EPS贸易数据分析

胡华平

2020-12-25

Table of Contents

# 目前进展

数据收集进展情况：

1. EPS数据库。

* HS2002子库（2002-2011年）：正在下载中，一天内可以完成数据下载！
* HS2012子库（2012-2016年）：已经完成下载！可以进行数据分析。

1. 竹云贸易数据库（2017-2020）。受到数据可系统查询设定的约束，下载工作量较大，需要较长时间。

# HS四位编码农产品数量  
n\_prod <- 197  
# 国家数（包括总值）  
n\_country <- 243  
# 年份数（2017-2020年）  
n\_year <- 4  
# 每次操作需要的平均秒数  
n\_operation <- 5  
  
# 估算全部操作需要的时长为 小时数  
(time\_total <- (n\_prod\*n\_country \*n\_year \*n\_operation)/3600/24)

[1] 11

# 面临问题

# 文献讨论

理论模型：

实证模型：

操作步骤和主要过程

主要结论

# 数据分析

## 比较HS国家编码变动

## 合并HS2012数据

下面合并操作只需要运行一次即可。

all\_files <- list.files("../data/eps/import/HS2012")  
n <- length(all\_files)  
  
path\_dir <- here::here("manuscript","data","eps", "import", "HS2012")  
path\_file <- paste0(path\_dir, "/",all\_files)  
  
tbl\_hs2012 <- NULL  
for (i in 1:n) {  
 tbl\_tem <- read\_rds(file = path\_file[i])  
 tbl\_hs2012 <- bind\_rows(tbl\_hs2012, tbl\_tem)  
 print(paste0("成功合并到了第", i,"个文件。共有",n,"个文件！"))  
}  
  
# 写出文件  
write\_rds(tbl\_hs2012, "../data/eps/tbl-hs2012.rds")

## 比较HS产品编码变动

## HS2012数据描述性分析

读取已经合并好的数据，并进行简单清洗，具体包括：

* 去掉原数据中的**千分位分隔符**（逗号,）。
* 正确变换列类型，例如列value（贸易额）的类型应该为number（数值型）。
* 变换数值单位。例如列value（贸易额）的单位由**美元**变换为**万美元**。
* 筛除不必要的数据行。例如，删除所有**贸易量**行（用不到），以及**贸易额**的无数据行（因为没有产生贸易活动）。
* 加入**国家**的编码等信息。需要用到tbl\_nation的数据表，先处理好匹配需要的key，然后使用left\_jion()函数进行匹配。

# 读取数据  
tbl\_hs2012 <- read\_rds("../data/eps/tbl-hs2012.rds")  
  
tbl\_nation <- read\_rds("../data/eps/list-nation-hs2012.rds") %>%  
 rename("country" = "nation", "country\_code"="code",  
 "country\_id"="index")  
#str(tbl\_nation)  
  
# 转换数据形式  
tbl\_hsclear <- tbl\_hs2012 %>%  
 mutate(value = str\_replace\_all(value, ",", "")) %>%  
 mutate(value = as.numeric(value),  
 value = 0.001\*value) %>%  
 filter(str\_detect(var, "美元"), !is.na(value)) %>%  
 left\_join(., tbl\_nation, by = "country")%>%  
 arrange(code\_prod, year,month, country\_id)  
  
#head(tbl\_hsclear)  
#unique(tbl\_hsclear$year)  
#str(tbl\_hsclear)

下面我们进行初步的数据汇总分析。

1.按产品和按月份，来统计加总进口额情况

这是一张表

表 1: 2012年1001产品分月进口总额

|  |  |  |
| --- | --- | --- |
| code\_prod | month | total |
| 1001 | 05 | 16.76 |
| 1001 | 03 | 16.23 |
| 1001 | 09 | 16.23 |
| 1001 | 10 | 11.09 |
| 1001 | 02 | 10.84 |
| 1001 | 07 | 7.98 |
| 1001 | 04 | 7.65 |
| 1001 | 08 | 6.80 |
| 1001 | 01 | 6.67 |
| 1001 | 06 | 5.62 |
| 1001 | 11 | 4.06 |
| 1001 | 12 | 0.21 |

