本次数据与首次数据差异较大，原因在于：

1. 对深交所2002年的文件审阅有问题，遗漏了部分数据
2. 对关联方交易以及关联方担保的纳入是在收集数据进程中的，遗漏了部分数据
3. 把年报未及时披露纳入在内，遗漏了部分数据（主要是早期）

申万宏源的造假年份区间选取以及数据参考：

08-15：136

03-15：171

我认为抛弃2002年数据的原因可能是：早期中国市场刚刚建立，很不规范，漏洞很多，所具有的参考意义不大。就深交所数据而言，2002年之前处理的数量是总数的约30%。