

价格离散测度及影响溯源研究

——基于在线大数据的经验证据

秦 杰,胡林瑶,孟 陆

(1. 南京审计大学 经济学院,江苏 南京 211815;2. 北京工商大学 经济学院,北京 100048;

3. 西南交通大学 经济管理学院,四川 成都 610031)

[摘 要]价格离散是衡量市场整合的重要指标,为全国统一大市场研究提供新的切入点。通过采集同一商品在不同平台商家的日度价格数据,对中国价格离散程度进行测度。结果表明:(1)中国线上市场的价格离散程度总体处于较高水平,且不同类别商品间存在显著异质性。基于此构建的高频价格离散指数(*iPDI*)发现,指数越低,说明市场统一程度越高,反之统一程度越低。(2)商品价格方差分解发现,价格离散主要源自跨平台竞争,并且构建实证模型,得出跨平台竞争对价格离散呈倒 U 型关系。(3)将价格离散的微观测度结果应用于宏观层面,基于价格离散特征编制的居民消费价格指数(*iCPI_pd*),更准确地反映商品价格的波动特征。

[关键词]价格离散;在线大数据;价格离散指数;跨平台竞争;居民消费指数

[中图分类号]F714.1 **[文献标志码]**A **[文章编号]**1004-4833(2025)03-0103-14

一、引言

党的二十大报告指出,“构建全国统一大市场,深化要素市场化改革,建设高标准市场体系”,着力“推动经济实现质的有效提升和量的合理增长”。以数据流通为支点助力构建全国统一大市场,通过优化要素配置、提升市场透明度,加速形成高效协同的市场体系,是深化要素市场化改革的重要路径^[1]。电商平台作为数据流通的核心载体,依托其大数据和算法优势,畅通生产端与消费端的信息对接,推动供应链和价值链的深度融合,成为实现经济高质量发展的重要着力点。

在此背景下,厘清影响电商平台效率的关键因素,对于推动数据要素高效流通,深化平台经济供需对接,加快构建高标准市场体系具有重要意义^[2]。理论上,相较于传统市场,以电商平台为代表的线上市场信息搜索成本更低,更接近完全竞争市场,市场价格应遵循“一定定律”^[3]。价格离散(Price Dispersion)作为衡量市场偏离完全竞争状态的关键指标,也是评估市场效率的重要工具。早期受限于样本因素,价格离散的微观研究相对缓慢,然而,电子商务的快速发展为价格离散研究拓宽了数据来源和提高了分析频率。线上市场数据能够提供更实时的价格信息,有助于捕捉价格的动态变化和市场反应速度,并且将其用于宏观经济指标的构建,能够弥补传统宏观模型在时间滞后方面的不足。因此,本文基于线上大数据,从商品价格角度识别全国统一大市场的构建情况,测度价格离散程度,厘清影响价格离散的主要因素,并挖掘价格离散在高频物价指标构建中的宏观应用价值。

本文的创新之处在于:第一,在研究数据上,本文采用高频微观价格大数据,这些数据覆盖国家统计局整个 CPI 篮子的商品和服务,为我们提供了一个全面、高频的视角来观察和分析中国商品价格的动态变化,更加细化地揭示价格离散的微观动态特性。第二,在研究方法上,本文将静态测度与动态测度相结合,观察商品价格的动态表现,并且构建高频价格离散指数(*iPDI*)。特别是,本文首次将价格离散运用到居民消费物价指数(CPI)的测度上,为实时反映市场统一程度提供了新工具。第三,在研究价值上,本文的价格离散测度结果为构建基于微观数据的宏观经济模型提供实证基础。通过深入分析价格离散与 CPI 之间的关系,增进对价格形成机制的认识,为理解和预测通货膨胀趋势提供新的工具和方法。

[收稿日期]2024-08-09

[基金项目]国家社会科学基金项目(23AJY018);国家自然科学基金项目(72302192)

[作者简介]秦杰(1993—),男,江苏泰州人,南京审计大学经济学院讲师,从事平台经济、大数据分析研究;胡林瑶(1988—),通信作者,女,安徽黄山人,北京工商大学经济学院师资博士后,从事平台经济、大数据分析研究,E-mail:lynn_0703@163.com;孟陆(1991—),男,天津人,西南交通大学经济管理学院教授,博士生导师,从事平台经济、大数据分析研究。

二、文献综述

自从1961年Stigler定义了价格离散,即相同商品在同一时期不同商家的价格分布^[4]。国内外学者基于此开展了大量研究工作。随着研究理论和方法不断完善,学者们逐渐认识到价格离散不仅存在于传统市场,还存在于线上市场。虽然互联网技术降低了信息搜索成本,但是同一商品价格依然偏离均衡价格,并未实现“一价定律”。为了更好地研究商品市场的价格离散情况,本文主要梳理价格离散的测度方法和影响因素的相关研究。

(一) 价格离散的测度方法

根据上述关于价格离散的基本事实,结合微观商品价格的可得性,学者们对价格离散进行了测度。通过对已有文献的梳理,主要得到五种价格离散的测度方法。

第一,通过价格标准差法计算价格离散^[5]。计算公式为 $\sigma = \sqrt{\frac{1}{N-1} \sum (p - \bar{p})^2}$,其中 σ^2 表示价格方差, p 表示线上商品价格, $\bar{p}$ 表示线上商品价格的平均值, N 表示样本总数。其优点在于能够充分利用价格数据,缺点是无法测度不同时间的价格离散程度。第二,采用变异系数衡量价格离散^[6]。计算公式为 $CV = \sigma/\mu$,其中 CV 表示变异系数, σ 表示标准差, μ 表示价格平均值。其优点在于能够跨时间测度价格离散,消除尺度和量纲的影响,但容易受到极端值的干扰,影响测度的精确性。第三,根据价格样本区间得到价格离散^[7]。计算公式为 $Range = p_{max}/p_{min}$,其中 $Range$ 表示价格区间, p_{max} 表示商品最高价格, p_{min} 表示商品最低价格。价格区间取决于商品最高价格和最低价格,价格区间越大,价格离散程度越高。此方法能够直观反映样本内价格的上下限差距,但其结果容易受到个别极值的过度影响,且无法细化至价格的内部分布。第四,通过最低价格之差计算价格离散。计算公式为 $G = p_{min,2} - p_{min,1}$,其中 G 表示最低价格之差, $p_{min,1}$ 表示最低的商品价格, $p_{min,2}$ 表示第二低的商品价格。优点在于克服了伯川德背景下商品价格非对称均衡的存在,但无法反映整体价格分布的离散程度,适用范围较为有限。第五,基于时间序列模型计算价格离散^[8]。此方法通过计算商品在不同时期价格的自回归系数,衡量价格差异的动态变化特征。其优点在于能够捕捉商品价格在外部冲击下的动态调整过程,反映价格差异的收敛或发散趋势。然而,该方法的缺点在于对样本数量和时间跨度的要求较高,易受数据质量的限制。以上离散指标中,第一至第四种方法反映商品在特定时间点上的价格分布特征,称之为静态方法;第五种方法考察处于波动过程中价格离散的动态表现,称之为动态方法。

(二) 价格离散的影响因素

随着互联网技术的发展,价格离散现象依然存在。价格分解法是分析这一现象的关键方法。通过对价格分解,能够将价格离散现象拆解为不同的组成部分,包括跨平台价格差异、同一平台上商家间的价格差异、随时间变化的商家内的价格差异、促销策略导致的价格差异四个部分,从而识别出价格差异的具体来源^[9]。基于此,本文将从这四个方面梳理价格离散的影响因素。

第一,跨平台价格差异。尽管互联网降低了信息获取的成本,但不同平台间的价格仍存在显著差异。这种差异不仅反映了平台之间的直接竞争,还与消费者在不同平台上感知的价格透明度和信息复杂度有关^[10]。第二,同一平台下商家间的价格差异。同一平台上的商家不同的成本结构、市场定位和库存管理策略,导致价格存在显著差异,并且平台的推荐算法和广告排序机制在无形中放大了商家间的价格差异,特别是在消费者行为受到平台干预的情况下,价格差异更为显著^[11]。第三,随时间变化的商家内价格差异。商家为实现利润最大化和有效管理库存,会根据市场需求、竞争对手的价格策略以及库存水平来调整价格,从而导致同一商品在不同时段的价格存在显著差异。随着微观数据的可获得性,商家利用大数据技术进行实时价格调整,导致同一商品在不同时间段的价格存在较大波动^[12]。第四,促销策略导致的价格差异。促销活动通过临时性降价、捆绑销售以及限时折扣等手段,明显改变了商品的价格分布。促销活动不仅直接影响了商品的即时价格,还通过改变消费者的价格预期影响了商家未来的定价策略^[13]。这些研究为理解价格离散进行了诸多探索,给后续研究提供了参考和借鉴。然而,现有研究从多种角度解释价格离散的影响因素,却无法确定当前影响价格离散的主要因素,导致这一研究领域众多理论“齐头并进”的局面。

