

A 股价格波动真的非平稳?

——大数据背景下个股数据的全样本实证

管河山, 闫文玉^①, 王 谦^②

(南华大学 经济管理与法学院, 湖南 衡阳 421001)

[摘 要] 传统研究对股票市场平稳性分析中大多采用市场指数抽样方式, 分析方法和数据采样的差异一定程度上解析了现有研究结论中的不一致性。从市场指数构建原理入手可探究抽样分析的局限, 作为系列个股的加权值, 市场指数无法精确刻画市场所有个股的波动情况。实证时采用 A 股市场所有个股数据开展全样本分析, 这更符合大数据分析理念, 统计出个股分析结果继而对 A 股市场波动进行判定将更加系统和可靠。在不考虑结构突变的情形下, 分别计算出 ADF 检验、KPSS 检验和 PP 检验三种方法的结果; 全样本分析结果表明 A 股市场年度股价数据更多地呈现出平稳性特点, 而且, 个股数据检验结果与市场指数检验结果存在显著差异。

[关键词] 市场指数; 个股; 价格; 平稳性分析

[中图分类号] F832.5 **[文献标识码]** A **[文章编号]** 1673-0755(2018)06-0081-08

2015 年 A 股市场的大起大落给金融监管部门和投资者带来了前所未有的大挑战, 为了维护金融市场的稳定, 国家队多次出手救市, 进一步显露出我国“政策市”的特点。市场指数功能“失真”和“妖股”横行等现象在 2015 年度的 A 股市场中比较普遍, 金融市场监管和谨慎投资意识逐渐强化, 因此, 很有必要加强对个股数据波动的探究。大盘指数在 2015 年 6 月和 8 月经历了两次大跌, 上证指数至今一直在 2800~3600 点区间波动, 维稳成为了现阶段我国金融市场监管工作的一个重点。一方面, 救市政策等外生冲击事件如何影响到市场指数和个股数据的走势仍值得深究, 这有必要对我国 A 股市场中数据的平稳性进行探究。另一方面, 市场指数是否能有效刻画股票市场波动特征值得深究, 这需要从市场指数构建原理和统计抽样角度展开分析。个股分析和市场指数分析的结论是否趋同需从实证角度给出具体的研讨, 这正是本研究的出发点。

本文不将检验方法作为重点分析对象, 改而深入探究个股的平稳性特征, 重点挖掘 A 股市场个股价格数据的波动规律。理论上, 这将不同于传统的市场指数抽样分析, 而是对个股数据进行微观分析, 归纳得出我国 A 股市场数据波动的特征。实践中,

通过对个股数据和市场指数进行对比分析, 探讨传统的市场指数抽样分析的不足, 进一步给出股票市场已有平稳性研究结论中不一致问题的解析。对上市公司的全样本分析更有利于全面、系统地了解 A 股市场中数据波动的本质, 为金融监管层和投资者的决策提供更精准、更有力的支持。

一 平稳性研究概述

(一) 股市平稳性分析的意义

经济时间序列分析中, 平稳序列和非平稳序列具有不同的经济含义和统计性质。传统观点认为: 如果当前冲击的影响是短暂的, 则经济时间序列数据长期运动是由确定性趋势决定的, 不会因为该冲击而改变, 称为趋势平稳 (Trend Stationary) 或确定趋势 (Deterministic Trend) 过程; 反之, 则表示时间序列数据是含有单位根过程。

给出股票价格数据生成方式的分析: 对判定结果为非平稳的价格数据而言, 表明该上市公司数据生成过程含有单位根, 扰动项的冲击具有永久效果, 并未衰减, 政府政策或公司重大事件容易被随机扰动逐渐抵消; 对判定结果为平稳的数据而言, 表明该上市公司数据生成过程是一种趋势平稳过程, 外生冲击改变了数据生成过程的概率分布特征, 但数据

[收稿日期] 2018-07-06

[基金项目] 湖南省教育厅优秀项目“我国 A 股市场个股数据的平稳性探究”资助 (编号: 16B234)

[作者简介] 管河山 (1981-), 男, 湖南祁东人, 南华大学经济管理与法学院副教授, 博士后。

^①南华大学经济管理与法学院硕士研究生。^②南华大学经济管理与法学院教师。

仍将沿着某种确定趋势演进,政府政策或公司重大事件对股票收益率产生积极作用。是否含有单位根将决定不同的时间序列预测理念与结果,简单来说平稳过程对未来的预测主要是其确定性趋势,而单位根过程对未来的最佳预测就是其当前值。

(二) 市场指数抽样分析

我国学者对 A 股市场中数据波动的传统研究中大多是针对市场指数的抽样分析。研究对象包括价格数、收益率数据和成交量等,研究方法包括传统 ADF 检验和结构突变检验等。

