

基于高频交易数据的“羊群行为”测度模型^①

朱菲菲¹ 唐 涯¹ 徐建国^{2,3} 李宏泰⁴

(1. 北京大学光华管理学院; 2. 北京大学国家发展研究院;
3. 北京大学数字金融研究中心; 4. 漳州市人民政府办公室)

研究目标: 对“羊群行为”测度进行理论建模和实证检验。**研究方法:** 通过结构化方法构建拓展的“羊群行为”测度模型, 使用递归方法和极大似然估计方法对参数进行估计, 并基于中国和美国市场的高频交易数据对“羊群行为”程度进行测量。**研究发现:** 拓展的“羊群行为”测度模型不仅可以对“羊群行为”的短期脆弱性特征加以刻画, 而且可以较好地地区分“伪羊群行为”和“基于信息的羊群行为”。拓展模型下, 知情交易者交易大额单的比例显著高于噪声交易者, 而且发生“羊群行为”的知情交易者占比也更高。**研究创新:** 创新性地考虑“交易单顺序”和“交易单规模”, 并使用高频数据对“羊群行为”程度进行度量。**研究价值:** 更加准确地识别和度量“羊群行为”, 提高金融市场定价效率。

关键词 结构模型 高频交易数据 交易单顺序 交易单规模

中图分类号 F831 **文献标识码** A

引 言

“羊群行为”(Herd Behavior), 又称“跟风行为”或者“从众行为”, 是指投资者在投资决策中放弃自己的私人信息, 依据其他参与者的行为来决定自己的投资决策。众所周知, 金融市场的基本功能之一是通过价格的变化汇集和传递信息, 从而实现资源的有效配置(Hayek, 1945)。投资者的信息能否全面、准确、迅速地反映在价格中, 是一个市场是否有效的重要依据(Fama, 1970)。由于在“羊群行为”中, 投资者会降低自己私人信息的权重, 更多地依照公共信息来进行市场交易, 因此这会阻碍有价值的私人信息迅速反映在股价中, 从而降低了资本市场的信息透明度、定价效率以及预测准确度(孙培源和施东晖, 2002; 许年行等, 2013; Bikhchandani 和 Sharma, 2000; Da 和 Huang, 2016)。严重时还可能导致资产价格偏离基本面, 降低资本市场效率, 扭曲资源的有效配置, 引发市场动荡等(董志勇和韩旭, 2007)。

光大证券的“乌龙指”事件, 是“羊群行为”的一个典型案例。2013年8月16日, 光大证券自营的策略交易系统出现错误, 生成巨量市价委托买单, 当日11点05分, 大批权重

^① 本文受到国家自然科学基金项目(71472006、71772004)的资助。

股瞬间被一两个大单拉升之后,大批巨额买单跟着涌入,导致71只权重股瞬间涨停,上证指数瞬间从2074点涨到2198.85点,两分钟交易额高达87亿元。在短暂的暴涨中,大量投资者在观测到巨量市价委托买单和大幅价格拉升后跟进买入,导致指数的持续上升。当天下午真相大白之后,股指便开始下跌,上证指数一路跌至2011点。“乌龙指”事件生动地反映了市场上投资者的“羊群行为”及其带来的市场剧烈波动。显然,为了更加深入地理解投资者的交易行为,完善我国股票市场的价格发现功能,建立透明高效的多层次资本市场,对投资者“羊群行为”的研究就显得十分必要。

虽然“羊群行为”在学术界备受关注,但现有的研究却存在明显的缺陷。一方面,大量关于“羊群行为”的理论研究和实证研究之间存在明显鸿沟,即理论研究者通常将“羊群行为”处理为一种“事前假定存在”的行为,而实证数据检验的往往是“事后实际存在”的结果(Cipriani和Guarino,2014)。另一方面,实证研究中,学者们通常将“交易的聚集性”^①等价于“羊群行为”(Grinblatt等,1995;Wermers,1999),虽然Lakonishok等(1992)提出的LSV方法能够测算个股的“羊群行为”大小,但其测算结果既可能是由不对称信息下所导致的“基于信息的羊群行为”(Informational Herding),即真实的“羊群行为”,也可能是由投资者对公开信息的集体反应所导致的“伪羊群行为”(Spurious Herding)。并且由于该方法基于低频交易数据进行测量,因此其也无法捕捉“羊群行为”的短期脆弱性特征(Bikhchandani等,1992)。

直到最近,Cipriani和Guarino(2014)才第一次完成了“羊群行为”理论模型和实证检验的对接工作。作者在Glosten和Milgrom(1985)以及Avery和Zemsky(1998)的序贯交易模型基础上,构建一个结构模型,并使用金融市场上的日内高频交易数据对“基于信息的羊群行为”进行了度量。这篇发表在《美国经济评论》(*American Economic Review*)的重要论文给出了一套如何使用金融市场数据去度量理论模型中“羊群行为”程度的标准方法论,因此在“羊群行为”文献中具有重要的奠基性意义。但是该文中的结构模型也存在一个不足,即作者并没有考虑到“交易单规模”对“羊群行为”的影响。而我们都知,大额交易单和小额交易单所蕴含的信息含量是不同的,一个大的交易单比几个小的交易单更可能向投资者传递关于股票的价值信息,因此交易单规模很可能是影响“羊群行为”测度的重要变量(Easley和O'Hara,1987、1992;Easley等,1997)。此外,通过交易数据估计模型参数时,仅有交易单的方向和规模是影响“羊群行为”程度的关键因素,而“交易单顺序”并不会影响“羊群行为”程度。然而,在现实交易中,交易单的顺序对于交易行为有着重要的影响,一连串的买单或卖单要比分散的交易单更有可能引起交易者的“羊群行为”(Glosten和Milgrom,1985;Cipriani和Guarino,2014)。

基于以上考虑,本文提出一个拓展的“羊群行为”测度模型,该模型对“羊群行为”的理论文献做出两方面重要改进:首先,借鉴Easley等(1997),本文在考虑“交易单顺序”的情况下,通过使用递归和极大似然估计方法,对“羊群行为”测度模型的参数进行了估计;其次,通过考虑“交易单规模”,本文拓展的“羊群行为”测度模型能够额外捕捉由单个大交易单而引起的“羊群行为”。

此外,本文还弥补了有关“羊群行为”实证研究的不足:首先,借鉴Christoffersen和

^① 交易的聚集性是指投资者会比较密集地进行投资决策,交易的频率远高于在独立交易假设下的频率(Lakonishok等,1992;Grinblatt等,1995;Wermers,1999)。

Tang (2010) 的研究, 本文使用日内高频交易数据对原有的 LSV 方法进行改进, 从而避免了现有国内研究中使用季度持仓数据对“羊群行为”衡量时可能出现的低估和忽略。其次, 由于 LSV 方法, 无法区分“伪羊群行为”和“基于信息的羊群行为”, 因此其估计结果存在高估。针对此, 我们提出了基于最大似然估计的 LSV 方法 (ML-LSV), 该方法可以更好地区分两类“羊群行为”, 从而使得度量结果更为精确。

一、基本模型

本文在 Easley 等 (1997)、Cipriani 和 Guarino (2014) 的研究基础上, 基于金融市场的交易数据构建了一个结构模型。模型的交易结构可以用图 1 来刻画。

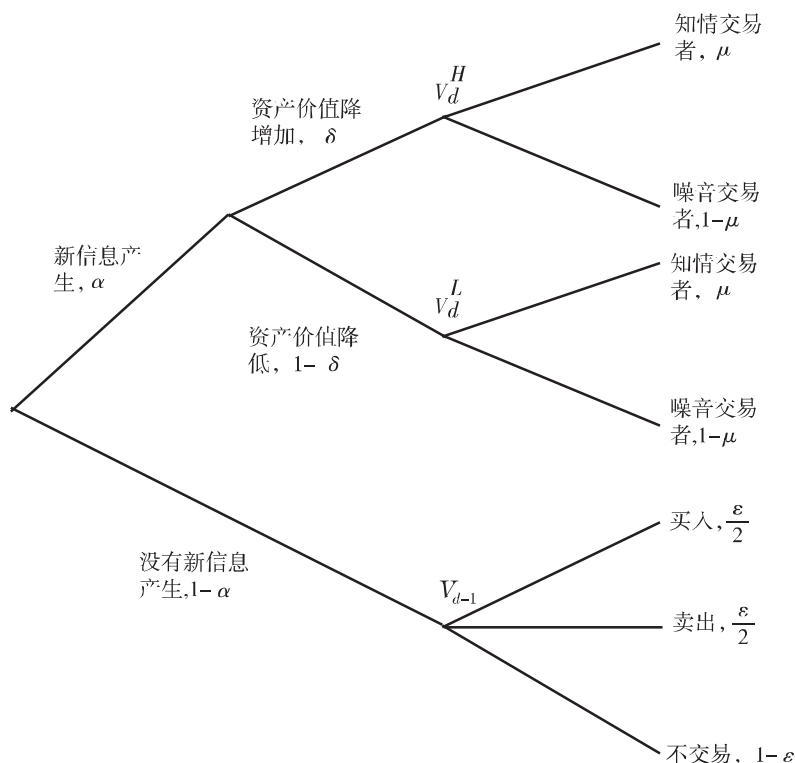


图 1 基本模型交易结构

假设市场中存在两类交易者（知情交易者和噪声交易者），所有交易者的交易对手都为造市商^①。交易日为 $d=1, 2, 3, \dots$ ，每个交易日的交易时间为离散的，并标记为 $t=1, 2, 3, \dots$ 。在第 d 个交易日开始时，有 $1-\alpha$ 的概率没有新信息产生，此时资产的基本价值 V_d 与前一天保持一样，即 $V_d=v_{d-1}$ ；而有 α 的概率有新信息产生，此时资产的基本价值会产生变化。在有新信息产生的情况下，有 $1-\delta$ 的概率发生利空消息，此时资产价值降为 $v_{d-1}-\lambda^L$ ，记 $v_d^L=v_{d-1}-\lambda^L$ ；有 δ 的概率发生利好消息，此时资产价值增加为 $v_{d-1}+\lambda^H$ ，记 $v_d^H=v_{d-1}+\lambda^H$ ，其中 $\lambda^L>0, \lambda^H>0$ 。假设信息事件在所有交易日中都独立分布，并且价格为鞅过程，即 $(1-\delta)\lambda^L=\delta\lambda^H$ 。

