

SSE-DP-2025-1-C

移动节假日与经济指标的季节性

国友直人（编辑责任人）

统计数理研究所

2025年12月

SSE—DP(Discussion Papers Series) can be downloaded without charge from: <https://stat-expert.ism.ac.jp/training/discussionpaper/>

Discussion Papers are a series of manuscripts in their draft form. They are not intended for circulation or distribution except as indicated by the author. For that reason, Discussion Papers may not be reproduced or distributed without the written consent of the author.

移动节假日与经济指标的季节性 *

国友直人（编辑责任人）†

2025 年 2 月

关键词 (Key Words)：移动节假日、春节、中国的宏观经济数据、存在缺失值的非平稳季节时间序列、X12-ARIMA、X13-ARIMA-SEATS、SarSIML (S-SIML) 季节调整

摘要

作为统计数学研究所推进的统计专家培训项目的一部分，我们在 2024 年度实施了以“移动节假日与经济指标的季节性”为主题的统计咨询专题研究讨论。在讨论会中，我们听取了内阁府（经济财政分析・海外担当）中实际负责分析中国经济动向的人员关于中国宏观经济数据分析中季节性相关问题的说明，并以“在中国发布的宏观经济数据”为题材，探讨了其中的季节性处理问题，以及美国人口普查局发布的“X-13ARIMA-SEATS”、“Kitagawa 的 DECOMP”、“Sato 的 S-SIML”等方法在实际使用中的问题。

本报告首先探讨了使用中国历（农历）处理宏观经济数据时所面临的问题，特别是与公历（阳历）和农历的差异、春节效应以及节假日效应等相关的话题。在实际应用中，尤其是涉及季节调整方法如 X-12-ARIMA 和 X-13ARIMA-SEATS 的使用问题显得尤为重要。关于中国的季节性和日历效应，本文开发了赵-虚拟变量（Zhao Dummy Variables），并通过分析月度数据的中国贸易统计中的季节性，进一步探讨了在包含季节性的非平稳时间序列中，当存在缺失值时的滤波问题，以及如何构造

*本报告是统计专家培训项目中的统计咨询专题研究讨论“移动节假日与经济指标的季节性”的成果。演习的参与成员包括国友小组的研修相关人员、内阁府政策统括官（经济财政分析担当）附参事官（海外担当）的负责人（如花垣贵司）、以及协作人员赵宇（东京理科大学）、高冈慎（琉球大学）、佐藤整尚（东京大学）。本报告的内容不代表内阁府或统计数学研究所的官方立场。

†统计数理研究所

季节调整后的系列等内容。此外，报告还介绍了一种基于频率分解的新型季节调整方法 SarSIML（S-SIML）法，指出该方法在处理具有季节性且存在缺失值的非平稳时间序列时也能较为方便地加以应用。

目录

1. 引言
2. 中国的日历与经济活动：赵-虚拟变量（赵宇・国友直人）
 <附录> 赵-虚拟变量
3. 中国贸易统计与季节调整（高冈慎）
4. 非平稳季节性经济时间序列中的缺失值问题（国友直人）
5. SarSIML 季节调整（佐藤整尚・国友直人）
6. 有关季节性与季节调整的讨论（国友直人）
7. 结语

引言

日本、美国、中国等国家的官方统计部门会定期公布大量宏观时间序列的季节调整值。这是因为在多数原始系列中显著存在季节性波动，为了正确理解经济现状、合理制定政策，需要去除原始数据中的季节变动成分。作为季节调整方法，日本、美国、中国等多数国家的官方统计部门原本使用的是由美国人口普查局开发的 X-11、X-12-ARIMA 方法，但截至 2024 年 12 月，X-12-ARIMA 已停止维护，人口普查局推荐使用 X-13ARIMA-SEATS。

在这一背景下，统计数理研究所主导的统计专家培养项目于 2024 年度组织了“移动节假日与经济指标的季节性”统计咨询专题研究讨论。这是由于在内阁府对海外经济形势进行分析的过程中，负责中国宏观经济分析的人员遇到了统计上的难题，从而发起了这项研讨。随着中国经济的扩张，其与日本经济、世界经济的联动性增强，中国的宏观经济动向，特别是近年来的大幅波动，已成为日本经济学家高度关注的焦点。然而，中国负责发布官方统计的机构出于若干原因，在数据发布形式上有别于主要发达国家，这使得掌握中国实体经济的实际动向存在困难。例如，关于重要的宏观经济指标，即便是月度数据，也经常不发布 1 月单月数据，而是在之后只公布 1 月和 2 月的合计数据，这一做法已经实行多年。

在本次“统计专家培养项目”中的“统计咨询演习”中，我们首先听取了内阁府负责中国宏观经济分析的相关人员对中国宏观经济数据中季节性处理问题的介绍。虽然参与研修者中包括中国出身的人员，但实际上没有中国经济动向熟悉的专家，因此我们借助了以下资料来加深理解：

- 丸川知雄《现代中国经济》（新版，有斐阁，2021 年）- 梶谷懷《中国经济讲义》（中公新书，2018 年）- 田中直毅・佐藤整尚《经济预测与因果解析》（日本评论社，2024 年）。

此外，咨询演习还邀请了熟悉季节调整问题的两位专家，佐藤整尚（东京大学）与高冈慎（琉球大学），在他们的协助下，演习开展了自由而积极的讨论。国友在 2024 年度为该项目开设了题为“统计时间序列解析”的课程（使用北川源四郎《用 R 入门时间序列建模》，岩波书店，2021 年，Zoom 授课 13 次），来自内阁府的多位成员参与了课程，借此机会掌握了统计时间序列分析的基础知识。在第 12、13 讲中，佐藤教授还特别讲授了利用 R 与 Google Colaboratory 进行前沿经济时间序列分析的方法。

在统计咨询演习的讨论基础上，本报告首先由中国出身的原研修生赵宇（现为东京理科大学教员）与国友直人一起探讨了“中国日历制度与经济活动”的相关问题。我们分析了目前广泛使用的季节调整程序（如 X-12-ARIMA、X-13ARIMA-SEATS）中，如何从原始序列中减去季节成分以构建季节调整序列，并指出由于这些方法通过 RegARIMA 模型估计季节性，需要引入包括星期效应在内的回归变量。因此，我们基于对中国制度性安排的检讨，构造了用于分析中国数据的虚拟变量，称为“赵虚拟变量”。利用该变量，高冈教授分析了中国的贸易统计数据，该数据是在中国与日本海关业务过程中直接记录的，相较其他宏观数据，估计误差可能较小（虽然不能完全排除走私等因素的影响）。

接下来，我们考虑了含季节性特征的非平稳宏观经济时间序列中的缺失值问题。这是一个在文献中几乎未被探讨但在实际应用中十分关键的问题。在此背景下，我们检讨了季节调整方法，特别是国友与佐藤正在共同开发的 SIML 分析方法。佐藤教授基于该方法，开发了一个新型的季节调整程序，称为 SarSIML 或 S-SIML 季节调整方法。与 X-12-ARIMA、DECOMP 等传统方法相比，该方法构想虽简单，但能直接从原始序列中去除 Gregorian 季节性，是一种操作上更为简便的季节调整方案。

此外，截至 2024 年，美国、日本、中国、欧洲等主要国家的官方统计普遍使用的是基于 Gregorian 季节性的 X-13ARIMA-SEATS 方法¹。然而，从统计科学的角度看，这些方法的机械性地使用存在诸多问题。既有方法尽管结合了传统时间序列分析与复杂的工程性设计，但其季节性定义不够明确，且在面对如中国阴历影响等特殊情形时存在局限。因此，本文尝试从更基本的统计学视角出发，提出可替代 X-12-ARIMA、X-13ARIMA-SEATS 的新方法——S-SIML。需要说明的是，S-SIML 方法并非专为中国经济设计，它也可广泛用于日本、美国等国家的经济时间序列分析。

我们希望本报告能为今后日本理解中国经济动态、乃至更广泛的季节调整问题提供有益参考。

2025 年 1 月

国友直人

¹有关季节性调整法 X-12-ARIMA, X-13ARIMA-SEATS, 请参考统计专家培训
DP(2023/03/SSE-DP-2023-1.pdf)
<https://stat-expert.ism.ac.jp/wp/wp-content/uploads/2023/03/SSE-DP-2023-1.pdf>

中文版引言

统计数理研究所推进的“统计专家培训项目”在2024年度的统计咨询演习中，策划了名为“移动假日与经济指标的季节性”的课题。这是应内阁府负责分析海外经济形势，特别是中国宏观经济的人员提出的统计咨询请求而策划的项目。该项目的部分研究结果已整理为日语报告书SSE-DP-2025-1并予以公布。

有幸的是，统计专家培训项目中有来自中国的学员参加。我们认为将其翻译成中文或许对中国今后的官方统计工作有所裨益，因此启动了翻译项目，并于近期完成了中文版的制作。翻译项目由赵宇（东京理科大学经营学部）担任翻译负责人，并按以下分工进行了翻译工作：

- “引言”：张文婷 (Zhang WenTing)
- “中国的日历与经济活动：赵-虚拟变量（赵宇・国友直人）”：赵宇 (Zhao Yu)
- “中国贸易统计与季节调整（高冈慎）”：何协 (He Xie)
- “非平稳季节性经济时间序列中的缺失值问题（国友直人）”：薛玉杰 (Xue YuJie)
- “SarSIML 季节调整（佐藤整尚・国友直人）”：薛玉杰 (Xue YuJie)
- “有关季节性与季节性调整的讨论（国友直人）”：张文婷 (Zhang WenTing)
- “结语”：张文婷 (Zhang WenTing)

此外，包括咨询负责人在内，由于我们并非中国经济的专家，不排除内容中可能存在对中国经济和公布数据的意想不到的误解或理解不充分之处。包括这些问题在内，我们诚挚欢迎相关各位的批评指正。如果本书能为理解今后关于中国官方统计的讨论及季节性调整问题提供些许帮助，我们将不胜荣幸。

在此，向参与本翻译项目的各位学员表示感谢。

2025年12月 国友直人

中国的日历与经济活动²

2024-10-15

赵宇³

国友直人⁴

1. 引言

近年来，随着中国经济的扩张，其与日本经济及世界经济的关联及其影响力也在不断增强。中国宏观经济的动向，特别是近期的剧烈波动及其走向，已成为日本经济学家极为关注的话题。然而，负责中国官方统计的部门有时会采取与主要发达国家不同的形式发布宏观信息，这使得把握中国实体经济的动向变得困难。例如，对于重要的宏观经济指标，统计部门会采取事后公布1月和2月合计数据的方式。这种做法可以解释为，统计部门考虑到春节等因素，为了避免因与其他主要发达国家不同的季节性波动而导致公布的数值在国内外引起误解所采取的措施。另一方面，在日本及其他发达国家的经济政策制定者和经济学家之间，利用主要宏观经济数据时，通常依赖于从原始序列生成的季节调整序列数值。各国负责统计编制和发布的人员通常使用美国普查局开发的X-12-ARIMA法或X-13ARIMA-SEATS法作为季节调整方法，但对于不一定遵循格里高利历（公历/阳历）的太阴历（农历/阴历）日历因素，或者原始数据存在缺失值的情况，应如何适当地处理季节性等基本的统计问题，其实并不明确。

本文首先作为分析和讨论已公布的宏观经济数据的前提，为了整理关于中国日历的变迁与经济活动的论点，总结了关于中国日历与宏观经济活动的基础事项和注意事项，并说明了为中国宏观经济分析而创建的季节虚拟变量（Dummy Variable）。不熟悉中国情况的国友直人向来自中国的赵宇（专业为运筹学 Operations Research）提出了一些朴素的问题，由于其中包含不少能加深对中国经济活动理解的信息，因此我们将两人实际交换意见的内容整理为(Q: 国友) - (A: 赵)的形式刊载于此。所

