

# 程序审查报告

北京师范大学国民核算研究院

## 0 程序简介

所审查的程序用于江西方面“横比价格指数”研究。

## 1 程序优点

- 1.1 在功能实现方面，对于相对应的平均价格与支出权重数据，程序可以实现基本分类 PPP 的计算与基本分类以上 PPP 的加总等基本操作，并且运行得比较流畅。
- 1.2 在代码编写方面，程序有明确的区段划分，各个区段程序的功能明确，尤其是文件读写操作齐全，每一区段的运行结果都有对应的数据文件可以手动查询；程序的注释与程序的区段划分相匹配，能够概括每一区段程序的具体功能；程序各个函数之间的衔接流畅自然，尤其是在构造 CPD 法所需的虚拟变量资料矩阵的时候。

## 2 程序缺点

- 2.1 程序是使用 Python 3 系列版本语言写成的。虽然 Python 是一款非常适用于数据统计分析的计算机语言，但是 Python 语言写成的“.py”文件在执行时要求计算机必须首先安装相应的 Python 版本，并且在运行时一般使用 Windows 命令提示符运行，操作起来具有一定难度，需要懂得这门计算机语言的人来进行。
- 2.2 程序在运行过程中对工作路径做出了明确规定，也就是说，对于程序运行过程中需要使用到的各种数据文件（如包含各地区平均价格和支出权重的.xlsx 数据文件），在不改写程序的情况下，程序要求这些文件必须被存储在计算机的特定位置。程序对文件存储位置的要求，使得除非程序自始至终只在一台电脑上运行，或者在不同电脑上运行时都建立相同的工作路径，否则必须经过一定的调整才能在不同的计算机上运行。
- 2.3 程序对所给的数据适应较好，不代表在其他数据情况下也能正常使用。
  - 1) 在计算基本分类 PPP 时，程序对数据文件的操作对文件内部的数据组织格式作出了明确规定。比如读取各地区的平均价格时，由于原始数据文件前三行是没有用的，所以程序设定为从第四行开始读取，这种情况下倘若其他地区的原始数据文件格式不同，就有可能导致读取遗漏或多余，严重时会导致接下来的一系列操作出错。要解决这一问题就得要求各地区在填写数据文件、上报数据时按照统一的文件模板填写。

2) 程序直接使用了数字代替某些有意义的变量,如地区数目,这些变量的具体取值往往依赖于数据文件。直接使用数字替代变量,会使得当前程序在应用于其他的数据文件时出现数字不匹配等问题,因此不应该直接使用数字。

3) 所给的数据是一个模拟数据。与现实数据相比,一般而言,模拟数据在数据完整度和数据中的奇异值等方面都有更好的性质。若将此程序运用于现实数据,其运行情况有待观察。示例所给的数据文件是一个完整的平均价格表,不存在某地区某规格品平均价格数据的缺失,而这种情况在现实中是很难出现的。由于数据文件是完整的价格表,程序没有考虑如何补全缺失数据,这就意味着未来在运用该程序时,必须通过其他方法及程序补全缺失数据,将不完整的价格表补成完整的价格表,然后才能调用该程序做进一步的计算。

2.4 “2.1 基本分类以上的 ppp (GEKSPPP).py”中“对先前输出的 ppp 表格在 excel 中修改了格式另存为 price 表格”,虽然程序中没有说明这些修改时如何做出的,但是通过比较 ppp.xlsx 和 price.xlsx 可以发现修改主要有三个部分:一是加入了对应的大类编码,二是加入了作为基准地区的南昌的基本分类 PPP,三是将原先各列的数字标签替换成了对应的地区名称。这些操作没有在程序内完成而是在程序外完成,就导致程序运行间隙需要工作人员手动调整,非常复杂。

2.5 程序对应的研究使用的是根据 CPI 产品清单改编而成的支出分类,该支出分类有 3 个层次,分别是 90 个“基本分类”(同时也是“中类”),11 个“大类”和 1 个最高的“GDP”层次,规格品在“基本分类”以下。从指数编制的原理上来讲这样做是可行的。但是这个支出分类与一般的 ICP 支出分类并不相同,有关 GDP 支出分解的一系列配套操作也有待说明。

2.6 在计算基本分类以上的 ppp 时,程序对于 GEKS 法的理解与公式套用是错误的。GEKS 法

$$GEKS_k = \left( \prod_{j=1}^K P_F(p^k, p^j, q^k, q^j) \right)^{1/k}, \quad k = 1, 2, \dots, K$$

是将目标地区 k 的价格向量固定为 Fisher 指数的分子上的价格向量,将不同地区的价格向量轮流带入分母中的价格向量,并将所有 Fisher 指数求等权重的几何平均作为 GEKS 化的 Fisher 指数。但是程序中做的恰恰相反,将目标地区 k 的价格向量固定为 Fisher 指数分母上的价格向量,而将不同地区的价格向量轮流带入分子中。

$$GEKS_k^{pseudo} = \left( \prod_{j=1}^K P_F(p^j, p^k, q^j, q^k) \right)^{1/k}, \quad k = 1, 2, \dots, K$$

这样计算不符合 GEKS 过程,计算得到的指数也就不具备所期望的优良性质。好在程序的问题不大,只要把相应计算过程中的分子与分母调换位置,即可修正。

### 3 其他须注意的问题

3.1 程序频繁的读写操作没有必要，可以考虑去除一部分多余的读写操作。

3.2 程序在数据的组织存储、运算方面还有优化的空间。