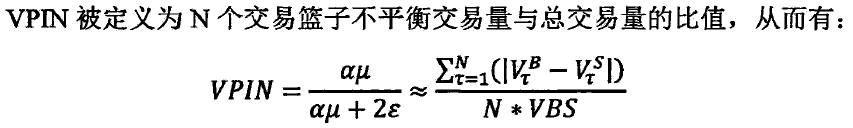
**阶段性结果**

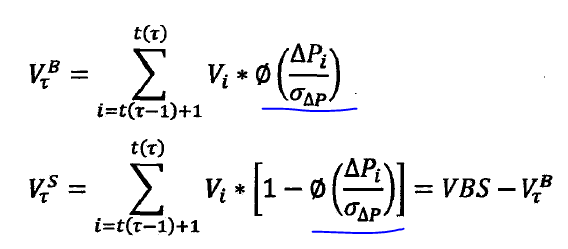
**一、 VPIN的计算**

1 文献结果

VPIN的计算公式：



不平衡交易量的计算是根据价差，历史价差数据累积正态分布参数，来求解的，见下述公式。



论文中交易篮子固定为50。

参数估计数据区间：2010年9月1日～2011年2月28日；

样本数据区间：2011年3月1日～2017年2月28日。

成交篮子大小固定，是使用参数估计数据来计算的，固定为**3579**。

正态分布标准差：**3.168881**；

正态分布均值：**0**；

2 参数结果

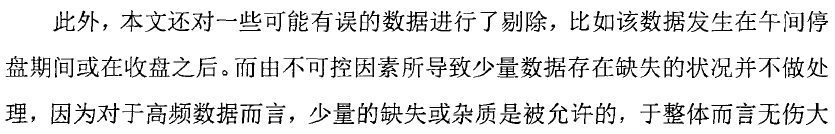
（1）成交篮子大小

使用和论文中相同数据区间，使用相同的数据“IF主力连续.csv”，验证成交篮子大小为**4067**；

（2）正态分布参数

标准差**2.7299**，均值**-0.0086**；

两组比较重要的参数均与文献上的有差别。可能原因是数据来源不同，数据清洗原则不同。论文中提及了



具体原则没有给出。

3 VPIN计算

（1）不平衡订单计算流程



1. 初始化

划分出参数估计区、样本计算区；根据参数估计区计算出成交篮子大小VBS **4067**，初始化篮子ib的cumV大小0；

1. 判断是否已经到达时间序列终点；
2. 如果未达到终点，cumV=cumV+V(i)；
3. 判断cumV是否大于设定的V（VBS），如果否，返回（2）；
4. 记录当前篮子数据

* \*如果cumV大于V，则记录当前篮子内所有的数据，包括取样点时间、价差；
* \*根据历史价差数据，估计价差数据分布的标准差和均值；
* \*根据价差和均值，当前价差估算当前价差所处累积分布概率的位置，然后计算买入和卖出的实际成交量；
* \*根据买卖的实际成交量，计算当前篮子内不平衡订单的总量。

