

文章编号: 1000-5013(2012)03-0330-07

基于半参数 LM-ARMAX 模型的 股价波动成因分析

陈春春, 胡日东

(华侨大学 经济与金融学院, 福建 泉州 362021)

摘要: 选择沪市 1991—2010 年所有上市公司的数据, 建立 LM-ARMAX 模型来实证股票价格波动的决定因素; 然后, 根据模型半参数估计的结果, 进行基于半参数估计的广义似然比检验和基于 Wild Bootstrap 的 Smirnov 检验. 研究表明: 市帐率和成交量是股票价格波动的主要因素, 而净资产收益率对股票价格波动的影响不显著; 相比起“指数研究”和“样本替代研究”而言, 实证的数据精确度更高, 说服力更强.

关键词: 股票价格; LM-ARMAX 模型; 波动率; 半参数估计; 非线性

中图分类号: O 212.7; F 830.91

文献标志码: A

目前, 研究股票价格的波动主要集中在针对股价波动自身的建模, 对其影响因素的分析较少, 而对此进行实证分析就更加稀缺. 已有的资料大多是利用综合指数来研究整个市场^[1], 而基于综合指数计算的波动与真实的股票市场价格波动毕竟存在偏差, 由此而产生的结论自然也是不够精确的. 不少学者就股票价格波动随时间的变化做过深入的研究, 并形成了“净资产收益率假说”、“市帐率假说”和“成交量假说”等理论. “净资产收益率假说”的学者认为, 股票价格是公司未来盈利的当前价值, 因此股价的波动与公司盈利的能力呈正相关关系. Vuolteenaho^[2]提出了一个线性模型来实证这种关系, 并用净资产收益率指代公司的盈利能力. 但是, 它所依据的线性模型也许只能在短期有效, 很难反映较长期股价的波动状况^[3]. Pastor 等^[4]也认为股价的波动与净资产收益率的确存在着正相关关系, 但其数据来自公司层面的截面数据, 结论只能是针对不同部门的波动差异, 不具有时变性质. “市帐率假说”的学者认为, 公司成长可能性的当前价值也应该是股价波动的重要原因. Cao 等^[5]指出, 市帐率可以很好地刻画公司成长的可能性, 较高的市帐率也意味着较高的成长可能性, 相应地股价的波动也必然为之而发生改变, 而且这种影响作用在发展较不稳定的中小企业及高科技企业中会体现得尤为突出. 但这个假说还仅仅只是理论上的推演, 缺乏实证的分析. “成交量假说”的学者认为, 股价的波动可能受到股票成交量变化的影响^[6-7]. 本文遵循实证的思路, 通过选择适当的指标, 建立半参数的模型, 对上述理论进行检验.

1 LM-ARMAX(α, d, β)模型的估计和检验

1.1 LM-ARMAX(α, d, β)模型

时变分整自回归移动平均 (autoregressive fractional integrated moving average, 简称 ARFIMA) 模型是差分自回归移动平均 (autoregressive integrated moving average, 简称 ARIMA) 模型在分数阶协整领域的扩展, 在具有长记忆性的金融时间序列中得到了广泛的运用; 而多变量自回归滑动平均 (autoregressive moving average with exogenous variables, 简称 ARMAX) 模型又是传统时间序列的自回归滑动平均 (autoregressive moving average, 简称 ARMA) 模型在多元变量基础上的延伸, 在分析响应变量时具有单一变量模型所无法比拟的优势.

收稿日期: 2011-04-27

通信作者: 胡日东 (1964-), 男, 教授, 主要从事金融工程的研究. E-mail: j_rdh@hqu.edu.cn.

基金项目: 国家软科学计划项目 (2008GXS5D130); 教育部科学技术研究重点基金资助项目 (209148)

为了研究股票市场价格波动变化的影响因素,同时又兼顾到高频数据的长记忆性,结合两种模型的特点,提出了 LM(long memory)-ARMAX(α, d, β)模型,即

$$\varphi(L)(1-L)^d Y_t = \theta(L)\varepsilon_t + f(X_t). \quad (1)$$

式(1)中: Y_t 为被解释变量的观测值序列; X_t 为解释变量的观测值序列; L 为滞后算子; $(1-L)^d$ 为分数差分算子($0 < d < 0.5$); $\varphi(L) = 1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_a L^a$; $\theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_\beta L^\beta$; f 为需要估计的未知函数, $\{\varepsilon_t\}$ 为白噪声序列. 在对 ARMAX(α, d, β)进行估计之前,要对分数差分阶数 d 、自回归阶数 α 、移动平均阶数 β 进行识别.

1.2 LM-ARMAX(α, d, β)模型的识别

GPH 估计法最早由 Geweke, Porter 和 Hudak^[8] 提出的,后经 Robinson^[9] 进一步发展,并由 Andrews 等^[10] 改进的方法. 对于分数差分阶数 d 的识别,常用的方法是 GPH 估计法.

