

季节调整中的春节模型

栾惠德 张晓峒*

摘要 如何估计并消除春节等移动假日的影响是我国季节调整工作中的一个重点和难点。本文首先借鉴 X-12-ARIMA 季节调整程序中的复活节模型建立了基本的春节模型,继而考虑到存量数据与流量数据在性质上的差异,提出了存量数据的春节模型。在此基础上进一步扩展,构造了三区段变权重春节模型。对社会消费品零售总额和货币供应量的实证检验表明,这一组春节模型能够很好地消除季节调整中的春节效应。

关键词 季节调整,春节效应,

X-12-ARIMA

一、引言

由于季节因素的存在,使得原始的月度或季度时间序列数据(统称为子年度数据)不具有可比性,不能直接进行比较。季节调整就是对子年度数据中隐含的由季节性因素引起的季节影响加以纠正的过程。季节调整的用途主要有两点:其一是出于制订生产计划和控制存货的目的,对季节变化预先作出估计,提前采取应对措施;其二是为了能够从总量变化中消除季节变化,从而能够更清晰地揭示其他类型的变动,特别是反映经济周期的运动规律。季节调整后的时间序列是趋势、循环和不规则因素的合成。与没有经过季节调整的数据相比,调整后数据最重要的一个特点,就是可以更及时地反映经济的瞬间变化,识别经济变化的转折点。

与季节因素相伴而来的还有日历效应,也就是各种与日历相关的影响因素,如闰年、交易日效应、移动假日效应等,它们是长期存在并可以预测的,其对子年度时间序列的影响与季节因素类似,同样会给经济周期的判断造成困难,因此常把它们和季节因素组合在一起考虑并加以消除。

移动假日定期出现,但不一定出现在每年的同一时间。随着假日在公历中的出现日期的推移,使得此类假日的影响在相邻月份(或季度)之间的分配每年都有所不同。在我国最重要的移动假日是春节。如何估计和消除春节的影响,是我国季节调整工作中的一个重点和难点。国内已有的研究,如齐

* 南开大学国际经济研究所。通讯作者及地址:栾惠德,南开大学西区公寓8号楼B座15-104,300071;电话:(022)23494529,13752503450;E-mail: luanderful@163.com。本文系中国人民银行项目“PBC版时间序列 X-12-ARIMA 季节调整原理研究”课题成果之一。作者感谢美国普查局 Brian. C. Monsell 的热情帮助和中国人民银行唐思宁、王毅、章丽盛的有益评论,感谢匿名审稿人提出的宝贵意见,当然作者文责自负。

东军(2004)、张鸣芳等(2004)和贾淑梅(2005)等均借鉴了美国普查局 X-12-ARIMA 季节调整程序中的复活节模型,根据所设定的假日影响的天数在相邻月份之间的分配比例构造回归变量,从而对春节效应作出估计。但在实证应用中未对流量数据和存量数据作出严格区分,简单套用流量春节模型对货币供应量等存量数据进行调整,显然是不合适的。在本文中,我们首先界定了流量和存量的概念差别,为流量和存量数据分别构造了基本的春节效应模型,进而借鉴美国普查局 genhol 程序所给出的三区段等权重移动假日模型,建立了扩展的变权重流量春节模型和存量春节模型。经实证检验,这两个模型能够很好地解决我国春节等移动假日的调整问题。本文的内容安排如下:第二节简要介绍基本的流量数据春节模型;第三节以社会消费品零售总额月度数据为例具体说明基本春节模型的计算原理;第四节构造存量数据基本春节模型;第五节借鉴 genhol 程序建立三区段变权重流量春节模型;第六节构造三区段变权重存量春节模型,并以货币供应量 M0 月度数据为例说明其计算原理;最后是结论部分。

二、流量数据的基本春节模型

X-12-ARIMA 是在世界范围内得到广泛应用的季节调整程序,由美国普查局历时五十余载开发、升级而成,根据美国的情况设定了复活节、劳动节和感恩节三种移动假日,用以消除流量数据中可能存在的移动假日效应。复活节和劳动节效应的基本模型假定:从节日之前的第 w 天开始,经济活动的水平发生变动,并保持在这一新水平上直至节日的前一天。感恩节效应的基本模型假定:日度活动水平从感恩节之前或之后的第 w 天开始发生变动,并保持在这一新水平上直到 12 月 24 日;其中, $w > 0$ 表示感恩节之前的 w 天, $w < 0$ 表示感恩节之后的 w 天。

据此构造回归变量,将对于给定月份 j ,受到节日影响的时段落在 j 月份的时间占整个受影响时段 w 的比例记为:

$$P(w, j) = \frac{1}{w} \times [\text{节前 } w \text{ 天中落在 } j \text{ 月份的天数}],$$

在节日影响不到的月份,该变量的取值为 0。

在实际计算时,回归变量应当采取中心化(centered)的形式,以保证在消除估计的假日影响之后,所得到的序列的年度总和大致等于调整之前的原序列的年度总和。如果没有进行中心化处理,这两个年度总和之间就会每年相差大致相同的幅度,这样,经过假日调整和季节调整之后得到的序列将是原观测序列的有偏估计。

在我国,最重要的移动假日就是春节。春节在现行公历中的变动范围是

从 1 月 21 日到 2 月 20 日，因此，农历新年与公历新年（元旦）之间的时滞从 21 天到 51 天不等。图 1 就是最近 40 年（1981—2020）春节在公历中的分布情况。