²本文报告的部分内容由赵宇于2024年2月通过网络等方式调查所得。此后，部分网络信息可能因各种原因有所更新。但由于通过同样的方法获得类似信息仍是可能的，故在此按原样刊载。需要注意引用文中的字体，如%，_等。

³东京理科大学

⁴统计数理研究所

创建的中国虚拟变量（赵-虚拟变量）已被例如高冈 (2025) 所使用，我们计划将其公开以供下载。

2. 中国日历的基本事项

(Q) 请说明一下中国日历与经济活动的基础事项。

(A) 在中国，有一些基于农历（太阴历）确定的传统节日，这些节日的日期每年都在变动。例如，春节⁵是农历的正月初一，因此日期每年不同。端午节在农历五月初五庆祝，这一天许多人会吃粽子。中秋节是农历八月十五，这一天有赏月吃月饼的习俗。这些节日的日期每年都会变化。另一方面，也有一些基于公历（阳历）的固定节日。例如，劳动节固定在每年的5月1日，国庆节固定在10月1日。清明节稍微复杂一些，通常在4月4日或5日。这一天人们有扫墓祭祖的习俗。然而，这些只是一般规则，实际的休假和调休（换班）日期，每年由中国国务院正式发布。例如，(公历) 2024年2月9日是2023年的“除夕”，通常是休息日，但根据国务院的决定，这一天变成了工作日。2024年的法定节假日可以通过以下网站确认：

https://english.www.gov.cn/policies/latestreleases/202310/25/content_WS65387be8c6d0868f4e8e0a04.html

(Q) 英语文献中有提到 the Dragon Boat Festival (May-June) 和 Mid-Autumn Festival (Sep-Oct)，前者是指端午节，后者是指中秋节吗？

(A) 是的，没错。

(Q) 关于春节，日本经常有报道，那么端午节和中秋节对经济活动的影响也很大吗？

(A) 端午节和中秋节也都是中国的法定节假日。因为分别有3天的假期，所以外出旅游的人也不在少数。关于对经济活动的影响，虽然可能不如春节、劳动节、国庆节那么大，但因为有各种活动和仪式，我认为其影响比普通的周末要大。（当然这只是个人观点。）

(Q) 对于普通上班族来说，周六周日是休息日吗？

(A) 是的，没错。

(Q) 能否请您总结并说明一下关于休假的情况？

(A) 如果我的经验和认知没有错的话，中国的休息日包括周六日和法定节假日。关于法定节假日，正如之前的邮件回复所述，由于很多节日基于农历，每年都由中国国务院正式发布。2024年的法定节假日可以从以下网站确认：

⁵春节, Spring Festival

https://english.www.gov.cn/policies/latestreleases/202310/25/content_WS65387be8c6d0868f4e8e0a04.html

关于之前的年份，可以通过万年历进行确认。过去通常通过小册子等确认，现在则可以在网上查看中国的万年历。不过，2024年及以后的节假日，以该年中国国务院发布的信息为准。例如，以下网站具有参考价值：

<http://www2s.biglobe.ne.jp/~yoss/worldtime/radnelac\%20ranul\%20esenlhc.html>

附言1：下表是中国国务院发布的2024年法定节假日。

Name Date Duration:

New Year's Day 元旦 Dec 30, 2023 — Jan 1, 2024 three days

Spring Festival 春节 Feb 10 — Feb 17 eight days

Qingming Festival 清明节 April 4 — April 6 three days

Labor Day 劳动节 May 1 — May 5 five days

Dragon Boat Festival 端午节 June 8 — June 10 three days

Mid-Autumn Festival 中秋节 Sept 15 — Sept 17 three days

National Day 国庆节 Oct 1 — Oct 7 seven days

附言2：例如有以下新闻

<https://www.reuters.com/world/asia-pacific/china-dragon-boat-festival-tourism-trips-above-pre-covid-levels-ministry-2023-06-25/>

<https://www.globaltimes.cn/page/202306/1293068.shtml>

由此可见，端午节也被认为对经济活动产生了重大影响。

<https://www.reuters.com/world/asia-pacific/china-dragon-boat-festival-tourism-trips-above-pre-covid-levels-ministry-2023-06-25/>

<https://www.globaltimes.cn/page/202306/1293068.shtml>

(Q) 感谢您提供的信息。您提到顺序是春节（日期每年变化）、劳动节、国庆节，那么假期的长度都是3天吗？关于春节，这是否意味着假期前后的影响有所不同？例如，日本引入消费税时定在4月1日（这是日期可以预先预测的情况，但两次实施的模式差异很大），因此产生了抢购需求及其反向作用，造成了相当大的影响（我记得特别是对住宅价格、耐用消费品价格以及消费者物价指数CPI的影响很大）。

(Q) 附言：刚才的资料里写了假期的长度。在日本，也有夹着连休休假的模式，但随

着时代变迁，季节性对企业和个人的影响也在发生变化。我预计直接利用 X-13 中的虚拟变量可能行不通。像 COVID-19 这样（意料之外的）突然变化点（统计学上非平稳时间序列中的 change point）的处理似乎也将成为本次的主题。看起来有很多研究课题，所以我接受了这次统计咨询的工作。

3. 关于赵-虚拟变量

关于中国的节假日虚拟变量，我附上了数据文件（Excel）。您委托的时间序列数据存储在“数据”工作表中。

注 1：如果修改“函数表”或“2002”等工作表，其他数据会发生联动，因此请勿更改。

注 2：2007 年之前，法定节假日为元旦、春节、劳动节、国庆节。2008 年节假日相关法律进行了修订，法定节假日变为元旦、春节、劳动节、清明节、端午节、中秋节、国庆节。

[https://en.wikipedia.org/wiki/Public\\$_holidays\\$_in\\$_China](https://en.wikipedia.org/wiki/Public$_holidays$_in$_China)

4. 关于中国日历与经济活动的一些理解

(Q) Wiki 上的“A notable feature of mainland Chinese holidays is that weekends are usually swapped with the weekdays next to the actual holiday to create a longer holiday period.”（中国大陆节假日的一个显著特征是，周末通常会与节假日附近的平日互换，以形成更长的假期。）以及“2002 年 1 月 天数 31, 周六日 8, 官方休日数 11, 营业日天数 20”“2002 年 2 月 天数 28, 周六日 8, 官方休日数 11, 营业日天数 17”这里的含义，并不是 $31-8-11=22, 28-8-17=13$ ，而是：(i) 假日和周末进行了互换（swap），(ii) 假日和周末重叠，意思是（典型的工厂）生产线上的工作日为 $31-8-11-2(\text{调整})=20, 28-8-11+4(\text{调整})=17$ 吗？

(Q) 与消费相关的百货商店、便利店等在节假日也是营业的吧？在生产线上，日本的钢铁高炉应该只有在盂兰盆节（Obon）和正月才会停止运作，但根据农林业等行业形态的不同，情况似乎差异很大，需要注意。国家和行业不同，休日和营业日数的计算方法似乎也多种多样。

(Q) 以日本为例，在矿工业生产指数（IIP）等情况下，经济产业省会采用独特的方法制作考虑了节假日天数的虚拟变量，并将其纳入 Reg-ARIMA 模型，通过 X-12-ARIMA 制作并公布季节调整值。

(A) “2002 年 1 月：天数 31, 周六日 8, 官方休日数 11, 营业日天数 20”“2002 年 2 月：天数 28, 周六日 8, 官方休日数 11, 营业日天数 17”周六日有时会因为调休等原因变更为工作日。因此，没有被变更为工作日的周六日，才作为休息日处理。在实际统计官方休日数时，使用“未变更为工作日的周六日 (A) + 法定节假日 (B) - A 与 B 重叠的天数”。统计营业日天数时，使用“该月周一至周五的天数 - 因法定节假日变更为休息日的周一至周五的天数 + 变更为工作日的周六日天数”。因此，“天数 = 未变更为工作日的周六日 (A) + 法定节假日 (B) - A 与 B 重叠的天数 + 周一至周五的天数 (C) - 因法定节假日变更为休息日的周一至周五的天数 (D) + 变更为工作日的周六日天数 (E)”这一公式成立。在此，注意到 $A + E = \text{周六日总天数}$, $(\text{法定节假日} (B) - A \text{ 与 } B \text{ 重叠的天数}) - (\text{因法定节假日变更为休息日的周一至周五的天数} (D)) = 0$, 天数可以整理为：天数 = 周六日总天数 + 周一至周五的天数。因此，天数 31 天 = 官方休日数 11 天 + 营业日天数 20 天。

(Q) 与消费相关的百货商店、便利店等在节假日也是营业的吧？

(A) 即使是节假日，大部分百货商店和便利店应该也是营业的。不过，我在外旅居多年，不确定这一信息是否符合现状。我的印象中即便是周六日或节假日，它们也都是营业的。

(A) (i) 假日和周末进行了互换，(ii) 假日和周末有时会重叠，关于 (i) 和 (ii)，可以在“2002”～“2014”的工作表中，通过 I 列和 J 列进行详细确认。I 列的“1”表示法定节假日期间。J 列的“1”意味着周六日被变更为了工作日。

(Q) 感谢您的说明。各国经济学家在判断中国经济景气时，似乎常参考制造业 PMI，但探讨星期效应 (Day-of-the-week effect) 似乎是有用的。至少日本内阁府的分析负责人朋友似乎也在参考默认 Eviews (即 X-11) 的季节调整值。既然那不能说是妥当的，那么至少进行“统计咨询”是有意义的。

附言：从营业天数来看，春节所在的 1 月～2 月的营业日并没有那么少，这一点也请确认。如果这是正确的，给人的印象是可能存在仅靠营业天数无法捕捉的问题。举例来说，“若以月球的月龄约 29.53 天为基准”，春节效应可以被视为非平稳季节时间序列中特定频域效应的测量。

5. 结语

中国的宏观经济数据提供了若干迄今为止在经济统计中未被讨论过的有趣论点。例如，在美国和欧洲许多国家，为了理解宏观经济状况，会制作季节调整序列，并通过

报纸、电视新闻等大众媒体公布，使用的是 X-12-ARIMA, X13ARIMA-SEATS 等季节调整程序。这种季节调整隐含地利用了格里高利历（公历），并定义了一年的季节性。然而在中国，太阴历等的影响不可小觑，这似乎提出了“什么是季节性？”、“什么是季节调整？”这一根本性的问题。

如果本文能对经济统计中季节性处理的未来发展有所贡献，我们将深感荣幸。

References

- [1] 高冈慎 (2025), “中国的贸易统计与季节调整”，未定稿.
- [2] Kunitomo, N. and S. Sato (2025), *The SIML Filtering Method for Noisy Non-stationary Economic Time Series*, Springer. forthcoming
- [3] 国友直人・櫻井智章・佐藤整尚 (2022), “经济时间序列的状态估计与宏观指标,” 统计研究汇报, 79, 1-20, 统计局.