设 Z_t 为被检验序列,令 $W_t = (1-L)^d Z_t$, $f_w(\omega)$ 和 $f_z(\omega)$ 分别是 $\{W_t\}$, $\{Z_t\}$ 谱函数,则有

$$\left. \begin{aligned} f_w(\omega) &= (\sigma_a^2/2\pi) |\theta_q(\exp(-i\omega))/\phi_p(\exp(-i\omega))|^2, \\ f_z(\omega) &= |1 - \exp(-i\omega)|^{-2d} f_w(\omega). \end{aligned} \right\} \quad (2)$$

对 $f_z(\omega)$ 两边取对数,可得到

$$\ln f_z(\omega) = \ln f_w(0) - d \ln |1 - \exp(-i\omega)|^2 + \ln(f_w(\omega)/\ln f_w(0)). \quad (3)$$

令 $\omega_j = 2\pi j/n$, $j=1, \dots, m$ 替代 ω , 把 $\ln f_z(\omega_j) = 1/2\pi n \left| \sum_{t=1}^n Z_t \exp(-i\omega t) \right|^2$ 带入式(3),可得到

$$\ln f_z(\omega_j) = \ln f_w(0) - d \ln |1 - \exp(-i\omega)|^2 + \ln(f_w(\omega_j)/\ln f_w(0)) + \ln(I_z(\omega_j)/f_z(\omega_j)). \quad (4)$$

当 $j=1, \dots, m \ll [n/2]$ 时,满足当 $n \rightarrow \infty$ 时 $m/n \rightarrow 0$, 所以 $\ln(f_w(\omega_j)/\ln f_w(0))$ 可以忽略. 再令 $\varepsilon_j = \ln(I_z(\omega_j)/f_z(\omega_j))$ 为随机误差项,于是利用线性估计可得到

$$d = \left\{ \sum_{j=1}^m (X_j - \bar{X})(Y_j - \bar{Y}) \right\} / \left\{ \sum_{j=1}^m (X_j - \bar{X})^2 \right\}. \quad (5)$$

式(5)中: $Y_j = \ln(I_z(\omega_j))$; $X_j = \ln|1 - \exp(-i\omega_j)|^2 = \ln(2 - 2\cos \omega_j)$.

对于半参数模型的自回归阶数 α 和移动平均阶数 β 的识别,可以利用 Hurvich 等^[11] 针对半参数模型所提出的 AIC_C 准则,但是考虑到 AIC_C 形式较为繁琐,故采用 Simonoff^[12] 改进的 AIC_C 方法. 即

$$\text{AIC}_C = \log(\hat{\sigma}^2) + (1 + \text{tr}(H)/n) / \{1 - [\text{tr}(H) + 2]/n\}. \quad (6)$$

式(6)中: $\hat{\sigma}^2$ 为模型随机误差项的方差估计值; $H = (\mathbf{Q}'\mathbf{Q})^{-1}\mathbf{Q}'$, \mathbf{Q} 为对模型进行正交序列估计时所使用的正交矩阵.

一般来说,自回归阶数 α 和移动平均阶数 β 的阶数值不会超过 2,故将 $\alpha=0, 1, 2$ 和 $\beta=0, 1, 2$ 共 9 种组合的 AIC_C 利用式(6)计算出来,选择 AIC_C 最小的 α 和 β 的组合,即得到 LM-ARMAX 模型最优的 α 和 β 阶数.

1.3 LM-ARMAX(α, d, β)模型的估计

对 LM-ARMAX(α, d, β)模型进行半参数估计,有以下 6 个步骤.

1) 假设 $\varphi(L)$ 已知,令 $Z_t = (1-L)^d Y_t$, 基于参数方程

$$\varphi(L)Z_t = \theta(L)\varepsilon_t + f(X_t) + u_t, \quad (7)$$

选择正整数 ξ (如 $\xi=1$), 可得到 $f(\cdot)$ 的正交序列估计为

$$\hat{f}(\cdot, \varphi(L)) = (\varphi_1(\cdot), \varphi_2(\cdot), \dots, \varphi_\xi(\cdot))(\mathbf{Q}'\mathbf{Q})^{-1}\mathbf{Q}'(\varphi(L)Z_t). \quad (8)$$

式(8)中: $\mathbf{Q} = (\mathbf{Q}_1, \dots, \mathbf{Q}_\xi)$; $\mathbf{Q}_i = \varphi_i(X_t)$; $\varphi_i(\cdot) \in [a, b]$ 并满足 $\int_a^b \varphi_i(\cdot)\varphi_j(\cdot)d(\cdot) = \begin{cases} 0, & i \neq j, \\ c_i, & i = j. \end{cases}$

2) 令 $\varepsilon_t = \hat{u}_t$, 基于参数方程

$$\varphi(L)Z_t - \theta(L)\varepsilon_t = \hat{f}(X_t, \varphi(L)) + v_t,$$

选择正交序列估计,可得到 $\hat{\varphi}(L)$ 和 $\hat{\mathbf{Q}}(L)$, 带入求出 $\hat{f}(\cdot, \varphi(L), \theta(L))$ 的估计 $\hat{f}(\cdot, \hat{\varphi}(L), \hat{\theta}(L))$ 及 \hat{v}_t .

