

A股市场交易毒性初探

证券分析师: 秦国文

电话: 0755-82130875

E-mail: qingw@guosen.com.cn

证券投资咨询执业资格证书编码: S0980511080001

联系人: 张璐楠

电话: 0755-82133115

E-mail: zhanglun@guosen.com.cn

证券分析师: 戴军

电话: 0755-82133136

E-mail: daijun@guosen.com.cn

证券投资咨询执业资格证书编码: S0980510120064

背景和摘要

交易毒性(VPIN)是指市场上知情交易者会利用不知情的交易者进行操作,当交易指令流逆向选择做市商,而且做市商没有意识到他们正在承受亏损的情况下提供流动性时,交易指令流被视为有毒。指令到达的过程一般能够为后续的价格移动提供信息,特别是能够给交易毒性的测算和识别提供信息。但是,由于高频市场结构中的交易原理,从高频指令流中提炼出这个信息非常复杂。大量研究证明,通过成交量捕捉的交易时间,比时钟时间能够传达出更多信息。

VPIN 对 PIN 的修正

在国外市场常用的知情交易概率(PIN)估算方法对无法观测的参数需要使用极大似然估计法拟合出每日股票买入卖出指令的三个泊松分布的混合体。最新的VPIN算法对此作了修正,对知情和不知情交易者的交易指令进行建模,基于VPIN的估计方法不需要对不可观测的参数进行估计,而且可以在动态时间更新。这个估计交易毒性的新方法还克服了在大成交量的市场估计传统PIN模型的难题,并且提供了一种建议处理高频数据来测量交易指令流毒性的方法。

构建成交量篮子估计 VPIN

计算VPIN是基于成交量划分的时间而非时钟时间。实现基于成交量的取样方法是将交易序列以均匀成交量分组,称为一个“成交量篮子”。一个成交量篮子是成交量加总的一组交易。完成篮子的最后一笔交易如比所需要的大,超过的成交量将被划分到下一个篮子,按照这样的成交量篮子取样后我们获取到一个观测值。构建成交量篮子时我们对每一笔买入和卖出成交量进行分类,成交量的方向与交易指令单毒性存在潜在联系,结合总成交量水平和成交量的方向可以指示出存在新信息的可能性。因此,买入成交量占多数说明毒性来自好的信息,反之亦然。我们通过观测买单和卖单之间交易的强度和不平衡性来估计出VPIN。

VPIN 的实证检验

VPIN估计值越大,说明市场毒性越高,知情交易者占的比例越高。另外,VPIN更新的速度越快,也说明市场上的知情信息越多。在实证检验中,通过估计50ETF的VPIN,展示了其对应的CDF函数在2009年7月29日崩盘前提前1小时预警提示大盘毒性极高,告诉投资者市场面临大幅波动的概率迅速增大。我们还展示了VPIN对于波动性有较为明显的预测能力,并且它对未来的高绝对收益提供了指示作用。

VPIN 的应用

因为上述的性质,我们认为VPIN可以作为一种高频交易领域的风险管理工具。通过交易指令估计交易毒性,并可通过高的VPIN值提示后续价格将有较高的波动,做市商可以通过VPIN调整做市价差。流动性风险不仅对做市商而言至关重要,对交易者来说,流动性可能会导致价格潜在的大幅波动,适时通过VPIN控制交易执行风险也会是一个不错的选择。后续我们将对不同的交易品种进行VPIN估计的系统性测算,同时进一步挖掘VPIN与标的资产波动性的潜在关系。

独立性声明:

作者保证报告所采用的数据均来自合规渠道,分析逻辑基于本人的职业理解,通过合理判断并得出结论,力求客观、公正,结论不受任何第三方的授意、影响,特此声明。

内容目录

背景	3
研究摘要	3
模型说明	4
VPIN 与参数估计	7
信息和时间在 VPIN 估计中的概念	7
成交量划分	8
VPIN 知情交易计量法	8
买入成交量和卖出成交量分类	10
VPIN 计量法的实证检验	10
VPIN 是 PIN 的近似值	10
VPIN 估计值的实证效果及其稳定性	11
毒性和交易强度	13
对 ETF 使用 VPIN 计量法	14
VPIN 与波动性	15
VPIN 与绝对收益的相关系数曲面	15
VPIN 水平对应标的资产的绝对收益	17
结论	18

图表目录

图 1: 在不确定事件下的交易序列	5
图 2: 8 月 25 日平均成交量档案	7
图 3: 50ETF 的 VPIN 估计与 VPIN 更新时间	9
图 4(a): 1 分钟频率取样的 50ETF 价格与 VPIN 估计值	12
图 4(b): 30 秒频率取样的 50ETF 价格与 VPIN 估计值	12
图 4(c): 10 秒频率取样的 50ETF 价格与 VPIN 估计值	13
图 5: 50ETF 价格走势、VPIN 估计值与 1 分钟成交量	15
图 6: VPIN 和 50ETF 绝对收益相关性的参数敏感度分析	16
图 7: VPIN 水平对应的 50ETF 绝对收益	17

背景

高频交易(High Frequency Trading)占到美国股票市场成交量的 70%，并且正在快速地接近期货市场成交量的 50%。这些高频交易公司作为做市商，通过在不同层次的电子指令簿(order book)中下单（被动指令单），为头寸接受者提供流动性。被动交易指令(passive order)被定义为不立即与市场已有报单成交(cross the market)的指令，即被动指令的发出者不能控制指令执行的时间。高频做市商通常不进行方向性的下注，而是努力从大量的交易中赢得买卖报价价差中的薄利。他们的这种盈利能力取决于对头寸风险的控制，最主要的就是他们执行被动指令单时控制逆向选择的能力。

市场投资者把“被动指令单在应该被缓慢执行时快速执行，或者应该快速执行时缓慢（或者完全不）执行”的这一市场现象定义为“交易指令的逆向选择” (Jeria 和 Sofianos, 2008)。而这个直观的现象和市场的微观结构模型里的理论是一致的 (Glosten and Milgrom [1985], Kyle [1985], and Easley and O'Hara [1987, 1992b])，即知情交易者会利用不知情的交易者。当交易指令流逆向选择做市商，而且做市商没有意识到他们正在承受亏损的情况下提供流动性时，交易指令流被视为有毒的，也就是我们定义的交易毒性。

国外对如何测量高频领域的指令流毒性发展了一种新的方法，像 Easley, Engle, O'Hara and Wu (2008) 讨论的那样，微观经济学中的一个基本观点是：指令到达过程一般能够为后续的价格移动提供信息，特别是能够给交易毒性的测算和识别提供信息。但是，由于高频市场结构中的交易原理，从高频指令流中提炼出这个信息是非常复杂的。大量研究证明，**通过成交量捕捉的交易时间，比时钟时间能够传达出更多信息**。这里的信息指的是能够在一个相对很短的时间中导致不平衡交易的潜在事件。信息事件的产生有可能来自不同的原因，有些是关于资产的回报，还可能与市场系统风险，或者某些基于高频交易的投资组合有关。受信息事件影响的一个典型应用例子是期货合约，信息可能直接造成市场系统性因素的变化，从而影响期货在对冲策略或投资组合中的应用。

研究摘要

本报告借鉴国外相关文献中提供的一个新的方法，基于市场成交量变化的过程，直接估算出指令流毒性，而我们把这种指令流毒性命名为**等成交量下的知情交易概率 (VPIN, Volume-Synchronized Probability of Informed Trading)**。原先的知情交易概率 (PIN) 估算方法 (Easley 等 1996) 对无法观测的参数需要使用极大似然估计法拟合出每日股票买入卖出指令的泊松分布。最新的这个静态估算 VPIN 的方法被 Easley 等 (2008) 用 GARCH 方法进行了扩展，对知情和不知情交易者的交易指令进行建模。这篇文章中基于 VPIN 的方法不需要对不可观测的参数进行估计，而且 VPIN 是在动态时间更新的，这个动态时间被定义为拥有相同成交量的每一个交易区间。因此，这个估计交易毒性的方法克服了在大成交量的市场估计传统 PIN 模型的难题，并且提供了一种建议处理高频数据来测量交易指令流毒性的方法。（传统 PIN 模型在此不做展开，可以参考公开的国外文献）