假设每次交易的资产数量为一单位，投资者可以买入（buy）、卖出（sell）或者选择不

^① Λ 股市场不是一个表面意义上的造市商市场，但是我们可以将 Λ 股市场看作一个由无数个小造市商构成的市场，投资者与这些无数的小造市商进行交易。

交易,即投资者的行动空间为 $A = \{buy, sell, notrade\}$ 。记第 d 个交易日时间 t 时投资者的决策为 X_t^d ,而之前所有的交易历史为 H_t^d 。在没有信息发生时,所有的投资者均为噪声交易者。由于流动性冲击等原因,噪声交易者买入、卖出和不交易的概率分别为 $\epsilon/2$, $\epsilon/2$ 和 $1-\epsilon$,其中 $0 < \epsilon < 1$ 。而在有信息的交易日中,知情交易者被选择参与交易的概率为 μ ,噪声交易者被选择参与交易的概率为 $1-\mu$,其中 $\mu \in (0, 1)$ 。

造市商根据知情交易者参与交易的概率设定买卖价格,记在交易日 d 时刻 t 时,造市商的卖价(投资者的买入价格)为 a_t^d ,造市商的买价(投资者卖出的价格)为 b_t^d 。由于造市商之间充分竞争^①,最终他们会获得零利润,即买卖价格等于基于既有信息下资产价值的条件期望,如式(1)所示:

$$\begin{aligned} a_t^d &= E(V_d | h_t^d, X_t^d = buy, a_t^d, b_t^d) \\ b_t^d &= E(V_d | h_t^d, X_t^d = sell, a_t^d, b_t^d) \end{aligned} \quad (1)$$

知情交易者具有关于资产价值的新的私有信息 S_t^d ,其密度函数由式(2)决定:

$$\begin{aligned} g^H(s_t^d | v_d^H) &= 1 + \tau(2s_t^d - 1) \\ g^L(s_t^d | v_d^L) &= 1 - \tau(2s_t^d - 1) \end{aligned} \quad (2)$$

其中 $\tau \in (0, \infty)$ 。该参数反映了新信息中所包含的信息量,如果 $\tau \rightarrow 0$,此时的密度函数为均匀分布,即新信号中不包含任何信息;随着 τ 的增加,信号中包含的信息量也随之增加。

当 $\tau \in (0, 1]$,密度函数的区间为 $[0, 1]$;当 $\tau > 1$, g^H 的区间为 $[\tau-1/2\tau, \tau-1+2\sqrt{\tau}/2\tau]$, g^L 的区间为 $[\tau+1-2\sqrt{\tau}/2\tau, \tau+1/2\tau]$ 。私有信号 S_t^d 满足单调似然比的性质,获得新信号后的似然比为:

$$\frac{\Pr(V_d = v_d^H | h_t^d, s_t^d)}{\Pr(V_d = v_d^L | h_t^d, s_t^d)} = \frac{g^H(s_t^d | v_d^H)}{g^L(s_t^d | v_d^L)} \frac{\Pr(V_d = v_d^H | h_t^d)}{\Pr(V_d = v_d^L | h_t^d)} \quad (3)$$

如果 $s_t^d > 0.5$,这一似然比将高于获取信号前的似然比;如果 $s_t^d < 0.5$,这一似然比将低于获取信号前的似然比。因此,我们将大于 0.5 的信号记为利好信号,将低于 0.5 的信号记为利空信号。

值得注意的是,不同于 Easley 等(1997)模型中所有知情交易者都知道关于资产价值的精确信息,在本文的信息结构下,由于每个人得到的信息准确度 τ 不同,知情交易者存在异质性。他们可能会忽略自己得到的新信息,依赖前人的交易信息进行决策,并参与到“羊群行为”之中,因此其他交易者的历史交易信息会对其关于资产价格的预期产生影响。此时,交易单的顺序就变得非常关键。鉴于此,在本文的结构模型中,我们不仅需要考虑交易单规模和方向,还需要考虑交易单顺序对“羊群行为”的影响。

知情交易者的支付函数 $U: \{v_d^L, v_d^H\} \times A \times \{v_d^L, v_d^H\}^2 \rightarrow R^+$ 由式(4)决定,知情交易者可以通过选择 X_t^d (买入,卖出或不交易)从而最大化自己的期望效用 $E[U(v_d, X_t^d, a_t^d, b_t^d) | h_t^d, s_t^d]$ 。

$$U(v_d, X_t^d, a_t^d, b_t^d) = \begin{cases} v_d - a_t^d & \text{if } X_t^d = buy \\ 0 & \text{if } X_t^d = notrade \\ b_t^d - v_d & \text{if } X_t^d = sell \end{cases} \quad (4)$$

① 由于我们将 A 股市场看作由无数个小造市商组成的市场,因此可以粗略地假设造市商之间满足充分竞争。

在时刻 t ，知情交易者的交易决策可以由卖出临界值 σ_t^d 和买入临界值 β_t^d 来刻画，当知情交易者获得的信号小于 σ_t^d 时，其会选择卖出；当获得的信号大于 β_t^d 时，其会选择买入；而获得的信号位于二者之间时，他会选择不交易。临界值由式 (5) 决定：

$$\begin{aligned} E(V_d | h_t^d, \sigma_t^d) &= b_t^d \\ E(V_d | h_t^d, \beta_t^d) &= a_t^d \end{aligned} \quad (5)$$

“羊群行为”产生的条件为：在交易日 d 时刻 t ，如果买入临界值 $\beta_t^d < 0.5$ 或者卖出临界值 $\sigma_t^d > 0.5$ ，则会有买入“羊群行为”或卖出“羊群行为”发生。具体而言，如果买入临界值 $\beta_t^d < 0.5$ ，那么即使在利空信号（信号小于 0.5）发生的情况下，获得信号位于买入临界值和 0.5 之间的投资者仍然会选择买入，由此引发买入“羊群行为”。如果卖出临界值 $\sigma_t^d > 0.5$ ，那么即使在利好信号（信号大于 0.5）发生的情况下，获得信号位于 0.5 和卖出临界值之间的投资者仍然会选择卖出，由此引发卖出“羊群行为”。

上述模型中，“羊群行为”产生的根本原因在于新信息发生的不确定性以及信息的不对称性。此外，信号中的信息量同样会影响“羊群行为”的发生。值得强调的是，由于该模型仅考虑由于新信息发生的不确定性以及信息的不对称性所导致的“基于信息的羊群行为”，而并没有考虑由于公共信息所导致的“伪羊群行为”，因此可以准确地区分两类“羊群行为”，从而可以对真实的“羊群行为”程度做出准确度量。

二、基本模型的估计和“羊群行为”测度

1. 模型估计

基础模型中，我们不仅考虑了每天买单和卖单的数量，而且还考虑了交易单的顺序。在考虑交易单顺序之后，模型参数的估计就变得更加困难。参考 Easley 等 (1997)，本文首先通过递归方法计算出历史交易概率，然后再使用极大似然方法对参数进行估计，从而最终对“羊群行为”程度进行测量。

由于每天的信息事件相互独立，且历史信息为公共信息，因此交易日 d 交易单的顺序只依赖于当天资产的价值，每天交易的似然函数可写为：

$$L(\Phi; \{h^d\}_{d=1}^D) = \Pr(\{h^d\}_{d=1}^D | \Phi) = \prod_{d=1}^D \Pr(h^d | \Phi) \quad (6)$$

交易日 d 时刻 t 的交易概率 $\Pr(x_t^d | h_t^d, \Phi)$ 依赖于给定交易历史后，知情交易者买入、卖出或者不交易的概率。具体如下式：

$$\Pr(h^d | \Phi) = \prod_{s=1}^t \Pr(x_s^d | h_s^d, \Phi), x_t^d \in \{buy, sell, no\ trade\} \quad (7)$$

其中待估参数 $\Phi = \{\alpha, \delta, \mu, \tau, \varepsilon\}$ ， h^d 为第 d 个交易日最终的交易历史。