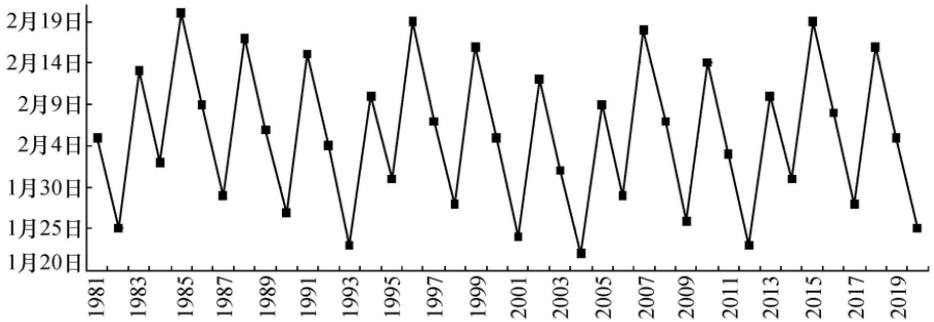


图 1 春节变动

参照 X-12-ARIMA 对复活节和感恩节的调整方法，我们首先建立基本的流量数据春节模型。假定春节对于待调整变量的影响主要发生在春节前的 w 天； $-14 \leq w \leq 20$ 且 $w \neq 0$ ； $w > 0$ 表示春节之前的 w 天； $w < 0$ 表示春节之后的 w 天； $w = -14$ 表示节后影响的时效是从正月初二到正月十五。对于给定的年份 i ，令 $w_{i,j}$ 表示春节之前 w 天中落在 j 月份的天数。定义变量 $P_{i,j}(w) = w_{i,j}/w$ （用 i, j 双下标表示时序 t ，等价于前文的 $P(w, j)$ ），给定对 w 的约束，该变量在除 1 月、2 月和 3 月以外的其他月份取值都是 0。然后，还需要对 $P_{i,j}(w)$ 进行中心化处理，消除季节性。从 $P_{i,j}(w)$ 中减掉月份 j 在全部样本年份的均值 $\bar{P}_{\cdot,j}(w)$ ，就得到最终的解释变量 $X_{i,j}(w)$ ，即 $X_{i,j}(w) = P_{i,j}(w) - \bar{P}_{\cdot,j}(w)$ 。这样，通过消除春节效应对所关心的特定月份的长期影响，就使得这一解释变量具有零均值、不含季节性，从而能够保持序列的原始水平。进而模型可以写作 $Y_{i,j} = bX_{i,j}(w) + z_{i,j}$ ，其中 $Y_{i,j}$ 是对应第 i 年第 j 期的待调整序列， $z_{i,j}$ 是回归误差。

对于加法分解模型，如果是通过 regARIMA 预调整模块向 X-12-ARIMA 程序导入用户自定义的回归变量，则一个一般的 regARIMA 回归模型形式为

$$\varphi(L)\Phi(L^s)(1-L)^d(1-L^s)^D(Y_{i,j} - a - bX_{i,j}(w) - \dots) = \theta(L)\Theta(L^s)\varepsilon_{i,j}. \quad (1)$$

其中 L 是滞后算子， s 是季节周期（月度数据 $s=12$ ，季度数据 $s=4$ ）。模型隐含着首先从 $Y_{i,j}$ 中减去回归效应，得到零均值的误差序列 $z_{i,j}$ ，然后对该序列 $z_{i,j}$ 取差分以得到一个平稳序列，比如说 $u_{i,j}$ ， $u_{i,j}$ 被假定服从平稳的 ARMA 模型，即 $\varphi(L)\Phi(L^s)u_{i,j} = \theta(L)\Theta(L^s)\varepsilon_{i,j}$ 。使用迭代广义最小二乘算法估计参数，得到估计的春节效应为 $\hat{S}_{i,j} = \hat{b}X_{i,j}(w)$ 。从 $Y_{i,j}$ 中减去春节回归效应，即 $Y_{i,j} - \hat{S}_{i,j}$ ，就得到了经过春节效应调整的序列。

对于乘法分解模型,各种分解成分之间是相乘的关系,因此需要先对原序列 $Y_{i,j}$ 取对数,即 $y_{i,j} = \log(Y_{i,j})$, 然后对 $y_{i,j}$ 建立 (1) 式形式的 regARIMA 模型,实际的春节模型为: $y_{i,j} = \log(Y_{i,j}) = bX_{i,j}(\omega) + z_{i,j}$, 估计得到的春节效应回归结果为 $\hat{s}_{i,j} = \hat{b}X_{i,j}(\omega)$, 再对其取指数变换,即得春节效应调整因子 $\hat{S}_{i,j} = e^{\hat{s}_{i,j}} = e^{\hat{b}X_{i,j}(\omega)}$, 用原序列 $Y_{i,j}$ 除以相应的调整因子,即 $Y_{i,j}/\hat{S}_{i,j}$, 就消除了春节的影响。

三、流量数据基本春节模型的应用

下面以中国社会消费品零售总额月度序列(1993年1月至2004年11月)为例对以上模型具体加以说明。数据引自 BvD 商业电子数据出版社 EIU 各国宏观经济指标数据库。原始序列的折线图显示该序列存在明显的季节波动。为了更方便地解释春节效应的计算过程,这里采用的是加法分解模型,并且没有考虑交易日效应。

首先不考虑春节效应进行季节调整,回归变量只包括常数项,令程序自动选取合适的 ARIMA 模型,自动探测 AO、LS 和 TC 三类异常值。¹ 估计得到的回归结果如表 1 所示。可以看到,在不考虑春节效应的情况下,程序标识出的异常值多达 7 个,其中在标记为 AO 异常值的 6 个时点中,1996 年 1 月、1998 年 1 月、2002 年 1 月和 2004 年 1 月都分布在 1 月份,具有明显的规律性,因此,有理由相信这一次季节调整在一定程度上受到了春节效应的影响。

表 1 回归模型

变量	参数估计	标准差	t 统计量	概率值
常数项	4.9345	3.57043	1.38	0.1702
自动识别的异常值				
AO1996. Jan	-97.8654	19.34832	-5.06	0.0000
AO1996. Dec	125.4431	20.04550	6.26	0.0000
AO1998. Jan	133.3739	19.34816	6.89	0.0000
AO2002. Jan	-138.9136	18.68786	-7.43	0.0000
AO2003. May	-162.8330	18.68786	-8.71	0.0000
LS2003. Nov	-159.5826	26.11280	-6.11	0.0000
AO2004. Jan	138.9818	27.68100	5.02	0.0000