本报告在回顾国外学者对 VPIN 模型的最新结论的基础上，为 VPIN 方法的统计学属性提供实证证据，我们展示了用成交量划分的方法（时间周期已经选定，所以每一个都有相同的成交量）能有效降低估算交易毒性中所用样本的波动性。因为大的价格移动都伴随着大的成交量，我们可以把按成交量的采样看作是按波动性采样的

代理。国外学者通过蒙特卡洛模拟展示出了这种方法能够对所有参数组合提供 VPIN 的准确估计，即模型对仅有的参数并不敏感。

我们展示了用 VPIN 估计股指期货和 ETF 基金的交易毒性，并通过实证结果发现交易毒性的快速上升以及观察其相对水平 (VPIN 的 CDF) 来识别出市场大跌前的市场环境。最后，我们还展示 VPIN 与资产的波动性存在重要的联系。

由于毒性对所有做市商都是有害的，高的 VPIN 值预示了高的波动性。在国外做市商盛行的市场里，VPIN 可以预测短期波动性，尤其是 VPIN 与大的价格移动相联系时。而国内的 ETF 也存在高 VPIN 伴随高波动性的现象，但并无美国市场那么明显。

对交易流毒性的估计有大量的直接应用，如我们在实证检验中使用 VPIN 作为实时的风险管理工具，展示了被 VPIN 捕捉到的交易流在 2009 年 7 月 29 日大盘在最后 1 小时曾一度暴跌 7.64% 的前数小时里面变得更加具有毒性，这一毒性导致许多交易者纷纷退出市场，造成大盘随后相继的大幅调整。跟踪 VPIN 可以让做市商控制风险并在波动的市场里把握主动权。监管者可以利用 VPIN 来监测流动性条款，积极地约束交易和实施市场管理。在高频领域，有效的管理需要在事前实施，需要在问题导致崩盘之前提前预见。监测 VPIN 可以指示在某一时间当前的流动性条款是否存在风险，并实施市场熔断以及其它一些防止崩盘的措施。交易员可以利用基于 VPIN 的方法设计控制执行风险的算法。市场微观结构模型指出日内周期性可以反映知情与不知情交易者的不同参与率 (如 Admati 等 1988)。根据 VPIN 设计的算法推迟或提早交易可以减少在劣势下执行交易单的风险。

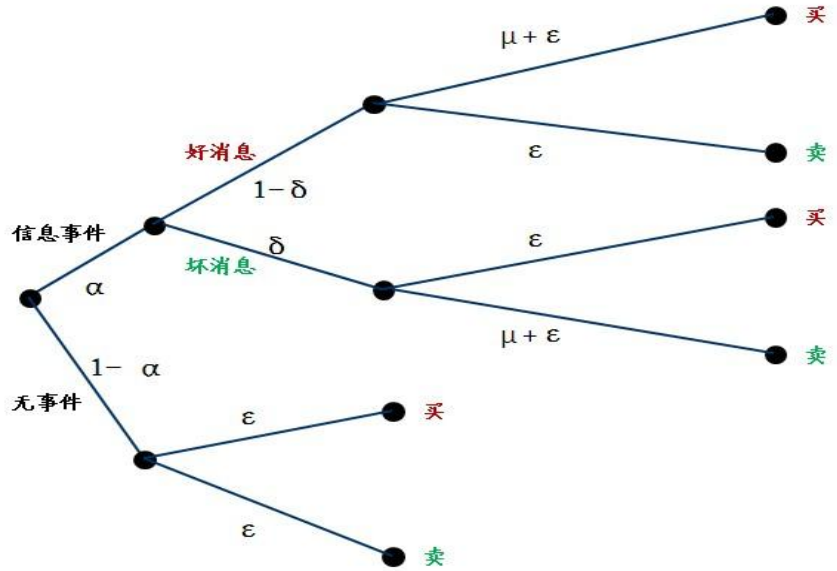
本报告第二部分讨论模型的理论框架并且展示 PIN 如何影响买卖价差。在第三部分提出我们估计 VPIN 的方法。在第四部分通过实证检验，评估 VPIN 的准确性并讨论 VPIN 估计的高毒性如何对市场进行风险揭示。在第五部分讨论 VPIN 与标的资产波动性的相关性。第六部分总结我们的研究成果。VPIN 的计算代码及样本数据的获取详情请联系相关分析师。

模型说明

在报告的这一部分，我们首先描述使用指令流数据推测指令流毒性的基本模型。我们用一个标准的微观结构模型 PIN 来推导衡量指令流毒性的方法，然后展示如何改进 PIN 来更好地适用于现代市场。

Easley 等人在一系列的文章中展示了如何使用微观结构模型利用交易数据对个股估计知情交易概率 PIN。微观结构模型将交易视为流动性提供者和交易者 (头寸接受者) 在交易时段的重复博弈。在每阶段的开始，自然选择信息事件是否出现。这些事件独立地以概率 α 出现。如果信息是好消息，那么知情的交易者就知道在交易阶段结束时，资产价值将会是 \bar{S}_i ；如果信息是坏消息，资产价值将会是 \underline{S}_i ， $\bar{S}_i > \underline{S}_i$ 。好消息的发生概率是 $1-\delta$ ，坏消息的发生概率是 δ 。在信息事件发生或不发生后，阶段 i 交易者的到达服从泊松过程。如果一个阶段有信息事件，知情交易者的交易指令以速率 μ 到达。这些知情者如果已经看到了好消息就买入，如果已经看到了坏消息就卖出。每阶段，非知情的买家和卖家均以 ϵ 的速率到达。

图 1: 在不确定事件下的交易序列



资料来源: 国信证券经济研究所整理

图 1 的结构模型让我们能够将可观测的市场结果（买、卖）与无法观测的信息和交易背后的指令流产生机制联系起来。之前的文章着重通过极大似然法估计来判断这些过程。模型直观地解释了不知情交易买卖在一般情况下的流程，并且用数据指出了不知情指令流的比率 ϵ 。异常买卖成交量被解释为基于信息的交易，并且用来计算 μ 。有异常买卖时间段数量被用来确定 α 和 δ 。流动性提供者用他对这些参数的了解决定他想做多的价格，即“买盘价”(Ask)，以及他想做空的价格，“卖盘价”(Bid)。这些价格是不同的，所以有“买卖盘价差”，因为流动性提供者不知道交易对手是否具备信息。这个价差是交易者从流动性提供商处买入资产的条件期望价值与交易者从流动性提供商处卖出资产的条件期望价值的差值。这些条件期望在不同时间是在变化的，因为市场中存在更知情的交易者的逆向选择问题导致流动性提供商需要不定期的调整买卖价差。

随着交易进行，流动性提供者观察交易，并使用贝叶斯规则建模以更新他们对交易流毒性的看法，这在我们的模型中被描述为参数估计。设

$$P(t) = (P_n(t), P_b(t), P_g(t))$$

是流动性提供者在 t 时对没有消息 (n)，坏消息 (b) 和好消息 (g) 的看法。流动性提供者在 $t=0$ 时相信:

$$P(0) = (1-\alpha, \alpha\delta, \alpha(1-\delta))$$

为了决定在 t 时的卖出价和买入价，流动性提供商在相关类型指令单到达时条件地更新他的看法。T 时间资产的以历史交易为条件的期望价值是:

$$E[S_i | t] = P_n(t)S_i^* + P_b(t)\underline{S}_i + P_g(t)\bar{S}_i$$

其中

$$S_i^* = \delta\underline{S}_i + (1-\delta)\bar{S}_i$$

是资产之前预期的价值。买入询价是别人想卖给流动性提供商的期望价值，所以是:

$$B(t) = E[S_i | t] - \frac{\mu P_b(t)}{\varepsilon + \mu P_b(t)} [E[S_i | t] - \underline{S}_i]$$