1. 初始化下一个篮子

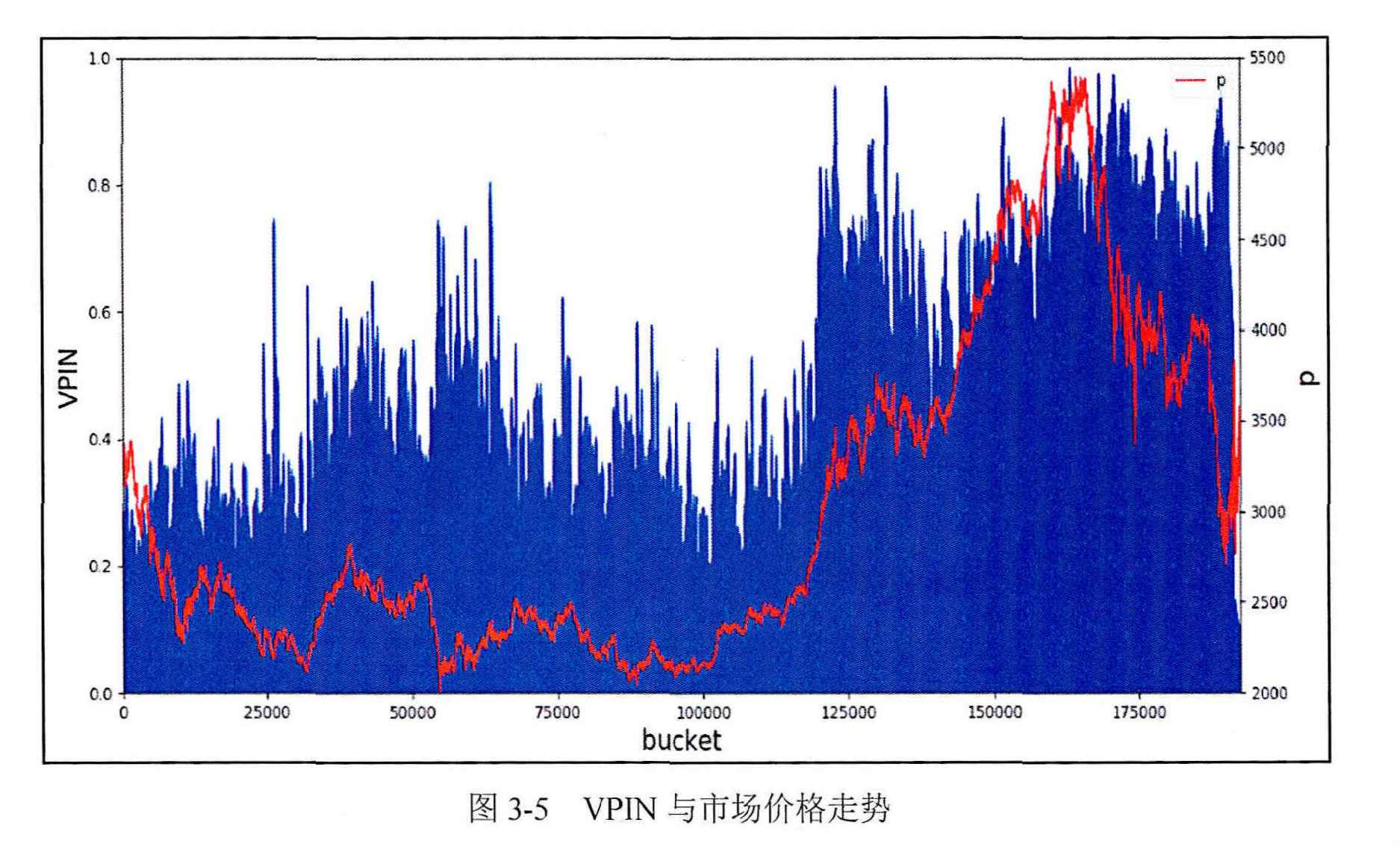
* 计算上一个篮子剩余成交量，篮子序号ib+1，cumV=上一个篮子剩余成交量；若上一个篮子剩余成交量大于V，则ib+1，cumV=cumV-V，根据V计算不平衡交易量，直接增加一个篮子的数据。若cumV仍然大于V，则继续执行上述步骤，直至cumV小于V；