3) 在步骤 2) 中,令 $\varepsilon_t = \hat{v}_t$, 重复步骤 2) 的过程,迭代若干次,直到 $\hat{\varphi}(L)$ 和 $\hat{\theta}(L)$ 中的参数 φ_i 和 θ_i 在前后两次迭代中的改变量小于 0.001.

(4) 在步骤 1) 中改变 ξ , 取 $\xi=1, \dots, 100$; 重复步骤 2), 3) 的过程, 选择使得残差平方和最小的正整

数 ξ^* , 得到函数 $\hat{f}(\cdot, \hat{\varphi}(L), \hat{\theta}(L))^*$, 以及 $\hat{\varphi}(L)^*$ 和 $\hat{\theta}(L))^*$. 带入式(7)中, 可以得到 Z_i 的估计值 \hat{Z}_i , 进而算出被解释变量 Y_i 的估计值 \hat{Y}_i .

5) 保持解释变量 X_i 中其他观测值不变, 调整 X_1 的取值并令 $X'_1 = X_1 + 10^{-5}$, 利用步骤 1)~4), 计算出解释变量调整后的函数 $\hat{f}(\cdot, \hat{\varphi}(L), \hat{\theta}(L))^*$, 以及 $\hat{\varphi}(L)_1^*$ 和 $\hat{\theta}(L)_1^*$; 然后, 带入式(7)可以得到 \hat{Z}'_1 , 进而算出 \hat{Y}'_1 . 那么, $\partial Y_i / \partial X_i |_{X_i = X_i} \approx (\hat{Y}'_i - \hat{Y}_i) / 10^{-5}$.

6) 将解释变量 X_i 中的其他观测值 $X_2 \sim X_n$, 按步骤 5) 重复, 可估计出所有 $\partial Y_i / \partial X_i |_{X_i = X_i}$ 的近似值; 最后, 把 $\partial Y_i / \partial X_i |_{X_i = X_i}$ 用三次样条方法进行插值, 可得到较为光滑的 $\partial Y_i / \partial X_i$ 函数图形.

1.4 半参数估计的广义似然比检验

目前, 对非参数回归的研究主要集中在对非参数函数的估计问题上, 而对非参数回归模型的检验却不多. 其中, 有代表性的是 Hart^[13] 提出的非线性检验, 以及 Fan^[14] 改进的半参数广义似然比检验.

考虑半参数回归模型 $Y_i = Z'_i \gamma + m(X_i) + \epsilon_i, i = 1, 2, \dots, n$. 其中: $Z_i = (z_{i,1}, z_{i,2}, \dots, z_{i,p})'$, $X_i = (x_{i,1}, x_{i,2}, \dots, x_{i,p})'$ 为 p 维和 q 维解释变量; γ 为 p 维未知参数向量 Y_i 为响应变量; $m(\cdot)$ 为未知光滑函数. 检验问题 $H_0 : m(X) = X'\beta, H_1 : m(X) \neq X'\beta$, 而 β 为 q 维未知参数向量.

假定 $\epsilon_i \sim N(0, \sigma^2 I_n)$ 服从正态分布, 得到上述半参数模型的对数似然函数为

$$L(f, \beta, \sigma^2) = -(n/2)(\log 2\pi\sigma^2) - \sum_{i=1}^n (Y_i - Z'_i \gamma - m(X_i))^2 / 2\sigma^2. \tag{9}$$

对式(9)求偏导, 并令 $\partial L(f, \beta, \sigma^2) / \partial \sigma^2 = 0$, 则可得到 $\hat{\sigma}^2 = \text{RSS} / n$. 其中, RSS 为残差平方和. 所以, 一方面, 当 H_0 成立时, 对线性模型进行估计, 并将 $\hat{\beta}, \hat{\gamma}_1$ 和 $\hat{\sigma}_1^2$ 代入式(9)可得到

$$L(H_1) = -n/2 - (n/2)\log 2\pi/n - (n/2)\log \text{RSS}_0; \tag{10}$$

另一方面, 当 H_1 成立时, 利用半参数估计方法得出 $\hat{m}(\cdot), \hat{\gamma}_2$ 和 $\hat{\sigma}_2^2$, 带入式(9)可得到

$$L(H_1) = -n/2 - (n/2)\log(2\pi/n) - (n/2)\log \text{RSS}_1. \tag{11}$$

构造统计量 $T_{H_0} = L(H_1) - L(H_0) = n/2(\log(\text{RSS}_0/\text{RSS}_1))$. Fan 等^[14] 证明当一定条件成立时, T_{H_0} 渐近于自由度为 q 的 χ^2 分布.