< 附录：赵-虚拟变量 >

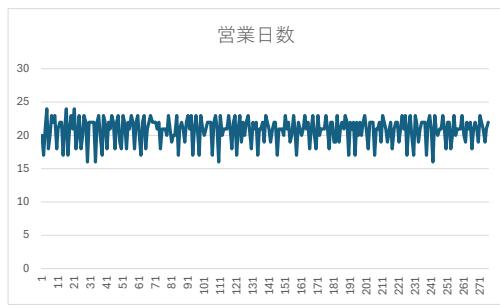
本附录刊载了基于“关于中国的日历与经济活动：赵-虚拟变量”（赵·国友）正文中所说明的方法构建的中国节假日效应变量（这里通称为赵-虚拟变量）。

開始日	終了日	月日数	土・日数	公式休日数	営業日日数	月曜日の数	火曜日の数	水曜日の数	木曜日の数	金曜日の数	土曜日の数	日曜日の数
2002/1/1	2002/1/31	31	8	11	20	4	5	5	5	4	4	4
2002/2/1	2002/2/28	28	8	11	17	4	4	4	4	4	4	4
2002/3/1	2002/3/31	31	10	10	21	4	4	4	4	5	5	5
2002/4/1	2002/4/30	30	8	6	24	5	5	4	4	4	4	4
2002/5/1	2002/5/31	31	8	13	18	4	4	5	5	5	4	4
2002/6/1	2002/6/30	30	10	10	20	4	4	4	4	4	5	5
2002/7/1	2002/7/31	31	8	8	23	5	5	5	4	4	4	4
2002/8/1	2002/8/31	31	9	9	22	4	4	4	5	5	5	4
2002/9/1	2002/9/30	30	9	7	23	5	4	4	4	4	4	5
2002/10/1	2002/10/31	31	8	13	18	4	5	5	5	4	4	4
2002/11/1	2002/11/30	30	9	9	21	4	4	4	4	5	5	4
2002/12/1	2002/12/31	31	9	9	22	5	5	4	4	4	4	5
2003/1/1	2003/1/31	31	8	9	22	4	4	5	5	5	4	4
2003/2/1	2003/2/28	28	8	11	17	4	4	4	4	4	4	4
2003/3/1	2003/3/31	31	10	10	21	5	4	4	4	4	5	5
2003/4/1	2003/4/30	30	8	6	24	4	5	5	4	4	4	4
2003/5/1	2003/5/31	31	9	14	17	4	4	4	5	5	5	4
2003/6/1	2003/6/30	30	9	9	21	5	4	4	4	4	4	5
2003/7/1	2003/7/31	31	8	8	23	4	5	5	5	4	4	4
2003/8/1	2003/8/31	31	10	10	21	4	4	4	4	5	5	5
2003/9/1	2003/9/30	30	8	6	24	5	5	4	4	4	4	4
2003/10/1	2003/10/31	31	8	13	18	4	4	5	5	5	4	4
2003/11/1	2003/11/30	30	10	10	20	4	4	4	4	4	5	5
2003/12/1	2003/12/31	31	8	8	23	5	5	5	4	4	4	4
2004/1/1	2004/1/31	31	9	13	18	4	4	4	5	5	5	4
2004/2/1	2004/2/29	29	9	9	20	4	4	4	4	4	4	5
2004/3/1	2004/3/31	31	8	8	23	5	5	5	4	4	4	4
2004/4/1	2004/4/30	30	8	8	22	4	4	4	5	5	4	4
2004/5/1	2004/5/31	31	10	15	16	5	4	4	4	4	5	5
2004/6/1	2004/6/30	30	8	8	22	4	5	5	4	4	4	4
2004/7/1	2004/7/31	31	9	9	22	4	4	4	5	5	5	4
2004/8/1	2004/8/31	31	9	9	22	5	5	4	4	4	4	5
2004/9/1	2004/9/30	30	8	8	22	4	4	5	5	4	4	4
2004/10/1	2004/10/31	31	10	15	16	4	4	4	4	5	5	5
2004/11/1	2004/11/30	30	8	8	22	5	5	4	4	4	4	4
2004/12/1	2004/12/31	31	8	8	23	4	4	5	5	5	4	4
2005/1/1	2005/1/31	31	10	11	20	5	4	4	4	4	5	5
2005/2/1	2005/2/28	28	8	11	17	4	4	4	4	4	4	4
2005/3/1	2005/3/31	31	8	8	23	4	5	5	4	4	4	4
2005/4/1	2005/4/30	30	9	8	22	4	4	4	4	5	5	4
2005/5/1	2005/5/31	31	9	13	18	5	5	4	4	4	4	5
2005/6/1	2005/6/30	30	8	8	22	4	4	5	5	4	4	4
2005/7/1	2005/7/31	31	10	10	21	4	4	4	4	5	5	5
2005/8/1	2005/8/31	31	8	8	23	5	5	5	4	4	4	4
2005/9/1	2005/9/30	30	8	8	22	4	4	4	5	5	4	4
2005/10/1	2005/10/31	31	10	13	18	5	4	4	4	4	5	5
2005/11/1	2005/11/30	30	8	8	22	4	5	5	4	4	4	4
2005/12/1	2005/12/31	31	9	8	23	4	4	4	5	5	5	4
2006/1/1	2006/1/31	31	9	12	19	5	5	4	4	4	4	5
2006/2/1	2006/2/28	28	8	10	18	4	4	4	4	4	4	4
2006/3/1	2006/3/31	31	8	8	23	4	4	5	5	5	4	4
2006/4/1	2006/4/30	30	10	8	22	4	4	4	4	4	5	5
2006/5/1	2006/5/31	31	8	13	18	5	5	5	4	4	4	4
2006/6/1	2006/6/30	30	8	8	22	4	4	4	5	5	4	4
2006/7/1	2006/7/31	31	10	10	21	5	4	4	4	4	5	5
2006/8/1	2006/8/31	31	8	8	23	4	5	5	4	4	4	4
2006/9/1	2006/9/30	30	9	8	22	4	4	4	4	5	5	4
2006/10/1	2006/10/31	31	9	13	18	5	5	4	4	4	4	5
2006/11/1	2006/11/30	30	8	8	22	4	4	5	5	4	4	4
2006/12/1	2006/12/31	31	10	8	23	4	4	4	4	5	5	5
2007/1/1	2007/1/31	31	8	11	20	5	5	5	4	4	4	4
2007/2/1	2007/2/28	28	8	11	17	4	4	4	4	4	4	4
2007/3/1	2007/3/31	31	9	9	22	4	4	4	5	5	5	4
2007/4/1	2007/4/30	30	9	7	23	5	4	4	4	4	4	5
2007/5/1	2007/5/31	31	8	13	18	4	5	5	5	4	4	4
2007/6/1	2007/6/30	30	9	9	21	4	4	4	4	5	5	4
2007/7/1	2007/7/31	31	9	9	22	5	5	4	4	4	4	5
2007/8/1	2007/8/31	31	8	8	23	4	4	5	5	5	4	4
2007/9/1	2007/9/30	30	10	8	22	4	4	4	4	4	5	5
2007/10/1	2007/10/31	31	8	9	22	5	5	5	4	4	4	4

2007/11/1	2007/11/30	30	8	8	22	4	4	4	5	5	4	4
2007/12/1	2007/12/31	31	10	10	21	5	4	4	4	4	5	5
2008/1/1	2008/1/31	31	8	9	22	4	5	5	5	4	4	4
2008/2/1	2008/2/29	29	8	11	18	4	4	4	4	5	4	4
2008/3/1	2008/3/31	31	10	10	21	5	4	4	4	4	5	5
2008/4/1	2008/4/30	30	8	9	21	4	5	5	4	4	4	4
2008/5/1	2008/5/31	31	9	10	21	4	4	4	5	5	5	4
2008/6/1	2008/6/30	30	9	10	20	5	4	4	4	4	4	5
2008/7/1	2008/7/31	31	8	8	23	4	5	5	5	4	4	4
2008/8/1	2008/8/31	31	10	10	21	4	4	4	4	5	5	5
2008/9/1	2008/9/30	30	8	11	19	5	5	4	4	4	4	4
2008/10/1	2008/10/31	31	8	11	20	4	4	5	5	5	4	4
2008/11/1	2008/11/30	30	10	10	20	4	4	4	4	4	5	5
2008/12/1	2008/12/31	31	8	8	23	5	5	5	4	4	4	4
2009/1/1	2009/1/31	31	9	14	17	4	4	4	5	5	5	4
2009/2/1	2009/2/28	28	8	7	21	4	4	4	4	4	4	4
2009/3/1	2009/3/31	31	9	9	22	5	5	4	4	4	4	5
2009/4/1	2009/4/30	30	8	9	21	4	4	5	5	4	4	4
2009/5/1	2009/5/31	31	10	12	19	4	4	4	4	5	5	5
2009/6/1	2009/6/30	30	8	8	22	5	5	4	4	4	4	4
2009/7/1	2009/7/31	31	8	8	23	4	4	5	5	5	4	4
2009/8/1	2009/8/31	31	10	10	21	5	4	4	4	4	5	5
2009/9/1	2009/9/30	30	8	7	23	4	5	5	4	4	4	4
2009/10/1	2009/10/31	31	9	14	17	4	4	4	5	5	5	4
2009/11/1	2009/11/30	30	9	9	21	5	4	4	4	4	4	5
2009/12/1	2009/12/31	31	8	8	23	4	5	5	5	4	4	4
2010/1/1	2010/1/31	31	10	11	20	4	4	4	4	5	5	5
2010/2/1	2010/2/28	28	8	11	17	4	4	4	4	4	4	4
2010/3/1	2010/3/31	31	8	8	23	5	5	5	4	4	4	4
2010/4/1	2010/4/30	30	8	9	21	4	4	4	5	5	4	4
2010/5/1	2010/5/31	31	10	11	20	5	4	4	4	4	5	5
2010/6/1	2010/6/30	30	8	9	21	4	5	5	4	4	4	4
2010/7/1	2010/7/31	31	9	9	22	4	4	4	5	5	5	4
2010/8/1	2010/8/31	31	9	9	22	5	5	4	4	4	4	5
2010/9/1	2010/9/30	30	8	8	22	4	4	5	5	4	4	4
2010/10/1	2010/10/31	31	10	14	17	4	4	4	4	5	5	5
2010/11/1	2010/11/30	30	8	8	22	5	5	4	4	4	4	4
2010/12/1	2010/12/31	31	8	8	23	4	4	5	5	5	4	4
2011/1/1	2011/1/31	31	10	10	21	5	4	4	4	4	5	5
2011/2/1	2011/2/28	28	8	12	16	4	4	4	4	4	4	4
2011/3/1	2011/3/31	31	8	8	23	4	5	5	5	4	4	4
2011/4/1	2011/4/30	30	9	10	20	4	4	4	4	5	5	4
2011/5/1	2011/5/31	31	9	10	21	5	5	4	4	4	4	5
2011/6/1	2011/6/30	30	8	9	21	4	4	5	5	4	4	4
2011/7/1	2011/7/31	31	10	10	21	4	4	4	4	5	5	5
2011/8/1	2011/8/31	31	8	8	23	5	5	5	4	4	4	4
2011/9/1	2011/9/30	30	8	9	21	4	4	4	5	5	4	4
2011/10/1	2011/10/31	31	10	13	18	5	4	4	4	4	5	5
2011/11/1	2011/11/30	30	8	8	22	4	5	5	4	4	4	4
2011/12/1	2011/12/31	31	9	8	23	4	4	4	5	5	5	4
2012/1/1	2012/1/31	31	9	14	17	5	5	4	4	4	4	5
2012/2/1	2012/2/29	29	8	8	21	4	4	5	4	4	4	4
2012/3/1	2012/3/31	31	9	8	23	4	4	4	5	5	5	4
2012/4/1	2012/4/30	30	9	11	19	5	4	4	4	4	4	5
2012/5/1	2012/5/31	31	8	9	22	4	5	5	4	4	4	4
2012/6/1	2012/6/30	30	9	10	20	4	4	4	4	5	5	4
2012/7/1	2012/7/31	31	9	9	22	5	5	4	4	4	4	5
2012/8/1	2012/8/31	31	8	8	23	4	4	5	5	4	4	4
2012/9/1	2012/9/30	30	10	9	21	4	4	4	4	4	5	5
2012/10/1	2012/10/31	31	8	13	18	5	5	5	4	4	4	4
2012/11/1	2012/11/30	30	8	8	22	4	4	4	5	5	4	4
2012/12/1	2012/12/31	31	10	10	21	5	4	4	4	4	5	5
2013/1/1	2013/1/31	31	8	9	22	4	5	5	4	4	4	4
2013/2/1	2013/2/28	28	8	11	17	4	4	4	4	4	4	4
2013/3/1	2013/3/31	31	10	10	21	4	4	4	4	5	5	5
2013/4/1	2013/4/30	30	8	9	21	5	5	4	4	4	4	4
2013/5/1	2013/5/31	31	8	9	22	4	4	5	5	4	4	4
2013/6/1	2013/6/30	30	10	11	19	4	4	4	4	4	5	5
2013/7/1	2013/7/31	31	8	8	23	5	5	5	4	4	4	4
2013/8/1	2013/8/31	31	9	9	22	4	4	4	5	5	5	4
2013/9/1	2013/9/30	30	9	9	21	5	4	4	4	4	4	5