下面我们来做一张柱状图：

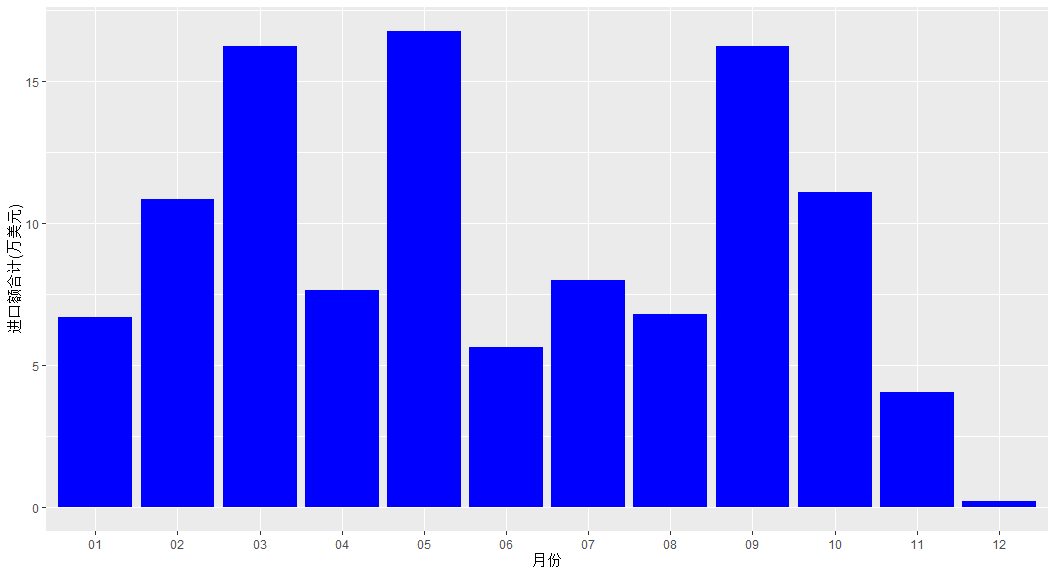


图 1: 一张丑丑的柱状图

上面的表1 和图1 共同表明，分月合计的产品进口贸易额呈现一定的双峰分布形态。

这是另外一张表。它表明什么社么

表 2: 2012年所有产品分月进口总额

|  |  |
| --- | --- |
| month | sum\_pro |
| 09 | 7920469 |
| 07 | 7678319 |
| 12 | 7602672 |
| 05 | 7412776 |
| 06 | 7258611 |
| 08 | 7212983 |
| 11 | 6946904 |
| 03 | 6911726 |
| 04 | 6801753 |
| 10 | 6728003 |
| 01 | 6010024 |
| 02 | 5683367 |

分析国家进出情况（有问题，未完成）。

tbl\_smry <- tbl\_hsclear %>%  
 ungroup() %>%  
 filter(year=="2012") %>%  
 filter(country !="总值") %>%  
 group\_by(code\_prod, month) %>%  
 summarise(n = dplyr::n(country))

# 得到分析数据表

## 贸易总额及其月度变动

### 理论表达

表示**分产品每个月**的加总贸易额，表示滞后1期变量。

表示**分产品**下的**月度贸易变动额**。

### 代码操作

**具体思路**如下：

1. 我们注意到原数据中实际上已经有**分产品**下的**月度贸易总额**。数据。也即country变量下包含有总值类别，因此可以直接使用filter()函数过滤得到**分产品**下的**月度贸易总额**。
2. **分产品**下的**月度贸易变动额**可以先计算得到滞后变量，然后再差分得到贸易变动额（$\Delta X}= X\_{l0} - X\_{l1}$）。具体使用group\_by() + dplyr::lag()函数操作得到滞后变量。