下面我们具体展示如何通过递归法求解每个交易日的交易历史概率 $\Pr(h^d | \Phi)$ ，然后写出似然函数式 (6)，从而最终估计模型参数 $\Phi = \{\alpha, \delta, \mu, \tau, \varepsilon\}$ 。

由于买入“羊群行为”的临界值 β_t^d 和卖出“羊群行为”的临界值 σ_t^d 满足下式^①：

$$E(V_d | h_t^d, \beta_t^d) = a_t^d = E(V_d | h_t^d, X_t^d = buy, a_t^d, b_t^d) \quad (8)$$

$$E(V_d | h_t^d, \sigma_t^d) = b_t^d = E(V_d | h_t^d, X_t^d = sell, a_t^d, b_t^d) \quad (9)$$

① 为简便起见，本文之后的所有条件概率公式都省略参数向量 Φ 。

经化简, 以及 $(1-\delta)\lambda^L = \delta\lambda^H$, 式(8)、式(9)可变为:

$$\Pr(v_d^H | h_t^d, \beta_t^d) - \Pr(v_d^H | h_t^d, buy_t^d) = \frac{\delta}{1-\delta} [\Pr(v_d^L | h_t^d, \beta_t^d) - \Pr(v_d^L | h_t^d, buy_t^d)] \quad (10)$$

$$\Pr(v_d^H | h_t^d, \sigma_t^d) - \Pr(v_d^H | h_t^d, sell_t^d) = \frac{\delta}{1-\delta} [\Pr(v_d^L | h_t^d, \sigma_t^d) - \Pr(v_d^L | h_t^d, sell_t^d)] \quad (11)$$

对于知情交易者而言, 在他获得信号 β_t^d 后, 他认为当天发生利好和利空事件的概率分别为:

$$\Pr(v_d^H | h_t^d, \beta_t^d) = \frac{g^H(\beta_t^d | v_d^H) \Pr(v_d^H | h_t^d, V_d \neq v_{d-1})}{g^H(\beta_t^d | v_d^H) \Pr(v_d^H | h_t^d, V_d \neq v_{d-1}) + g^L(\beta_t^d | v_d^L) \Pr(v_d^L | h_t^d, V_d \neq v_{d-1})} \quad (12)$$

$$\Pr(v_d^L | h_t^d, \beta_t^d) = \frac{g^L(\beta_t^d | v_d^L) \Pr(v_d^L | h_t^d, V_d \neq v_{d-1})}{g^H(\beta_t^d | v_d^H) \Pr(v_d^H | h_t^d, V_d \neq v_{d-1}) + g^L(\beta_t^d | v_d^L) \Pr(v_d^L | h_t^d, V_d \neq v_{d-1})} \quad (13)$$

对于造市商而言, 在其观察到投资者的买入行为后, 认为当天发生利好或利空事件的概率分别为:

$$\Pr(v_d^H | h_t^d, buy_t^d) = \frac{\Pr(buy_t^d | v_d^H, h_t^d) \Pr(v_d^H | h_t^d)}{\Pr(buy_t^d | v_d^H, h_t^d) \Pr(v_d^H | h_t^d) + \Pr(buy_t^d | v_{d-1}, h_t^d) \Pr(v_{d-1} | h_t^d) + \Pr(buy_t^d | v_d^L, h_t^d) \Pr(v_d^L | h_t^d)} \quad (14)$$

$$\Pr(v_d^L | h_t^d, buy_t^d) = \frac{\Pr(buy_t^d | v_d^L, h_t^d) \Pr(v_d^L | h_t^d)}{\Pr(buy_t^d | v_d^H, h_t^d) \Pr(v_d^H | h_t^d) + \Pr(buy_t^d | v_{d-1}, h_t^d) \Pr(v_{d-1} | h_t^d) + \Pr(buy_t^d | v_d^L, h_t^d) \Pr(v_d^L | h_t^d)} \quad (15)$$

将式(12)~式(15)代入式(10)得到:

$$\begin{aligned} & \frac{g^H(\beta_t^d | v_d^H) \Pr(v_d^H | h_t^d, V_d \neq v_{d-1}) - \frac{\delta}{1-\delta} g^L(\beta_t^d | v_d^L) \Pr(v_d^L | h_t^d, V_d \neq v_{d-1})}{g^H(\beta_t^d | v_d^H) \Pr(v_d^H | h_t^d, V_d \neq v_{d-1}) + g^L(\beta_t^d | v_d^L) \Pr(v_d^L | h_t^d, V_d \neq v_{d-1})} \\ &= \frac{\Pr(buy_t^d | v_d^H, h_t^d) \Pr(v_d^H | h_t^d) - \frac{\delta}{1-\delta} \Pr(buy_t^d | v_d^L, h_t^d) \Pr(v_d^L | h_t^d)}{\Pr(buy_t^d | v_d^H, h_t^d) \Pr(v_d^H | h_t^d) + \Pr(buy_t^d | v_{d-1}, h_t^d) \Pr(v_{d-1} | h_t^d) + \Pr(buy_t^d | v_d^L, h_t^d) \Pr(v_d^L | h_t^d)} \end{aligned} \quad (16)$$

通过式(16)可解出 $\beta_t^d = f(\Phi)$ ^①。

同样的方法计算式(11)可解出 $\sigma_t^d = f(\Phi)$ ^②。

在求出 β_t^d 和 σ_t^d 后, 我们可进一步计算投资者买入、卖出或不交易的概率。具体见表1。

接下来, 我们将交易日 d 时刻的交易记为 $x_{t-1}^d \in \{buy_t^d, sell_t^d, no\ trade_t^d\}$, 由此, 在观察到投资者买入后, 造市商认为当天发生利好、利空以及没有信息发生的概率分别为:

① 限于篇幅原因, 具体的计算过程在文中省略, 感兴趣的读者可向作者索取。

② 限于篇幅原因, 具体的计算过程在文中省略, 感兴趣的读者可向作者索取。

表 1 不同情形下投资者交易概率

	买入概率	卖出概率	不交易概率
Panel A: 利好信息发生			
$\tau \in (0, 1)$	$(1-\beta_t^d)(1+\tau\beta_t^d)\mu + (1-\mu)\frac{\varepsilon}{2}$	$[(1-\tau)\sigma_t^d + \tau\sigma_t^{d^2}]\mu + (1-\mu)\frac{\varepsilon}{2}$	$1 - (1-\beta_t^d)(1+\tau\beta_t^d)\mu - [(1-\tau)\sigma_t^d + \tau\sigma_t^{d^2}]\mu - (1-\mu)\varepsilon$
$\tau > 1$	$\left(\frac{\tau-1+2\sqrt{\tau}}{2\tau} - \beta_t^d\right) \left(\sqrt{\tau} - \frac{\tau-1}{2} + \tau\beta_t^d\right)\mu + (1-\mu)\frac{\varepsilon}{2}$	$\tau\left(\sigma_t^d - \frac{\tau-1}{2\tau}\right)^2 \mu + (1-\mu)\frac{\varepsilon}{2}$	$1 - \left(\frac{\tau-1+2\sqrt{\tau}}{2\tau} - \beta_t^d\right) \left(\sqrt{\tau} - \frac{\tau-1}{2} + \tau\beta_t^d\right)\mu - \tau\left(\sigma_t^d - \frac{\tau-1}{2\tau}\right)^2 \mu - (1-\mu)\varepsilon$
Panel B: 利空信息发生			
$\tau \in (0, 1)$	$(1-\beta_t^d)(1-\tau\beta_t^d)\mu + (1-\mu)\frac{\varepsilon}{2}$	$[(1+\tau)\sigma_t^d - \tau\sigma_t^{d^2}]\mu + (1-\mu)\frac{\varepsilon}{2}$	$1 - (1-\beta_t^d)(1-\tau\beta_t^d)\mu - [(1+\tau)\sigma_t^d - \tau\sigma_t^{d^2}]\mu - (1-\mu)\varepsilon$
$\tau > 1$	$\tau\left(\frac{\tau+1}{2\tau} - \beta_t^d\right)^2 \mu + (1-\mu)\frac{\varepsilon}{2}$	$\left(\sigma_t^d - \frac{\tau-1+2\sqrt{\tau}}{2\tau}\right) \left(\sqrt{\tau} + \frac{\tau+1}{2} - \tau\sigma_t^d\right)\mu + (1-\mu)\frac{\varepsilon}{2}$	$1 - \tau\left(\frac{\tau+1}{2\tau} - \beta_t^d\right)^2 \mu - \left(\sigma_t^d - \frac{\tau-1+2\sqrt{\tau}}{2\tau}\right) \left(\sqrt{\tau} + \frac{\tau+1}{2} - \tau\sigma_t^d\right)\mu - (1-\mu)\varepsilon$
Panel C: 无信息发生			
	$\varepsilon/2$	$\varepsilon/2$	$1-\varepsilon$