在这里, 一定条件指 $E(\epsilon_i) = 0, E(\epsilon_i^2) = \sigma_i^2, E(\epsilon_i^4) < \infty, E(X'_i X_i) = \Sigma, E(X_{i,r}^4) < \infty$. 其中: $i = 1, \dots, n; r = 1, \dots, q; X_{i,r}$ 为 X_i 的第 r 个分量. 因此, 可得到显著性水平为 $1 - \alpha$ 拒绝域为

$$C_{H_0} = \{\hat{T}_{H_0} \geq \chi_q^2(1 - \alpha)\}.$$

1.5 基于 Wild Bootstrap 的 Smirnov 检验

Smirnov 检验是一种极为重要的非参数检验规则, 在社会统计学中得到了广泛的运用. 但是, 由于时间的不可重复性, 计量经济学领域的样本极为稀缺, 大多数统计学的检验思想都难以在计量领域实现, 而 Wild Bootstrap 方法为大样本检验提供了有力的帮助. 因此, 基于 Wild Bootstrap 模拟对 LM-ARMAX(α, d, β) 模型进行 Smirnov 检验, 具体有如下 4 个主要步骤.

1) 产生一组 Wild Bootstrap 残差 $\{\eta_q\}$. 其中: $\eta_q = \tilde{u}_q \cdot \lambda_q, \{\lambda_q\}$ 是一组独立同分布的随机变量, 服从标准正态分布; $\tilde{u}_q = u_q - \bar{u}_q$ 为基于 LM-ARMAX 模型计算的残差与其均值的差额.

2) 利用 LM-ARMAX(α, d, β) 模型的半参数估计值和 Wild Bootstrap 残差 $\{\eta_q\}$ 来构造 Wild Bootstrap 样本 $\{Y'_i\}$. 其中, $Y'_i = Y_i + \eta_q$.

3) 针对 $\{Y_i\} \{Y'_i\}$ 进行 Smirnov 检验.

4) 重复上述步骤 10 000 次, 累计 Smirnov 检验的结果为存在显著差异的频率 f . 当 $f < 5\%$ 时, 可以认为 LM-ARMAX(α, d, β) 模型设定正确; 反之, 则认为模型的设定不正确. Kreiss 等^[15] 研究表明, Bootstrap 类方法在模拟的样本容量超过 1 000 时才是有效的, 故文中为了获取更为精确地结果, 设定模拟的样本容量为 10 000.

2 实证分析

2.1 数据来源

所使用的数据全部来源于北京聚源锐思数据科技有限公司 RESSET 金融研究数据库和北京大学

中国经济研究中心 CCER 经济金融数据库. 被解释变量为股票价格, 其样本包含 1991 年 1 月 2 日至 2010 年 12 月 31 日共 4 895 个交易日数据(去除 1992 年 9 月 2 日;1992 年 5 月 20 日、21 日、25 日;1994 年 8 月 9 日;1995 年 5 月 18 日、19 日的异常数据);解释变量为净资产收益率、市帐率和成交量, 样本包含 1991 年一季度至 2010 年四季度共 80 个季度数据.

2.2 被解释变量的计量

以股票价格的波动作为被解释变量,是不能直接从证券市场上观测到的. 为了研究此问题,Boller-slev 等^[16]提出了“已实现”波动(RV)来计算波动,但此方法难以满足统计意义上的稳健性、有效性等要求. Nielsen 等^[17]提出了“已实现”双幂次变差波动(RBV),再考虑到“日历效应”对股票价格的影响. 文中利用李胜歌等^[18]提出的赋权“已实现”双幂次变差波动(WRBV)来计量股价的波动,其计算式为

$$\left. \begin{aligned} x_{i,t} &= \ln y_{i,t} - \ln y_{i,t-1}, \\ \text{WRBV}_{i,t}(r,s) &= \omega_{i,t} \mid x_{i,t} \mid^r \mid x_{i,t-1} \mid^s, \\ \omega_{i,t} &= \arg \min \left\{ (1/n)^{1-(r+s)/2} \left[\sum_{i=1}^n \omega_{i,t} \mid x_{i,t} \mid^r \mid x_{i,t-1} \mid^s \right] \right\}, \\ \text{WRBV}_{m,t} &= \sum_{i=1}^k \omega_{i,t} \text{WRBV}_{i,t}. \end{aligned} \right\} \quad (12)$$

式(12)中: $y_{i,t}$ 为股票 i 在交易日 t 的价格; $x_{i,t}$ 为对数价格的增长率; $\text{WRBV}_{i,t}$ 为股票 i 在交易日 t 的波动; $\omega_{i,t}$ 为使 $\text{WRBV}_{i,t}$ 在所有 n 个交易日中加权平均后取得最小值时的权重,并满足 $\sum_{i=1}^n \omega_{i,t} = 1$; $\text{WRBV}_{m,t}$ 为整个股票市场在交易日 t 内的波动; $\omega_{i,t}$ 为股票 i 在交易日 t 的市值权重,并满足 $\sum_{i=1}^k \omega_{i,t} = 1$, k 为交易日 t 内整个市场的股票总数; r,s 为可变参数,文中取 $r=s=1$. 李胜歌等^[18]研究表明:在 $r=s=1$ 时, $\text{WRBV}_{i,t}(r,s)$ 的数值比 r,s 取其他值时的有效性都更高.