早期研究以不考虑结构突变为重。比如,李京对沪深两市的 A 股指数序列进行了系统的单位根检验,ADF 检验结果支持两市的指数序列不能排除单位根的存在,但其差分序列可以排除单位根的存在^[1]。孙便霞等对上证指数收益率研究结果表明该过程有单位根,波动呈现出很强的持续性^[2]。刘田对股市的周收盘价研究表明沪深综合指数序列包含高阶趋势并且是单位根过程^[3]。孙建明对上海证券指数收市价及成交量作了 ADF 检验,发现指数收市价存在有单位根,价格服从随机行走模型,成交量趋势平稳,市场是有效的^[4]。夏南新对上证综指和深证成指的研究中根据 ADF 检验和 PP 检验均表明股市中存在单位根^[5]。李双琦对上证国债指数的收益率和成交量增长率、上证基金指数的收益率和成交量增长率、上证综合指数的收益率和成交量增长率进行单位根检验后得出了不存在单位根的结论^[6]。王丽采用上证综合指数为样本数据,对数据建模进行自相关函数检验和单位根检验,从不同角度得出上证综指服从了随机游走过程,不具有确定的时间趋势^[7]。综上可知,在未考虑结构突变的研究中,大多数学者支持 A 股市场股价数据非平稳、收益率数据平稳的结论。

对 A 股市场数据平稳性分析的近期研究中更多地引入结构突变因素。这是因为结构突变是一种持续性的影响,在单位根检验之前需要去除掉其带来的影响,由此来确定持久性的成分来源于新息成分,即使是中小型的傅里叶型结构突变,也会严重影响 DF 检验的效果,从而使得含傅里叶型平滑结构突变的平稳过程被误判为单位根过程^[8]。比如,佟孟华等采用突变理论分析了上证指数,结果表明上证指数是结构突变的趋势平稳过程,若不考虑结构突变,ADF 检验结果表明上证指数数据服从单位根过程^[9]。周宏山、路维春根据 Zivot 和 Andrew 结构突变断点检测出股市不能拒绝存在单位根的假设,说明上证指数符合随机游走假说^[10]。巴曙松等研

究表明上证指数的 ADF 检验结果为非平稳过程,若采用结构突变理论来分析,则可以发现其中存在一个结构突变,是分段趋势平稳序列^[11]。杨利雄、张春丽分析了上证指数数据的 ADF 检验结果为非平稳的,而低频傅里叶变换能较处理好序列中的结构突变问题,提高分析效率^[12]。韩青采用傅里叶级数逼近展开结构突变研究,结果表明上证指数和深圳指数是含结构突变的随机游走,采用单位根过程建模分析更合适^[13]。侯春艳等以股票资金流强弱指数为研究对象,结果证明上证 A 股存在结构突变点且上证 A 股资金流强弱指数是平稳的非正态时间序列^[14]。上述研究都是针对价格数据展开的分析。现有结构突变研究中存在结论不一致的问题,部分学者认为是平稳的,部分学者认为是非平稳的。

综上所述,可得 3 点结论。第一,早期对 A 股市场的波动研究中,大多认为股价波动是非平稳的。第二,对 A 股市场结构突变的研究中,研究结论仍存在不一致性,尚且无法给出我国 A 股市场中数据波动的最终结论,这可能是分析方法和数据采样方法差异所致,其具体原因仍需深究。第三,现有文献研究过程中大都是对市场指数的抽样分析,并未对个股数据波动展开探讨,该方式能否有效地反映个股波动情况仍值得深究,目前也未有文献对此进行系统探讨。然而,2015 年的 A 股市场波动中,市场指数功能“失真”和“妖股”横行现象比较普遍,有必要从理论角度对市场指数分析途径进行探讨。

以上海证券综合指数为例,样本股是在上海证券交易所全部上市股票,包括 A 股和 B 股,用于测量上交所上市公司价格变动的整体情况。虽然入选的样本是全体股票,然而加权求解上证指数只是宏观测量市场行情,理论上是无法准确反映出个股波动的情况,这一点将在本文实证部分得到验证。以上证 50 指数为例,该指数挑选上交所中规模大、流动性好的最具代表性的 50 只股票组成样本股,力求综合反映上交所最具市场影响力的一批龙头企业的整体状况。然而,成分股选取是变动的,加权求解指数也无法精确反映出个股波动情况。理论上而言,对市场指数的抽样分析虽然能宏观描述市场波动的特点,然而并不能精准刻画个股的波动实情。因此,市场指数抽样分析可能不是最有效途径。现实中,投资者对个股涨跌预判给予了较多的关注,技术投资仍是一项重要的手段,因此,对个股数据的平稳性探究是很有必要的。另一方面,随着大数据理论和实践的发展,对 A 股市场所有个股展开系统分析势在必行。

(三) 大数据分析理念

传统抽样调查是通过局部来进行统计推断, 并由此来探究整体的规律性。在收集数据困难和信息匮乏的时代, 为了体现总体的特征, 抽样方法显得极为重要。这包括随机抽样和非随机抽样, 其中使用较多的方法为随机抽样, 随机抽样是总样本中的每个样本都有一定的机率被抽取, 然后用样本统计量对总体来进行推断。但是, 抽样调查具有其自身的缺陷性, 不管是多完美的随机抽样, 所选取的也只是总样本的一部分, 其统计量特征能否有效地代表总样本特征仍然需要探究。所以统计学需要在大数据背景下进行变革, 进行抽样创新, 否则统计学存在就可能没有显著的意义^[15]。