$$\begin{aligned} \Pr(v_d^H | h_t^d, \Phi) &= \Pr(v_d^H | h_{t-1}^d, x_{t-1}^d) = \\ &= \frac{\Pr(x_{t-1}^d | v_d^H, h_{t-1}^d) \Pr(v_d^H | h_{t-1}^d)}{\Pr(x_{t-1}^d | v_d^H, h_{t-1}^d) \Pr(v_d^H | h_{t-1}^d) + \Pr(x_{t-1}^d | v_{d-1}^d, h_{t-1}^d) \Pr(v_{d-1}^d | h_{t-1}^d) + \Pr(x_{t-1}^d | v_d^I, h_{t-1}^d) \Pr(v_d^I | h_{t-1}^d)} \end{aligned} \quad (17)$$

$$\begin{aligned} \Pr(v_d^I | h_t^d, \Phi) &= \Pr(v_d^I | h_{t-1}^d, x_{t-1}^d) = \\ &= \frac{\Pr(x_{t-1}^d | v_d^I, h_{t-1}^d) \Pr(v_d^I | h_{t-1}^d)}{\Pr(x_{t-1}^d | v_d^H, h_{t-1}^d) \Pr(v_d^H | h_{t-1}^d) + \Pr(x_{t-1}^d | v_{d-1}^d, h_{t-1}^d) \Pr(v_{d-1}^d | h_{t-1}^d) + \Pr(x_{t-1}^d | v_d^I, h_{t-1}^d) \Pr(v_d^I | h_{t-1}^d)} \end{aligned} \quad (18)$$

$$\begin{aligned} \Pr(v_{d-1} | h_t^d, \Phi) &= \Pr(v_{d-1} | h_{t-1}^d, x_{t-1}^d) = \\ &= \frac{\Pr(x_{t-1}^d | v_{d-1}^d, h_{t-1}^d) \Pr(v_{d-1}^d | h_{t-1}^d)}{\Pr(x_{t-1}^d | v_d^H, h_{t-1}^d) \Pr(v_d^H | h_{t-1}^d) + \Pr(x_{t-1}^d | v_{d-1}^d, h_{t-1}^d) \Pr(v_{d-1}^d | h_{t-1}^d) + \Pr(x_{t-1}^d | v_d^I, h_{t-1}^d) \Pr(v_d^I | h_{t-1}^d)} \end{aligned} \quad (19)$$

知情交易者在观察到信号后，认为当天发生利好和利空消息的概率分别为：

$$\begin{aligned} \Pr(v_d^H | h_t^d, V_d \neq v_{d-1}) &= \Pr(v_d^H | h_{t-1}^d, x_{t-1}^d, V_d \neq v_{d-1}) = \\ &= \frac{\Pr(x_{t-1}^d | v_d^H, h_{t-1}^d) \Pr(v_d^H | h_{t-1}^d, V_d \neq v_{d-1})}{\Pr(x_{t-1}^d | v_d^H, h_{t-1}^d) \Pr(v_d^H | h_{t-1}^d, V_d \neq v_{d-1}) + \Pr(x_{t-1}^d | v_d^I, h_{t-1}^d) \Pr(v_d^I | h_{t-1}^d, V_d \neq v_{d-1})} \end{aligned} \quad (20)$$

$$\Pr(v_d^L | h_t^d, V_d \neq v_{d-1}) = \Pr(v_d^L | h_{t-1}^d, x_{t-1}^d, V_d \neq v_{d-1}) = \frac{\Pr(x_{t-1}^d | v_d^L, h_{t-1}^d) \Pr(v_d^L | h_{t-1}^d, V_d \neq v_{d-1})}{\Pr(x_{t-1}^d | v_d^H, h_{t-1}^d) \Pr(v_d^H | h_{t-1}^d, V_d \neq v_{d-1}) + \Pr(x_{t-1}^d | v_d^L, h_{t-1}^d) \Pr(v_d^L | h_{t-1}^d, V_d \neq v_{d-1})} \quad (21)$$

由式(17)~式(21)可以看出, t 期概率的计算依赖于 $t-1$ 期的概率, 而 $t-1$ 期又依赖 $t-2$ 期, 依次递推, 因此我们需要用递归的方法对交易概率进行求解。

由于在初始时刻 $t=1$ 时, h_1^d 为空集, 即没有历史交易信息。此时买入、卖出和不交易的概率分别为:

$$\begin{aligned} \Pr(buy_1^d | \Phi) &= \Pr(buy_1^d | v_d^H, \Phi) \Pr(V_d = v_d^H | \Phi) + \\ &\Pr(buy_1^d | v_d^L, \Phi) \Pr(V_d = v_d^L | \Phi) + \Pr(buy_1^d | v_{d-1}, \Phi) \Pr(v_{d-1} | \Phi) \\ \Pr(sell_1^d | \Phi) &= \Pr(sell_1^d | v_d^H, \Phi) \Pr(V_d = v_d^H | \Phi) + \\ &\Pr(sell_1^d | v_d^L, \Phi) \Pr(V_d = v_d^L | \Phi) + \Pr(sell_1^d | v_{d-1}, \Phi) \Pr(v_{d-1} | \Phi) \\ \Pr(notrade_1^d | \Phi) &= 1 - \Pr(buy_1^d | \Phi) - \Pr(sell_1^d | \Phi) \end{aligned} \quad (22)$$

则初始时刻, 买入“羊群行为”的临界值 β_1^d 满足:

$$\begin{aligned} &\frac{[1 + \tau(2\beta_1^d - 1)] \cdot \delta - \frac{\delta}{1-\delta} [1 - \tau(2\beta_1^d - 1)] \cdot (1-\delta)}{[1 + \tau(2\beta_1^d - 1)] \cdot \delta + [1 - \tau(2\beta_1^d - 1)] \cdot (1-\delta)} = \\ &\frac{\left\{ [(1-\tau)(1-\beta_1^d) + \tau(1-\beta_1^{\frac{2}{d}})]\mu + (1-\mu) \cdot \frac{\varepsilon}{2} \right\} \cdot \alpha \cdot \delta - \frac{\delta}{1-\delta} \left\{ [(1+\tau)(1-\beta_1^d) - \tau(1-\beta_1^{\frac{2}{d}})]\mu + (1-\mu) \cdot \frac{\varepsilon}{2} \right\} \cdot \alpha \cdot (1-\delta)}{\left\{ [(1-\tau)(1-\beta_1^d) + \tau(1-\beta_1^{\frac{2}{d}})]\mu + (1-\mu) \cdot \frac{\varepsilon}{2} \right\} \cdot \alpha \cdot \delta + \frac{\varepsilon}{2} \cdot (1-\alpha) + \left\{ [(1+\tau)(1-\beta_1^d) - \tau(1-\beta_1^{\frac{2}{d}})]\mu + (1-\mu) \cdot \frac{\varepsilon}{2} \right\} \cdot \alpha \cdot (1-\delta)} \end{aligned} \quad (23)$$

初始时刻, 卖出“羊群行为”的临界值 δ_1^d 满足:

$$\begin{aligned} &\frac{[1 + \tau(2\delta_1^d - 1)] \cdot \delta - \frac{\delta}{1-\delta} [1 - \tau(2\delta_1^d - 1)] \cdot (1-\delta)}{[1 + \tau(2\delta_1^d - 1)] \cdot \delta + [1 - \tau(2\delta_1^d - 1)] \cdot (1-\delta)} = \\ &\frac{\left\{ [(1-\tau)\sigma_1^d + \tau\sigma_1^{\frac{2}{d}}]\mu + (1-\mu) \cdot \frac{\varepsilon}{2} \right\} \cdot \alpha \cdot \delta - \frac{\delta}{1-\delta} \left\{ [(1+\tau)\sigma_1^d - \tau\sigma_1^{\frac{2}{d}}]\mu + (1-\mu) \cdot \frac{\varepsilon}{2} \right\} \cdot \alpha \cdot (1-\delta)}{\left\{ [(1-\tau)\sigma_1^d + \tau\sigma_1^{\frac{2}{d}}]\mu + (1-\mu) \cdot \frac{\varepsilon}{2} \right\} \cdot \alpha \cdot \delta + \frac{\varepsilon}{2} \cdot (1-\alpha) + \left\{ [(1+\tau)\sigma_1^d - \tau\sigma_1^{\frac{2}{d}}]\mu + (1-\mu) \cdot \frac{\varepsilon}{2} \right\} \cdot \alpha \cdot (1-\delta)} \end{aligned} \quad (24)$$

从式(23)和式(24)可解出 $\beta_1^d = f(\Phi)$ 和 $\sigma_1^d = f(\Phi)$, 由此得到 $t=1$ 时各类情形下的买入行为和卖出行为发生概率。在计算出 $t=1$ 时刻的交易概率后, 可计算时刻 $t=2$ 的交易概率, 依次类推, 直至每个交易日的最后时刻。由此我们可以得到式(7)中的 $\Pr(h^d | \Phi)$, 并写出似然函数式(6), 并最终使用极大似然方法对参数 $\Phi = \{\alpha, \delta, \mu, \tau, \varepsilon\}$ 进行估计。

2. “羊群行为”测度

在用极大似然估计方法完成参数估计后, 进一步地, 我们使用“羊群行为的发生频率”“羊群行为的知情交易者比例”以及“基于极大似然估计的 LSV 方法 (ML-LSV)”来对“羊群行为”进行度量。