据此,本文采用在线价格大数据,通过静态和动态两种视角测度价格离散程度,并将测度结果用于构建价格离散指数($iPDI$),同时,对商品价格进行方差分解,厘清和验证价格离散的最主要影响因素。进一步,将价格离

散结果首次用于居民消费物价指数(CPI)的估算,为建立基于微观数据的宏观经济模型提供经验证据。

三、数据、变量与指标

(一)数据来源

本文采用网页抓取技术,采集同一商品^①在各大电商平台的价格数据,其囊括国家统计局八大类商品的价格情况,具有一定的代表性。基本过程如下:首先,利用 Python 程序访问电商平台网站,获取商品名称、分类以及链接等信息,并将商品信息存储于 MySQL 数据库中;其次,调取数据库中同一商品链接,在比价网站上进行逐一匹配,获取商品在相应平台的历史价格;最后,通过 Python 程序对价格数据进行处理,测度价格离散程度。

(二)数据预处理

在测度价格离散之前,本文先对原始价格数据进行预处理。

1. 缺失值处理。本文采取对商品价格数据的连续观测。然而,偶发因素(如软件故障、信号中断等)导致部分价格信息缺失。为了解决这一问题,本文采用前值填补法,即在某一期价格缺失时,使用前一期的价格进行替代^[14]。

2. 异常值处理。在处理异常值时,借鉴 Cavallo 等的研究^[15],并考虑到突发卫生事件对线上商品价格的冲击,因此价格上涨超过 500% 或价格下降超过 90% 的数据视为异常值。考虑到本文中的异常值比例较低,但其存在可能会影响统计结果的可靠性,因此予以剔除。

3. 样本期处理。样本期是指商品价格信息从首次记录到最后一次记录所覆盖的时间长度。部分商品在观测期内存在下架等情况,不同商品的实际观测天数可能存在显著差异,导致无法全面反映价格的动态变化特征。因此,本文设定最低观测天数阈值为 30 天,即剔除观测天数少于 30 天的商品样本。

4. 名义价格处理。在处理名义价格时,本文定义其为不考虑促销策略的商品价格,而促销价格则是指商品在促销期间调整后的价格。具体方法是:如果某商品在某期间参与促销,将该期间的促销价格替换为促销前的名义价格,直至促销活动结束。

5. 销售商家处理。本文主要关注同一商品在不同商家中的销售情况。由于在数据采集过程中,同一商品所在电商平台的数量可能会发生变化,尤其是在样本期间内某商品仅在一个平台销售的情况,无法有效测度价格离散。因此,借鉴 Hitsch 等的处理^[9],本文要求在价格离散度测度中,同一商品需在两个或以上的商家进行销售。对于那些在样本期间内仅在一个商家销售的商品,予以剔除。

(三)数据描述

本文所用数据集来自电商平台 186853 种商品和服务的信息,时间跨度从 2020 年 4 月至 2022 年 5 月,共有 759 天 74032858 条观测值。本文将线上商品的网络分类结果与国家统计局八大类商品进行匹配^②。发现八大类商品数量占全部商品数量的比例存在差异,其中,生活类(30.13%)、食品类(25.57%)、教育类(23.87%)、其他类(13.53)占比较高,居住类(2.71%)、交通类(2.33%)、衣着类(1.40%)、医疗类(0.47%)类占比较低。但从绝对数量看,每个类别均包含了较多的商品数量,即便是商品数量较少的居住、交通、衣着和医疗类也至少有超过 7000 种商品,保证每类商品有着充足的样本观测值。总体上,本文数据集的主要特点是产品种类丰富、样本量大和时间频率高。经过对原始数据进行缺失值、异常值、样本期、促销处理及销售商家处理后,最终本文得到 110414 种商品,观测样本值为 33304031 个,见表 1。

表 1 商品数据描述

总观测值	74032858
预处理后观测值	33304031
产品数量	186853 种
预处理后产品数量	110414 种
起止日期	2020 年 4 月 28 日 - 2022 年 5 月 27 日
商品信息	商品名称、商品 ID、平台名称、商家名称、商品链接、商品类别等
价格信息	商品促销价格、优惠方式等

(四)价格离散衡量指标

本部分结合在线价格大数据,通过静态和动态两种方法,测度在线商品价格离散。

①本文将同一商品定义为商品在品牌层面和物理层面(如:重量、体积等)上完全相同。

②按照国家统计局分类将商品分为八类:食品烟酒、衣着、居住、生活用品及服务、交通和通信、教育文化和娱乐、医疗保健、其他用品和服务。为便于表述,本文简写为食品、衣着、居住、生活、交通、教育、医疗和其他。

1. 静态方法

本部分通过静态方法测度价格离散,以标准差和百分位比两个统计指标反映价格离散。具体过程如下:

- 第一步,计算属于 d 类^①的商品 i 在 t 期的价格离散。式(1)基于标准差方法计算价格离散,简称 SD 法,式(2)基于百分位比指标反映价格离散情况,简称 EV 法。

$$pd_sd_{dit} = \sqrt{\frac{1}{N_{dit} - 1} \sum_{s \in S_{it}} (\log(p_{dist}) - \overline{\log(p_{dist})})^2} \quad (1)$$

其中, pd_sd_{dit} 表示根据标准差指标,商品 i 在 t 期的价格离散程度, N_{dit} 表示 d 类商品 i 在 t 期的观测值, p_{dist} 表示商品 i 在商家 s 中 t 期的价格, S_{it} 表示商品 i 在 t 期的所有商家集合。

$$pd_ev_{dit} = p_{dit}(0.95)/p_{dit}(0.05) \quad (2)$$

其中, pd_ev_{dit} 表示根据百分位比指标,商品 i 在 t 期的价格离散程度, $p_{dit}(0.95)$ 表示商品 i 在 t 期第 95 百分位的价格, $p_{dit}(0.05)$ 表示商品 i 在 t 期第 5 百分位的价格。

第二步,计算 d 类的价格离散。由每个商品 i 到 d 类商品的价格离散,考虑到数据获取的有限性,无法得到商品在相应期间的销量,因此每种商品在 d 类中权重难以确定,在这一层面没有权重,统一采用算数平均法,计算出 d 类价格离散,即^②:

$$pdme_sd_d = \text{mean}(pd_sd_{dit}) \quad (3)$$

第三步,计算总体的价格离散。由 d 类商品到整体商品的价格离散,可采用算数平均法、中位数法、加权平均法,分别计算出整体的价格离散,其中,八大类的权重确定方法来自金雪军(2013)^③。即:

$$\begin{cases} pdme_sd_w = \text{mean}(pdme_sd_d) \\ pdmd_sd_w = \text{median}(pdme_sd_d) \\ pdwe_sd_w = \sum w_d \times pdme_sd_d \end{cases} \quad (4)$$

2. 动态方法

本部分通过动态方法观察价格差异在波动过程中的动态表现。如果商流活动活跃,则同一商品在不同平台的价格差异 q_{it}^k 应具有收敛特征。即便外生冲击拉大了价差,也会在短时间内恢复。具体而言,如果商家 i, j 均在线销售同一商品,则商家间的套利行为就会使价差实现收敛;如果目前商家 i 在平台 x 调整价格,那么价格调整信息也会传递给同一商品的其他平台商家, q_{it}^k 不会一直处于较高水平。相反,如果商家的套利行为不活跃甚至市场完全分割,那么价格差异 q_{it}^k 将表现为发散特征,在受到外生冲击后还会表现出放大趋势。可以看出,相较于静态方法通过指标衡量价格离散,动态方法综合考虑了平台商家间的交易效率和套利行为,反映出价格离散的动态特征。具体过程如下:

第一步,计算属于 d 类的商品 i 在 t 期的价格离散。借鉴张昊的动态测度方法^[8],并且考虑到线上商品价格的特征,本文对商品 i 在 t 期不同商家下的价格计算其方差,以便于与现有文献结果进行比较,即:

$$pd_var_{dit} = \text{var}(p_{di1t}, p_{di2t}, \dots, p_{dint}) \quad (5)$$

其中, pd_var_{dit} 表示商品 i 在 t 期的价格离散程度, p_{dist} 表示商品 i 在商家 s 中 t 期的价格 ($s = 1, 2, \dots, n$)。

第二步,计算属于 d 类的商品 i 的价格差异波动特征。使用时间序列回归模型计算的 pd_ar1_{di} 和 pd_sar_{di} 来衡量^④,即:

$$pd_var_{dit} = \mu + \gamma_1 pd_var_{di,t-1} + \varepsilon_t \quad (6)$$

$$pd_var_{dit} = \mu + \gamma_1 pd_var_{di,t-1} + \dots + \gamma_p pd_var_{di,t-p} + \varepsilon_t \quad (7)$$

其中,式(6)表示 pd_var_{dit} 服从 $AR(1)$ 过程, $pd_ar1_{di} = \gamma_1$, 式(7)表示 pd_var_{dit} 服从 $AR(p)$ 过程^⑤, $pd_sar_{di} = \sum_{i=1}^p \gamma_i$ 。

①按照国家统计局分类的八类商品:食品、衣着、居住、生活、交通、教育、医疗以及其他。

②篇幅原因,正文反报告标准差方法的结果,百分比的详细结果可向作者索取,下同。

③受限于价格数据内容,本文无法获取商品在每期的销售数量作为权重,文章以八大类商品的人均消费支出占比作为八大类商品权重的替代变量以权重估算方法作为本文样本期内的分类权重。

④假定 pd_var_{dit} 服从 $AR(1)$ 和 $AR(P)$ 过程。

⑤ p 取 1-5 阶,最优滞后阶数根据 BIC 准则确定。

第三步,计算 d 类的价格差异波动特征。考虑到数据获取的有限性,采用算术平均法,观察 d 类价格离散,即^①:

$$pdme_ar1_d = mean(pd_ar1_{di}) \quad (8)$$

$$pdme_sar_d = mean(pd_sar_{di}) \quad (9)$$

第四步,计算总体的价格离散。由 d 类商品到整体商品的价格离散,可采用算术平均法、中位数法、加权平均法^②,分别计算出整体的价格离散,即:

$$\begin{cases} pdme_ar1_w = mean(pdme_ar1_d) \\ pdmd_ar1_w = median(pdme_ar1_d) \\ pdwe_ar1_w = \sum w_d \times pdme_ar1_d \end{cases} \quad (10)$$

3. 价格离散指数

作为反映市场统一和市场效率的重要指标,价格离散的实时化和高频化程度较低。目前,电子商务的快速发展和微观价格数据的获取为价格离散指标的构建提供了可行性。据此,本部分将构建基于互联网在线数据的价格离散指数(internet-based Price Dispersion Index),简称 $iPDI$ 。借鉴刘涛雄的指数设计思路^[16], $iPDI$ 的主要编制过程分为四步,具体如下:

第一步,计算在线商品 i 在 t 期的价格离散变化比。等于 t 期的价格离散 dp_i 除以 $t-1$ 期的价格 dp_{i-1} ^③。

$$dpr_{di} = dp_{di} / dp_{di-1} \quad (11)$$

其中, dpr_{di} 表示属于 d 类商品 i 在 t 期的价格离散变化比。 dp_{di} 和 dp_{di-1} 分别表示商品 i 在 t 期和 $t-1$ 期的价格离散程度。

第二步,计算 d 类在 t 期的价格离散变化比。由每个商品到 d 类商品的价格离散,考虑到数据获取的有限性,在这一层次采用非加权的平均计算方法。考虑指数检验功效和商品替代效果等问题,本文采用 Jevons 几何指数公式计算 d 类的价格离散指数。

$$dpr_{dt} = \sqrt[n]{dpr_{d1t} \times dpr_{d2t} \cdots dpr_{dnt}} \quad (12)$$

其中, dpr_{dt} 表示属于 d 类商品在 t 期的价格离散变化比。

第三步,计算整体在线商品的价格离散 $iPDI$ 。由 d 类商品到整体商品的价格离散,其中八大类的权重数据来自八大类商品的人均消费支出占比。

$$iPDI_t = \sum w_d \times dpr_{dt} \quad (16)$$

四、价格离散的测度

本部分基于微观商品价格数据对价格离散进行测度。参考张昊的研究方法^[8],本文分别采用静态和动态两种方法判断商品价格的离散程度。同时,构建价格离散指数($iPDI$),为价格离散程度提供更为直观的指标。

(一) 价格离散测度:静态视角

本部分结合在线价格大数据测度线上商品价格离散,分别得到总体和八大类的价格离散指标,具体见表 2 和表 3。

表 2 报告了总体价格离散的测算结果。一方面,基于 SD 法计算价格离散时,当商品价格包含促销,通过中位数法、算术平均法、加权平均法计算得到总体商品的价格离散程度分别是 0.149、0.262、0.220,由结果可知,对于 95% 商品价格,包含整体商品平均价格的抽样误差为 $\pm 29\%$ 。可以看出,总体价格离散程度较大,也就是同一商品在不同商家之间存在较大差异。进一步与现有研究对比,本文发现与 Hitsch 等发现的“美国线上商品价格的离散程度为 0.161(中位数法)、0.163(算术平均法)”有相似之处^[9]:一方面,测度方法一致性。两者均基于

表 2 价格离散测算结果:总体

算法	包含促销		剔除促销	
	SD 法	EV 法	SD 法	EV 法
中位数法	0.149	1.272	0.148	1.269
算数平均	0.262	1.957	0.260	1.954
加权平均	0.220	1.632	0.218	1.626

数据来源:作者根据测算结果,自行整理。

①正文仅报告 $AR(1)$ 的结果, $AR(2)$ 的详细结果可向作者索要,下同。

②与静态方法相同,以八大类商品的人均消费支出占比作为权重的替代变量。

③利用标准差法计算得到商品在 t 期的商品 i 在 t 期的价格离散。

静态视角,采用标准差和对数价格标准差等指标,保证了结果的可比性,并且两者测度结果中,价格离散的中位数均小于平均值,说明测度结果存在极大值。第二,商品结构相似性。研究样本均涵盖快速消费品和日用消费品等核心品类,商品属性的重叠增强了测算结果的适用性。为了进一步研究线上商品的价格离散情况,将商品按照国家统计局标准进行分类,结果见表3。可以看出,八大类商品与总体商品的价格离散情况相似,也就是八大类商品的价格离散程度较大。其中,基于SD算法下,无论是包含促销还是剔除促销情况,教育类商品价格离散程度最大,两者结果均为0.317。居住类商品价格离散程度最小,分别是0.184和0.186。同时,基于EV法的价格离散结果验证了这一结果。

同样,基于SD法对线上商品按照八大类来观察价格离散的分布情况^①,可以看出:一方面,八大类商品的价格离散均呈现右偏分布,与整体结果相同。另一方面,结合偏度系数^②,教育、生活以及食品类商品呈现明显的右端拖尾形态,表明在这三类商品中,价格离散出现极大值的可能性较高,与整体表现较为相似。此外,基于EV法得到八大类商品的价格离散分布情况,验证了上述结论,也就是对于八大类商品,价格离散程度是较大的。本文通过静态法测度价格离散发现,中国线上市场的价格离散程度是较高的,不同类别商品的价格离散程度存在异质性。

(二) 价格离散测度:动态视角

从总体上看,中国线上价格离散程度较高。表4是基于动态法的价格离散估计结果。在包含促销的情况下,通过中位数法、算术平均法、加权平均法得到AR(1)系数分别为0.940、0.852和0.825,表明商品价格离散处于较高水平。剔除促销后,总体价格水平的AR(1)系数均得到提高,分别为0.968、0.894和0.877,说明价格调整对线上商品的价格离散起到一定程度的抑制作用,反映出价格粘性对价格离散的动态过程有着重要影响。SAR系数结果也证明了这一点。