大数据时代, 对样本抽样和总体的探讨颇为频繁。在国外研究中, Viktor 等人提出关于抽样与全样本的三个论断“不是随机样本, 而是全体数据; 不是精确性, 而是混杂性; 不是因果关系, 而是相关关系”; 而其中“不是精确性”特点也进一步展现了市场指数分析的局限性^[16]。美国科学院全国研究理事会的“大数据分析委员会”指出了在大数据处理中应用抽样的一些问题, 大数据涉及数据分析及编程, 其中最重要的就是处理抽样偏差和数据的非均质性^[17]。Harford 提出人与自然环境都是在不断变化的, 大数据存在其问题, 如果数据的假象影响了其本质, 那么大数据等于总体的观点也只是假设^[18]。国内研究中, 朱建平等指出, 统计学涉及到数据, 需要随着时代的发展而变革, 抽样再准确也只是整体的一部分, 同时在如今大数据时代, 抽样调查中抽样框不稳定、调查数据受限于调查目的、抽样结果经不起细分以及纠正偏误的成本高等都能够解决^[19]。耿直认为由于大数据自身的特点, 必然会面对抽样偏倚、随机误差、非随机误差和数据量过大等问题, 而采用随机抽样、案例抽样、事件抽样等概率抽样和非概率抽样的方法能够避免这些问题的出现, 这也就使得人们不得不学习大数据所涉及的相关领域的内容, 更要研究出一种新的抽样调查方法^[20]。潘利明认为抽样调查相对于普查而言具有经济性、时效性、准确性和灵活性的优势, 但是对于调查统计信息的挖掘与公开又带有滞后性的缺点, 大数据的到来能够对抽样调查进行相互验证, 同样在大数据无法观测到总体特征时, 抽样调查可以发挥其作用^[21]。综上所述, 抽样分析存在其自身的局限性, 大数据处理技术能够对超大规模的数据进行分析处理, 对研究对象的特征既能做到总体把握, 又能了解局部情况, 使得大数据成为普受关注的数据采集方式。获

取大数据是开展分析的前提, 拥有大数据后如何开展分析就是应该考虑的重点^[22]。

金融股市中大量时间序列数据被收集, 给时间序列研究和应用提供了丰富的数据资源和广阔的研究空间。以单位根检验为代表的平稳性分析可以通过编译循环语句处理批量时间序列数据。尽管如此, 股市数据平稳性分析中仍大量采用了抽样分析的方式, 即侧重对市场指数的宏观分析, 淡化了对个股数据的探究; 这并不能精准地反映市场波动的精确情形, 研究是否抽样存在偏差、随机误差、或者非随机误差等问题仍然需要进一步的探究, 所得出结论也需要进一步验证。因此, A股市场股票数据波动研究结论仍存在不一致性就有其必然性。另一方面, 随着大数据时代的到来, 对数据总体分析的特点愈加明显, 因此, 有必要对我国上市公司个股开展系统的分析, 通过对所有个股数据的平稳性探究, 上升到对市场整体的宏观判定, 可以得到更加合理的结论。所以本文将对所有个股的数据进行收集和分析, 这更符合大数据分析理念。

一 平稳性分析理论及方法

(一) 平稳性分析方法概论

主观检验方法的典型方式是借助图形判断时间序列的平稳性特征, 以时间路径图法和自相关函数 (Autocorrelation Function, ACF) 图等方法为代表。时间路径图法可以绘制出时间序列的散点图, 然后观察散点是否围绕其均值上下波动, 如果是, 则判断序列为平稳的, 否则, 判断序列为非平稳的, 该方法受判断者的主观意识影响比较强。以自相关函数图为例, 给定某个时间序列, 分别计算其滞后 k 阶的自相关函数, 如果自相关函数图快速收敛到 0, 则时间序列判定为平稳的, 反之, 则判定为非平稳的。这些方法可进行平稳性的粗略判定, 但面对海量序列数据处理时的工作量极大, 因为该类方法缺少必要的量化分析原理和方法, 至今仍停留在主观分析层面, 而且检验结果因人而异, 得不到一致的结论。此类方法不适合分析我国 A 股市场中的个股数据, 因为个股数据的样本量较大, 主观判定耗时较大, 也难以得到一致的结论。

单位根检验方法因其较完整的理论体系和客观的检验方式成为了平稳性检验的主流方法。单位根检验 (Dickey-Fuller test, DF 检验) 是平稳性检验中常用到的一种方法。ADF 检验是一种扩充的 DF 检验, 可以解决 DF 检验中对随机项误差白噪声假设的问题。其后续研究也很多。此外, PP 检验、KPSS