(1) “羊群行为”的发生频率。如前文所述, 在交易日 d 时刻 t , “羊群行为”产生的条件为买入临界值 $\beta_t^d < 0.5$ 或卖出临界值 $\sigma_t^d > 0.5$ 。利用参数估计结果, 我们可以得到每个交易时刻所对应的 β_t^d 和 σ_t^d 。然后, 进一步可计算每个交易日买入临界值低于 0.5 和卖出临界

值高于 0.5 的概率,并以此作为“羊群行为”发生频率的度量。

(2) 知情交易者发生“羊群行为”的比例。由于“羊群行为”的发生频率无法反映“羊群行为”是由知情交易者还是由噪声交易者产生的,因此本文进一步计算知情交易者发生“羊群行为”的比例。在计算这一比例之前,首先需要计算每个交易日发生利好或利空事件的概率,我们使用交易日结束时资产价值的后验概率来代替,如果 $\Pr(V_d = v^H | h^d) > 0.9$,则认为该天发生了利好事件,如果 $\Pr(V_d = v^L | h^d) > 0.9$,则认为该天发生了利空事件。随后,我们计算每个交易时刻,知情交易者会在利空信号发生时产生买入“羊群行为”的比例以及利好信号发生时产生卖出“羊群行为”的比例。在发生利好事件的交易日中,信号的分布为 $g^H(s_t^d | v_d^H) = 1 + \tau(2s_t^d - 1)$,针对买入“羊群行为”($\beta_t^d < 0.5$ 时),我们计算“位于 $\beta_t^d \sim 0.5$ 之间的信号”占“位于 $0 \sim 0.5$ 之间的信号”的比例,并将其记为知情交易者发生买入“羊群行为”的比例。在发生利空事件的交易日中,信号的分布为 $g^L(s_t^d | v_d^L) = 1 - \tau(2s_t^d - 1)$,针对卖出“羊群行为”($\sigma_t^d > 0.5$ 时),我们计算“位于 $0.5 \sim \sigma_t^d$ 之间的信号”占“位于 $0.5 \sim 1$ 之间的信号”的比例,并将其记为知情交易者发生卖出“羊群行为”的比例。最后,我们计算在不同的参数取值下,买入“羊群行为”和卖出“羊群行为”的知情交易者比例。表 2 展示了不同情况下的计算结果。

表 2 不同情形下发生“羊群行为”的知情交易者比例

	买入“羊群行为”的知情交易者比例	卖出“羊群行为”的知情交易者比例
Panel A: $\tau \in (0, 1)$		
利好信息	$1 - \frac{(1-\tau) \beta_t^d + \tau \beta_t^{d^2}}{0.5 - 0.25\tau}$	$\frac{(1-\tau) \sigma_t^d + \tau \sigma_t^{d^2} - 0.5 + 0.25\tau}{0.5 + 0.25\tau}$
利空信息	$1 - \frac{(1+\tau) \beta_t^d - \tau \beta_t^{d^2}}{0.5 + 0.25\tau}$	$\frac{(1+\tau) \sigma_t^d - \tau \sigma_t^{d^2} - 0.5 - 0.25\tau}{0.5 - 0.25\tau}$
Panel B: $\tau > 1$		
利好信息	$\frac{(1-\tau)(0.5 - \beta_t^d) + \tau(0.25 - \beta_t^{d^2})}{(1-\tau) \frac{(2\sqrt{\tau}-1)}{2\tau} + \tau \left[0.25 - \left(\frac{\tau+1-2\sqrt{\tau}}{2\tau} \right)^2 \right]}$	$\frac{(1-\tau)(\sigma_t^d - 0.5) + \tau(\sigma_t^{d^2} - 0.25)}{(1-\tau) \left(\frac{\tau-1+2\sqrt{\tau}}{2\tau} - 0.5 \right) + \tau \left[\left(\frac{\tau+1-2\sqrt{\tau}}{2\tau} \right)^2 - 0.25 \right]}$
利空信息	$\frac{(1+\tau)(0.5 - \beta_t^d) - \tau(0.25 - \beta_t^{d^2})}{(1+\tau) \frac{(2\sqrt{\tau}-1)}{2\tau} - \tau \left[0.25 - \left(\frac{\tau+1-2\sqrt{\tau}}{2\tau} \right)^2 \right]}$	$\frac{(1+\tau)(\sigma_t^d - 0.5) + \tau(\sigma_t^{d^2} - 0.25)}{(1+\tau) \left(\frac{\tau-1+2\sqrt{\tau}}{2\tau} - 0.5 \right) - \tau \left[\left(\frac{\tau-1+2\sqrt{\tau}}{2\tau} \right)^2 - 0.25 \right]}$

(3) 基于极大似然估计的 LSV 方法 (ML-LSV)。如前文所述,目前关于“羊群行为”的实证研究多采用 LSV 方法对“羊群行为”进行度量。但是该方法存在两个较为明显的缺陷。首先,LSV 方法并没有对“伪羊群行为”和“基于信息的羊群行为”做出区分,因此该方法对真实“羊群行为”程度的测量很可能存在高估。其次,现有的基于 LSV 方法的实证检验,多基于基金经理的月度或季度持仓数据,而低频数据下的估计会忽略“羊群行为”的短期脆弱性特征,造成了对真实“羊群行为”估计的扭曲。

鉴于此,我们参考 Christofferson 和 Tang (2010) 的研究,使用日度内的高频交易数据对传统的 LSV 方法进行改进,提出基于极大似然估计的 LSV 方法 (ML-LSV)。ML-LSV 方法的本质是,用真正的交易单比例减去投资者在不发生“羊群行为”情况下的交易单比例。由于该

方法可以直接计算不发生“羊群行为”情况下的交易单比例，而不需要用交易日当天所有股票交易行为的截面平均来近似替代这一比例，这使得 ML-LSV 方法比 LSV 方法更加准确。

基于 ML-LSV 方法，如果在交易日 d 发生了利好事件，“羊群行为”的计算公式为：

$$H_{i,d}^{Good} = \left| \frac{B_{i,d}^{Good}}{B_{i,d}^{Good} + S_{i,d}^{Good}} - p_{i,Good}^{ML} \right| - AF(i,d) \quad (25)$$

其中 $B_{i,d}^{Good}$ 和 $S_{i,d}^{Good}$ 为利好事件交易日时的买单和卖单数量， $p_{i,Good}^{ML}$ 为利好事件交易日时，不发生“羊群行为”情况下的买单比例。

如果在交易日 d 发生的是利空事件，则“羊群行为”的计算公式为：

$$H_{i,d}^{Bad} = \left| \frac{B_{i,d}^{Bad}}{B_{i,d}^{Bad} + S_{i,d}^{Bad}} - p_{i,Bad}^{ML} \right| - AF(i,d) \quad (26)$$

其中 $B_{i,d}^{Bad}$ 和 $S_{i,d}^{Bad}$ 为利空事件交易日时的买单和卖单数量， $p_{i,Bad}^{ML}$ 为利空事件交易日时，不发生“羊群行为”情况下的买单比例。

当不发生“羊群行为”时，获得利好信号的交易者会选择买入股票，获得利空信号的投资者会选择卖出股票。因此，利好事件交易日时，不发生“羊群行为”情况下的买单比例为：

$$p_{i,Good}^{ML} = \frac{\mu(0.5 + 0.25\tau) + (1 - \mu) \frac{\varepsilon}{2}}{\mu + (1 - \mu)\varepsilon} \quad (27)$$

在利空事件交易日时，不发生“羊群行为”情况下的买单比例为：

$$p_{i,Bad}^{ML} = \frac{\mu(0.5 - 0.25\tau) + (1 - \mu) \frac{\varepsilon}{2}}{\mu + (1 - \mu)\varepsilon} \quad (28)$$

式 (25) 和式 (26) 中的减项是为了调整超出原假设下所有交易者都独立交易的买单比例，原假设下：每个投资者的交易决策相互独立，买方驱动单服从二项分布。因此，在发生利好事件的交易日中，调整项 $AF(i, d)$ 为：

$$AF(i, d) = \sum_{k=0}^{N_{i,d}} \binom{N_{i,d}}{k} p_{i,Good}^{ML k} (1 - p_{i,Good}^{ML})^{N_{i,d}-k} \left| \frac{k}{N_{i,d}} - p_{i,Good}^{ML} \right| \quad (29)$$

在发生利空事件的交易日中，调整项 $AF(i, d)$ 为：

$$AF(i, d) = \sum_{k=0}^{N_{i,d}} \binom{N_{i,d}}{k} p_{i,Bad}^{ML k} (1 - p_{i,Bad}^{ML})^{N_{i,d}-k} \left| \frac{k}{N_{i,d}} - p_{i,Bad}^{ML} \right| \quad (30)$$

最终，我们得到一段时间内“羊群行为”平均大小的计算公式：

$$H_i^{ML} = \alpha \delta \frac{\sum_{d \in \mathbf{G}} H_{i,d}^{Good}}{N_i^{Good}} + \alpha (1 - \delta) \frac{\sum_{d \in \mathbf{B}} H_{i,d}^{Bad}}{N_i^{Bad}} \quad (31)$$