¹ AO(Additive Outlier, 单点异常值), 表示发生在 t_0 时刻的异常值点; LS(Level Shift, 水平漂移), 表示从 t_0 时刻起变量瞬间变化到一个新的水平上并保持这一水平; TC(Temporary Change, 暂时变化), 表示在 t_0 时刻变量瞬间变化到一个新的水平上, 随后经指数衰减回到原有水平。X-12-ARIMA 季节调整程序在进行异常值自动探测的过程中, 使用残差标准偏差的稳健估计计算 t 统计量, 即 $\sigma^R = 1.48 \times$ 残差绝对偏差的中位数, 同时默认临界值由样本容量决定(均大于通常使用的 t 分布临界值), 以保证检测程序只把最为显著的异常值标识出来, 减少由季节调整选项的设置差异而引起的异常值识别的不稳定性。在本例中, 样本数为 143, 程序默认的异常值探测临界值为 3.89, 即只有当回归得到的 t 统计量的绝对值大于 3.89 时, 该时点才会被标识为异常值。在此基础上自动在模型中引入对应的虚拟变量, 以保证模型的参数估计具有稳健性。

接下来引入春节模型消除春节效应。在本例中，我们假定 $w=20$ 。除 1 月和 2 月以外，即当 $j \neq 1$ 或 2 时， $P_{i,j}$ 的取值均为 0。提供 1993 年至 2005 年的春节变量数据，其中 1、2 月份的具体数值见表 2。例如 1999 年，在春节（2 月 16 日）之前的 20 天里，有 5 天落在 1 月份、15 天落在 2 月份。因此 $P_{1999,1}(20)=5/20=0.25$ ， $P_{1999,2}(20)=15/20=0.75$ 。 $P_{i,1}(20)$ 和 $P_{i,2}(20)$ 的样本平均值分别为 $\bar{P}_{\cdot,1}(20)=0.7269$ 和 $\bar{P}_{\cdot,2}(20)=0.2731$ ，即平均而言，节前 20 天的影响落在 1 月份的比率是落在 2 月份的比率的近 3 倍。用这些月度均值去修正变量 $P(w)$ ，就得到了最终的解释变量 $X(w)$ ，列在表 2 的最后两列。程序估计得到的回归结果如表 3 所示。

表 2 春节模型的数据 ($w=20$)

年份	春节	20 天之前	# 天在 1 月 ($w_{i,1}$)	# 天在 2 月 ($w_{i,2}$)	$P_{i,1}(w)$	$P_{i,2}(w)$	$X_{i,1}(w)$	$X_{i,2}(w)$
1993	1 月 23 日	1 月 3 日	20	0	1	0	0.2731	-0.2731
1994	2 月 10 日	1 月 21 日	11	9	0.55	0.45	-0.1769	0.1769
1995	1 月 31 日	1 月 11 日	20	0	1	0	0.2731	-0.2731
1996	2 月 19 日	1 月 30 日	2	18	0.1	0.9	-0.6269	0.6269
1997	2 月 7 日	1 月 18 日	14	6	0.7	0.3	-0.0269	0.0269
1998	1 月 28 日	1 月 8 日	20	0	1	0	0.2731	-0.2731
1999	2 月 16 日	1 月 27 日	5	15	0.25	0.75	-0.4769	0.4769
2000	2 月 5 日	1 月 16 日	16	4	0.8	0.2	0.0731	-0.0731
2001	1 月 24 日	1 月 4 日	20	0	1	0	0.2731	-0.2731
2002	2 月 12 日	1 月 23 日	9	11	0.45	0.55	-0.2769	0.2769
2003	2 月 1 日	1 月 12 日	20	0	1	0	0.2731	-0.2731
2004	1 月 22 日	1 月 2 日	20	0	1	0	0.2731	-0.2731
2005	2 月 9 日	1 月 20 日	12	8	0.6	0.4	-0.1269	0.1269
$\bar{P}_{\cdot,j}(w)$					0.7269	0.2731		

表 3 回归模型

变量	参数估计	标准差	t 统计量	概率值
常数项	4.6399	2.97013	1.56	0.1213
用户自定义变量				
spring	88.1090	11.14360	7.91	0.0000
自动识别的异常值				
AO1996. Dec	138.5643	23.34831	5.93	0.0000
AO2003. May	-174.7068	23.35510	-7.48	0.0000
LS2003. Nov	-164.1717	29.20858	-5.62	0.0000

可以看出春节效应变量 spring 在 1% 的检验水平上具有显著性。采用这一春节模型，程序标识出的异常值只有 3 个，不具有明显的规律性。通过公式 $\hat{S}_{i,j} = \hat{b}X_{i,j}(w)$ 就可以估计得到春节效应，输出序列记为 sales_holl，具体数值见表 4。经过对照可以发现，表 1 中识别到的 4 个发生在 1 月份的 AO 异常值的估计系数，与表 4 中对应年份的春节效应的符号完全一致，由此判断，正是春节效应的存在导致最初的季节调整将这些时点标识为异常值。

表4 由春节模型估计得到的春节效应。1月和2月的值(其他月份的值为0)

年份	1月	2月
1993	24.0605	-24.0605
1994	-15.5885	15.5885
1995	24.0605	-24.0605
1996	-55.2376	55.2376
1997	-2.3722	2.3722
1998	24.0605	-24.0605
1999	-42.0212	42.0212
2000	6.4387	-6.4387
2001	24.0605	-24.0605
2002	-24.3994	24.3994
2003	24.0605	-24.0605
2004	24.0605	-24.0605
2005	-11.1831	11.1831

事实上,对于该序列而言,同季节影响相比,春节效应的影响幅度相对很小。在本例中,春节效应对1、2月份的平均调整幅度分别相当于原序列水平的1.12%和1.17%,而同期季节因素的平均调整幅度分别为5.48%和2.49%。使用乘法分解模型可以得出类似的结论,限于篇幅这里不再给出具体结果。春节效应(sales_hol1)、季节因素(sales_sf1)与异常值冲击(sales_otl1)的对比图见图2。

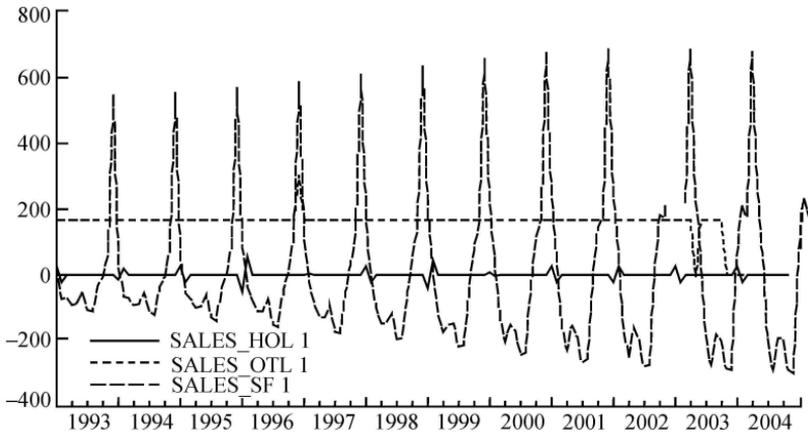


图2 春节效应(sales_hol1)、季节因素(sales_sf1)与异常值冲击(sales_otl1)

