使得收盘收益变低甚至为负。但是,我们知道中国股票市场是不允许卖空的,因此没有动机使投资者操纵股票价格走低。

至于所有的指数显示 10:00 收益明显为负的事实,这是个令人困惑的现象。一个可能的解释是投资者的交易习惯。例如尽管卖空被法律禁止,但是某些投资者依然可以通过某些渠道在交易开始时取得空头头寸,这样便使价格走低,为了平仓,投资者会在结束时买进从而导致收盘价格升高,因为根据结算规则及程序,他们不能持有空头头寸超过一个交易日。当然利用更多的比如买卖命令的类型、成交量等其他的日内交易信息可能会更好地解释这种日内价格模式。另外一个可能地解释是从投资者情绪的方向展开,关于这个方向的研究目前还处于探索阶段。

参考文献

- [1] 陈小悦、陈晓、顾斌:中国股市弱式效率的实证研究 [J],《会计研究》,1997.9。
- [2] 戴国强、吴林祥:《金融市场微观结构理论》[M],上海财经大学出版社,1999。
- [3] 范龙振、张子刚:深圳股票市场的弱有效性 [J],《管理工程学报》,1998.2。
- [4] 高鸿桢:关于上海股市效率性的探讨 [J],《厦门大学学报》,1996.4。
- [5] 沈艺峰、吴世农:中国证券市场过度反应了吗? [J],《经济研究》,1999.2。
- [6] 宋颂兴、金伟根:上海股市市场有效性实证研究 [J],《经济学家》,1995.4。
- [7] 俞乔:市场有效、周期异常与股价波动 [J],《经济研究》,1994.8。
- [8] 张人骥,朱平方,王怀方:上海证券市场过度反应的实证检验 [J],《经济研究》,1998.5。
- [9] 张亦春、周颖刚:中国股市弱式有效吗? [J],《金融研究》,2001.3。
- [10] Cooter, D. M., Poterba, J. M. and Summers, L. H. (1989): What moves stock prices, *Journal of Portfolio Management* 15, 4 ~ 12.
- [11] Cooter, D. M., Poterba, J. M. and Summers, L. H. (1991): Speculative dynamics. *Review of Economic Studies* 58, 529 ~ 41.
- [12] Fama, E. (1965): The behaviour of stock market prices, *Journal of Business* 38, 34 ~ 105.
- [13] Fama, E. (1970): Efficient capital markets: a review of the theory and empirical work, *Journal of Finance* 25, 383 ~ 416.
- [14] Fama, E. (1991): Efficient capital markets, *Journal of Finance* 1575 ~ 617.
- [15] Foster, F. D. and Viswanathan, S. (1993): Variations in trading volume, return volatility and trading costs: evidence on recent price formation models, *Journal of Finance* 48, 187 ~ 211.

- [16] Harris, L. (1986): A transaction data study of weekly and intraday patterns in stock returns, *Journal of Financial Economics* 16, 99 ~ 117.
- [17] Jain, P. C and Joh, G. H. (1988): The dependence between hourly prices and trading volume, *Journal of Financial and Quantitative Analysis* 23, 269 ~ 83.
- [18] Wood, R. A. , McInish, T. H. and Ord, J. K. (1985): An investigation of transactions data for NYSE stocks, *Journal of Finance* 40, 723 ~ 41.
- [19] A. Niarchos, and C. A. Alexakis (2003): Intraday stock price patterns in the Greek stock exchange, *Journal of Applied Financial Economics* 13, 13 ~ 22.

基于高频数据的套利研究——对中国股市弱式有效的一个检验

作者：[仪垂林](#)，[王家琪](#)
作者单位：[南京财经大学金融学院](#)

本文链接：http://d.g.wanfangdata.com.cn/Conference_6567882.aspx