检验、LMC 检验等检验方法也得到了广泛的研究和应用。国内外学者的后续研究极大地丰富了时间序列单位根检验理论和应用研究,但是这些研究本质上还是继承了单位根检验的思想,研究过程中大量地采用了统计分布的概念。该类方法给出了平稳性分析的量化方式,因此可以处理海量时间序列数据,本文实证部分将采用此类方法展开分析。

经济时间序列中的结构突变检验是近期研究的重点。结构突变就是考察外生冲击是否改变了时间序列数据生成过程,它模糊了单位根过程与趋势平稳过程的鲜明界限。一个显著的特点是,结构突变检验通常假定时间序列波动是非线性的趋势,其分析途径主要由内生和外生两种。然而,本文实证部分对结构突变不做重点考虑,在此,不过多探讨。实证部分重在对上市公司全体样本进行系统性的分析,检验方法采用传统的 ADF 检验、PP 检验和 KPSS 检验。

(二) ADF 检验

单位根检验已成为时间序列平稳性检验的主流方法。DF 单位根检验和 ADF 单位根检验是通过建立时间序列的自回归方程,引入滞后算子构建特征方程,根据其特征根的绝对值来判断序列的平稳性,其准确性主要受到自回归方程精确性的影响,实践中只能从统计建模角度给出一个最优的模型。Dickey 和 Fuller 使用蒙特卡洛模拟方法计算出了 ADF 检验统计量的临界值表,并用于判断序列的平稳性,其判定结果也是统计意义上的某种近似结果。

用 $\{x_t\}$ 表示一个 AR(p) 时间序列,为了验证序列是否存在单位根,可以采用 ADF 检验来验证,即可以用如下回归来进行假设检验 ($H_0: \beta = 1; H_1: \beta < 1$):

$$x_t = c_t + \beta x_{t-1} + \sum_{i=1}^{p-1} \phi_i \Delta x_{t-i} + e_t \quad (1)$$

其中 c_t 是关于时间 t 的确定性函数, Δx_{t-i} 是 x_{t-i} 的一阶差分序列。 $\hat{\beta}$ 的 t -比为 $(\hat{\beta}-1)/\sigma(\hat{\beta})$, 其中, 分母表示标准差, $\hat{\beta}$ 为 β 最小二乘估计, 该比值为 ADF 检验。注意到由于一阶差分, 上面 x_t 的式子等价于一个带确定性函数 c_t 的 AR(p) 模型, 可以改写为:

$$\Delta x_t = c_t + \beta_c x_{t-1} + \sum_{i=1}^{p-1} \phi_i \Delta x_{t-i} + e_t \quad (2)$$

其中 $\beta_c = \beta - 1$, 相应的检验为 $H_0: \beta_c = 0; H_1: \beta_c < 0$ 。

实际中, c_t 可以是常数或者 $c_t = \omega_0 + \omega_1 t$ 。针对 c_t 取值不同, ADF 检验方法有 3 种形式, 即为: 原始检

验方程 ($c_t = 0$)、包含常数项的检验方程 (c_t 为常数) 和包含常数及趋势项的检验方程 ($c_t = \omega_0 + \omega_1 t$)。实证时, 依次给出 3 个模型的检验结果, 并查验 ADF 检验临界表, 给出最终检验结果。如果 3 个模型的检验结果都不能拒绝原假设, 则认为时间序列是非平稳的。具体步骤可以参见李子奈的《高级计量经济学》^[23]。

(三) PP 检验

PP 检验的原理是根据非参数估计法来估计出多余参数, 然后去修正 DF 统计量由此来判断是否含有单位根。PP 检验假定数据生成过程 $\{x_t\}$ 不是特殊的而是一般的过程。ADF 中检验过程假定为自回归过程, 但在 PP 检验中则是 $MA\{\infty\}$ 过程。PP 检验时将数据生成时的误差当做是非相关序列, 然后根据最小二乘法构造初始统计量, 最后再考虑序列的相关性, 对统计量进行修改。

假定时间序列 $x_t = c_t + \beta x_{t-1} + u_t$, 其中 $u_t = \varphi(L)e_t = \sum_{j=0}^{\infty} \varphi_j e_{t-j}$, e_t 独立同分布, 均值为 0, 方差为 σ , $\varphi(L)$ 为无穷阶的滞后多项式。假设检验 $H_0: \beta = 1; H_1: \beta < 1$ 。构建检验量如下:

$$Z = T(\hat{\rho}_T - 1) - \frac{1}{2}(\sigma^2 \varphi(1)^2 - r_0) \left(\frac{T^2 \hat{\eta}_T^2}{S_T^2} \right) \\ \Rightarrow \frac{\frac{1}{2}(W^2(1) - 1) - W(1) \int_0^1 W(r) dr}{\int_0^1 W^2(r) dr - (\int_0^1 W(r) dr)^2} \quad (3)$$

$$t_T = \frac{T(\hat{\rho}_T - 1)}{(T^2 \hat{\eta})^{\frac{1}{2}}} \\ \Rightarrow \frac{\frac{1}{2}(W^2(1) - 1) - W(1) \int_0^1 W(r) dr}{\left\{ \int_0^1 W^2(r) dr - (\int_0^1 W(r) dr)^2 \right\}^{\frac{1}{2}}} \quad (4)$$