从八大类来看,基于不同估算方法得到的各大类商品的离散程度均存在较大差异,但是剔除促销后的离散指数均一致地高于包含促销的情况。基于中位数法,八大类商品包含促销的AR(1)系数在0.867至0.982之间,其中,其他类商品的自回归系数最小,说明价格收敛特征明显,离散程度较小,而衣着类商品的自回归系数最大,价格离散程度较大。基于算术平均法,结果也基本相同^③。

基于动态方法,本文得出:一方面,价格离散程度总体上处于较高的水平,并且不同类别商品的价格离散有着异质性,与静态法结果相似。另一方面,在长期动态过程中,促销在一定程度上削弱了价格离散程度,其可能原因:促销增加了消费者对商品价格的敏感性,使得促销后的商品价格趋于相同^[17]。

(三) 价格离散指数

随着大数据技术的发展,宏观经济指标开始具备高频化和实时化的特征。高频化和实时化的结合能够更加

表3 价格离散测算结果:八大类

商品列表	包含促销		剔除促销	
	SD法	EV法	SD法	EV法
食品	0.203	1.499	0.194	1.476
衣着	0.311	1.649	0.309	1.644
居住	0.184	1.443	0.186	1.445
生活	0.247	1.731	0.247	1.730
交通	0.205	1.583	0.206	1.583
教育	0.317	2.579	0.317	2.581
医疗	0.210	1.544	0.210	1.547
其他	0.217	1.532	0.208	1.518

数据来源:作者根据测算结果,自行整理。

表4 AR和SAR计算结果

算法	商品类别	包含促销		剔除促销	
		AR(1)	SAR	AR(1)	SAR
中位数法	整体	0.940	0.941	0.968	0.965
	食品	0.871	0.906	0.945	0.945
	衣着	0.982	0.977	0.987	0.991
	居住	0.902	0.908	0.942	0.946
	生活	0.928	0.933	0.965	0.962
	交通	0.914	0.932	0.946	0.956
	教育	0.969	0.963	0.977	0.976
	医疗	0.914	0.917	0.94	0.955
	其他	0.867	0.898	0.943	0.946
	其他	0.867	0.898	0.943	0.946
算术平均	整体	0.852	0.884	0.894	0.912
	食品	0.778	0.841	0.858	0.889
	衣着	0.966	0.957	0.973	0.975
	居住	0.812	0.848	0.86	0.887
	生活	0.842	0.877	0.891	0.91
	交通	0.834	0.872	0.882	0.905
	教育	0.898	0.916	0.918	0.929
	医疗	0.836	0.859	0.865	0.887
	其他	0.755	0.82	0.836	0.86
	其他	0.755	0.82	0.836	0.86
加权平均	整体	0.825	0.864	0.877	0.900

数据来源:作者根据测算结果,自行整理。

①基于SD法对八大类商品价格离散分布图可向作者索取。
②该偏度系数为包含促销价格情况计算得出,高到低依次是:教育、生活、食品、其他、居住、交通、医疗以及衣着,分别是:33.996、19.112、14.150、9.205、7.329、7.120、4.981、1.565。
③利用SAR系数也可以得到同样的结论。

清晰地观测市场行为,更好地放大宏观经济。高频实时的价格离散指标能够更好地观测市场运行情况以及衡量全国统一大市场的构建程度。目前尚未关于价格离散的相关高频实时指标,因此,本文将观察 2020 年 4 月 28 日至 2022 年 5 月 27 日的 *iPDI* 的表现,并且本文中编制的指数都是环比指数,即以“上一期(上日/上周/上月) $=100$ ”,在分析中采用指数环比变化率(见图 1^①)。

图 1 显示了整体价格离散指数的走势。可以看出:价格离散程度是存在的,随着时间推移而不断变化,指数越大,说明当期价格离散程度越高。相较于剔除促销,当商品价格包含促销时,价格离散的变化幅度进一步扩大,尤其在电商节日(6.18 和 11.11)期间,平台商家的促销行为进一步加深了价格离散程度,并且,在电商节日后指数的迅速回落,说明促销对价格离散影响的“短暂性”,例如,价格离散指数在 2020 年 11 月 11 日为 103.74,而在 2020 年 11 月 12 日为 87.41。在构建统一大市场的背景下,理想状态下的价格离散指数应当相对较低且稳定,这意味着市场中的商品价格在平台之间趋于一致,体现出较高的市场整合度和价格传导效率。然而,从价格离散指数波动来看,市场尚未实现充分的一体化,尤其是在电商节日期间,价格差异显著扩大,这表明市场整合度尚有提升空间。

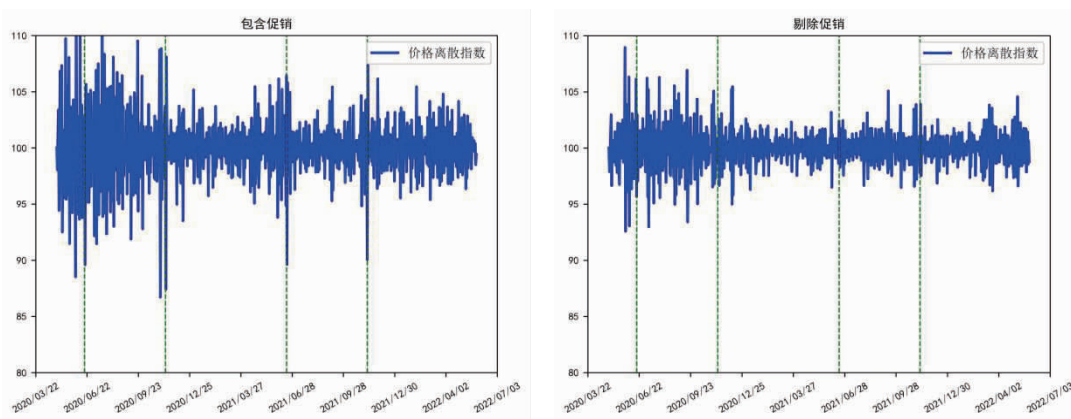


图 1 整体价格离散指数

同时,对商品进行分类来观察价格离散情况^②,可以看出:食品、衣着、居住、生活、交通、教育、医疗以及其他八类商品的价格离散与整体情况一致。八大类商品受促销策略的影响,价格离散程度不断变化,并且在电商节日的催化下,价格离散程度进一步提高,其中,根据价格离散指数的变化幅度^③由高到低依次是食品、医疗、居住、其他、生活、教育、交通以及衣着类商品。同样,八类商品的价格离散指数的波动反映市场中存在较大的差异,市场尚未达到完全统一。

以上可知,基于当期离散测度结果构建的价格离散指数(*iPDI*),不仅更直观地反映出同一商品的差异程度,即价格离散指数越大,价格差异程度越高,还从价格视角衡量全国统一大市场的构建程度,市场统一程度越高,同一商品价格越趋于一致。

五、价格离散的方差分解与实证检验

由上文可知,价格离散程度存在较大差异,并且在促销策略的影响下,价格离散程度会进一步扩大。接下来,本文将量化这些来源对整体价格离散的相对贡献。

(一) 价格方差分解

第一步,计算商品 *i* 的价格方差分解。本文将名义价格和促销价格纳入方差分析中,区分平台商家的促销

①为了观察价格离散指数在电商节日的情况,图 1 中的绿色虚线对应的日期依次是:2020 年 6 月 18 日、2020 年 11 月 11 日、2021 年 6 月 18 日、2022 年 11 月 11 日。

②篇幅原因,按八大类商品的价格指数分布可向作者索要。

③计算 2020 年 6 月 18 日、2020 年 11 月 11 日、2021 年 6 月 18 日、2022 年 11 月 11 日四个电商节日当天离散指数的变化幅度,取算数平均,得出对应类的平均变化幅度。