四、存量数据的基本春节模型

X-12-ARIMA 程序对存量数据中可能存在的移动假日影响没有提供内建的解释变量。经过与美国普查局 Brian. C. Monsell 联系确认,他们认为在实践中存量数据很少需要考虑移动假日效应,如果确实存在移动假日效应,也

是通过导入用户自定义回归变量进行处理的。

实际上，存量数据的统计同样会受到移动假日的影响。比如在我国，人们习惯上仍然大量使用现金进行消费，在春节等假日前夕人们往往会持有更多的现金，集中购物，这就必将影响到春节期间货币供应量的统计。因此，有必要研究针对存量数据如何进行移动假日效应的调整。

在概念上，流量数据是日度数据的累加值，在移动假日影响区间内，经济活动水平不同于非假日，将这段时期内每一天的数据加总，就构成了待调整变量总的移动假日效应，因此可以基于对这一区间的分解来分离移动假日效应，构造解释变量。这一解释变量必须满足互补性 (reciprocal) 约束，即移动假日对流量数据年度总和的净影响从长期来看保持某一定值，只不过在相邻月份 (或季度) 之间的分配比率会随着这一假日在公历中出现日期的移动而发生变动。比如在前面的例子中，1999 年春节出现在 2 月 16 日，则春节效应对当年 2 月的影响高于历史平均水平 $0.4769 \times 88.1090 \approx 42$ 亿元，相应地，当年 1 月的春节效应就低于历史平均水平约 42 亿元。

存量数据对应于观测登记日这一时点上的统计值，如果观测登记日落在移动假日的影响区间内，则待调整变量的统计就受到了假日的影响，否则，变量的统计就不受影响。因此对存量数据构造移动假日效应解释变量就不存在互补性约束。如果简单套用流量数据的移动假日模型，就会造成错误，人为引起过度调整 (参见 Hong Kong Monetary Authority, 2000)。

对存量数据构造移动假日回归变量，比流量数据更为复杂，需要涉及两个参数，即假日效应的影响期 (ω_1) 和存量数据的观测登记日 (ω_2)。仍然以春节为例加以说明。

假定春节对于待调整变量的影响主要发生在春节之前的 ω_1 天， $1 \leq \omega_1 \leq 20$ 。变量在每月的 ω_2 这一天进行观测登记， $\omega_2 = 31$ 表示每月的最后一天。对于给定的年份 i ，如果观测登记日 ω_2 落在影响期内，则定义变量 $P_{i,j}(\omega_1, \omega_2) = 1$ ，否则， $P_{i,j}(\omega_1, \omega_2) = 0$ 。给定 ω_1 的取值范围，该变量在除 1 月、2 月以外的其他月份的值都是 0。例如，当 $\omega_1 = 20$ 、 $\omega_2 = 31$ 时，变量在除 1 月以外的其他月份取值都是 0。对 $P_{i,j}(\omega_1, \omega_2)$ 进行中心化处理，减掉月份 j 在全部样本年份的均值 $\bar{P}_{.,j}(\omega_1, \omega_2)$ 得到实际的回归变量 $X_{i,j}(\omega_1, \omega_2)$ ，即 $X_{i,j}(\omega_1, \omega_2) = P_{i,j}(\omega_1, \omega_2) - \bar{P}_{.,j}(\omega_1, \omega_2)$ 。 $X_{i,j}(\omega_1, \omega_2)$ 表示春节影响时效为 ω_1 、观测登记日为每月 ω_2 时，第 i 年第 j 月对应的春节影响。模型可以写作 $Y_{i,j} = bX_{i,j}(\omega_1, \omega_2) + z_{i,j}$ ，其中 $Y_{i,j}$ 是对应第 i 年第 j 期 (月份或季度) 的待调整序列， $z_{i,j}$ 是回归误差。最后估计得到的春节效应为 $\hat{S}_{i,j} = \hat{b}X_{i,j}(\omega_1, \omega_2)$ 。

五、扩展的变权重流量数据春节模型

针对美国以外的国家 (地区) 使用 X-12-ARIMA 进行移动假日调整的需

要,美国普查局 Brian. C. Monsell 提供了一个专门生成移动假日解释变量的 genhol 程序,用户只需提供本地特定的移动假日所对应的在公历中的日期,同时设定假日效应的影响时效,程序就能够自动产生供 X-12-ARIMA 使用的移动假日解释变量的数据值。

genhol 程序将移动假日效应进一步细分为节前影响、节后影响和节日期间的影 响三类,对应引入三个解释变量,如图 3 所示。基本假设和前面一样,假定经济活动水平在影响期内保持不变,即服从均匀分布。对三类影响的时间长度可以分别加以设定,并通过最终的各种诊断检验方法,如 AIC 准则、样本外预测表现等选取最合适的影响时效。Lin and Liu (2003) 利用这一程序导入春节变量,对我国台湾地区十个重要的经济变量进行了分析。



图 3

上述春节模型均假定经济活动水平在影响期内保持不变,即服从均匀分布,这无疑是对现实情况的一种合理简化。通过对参数 w 的调整,一般总可以得到比较满意的季节调整结果。在没有日度数据作为建模依据的情况下,该模型保证了 X-12-ARIMA 程序具有广泛的适应性。但事实上有些经济变量所受到的春节影响并不服从这种均匀分布,如图 4 所示的香港金融管理局根据日度数据绘制的负债权证(certificates of indebtedness, CIs, 类似于基础货币)(除以年度平均数)在春节前后的变动情况。