将 $\text{WRBV}_{m,t}$ 计算出来并构成时间序列,如图 1 所示. 从图 1 可以看出,1992 年年末至 1993 年 $\text{WRBV}_{m,t}$ 波动较为强烈,在此后的 1995 年和 1997 年, $\text{WRBV}_{m,t}$ 变化更为加剧;经过相对平稳的 10 a 发展,到 2007 年、2008 年年末及 2009 年年初, $\text{WRBV}_{m,t}$ 又有波动上升的趋势,最后直到 2010 年 $\text{WRBV}_{m,t}$ 波动才回落.

由于解释变量的数据均来源于上市公司的季度报表,与 $\text{WRBV}_{m,t}$ 的日间数据不具有可比性,因此将 $\text{WRBV}_{m,t}$ 调整成季度数据进行分析,调整式为

$$\text{WRBV}_{m,q} = (1/g) \sum_{t=1}^g \text{WRBV}_{m,t}.$$

其中: $\text{WRBV}_{m,q}$ 为市场在季度 q 的价格波动, g 为季度 q 的交易天数.

2.3 解释变量的选择

为了检验“净资产收益率假说”、“市帐率假说”和“成交量假说”,选选择解释变量有净资产收益率、市帐率和股票成交量. 这里用 $\text{ROE}_{i,q}$ 表示股票 i 在季度 q 的净资产收益率(年化收益率),用 $\text{MABA}_{i,q}$ 表示股票 i 在季度 q 的市帐率,用 $\text{VOLU}_{i,q}$ 表示股票 i 在季度 q 的成交量;而市场在季度 q 的净资产收益率、市帐率和成交量分别由上述个股指标经市值加权而得,分别用 $\text{ROE}_{m,q}$, $\text{MABA}_{m,q}$ 和 $\text{VOLU}_{m,q}$ 表示.

2.4 实证结果

考虑财务报表的发布对股票投资者决策的滞后性,所取为解释变量的滞后一期,即 $\text{ROE}_{m,q}(-1)$, $\text{MABA}_{m,q}(-1)$ 和 $\text{VOLU}_{m,q}(-1)$;然后,再将其逐一换成式(1)中的解释变量 X_t ,构造 3 个单变量方程来检验它们各自对 $\text{MABA}_{m,q}$ 影响.

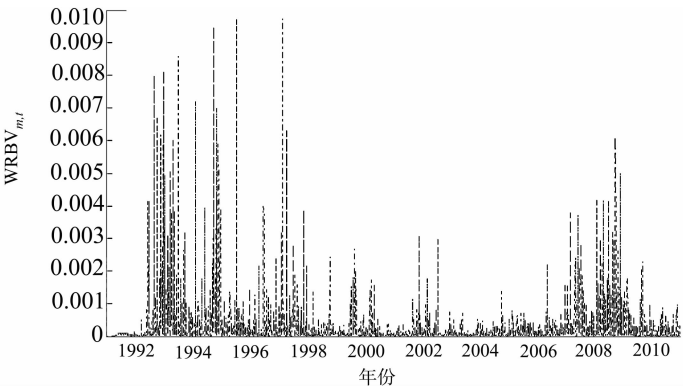


图 1 $\text{WRBV}_{m,t}$ 随时间的变化
Fig. 1 $\text{WRBV}_{m,t}$ changes with time

在每个单变量方程中,首先利用对分数差分阶数 d 、自回归阶数 α 、移动平均阶数 β 识别的公式,识别其 (α, d, β) 的类型;其次,按照正交序列半参数方法估计各个方程;最后,再根据广义似然比检验和 Smirnov 检验来衡量模型估计的效果,结果如表 1 所示.

需要说明的是, d 的识别结果只与被解释变量 $WRBV_{m,q}$ 有关,而无论解释变量为 $ROE_{m,q}(-1)$, $MABA_{m,q}(-1)$, 还是 $VOLU_{m,q}(-1)$, d 的识别结果都是一样的.