类似的，卖出询价是别人想从流动性提供商处买入资产的期望价值，所以是：

$$A(t) = E[S_i | t] - \frac{\mu P_g(t)}{\varepsilon + \mu P_g(t)} [\bar{S}_i - E[S_i | t]]$$

上面的两个方程展示了知情和不知情交易者到达对报价的影响。如果没有知情交易者 $\mu=0$ ，那么交易者不携带信息，因此买卖询价都等于之前的期望价格。或者说，如果没有不知情交易者 $\varepsilon=0$ ，买入询价是最小的价格和卖出询价是最大的价格。在这些价格的情形，没有知情交易者会交易，市场也将会关闭。通常，知情不知情的交易者都会在市场中出现，所以买入询价会低于 $E[S_i | t]$ ，而卖出询价高于 $E[S_i | t]$ 。

在 t 时刻买入卖出询价的价差记为 $\Sigma(t) = A(t) - B(t)$ 。计算显示这个价差是，

$$\Sigma(t) = \frac{\mu P_g(t)}{\varepsilon + \mu P_g(t)} [\bar{S}_i - E[S_i | t]] + \frac{\mu P_b(t)}{\varepsilon + \mu P_b(t)} [E[S_i | t] - \underline{S}_i]$$

价差公式中的第一项为一个买入是知情的概率乘以输给知情买家的期望损失，第二项是卖出的对称。对于好坏事件等可能出现时，阶段初始的价差有一个非常简单的形式：

$$\Sigma = \frac{\alpha\mu}{\alpha\mu + 2\varepsilon} [\bar{S}_i - \underline{S}_i]$$

这个模型的关键点是初始指令来自知情交易者的概率，知情交易的概率记作 PIN。可以直观地看出一段时间的第一笔交易“是知情的”的概率由下式给出：

$$PIN = \frac{\alpha\mu}{\alpha\mu + 2\varepsilon}$$

其中，分母是所有指令单的发生率，分子是基于信息的指令单的发生率。PIN 是知情指令单与全体指令的比例。价差方程显示 PIN 是价差的决定因素。

这些方程描述了以下事实：流动性提供者需要正确地估计他们的 PIN 以发现进入市场的最优程度。在 PIN 的非预期增长后，不调节价格的流动性提供商将会产生损失。

VPIN 与参数估计

计算 PIN 模型的标准方法是使用极大似然法估计出驱动交易的不可观测的参数 (α , μ , δ , ϵ)，然后根据这些参数的估计值推断 PIN。国外学者在今年初的相关研究中提出另一种可以直接估计 PIN 的方法，该方法不需要对不可观测的参数进行中间估计。同时，它还能够随机时间更新（传统估算 PIN 的方法并非随机更新，而是通过历史数据定期用 MLE 做参数优化），与市场上新信息的到达速率相匹配。这种基于成交量的方法，我们记为 VPIN，对高频环境下的交易指令流毒性提供了一种简单的度量。首先，我们从高频交易中信息和时间角色的讨论开始。

信息和时间在 VPIN 估计中的概念

标准交易序列模型中的“信息”通常被认为是资产未来价值有指示作用的数据。因此，在股票市场中，可以很自然地认为信息是资产的未来回报，而这个信息与公司的前景有关，例如公司产品的市场环境和销售情况等。在有效市场中，由于知情交易者利用他们所掌握的信息通过交易获利，资产的价值应当趋于完全信息公开化后的资产价值。因为做市商可以做多或者做空股票，资产价值的未来移动将会影响他们的获利能力，因此他们尝试着从交易类型中找出一切潜在的新信息。他们即时掌握的观点反映到了他们的买卖价差上。

在高频领域，尽管做市商的运营行为是在不断改变着市场，他们也面临着关于买卖价差的问题，一个预期仅持有股票数分钟的高频做市商也被这段时间里影响价值的信息所影响。这些信息有些与潜在资产的基本面有关，但是也会在一个特殊区间的流动性需求中反映出来。例如，在期货合约中，一个信息的出现增加了期货合约用作对冲投资组合的需求，也会影响期货当时的价格，并且与做市商是相关的。这个对信息的广义定义意味着在一天之中信息事件将会频繁出现，对于期货价格变化，它们的重要性也会不同。无论如何，它们的存在同样可以被交易的性质与时间标识出来。

高频交易模型最重要的一点是交易并不是在时间上平均分布的。交易以不规则的频率到达，有些交易比其它交易更加重要因为它们能够揭示不同数量的信息。

例如图 2 所示，交易上证 50ETF 和 180ETF 展示了不同的日内周期性，50ETF 成交量在开盘后一路。新信息的到达导致了一波一波的投资决定，这将转化为成交量的猛增。不同产品的信息在不同的时间到达，因此产生了不同的日内成交量周期性。

图 2: 8 月 25 日平均成交量档案



资料来源: 天软资讯, 国信证券经济研究所整理

Easley 和 O'Hara (1992b) 提出一种想法，交易成交的时间间隔与新信息存在相关性，这为后面的研究中不使用“时钟时间”而是“交易时间”提供了一个基础。在本报告的研究中，我们使用成交量时间的方法进行研究，而不是时钟时间。一个较为直观的结论是，一条信息相关系数越高，它引起的成交量越高。我们对于每次市场交易某一固定的成交量后进行采样，就可以模拟相关性可比的市场消息的到达。如果一条消息产生的成交量是另一条消息的两倍，我们就能得到两倍数量的观测结果，从而将前一条信息在样本中的权重翻倍。

成交量划分

为了实现基于成交量的取样，我们将交易序列以均匀成交量分组，称为一个“成交量篮子”，大小定义为一个外生变量 V 。一个成交量篮子是成交量加总为 V 的一组交易。如果完成篮子的最后一笔交易比所需要的大，超过的成交量将被划分到下一个篮子，按照这样的成交量篮子取样后，在每一个篮子完成后我们获取到一个观测值。这让我们能够将交易过程划分成区间，在区间上的交易失衡对流动性提供者有明确的指导意义。

设 $t=1$ 到 n 是成交量篮子的下标。我们将每个篮子的交易分类成买 V_t^B 和卖 V_t^S 。对于每个 t ， $V = V_t^B + V_t^S$ 。

VPIN 知情交易计量法

标准的 PIN 模型仅根据买卖成交的数量来推测潜在的信息结构，成交量在模型中没有明显的角色。在我们分析的高频领域，交易数量是有问题的，很多时候看到的交易数量都是经过合成得出的。回到 PIN 的理论基础，我们真正想要的信息是知情和非知情交易者的交易目的。交易目的和交易数据间的联系是充满噪音的，因为交易目的可以被切分成很多块从而对市场的影响最小化。一个指令单也许最终被执行为多个交易指令，知情交易可以以不同的形式出现。基于这些原因，我们将每一个交易视为单位大小的加总。所以，在价格 p 的五个交易合约被视为五个以价格 p 交易的单位合约。这个惯例将交易密集程度融入分析中。后面，我们探讨如何将这些交易与理论联系起来。

根据 Easley 等的研究，预期的交易失衡 (trade imbalance) 是 $E[|V_t^S - V_t^B|] \approx \alpha\mu$ ，所有交易的预期到达速率是

$$\begin{aligned} \frac{1}{n} \sum_{t=1}^n (V_t^B + V_t^S) &= V = \underbrace{\alpha(1-\delta)(\varepsilon + \mu + \varepsilon)}_{\text{volume from up event}} + \underbrace{\alpha\delta(\mu + \varepsilon + \varepsilon)}_{\text{volume from down event}} + \underbrace{(1-\alpha)(\varepsilon + \varepsilon)}_{\text{volume from no event}} \\ &= \alpha\mu + 2\varepsilon \end{aligned}$$

成交量划分使我们能非常简单地估计。回忆我们之前将交易日划分成成交量相等的篮子。就是说， $V_t^B + V_t^S$ 是一个定值，对于所有的 t 都等于 V 。