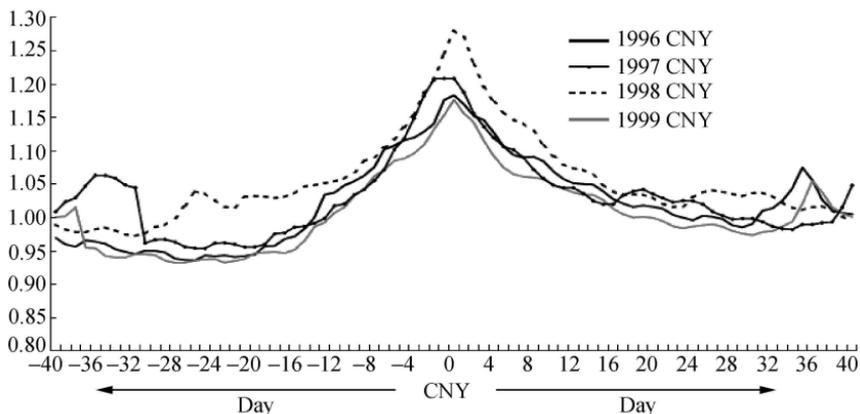


图 4 香港 CIs 对年度平均 CIs 的比率 (1996—1999)

(引自 Hong Kong Monetary Authority, 2000)

图 4 中 CNY 表示中国春节。从图中可以看出,在春节之前 CIs 持续攀升,到节日期间达到最大值,随后又持续下降。影响时效约为春节前后各 20 天。有鉴于此,结合 genhol 程序的三区段建模思想,我们假设春节效应在节

前和节后的日度权重各不相同，距离春节越近，影响强度越大，赋予的权重也越大；节日期间变量服从均匀分布，日度权重相等。构造扩展的三区段三变量流量数据春节模型如下。

假定：节前春节效应影响天数共 w_b 天，其影响强度呈线性增长，即节前第 w_b 天的影响权数为 $\frac{1}{w_b}$ ，节前第 $w_b - 1$ 天的影响权数为 $\frac{2}{w_b}$ ， \dots ，节前第 2 天的影响权数为 $\frac{w_b - 1}{w_b}$ ，节前第 1 天的影响权数为 $\frac{w_b}{w_b}$ 。

$$\sum_{i=1}^{w_b} i \qquad \sum_{i=1}^{w_b} i$$

节日期间的春节效应影响天数共 w_d 天，其影响强度保持不变，即每一天的影响权数相同。

节后春节效应影响天数共 w_a 天，其影响强度呈线性衰减，即节后第 1 天的影响权数为 $\frac{w_a}{w_a}$ ，节后第 2 天的影响权数为 $\frac{w_a - 1}{w_a}$ ， \dots ，节后第 $w_a - 1$ 天的影响权数为 $\frac{2}{w_a}$ ，节后第 w_a 天的影响权数为 $\frac{1}{w_a}$ 。

$$\sum_{i=1}^{w_a} i \qquad \sum_{i=1}^{w_a} i$$

然后根据节前、节中和节后影响天数落在 1 月、2 月和 3 月的具体情况，将各天的影响权数加总得到比例变量 $P_b(w_b)$ 、 $P_d(w_d)$ 和 $P_a(w_a)$ ，再分别进行中心化处理，构造节前、节中和节后回归变量 $X_b(w_b)$ 、 $X_d(w_d)$ 和 $X_a(w_a)$ ，导入 regARIMA 模型。若对应的 t 统计量不具有统计显著性，即表明该解释变量不具有解释力，从回归模型中消去，再重新估计模型，直到最后获得满意的模型。

使用节前、节中和节后三个变量进行三区段建模，由于每个变量的系数是分别估计的，这样就可以精确计算节前、节中和节后不同的春节影响；如果其中某个区段上的春节影响的方向与其他区段上的影响方向不同，变化不连续，则三个变量的估计系数就可能具有不同的符号。具有这种特征的流量数据是现实存在的，比如工业生产，在节日之前可能会出现加速，节日期间几乎完全停滞，节日之后缓慢恢复到正常水平。又比如春运期间的铁路客流量，在春节之前持续攀升，是正向变化；节日期间急剧缩减，低于正常水平，是负向变化；节后又从较高的峰值上逐渐减少，恢复到正常水平，春节影响也是正向的。如图 5 所示。因此，这一三变量三区段模型完全可以适用于具有 μ 型双峰分布特征的数据，具有良好的通用性。

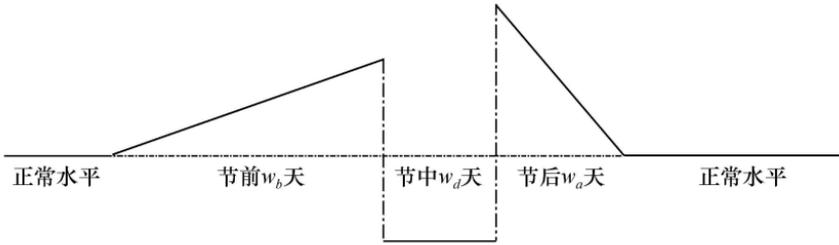


图 5

此外,使用这一模型,可以单独调整节前或节后某一个变量的参数(w_b 或 w_a),而不必重新计算其他两个变量的数据值;若其中某一个变量不具有显著性,也可以将其单独剔除,而保留其他具有解释力的变量,从而模型具有高度的灵活性。

仍以社会消费品零售总额时间序列为例加以说明。考虑节前影响期为 $w_b=20$ 天,节日期间影响期为 $w_d=7$ 天,节后影响期为 $w_a=20$ 天²,计算1993年至2005年的春节变量数据。表5给出了经过中心化的最终解释变量在1、2、3月份的具体数值。

表 5 扩展的流量数据春节模型的部分数据($w_b=20, w_d=7, w_a=20$)