在 PP 检验中, 可以对它们进行调整, 修正后极限与 DF 极限相同。PP 检验作为非参数估计方法, 本文实证时将以此进行对比。

(四) KPSS 检验

在 DF 检验、ADF 检验和 PP 检验中, 单位根都对应于零假设, 但是我们需要注意的是如果不能拒绝零假设并不意味着零假设就一定能成立, 即该时间序列是含有单位根的, 这是因为我们所得到的信息还不够完整。在 KPSS 检验中原假设换成了时间序列数据是平稳的, 备择假设则是换成了单位根序列。KPSS 检验的主要内容是假设原时间序列数据是平稳序列, 那么可以分解成为确定的时间序列数

据, 在受到随机游走或者是稳定过程的干扰之后, 随机游走部分的方差应该为 0。

通过下面几个步骤来实现:

(1) 用 x_t 做对截距项和时间趋势的 OLS 回归,
 $x_t = c + \beta t + e_t$

(2) 计算该回归的残差部分和 $\hat{S}_t = \sum_{t=1}^T e_t, e_t = x_t - \hat{c} - \hat{\beta}t$

(3) 构造检验统计量: $v_2(r) = W(r) - (3r^2 - 2r)W(1) + (6r^2 + 6r) \int_0^1 w(s) ds, \sigma_e^2$ 是残差平方和的估算, $\sigma_e^2 = T^{-1} \sum_{i=1}^T e_i^2$

其中, 统计量 η_T 的渐进分布并不是标准的分布, 它的极限分布为 $\eta_T \Rightarrow \int_0^1 [v_2(r)]^2 dr$, 且 $v_2(r) = W(r) - (3r^2 - 2r)W(1) + (6r^2 + 6r) \int_0^1 w(s) ds$ 。根据计算, Kwiatkowski 得出了统计量 5% 的临界值是 0.146。需要注意的是如果原假设是一般平稳而不是趋势平稳假设, 则在第一个回归的时间应该把趋势变量去掉, 得出此时 5% 的临界值为 0.463。

三 实证分析

(一) 数据采集说明

采集 A 股市场上市公司的历史数据进行分析, 采集对象为日价格数据(考虑现金红利再投资的收盘价的可比价格), 数据采样时间段为 1990 年 1 月 1 日至 2017 年 1 月 1 日, 共计 3099 只个股, 数据来自国泰安 CSMAR 数据库。然而, 考虑到我国证券市场自 1995 年沿用 T+1 制度, 以及从 1996 年 12 月开始实施涨跌停板制度, 制度更改对个股价格涨跌造成了明显的影响, 数据采样时间段修正为 1997 年 1 月 1 日至 2017 年 1 月 1 日。此外, 上市公司因为中途停牌, 或者上市时间差异等影响, 需要对个股采样时间长度进行分析。本文按年度分别采集个股数据, 并剔除了采样时间长度小于 100 天的股票, 得到每个年度的数据, 如表 1 所示。

表 1 各年度的个股数量 (单位: 支)

年份(年)	个股数	年份(年)	个股数	年份(年)	个股数
2016	2840	2009	1569	2002	1162
2015	2714	2008	1541	2001	1103
2014	2510	2007	1405	2000	997
2013	2453	2006	1338	1999	886
2012	2413	2005	1351	1998	789
2011	2184	2004	1329	1997	684
2010	1853	2003	1227		

为了分析个股数据和市场指数之间的波动性差异, 本文采集了上证综合指数(000001)、深证成份指数(399001) 和沪深 300 指数(000300) 进行对比分析, 数据采样时间段也为 1997 年 1 月 1 日至 2017 年 1 月 1 日。考虑到 3 个市场指数的实际使用时间, 其数据时间长度仍存在差异, 如表 2 所示。

表 2 市场指数采样时间长度 T (单位: 天)

市场指数	样本实际时间段	时间长度 T
上证综合指数(000001)	1997/1/1-2017/1/1	4843
深证成份指数(399001)	1997/1/1-2017/1/1	4843
沪深 300(000300)	2005/4/8-2017/1/1	2854

(二) 相关系数分析

个股走势与市场指数走势通常呈现出不同的步调, 这可以通过求解两者之间的相关系数进行分析。相关系数取值位于 $[-1, 1]$ 区间, 取值越大表明两者之间的差异越小, 反之, 则表明两者之间的差异越明显。对个股和市场指数分析, 汇总 20 年数据为一个整体样本, 并分别计算个股与 3 个市场指数之间的相关系数, 统计相关系数的极值点、四分位点和缺失样本的情况, 如表 3 所示。有 28 家上市公司与沪深 300 指数之间的相关系数缺失, 究其原因不难发现, 沪深 300 指数的采样时间起点为 2005 年 4 月 8 日, 而部分上市公司的交易数据在这之前就已经结束, 因此, 两者之间的相关系数无法计算, 故而缺失。