表 1 单变量方程的估计与检验

Tab. 1 Estimation and tests of single variable equation

被解释变量	解释变量	(α, d, β) 的识别类型	广义似然比检验	Smirnov 检验
$ROE_{m,q}$	$ROE_{m,q}(-1)$	$(0, 0.272, 2)$	$1.13 \times 10^{-4} \text{①}$	0.980 5
$MABA_{m,q}$	$MABA_{m,q}(-1)$	$(2, 0.272, 2)$	$1.75 \times 10^{-2} \text{①}$	0.004 7 ^①
$VOLU_{m,q}$	$VOLU_{m,q}(-1)$	$(2, 0.272, 2)$	0.031 2*	0.014 5 ^①

① 表示在 5% 的置信水平下显著

从表 1 的检验结果可看出:当解释变量为 $ROE_{m,q}(-1)$ 时,模型的识别结果为 $(0, 0.272, 2)$, 估计的结果通过了广义似然比检验,但是没能通过 Smirnov 检验,说明 $ROE_{m,q}(-1)$ 作为解释变量不成立;当解释变量为 $MABA_{m,q}(-1)$ 时,模型的识别结果为 $(2, 0.272, 2)$, 估计的结果通过了广义似然比检验和 Smirnov 检验,表明可以用 $MABA_{m,q}(-1)$ 来解释 $WRBV_{m,q}$;当解释变量为 $VOLU_{m,q}(-1)$ 时,模型的识别结果为 $(2, 0.272, 2)$, 估计的结果也通过广义似然比检验和 Smirnov 检验,意味着 $MABA_{m,q}(-1)$ 对 $WRBV_{m,q}$ 影响显著.

经过分析,认为 $MABA_{m,q}(-1)$ 和 $VOLU_{m,q}(-1)$ 是 $WRBV_{m,q}$ 变化的主要原因,而 $ROE_{m,q}(-1)$ 对 $WRBV_{m,q}$ 的影响不显著. 构建了二元的 LM-ARMAX (α, d, β) 模型,经过分数差分阶数 d 、自回归阶数 α 、移动平均阶数 β 识别的公式,模型的识别结果为 $(2, 0.272, 2)$. 因此,模型的具体形式为

$$(1 - \phi_1 L - \phi_2 L^2)WRBV_{m,q} = (1 - \theta_1 L - \theta_2 L^2)\epsilon_q + f(MABA_{m,q}(-1), VOLU_{m,q}(-1)).$$

采用正交序列的半参数估计,可得到 $\phi_1 = 0.409\ 7, \phi_2 = 0.123\ 2, \theta_1 = 0.440\ 7$ 和 $\theta_2 = -0.057\ 5$. 经计算相关检验指标,可得到广义似然比检验 p 值为 $3.946\ 7 \times 10^{-11}$, Smirnov 检验 p 值为 0.002 7. 由此可见,包含 $MABA_{m,q}(-1)$ 和 $VOLU_{m,q}(-1)$ 的 LM-ARMAX (α, d, β) 模型,显著地通过了检验,并且双变量模型的显著性要远远高于单变量模型. 这表明,用 $MABA_{m,q}(-1)$ 和 $VOLU_{m,q}(-1)$ 来解释 $WRBV_{m,q}$ 的变化非常合适;而且, $\phi_1 > \phi_2, |\theta_1| > |\theta_2|$, 这就说明 $WRBV_{m,q-1}$ 和 ϵ_{q-1} 对 $WRBV_{m,q}$ 影响比 $WRBV_{m,q-2}$ 和 ϵ_{q-2} 更为深刻.

此外,还得到了 $\partial WRBV_{m,q} / \partial MABA_{m,q}(-1)$ 和 $\partial WRBV_{m,q} / \partial VOLU_{m,q}(-1)$ 的偏导数图形,以及 $\widehat{WRBV}_{m,q}$ 与 $MABA_{m,q}$ 的拟合情况,如图 2, 3 所示. 图 3 中的纵坐标刻度为负值,这是因为 $WRBV_{m,q}$ 考虑到数据的数量级匹配问题,在建模之前对 $WRBV_{m,q}$ 进行过对数处理,而解释变量也进行过类似处理.