年份	$Pb_{i,1}$	$Pb_{i,2}$	$Xb_{i,1}$	$Xb_{i,2}$	$Xd_{i,1}$	$Xd_{i,2}$	$Xa_{i,1}$	$Xa_{i,2}$	$Xa_{i,3}$
1993	1	0	0.374	-0.374	0.714	-0.714	0.143	-0.015	-0.128
1994	0.31	0.69	-0.311	0.311	-0.286	0.286	-0.042	-0.001	0.043
1995	1	0	0.374	-0.374	-0.143	0.143	-0.042	0.171	-0.128
1996	0.01	0.99	-0.611	0.611	-0.286	0.286	-0.042	-0.477	0.519
1997	0.5	0.5	-0.126	0.126	-0.286	0.286	-0.042	0.099	-0.057
1998	1	0	0.374	-0.374	0.286	-0.286	-0.042	0.171	-0.128
1999	0.07	0.93	-0.554	0.554	-0.286	0.286	-0.042	-0.329	0.372
2000	0.65	0.35	0.022	-0.022	-0.286	0.286	-0.042	0.156	-0.114
2001	1	0	0.374	-0.374	0.714	-0.714	0.053	0.075	-0.128
2002	0.21	0.79	-0.411	0.411	-0.286	0.286	-0.042	-0.091	0.134
2003	1	0	0.374	-0.374	-0.286	0.286	-0.042	0.171	-0.128
2004	1	0	0.374	-0.374	0.714	-0.714	0.229	-0.101	-0.128
2005	0.37	0.63	-0.254	0.254	-0.286	0.286	-0.042	0.171	-0.128

² 由于缺少日度数据作为建模依据,春节变量参数值 w_b 、 w_d 和 w_a 应当通过对季节调整总体效果的各种诊断检验进行选取,包括sliding spans(平移区间检验)、revision history(修正历史检验)、样本外预测效果以及AICC值(修正的赤池信息准则)等指标。在本例中参数值的确定主要是出于方便计算及比较的考虑。

表 6 回归模型

变量	参数估计	标准差	t 统计量	概率值
常数项	5.7130	3.48244	1.64	0.1036
用户自定义变量				
springb	67.1720	10.03807	6.69	0.0000
springd	46.1008	9.50848	4.85	0.0000
springa	29.0146	14.85066	1.95	0.0535
自动识别的异常值				
AO1993. Jan	-177.2337	39.70053	-4.46	0.0000
AO1996. Dec	127.6255	17.11331	7.46	0.0000
AO2003. May	-163.9596	17.15321	-9.56	0.0000
LS2003. Nov	-158.0010	23.86389	-6.62	0.0000

程序估计得到的回归结果见表 6。可以看出节前效应变量 springb 和节中效应变量 springd 都具有很强的显著性，而节后春节效应变量 springa 的 t 值也达到了 1.95，接近于 5% 的显著性水平，故在模型中予以保留。节前效应的显著性最强，影响最大。

分别计算各个月份的节前、节中和节后效应： $\hat{b}_1 X_b(\omega_b)$ 、 $\hat{b}_2 X_d(\omega_d)$ 和 $\hat{b}_3 X_a(\omega_a)$ ，将其加总就得到了综合的春节效应 $\hat{S} = \hat{b}_1 X_b(\omega_b) + \hat{b}_2 X_d(\omega_d) + \hat{b}_3 X_a(\omega_a)$ ，其中 1、2、3 月份的数值见表 7。经对照可以看出，由扩展的春节模型所得到的春节效应与基本模型的差别并不很大，再次印证了基本模型的结论，即对于社会消费品零售总额序列而言，同季节影响或异常值冲击相比，春节效应的影响幅度相对很小。从而也说明基本春节模型对于一般的应用而言能够提供较为满意的调整效果。

表 7 由扩展的春节模型估计得到的综合春节效应：1 月、2 月和 3 月的值
(其他月份的值为 0)

年份	1 月	2 月	3 月
1993	62.2311	-58.5113	-3.7198
1994	-35.3189	34.0647	1.2541
1995	17.3277	-13.6079	-3.7198
1996	-55.4704	40.3998	15.0706
1997	-22.8441	24.4914	-1.6473
1998	37.0852	-33.3654	-3.7198
1999	-51.6320	40.8445	10.7875
2000	-12.9282	16.2335	-3.3053
2001	59.6060	-55.8862	-3.7198
2002	-42.0361	38.1568	3.8793
2003	10.7419	-7.0221	-3.7198
2004	64.7181	-60.9983	-3.7198
2005	-31.4805	35.2003	-3.7198

六、扩展的变权重存量数据春节模型

采取和流量数据类似的思路,分节前、节中和节后三个区段进行建模,构造春节模型。

假定:节前春节影响呈线性增长,影响天数共 w_b 天,则节前第 w_b 天的影响权数为 $\frac{1}{w_b}$, 节前第 $w_b - 1$ 天的影响权数为 $\frac{2}{w_b}$, ..., 节前第 2 天的影响权数为 $\frac{w_b - 1}{w_b}$, 节前第 1 天的影响权数为 $\frac{w_b}{w_b} = 1$ 。

节日期间的春节影响保持不变,即每一天的影响权数相同,均为 1。

节后春节影响呈线性衰减,影响天数共 w_a 天,则节后第 1 天的影响权数为 $\frac{w_a}{w_a} = 1$, 节后第 2 天的影响权数为 $\frac{w_a - 1}{w_a}$, ..., 节后第 $w_a - 1$ 天的影响权数为 $\frac{2}{w_a}$, 节后第 w_a 天的影响权数为 $\frac{1}{w_a}$ 。

然后考察观测登记日 w_2 具体落在节前、节中和节后哪一个影响区段的哪一天,计算对应的影响权数,得到变量 $P(w_b, w_d, w_a, w_2)$, $w_2 = 31$ 表示每月的最后一天,再进行中心化处理,构造回归变量 $X(w_b, w_d, w_a, w_2)$, 加入 regARIMA 模型,若对应的 t 统计量不具有统计显著性,即表明该解释变量不具有解释力,则调整参数 w_b 、 w_d 和 w_a 的取值,再重新估计模型,直到最后获得满意的模型。

考虑到存量数据,如人口总量,一般而言都是连续变化的,具有较强的惯性,所受到的春节影响通常是同向的,因此这里采用的是单变量三区段的形式。

以中国货币供应量月度时间序列 M0 为例加以说明(1995 年 1 月—2005 年 7 月,数据来源:中国人民银行调查统计司)。未调整的 M0 序列如图 6 所示,可以看出其中存在明显的季节性波动。首先不考虑春节影响,使用乘法分解模型进行调整,考虑存量交易日效应,令程序自动选取合适的 ARIMA 模型、自动探测异常值。³得到的调整后序列如图 7 所示,可以看出调整结果并不理想。与此相对应,输出结果中程序自动标识出的异常值多达 6 个,分别为 AO1996.1、AO1996.2、AO1999.1、AO1999.2、AO2002.1 和 AO2003.1,全部分布在 1 月和 2 月。说明春节效应的存在严重影响季节调整的效果。