2013/10/1	2013/10/31	31	8	12	19	4	5	5	5	4	4	4
2013/11/1	2013/11/30	30	9	9	21	4	4	4	4	5	5	4
2013/12/1	2013/12/31	31	9	9	22	5	5	4	4	4	4	5
2014/1/1	2014/1/31	31	8	9	22	4	4	5	5	5	4	4
2014/2/1	2014/2/28	28	8	11	17	4	4	4	4	4	4	4
2014/3/1	2014/3/31	31	10	10	21	5	4	4	4	4	5	5
2014/4/1	2014/4/30	30	8	9	21	4	5	5	4	4	4	4
2014/5/1	2014/5/31	31	9	10	21	4	4	4	5	5	5	4
2014/6/1	2014/6/30	30	9	10	20	5	4	4	4	4	4	5
2014/7/1	2014/7/31	31	8	8	23	4	5	5	5	4	4	4
2014/8/1	2014/8/31	31	10	10	21	4	4	4	4	5	5	5
2014/9/1	2014/9/30	30	8	8	22	5	5	4	4	4	4	4
2014/10/1	2014/10/31	31	8	12	19	4	4	5	5	5	4	4
2014/11/1	2014/11/30	30	10	10	20	4	4	4	4	5	5	5
2014/12/1	2014/12/31	31	8	8	23	5	5	5	4	4	4	4
2015/1/1	2015/1/31	31	9	10	21	4	4	4	5	5	5	4
2015/2/1	2015/2/28	28	8	11	17	4	4	4	4	4	4	4
2015/3/1	2015/3/31	31	9	9	22	5	5	4	4	4	4	5
2015/4/1	2015/4/30	30	8	9	21	4	4	5	5	4	4	4
2015/5/1	2015/5/31	31	10	11	20	4	4	4	4	5	5	5
2015/6/1	2015/6/30	30	8	9	21	5	5	4	4	4	4	4
2015/7/1	2015/7/31	31	8	8	23	4	4	5	5	4	4	4
2015/8/1	2015/8/31	31	10	10	21	5	4	4	4	4	5	5
2015/9/1	2015/9/30	30	8	8	22	4	5	5	4	4	4	4
2015/10/1	2015/10/31	31	9	13	18	4	4	4	5	5	5	4
2015/11/1	2015/11/30	30	9	9	21	5	4	4	4	4	4	5
2015/12/1	2015/12/31	31	8	8	23	4	5	5	5	4	4	4
2016/1/1	2016/1/31	31	10	11	20	4	4	4	4	5	5	5
2016/2/1	2016/2/29	29	8	11	18	5	4	4	4	4	4	4
2016/3/1	2016/3/31	31	8	8	23	4	5	5	5	4	4	4
2016/4/1	2016/4/30	30	9	10	20	4	4	4	4	5	5	4
2016/5/1	2016/5/31	31	9	10	21	5	5	4	4	4	4	5
2016/6/1	2016/6/30	30	8	9	21	4	4	5	5	4	4	4
2016/7/1	2016/7/31	31	10	10	21	4	4	4	4	5	5	5
2016/8/1	2016/8/31	31	8	8	23	5	5	5	4	4	4	4
2016/9/1	2016/9/30	30	8	9	21	4	4	4	5	5	4	4
2016/10/1	2016/10/31	31	10	13	18	5	4	4	4	4	5	5
2016/11/1	2016/11/30	30	8	8	22	4	5	5	4	4	4	4
2016/12/1	2016/12/31	31	9	9	22	4	4	4	5	5	5	4
2017/1/1	2017/1/31	31	9	12	19	5	5	4	4	4	4	5
2017/2/1	2017/2/28	28	8	9	19	4	4	4	4	4	4	4
2017/3/1	2017/3/31	31	8	8	23	4	4	5	5	4	4	4
2017/4/1	2017/4/30	30	10	11	19	4	4	4	4	4	5	5
2017/5/1	2017/5/31	31	8	10	21	5	5	5	4	4	4	4
2017/6/1	2017/6/30	30	8	8	22	4	4	4	5	5	4	4
2017/7/1	2017/7/31	31	10	10	21	5	4	4	4	4	5	5
2017/8/1	2017/8/31	31	8	8	23	4	5	5	5	4	4	4
2017/9/1	2017/9/30	30	9	8	22	4	4	4	4	5	5	4
2017/10/1	2017/10/31	31	9	14	17	5	5	4	4	4	4	5
2017/11/1	2017/11/30	30	8	8	22	4	4	5	5	4	4	4
2017/12/1	2017/12/31	31	10	10	21	4	4	4	4	5	5	5
2018/1/1	2018/1/31	31	8	9	22	5	5	5	4	4	4	4
2018/2/1	2018/2/28	28	8	11	17	4	4	4	4	4	4	4
2018/3/1	2018/3/31	31	9	9	22	4	4	4	5	5	5	4
2018/4/1	2018/4/30	30	9	10	20	5	4	4	4	4	4	5
2018/5/1	2018/5/31	31	8	9	22	4	5	5	5	4	4	4
2018/6/1	2018/6/30	30	9	10	20	4	4	4	4	5	5	4
2018/7/1	2018/7/31	31	9	9	22	5	5	4	4	4	4	5
2018/8/1	2018/8/31	31	8	8	23	4	4	5	5	5	4	4
2018/9/1	2018/9/30	30	10	9	21	4	4	4	4	4	5	5
2018/10/1	2018/10/31	31	8	13	18	5	5	5	4	4	4	4
2018/11/1	2018/11/30	30	8	8	22	4	4	4	5	5	4	4
2018/12/1	2018/12/31	31	10	9	22	5	4	4	4	4	5	5
2019/1/1	2019/1/31	31	8	9	22	4	5	5	4	4	4	4
2019/2/1	2019/2/28	28	8	11	17	4	4	4	4	4	4	4
2019/3/1	2019/3/31	31	10	10	21	4	4	4	4	5	5	5
2019/4/1	2019/4/30	30	8	9	21	5	5	4	4	4	4	4
2019/5/1	2019/5/31	31	8	9	22	4	4	5	5	5	4	4
2019/6/1	2019/6/30	30	10	11	19	4	4	4	4	4	5	5
2019/7/1	2019/7/31	31	8	8	23	5	5	5	4	4	4	4
2019/8/1	2019/8/31	31	9	9	22	4	4	4	5	5	5	4

2019/9/1	2019/9/30	30	9	9	21	5	4	4	4	4	4	5
2019/10/1	2019/10/31	31	8	12	19	4	5	5	5	4	4	4
2019/11/1	2019/11/30	30	9	9	21	4	4	4	4	5	5	4
2019/12/1	2019/12/31	31	9	9	22	5	5	4	4	4	4	5
2020/1/1	2020/1/31	31	8	13	18	4	4	5	5	5	4	4
2020/2/1	2020/2/29	29	9	8	21	4	4	4	4	4	5	4
2020/3/1	2020/3/31	31	9	9	22	5	5	4	4	4	4	5
2020/4/1	2020/4/30	30	8	8	22	4	4	5	5	4	4	4
2020/5/1	2020/5/31	31	10	12	19	4	4	4	4	5	5	5
2020/6/1	2020/6/30	30	8	9	21	5	5	4	4	4	4	4
2020/7/1	2020/7/31	31	8	8	23	4	4	5	5	5	4	4
2020/8/1	2020/8/31	31	10	10	21	5	4	4	4	4	5	5
2020/9/1	2020/9/30	30	8	7	23	4	5	5	4	4	4	4
2020/10/1	2020/10/31	31	9	14	17	4	4	4	5	5	5	4
2020/11/1	2020/11/30	30	9	9	21	5	4	4	4	4	4	5
2020/12/1	2020/12/31	31	8	8	23	4	5	5	5	4	4	4
2021/1/1	2021/1/31	31	10	11	20	4	4	4	4	5	5	5
2021/2/1	2021/2/28	28	8	11	17	4	4	4	4	4	4	4
2021/3/1	2021/3/31	31	8	8	23	5	5	5	4	4	4	4
2021/4/1	2021/4/30	30	8	8	22	4	4	4	5	5	4	4
2021/5/1	2021/5/31	31	10	12	19	5	4	4	4	4	5	5
2021/6/1	2021/6/30	30	8	9	21	4	5	5	4	4	4	4
2021/7/1	2021/7/31	31	9	9	22	4	4	4	5	5	5	4
2021/8/1	2021/8/31	31	9	9	22	5	5	4	4	4	4	5
2021/9/1	2021/9/30	30	8	8	22	4	4	5	5	4	4	4
2021/10/1	2021/10/31	31	10	14	17	4	4	4	4	5	5	5
2021/11/1	2021/11/30	30	8	8	22	5	5	4	4	4	4	4
2021/12/1	2021/12/31	31	8	8	23	4	4	5	5	5	4	4
2022/1/1	2022/1/31	31	10	10	21	5	4	4	4	4	5	5
2022/2/1	2022/2/28	28	8	12	16	4	4	4	4	4	4	4
2022/3/1	2022/3/31	31	8	8	23	4	5	5	4	4	4	4
2022/4/1	2022/4/30	30	9	9	21	4	4	4	4	5	5	4
2022/5/1	2022/5/31	31	9	11	20	5	5	4	4	4	4	5
2022/6/1	2022/6/30	30	8	9	21	4	4	5	5	4	4	4
2022/7/1	2022/7/31	31	10	10	21	4	4	4	4	5	5	5
2022/8/1	2022/8/31	31	8	8	23	5	5	5	4	4	4	4
2022/9/1	2022/9/30	30	8	9	21	4	4	4	5	5	4	4
2022/10/1	2022/10/31	31	10	13	18	5	4	4	4	4	5	5
2022/11/1	2022/11/30	30	8	8	22	4	5	5	4	4	4	4
2022/12/1	2022/12/31	31	9	9	22	4	4	4	5	5	5	4
2023/1/1	2023/1/31	31	9	13	18	5	5	4	4	4	4	5
2023/2/1	2023/2/28	28	8	8	20	4	4	4	4	4	4	4
2023/3/1	2023/3/31	31	8	8	23	4	4	5	5	5	4	4
2023/4/1	2023/4/30	30	10	10	20	4	4	4	4	4	5	5
2023/5/1	2023/5/31	31	8	10	21	5	5	5	4	4	4	4
2023/6/1	2023/6/30	30	8	9	21	4	4	4	5	5	4	4
2023/7/1	2023/7/31	31	10	10	21	5	4	4	4	4	5	5
2023/8/1	2023/8/31	31	8	8	23	4	5	5	5	4	4	4
2023/9/1	2023/9/30	30	9	10	20	4	4	4	4	5	5	4
2023/10/1	2023/10/31	31	9	12	19	5	5	4	4	4	4	5
2023/11/1	2023/11/30	30	8	8	22	4	4	5	5	4	4	4
2023/12/1	2023/12/31	31	10	10	21	4	4	4	4	5	5	5
2024/1/1	2024/1/31	31	8	9	22	5	5	5	4	4	4	4
2024/2/1	2024/2/29	29	8	11	18	4	4	4	5	4	4	4
2024/3/1	2024/3/31	31	10	10	21	4	4	4	4	5	5	5
2024/4/1	2024/4/30	30	8	8	22	5	5	4	4	4	4	4
2024/5/1	2024/5/31	31	8	10	21	4	4	5	5	5	4	4
2024/6/1	2024/6/30	30	10	11	19	4	4	4	4	4	5	5
2024/7/1	2024/7/31	31	8	8	23	5	5	4	4	4	4	4
2024/8/1	2024/8/31	31	9	9	22	4	4	4	5	5	5	4
2024/9/1	2024/9/30	30	9	9	21	5	4	4	4	4	4	5
2024/10/1	2024/10/31	31	8	12	19	4	5	5	5	4	4	4
2024/11/1	2024/11/30	30	9	9	21	4	4	4	5	5	4	4
2024/12/1	2024/12/31	31	9	9	22	5	5	4	4	4	4	5