通过上面计算的数值，我们可以推导 VPIN:

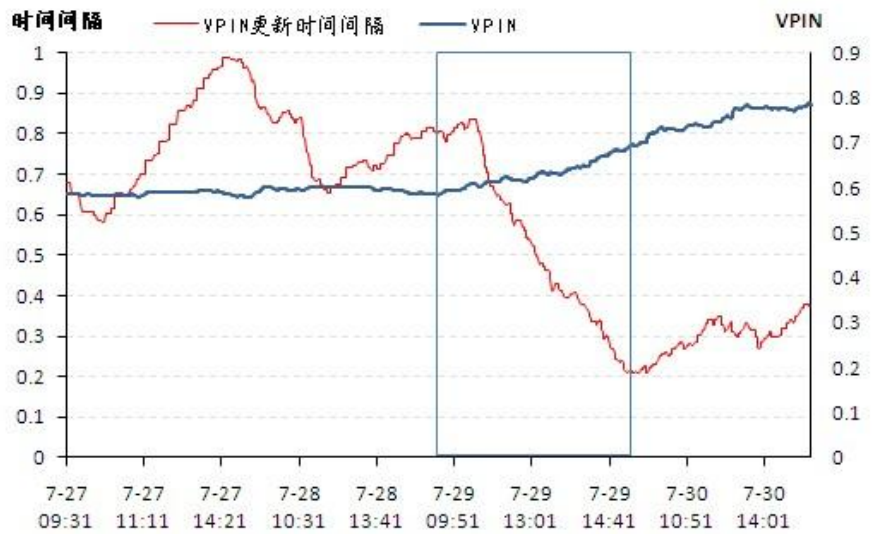
$$VPIN = \frac{\alpha\mu}{\alpha\mu + 2\varepsilon} = \frac{\alpha\mu}{V} \approx \frac{\sum_{\tau=1}^n |V_{\tau}^S - V_{\tau}^B|}{nV}$$

估计 VPIN 需要对每一个篮子的成交量 V 以及用来估计预期交易失衡和密集程度的篮子的数量 n 进行选择。作为初始特值，我们将 V 设定为每日交易量的五分之一。如果我们选择 n=50，我们将会对 50 个篮子计算 VPIN，在一个平均交易量的交易日，这就是日 VPIN。在第六章中将会讨论，我们的结果对大量的 V 和 n 具有鲁棒性。我们也会展示最优的 V 和 n 值可以通过一个最优化的方法找到。

VPIN 估计法在每个交易量篮子后更新。因此，当篮子 51 号满了，我们丢掉 1 号篮子，用篮子 2 号到 51 号计算新的 VPIN。我们在成交量时间更新 VPIN 有两个原因。首先，我们希望更新 VPIN 的速度能够模拟信息到达的速率。为了实现这个目的，我们使用成交量作为信息到达的一个代理变量。第二，我们希望每一次的更新基于可比的信息量。成交量在参与度低的区间非常失衡，因此低的成交量很难像看上去有新信息。因此，在时钟时间更新 VPIN 将会导致信息数量的不同。

以 2009 年 7 月 29 日的上证 50ETF 作为例子。这一天的成交量极高，所以我们的方法产生了 137 个 VPIN 计量结果，而每天平均是 50 个估计值。因为我们的样本容量 n 也是 50，2009 年 7 月 29 日用于估计 VPIN 的时间跨度仅仅是 1 个多小时，而之前平均是 4 小时。

图 3: 50ETF 的 VPIN 估计与 VPIN 更新时间



资料来源：天软资讯，国信证券经济研究所整理

图 3 展示了时间跨度随着交易密度变得有弹性的一个情况，交易密度是信息到达速度的一个代理变量。在早上 9 点 30 分，用于计算 VPIN 的数据跨度约为 1 天。但是随着交易所开市没多久，算法检测到需要更频繁地更新 VPIN，最新的 VPIN 估计建立在更短的时钟时间周期内。在下午 1 点开市时，VPIN 计量值采用半天的数据进行计算（时间间隔为 0.5 天）。注意，减少样本的时间段并未导致估计噪音的增加。相反，VPIN 计量值服从连续的变化趋势。理由是，时间范围不包含可比的信息量。取而代之的是，每次更新的成交量产生了可比信息。

GARCH 提供处理时钟时间高频数据采样中波动性聚集的另一种方法。采用成交量

时间降低了成交量聚集的影响，因为我们每个样本的成交量都相等。大的价格移动跟大的成交量相联系，所以按照成交量采样可以被认为按照波动性采样的一个代理。结果是近似服从正态分布的一组观测值，并且异方差性相对于时钟时间减少。因此，使用成交量时间可以看作是一个简单的对 GARCH 方法的替代。

买入成交量和卖出成交量分类

我们目前还未讨论的问题是如何对买入和卖出成交量进行分类。成交量的方向与交易指令单毒性存在潜在联系，结合总成交量水平和成交量的方向可以指示出存在新信息的可能性。因此，买入成交量占多数说明毒性来自好的信息，反之亦然。

微观结构的研究一般使用打钩算法 (tick-based algorithm) 去标记交易。但是交易方向的划分一直是有问题，在研究中存在着争论。一个问题就是，市场交易回报的惯例对指令单是买单还是卖单可能会区别对待。以纽交所为例的话，对于一个大的卖单与多个小买单相成交，只报告成一笔交易；但是对于一个大买单与多个大卖单相成交，则报告成多笔成交。类似的，将大单分拆成多个小单意味着在很短一段周期内产生的交易不一定是相互独立的，而可能只是某策略的算法所致。而我们在 A 股获得的高频数据交易回报，根据不同软件的算法得出的结果也并非一致。实证研究者解决对应问题所采用的惯例是将一段周期同一方向的交易加总成为一个观测值。

另一个困难是，标记交易需要将交易价格与报价相联系起来。接受做市商买入价的交易员被认为是卖家，反之亦然。在做市商买入价与卖出价之间的交易一般采用打钩算法标记方向。Lee-Ready (1991) 算法建议在报价与交易价格间使用 5 秒的延迟以反映事实——记录下来的报价与交易报告机制不完全一样。即使是统计一些十分简单的低频交易策略，交易方向的判断也会存在错误。

在高频交易的环境下，交易方向的判定就更加困难了。在我们了解过的期货市场或者是涉及到套利的程序化 ETF 交易，没有交易员或专家交易，流动性主要来自于电子指令簿 (电脑程序化交易)，其中包括大量交易员事先设定在计算机里的限价指令。在这种电子化的市场，知情交易者可以以即时的报价成交或者提交一个限价单提高报价，甚至设定一个在当前报价之后的价格加入到指令簿。如果一个主动的指令单与知情交易者的报价成交，如果按照打钩规则进行判定，那么知情交易者将会成为交易对手方。另外，拆单、撤单是任意的。最优的报价 (Best-Bid-Offer) 在每一单交易成交之前可能会变化好几次，许多在同样价格下成交的合约事实上与原先的交易方向相反。在高频领域，对于逐个交易使用打钩算法是无效的。

在本报告的分析里，我们使用的方法是对短周期的交易进行加总，然后使用区间始末的价格变化来对总的成交量判定方向。加总缓和了拆单的影响，并且，使用区间末价格变化进行交易方向判断是一个减少噪音的计量方法。我们使用一分钟柱线图计算成交量，尽管我们的分析在其它频率的时间加总上也同样有效。这种方法将会误判一些成交量，但是在估计指令流毒性时，原始数据的使用都是采取加总的方式。

VPIN 计量法的实证检验

VPIN 是 PIN 的近似值

VPIN 计量法是对 PIN 模型进行的一种估算。在进行估算之前，应该确定其近似值的统计特性。假设买单和卖单是第二部分的交易模型生成的，模型使用的参数是 $(\alpha, \delta, \mu, \epsilon) = (2/5, 2/5, 10, 3)$ 。概率出现的情况主要集中在三个方面：没有事件，好的事件，坏的事件。模型预测这些事件将会促进交易，而且卖单数比买单数多的可能