³ 该序列样本数为 127,故程序默认的异常值探测的临界值为 3.86。

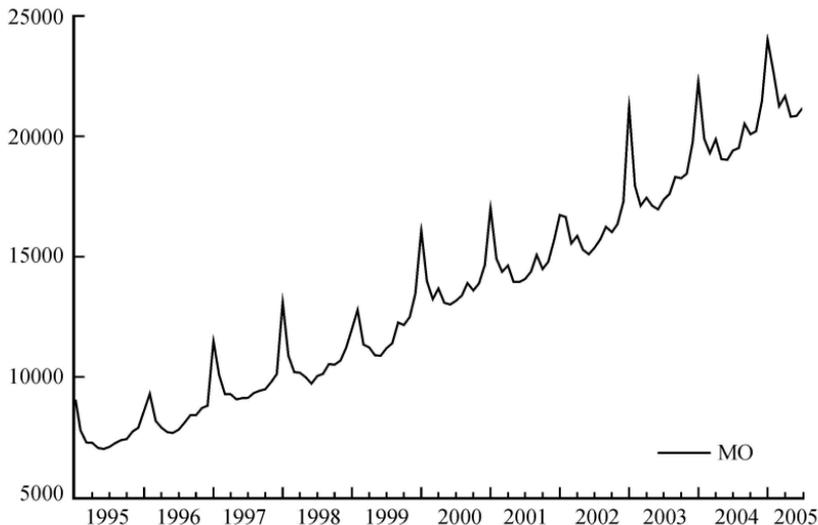


图 6 未调整的 M0 序列 (1995 年 1 月—2005 年 7 月)

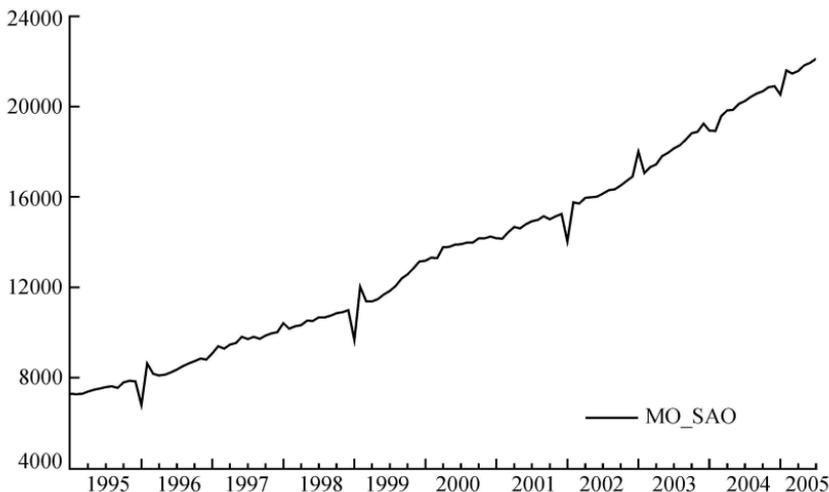


图 7 未考虑春节效应的季节调整后序列

下面引入扩展的存量春节模型。经过对多组参数值调整效果的综合比较，最终确定节前影响期为 $\tau_{w_b}=15$ 天，节日期间影响期为 $\tau_{w_d}=3$ 天，节后影响期为 $\tau_{w_a}=20$ 天，提供 1970—2020 年的春节变量 $P(15, 3, 20, 31)$ 的数据（程序在进行中心化处理时也将采用 1970—2020 年的样本均值）。其中 1995 年至 2005 年 1 月和 2 月的部分数值见表 8。程序估计得到的回归结果见表 9。

注意到自定义的三区段存量春节效应变量 spring 的 t 统计量达 29.05，具有极强的显著性。程序自动识别出的异常值只有 1997 年 1 月的 1 次暂时变动，体现了当时的货币政策的影响。分别计算各年份的春节效应调整因子 $e^{\hat{b}X(w_b, w_d, w_a, w_2)}$ ，其中 1 月和 2 月的数值列在表 8 的最后两列。春节效应与季节

因素、异常值冲击和交易日效应的对比图见图 8。从图中可以看出,对于中国货币供应量 M0 序列而言,春节效应的影响是相当大的,若不能予以适当消除,势必不能取得满意的调整效果。经计算可知,春节效应对 1、2 月份的平均调整幅度分别约为原序列水平的 4.71% 和 3.10%。

表 8 扩展的存量数据春节模型的数据 ($w_b=15, w_d=3, w_a=20, w_2=31$)

年份	$P_{i,1}$	$P_{i,2}$	$X_{i,1}$	$X_{i,2}$	$e^{\beta X_{i,1}(w_b, w_d, w_a, w_2)}$	$e^{\beta X_{i,2}(w_b, w_d, w_a, w_2)}$
1995	1.0000	0.0000	0.4239	-0.1833	1.0675	0.9722
1996	0.0000	0.6500	-0.5761	0.4667	0.9151	1.0745
1997	0.6000	0.1000	0.0239	-0.0833	1.0037	0.9872
1998	1.0000	0.0000	0.4239	-0.1833	1.0675	0.9722
1999	0.0000	0.5500	-0.5761	0.3667	0.9151	1.0581
2000	0.7333	0.0000	0.1572	-0.1833	1.0245	0.9722
2001	0.8000	0.0000	0.2239	-0.1833	1.0351	0.9722
2002	0.2667	0.3500	-0.3095	0.1667	0.9535	1.0260
2003	1.0000	0.0000	0.4239	-0.1833	1.0675	0.9722
2004	0.7000	0.0000	0.1239	-0.1833	1.0193	0.9722
2005	0.4667	0.2000	-0.1095	0.0167	0.9833	1.0026
2006	1.0000	0.0000	0.4239	-0.1833	1.0675	0.9722

表 9 回归模型

变量	参数估计	标准差	t 统计量	概率值
存量交易日效应				
Mon	0.0007	0.00194	0.37	0.7121
Tue	-0.0063	0.00191	-3.29	0.0014
Wed	-0.0011	0.00196	-0.59	0.5565
Thu	-0.0002	0.00193	-0.11	0.9126
Fri	0.0008	0.00192	0.39	0.6973
Sat	0.0027	0.00195	1.38	0.1705
* Sun (derived)	0.0035	0.00181	1.92	0.0576
用户自定义变量				
spring	0.1540	0.00530	29.05	0.0000
自动识别的异常值				
TC1997. Jan	0.0740	0.01177	6.28	0.0000