以下为具体R代码操作：

xc <- tbl\_hsclear %>%  
 arrange(code\_prod, year, month) %>%  
 filter(country\_id==1) %>%  
 #unite(col = "ym", year, month, sep = "-", remove = FALSE) %>%  
 select(code\_prod,product, year, month, value) %>%  
 group\_by(product) %>%  
 mutate(x\_l0 = value,   
 x\_l1 = dplyr::lag(value,n=1)) %>%  
 ungroup() %>%  
 mutate(delta\_x = x\_l0 - x\_l1)

下面表3简单展示了得到的数据表结果：

表 3: （其中5类产品）分产品的月度贸易额及变动（单位：万美元）

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| index | code\_prod | product | year | month | x\_l0 | x\_l1 | delta\_x |
| 344 | 0201 | 0201 - 鲜、冷牛肉 | 2016 | 09 | 5772 | 4559 | 1213 |
| 345 | 0201 | 0201 - 鲜、冷牛肉 | 2016 | 10 | 5375 | 5772 | -397 |
| 346 | 0201 | 0201 - 鲜、冷牛肉 | 2016 | 11 | 5569 | 5375 | 194 |
| 347 | 0201 | 0201 - 鲜、冷牛肉 | 2016 | 12 | 7262 | 5569 | 1694 |
| 4382 | 1001 | 1001 - 小麦及混合麦 | 2016 | 09 | 102137 | 87307 | 14830 |
| 4383 | 1001 | 1001 - 小麦及混合麦 | 2016 | 10 | 60288 | 102137 | -41850 |
| 4384 | 1001 | 1001 - 小麦及混合麦 | 2016 | 11 | 6282 | 60288 | -54006 |
| 4385 | 1001 | 1001 - 小麦及混合麦 | 2016 | 12 | 47432 | 6282 | 41150 |
| 4568 | 1005 | 1005 - 玉米 | 2016 | 09 | 5160 | 6933 | -1773 |
| 4569 | 1005 | 1005 - 玉米 | 2016 | 10 | 5303 | 5160 | 143 |
| 4570 | 1005 | 1005 - 玉米 | 2016 | 11 | 11522 | 5303 | 6219 |
| 4571 | 1005 | 1005 - 玉米 | 2016 | 12 | 38729 | 11522 | 27207 |
| 4628 | 1006 | 1006 - 稻谷、大米 | 2016 | 09 | 101107 | 87018 | 14088 |
| 4629 | 1006 | 1006 - 稻谷、大米 | 2016 | 10 | 95166 | 101107 | -5941 |
| 4630 | 1006 | 1006 - 稻谷、大米 | 2016 | 11 | 142753 | 95166 | 47587 |
| 4631 | 1006 | 1006 - 稻谷、大米 | 2016 | 12 | 199596 | 142753 | 56843 |
| 5362 | 1201 | 1201 - 大豆，不论是否破碎 | 2016 | 09 | 3078517 | 3292747 | -214230 |
| 5363 | 1201 | 1201 - 大豆，不论是否破碎 | 2016 | 10 | 2248721 | 3078517 | -829795 |
| 5364 | 1201 | 1201 - 大豆，不论是否破碎 | 2016 | 11 | 3382301 | 2248721 | 1133579 |
| 5365 | 1201 | 1201 - 大豆，不论是否破碎 | 2016 | 12 | 3927236 | 3382301 | 544936 |

## 得到变动因素1：由于新加入国家引起贸易变动

### 理论表达

其中：

### 代码操作

rm\_list <- c("var", "ym", "country\_code")  
tbl\_newer <- tbl\_hsclear %>%  
 select(-all\_of(rm\_list)) %>%  
 filter(country\_id!=1) %>%  
 group\_by(code\_prod, product, year, month) %>%  
 nest() %>%  
 ungroup() %>%  
 rename("dt\_l0"="data")