季节调整方法：理论与官方统计中的应用

以中国贸易统计的季节调整为例

高冈慎⁶

目录

1 引言	20
2 中国节日效应的虚拟变量	20
2.1 中国的法定节假日	20
2.2 虚拟变量的设置	21
3 基于 X-12-ARIMA 和 X-13ARIMA-SEATS 的季节调整	22
3.1 X-12-ARIMA 与 X-13ARIMA-SEATS	22
3.2 RegARIMA 模型	24
3.3 X-12-ARIMA 与 X-13ARIMA-SEATS 的处理流程	24
4 模型选择与季节调整的估计	25
4.1 数据	25
4.2 估计流程	26
4.3 结果图表说明	26
4.4 结果 1：中国对外贸易额	26
4.5 结果 2：日本的贸易统计（数量指数）	30
4.6 结果 3：日本的贸易统计（实际金额）	35
4.7 小结	57
5 修正历史检验（Revision History）	58
5.1 Revision History	58
5.2 估计结果	59
5.3 小结	60

⁶琉球大学国际地域创造学院

6 结论	60
6.1 结果总结	60
6.2 季节调整所涉及的问题	84
6.3 今后的研究方向	85

1 引言

近年来，随着与中国的经济联系日益紧密，日本的部分经济指标（如贸易、生产、入境旅游等）也开始受到中国经济统计特有的波动的影响，导致季节调整等统计处理出现问题。由于中国的经济指标受到按农历设定的节假日影响较大，呈现出特殊的波动模式，因此在使用如 X-12-ARIMA 等季节调整软件时，如果采用默认设置，可能无法进行充分的调整。

本文主要以贸易统计为对象，展示了通过设置春节相关的虚拟变量，能够在多大程度上调整中国经济统计中所特有的周期性波动的估计结果。

2 中国节日效应的虚拟变量

2.1 中国的法定节假日

中国的法定节假日是根据《中华人民共和国法律》及相关行政规定设定的公众假期，指的是大多数企业、机关、学校等停止办公的日子。法定节假日的具体安排通常由国务院在前一年的 12 月公布。但由于许多节假日是根据农历设定的，因此其时间和持续天数每年可能会有较大变动。主要的节假日如下所示。

- 元旦（公历 1 月 1 日）
- 春节（农历 1 月 1 日）
- 清明节（农历 4 月 4 日或 5 日）
- 劳动节（公历 5 月 1 日）
- 端午节（农历 5 月 5 日）
- 中秋节（农历 8 月 15 日）
- 国庆节（公历 10 月 1 日起）

此外，为了延长长假，中国通常会将前后的双休日调整为工作日，以便形成连续的假期。这种做法已相当普遍。例如，将周末设为上班日，进而将工作日调为休息日，从而安排出较为集中的假期。

其中影响最大的是春节。春节根据农历日期决定，每年一般在公历 1 月下旬至 2 月中旬之间变动。这种变动会对中国经济活动产生显著影响，也使得日本在 1 至 3 月期间的贸易统计与生产统计呈现出明显的月度波动。图 1 展示了近年来

中国节假日的安排情况。

年份不同导致的所跨月份变化的连休

	春节	劳动节	中秋节	国庆节
2018 年	2/15~21	4/29~5/1	9/22~24	10/1~7
19 年	2/4~10	5/1	9/13~15	10/1~7
20 年	1/24~30	5/1~5		10/1~8
21 年	2/11~17	5/1~5	9/19~21	10/1~7
22 年	1/31~2/6	4/30~5/4	9/10~12	10/1~7
23 年	1/21~27	4/29~5/3		9/29~10/6
24 年	2/10~17	5/1~5	9/15~17	10/1~7

图 1: 近年来中国的节假日情况

2.2 虚拟变量的设置

连休虚拟变量

关于每年时期不同的长期连休的影响, Bell 和 Hillmer (1983) 提出了使用虚拟变量的如下形式。

$$D_{y,m} = \frac{\tau_{y,m}}{\tau_y}$$

$D_{y,m}$ 表示对应于 y 年 m 月的虚拟变量, 其中, τ_y 表示 y 年特定连休的假日天数, $\tau_{y,m}$ 表示该连休假期中落在 m 月的天数。当假期跨月时, $D_{y,m}$ 是按假期天数进行分摊的变量, 用于表达假期对各月份的影响。

例如, 在表示春节这一假期对 1 月到 2 月之间的影响时, $D_{y,m}$ 会在 1 月和 2 月中取 0 到 1 之间的实数值, 而在其他月份中则取值为 0。

考虑前后影响的长期连休虚拟变量

另一方面, 长期连休不仅对假期期间的经济活动产生影响, 还常常对假期前后时期的经济活动水平造成显著影响。通常可观察到的模式是: 假期前经济活动大幅上升, 假期间活动水平下降, 假期结束后逐渐恢复。

Lin 和 Liu (2003) 考虑到这一点, 提出了将 $D_{y,m}$ 拆分为三部分的方法。该方法

可以具体表示为：

$$D_{y,m}^{(p)} = \frac{\tau_{y,m}^{(p)}}{\tau_y^{(p)}}, \quad p = 1, 2, 3$$

其中， $p = 1$ 表示假期前的影响， $p = 2$ 表示假期期间的影响， $p = 3$ 表示假期后的影响。 $\tau_y^{(p)}$ 表示 y 年假期第 p 阶段的总天数， $\tau_{y,m}^{(p)}$ 表示该阶段中落在 m 月的天数。

图 2 展示了这些变量的构成示意。

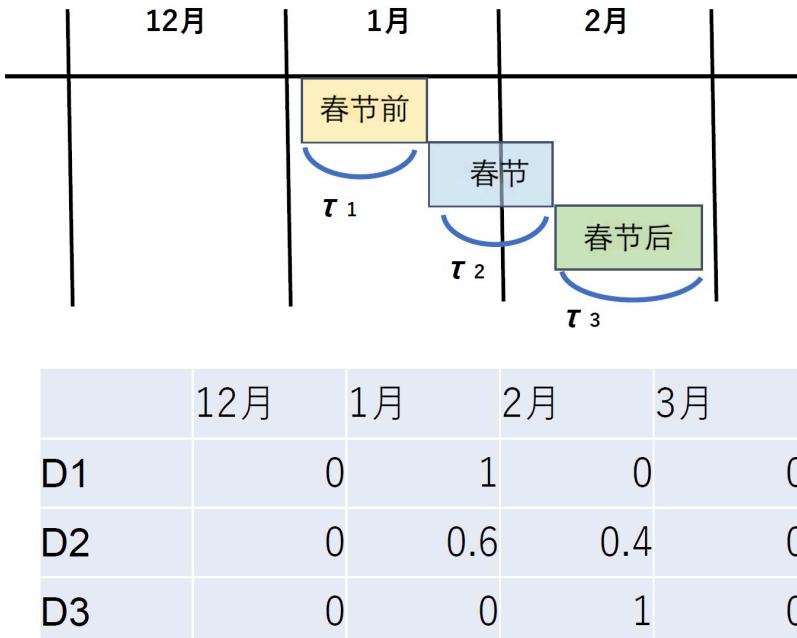


图 2: 对应长期连休的虚拟变量

可以看出，关于假期前后影响所及的天数 $\tau_y^{(1)}$ 和 $\tau_y^{(3)}$ 如何设定是一个问题，本文采用的方法是如后所述，通过数据使用 AIC 进行选择。另外，关于 $\tau_y^{(2)}$ ，则简单地采用假期的天数。

此外，春节对应的虚拟变量的具体形式如图 3 所示。

3 基于 X-12-ARIMA 和 X-13ARIMA-SEATS 的季节调整

3.1 X-12-ARIMA 与 X-13ARIMA-SEATS

目前，在包括日本在内的许多国家的官方统计中，用于季节调整的主要软件是由美国商务部普查局开发的 X-12-ARIMA 及其后继版本 X-13ARIMA-SEATS。

	A	B	C	D	E	F
1	Y	M	D1	D2	D3	
2	2002		1	0	0	0
3	2002		2	1	1	1
4	2002		3	0	0	0
5	2002		4	0	0	0
6	2002		5	0	0	0
7	2002		6	0	0	0
8	2002		7	0	0	0
9	2002		8	0	0	0
10	2002		9	0	0	0
11	2002		10	0	0	0
12	2002		11	0	0	0
13	2002		12	0	0	0
14	2003		1	0.857	0	0
15	2003		2	0.143	1	1
16	2003		3	0	0	0
17	2003		4	0	0	0
18	2003		5	0	0	0
19	2003		6	0	0	0
20	2003		7	0	0	0
21	2003		8	0	0	0
22	2003		9	0	0	0
23	2003		10	0	0	0
24	2003		11	0	0	0
25	2003		12	0	0	0
26	2004		1	1	1	0.429
27	2004		2	0	0	0.571
28	2004		3	0	0	0
29	2004		4	0	0	0

图 3: 春节对应的虚拟变量

这些软件的结构是通过反复应用相对简单的移动平均滤波器，时间序列分解为趋势成分、季节成分和不规则成分。但是，为了在数据端点附近也能适用对称的移动平均滤波器，在应用滤波器之前，作为预处理步骤，软件会通过时间序列模型对未来的数据进行预测并将其接续在端点之后。因此，在端点附近的季节调整值中会含有来自时间序列模型的预测值，这就使得在实际操作中选择符合数据特性的时间序列模型尤为重要。

3.2 RegARIMA 模型

用于预测的内部模型被称为“RegARIMA 模型”，可用下式表示：

$$\phi_p(B)\Phi_P(B^s)(1-B)^d(1-B^s)^D \left(y_t - \sum_{j=1}^r \beta_j z_{jt} \right) = \theta_q(B)\Theta_Q(B^s)u_t$$

这是在季节性 ARIMA 模型的基础上，扩展引入回归变量而成的形式，其结构通过回归变量来处理时序模型难以刻画的变动。RegARIMA 模型通过对数据及其解释变量进行适当次数的差分，可以归结为一个以误差项为平稳序列的回归模型，其中外生性的变动因素作为回归变量来处理。回归变量通常包括表示异常值、水平变动、星期效应等的虚拟变量。

3.3 X-12-ARIMA 与 X-13ARIMA-SEATS 的处理流程

在对原始序列进行处理时，首先会应用 RegARIMA 模型，将由回归变量表示的部分与由 ARIMA 模型表示的部分进行分离。其中，对由 ARIMA 模型表示的部分连接预测值，并通过 X-11 部分（基于移动平均的处理）将其分解为各成分。在默认设置下，回归部分会被整合到趋势成分或不规则成分中，最终的季节调整序列作为长期趋势与不规则变动之和被输出。

图 4 展示了在 RegARIMA 模型中不包含与包含春节效应时的处理流程。

根据图 4 所示的处理流程，如果存在本应由回归变量处理的外生性变动，却未在模型中纳入适当的回归变量，那么这些外生性变动的影响将渗透到 RegARIMA 模型的估计与预测过程、以及后续 X-11 分解中，并最终会被分散到趋势、季节变动与不规则变动三个成分中。