策略带来的方差贡献。通过价格方差分解^①,整体价格离散由跨平台价格差异($ampd_{di}$)、同一平台下商家间的价格差异($aspd_{di}$)、随时间变化的商家内的价格差异($wspd_{di}$)、商品促销折扣的价格差异(pdd_{di})以及商品促销策略差异(psd_{di})五部分组成,即:

$$\begin{aligned} \text{var}(p_{msit}) &= \text{var}(\bar{p}_{mi}) \text{ (跨平台促销价格差异)} \\ &+ \frac{1}{N} \sum_{m \in M} N_m \text{var}(\bar{b}_{si} | m) \text{ (商家间名义价格差异)} \\ &+ \frac{1}{N} \sum_{s \in S} N_s \text{var}(b_{msit} | s) \text{ (商家内名义价格差异)} \\ &+ \frac{1}{N} \sum_{m \in M} N_m \text{var}(b_{msit} - p_{msit} | m) \text{ (商品促销折扣差异)} \\ &- 2 \frac{1}{N} \sum_{m \in M} N_m \text{cov}(b_{msit} - p_{msit}, b_{msit} | m) \text{ (商品促销策略差异)} \end{aligned} \quad (14)$$

其中, $\text{var}(\bar{p}_{mi})$ 表示各平台之间的平均促销价格方差, $\text{var}(\bar{b}_{si} | m)$ 表示商家之间的平均名义价格方差, $\text{var}(b_{msit} | s)$ 表示商家内部随时间变化的名义价格方差, $\text{var}(b_{msit} - p_{msit} | m)$ 表示商家之间名义价格与促销价格差额的方差, $\text{cov}(b_{msit} - p_{msit}, b_{msit} | m)$ 表示商家之间名义价格与促销价格差额和名义价格的协方差, N 表示所有样本观测值, N_m 表示平台 m 下的样本观测值, N_s 表示商家 s 的样本观测值, M 表示所有平台的集合($m \in M$), S 表示所有商家的集合($s \in S$), $\bar{p}_{mi}$ 表示平台 m 的平均促销价格, $\bar{b}_{si}$ 表示商家 s 的平均名义价格, p_{mist} 、 b_{mist} 分别表示属于平台 m 下商家 s 的商品 i 在 t 期促销价格和名义价格。

第二步,计算 d 类的价格方差分解。由于商品 i 在 d 类中权重难以确定,因此对商品 i 在平台间价格差异($ampd_{di}$)、同一平台下商家间的价格差异($aspd_{di}$)以及随时间变化的商家内的价格差异($wspd_{di}$)这三部分的结果,采用算术平均法计算 d 类方差分解结果,即:

$$\begin{cases} ampd_d = \text{mean}(ampd_{di}) \\ aspd_d = \text{mean}(aspd_{di}) \\ wspd_d = \text{mean}(wspd_{di}) \end{cases} \quad (15)$$

第三步,计算总体的价格方差分解。由 d 类商品到整体商品的价格离散,采用加权平均法,计算出整体的价格方差分解,即:

$$\begin{cases} ampd = \sum w_d \times ampd_d \\ aspd = \sum w_d \times aspd_d \\ wspd = \sum w_d \times wspd_d \end{cases} \quad (16)$$

由表5可知,平台间价格差异对整体价格离散影响最大,为88.87%。同样,随时间变化的价格差异和商家间价格差异位于第二、第三名,分别为8.49%和2.22%。商品促销折扣差异占整体价格离散的1.3%,促销策略差异仅占-0.87%。与现有研究结果对比,Hitsch等发现“美国的平台间价格差、同一平台下商家间的价格差异、随时间变化的商家内的价格差异、商品促销折扣的价格差异以及商品促销策略差异的占比分别是32.7%、31.3%、12.3%、36%、-12.3%^[9],相比较而言,中国商品价格离散的来源主要来自跨平台的价格差异。

由上文可知,跨平台的价格差异是影响价格离散的最主要因素,也就是跨平台竞争对价格离散度有着显著影响。处于初期发展阶段的平台通常会采取降低抽成、促销和补贴等策略,以快速积累用户基础。随着用户群体的扩大,平台逐渐通过交叉网络外部性吸引更多用户,解决早期市场进入的“鸡蛋相生”问题^[18]。在这一阶段,平台数量的增加导致价格离散度上升,因为每个平台都在采取激进的价格策略以抢占市场份额。然而,当平

表5 价格方差:基本分解

价格差异	方差占比(%)
跨平台价格差异	88.87
商家间价格差异	2.22
商家内价格差异	8.49
商品促销折扣差异	1.30
商品促销策略差异	-0.87

数据来源:作者根据测算结果,自行整理。

①篇幅原因,价格方差分解的详细推导可向作者索要。

台企业从生存目标转向盈利目标时,其价格策略会发生变化。许多平台开始模仿市场中占主导地位的平台,价格竞争逐渐趋于一致,导致价格离散度下降。此外,由于促销和价格战具有短期效应,平台在长期内会选择放弃低价竞争,转而制定与竞争对手相近的价格区间。这意味着价格离散度在平台数量增加到一定程度后会开始下降。因此,本文认为跨平台的价格差异与价格离散度之间存在“倒 U 型”关系。

(二) 模型设计

在此基础上,本文通过构建商品层面的固定效应面板回归模型,分析跨平台竞争对价格离散的影响,也就是平台数目对价格离散的影响^[19]。基准回归模型如下:

$$pd_{it} = \alpha_1 \times pf_{it} + \alpha_2 \times pf_{it}^2 + \omega \times X_{it} + \sigma_i + \theta_t + \varepsilon_{it} \quad (17)$$

其中, pd_{it} 表示商品 i 在 t 期的价格离散, pf_{it} 表示商品 i 在 t 期的平台数目, X_{it} 表示控制变量,包括价格粘性(stick)、注意力配置(alloc)、突发卫生事件(emerg)。 σ_i 表示商品固定效应, θ_t 表示时间固定效应^①, ε_{it} 表示随机误差。 α_1 、 α_2 、 ω 为回归系数。

进一步,为检验平台势力在平台数目对价格离散影响中的调节效应,本文在模型(22)中引入平台数目($influ_{it}$)及其平方项与平台势力的交互项($pf_{it}^2 \times influ_{it}$),见式(23)。

$$pd_{it} = \alpha_1 \times pf_{it} + \alpha_2 \times pf_{it}^2 + \beta_1 \times influ_{it} + \beta_2 \times influ_{it} \times pf_{it}^2 + \beta_3 \times influ_{it} \times pf_{it} + \omega \times X_{it} + \sigma_i + \theta_t + \varepsilon_{it} \quad (18)$$

其中, pd_{it} 表示商品 i 在 t 期的价格离散, pf_{it} 表示商品 i 在 t 期的平台数目, X_{it} 表示控制变量,包括价格粘性、注意力配置、突发卫生事件。 σ_i 表示商品固定效应, θ_t 表示时间固定效应, ε_{it} 表示随机误差。 α_1 、 α_2 、 ω 为回归系数。

(三) 主要变量说明

1. 价格离散(pd)

本文将标准差作为衡量价格离散的统计指标。具体地,标准差法对价格均值附近的波动尤为敏感,能够有效捕捉价格波动的细微变化。考虑到本文选取在多平台(大于两个)销售的商品,并且同一商品在不同平台的价格差异通常较小,因此,本文主要使用标准差法来衡量价格离散。

$$pd_{it} = \sqrt{\frac{1}{N_{it} - 1} \sum_{s \in S_{it}} (\log(p_{ist}) - \overline{\log(p_{it})})^2} \quad (19)$$

其中, i 、 t 分别表示商品和日期,被解释变量 pd_{it} 表示在 t 期商品 i 的价格离散, N_{it} 表示在 t 期商品 i 的观测值, p_{it} 表示在 t 期商品 i 的价格,考虑到促销方式,商品价格分为包含和剔除促销两种情况, S_{it} 表示在 t 期商品 i 的平台商家集合。

2. 平台数目(pf)

本文通过同一商品在售平台的数量衡量平台数目,见公式(20)。考虑到同一商品在同一平台上通常有多个网店售卖,网店间不同的定价策略会影响价格离散度。消费者在购买决策时,受消费习惯、商家信誉(如销量、评价)等因素影响,即使非官方旗舰店也能有较高销量。因此,本文考虑同一平台上不同商家的定价行为。

$$pf_{it} = \text{count}(\text{platform}_{it}) \quad (20)$$

其中, pf_{it} 表示在 t 期商品 i 的平台数目, platform_{it} 是虚拟变量,如果商品 i 在 t 期处在平台商家,则取值为1,否则为0。

3. 平台势力(influ)