其中 $\mathbf{G}$ 表示所有发生利好事件的交易日的集合， $\mathbf{B}$ 表示所有发生利空事件的交易日的集合。

三、考虑交易单规模之后的拓展模型

上述基础模型中，我们只考虑了投资者买入或卖出交易单的数量和顺序，但并没有考虑交易

单规模。众所周知，大额交易单和小额交易单所蕴含的信息含量是不同的，一个大的交易单比几个小的交易单更有可能向投资者传递关于股票的价格信息，因此交易单规模是影响“羊群行为”测度的重要变量（Easley 和 O'Hara, 1987、1992；Easley 等, 1997）。鉴于此，我们在基础模型中创新性地加入“交易单规模”，并提出拓展的“羊群行为”测度模型。该模型不仅能捕捉由一系列同方向的交易单所引起的“羊群行为”，还能额外捕捉由单个大交易单而引起的“羊群行为”，因而提高了模型的精度和准确度。显然，这是本文对“羊群行为”理论研究的一大贡献。

具体地，我们记 SB 为小额买入交易单， LB 为大额买入交易单， SS 为小额卖出交易单， LS 为大额卖出交易单。噪声交易者交易大额单的概率为 φ ，交易小额单的概率为 $1-\varphi$ 。知情交易者交易大额单的概率为 ω ，交易小额单的概率为 $1-\omega$ 。在模型中加入参数 (φ, ω) 后，现有的模型变得更为复杂。一方面，拓展模型的待估参数变为 $\Phi = \{\alpha, \delta, \mu, \tau, \varepsilon, \varphi, \omega\}$ ；另一方面，投资者的行动空间变为 $A = \{SB, LB, SS, LS, notrade\}$ 。拓展模型的交易结构可以用图 2 进行刻画。

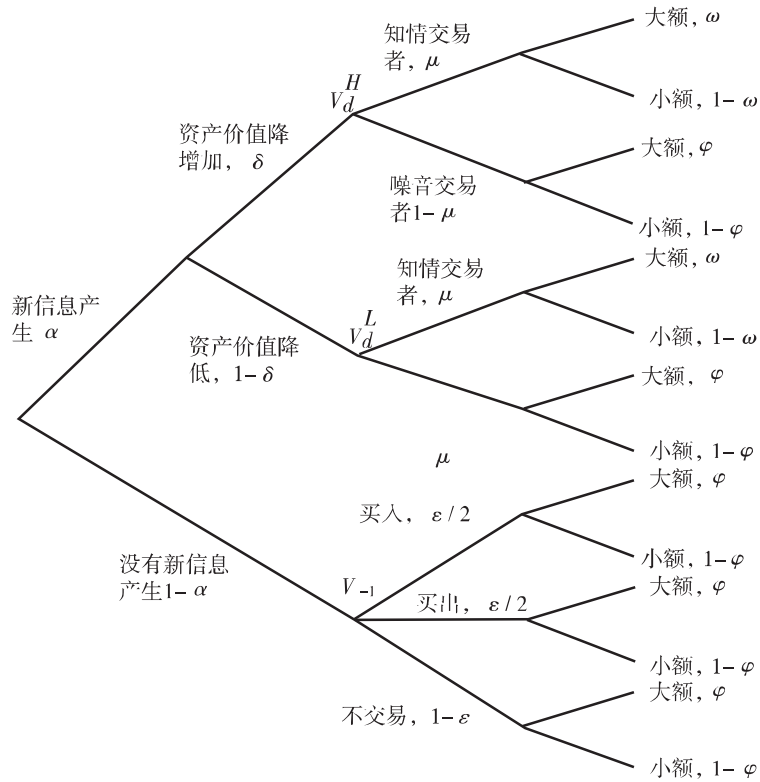


图 2 拓展模型交易结构

拓展模型的似然函数为：

$$L(\Phi; \{h^d\}_{d=1}^D) = \Pr(\{h^d\}_{d=1}^D | \Phi) = \prod_{d=1}^D \Pr(h^d | \Phi)$$

$$\Pr(h^d | \Phi) = \prod_{s=1}^t \Pr(x_s^d | h_s^d, \Phi), x_t^d \in \{SB, LB, SS, LS, notrade\} \quad (32)$$

其中参数 $\Phi = \{\alpha, \delta, \mu, \tau, \varepsilon, \varphi, \omega\}$ ， $h^d \in \{SB, LB, SS, LS, notrade\}^d$ 为第 d 个交易日最后交易时刻的交易历史。

买入和卖出“羊群行为”的临界值 β_t^d 和 σ_t^d 由下面两式决定：

$$E(V_d | h_t^d, \beta_t^d) = a_t^d = E(V_d | h_t^d, \{SB, LB\}, a_t^d, b_t^d) \quad (33)$$

$$E(V_d | h_t^d, \sigma_t^d) = b_t^d = E(V_d | h_t^d, \{SS, LS\}, a_t^d, b_t^d) \quad (34)$$

其中式(33)最终可写成:

$$\begin{aligned}
 & \frac{g^H(\beta_t^d | v_t^d) \Pr(v_t^d | h_t^d, V_d \neq v_{d-1}) - \frac{\delta}{1-\delta} g^L(\beta_t^d | v_t^d) \Pr(v_t^d | h_t^d, V_d \neq v_{d-1})}{g^H(\beta_t^d | v_t^d) \Pr(v_t^d | h_t^d, V_d \neq v_{d-1}) + g^L(\beta_t^d | v_t^d) \Pr(v_t^d | h_t^d, V_d \neq v_{d-1})} \\
 &= \frac{\Pr(\{SB, LB\}^d | v_t^d, h_t^d) \Pr(v_t^d | h_t^d) - \frac{\delta}{1-\delta} \Pr(\{SB, LB\}^d | v_t^d, h_t^d) \Pr(v_t^d | h_t^d)}{\Pr(\{SB, LB\}^d | v_t^d, h_t^d) \Pr(v_t^d | h_t^d) + \Pr(\{SB, LB\}^d | v_{d-1}, h_t^d) \Pr(v_{d-1} | h_t^d) + \Pr(\{SB, LB\}^d | v_t^d, h_t^d) \Pr(v_t^d | h_t^d)}
 \end{aligned} \quad (35)$$

通过式(35)可解出 $\beta_t^d = f(\Phi)$ 。

同理,通过式(34)最终可解出 $\sigma_t^d = f(\Phi)$ 。

在求出 β_t^d 和 σ_t^d 后,我们进一步使用递归方法计算 $\{SB, LB, SS, LS, notrade\}$ 各自所对应的概率,然后计算 $\Pr(x_t^d | h_t^d, \Phi)$,写出似然函数,并最终使用极大似然方法完成参数估计^①。

四、实证检验

1. 结构模型的“羊群行为”估计结果

本节将基于中国A股市场的股票交易数据对基本模型进行实证检验,数据来源为RESSET高频数据库。首先,我们使用高频交易数据对基本模型的参数进行估计,其次,我们分别使用“羊群行为的发生频率”“知情交易者发生羊群行为的比例”以及“基于最大似然估计的LSV方法”(ML-LSV方法)对“羊群行为”程度进行度量,并重点比较LSV方法和ML-LSV方法估计结果的差异。

由于模型估计需要使用逐笔的交易数据,但是RESSET高频数据库仅提供平均每6秒累加的分笔数据,这与我们对数据的要求有一定的差距。为了减少数据质量对模型估计的影响,我们选择了交易频率较低的股票进行估计,从而将数据缺陷带来的不利影响降到最低^②。经过筛选,本文最终选用“S中纺机”(股票代码600610)2012年的交易情况作为示例^③。该股票2012年的交易概况如表3所示。从表3可以看出,2012年该股票共有236个交易日,每笔交易的平均时间间隔为27秒,最小的也有12秒,远大于RESSET高频数据库中分笔数据的累加时间,因而该股票可以满足我们对数据结构的要求。

表3 “S中纺机”2012年的交易概况

变 量	N	均值	标准差	最小值	最大值
总交易单数量	236	593	199	172	1182
买单数量	236	281	111	57	701
卖单数量	236	312	114	0	648
交易间隔时间	236	27	10	12	84

此外,参数估计还需要用到“不交易”的信息,但是原始数据中并不包含该信息,因此我们需要手工决定什么时候股票发生了“不交易”。由表3可知,2012年该股票平均每

① 由于拓展模型参数的估计方法和第三部分内容完全类似,限于篇幅原因,在这里不再赘述,具体估计过程可向作者索取。

② 对于交易频率较低的股票而言,每个交易单则更可能是真正的逐笔交易数据。

③ 目前该股票更名为“上海中毅达股份有限公司(中毅达)”。

天的交易单为 593 笔。由于 A 股市场 A 股市场采用竞价交易方式, 每个交易日的 9:15~9:25 为开盘集合竞价时间, 9:30~11:30 以及 13:00~15:00 为连续竞价时间, 对于大多数交易者而言, 只能参与连续竞价, 因此我们将 A 股市场每天股票的交易时间记为 4 小时, 由此可知该股票每笔交易的平均时间为 24 秒。如果两笔交易之间的时间间隔大于 24 秒, 我们则认为这段时间内产生了“不交易”, 相应地需要插入“不交易”的记号, 插入记号的个数是时间间隔 24 秒的整数倍。在对股票交易加入“不交易”的信息后, 我们再次对“S 中纺机”股票 2012 年的交易概况进行了描述, 结果见表 4。从表 4 可以看出, 在插入“不交易”的记号后, 平均每天的总交易单数量从 593 笔上升为 934 笔, 即平均每天“不交易”的单数为 341 笔, 这一数量稍大于平均每天的买单数量 (281 笔) 和卖单数量 (312 笔)。此外, 每天插入“不交易”的最小数量为 31 笔, 最大数量为 505 笔。