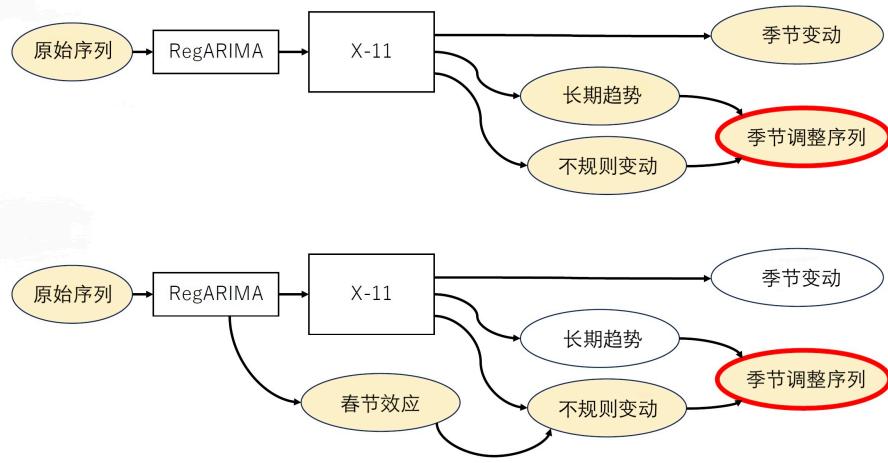


图 4: X-12-ARIMA 的处理流程

另一方面，若将春节效应作为回归变量纳入模型，则春节效应最终会被合并到不规则成分中。由于季节调整值被定义为趋势成分与不规则成分的总和，导致最终的季节调整值中也包含了春节效应，因此需注意这并不表示春节所带来的波动被剔除。

但是，若未使用适当的虚拟变量，将影响 RegARIMA 的估计与预测结果，从而可能导致季节调整序列的稳定性下降。

4 模型选择与季节调整的估计

4.1 数据

此次估计中使用的数据如下：

- 中国贸易统计：进出口实额（以百万元人民币计）
- 日本贸易统计：进出口数量指数（以 2020 年为基准 = 100）
- 日本贸易统计：按主要商品分类的进出口额（对中国，按月统计，单位：千日元）
 - <https://www.customs.go.jp/toukei/suui/html/time.htm>

关于中国的贸易统计，使用的是以“100 万元人民币为 1”的进出口实额。日本方面的统计使用的是来自财务省的贸易统计中的进出口数量指数和进出口实额。

4.2 估计流程

为了设定 RegARIMA 模型，有必要确定 ARIMA 部分的阶数，以及春节虚拟变量中所包含的 $\tau_y^{(1)}$ 和 $\tau_y^{(3)}$ 。在本次估计中，模型的选择是按照以下步骤进行的。

- (1) 将 ARIMA 部分固定为“航空公司模型”：即 (011)(011) 模型。
- (2) 假设假期前效应天数 $\tau_y^{(1)}$ 与假期后效应天数 $\tau_y^{(3)}$ 不随年份 y 变化，从若干候选值中选择使 AIC 最小的组合作为最终值。
- (3) 固定所选的 $\tau_y^{(1)}$ 和 $\tau_y^{(3)}$ ，再通过 AIC 选择 ARIMA 的阶数。

此外，在执行上述步骤时，采用了以下设定：

- 采用对数变换的乘法型季节调整。
- 春节虚拟变量的设定范围为 2002 年 1 月至 2024 年 12 月，分析中仅使用该范围内的数据。
- 同时使用内置变量 TD（星期效应虚拟变量）。
- 对于 $\tau_y^{(1)}$ 与 $\tau_y^{(3)}$ ，从集合 $\{0, 3, 6, 9, \dots, 30\}$ 中选择。

4.3 结果图表说明

首先，对接下来各小节中所示的图表进行说明。

从表 1 到表 4 是针对各数据执行模型选择结果的汇总表。此外，从图 5 到图 30 的各图，其具体内容如下：

- 图 A：就设置春节虚拟变量所进行的估计结果，对原始序列与所估计的春节效应进行比较。
- 图 B：就设置春节虚拟变量所进行的估计结果，对季节调整序列与从中剔除春节效应后的序列进行比较。
- 图 C：比较未设置与设置春节虚拟变量的估计所得的季节调整值。
- 图 D：比较未设置与设置春节虚拟变量的估计所得的趋势成分。
- 图 E：比较未设置与设置春节虚拟变量的估计所得的季节成分。
- 各图中纵向刻度线表示的是 1 月。

4.4 结果 1：中国对外贸易额

使用中国贸易统计的结果如表 1 以及图 5、图 6 所示。

表 1: 估计值: 中国进出口额

		輸出額 (100 万人民元)	輸入額 (100 万人民元)
未设置春节虚拟变量	AIC	3484.4	3336.0
设置春节虚拟变量	AIC	3401.0	3286.9
	春节前	(τ : 27) 0.156 [4.621]	(τ : 6) 0.019 [0.901]
	春节期间	0.270 [12.798]	-0.136 [-3.833]
	春节后	(τ : 9) -0.416 [-8.933]	(τ : 3) -0.035 [-1.047]

□) 虚拟变量一栏表示系数估计值及其 t 值 (中括号内为 t 值)。

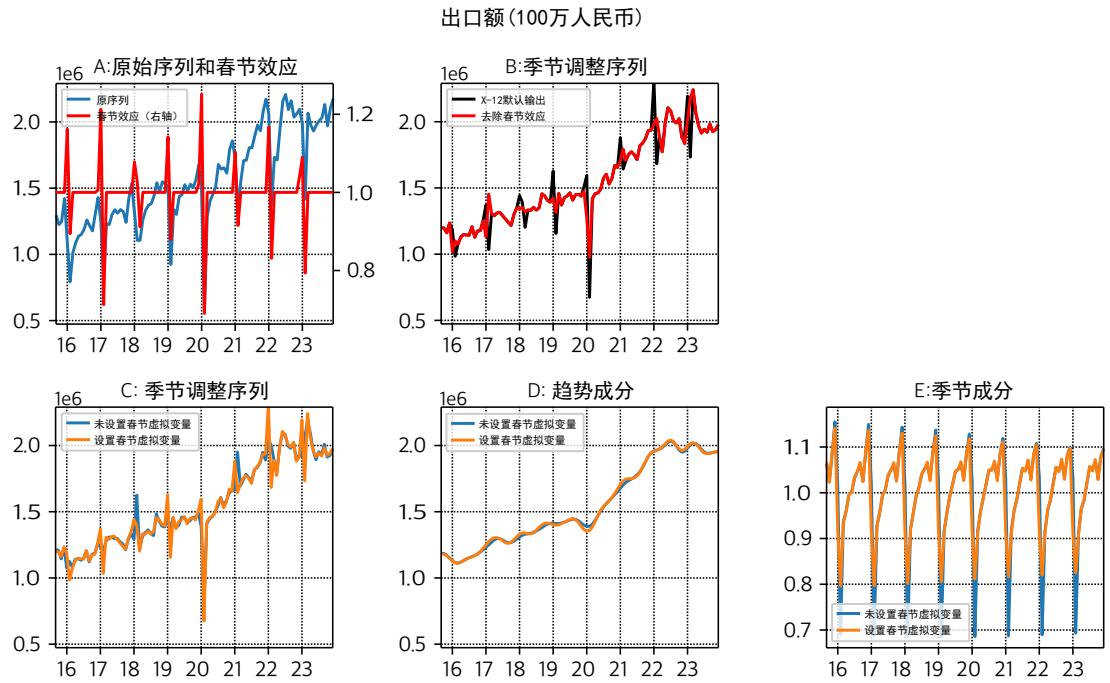


图 5: 出口额 (以百万元人民币计)

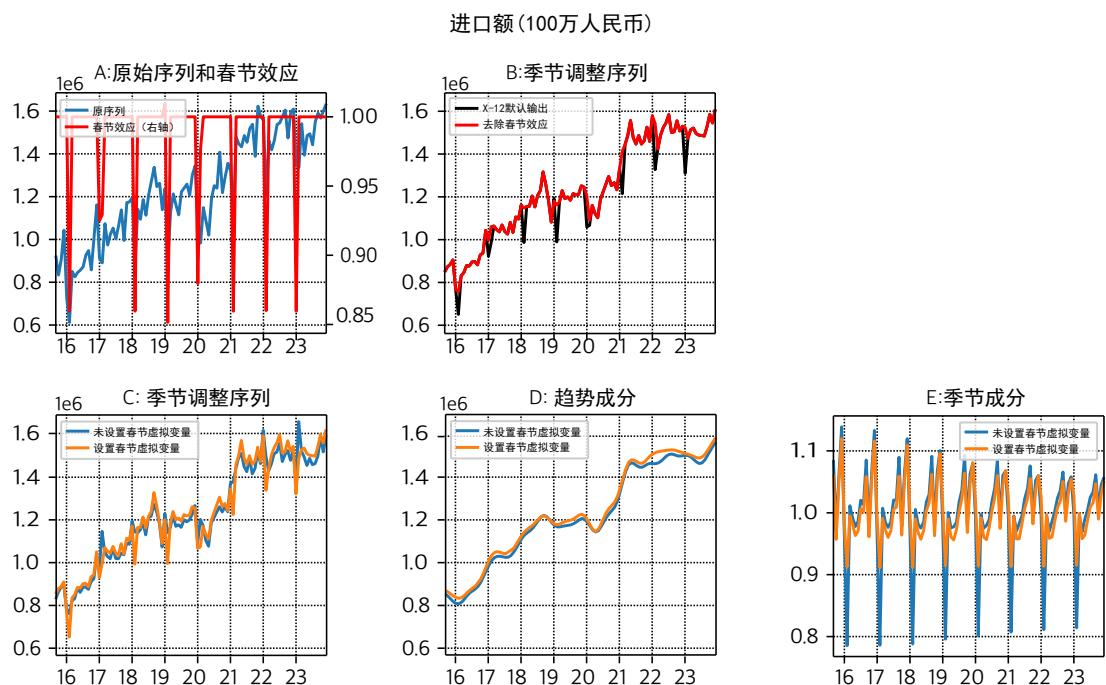


图 6: 进口额 (以百万元人民币计)

4.5 结果 2：日本的贸易统计（数量指数）

表 2: 估计值: 进出口数量指数 (对中国)

		出口数量指数 (中国)	进口数量指数 (中国)	出口数量指数 (世界)	进口数量指数 (世界)
未设置春节虚拟变量	AIC	753.1	810.6	648.6	617.9
设置春节虚拟变量	AIC	614.7	735.1	621.6	589.0
	春节前	(τ : 12) -0.177 [-10.344]	(τ : 27) 0.153 [3.528]	(τ : 18) -0.075 [-4.839]	(τ : 27) 0.041 [1.884]
	春节期间	-0.001 [-0.067]	0.268 [10.922]	-0.001 [-0.107]	0.066 [4.601]
	春节后	(τ : 27) 0.053 [2.479]	(τ : 15) -0.255 [-6.095]	(τ : 12) -0.022 [-1.102]	(τ : 15) -0.074 [-3.153]