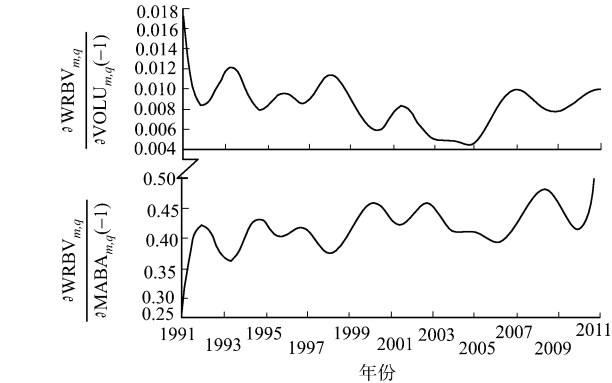


图 2 偏导数的估计结果

Fig. 2 Estimates of partial derivative

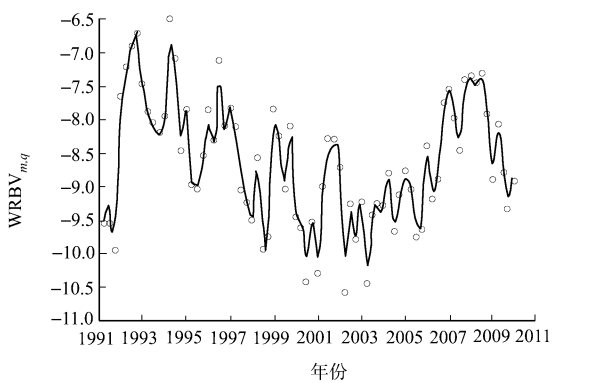


图 3 $\widehat{WRBV}_{m,q}$ 与 $WRBV_{m,q}$ 的拟合情况

Fig. 3 Fitting of $WRBV_{m,q}$ and $\widehat{WRBV}_{m,q}$

从图 2 可知, $MABA_{m,q}(-1)$ 对 $WRBV_{m,q}$ 的影响自 1991—2005 年左右有迂回下滑的迹象,从最开始的 0.016 0 下降到最低点的 0.004 2;但从 2006 年后又开始爬行回升,到了 2010 年底,几乎接近 0.01

的水平.与此不同的是, $VOLU_{m,q}(-1)$ 对 $WRBV_{m,q}$ 的影响,尽管起伏跌宕,但始终保持缓缓上升的趋势.此外,从参数的数值也可看出, $VOLU_{m,q}(-1)$ 对 $WRBV_{m,q}$ 的影响更为突出,几乎是 $MABA_{m,q}(-1)$ 影响程度的 50 倍左右.

从图 3 模型估计值与真实值拟合的情况来看,所建立的半参数模型模拟效果非常良好,尽管没有穿过所有的样本点,但是大体上已经描述出了股票价格的波动的变化趋势.如此高效地反映样本起伏跌宕的变化,正是半参数模型相对于线性模型所无法比拟的优势.

3 结论

选择沪市 1991—2010 年所有上市公司的数据,利用 LM-ARMAX 模型进行半参数估计,经过相关检验,可以得出以下 3 各主要结论.

1) 对股价的波动进行 GPH 分析,可以得到 $\hat{d}=0.272$. 这说明股价的波动存在十分显著的长记忆性,资本市场的分形结构非常显著.这个结论与很多相关文献的观点不太一致,导致这种差异可能有以下 3 个主要原因:a) 样本的时间跨度.文中选择的时间跨度长达 20 a,而许多相关文献考察的时间跨度较短,如李明等^[19]样本数据仅为 2005 年 6 月 8 日到 2008 年 9 月 4 日;b) 样本数据的来源.采用的是所有上市公司的股价波动加权,而目前绝大部分文献的数据都是股市指数,如王相宁等^[20]、苏桂芳等^[21]使用的就是上证综合指数;c) 估计方法.采用的是 GPH 周期图估计,而现有大量研究使用的 R/S 检验,当短记忆和异方差同时存在时不具备稳健性.应该说,文中对上述问题的处理,是具有一定优越性的.

2) 根据 LM-ARMAX(α, d, β)模型及其检验的结果,决定股价的波动有 3 个原因:一是自身的历史波动,二是股票的成交量,三是股票的市帐率(背后隐藏的是公司盈利能力).而且,前两个是决定股价波动的主要因素,后者重要程度略低.对于自身的历史波动,可以认为股价的波动具有“惯性”,在不存在任何外在因素使之偏离轨道之前,会一直保持相当的波动继续下去.对于股票的成交量,其上升背后反映的是资本市场交易费用的减小,公司财务数据甚至是内幕消息更为容易的获取,还有铺天盖地讲说投资的股评家和分析师.显然,这些因素都会导致股价波动的加大,而股票成交量正好反映了上述这些不可观测的因素.对于股票的市帐率就更加明显了.

此外,文中的模型并不支持“净资产收益率假说”.文中的设想是:投资主体在进行相关交易时,可能并不非常在意净资产收益率的绝对水平,而是给予净资产收益率的增量部分更多的关注.所以,资产收益率与股价波动关系不大,但是净资产收益率的变化率可能与股价波动相关.当然,这也是后续研究的重点.

3) 就模型所估计的参数来看,过去的股价波动确实会对现时的股价波动产生影响,但是这种作用的会随着时间的推移而减小.这表明,投资者在进行股票交易的时候重点考虑的是该股票近两个季度内的波动情况,但是不会关注得太长久.也即资本市场具有马尔科夫链似的“无后效性”,历史的信息已经堆积到过去的半年,考察长远的曾经意义不大.