性更大，也就是说 $\delta < 1/2$ 。

关于模型中事件出现的概率的讨论我们已经在第二部分的模型说明中介绍到了，假设交易指令单的到达服从泊松分布，其参数为 (μ, ε) ，我们可以考虑将所有可能出现的情况和它们的概率 (α, δ) 结合起来，得到一个总的混合分布：

$$P(V^B, V^S) = (1 - \alpha)P(V^B, \varepsilon)P(V^S, \varepsilon) + \alpha[\delta P(V^B, \varepsilon)P(V^S, \mu + \varepsilon) + (1 - \delta)P(V^B, \mu + \varepsilon)P(V^S, \varepsilon)]$$

PIN 是一个基于这些参数的统计值，它将包含在分布里的所有信息总结起来，得到知情交易者报价成功成交的概率。在上面的例子中，它有精确值：

$$PIN = \frac{\alpha\mu}{\alpha\mu + 2\varepsilon} = \frac{1}{6}$$

实际上我们不知道这些参数的值，因此也不知道 PIN 和 VPIN。但是，我们可以换一种方法，根据观测的买单和卖单之间的交易强度和不平衡性来推测出 VPIN。我们可以估算出信息出现的概率，以便于估算等成交量下的知情交易概率：

$$VPIN = \frac{\alpha\mu}{\alpha\mu + 2\varepsilon} \approx \frac{\sum_{\tau=1}^n |V_{\tau}^S - V_{\tau}^B|}{nV}$$

详细讨论这种估算的准确性方法请参考 Easley 等人（2011）文献的蒙特卡洛检验。

VPIN 估计值的实证效果及其稳定性

对于估计的 VPIN 来说，选择如何将交易分类比上面讨论的估算更加重要。尤其是因为 VPIN 涉及到观察交易的不平衡性和强度，不断地聚集柱线将会减少这个变量中的噪音。我们应该观察 VPIN 的相对水平而不是 VPIN 的绝对水平来作为它们的累积分布函数。

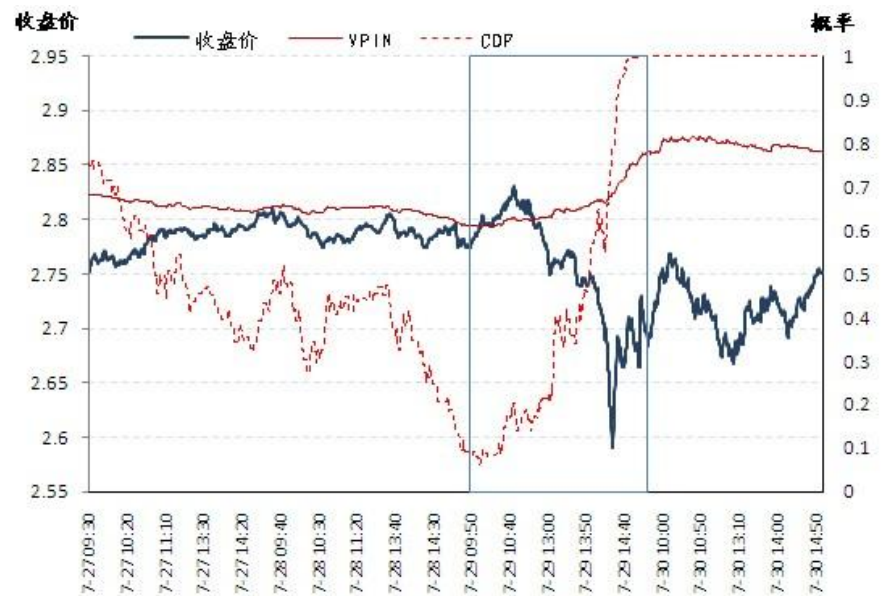
这一点可以通过观察在某一天不同的分类算法下 VPIN 的变化来解释。2009 年 7 月 29 日的尾盘大跳水在分析检验 VPIN 的有效性上有非常重要的意义，因此我们使用不同的交易分类方法来解释这一天 VPIN 的变化。图 4 (a) 是用一分钟柱线计算的 VPIN，图 4 (b) 用的是三十秒钟柱线，图 4 (c) 用的是十秒钟柱线。

图 4(a): 1 分钟频率取样的 50ETF 价格与 VPIN 估计值



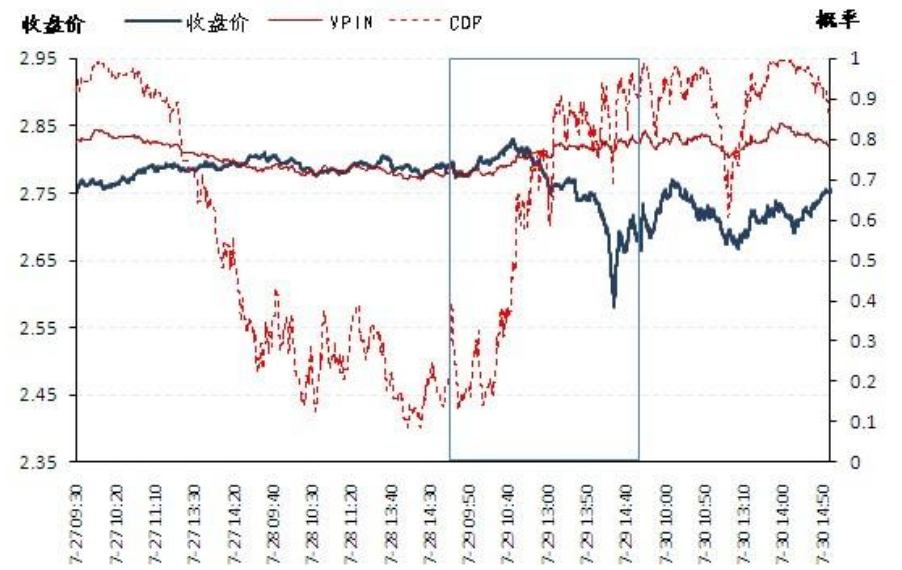
资料来源：天软资讯，国信证券经济研究所整理

图 4(b): 30 秒频率取样的 50ETF 价格与 VPIN 估计值



资料来源：天软资讯，国信证券经济研究所整理

图 4(c): 10 秒频率取样的 50ETF 价格与 VPIN 估计值



资料来源：天软资讯，国信证券经济研究所整理

上面的三个图解释了在 VPIN 的绝对水平下对市场交易的指导意义。观察 VPIN 的变动走势可以发现，当市场成交量减少时，VPIN 的值也会大大下降。同时它们解释了在 7 月 29 日崩盘的那天，当我们构建 VPIN 的累计密度函数 CDF 并将其与同时间轴的 ETF 走势作比较时，可以发现 VPIN 实际上已达到了一个异常高的值。用一分钟柱线和三十秒柱线计算出来的 VPIN 的相对水平基本上是一样的。在所有的情况下，它都在崩盘之前上升，并且在当天余下的时间保持着一个很高的水平，其中一分钟取样频率估算的 VPIN CDF 函数存在相对较好的提前效应。如果把取样频率提高的话，按 10 秒钟分类的结果有些许不同。在这种情况下，VPIN 的估计值在崩盘之前上升，但是它在崩盘之后迅速降低，并且在当天余下的时间保持上下震荡。但是这种上下的惯性是在一个接近市场指数的品种上升了超过 3% 的时期之内的，而实际上成交量在这段时期之内是不可能平衡的。我们觉得这应该是因为交易的分类错误，因此我们没有使用按 10 秒分类法。我们认为最优的结果是使用一分钟柱线计算的结果，同时第二幅实证图表也说明了使用三十秒钟柱线得出的相对结果是基本上相同的。