□) 虚拟变量一栏表示系数估计值及其 t 值 (中括号内为 t 值)。

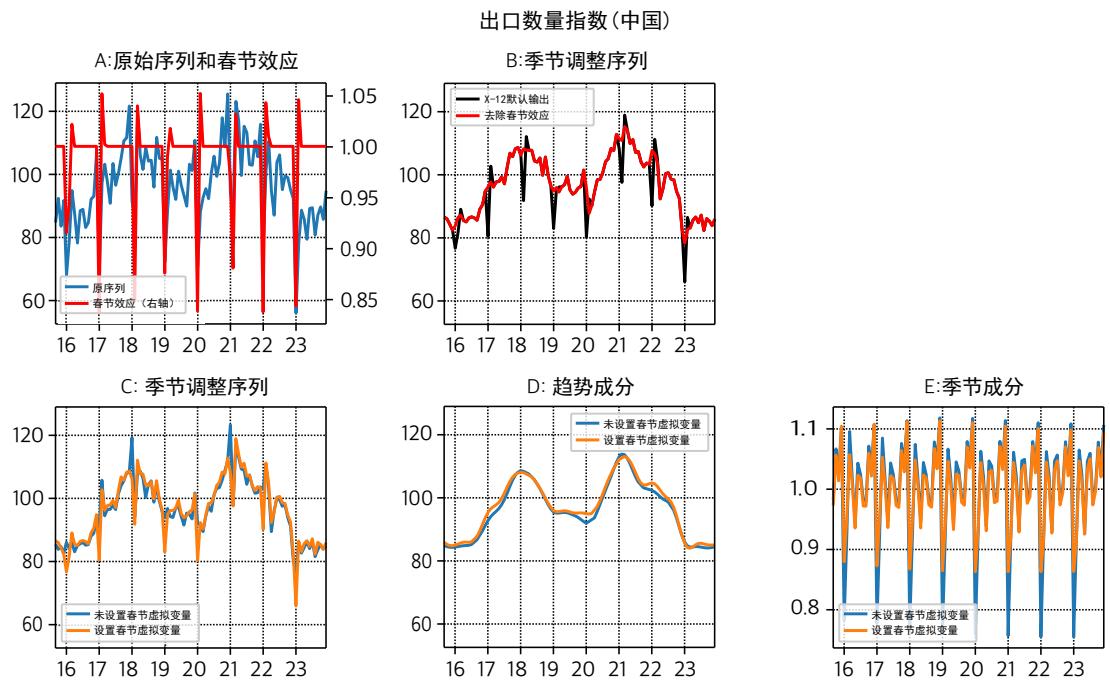


图 7: 出口数量指数 (中国)

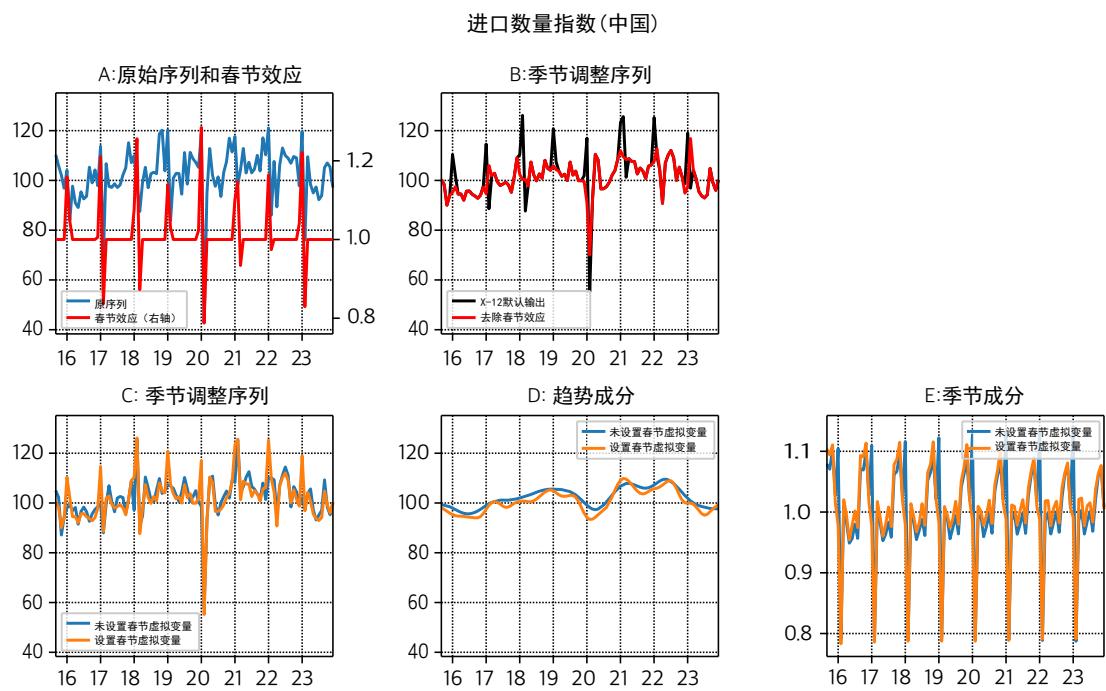


图 8: 进口数量指数 (中国)

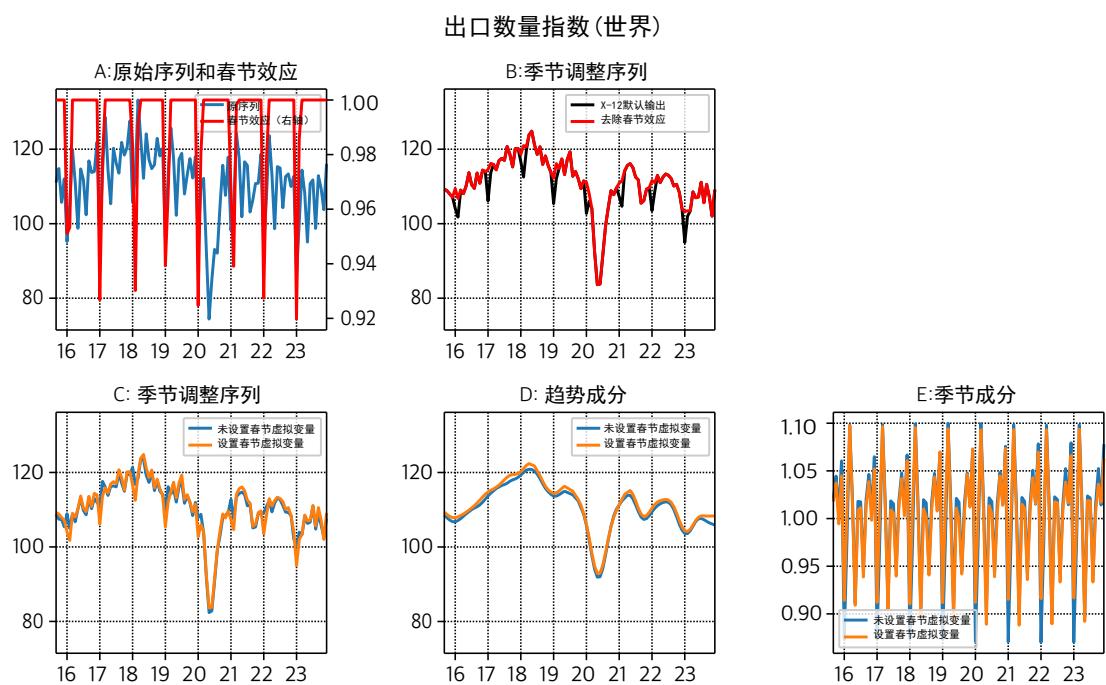


图 9: 出口数量指数 (世界)

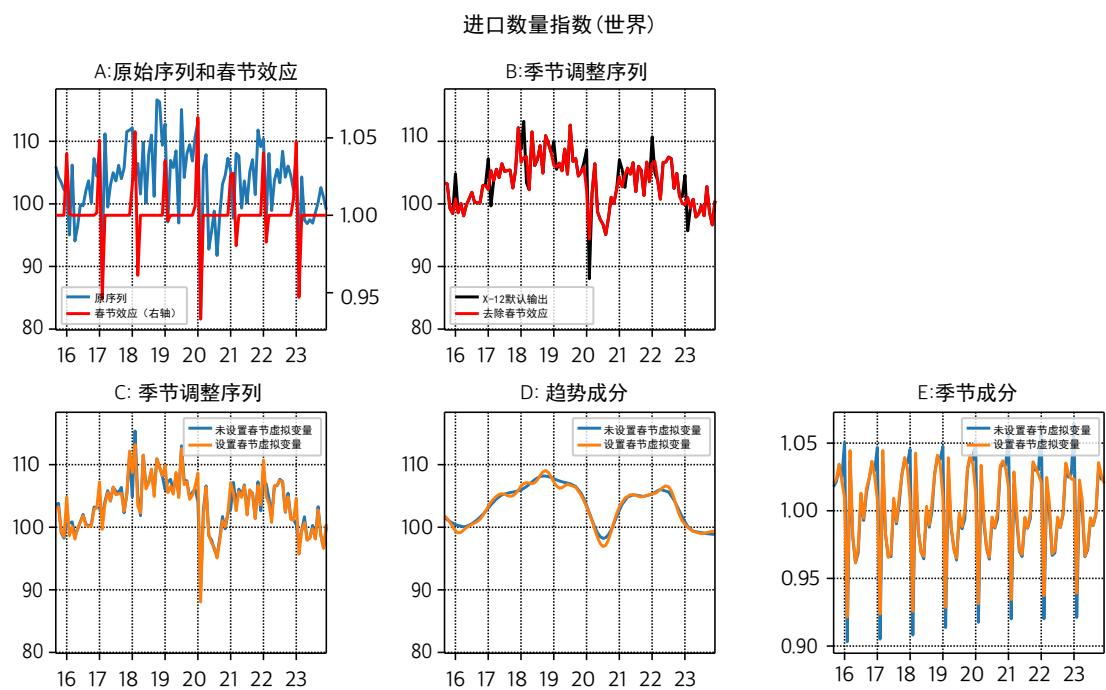


图 10: 进口数量指数 (世界)

4.6 结果 3：日本的贸易统计（实际金额）

使用日本贸易统计中的进出口实额（按品目分类）所得到的结果如表 3、表 4 以及图 11 至图 30 所示。

表 3: 估计值: 出口实额（对中国）

		总额(出口)	食品(出口)	原料(出口)	矿物燃料(出口)
未设置春节虚拟变量	AIC	9899.9	7805.1	8485.9	8510.3
设置春节虚拟变量	AIC	9750.0	7762.6	8455.5	8510.3
	春节前	(τ : 12) -0.168 [-14.247]	(τ : 15) -0.454 [-6.143]	(τ : 9) -0.187 [-6.291]	
	春节期间	0.017 [1.473]	0.074 [1.360]	0.128 [2.871]	
	春节后	(τ : 12) -0.015 [-0.809]	(τ : 21) -0.120 [-1.472]	(τ : 3) -0.040 [-0.910]	
		化学制品(出口)	按原料分类的制成品(出口)	一般机械(出口)	
未设置春节虚拟变量	AIC	9093.6	8949.1	9284.8	
设置春节虚拟变量	AIC	9020.4	8820.7	9181.5	
	春节前	(τ : 12) -0.183 [-8.530]	(τ : 9) -0.168 [-12.845]	(τ : 15) -0.247 [-11.102]	
	春节期间	0.040 [1.936]	0.048 [2.449]	0.044 [1.608]	
	春节后	(τ : 12) -0.054 [-1.625]	(τ : 3) -0.032 [-1.713]	(τ : 3) 0.012 [0.417]	
		电气设备(出口)	运输设备(出口)	其他(出口)	
未设置春节虚拟变量	AIC	9210.7	9007.2	9045.2	
设置春节虚拟变量	AIC	9110.3	8979.6	8961.0	
	春节前	(τ : 6) -0.107 [-8.145]	(τ : 24) -0.178 [-5.000]	(τ : 6) -0.165 [-8.160]	
	春节期间	-0.020 [-0.949]	-0.018 [-0.556]	0.037 [1.325]	
	春节后	(τ : 3) 0.041 [2.038]	(τ : 3) -0.097 [-2.367]	(τ : 6) -0.046 [-1.072]	

□) 虚拟变量一栏表示系数估计值及其 t 值（中括号内为 t 值）。

表 4: 估计值: 进口实额 (对中国)