本文根据同一商品在售大电商平台^②数量与所有电商平台的比率来衡量平台势力。当大电商平台占比越大,说明对市场的控制程度越高,平台势力越大^[20]。

$$influ_{it} = \text{big_pf}_{it} / \text{all_pf}_{it} \quad (21)$$

其中, big_pf_{it} 表示商品 i 在 t 期所在大电商平台的数目, all_pf_{it} 表示商品 i 在 t 期所处的全部平台数目。

(四) 实证分析

1. 基准回归

在模型估计方法的选取上,基于 Hausman 检验的结果,在 1% 的显著性水平下拒绝原假设,从而选择固定效

①本文采用从研究起始日到指定日期的天数来衡量时间固定效应,吸收价格离散随时间变化的线性趋势。

②本文依据样本期内的平台所占市场比重,将天猫商城、京东商城以及苏宁易购定义为大电商平台。

应(FE)模型进行回归分析。此外,本文同时考虑包含促销和剔除促销的情况,以便于对不同情况下的结果进行比较和分析。

表6第(1)列显示平台数目的平方项(pf^2)回归系数为-0.058,且通过了1%水平上的显著性检验。第(2)列)引入控制变量(X)后,提高了拟合优度 R^2 。同样地,平台数目的平方项显著为负,拐点为3.652,说明平台数目与价格离散之间存在显著倒U型关系。这意味着,当平台数目较少时,会对线上商品的价格离散有着正向影响,此时平台数目越多,价格离散程度越高,而当平台数目跨过临界点3.652后,才对价格离散起到抑制作用。此外,由表6第(3)和(4)可知,当商品价格剔除促销时,平台数目对价格离散依旧保持显著倒U关系,并且拐点值为3.682,高于包含促销的拐点值3.652,说明在剔除促销的情况下,平台数目与价格离散的倒U关系对平台数目更不敏感,也就是新平台的进入并不能更好地缓解同一商品在不同平台的价格差异情况,在一定程度上反映出促销在平台数目与价格离散关系中发挥重要作用,加快了平台数目对价格离散的负向影响。

2. 内生性检验

为解决内生性问题,本文以社会对垄断的注意力配置作为跨平台价格差异的工具变量。主要过程如下:以关键词“反垄断”和“反垄断法”作为统计对象,获取其在样本期间的日度频数,数据来源于百度指数(<https://index.baidu.com>)。考虑到社会对反垄断的关注度可能会影响平台的竞争行为。更高的关注度可能导致政府更严格的监管和更多的反垄断执法,从而影响平台的市场行为,减少价格差异^[21]。而社会对反垄断的关注度本身可能与价格离散没有直接关系,而是通过影响平台竞争间接影响价格离散。因此,选用社会对垄断的注意力配置作为平台竞争的工具变量是合理的。

借鉴李芳等的研究^[22],本文采用两阶段最小二乘法(2SLS)进行工具变量回归,表7的第(1)列和第(2)列是考虑促销和剔除促销两种情况的回归结果。一方面,Anderson LM统计量均在1%水平上拒绝了工具变量识别不足这一原假设。另一方面,Cragg-Donald Wald F统计量分别为214.486,高于10%水平临界值7.03,说明本文工具变量有着较强的解释力,不存在弱工具变量问题。列(1)和(2)显示平台数目平方项的回归系数显著为负,表明符合相关性要求。以上回归结果表明,平台数目对价格离散具有显著的倒U型影响。

3. 机制检验

由上文可知,平台数目对价格离散的影响呈倒U型,这一现象是基于交叉网络外部性的平台势力作用的结果,也就是平台通过连接不同用户群体(如消费者与商家),利用交叉网络外部性,一方用户的效用随另一方用户规模增加而提升的特性,形成对市场的控制能力。具体而言,在平台发展的初期阶段,市场处于增量竞争状态^①,平台的首要目标是迅速扩大用户规模以突破最低网络规模^[23]。在交叉网络外部性的影响下,平台通过高

表6 平台数目与价格离散

变量	包含促销		剔除促销	
	(1)	(2)	(3)	(4)
平台数目×平台数目 (pf^2)	-0.058 *** (0.000)	-0.056 *** (0.000)	-0.058 *** (0.000)	-0.055 *** (0.000)
平台数目(pf)	0.453 *** (0.000)	0.409 *** (0.000)	0.452 *** (0.000)	0.405 *** (0.000)
价格粘性($stick$)	—	0.017 *** (0.000)	—	0.018 *** (0.000)
注意力配置 ($alloc$)	—	-0.0004 *** (0.000)	—	-0.0004 *** (0.000)
突发卫生事件 ($emerg$)	—	0.001 *** (0.003)	—	0.001 *** (0.001)
样本量	12338950	12338950	12338950	12338950
组内 R^2	0.134	0.144	0.135	0.146

数据来源:作者根据测算结果,自行整理。*、**、*** 分别代表10%、5%、1%水平上显著。下同。

表7 内生性检验

变量	(1)	(2)
	包含促销	剔除促销
平台数目×平台数目(pf^2)	-0.278 *** (0.000)	-0.248 *** (0.000)
平台数目(pf)	1.379 *** (0.000)	1.243 *** (0.000)
价格粘性($stick$)	0.021 *** (0.000)	0.021 *** (0.000)
注意力配置($alloc$)	0.001 *** (0.000)	0.0004 *** (0.000)
突发卫生事件($emerg$)	0.001 *** (0.000)	0.001 *** (0.000)
样本量	12338950	12338950
Anderson LM 统计量 P-value	608.382 (0.000)	608.406 (0.000)
Cragg-Donald Wald F 统计量 10% maximal IV	214.486 7.03	214.498 7.03

数据来源:作者根据测算结果,自行整理。

①参考曲创和王夕琛的研究,增量竞争表示市场处于扩张阶段,各平台通过争夺新增用户实现规模增长,竞争主要围绕吸引新用户展开。存量竞争表示市场趋于饱和,用户规模接近上限,各平台通过争夺现有用户资源维持竞争优势,竞争集中在抢占已有用户上。

额补贴和低价促销吸引用户,商品价格因促销力度和商家定价策略的差异而表现出较大的差异^[24]。该阶段的用户竞争策略符合平台发展规律,此时平台势力在充分竞争的市场中表现对资源分配和市场规则的深度参与。同时,不同平台之间为了争夺用户展开激烈竞争,差异化的定价策略进一步加剧了价格离散的现象。因此,平台势力在这一阶段推动了价格离散水平的上升。

随着市场逐步进入成熟阶段,平台数目增加,用户数量接近上限,平台竞争由增量竞争转向存量竞争。此时,平台市场逐渐形成垄断结构,垄断平台凭借基于交叉网络外部性的平台势力,利用其规模效应和数据优势巩固市场地位,并通过数据和算法引导其他平台采用跟随定价行为,削弱了其差异化定价能力,抑制价格充分竞争^[25]。垄断平台往往通过统一定价策略削弱商家的定价灵活性,使市场价格趋于一致,价格离散水平随之显著下降。因此,平台势力对平台竞争与价格离散两者关系具有倒 U 型调节作用。

表 8 第(1)列表明,在包含促销的情况下,分别控制平台数量及其平方项与平台势力的交互项之后,一方面,平台数目与价格离散仍保持倒 U 型关系,即平台数目的平方项(pf^2)显著为负。另一方面,平台数目平方项与平台势力交互项($pf^2 \times influ$)的系数在 1% 水平下显著为负。这一结果表明,平台势力在平台数目与价格离散之间具有倒 U 型调节作用,当平台势力较低时,平台数目对价格离散的促进作用会增强;反之,当平台势力较高时,削弱了两者的正反馈关系。同样,在剔除促销情况下,平台势力对平台数目与价格离散之间具有倒 U 型调节作用,保证结果的稳健性。

4. 稳健性检验

为了检验之前估计结果的稳健性,本文从变量定义进行讨论。在价格离散程度的测量中,前文采用了标准差法,并利用该结果分析了平台数量对价格离散的影响。接下来,本文将对价格离散进行替换,使用极值法计算的价格离散度进行稳健性检验。

由表 9 可知,平台数目的平方项系数显著为负,表明平台数量对价格离散具有显著的倒 U 型影响。各变量系数的估计值与表 6 的结果基本一致,表明前述估计结果具有较高的稳健性和准确性。