毒性和交易强度

正如在第三部分“信息和时间在 VPIN 估计中的概念”里讨论的那样，一段长的交易期被划分为许多较短的交易期，交易的不平衡性对流动性提供者有着重要的经济意义。一个成交量篮子 (Volume Buckets) 在很短的时期内被填满意味着重要的或者超出预期的信息已经进入市场。比如说，像图 3 显示的那样，在 2009 年 7 月 29 日，当崩盘的时刻越来越临近时，添加到样本中的交易期的时间范围在不断减小。交易证据数量的上升也提高了知情交易者的参与度。如果 VPIN 是一个有效的衡量交易流毒性的指标，我们应该期待交易强度的上升会导致更高的 VPIN。在我们的分析中，大的交易强度是跟交易量有关的，因此它会被更多的交易者和更大的交易规模所影响。

我们可以看到，一个交易期包括许多变化着的聚集交易（成交量柱线 bars，数量用 K 表示）。如果在一个交易期交易数很大，那么 K 的值就很小。K 越小，VPIN

取得较高值的可能性越大。比如，在一个交易期之内只有一个柱线，使得 50 ETF 的价格移动了 20 个单位，从 2.000 到 1.980；所有的交易都会被记为“卖出”。即使下一个柱线是有着同样交易量的上涨，市场仍然远远不能达到平衡。做市商不能在这样的毒性的环境下进行运作，因为这种环境下不对称的回报会使他们在市场行情好转之前迅速地承受损失。实际上，从一个做市商的角度来考虑，这种情况是非常有毒性的，因为每种偏置的 bar 没有使得他们挽回损失，反而使得情况更加恶化。

VPIN 准确地抓住了毒性的来源，因为一连串的柱线会迅速提高它的价值。例如对于一日的成交量来说，即时当日买单的交易量接近与卖单的交易量，由于单根成交量柱线已足够提供一个或多个不平衡交易的成交量篮子，因此会使毒性不断升高。住的是，由于较小的交易量篮子的尺寸将使 VPIN 分布移动到右边，因此不同标准的 VPIN 之间的比较应该通过它们的累计密度函数 (CDF) 来进行，就如同图 4 中那样。

对 ETF 使用 VPIN 计量法

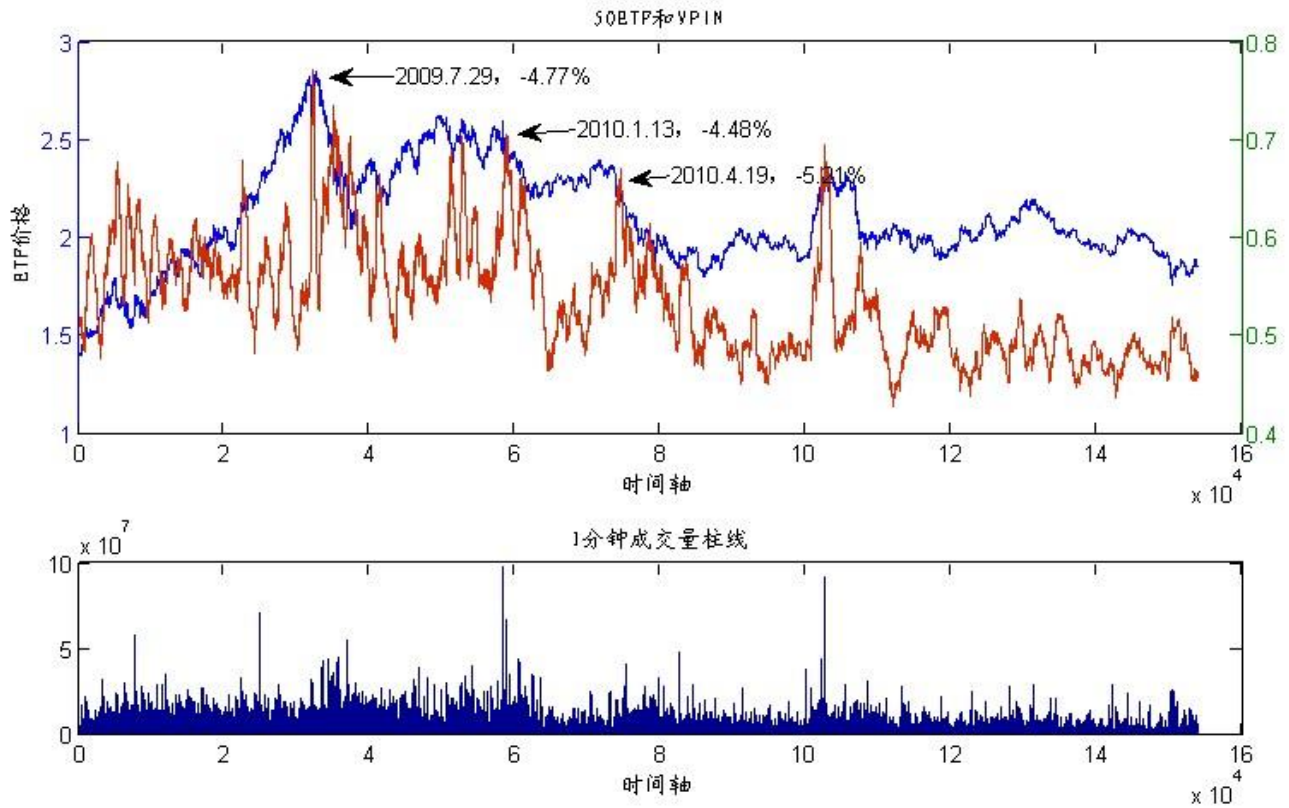
我们已对 VPIN 计量法的估计运算过程做了稳定性分析，下面通过历史数据在上证 50ETF 上分析 VPIN 的性质和用法。样本为上证 50ETF 一分钟成交量柱线：我们的样本期是从 2009 年 1 月 1 日到 2011 年 8 月 29 日。那么，这个数据平均每天有 50 个交易日 (V)，参数在一个 n=50 的滚动的窗口 (也就是平均一天的交易量) 进行估计。VPIN 可以用以下公式进行估算：

$$VPIN = \frac{\alpha\mu}{\alpha\mu + 2\varepsilon} \approx \frac{\sum_{\tau=1}^n |V_{\tau}^S - V_{\tau}^B|}{nV}$$

其中 $\tau=1, \dots, 50$ 表示每一个相同的交易量篮子， V_{τ}^S 是在交易期 τ 中与买单相反的交易量，而 V_{τ}^B 是在交易期 τ 中与卖单相反的交易量。

下面展示了上证 50ETF (红线，2 级市场价格) 和它的 VPIN (绿线) 的变化情况，下方的柱线图为 50ETF 1 分钟成交量柱线图。VPIN 基本是稳定在 0.4-0.6 之间波动。我们主要关注到，VPIN 在 2009 年 7 月 29 日达到了最大值，这为定性观测的“流动性蒸发”提供了实证上的解释。如此高的 VPIN 几乎相当于在当天出现了所有交易指令单都是单向的情况，如市场有做市商存在，这种过高的毒性将导致它们成为流动性消费者而不是流动性提供者，伴随着股指运行到相对的高位，高毒性也意味着崩盘概率的升高，结合之前 VPIN 的 CDF 走势，我们也许可以通过类似方法提前预测到市场崩盘的迹象。

图 5: 50ETF 价格走势、VPIN 估计值与 1 分钟成交量



数据来源: 天软资讯, 国信证券经济研究所整理

VPIN 与波动性

计算 VPIN 的目的不是预测波动性, 而是像微观结构理论解释的那样, 毒性是解释价格变动的本质和波动性的。当订单毒性上升时, 做市商面临着潜在的损失风险, 因此会选择减少甚至放弃市场庄家活动。因此, 这意味着, 高水平的 VPIN 应该预示着较为剧烈价格波动。我们的讨论将主要集中在成交密集时期 VPIN 与绝对收益的相关性。高 VPIN 高绝对收益率的关系也说明了, 在高频交易领域里, 基于成交量划分的波动性预期要比按时间划分更有意义。