表 4 调整之后“S 中纺机”2012 年的交易概况

变 量	N	均值	标准差	最小值	最大值
总交易单数量	236	934	161	249	1370
买单数量	236	281	111	57	704
卖单数量	236	312	114	0	648
不交易	236	341	76	31	505

根据前文介绍的方法, 我们写出模型的似然函数, 并用极大似然方法对其进行估计, 最终得到 2012 年“S 中纺机”股票“羊群行为”的参数估计结果, 如表 5 所示。

表 5 “S 中纺机”“羊群行为”的参数估计结果

参 数	α	δ	μ	τ	ϵ
估计结果	0.598	0.465	0.119	0.907	0.592

从表 5 的估计结果来看, 交易日发生新信息的概率 α 为 0.598, 表明大约有 60% 的概率, 股票的基本价值发生了变化, 而有 40% 的概率股票的基本价值与前一交易日相同。 δ 的估计值为 0.465, 表明在发生新信息情况下, 信息为利好的概率为 46.5%, 利空的概率为 53.5%。 μ 的估计值为 0.119, 说明知情交易者参与交易的概率为 11.9%。 τ 的值为 0.907, 说明新信息中所包含的信息量为 90.7%, 知情交易者的信息质量比较高。此外, 噪声交易者参与交易的概率 ϵ 为 0.592, 说明有 29.6% 的概率噪声交易者会选择买入, 有 29.6% 的概率噪声交易者会选择卖出, 而有 40.8% 的概率其会选择“不交易”。

估计出基本模型的参数后, 我们将使用“羊群行为的发生频率”“知情交易者发生羊群行为的比例”以及“基于最大似然估计的 LSV 方法” (ML-LSV 方法) 分别对“羊群行为”程度进行度量。首先, “羊群行为”的发生频率是指每个交易日买入临界值低于 0.5 或卖出临界值高于 0.5 的概率, 根据该标准, 我们估计了“S 中纺机”的“羊群行为”程度, 结果见表 6。从表 6 中可以看出, 在 236 个交易日中, 有 34.3% 的交易日中买入临界值小于 0.5 的频率大于 10%, 有 22.5% 的交易日中买入临界值小于 0.5 的频率大于 30%, 有 16.9% 的交易日中买入临界值小于 0.5 的频率大于 50%, 表明在 10%、30% 和 50% 的概率水平下, 分别有 81 天、53 天和 40 天出现了买入“羊群行为”; 卖出“羊群行为”中, 有 56.4% 的交易日中卖出临界值大于 0.5 的频率大于 10%, 有 47.9% 的交易日中卖出临界值大于 0.5 的

频率大于 30%，有 35.6% 的交易日中卖出临界值大于 0.5 的频率大于 50%，表明在 10%、30% 和 50% 的概率水平下，分别有 133 天、113 天和 84 天出现了卖出“羊群行为”。

表 6 “S 中纺机”“羊群行为”的发生频率 (单位: %)

	>10%	>30%	>50%
Herd-Buy	34.3	22.5	16.9
Herd-Sell	56.4	47.9	35.6

接下来我们使用“知情交易者发生羊群行为的比例”来估计“羊群行为”程度，估计结果见表 7。从表 7 中可以看出，在 236 个交易日中，有 121 个交易日发生了新信息，在发生新信息的情况下，发生买入“羊群行为”的知情交易者比例平均为 3.9%，最大值为 84.2%，发生卖出“羊群行为”的知情交易者比例平均为 6.7%，最大值为 48.5%。

表 7 “S 中纺机”发生“羊群行为”的知情交易者比例 (单位: %)

	交易日	均值	标准差	最大值
Herd-Buyers	121	3.9	12.1	84.2
Herd-Sellers	121	6.7	10.3	48.5

最后，我们“基于极大似然估计的 LSV 方法”(ML-LSV 方法)对股票的“羊群行为”进行度量，并将其与传统的 LSV 方法进行比较。由于 ML-LSV 方法认为只有当交易日当天有新信息产生时，才会发生“羊群行为”，而传统的 LSV 方法并没有考虑这一点。为了使估计结果具有可比性，我们将两种方法的样本进行了统一，都只保留了发生新信息的交易日样本，估计结果见表 8。表 8 显示，对于产生新信息的交易日样本，LSV 方法得到的“羊群行为”平均值为 7.31%，该数值大于使用 ML-LSV 方法得到的 4.97%。如果将买入“羊群行为”和卖出“羊群行为”分开来看，该结论仍然成立。这意味着，ML-LSV 方法较好地剔除了公共信息所引起的“伪羊群行为”，本文的结构模型可以对“羊群行为”程度有更为精确的度量。

表 8 LSV 和 ML-LSV 方法所估计的“羊群行为” (单位: %)

	均值	中位数	标准差	最小值	最大值
Panel A LSV 方法的“羊群行为”					
Herd-Buy	7.31	6.59	5.22	0.18	25.43
Herd-Buy	8.13	7.18	6.82	0.38	25.43
Herd-Sell	6.86	6.13	4.08	0.18	16.34
Panel B ML-LSV 方法的“羊群行为”					
Herd-Buy	4.97	2.72	7.00	0.00	44.22
Herd-Buy	6.68	2.34	10.87	0.00	44.22
Herd-Sell	4.13	2.94	3.74	0.01	14.39

2. 考虑“交易单规模”后的“羊群行为”估计结果

在本节，我们将对基本模型和加入“交易单规模”后的拓展模型进行比较。为了显示“交易单规模”对“羊群行为”测度的影响程度，本文遵循 Easley 等 (1997) 以及 Cipriani 和 Guarino (2014) 的样本选择 (纽约证券交易所 Ashland 股票 1995 年的交易数据)，从而使估计结果具有可比性，本节的数据来源为 TAQ 数据库。

基本模型和拓展模型的参数估计结果如表 9 所示。表 9 显示,在考虑了交易规模后,原有 5 个参数的估计结果并没有产生太大变化。而知情交易者交易大额单的概率为 0.412,噪声交易者交易大额单的概率为 0.273,说明知情交易者更有可能参与大额单的交易。

表 9 基本模型和拓展模型的参数估计结果

参数	基本模型	拓展模型
α	0.306	0.311
δ	0.485	0.400
μ	0.459	0.430
τ	0.585	0.577
ϵ	0.564	0.561
ω	—	0.412
φ	—	0.273
Log-likelihood	-40048.78	-54098.67

表 10 展示了使用基本模型和拓展模型估计的“羊群行为”发生频率。我们发现,两个模型的结果没有太大差别,说明交易规模并没有对买入临界值和卖出临界值产生较大影响。

表 10 “羊群行为”发生频率 (单位: %)

临界值	>10%	>30%	>50%
Panel A 基本模型			
Herd-Buy	72.2	43.3	20.6
Herd-Sell	80.6	53.2	29.8
Panel B 拓展模型			
Herd-Buy	69.8	41.7	22.6
Herd-Sell	76.2	54.0	31.3

表 11 为使用基本模型和拓展模型估计的发生“羊群行为”的知情交易者比例,我们发现交易单规模对发生“羊群行为”的知情投资者比例产生了较大影响。基本模型中发生买入“羊群行为”的投资者比例的均值为 1.71%、中位数为 0.49%,发生卖出“羊群行为”的投资者比例均值为 4.36%、中位数为 1.43%,而拓展模型中发生买入“羊群行为”的投资者比例均值为 2.97%、中位数为 0.84%,发生卖出“羊群行为”的投资者比例均值为 5.04%、中位数为 3.11%。该结果表明,在考虑交易单规模后,“羊群行为”变得更加显著。由于知情交易者参与大单买卖的概率要高于噪声交易者,并且大额交易单会强化未来投资者对股票价值的判断,因此本文的拓展模型不仅能捕捉到由一连串交易单所引起的“羊群行为”,还能额外捕捉到由单个大额交易单所引起的“羊群行为”,从而使得拓展模型下的估计结果要大于基础模型。

表 11 发生“羊群行为”的知情交易者比例 (单位: %)

	均值	中位数	标准差	最小值	最大值
Panel A 基本模型					
Herd-Buyers	1.71	0.49	2.49	0.00	10.47
Herd-Sellers	4.36	1.43	7.87	0.00	43.02

(续)

	均值	中位数	标准差	最小值	最大值
Panel B 拓展模型					
Herd-Buyers	2.97	0.84	4.94	0.00	22.93
Herd-Sellers	5.04	3.11	7.44	0.00	37.83