六、进一步讨论

价格离散显著影响居民消费物价指数(CPI)的估算与实际应用。目前对于 CPI 的计算,通常假定某一时期内每种商品价格为单一值^[26]。然而,现实中商品价格往往存在离散分布,特别是在不同平台和商家间的差异。因此,借鉴国家统计局编制 CPI 的理论方法,本文提出一种基于价格离散特征的高频物价指数 $iCPI_{pd}$,以期更准确反映消费者的支出水平。

(一)基于价格离散的居民消费指数构建

居民消费物价指数(CPI)是衡量一篮子商品和服务价格随时间变化的经济指标,反映居民消费水平的变化^①。其计算公式如下:

①CPI 由国家统计局负责编制和发布,通过监测一篮子代表性消费品和服务的价格变动,反映居民消费支出的总体价格水平,见网站 <https://data.stats.gov.cn/easyquery.htm?cn=A01>。

表 8 机制检验

变量	包含促销 (1)	剔除促销 (2)
平台数目 ² (pf^2)	-0.091 *** (0.000)	-0.090 *** (0.000)
平台数目(pf)	0.542 *** (0.000)	0.537 *** (0.000)
平台势力($influ$)	0.181 *** (0.000)	0.183 *** (0.000)
平台数目×平台势力 ($pf \times influ$)	-0.120 *** (0.000)	-0.119 *** (0.000)
平台数目 ² ×平台势力 ($pf^2 \times mc$)	-0.167 *** (0.000)	-0.166 *** (0.000)
价格粘性($stick$)	0.017 *** (0.000)	0.018 *** (0.000)
注意力配置($alloc$)	-0.001 *** (0.000)	-0.001 *** (0.000)
突发卫生事件($emerg$)	0.001 *** (0.000)	0.002 *** (0.000)
样本量	12338950	12338950
组内 R ²	0.154	0.157

数据来源:作者根据测算结果,自行整理。

表 9 稳健性检验

变量	(1) 包含促销	(2) 剔除促销
平台数目×平台数目 (pf^2)	-0.911 ** (0.028)	-0.847 ** (0.040)
平台数目(pf)	4.881 *** (0.006)	4.601 *** (0.009)
价格粘性($stick$)	0.389 *** (0.000)	0.403 *** (0.000)
注意力配置($alloc$)	0.004 ** (0.012)	0.004 ** (0.020)
突发卫生事件($emerg$)	0.016 ** (0.024)	0.016 ** (0.024)
样本量	12338950	12338950
Anderson LM 统计量	608.382	608.406
P-value	(0.000)	(0.000)
Cragg-Donald Wald F 统计量	214.486	214.498
10% maximal IV	7.03	7.03

数据来源:作者根据测算结果,自行整理。

CPI_t = \frac{\sum_{i=1}^N w_i \times p_{it}}{\sum_{i=1}^N w_i \times p_{i0}} \times 100 \tag{22}

其中, w_i 为商品 i 的权重, p_{it} 和 p_{i0} 表示商品 i 在 t 期和基期的价格。
在实际中,由于平台竞争、促销策略等因素,同一商品价格呈现离散分布。式(22)难以准确反映消费者实际支出情况。为此,本文假设商品 i 的价格为随机变量,其概率分布为 $f(p_{it})$,商品 i 的期望价格表示为:

E(p_{it}) = \int p_{it} \times f(p_{it}) dp \tag{23}

借鉴 Fava 的研究思路,商品期望价格 $E(p_{it})$ 可简化为:

E(p_{it}) = \bar{p}_{it} - k \times pd_{it} \tag{24}

其中, $\bar{p}_{it}$ 表示商品 i 在 t 期的平均价格, pd_{it} 表示商品 i 在 t 期的价格离散, k 表示价格离散程度对期望价格的修正系数,设为 0.5,其原因:一方面,消费者的低价偏好促使他们在面对价格离散时寻求成本效益^[39];另一方面,对低价商品质量的担忧限制了这种偏好,过低的价格可能会触发对商品质量的质疑^[40]。因此,本文采用中点估计,将 k 设为 0.5,用于平衡消费者对价格的敏感性和对质量的中性预期。

接下来,本文为减少价格离散带来的系统性偏差,通过商品期望价格 $E(p_{it})$,计算基于价格离散的居民消费价格指数 $iCPI_pd$,主要步骤如下:

第一步,相对价格计算。计算每个商品在 t 期的相对价格,即 t 期的期望价格与基期价格之比:

rp_{it} = \frac{E(p_{it})}{E(p_{i,t-1})} \tag{25}

其中, $E(p_{it})$ 表示商品 i 在 t 期的期望价格, rp_{it} 表示商品 i 在 t 期的相对价格。

第二步,基于八大类商品的几何平均。对八大类商品分别计算其相对价格的几何平均:

gm_{jt} = (\prod_{i \in j} rp_{it})^{\frac{1}{N_j}} \tag{26}

其中, gm_{jt} 表示 j 类商品在 t 期的几何平均, j 表示八大类商品中的某一分类, N_j 表示 j 类中的商品数量。

第三步,类别权重调整。根据八大类商品的权重,对各类别的几何平均值进行加权处理:

wgm_t = \prod_{j=1}^8 w_j \times gm_{jt} \tag{27}

其中, gm_{jt} 表示整体商品在 t 期的权重调整后的几何平均, w_j 表示 j 类商品权重。

第四步,指数化处理。将式(27),进行指数化处理,得出:

iCPI_pd_t = 100 \times \prod_{d=1}^{t-1} CPI_adjust_{t-d} \tag{28}

(二)居民消费指数分析与比较

本部分利用样本期内在线商品的价格数据,对基于价格离散的居民消费价格指数($iCPI_pd$)进行分析与比较,见表10。

第一, $iCPI_pd$ 的均值为 107.98,高于基期水平,表明样本期间内价格水平呈现温和上升趋势。这种增长态势不仅反映市场中有着更多的消费需求,也在一定程度上说明存在温和的通胀压力。

第二, $iCPI_pd$ 标准差为 1.84,说明价格指数波动幅度较小,市场价格整体较为稳定。其中,最大值与最小值分别为 110.97 和 99.33,价格在样本期间波动范围为 11.64%,表明短期内存在一定的价格波动风险,但整体波动可控。

第三,比较 $iCPI_pd$ 、 $iCPI^{\text{①}}$ 与 CPI 之间的相关系数。由表10可知, $iCPI_pd$ 与 CPI 的相关系数为 0.115,而 $iCPI$ 与 CPI 的相关系数为 0.025,两者均为正相关,但相关程度差异。由图2可知,相较于 $iCPI$, $iCPI_pd$ 在反映 CPI 波动趋势,具有更强的线性关系,这意味着通过在线大数据测度居民消费指数时,也就是分别基于商品价格

表 10 iCPI_pd 的测度结果

iCPI_pd				iCPI_pd 与 CPI	iCPI 与 CPI
均值	标准差	最大值	最小值	相关系数	相关系数
107.98	1.84	110.97	99.33	0.115	0.025

数据来源:作者根据测算结果,自行整理。

①iCPI 由清华大学 iCPI 项目组,利用在线数据构造实时高频的消费价格指数,见网站 <http://www.bdecon.com>。

单一值和离散特征计算的 $iCPI$ 和 $iCPI_{pd}$, 考虑商品价格的离散分布, 能够提升价格指数的代表性, 更多能够反映真实价格变化。

七、结论与启示

(一) 主要结论

本文利用网络文本挖掘技术, 收集来自电商平台的商品信息, 构建产品层面微观数据库, 首次利用大样本产品级价格数据对中国线上商品市场的价格离散的测度以及影响因素进行探讨。主要研究结论如下:

第一, 通过静态和动态两种方法测度价格离散, 中国线上商品的价格离散程度较大。不同类别商品之间的价格离散程度具有异质性。基于此构建的高频价格离散指数 ($iPDI$) 表明, 当商品价格包含促销时, 价格离散的变化幅度进一步扩大, 尤其在电商节日 (6.18 和 11.11) 期间, 平台商家的促销行为为进一步加深价格离散程度。第二, 通过价格离散的方差分解发现, 跨平台价格差异是价格离散的主要来源。在此基础上构建实证模型, 发现平台数目对价格离散的影响存在倒 U 型关系。当平台数目低于临界点时, 新平台进入会加剧价格离散; 超过临界点后, 价格离散程度开始下降。进一步研究发现, 平台势力通过交叉网络外部性对这一倒 U 型关系产生调节作用, 体现了市场从增量竞争向存量竞争转变的过程。第三, 将价格离散的微观测度结果拓展至宏观应用, 创新性地构建了基于价格离散特征的居民消费价格指数 ($iCPI_{pd}$)。研究发现, 相较于 $iCPI$, $iCPI_{pd}$ 与 CPI 具有更强的相关性, 能够更准确地反映商品价格的波动特征, 实现了价格离散微观测度与宏观应用的有效衔接。