表 3 市场指数与个股的相关系数

市场指数	最大值	第三四分位数(Q3)	第二四分位数(中位数)	第一四分位数(Q1)	最小值	缺失值(个)
上证综合指数	0.980	0.805	0.681	0.509	-0.926	0
深证成份指数	0.978	0.755	0.614	0.434	-0.937	0
沪深 300	0.991	0.794	0.674	0.499	-0.963	28

对 3 个市场指数而言, 第一四分位数都低于 0.51, 这表明至少有 25% 的个股与市场指数之间的

相关系数小于 0.51,当然,也有部分个股与市场指数之间是负相关的。第二四分位数都低于 0.69,这表明至少有 50%的股票样本与市场指数之间的相关系数小于 0.69。由表 3 可以推断出 A 股市场的股票走势与市场指数走势呈现出差异。因此,有必要对个股和市场指数同时展开分析,以更好地探究 A 股市场价格数据波动的特点。

(三)个股平稳性分析

分别采用 ADF 检验、PP 检验和 KPSS 检验对各年度上市公司的股价数据进行单位根检验。对 ADF 检验和 PP 检验,实证时,依次给出 3 个模型(不含常数项,含常数项,含趋势项)的检验结果,并查检验临界表,给出最终检验结果。显著性水平通常取 0.01、0.05 和 0.1 三种值。当检验统计量的概率 P 值小于给定的显著性水平值时,则拒绝原假设,序列是平稳的,当概率 P 值大于显著性水平值时,则接受原假设,序列非平稳。统计出检验结果如表 4。

表 4 个股平稳性分析结果(取值为样本占比)

年份	ADF 检验		KPSS 检验		PP 检验	
	平稳	非平稳	平稳	非平稳	平稳	非平稳
1997	0.72	0.28	0	1.00	0.72	0.28
1998	0.88	0.12	0	1.00	0.88	0.12
1999	0.90	0.10	0	1.00	0.90	0.10
2000	0.94	0.06	0.03	0.97	0.94	0.06
2001	0.98	0.02	0.01	0.99	0.98	0.02
2002	0.97	0.03	0.01	0.99	0.97	0.03
2003	0.97	0.03	0.08	0.92	0.97	0.03
2004	0.96	0.04	0.01	0.99	0.96	0.04
2005	0.98	0.02	0	1.00	0.98	0.02
2006	0.93	0.07	0	1.00	0.93	0.07
2007	0.95	0.05	0	1.00	0.95	0.05
2008	0.97	0.03	0	1.00	0.97	0.03
2009	0.99	0.01	0.02	0.98	0.99	0.01
2010	0.98	0.02	0.01	0.99	0.98	0.02
2011	0.96	0.04	0.04	0.96	0.96	0.04
2012	0.92	0.08	0.02	0.98	0.92	0.08
2013	0.37	0.63	0	1.00	0.37	0.63
2014	0.32	0.68	0	1.00	0.32	0.68
2015	0.28	0.72	0	1.00	0.28	0.72
2016	0.93	0.07	0.03	0.97	0.93	0.07

注:采用 MATLAB 软件分析,KPSS 检验选择了默认参数值,ADF 检验和 PP 检验对三个模型进行了综合分析

表 5 市场指数平稳性分析结果
(取值为样本占比)

年份	上证综合指数			深证成份指数			沪深 300		
	ADF	KPSS	PP	ADF	KPSS	PP	ADF	KPSS	PP
1997	0	0	0	0	0	0			
1998	0	0	0	0	0	0			
1999	0	0	0	0	0	0			
2000	0	0	0	1	0	1			
2001	0	0	0	0	0	0			
2002	0	0	0	0	0	0			
2003	0	0	0	0	0	0			
2004	0	0	0	0	0	0			
2005	0	0	0	0	0	0	0	0	0
2006	0	0	0	0	0	0	0	0	0
2007	0	0	0	0	0	0	0	0	0
2008	1	0	1	1	0	1	1	0	1
2009	0	0	0	0	0	0	0	0	0
2010	0	0	0	0	0	0	0	0	0
2011	0	0	0	0	0	0	0	0	0
2012	0	0	0	0	0	0	0	0	0
2013	0	0	0	0	0	0	0	0	0
2014	0	0	0	0	0	0	0	0	0
2015	0	0	0	0	0	0	0	0	0
2016	1	0	1	1	0	1	1	0	1

注:取值 1 表示平稳,取值 0 表示非平稳,统计出样本比例

由表 4 和表 5 分析,可以得出如下结论:

第一,对个股全样本分析和市场指数抽样分析结果进行比较。根据表 4 分析结果可知,ADF 检验和 PP 检验的结果相似,这表明,对 A 股市场个股数据波动的平稳性分析中,作为非参数检验的 PP 检验,其实际效果与 ADF 检验相当,且大部分个股的年度数据都判定为平稳序列。KPSS 检验的结果恰恰相反,大部分个股的年度数据都判定为非平稳序列。究其原因,KPSS 检验原假设是平稳的,实证结果大概率拒绝原假设。ADF 检验是经济时间序列单位根检验中使用频率最大的方法,由此可以推断出,个股价格数据呈现平稳性的概率较大。然而,根据表 5 分析结果可知,3 种检验方法的结论基本上一致,基于市场指数的单位根检验大多支持股价数据非平稳。综上所述,A 股市场上市公司全样本分析与传统基于市场指数抽样分析之间的实证结果相差较大。基于全样本的上市公司个股分析结果更为微观,其结果更能系统描述 A 股市场股价波动的特征,而基于市场指数的分析依赖于指数构建方式,其

波动特征能否刻画所有个股波动理论上仍缺少支撑依据。这一分析结果的对比,表明传统基于市场指数分析的局限所在,而对所有个股的分析更符合大数据理念。对个股分析结果进行统计,表明 A 股市场股价数据(年度)更趋向于平稳。处于成长周期、成熟周期或衰退周期的公司股票可能呈现某种稳定的变动趋势,政策利好(或利空)可能也会给上市公司股价带来某种稳定的变动趋势。因此从企业的生命周期和宏观政策导向而言,股价年度数据平稳的结论是比较合理的,这也是技术投资广泛存在的有力支撑。

第二,对股市不同阶段(年度)的波动特征进行比较分析。由表 4 可知,在股市出现熊市的当年及其前 1 年,比如 1998 年、2007 年和 2015 年等,其个股数据的检验结果与其它年份存在差异。对 ADF 检验、PP 检验和 KPSS 检验 3 种检验方法而言,个股波动的判定结果更多趋向于非平稳。究其原因,这些采样年份,市场指数和个股波动大多是由上涨转变为下跌的阶段,其股价波动幅度较大,且可能呈现出明显的非线性特点。而 3 种检验方法通常是假定为线性趋势项,其检验效果仍值得深入探讨,结构突变检验值得尝试。对比表 5,仅在 2008 年和 2016 年出现平稳特征,其它时刻都是非平稳的,这两年都

是股市大跌之后的第 2 年,其走势呈现出明显的直线下滑,可能形成趋势平稳特点。

第三,为了更为准确地对比两种分析方式的差异,我们对个股分析和市场指数分析进行统计分组检验。采用的方法两样本对比,分别分析均值(H_0 : 两组样本均值差为 0)和方差(H_0 : 两组样本方差比值为 1),统计 3 种检验方法所对应的计算结果如表 6。分别给出 3 个市场指数与个股数据之间的平稳性概率对比分析:上证指数与个股之间的概率方差无显著差异($Pr>F$ 取值大于 0.1、0.05 和 0.01 三种显著性水平),两者的概率均值显著不同($Pr>t$ 取值小于 0.0001,因此小于 0.1、0.05 和 0.01 三种显著性水平);深证指数与个股之间的概率方差有显著差异(假定显著性水平默认值为 0.05,如果显著性水平为 0.1 则结果相反),两者的概率均值显著不同($Pr>t$ 取值都小于 0.1、0.05 和 0.01 三种显著性水平);沪深 300 指数与个股之间的概率方差无显著差异($Pr>F$ 取值大于 0.1、0.05 和 0.01 三种显著性水平),两者的概率均值显著不同(假定显著性水平默认值为 0.05,如果显著性水平为 0.01 则结果相反)。通过上述分析,3 种不同检验方法得到的结果都表明市场指数与个股数据平稳性分析存在差异。

表 6 两样本 t 检验

市场指数	ADF 检验		KPSS 检验		PP 检验	
	均值($Pr>t$)	方差($Pr>F$)	均值($Pr>t$)	方差($Pr>F$)	均值($Pr>t$)	方差($Pr>F$)
上证综合指数	<0.0001	0.2324	0.0061	.	<0.0001	0.2324
深证成份指数	<0.0001	0.0549	0.0091	.	<0.0001	0.0549
沪深 300	0.0002	0.3317	0.0323	.	0.0002	0.3317

注:采用 SAS 软件分析,沪深 300 从 2005 年开始分析;先进行方差检验,然后再进行均值检验。其中 KPSS 检验所对应的方差分析结果缺失

综上所述,归纳个股数据集平稳性特征上升到 A 股市场波动特征是一种更为可靠的的实证途径,对个股数据的系统分析更为合理,这正是本文研究与传统研究最大的区别所在。

四 研究结论

针对 A 股市场股价波动的平稳性展开分析,重点探讨传统市场指数抽样分析和大数据背景下个股全样本实证的差异,检验理论方法的创新不是本文的着重点。理论上,从市场指数构建原理角度分析了其性能,指出市场指数波动无法精准刻画个股波动的特点。实践中,对 A 股市场所有个股进行全样本分析,并与 3 个市场指数抽样分析进行比较。相