三种衡量波动性关系的指标是: 相关系数, 临界值相关系数和条件概率。我们已经测试了不同的资产分类之间的这些关系, 并且发现所有的美国指数期货中存在着一致的效应。为了简明起见, 我们报告了 2008 年 1 月 1 日到 2010 年 11 月 26 日之间 E-mini S&P 500 指数期货的结果, 其他指数期货的结果只是稍微被提及。

VPIN 与绝对收益的相关系数曲面

我们开始研究每日成交量篮子数量以及用于构建 VPIN 的篮子数量的不同的组合对 VPIN 及波动性预测能力的影响。我们先用 VPIN 对数和 ETF 的绝对收益之间简单的相关系数

$$\rho\left(\ln(VPIN_{\tau-1}), \left|\frac{S_{\tau}}{S_{\tau}-1}-1\right|\right)$$

来观察它们之间的相关性。我们使用绝对收益

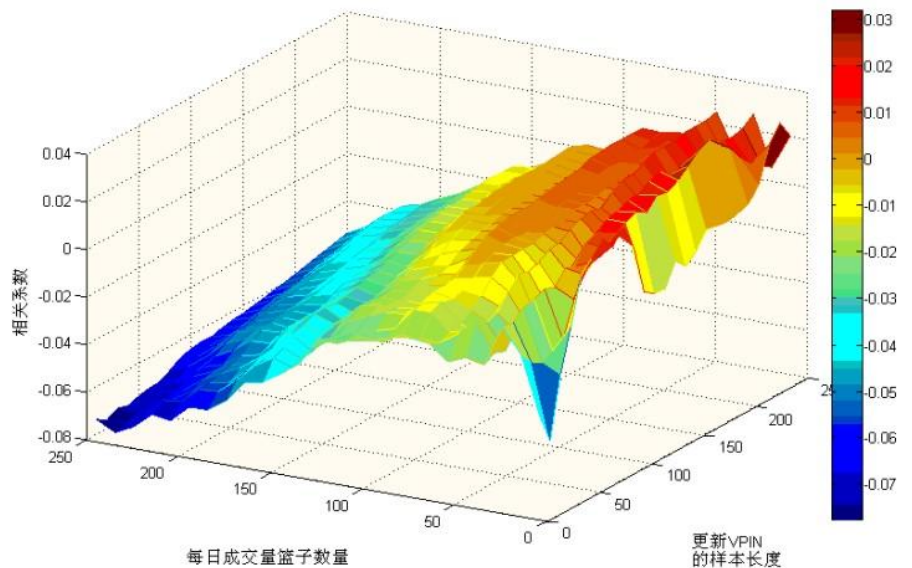
$$\left|\frac{S_{\tau}}{S_{\tau}-1}-1\right|$$

作为价格波动的代理变量。图 6 中的每一点都是用很大的样本进行计算，有些甚至超过了 80000 个观测数据。从图表中可以很明显地看到，VPIN 和绝对收益之间的相关性在不同的参数值之间平稳地移动。而通过观察图中的相关系数平面，我们发现 (50, 250) 组合对于 50ETF 来说是一个好的选择，它还可以简单地解释为“一周”的数据。（平均每天更新 50 个成交量篮子，每周 5 个交易日）。对于每天 50 个成交量篮子，250 个样本长度，我们可以从超过 80000 个观测数据中得到相关系数

$$\rho\left(\ln(VPIN_{\tau-1}), \left|\frac{P_{\tau}}{P_{\tau}-1}-1\right|\right) = 0.0248$$

正相关性可以看做 VPIN 的值越高，标的资产绝对回报越高的表现。

图 6: VPIN 和 50ETF 绝对收益相关性的参数敏感度分析



资料来源：天软资讯，国信证券经济研究所整理

由于有大量的参数组合可以得到相似的预测能力，过度拟合这些参数基本没有优点。因此，我们推荐使用“次最优”的组合 (50,250) 来计算不同的 VPIN 值对应绝对收益的期望值

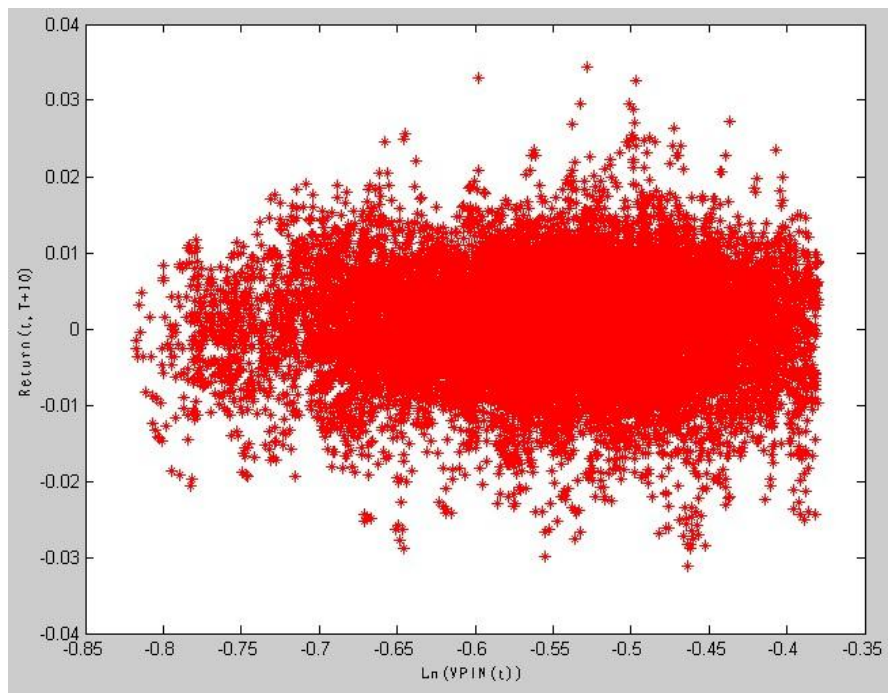
$$E \left[\left. \frac{P_{\tau}}{P_{\tau-1}} - 1 \right| VPIN_{\tau-1} \right],$$

发现 VPIN 和波动性存在同样的正向关系。我们可以注意到，对于我们实证检验里的 ETF 来说，当我们估计的 VPIN 值增加的同时，其波动性也会随之增加。

VPIN 水平对应标的资产的绝对收益

下一幅图展示了 50 ETF 收益 $Return(t,t+10)$ 对应更新 10 个成交量篮子（每日成交量的 1/5）VPIN 的关系，即横轴为相差 10 个交易日 VPIN 值的变化幅度，纵轴为更新 10 个 VPIN 值的时间内收益率的变化情况。图中的散点分布可以看出，毒性越高所对应的收益绝对值也越高（靠右边的点），相反毒性较低时相应的绝对收益较低（靠左边的点）。

图 7: VPIN 水平对应的 50ETF 绝对收益



资料来源：天软资讯，国信证券经济研究所整理

结论

本报告介绍了交易指令的逆向选择、交易流毒性和 PIN 之间的联系。在讨论完 PIN 的理论框架之后，我们解释了 PIN 如何影响流动性提供者的机制。然后，我们提出了一个新的 VPIN 方法。

VPIN 相对之前的估计方法的一个重要的优势就是它的更新频率与成交量契合，这样更新频率可以与信息到达时间同步。它是一个不依赖于不可观测参数的交易毒性估计模型。VPIN 估计值越大，说明市场毒性越高，知情交易者占的比例越高。另外，VPIN 更新的速度越快，说明市场上的知情信息越多。在实证检验中，通过估计 50ETF 的 VPIN，展示了其对应的 CDF 函数在 2009 年 7 月 29 日崩盘前提前 1 小时预警提示大盘毒性极高，告诉投资者市场面临大幅波动（在高位一般指大幅调整）的概率迅速增大。我们还展示了 VPIN 对于波动性有较为明显的预测能力，并且它对未来的高绝对收益提供了指示作用。