(二) 政策启示

一方面, 价格离散衡量了真实市场相对于完全竞争市场的偏离程度。本文的研究表明, 中国线上商品的价格离散程度较大, 说明线上市场虽然打破了地域限制, 但是同一产品在相同时期的价格分布仍未趋同。其中, 跨平台竞争是影响价格离散的重要因素。政府应加强对平台竞争行为的监管, 大力反对平台市场分割策略等不正当竞争行为, 维护统一的线上大市场。例如, 政府应当严格限制平台对于用户进行锁定的商业行为, 限制平台对用户进行画像进而采用如“大数据杀熟”等的价格歧视策略, 从根源上限制平台基于用户需求异质性对市场的分割。另一方面, 高频价格离散指数 ($iPDI$) 能够实时反映价格信息, 体现商品市场供需的变化情况, 以及基于价格离散特征的居民消费物价指数 ($iCPI_{pd}$), 更准确地反映消费品价格的变动特征, 对于及时地分析宏观经济问题和微观经营策略具有十分重要的参考价值。政府应加强构建实时高频价格离散指标, 编制高频物价指标, 用于实时监测宏观经济形势和物价变化, 特别是网络购物节、春节、突发卫生事件等价格的短期剧烈波动的情况。

参考文献:

- [1] 杨宏伟, 石卓达. 数字经济何以促进全国统一大市场建设——基于数据要素流通视角[J]. 调研世界, 2024(12): 31-43.
- [2] 李陈华, 刘怡嘉. 电商示范城市试点政策赋能包容性增长: 理论机制与实证检验[J]. 贵州财经大学学报, 2025(1): 90-99.
- [3] Baye M R, Morgan J. Information gatekeepers on the internet and the competitiveness of homogeneous product markets[J]. American Economic Review, 2001, 91(3): 454-474.
- [4] Stigler G J. The economics of information[J]. Journal of Political Economy, 1961, 69(3): 213-225.
- [5] Cason T N, Friedman D, Hopkins E. An experimental investigation of price dispersion and cycles[J]. Journal of Political Economy, 2021, 129(3): 789-841.
- [6] Mohapatra D, Mohapatra D P, Sewak Dubey R. Price dispersion across online platforms: evidence from hotel room prices in London(UK)[J]. Applied Economics, 2024, 56(52): 6598-6610.
- [7] 严玉珊. 电商异质性与线上价格离散[J]. 商业经济与管理, 2022(2): 17-28.
- [8] 张昊. 地区间生产分工与市场统一度测算: “价格法”再探讨[J]. 世界经济, 2020(4): 52-74.
- [9] Hitsch G J, Hortacsu A, Lin X. Prices and promotions in US retail markets[J]. Quantitative Marketing and Economics, 2021, 19(3): 289-368.
- [10] 端利涛, 吕本富. 在线购物是否存在“反戴蒙德悖论”现象? [J]. 管理评论, 2022(9): 134-146.

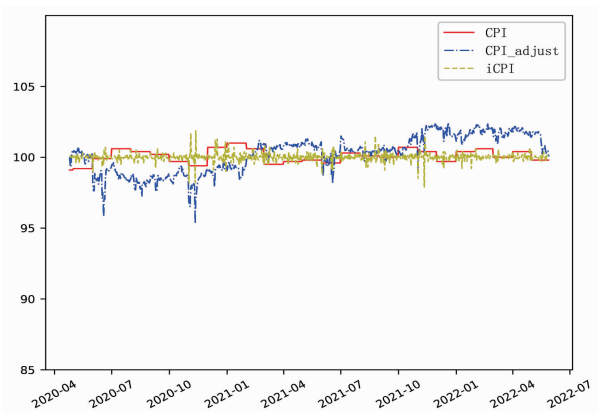


图2 $iCPI_{pd}$ 、 $iCPI$ 与 CPI

- [11] 李莉,张华. 考虑商家定价和广告投资的电商平台价格折扣决策[J]. 运筹与管理,2024(8):115-121.
- [12] 姜婷凤,汤珂,刘涛雄. 基于在线大数据的中国商品价格粘性研究[J]. 经济研究,2020(6):56-72.
- [13] Grewal D, Roggeveen A L, Compeau L D, Levy M. Retail value-based pricing strategies: New times, new technologies, new consumers[J]. Journal of Retailing, 2012, 88(1):1-6.
- [14] Klenow P J, Oleksiy K. State-dependent or time-dependent pricing: Does it matter for recent U. S. inflation? [J]. The Quarterly Journal of Economics, 2008, 123(3):863-904.
- [15] Cavallo E, Powell A, Becerra O. Estimating the direct economic damages of the earthquake in Haiti [J]. The Economic Journal, 2010, 120(546):298-312.
- [16] 刘涛雄,汤珂,姜婷凤. 一种基于在线大数据的高频 CPI 指数的设计及应用[J]. 数量经济技术经济研究,2019(9):81-101
- [17] Pan X, Ratchford B T, Shankar V. Price dispersion on the internet: a review and directions for future research[J]. Journal of Interactive Marketing, 2004, 18(4):116-135.
- [18] 于左,张芝秀,王昊哲. 交叉网络外部性、独家交易与互联网平台竞争[J]. 改革,2021(10):131-144.
- [19] 孙震,刘健平,刘涛雄. 跨平台竞争与平台市场分割——基于中国线上市场价格离散的证据[J]. 中国工业经济,2021(6):118-136.
- [20] Hackl F, Kummer M E, Winter-Ebmer R, Zulehner C. Market structure and market performance in E-commerce[J]. European Economic Review, 2014, 68:199-218.
- [21] Kovacic W E, Shapiro C. Antitrust policy: A century of economic and legal thinking[J]. Journal of Economic Perspectives, 2000, 14(1):43-60.
- [22] 李芳,刘洋,王松,等. 多主体外部监管合力可以助力企业提升全要素生产率吗? ——基于多源监管大数据的视角[J]. 审计与经济研究,2024(6):117-126.
- [23] 潘小军. 基于消费者显著性偏好的双边平台转换成本和定价策略研究[J]. 中国管理科学,2021(11):45-54.
- [24] 唐方成,郭欢,程立燕. 考虑交叉网络外部性下平台竞争的定价策略研究[J]. 中国管理科学,2024(6):219-228.
- [25] 曲创,王夕琛. 互联网平台垄断行为的特征、成因与监管策略[J]. 改革,2021(5):53-63.
- [26] Fava V L. Price dispersion and price indexes[J]. Applied Economics, 2010, 42(1):23-36.

[责任编辑:杨志辉]

Measurement and Source Identification of Price Dispersion: Empirical Evidence from Online Big Data

QIN Jie¹, HU Linyao², MENG Lu³

(1. School of Economics, Nanjing Audit University, Nanjing 211815, China;

2. School of Economics, Beijing University of Industry and Business, Beijing 100048, China;

3. School of Economics and Management, Southwest Jiaotong University, Chengdu 610031, China)

Abstract: Price dispersion serves as a crucial indicator for measuring market integration, providing a novel perspective for studying the unified national market. Through collecting daily price data of identical products from different platforms, this study measures the degree of price dispersion in China. The findings reveal: (1) The overall level of price dispersion in China's online market remains relatively high, with significant heterogeneity observed across different product categories. The constructed high-frequency price dispersion index (*iPDI*) indicates that a lower index value suggests higher market integration, while a higher value implies lower integration. (2) Price variance decomposition analysis reveals that price dispersion primarily stems from cross-platform competition. Furthermore, empirical modeling demonstrates an inverted U-shaped relationship between cross-platform competition and price dispersion. (3) Applying microscopic measurements of price dispersion at the macroeconomic level, the Consumer Price Index (*iCPI_{pd}*) compiled based on price dispersion characteristics more accurately reflects the fluctuation patterns of commodity prices.

Key Words: price dispersion; online big data; price dispersion index; cross-platform competition; consumer price index