关系数研究结果表明部分个股与市场指数的走势存在显著差异,平稳性检验统计结果也表明个股波动(年度数据)大部分呈现出平稳特征,而市场指数呈现出非平稳特征,市场指数平稳性分析与个股平稳性分析结论存在显著差异,两组样本对比检验结果进一步表明个股数据和市场指数的分析结果存在显著差异。

本文对个股分析更能系统捕捉市场波动的微观特点,在数据获取和技术分析可行的前提下,对 A 股市场个股数据的全样本分析更符合大数据理念,对市场指数抽样和部分个股抽样分析难以系统刻画股价波动的特征,实证得出股价平稳结论与现实中大量技术投资研究不谋而合(个股分析和技术投资

决策)。正如本文文献综述所示,传统的市场指数抽样分析却难以得到一致的结论,在大数据时代更显露出其局限性,大数据背景下数据总体分析将发挥更加重要的作用。

[参考文献]

- [1] 李京.深沪 A 股指数单位根现象的实证研究[D].厦门:厦门大学硕士学位论文,2002.
- [2] 孙便霞,韩鹏,王明进.随机波动率模型下的单位根检验及对我国股市的实证研究[J].数理统计与管理,2008,27(1):148-155.
- [3] 刘田.非线性趋势单位根检验研究[D].成都:西南财经大学博士学位论文,2009.
- [4] 孙建明.股市价格指数成交量的单位根及因果关系检验[J].华中科技大学学报(自然科学版),2003(10):106-108.
- [5] 夏南新.单位根的 DF、ADF 检验与 PP 检验比较研究[J].数量经济技术经济研究,2005(9):130-136.
- [6] 李双琦.基于分形理论的上海证券市场有效性研究[D].重庆:重庆工商大学硕士学位论文,2016.
- [7] 王丽.上海股票市场有效性实证研究[J].时代金融,2017(26):195-196.
- [8] 张春丽,杨利雄,李庆男.傅里叶型平滑结构突变下 Dickey-Fuller 检验的功效——兼论 Shibor 的平稳性[J].统计与信息论坛,2016,31(5):21-26.
- [9] 佟孟华,钟春仿,郭多祚.结构突变及其对上证指数数的实证研究[J].财经问题研究,2004(3):76-79.
- [10] 周宏山,路维春.中国股票指数的结构断点根检验研究[J].经济问题,2007(2):105-106.
- [11] 巴曙松,覃川桃,朱元倩.中国股票市场与房地产市场的联动关系[J].系统工程,2009,27(9):22-28.
- [12] 杨利雄,张春丽.基于傅里叶变换的含确定性趋势结构突变的协整回归模型和不等方差检验[J].统计研究,2014,3(11):96-100.
- [13] 韩青.中国宏观经济时序的平稳性再考察——内生突变与平滑转换[J].经济学(季刊),2015,14(2):801-826.
- [14] 侯春艳,李俊林,董安强.股票资金流强弱指数的结构突变研究[J].科技和产业,2017,17(2):168-172.
- [15] 纪思亮.数据圈宫心计:大数据与随机抽样——统计学必须进行技术革命[J].中国统计,2016(8):16-19.
- [16] VIKTOR M S. Big data: a revolution that will transform how we live, work, and think[M]. London: John Murray publisher, 2013.
- [17] Committee on the Analysis of Massive Data, et al. Frontiers in Massive Data Analysis [M]. Washington: National Academies Press, 2013.
- [18] HARFORD T. Big data: A big mistake? [J]. Significance, 2015, 11(5):14-19.
- [19] 朱建平,章贵军,刘晓葳.大数据时代下数据分析理念的辨析[J].统计研究,2014,31(2):10-19.
- [20] 耿直.大数据时代统计学面临的机遇与挑战[J].统计研究,2014,31(1):5-9.
- [21] 潘利明.简述抽样调查在“大数据”时代下的意义[J].现代经济信息,2016(1):103.
- [22] 程开明,宋艺璇.大数据还需要抽样吗[J].中国统计,2017(11):10-13.
- [23] 李子奈.高级应用计量经济学[M].北京:清华大学出版社,2012.

Is the Price Volatility of China's A Share Market Really Nonstationary?

——Full sample empirical analysis of individual stock data in the big data background

GUAN He-shan, YAN Wen-yu, WANG Qian

(University of South China, Hengyang 421001, China)

Abstract: The traditional research on the stock market stability analysis mostly uses sampling mode of the market index, so the differences between analysis method and data sampling to some extent give an explanation to the inconsistency of existing research conclusions. Starting with the principle of market index construction, the paper analyzes the limitation of sampling method. As a weighted value of a series of individual stocks, market index may not accurately depict the fluctuation of all individual stocks in the market. In order to be more in line with big data's analysis concept, the article uses all the A-share market data in China to carry out full-sample analysis, and work out all individual stock results which be more systematic and reliable to judge the fluctuation of A-share market. And then the results of ADF test, KPSS test and PP test were calculated without considering the structural mutation. The results of full-sample analysis show that the annual stock price data of A-share market present more stationary, moreover, the test results of individual stock data are significantly different from the market index's.

Key words: market index; individual stock; price; stationarity analysis