1. 终止。

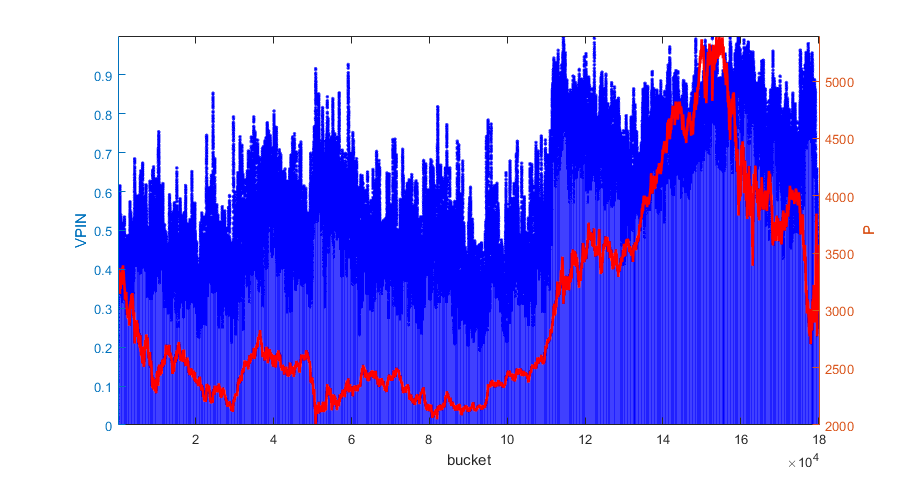
（2）VPIN计算

按照论文中叙述，以50为移动窗口，统计窗口内不平衡订单量之和，然后除以总的订单量计算出VPIN。

4 计算结果和论文结果比较



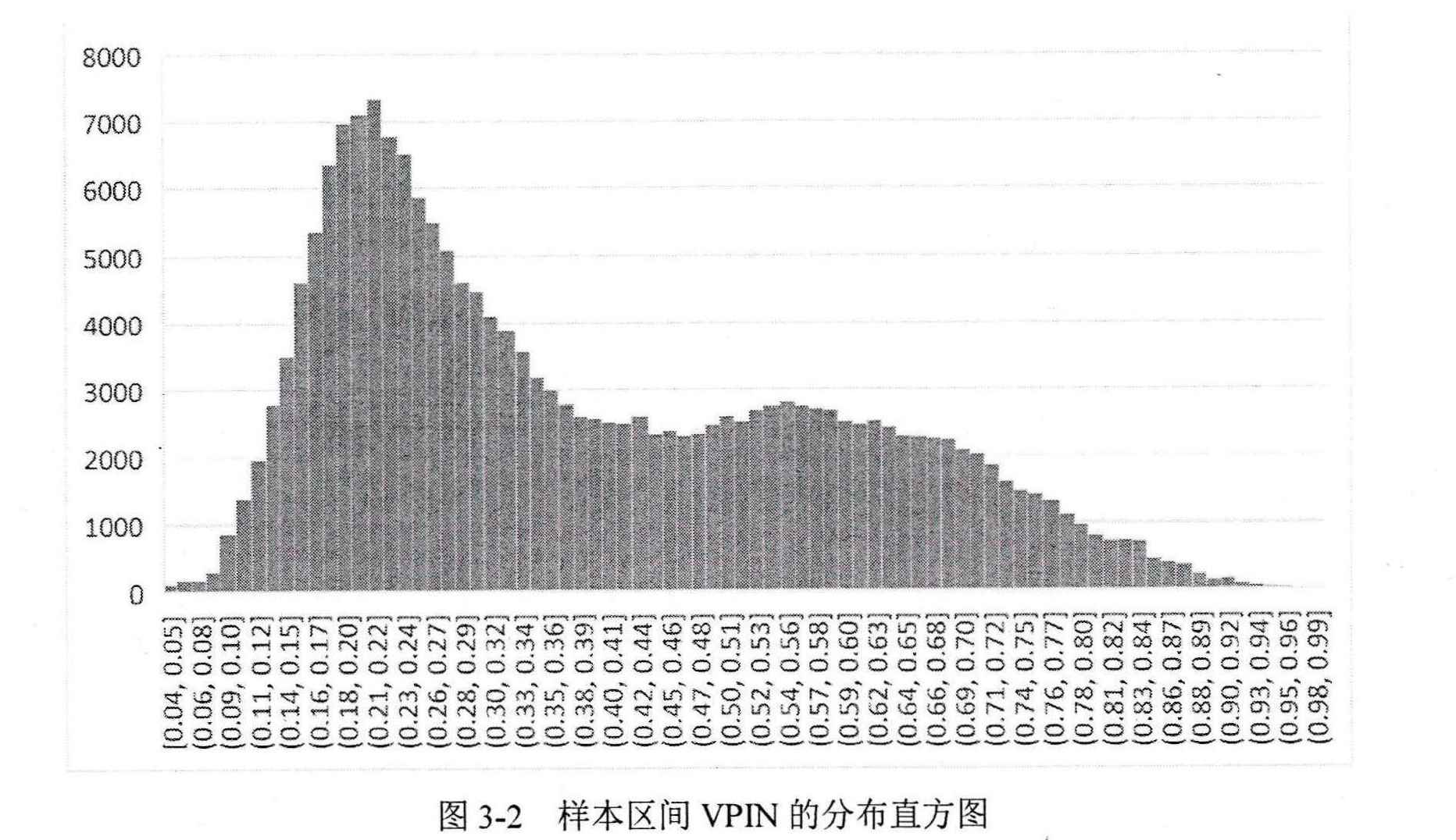
论文中结果



验证结果

直观上看，验证结果和论文结果相似性很高。个别区域也有差异。具体效果如何等待后续验证。

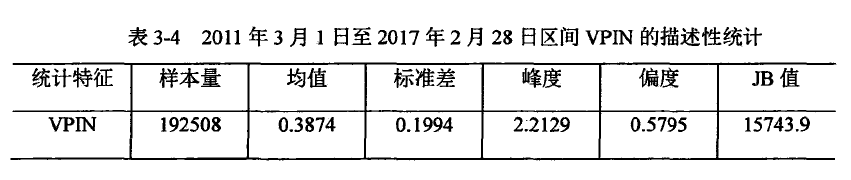
VPIN分布



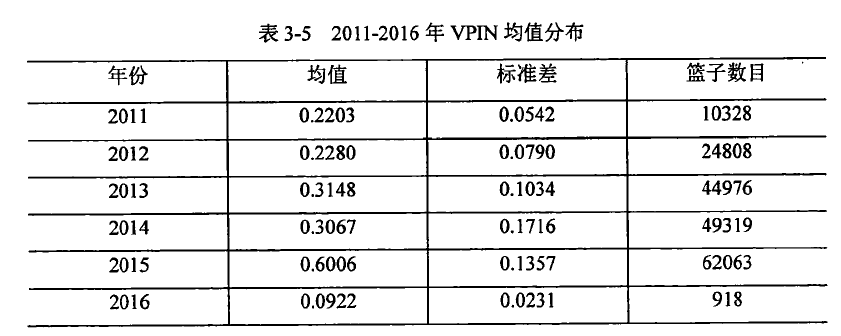
论文中结果



验证结果



|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 统计特征 | 样本量 | 均值 | 标准差 | 峰度 | 偏度 | JB值 |
| VPIN | 180359 | 0.5782 | 0.1742 | 2.0246 | 0.2305 | 8746.6 |



|  |  |  |  |
| --- | --- | --- | --- |
| 年份 | 均值 | 标准偏差 | 篮子数 |
| 2011 | 0.4557 | 0.0713 | 9599 |
| 2012 | 0.4315 | 0.0925 | 22931 |
| 2013 | 0.5261 | 0.1055 | 41941 |
| 2014 | 0.5060 | 0.1722 | 46169 |
| 2015 | 0.7514 | 0.1010 | 58830 |
| 2016 | 0.4313 | 0.1025 | 806 |