综上所述,实证检验中,我们使用日内高频交易数据,而非季度低频数据,对模型参数进行了估计,从而更好地刻画了“羊群行为”的“短期脆弱性”特征。基本模型下,我们基于中国A股市场股票交易数据,分别使用三种测度对“羊群行为”程度进行了测量,“羊群行为的发生频率”测度下,我们发现,在10%、30%和50%的概率水平下,分别有81天、53天,以及40天会出现买入“羊群行为”,有133天、113天、84天会出现卖出“羊群行为”。在“知情交易者发生羊群行为的比例”测度下,我们发现,新信息产生后,发生买入“羊群行为”的知情交易者比例平均为3.9%,发生卖出“羊群行为”的知情交易者比例平均为6.7%。在“ML-LSV方法”测度下,“羊群行为”、买入“羊群行为”和卖出“羊群行为”的平均值分别为4.97%、6.68%和4.13%,该结果显著小于由LSV方法得到的“羊群行为”估计结果(分别为7.31%、8.13%和6.86%),这表明ML-LSV的估计方法能较为准确地区分“伪羊群行为”和“基于信息的羊群行为”,更好地度量真实的“羊群行为”大小。

基于纽约证券交易所股票的交易数据,我们对基本模型和拓展模型进行了比较,比较结果显示,在考虑交易单规模后,虽然发生买入或卖出“羊群行为”的临界值并未发生明显变化,但知情交易者发生“羊群行为”的比例会显著提高,具体而言,基本模型中发生买入或卖出“羊群行为”的投资者比例分别为1.71%和4.36%,而拓展模型中发生买入或卖出“羊群行为”的投资者比例分别为2.97%和5.04%。由于本文的拓展模型不仅能够捕捉到由一系列交易单所引起的“羊群行为”,还能额外捕捉到由单个大额交易单所引起的“羊群行为”,显然上述结果符合预期。

五、结 论

本文在Easley等(1997)、Cipriani和Guarino(2014)的研究基础上,构建了一个度量“羊群行为”程度的结构模型,并使用金融市场高频交易数据对“羊群行为”进行了实证检验。结构模型中,新信息产生的不确定性以及不对称信息会导致“羊群行为”。因此,该模型能够较好地地区分由公共信息所引起的“伪羊群行为”和由信息不确定性所导致的“基于信息的羊群行为”,并对真实的“羊群行为”程度做更为精确的度量。

值得一提的是,本文创新性地在基本模型之上引入“交易单规模”,提出了拓展的“羊群行为”测度模型。在考虑“交易单规模”之后,本文的结构模型不仅能捕捉由一系列交易单所引起的“羊群行为”,还能额外捕捉由单个大额交易单所引起的“羊群行为”,因而提高了模型的精度和准确度,这也是本文对“羊群行为”研究的重要贡献。

实证研究中,我们使用A股市场和纽约证券交易市场的高频交易数据对本文的结构模型进行了实证检验,我们发现,结构模型能够较好地刻画“羊群行为”的“短期脆弱性”特征。基本模型中,在10%的概率水平下,一年中有81个交易日投资者会发生买入“羊群行为”,有133个交易日投资者会发生卖出“羊群行为”;在新信息产生的交易日中,发生买入或卖出“羊群行为”的知情交易者比例分别为3.9%和6.7%。此外,使用本文提出的ML-LSV

方法估计的“羊群行为”程度为 4.97%，显著低于传统的 LSV 方法所得到的 7.31%，表明 ML-LSV 方法能较好地区分“伪羊群行为”和“基于信息的羊群行为”，纠正传统 LSV 方法下对“羊群行为”高估的误差。

在对基本模型和拓展模型的比较中，我们发现，知情交易者交易大额单的概率为 0.412，显著高于噪声交易者交易大额单的概率 0.273。同时，由于大额交易会强化未来投资者对股票价值的判断，因此拓展模型估计出来的“羊群行为”要大于基础模型，两类模型对买入“羊群行为”中知情交易者的比例估计分别为 4.36% 和 1.71%，对卖出“羊群行为”中知情交易者的比例估计分别为 5.04% 和 2.97%。

总体而言，本文提出一个高精度的“羊群行为”测度模型，并使用金融市场交易数据对“羊群行为”进行了更准确的估算。在未来的研究中，学者们可以将本文的方法拓展到更多的股票和市场情形之上，如研究不同特征的股票（如大股票和小股票、价值股和增长股等）是否具有不同的“羊群行为”程度，不同的市场环境（如牛市、熊市、震荡市等）是否会对投资者的“羊群行为”产生影响等。此外，研究者们还可以将“羊群行为”拓展到信息经济学和资产定价领域当中，如量化“羊群行为”对市场信息有效性所造成的影响，探究“羊群行为”能否作为一个风险因子被金融市场定价等。

参 考 文 献

- [1] Avery C., Zemsky P., 1998, *Multidimensional Uncertainty and Herd Behavior in Financial Markets* [J], *American Economic Review*, 88 (4), 724~748.
- [2] Bikhchandani S., Hirshleifer D., Welch I., 1992, *A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades* [J], *Journal of Political Economy*, 100 (5), 992~1026.
- [3] Bikhchandani S., Sharma S., 2000, *Herd Behavior in Financial Markets* [J], *IMF Staff Papers*, 47 (3), 279~310.
- [4] Christoffersen S., Tang Y., 2010, *Institutional Herding and Information Cascades: Evidence from Daily Trades* [R], NFA Working Paper, August 30, 2009.
- [5] Cipriani M., Guarino A., 2014, *Estimating a Structural Model of Herd Behavior in Financial Markets* [J], *American Economic Review*, 104 (1), 224~251.
- [6] Da Z., Huang X., 2016, *Harnessing the Wisdom of Crowds* [DB/OL], https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2731884.
- [7] Easley D., Kiefer N. M., O'Hara M., 1997, *One Day in the Life of a Very Common Stock* [J], *Review of Financial Studies*, 10 (3), 805~835.
- [8] Easley D., O'Hara M., 1987, *Price, Trade Size, and Information in Securities Markets* [J], *Journal of Financial Economics*, 19 (1), 69~90.
- [9] Easley D., O'Hara M., 1992, *Time and the Process of Security Price Adjustment* [J], *Journal of Finance*, 47 (2), 577~605.
- [10] Fama E. F., 1970, *Efficient Capital Markets: A Review of Theory and Empirical Work* [J], *Journal of Finance*, 25 (2), 383~417.
- [11] Glosten L., Milgrom P., 1985, *Bid, Ask, And Transaction Prices in A Specialist Market with Heterogeneously Informed Traders* [J], *Journal of Financial Economics*, 14 (1), 71~100.
- [12] Grinblatt M., Titman S., Wermers R., 1995, *Momentum Investment Strategies, Portfolio Performance, and Herding: A Study of Mutual Fund Behavior* [J], *American Economic Review*, 85 (5), 1088~1105.
- [13] Hayek F. A., 1945, *The Use of Knowledge In Society* [J], *American Economic Review*, 35 (4),

519~530.

[14] Lakonishok J., Shleifer A., Vishny R. W., 1992, *The Impact of Institutional Trading on Stock Prices* [J], Journal of Financial Economics, 32 (1), 23~43.

[15] Wermers R., 1999, *Mutual Fund Herding and the Impact on Stock Prices* [J], Journal of Finance, 54 (2), 581~622.

[16] 董志勇、韩旭:《基于 GCAPM 的“羊群行为”检测方法与中国股市中的实证依据》[J],《金融研究》2007 年第 5 期。

[17] 孙培源、施东晖:《基于 CAPM 的中国股市“羊群行为”研究——兼与宋军、吴冲锋先生商榷》[J],《经济研究》2002 年第 2 期。

[18] 许年行、于上尧、伊志宏:《机构投资者“羊群行为”与股价崩盘风险》[J],《管理世界》2013 年第 7 期。

A Measurement Model of Herd Behaviour Based on High Frequency Data

Zhu Feifei¹ Tang Ya¹ Xu Jianguo^{2,3} Li Hongtai⁴

(1. Guanghua School of Management, Peking University;

2. National School of Development, Peking University;

3. Institute of Digital Finance, Peking University;

4. Zhangzhou Municipal People's Government Office)

Research Objectives: Estimating the herding behavior using structural model and empirical analysis. **Research Methods:** Construct a structural model of herding behavior, estimate model's parameters using recursive method and maximum likelihood estimation. Besides, test the model empirically by using the high frequency data of China and American stock market. **Research Findings:** The extend structural model not only can capture the short-lived and fragile characteristics of herding, but also distinguish “pseudo herding” and “informational herding”. Under the new model, the probability of “large orders by informed traders” is much higher than the probability of “larger orders by noise traders”, and more informed traders are participated in herding. **Research Innovations:** By taking “the order of transaction” and “transaction size” into consideration, we can capture herding behavior more accurately. **Research Value:** Accurately identify and measure herding behavior and improve the price discovery function of the stock market.

Key Words: Structural Model; High Frequency Data; Order of Transactions; Transaction Size

JEL Classification: C51; D82; G12

(责任编辑:王喜峰)