因为上述的性质，我们认为 VPIN 可以作为一种高频交易领域的风险管理工具。通过交易指令估计交易毒性，并通过高的 VPIN 值可以提示后续价格将有较高的波动，做市商可以通过 VPIN 来调整做市价差。流动性风险不仅对做市商而言至关重要，对交易员来说，流动性可能会导致价格潜在的大幅波动，适时通过 VPIN 控制交易执行风险也会是一个不错的选择。后续我们将对不同的交易品种进行 VPIN 估计的系统性测算，同时进一步挖掘 VPIN 与标的资产波动性的潜在关系。

国信证券投资评级

类别	级别	定义
股票 投资评级	推荐	预计 6 个月内，股价表现优于市场指数 20%以上
	谨慎推荐	预计 6 个月内，股价表现优于市场指数 10%-20%之间
	中性	预计 6 个月内，股价表现介于市场指数 $\pm 10\%$ 之间
	回避	预计 6 个月内，股价表现弱于市场指数 10%以上
行业 投资评级	推荐	预计 6 个月内，行业指数表现优于市场指数 10%以上
	谨慎推荐	预计 6 个月内，行业指数表现优于市场指数 5%-10%之间
	中性	预计 6 个月内，行业指数表现介于市场指数 $\pm 5\%$ 之间
	回避	预计 6 个月内，行业指数表现弱于市场指数 5%以上

风险提示

本报告信息均来源于公开资料，我公司对这些信息的准确性和完整性不作任何保证。报告中的内容和意见仅供参考，并不构成对所述证券买卖的出价或询价。我公司及其雇员对使用本报告及其内容所引发的任何直接或间接损失概不负责。我公司或关联机构可能会持有报告中所提到的公司所发行的证券头寸并进行交易，还可能为这些公司提供或争取提供投资银行业务服务。本报告版权归国信证券所有，未经书面许可任何机构和个人不得以任何形式翻版、复制、刊登。

证券投资咨询业务的说明

证券投资咨询业务是指取得监管部门颁发的相关资格的机构及其咨询人员为证券投资者或客户提供证券投资的相关信息、分析、预测或建议，并直接或间接收取服务费用的活动。

证券研究报告是证券投资咨询业务的一种基本形式，指证券公司、证券投资咨询机构对证券及证券相关产品的价值、市场走势或者相关影响因素进行分析，形成证券估值、投资评级等投资分析意见，制作证券研究报告，并向客户发布的行为。

国信证券经济研究所团队成员

宏观		固定收益		策略	
周炳林	0755-82130638	李怀定	021-60933152	黄学军	021-60933142
林松立	010-66026312	侯慧梯	021-60875161	林丽梅	021-60933157
崔嵘	021-60933159	张旭	010-66026340		
				技术分析	
交通运输		银行		闫莉	010-88005316
郑武	0755-82130422	邱志承	021-60875167	房地产	
陈建生	0755-82133766	黄飙	0755-82133476	方焱	0755-82130648
岳鑫	0755-82130432	谈焯	010-66025229	区瑞明	0755-82130678
周俊	0755-82130833-6215			黄道立	0755-82133397
糜怀清					
商业贸易		汽车及零配件		钢铁及新材料	
孙菲菲	0755-82130722	左涛	021-60933164	郑东	010-66025270
祝彬	021-60933156			秦波	010-66026317
常伟				郭莹	010-88005303
机械		基础化工		医药	
郑武	0755-82130422	刘旭明	010-66025272	贺平鸽	0755-82133396
陈玲	0755-82130646	张栋梁	0755-82130532	丁丹	0755-82139908
杨森	0755-82133343	罗洋	0755-82150633	杜佐远	0755-82130473
后立尧	010-88005327	吴琳琳	0755-82130833-1867	谭权胜	0755-82136019
		梁丹	0755-82134323	胡博新	0755-82133263
				刘勍	0755-82130833-1845
电力设备与新能源		传媒		有色金属	
杨敬梅	021-60933160	陈财茂	010-88005322	彭波	0755-82133909
张弢	010-88005311	刘明	010-88005319	谢鸿鹤	0755-82130646
				龙飞	
电力与公用事业		非银行金融		通信	
徐颖真	021-60875162	邵子钦	0755-82130468	严平	021-60875165
谢达成	021-60933161	田良	0755-82130513	唐俊杰	021-60875160
		童成敦	0755-82130513		
造纸		家电		计算机	
李世新	0755-82130565	王念春	0755-82130407	段迎晟	0755-82130761
邵达	0755-82130706	黄海培	021-60933150	欧阳仕华	0755-82151833
电子元器件		纺织服装		农业	
段迎晟	0755-82130761	方军平	021-60933158	张如	021-60933151
高耀华	0755-82130771				
熊丹					
建材		旅游		食品饮料	
郑东	010-66025270	曾光	0755-82150809	黄茂	0755-82138922
马彦	010-88005304				
建筑		新兴产业		研究支持	
邱波	0755-82133390	陈健	010-66022025	沈瑞	0755-82132998
刘萍	0755-82130678	李筱筠	010-66026326	雷达	0755-82132098
		孙伟	010-66026320	余辉	0755-82130741
				王越明	0755-82130478
量化投资产品		基金评价与研究		量化投资策略	
焦健	0755-82133928	杨涛	0755-82133339	葛新元	0755-82133332
阳瑾	0755-82133538	康亢	010-66026337	董艺婷	021-60933155
周琦	0755-82133568	刘舒宇	0755-82133568	程景佳	010-88005326
邓岳	0755-82150533	李腾	0755-82130833-6223	郑云	021-60875163
		刘洋	0755-82150566	毛甜	021-60933154
		潘小果	0755-82130843	李荣兴	021-60933165
		蔡乐祥	0755-82130833-1368	郑亚斌	
		钱晶	0755-82130833-1367		
量化交易策略与技术		数据与系统支持			

戴 军	0755-82133129	赵斯尘	021-60875174
黄志文	0755-82133928	徐左乾	0755-82133090
彭甘霖	0755-82133259	李扬之	0755-82136165
秦国文	0755-82133528	陈爱华	0755-82133397
韦 敏	0755-82130833-3772	袁 剑	0755-82139918
张璐楠	0755-82130833- 1379		

国信证券机构销售团队

华北区（机构销售一部）		华东区（机构销售二部）		华南区（机构销售三部）	
王立法	010-66026352 13910524551 wanglf@guosen.com.cn	盛建平	021-60875169 15821778133 shengjp@guosen.com.cn	魏 宁	0755-82133492 13823515980 weining@guosen.com.cn
王晓建	010-66026342 13701099132 wangxj@guosen.com.cn	马小丹	021-60875172 13801832154 maxd@guosen.com.cn	邵燕芳	0755-82133148 13480668226 shaoyf@guosen.com.cn
焦 戡	010-66026343 13601094018 jiaojian@guosen.com.cn	郑 毅	021-60875171 13795229060 zhengyi@guosen.com.cn	林 莉	0755-82133197 13824397011 linli2@guosen.com.cn
徐文琪	010-66026341 13811271758 xuwq@guosen.com.cn	黄胜蓝	021-60875166 13761873797 huangsl@guosen.com.cn	王昊文	0755-82130818 18925287888 wanghaow@guosen.com.cn
李文英	010-88005334 13910793700 liwying@guosen.com.cn	刘 塑	021-60875177 13817906789 liusu@guosen.com.cn	甘 墨	0755-82133456 15013851021 ganmo@guosen.com.cn
赵海英	010-66025249 13810917275 zhaohy@guosen.com.cn	叶琳菲	021-60875178 13817758288 yelf@guosen.com.cn	段莉娟	0755-82130509 18675575010 duanlj@guosen.com.cn
原 祎	010-88005332 15910551936 yuanyi@guosen.com.cn	孔华强	021-60875170 13681669123 konghq@guosen.com.cn	徐 冉	0755-82130655 13632580795 xuran1@guosen.com.cn
				颜小燕	0755-82133147 13590436977 yanxy@guosen.com.cn
				赵晓曦	15999667170 zhaoxxi@guosen.com.cn
				温 馨	13612901872 wenxin@guosen.com.cn
				郑 灿	0755-82133043 13421837630 zhengcan@guosen.com.cn