接下来进一步通过季节调整后序列来考察春节效应的相对大小及季节调整的效果。程序自动输出的季节调整后序列 (M0 __sa) 消除了季节因素、交易日效应和移动假日效应等日历效应,但并未消除异常值的影响。用原序列 (M0) 除以季节因素 (M0 __sf), 得到的是只消除了季节影响的序列, 记为 M0 __sp。绘制折线图如图 9 所示。从图中可以看出,不消除春节效应的季节调整后序列 (M0 __sp) 是不光滑的,所有的犄角点都分布在 1 月和 2 月;而利用我们所建立的存量数据春节模型能够很好地消除春节影响,所得到的调整后序列 (M0 __sa) 充分光滑,完全可以满足季节调整的要求,唯一一个明显的犄角点 (TC1997. Jan) 可以解释为一次异常值冲击;进一步消除异常值

之后的序列 ($M0_so$) 已经没有明显的犄角点。

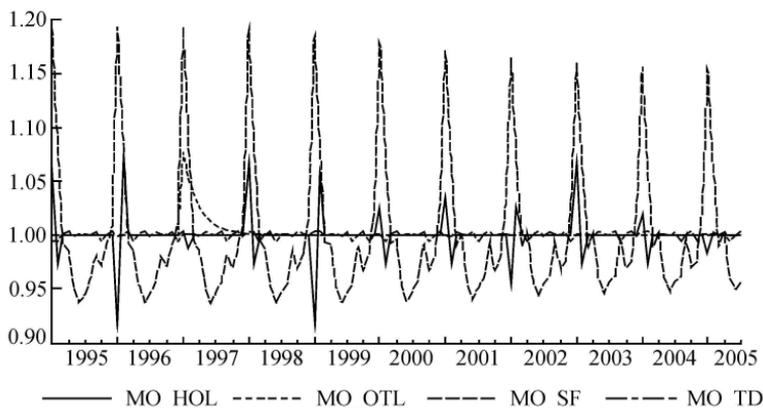


图 8 春节效应 ($M0_hol$)、季节因素 ($M0_sf$)、异常值冲击 ($M0_otl$) 与交易日效应 ($M0_td$)

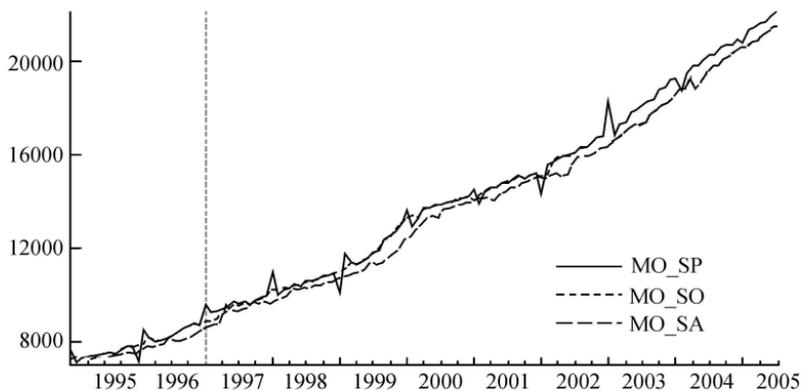


图 9 三种季节调整后序列的比较

七、结 论

本文首先在美国普查局 X-12-ARIMA 季节调整程序的复活节模型基础上，建立了基本的流量数据春节效应模型，进而界定了流量数据与存量数据在性质上的差别，构造了存量数据春节模型。在此基础上对春节效应的影响区间进一步细分，得到了扩展形式的多区段变权重流量春节模型和存量春节模型。分别运用这两个模型对社会消费品零售总额和货币供应量 $M0$ 月度时间序列进行季节调整，结果表明，这两个模型能够很好地对春节效应作出估计，因此可以用于解决我国春节等移动假日的调整问题。

参 考 文 献

- [1] Ladiray, D., and B. Quenneville, *Seasonal Adjustment with the X-11 Method*. New York: Springer-Verlag, 2001.
- [2] Hong Kong Monetary Authority, "Seasonal Adjustment of Hong Kong's Monetary Statistics", *Hong Kong Monetary Authority Quarterly Bulletin*, 2000, 11, 1—20.
- [3] 贾淑梅, "货币供应量季节调整中消除春节因素的实证研究", 《统计研究》, 2005年第10期, 第63—68页。
- [4] Lin, Jin-Lung., and T. S. Liu, "Modeling Lunar Calendar Holiday Effects in Taiwan", 《台湾经济预测与政策》, 2003年第33卷第2期, 第1—37页。
- [5] McIntire, R. J., "A Procedure to Control for Moving-holiday Effects in Seasonally Adjusting Employment and Hours Series", ASA Proceedings of the Business and Economic Statistics Section, American Statistical Association (Alexandria, VA), 1990, 443—448.
- [6] 齐东军, "季节调整方法在货币供应量中的应用", 《数量经济技术经济研究》, 2004年第6期, 第147—155页。
- [7] U. S. Census Bureau, *X-12-ARIMA Reference Manual*, Version 0.2.10, 2002.
- [8] 夏春, "实际经济时间序列的计算、季节调整及相关经济含义", 《经济研究》, 2002年第3期, 第36—43页。
- [9] 张鸣芳、项燕霞和齐东军, "居民消费价格指数季节调整实证研究", 《财经研究》, 2004年第3期, 第133—144页。

A Spring Festival Model for Seasonal Adjustments

HUIDE LUAN XIAOTONG ZHANG

(Nankai University)

Abstract It is a challenging issue in seasonal adjustments for China's monthly data to estimate and eliminate holiday effects such as the Chinese Spring Festival. This paper constructs a Spring Festival Model for flow data, and then proposes a Spring Festival Model for stock data allowing for the difference between flow data and stock data. On the basis of these two models, we formulate 3-period and varying-weights augmented Spring Festival Models for both flow data and stock data. Applying these models to retail sales and M0 series proves that they can eliminate Spring Festival effect in a valid way.

JEL Classification C87, C14, C22