此外, $MABA_{m,q}(-1)$ 和 $VOLU_{m,q}(-1)$ 对股价波动的影响呈现非线性的特点,随着时间的变化, $MABA_{m,q}(-1)$ 的影响程度先下降后上升,而 $VOLU_{m,q}(-1)$ 的影响程度则一致保持持续上升趋势.这个结论,与建立半参数模型的最初设想所吻合.在金融政策反复调控的传导机制下,在上市公司不断更新的资本市场里,在投资主体关注重点的持续变化中,传统的既定参数计量理论似乎与现实偏离较远,因此需要更多、更好、更全面的变参数半参数模型.

参考文献:

- [1] 郝军红,高丽峰,李平,等.上海证券市场股票价格波动的因素分析[J].沈阳工业大学学报,2007,29(3):355-360.
- [2] VUOLTEENAHO T. What drives firm-level stock return? [J]. Journal of Finance,2002,57(1):233-264.
- [3] CORIS F. A Simple long memory model of realized volatility[J]. Social Science Research Network,2004,7(1):174-196.
- [4] PÁSTPR L, VERONESI P. Stock valuation and learning about profitability[J]. Journal of Finance,2003,58(5):

1749-1790.

[5] CAO C, SIMIN T, ZHAO J. Can growth options explain the trend in idiosyncratic risk? [J]. Review of Financial Studies, 2008, 21(6): 2599-2633.

[6] SCHWERT G W. Stock volatility in the new millennium: How wacky is NASDAQ? [J]. Journal of Monetary Economics, 2002, 49(1): 3-26.

[7] WEI S X, ZHANG C. Why did individual stocks become more volatile? [J]. Journal of Business, 2006, 79(1): 259-292.

[8] GEWEKE J, PORTER-HUDAK S. The estimation and application of long memory time series models[J]. Journal of Time Series Analysis, 1983, 4(4): 221-238.

[9] ROBINSON P M. Gaussian semiparametric estimation of long range dependence[J]. The Annals of Statistics, 1995, 23(5): 1630-1661.

[10] ANDREWS D W K, GUGGENBERGER P. A bias-reduced log-periodogram regression estimator for the long-memory parameter[J]. Econometrica, 2003, 71(2): 675-712.

[11] HURVICH C M, TSAI C L. Regression and time series model selection in small samples[J]. Biometrika, 1989, 76(2): 297-307.

[12] SIMONOFF J S. Nonparametric regression using an improved akaike information criterion[J]. Journal of Statist, 2006, 60: 271-293.

[13] HART J D. Nonparametric smoothing and lack-of-fit tests[M]. New York: Springer, 1997.

[14] FAN Jian-qing, HUANG Tao. Profile likelihood inferences on semiparametric varying-coefficient partially linear models[J]. Bernoulli, 2005, 11(6): 1031-1057.

[15] KREISS J P, NEUMANN M, YAO Qi-wei. Bootstrap tests for simple structures in nonparametric time series regression[J]. Statistics and its Interface, 2008, 1(2): 367-380.

[16] ANDERSEN T G, BOLLERSLEV T, DIEBOLD F X, et al. Exchange rate returns standardized by realized volatility are (nearly) Gaussian[J]. Multinational Finance Journal, 2000, 4(3/4): 155-157.

[17] BARNDORFF-NIELSEN O E, SHEPHARD N. Power and bipower variation with stochastic volatility and jumps [J]. Journal of Financial Econometrics, 2004, 2(1): 1-37.

[18] 李胜歌, 张世英. “已实现”双幂次变差与多幂次变差的有效性分析[J]. 系统工程学报, 2007, 22(3): 280-286.

[19] 李明, 赵焯. 证券市场的长记忆特征比较研究[J]. 西部金融, 2010(7): 17-18.

[20] 王相宁, 邹佳. 可加模型及其在金融市场波动率估计中的应用[J]. 系统管理学报, 2009, 18(1): 82-88.

[21] 苏桂芳, 胡日东, 陈家干. 中国股市长记忆检验的滑动分块自助法仿真[J]. 华侨大学学报: 自然科学版, 2009, 30(3): 338-342.

Stock Price Volatility Analysis Based on Semi-Parametric LM-ARMAX Model

CHEN Chun-chun, HU Ri-dong

(College of Economics and Finance, Huaqiao University, Quanzhou 362021, China)

Abstract: We selected all listed companies data of 1991—2010 in Shanghai stock exchange, established long memory-autoregressive moving average with exogenous variables (LM-ARMAX) model to study the determining factors of stock price volatility, and then, had a generalized likelihood ratio test based on the semi-parametric estimation and Smirnov test based on Wild Bootstrap according to the results of semi-parametric estimation of the model. The results show that the major factors of the stock price volatility are the book ratio and the volume, not the income ratio of net assets; compared to “index research” and “sample alternative research”, the empirical data is more accurate and more convincing.

Keywords: stock prices; LM-ARMAX model; volatility; semi-parametric estimation; nonlinear