		总额(进口)	食品(进口)	原料(进口)	矿物燃料(进口)
未设置春节虚拟变量	AIC	10188.0	8744.2	8095.1	8571.0
设置春节虚拟变量	AIC	10012.9	8484.0	7997.2	8568.1
	春节前	(τ : 27) 0.105 [3.791]	(τ : 21) 0.087 [3.989]	(τ : 27) 0.110 [2.746]	(τ : 15) 0.174 [1.315]
	春节期间	0.186 [12.526]	0.103 [7.853]	0.206 [8.291]	0.085 [0.580]
	春节后	(τ : 15) -0.243 [-8.503]	(τ : 27) -0.334 [-14.613]	(τ : 15) -0.311 [-6.744]	(τ : 6) 0.200 [0.846]
		化学制品(进口)	按原料分类的制成品(进口)	一般机械(进口)	
未设置春节虚拟变量	AIC	8813.5	9242.6	9281.5	
设置春节虚拟变量	AIC	8720.4	9067.4	9211.2	
	春节前	(τ : 27) 0.117 [3.374]	(τ : 27) 0.116 [3.224]	(τ : 27) 0.094 [3.035]	
	春节期间	0.223 [10.320]	0.233 [12.588]	0.123 [7.396]	
	春节后	(τ : 9) -0.268 [-5.554]	(τ : 15) -0.318 [-8.753]	(τ : 12) -0.173 [-4.836]	
		电器设备(进口)	运输设备(进口)	其他(进口)	
未设置春节虚拟变量	AIC	9566.0	8339.8	9642.8	
设置春节虚拟变量	AIC	9523.4	8254.7	9444.6	
	春节前	(τ : 27) 0.107 [2.739]	(τ : 27) 0.087 [2.320]	(τ : 27) 0.126 [3.680]	
	春节期间	0.105 [4.846]	0.192 [9.249]	0.274 [14.431]	
	春节后	(τ : 15) -0.136 [-3.256]	(τ : 15) -0.227 [-5.669]	(τ : 18) -0.332 [-10.031]	

□) 虚拟变量一栏表示系数估计值及其 t 值 (中括号内为 t 值)。

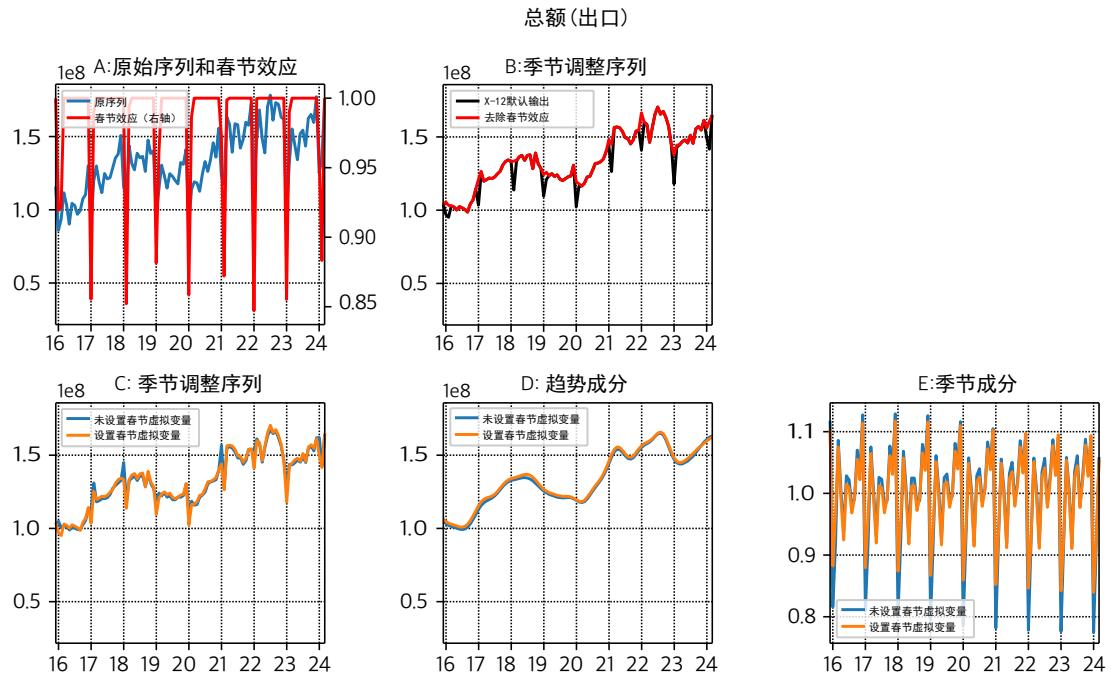


图 11: 总额 (出口)

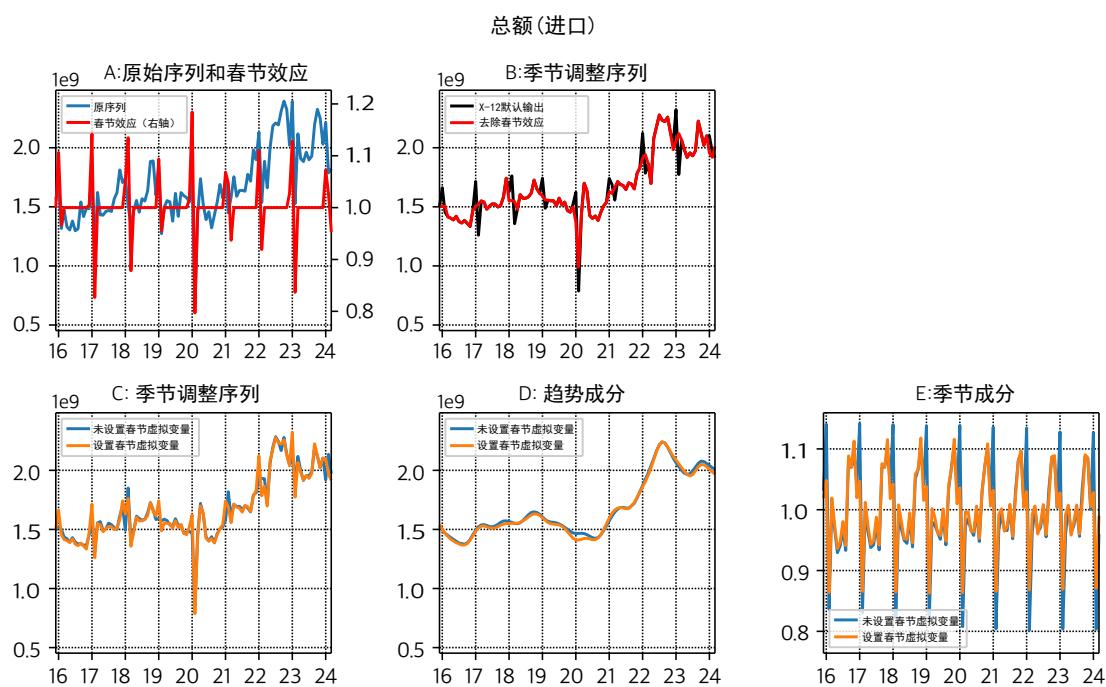


图 12: 总额 (进口)

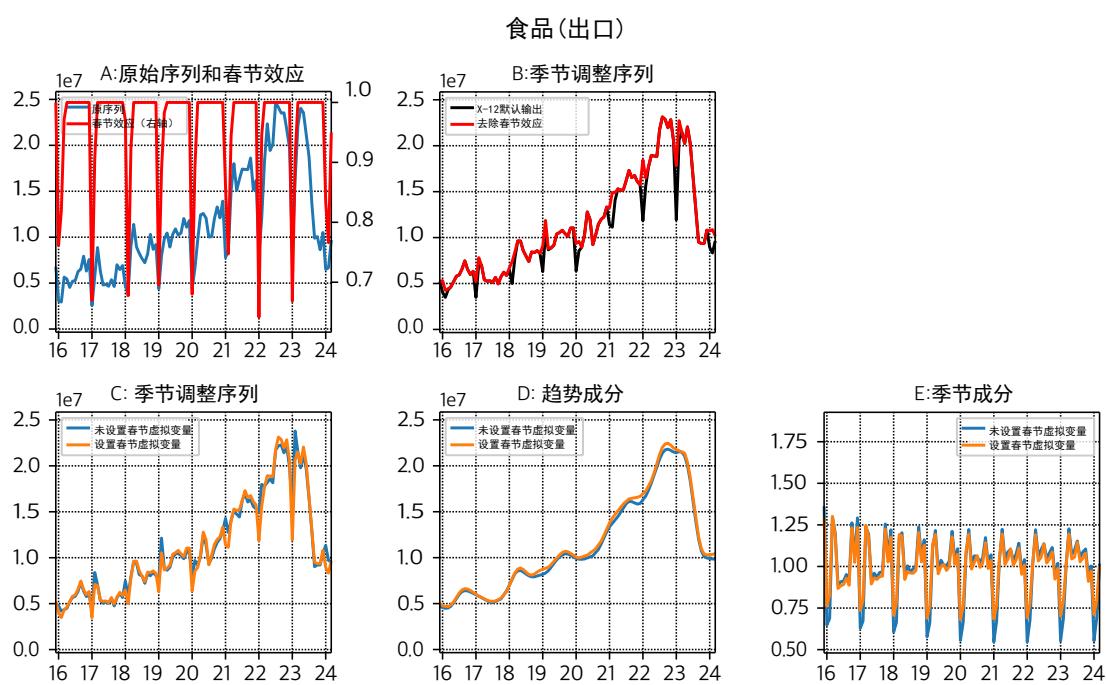


图 13: 食品(出口)

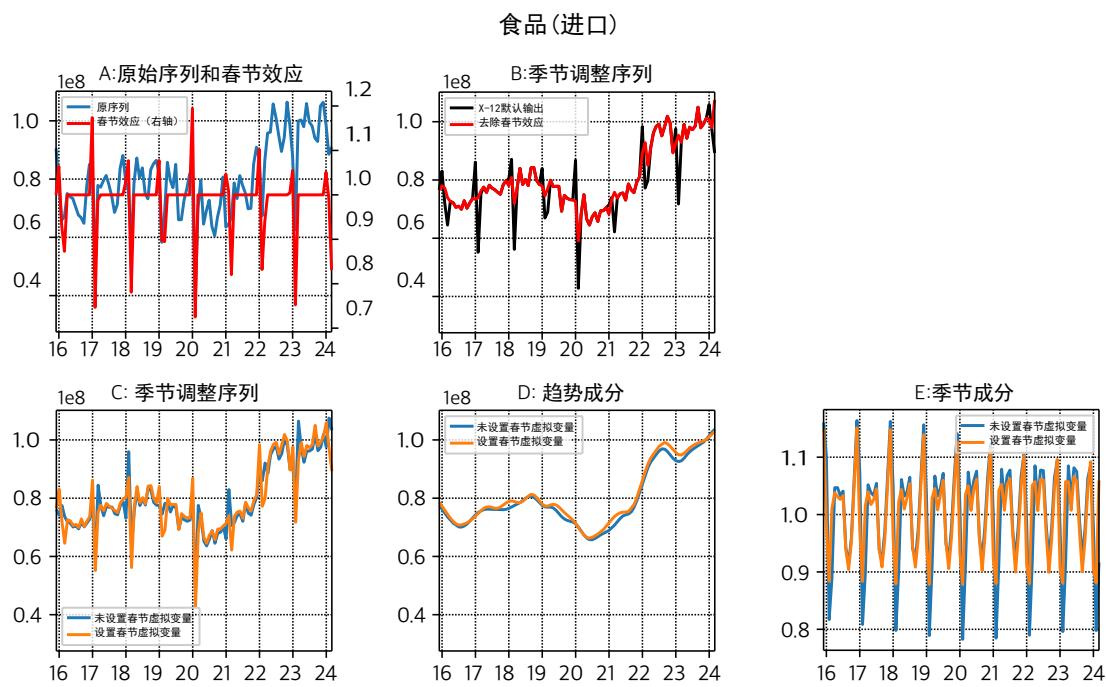


图 14: 食品(进口)

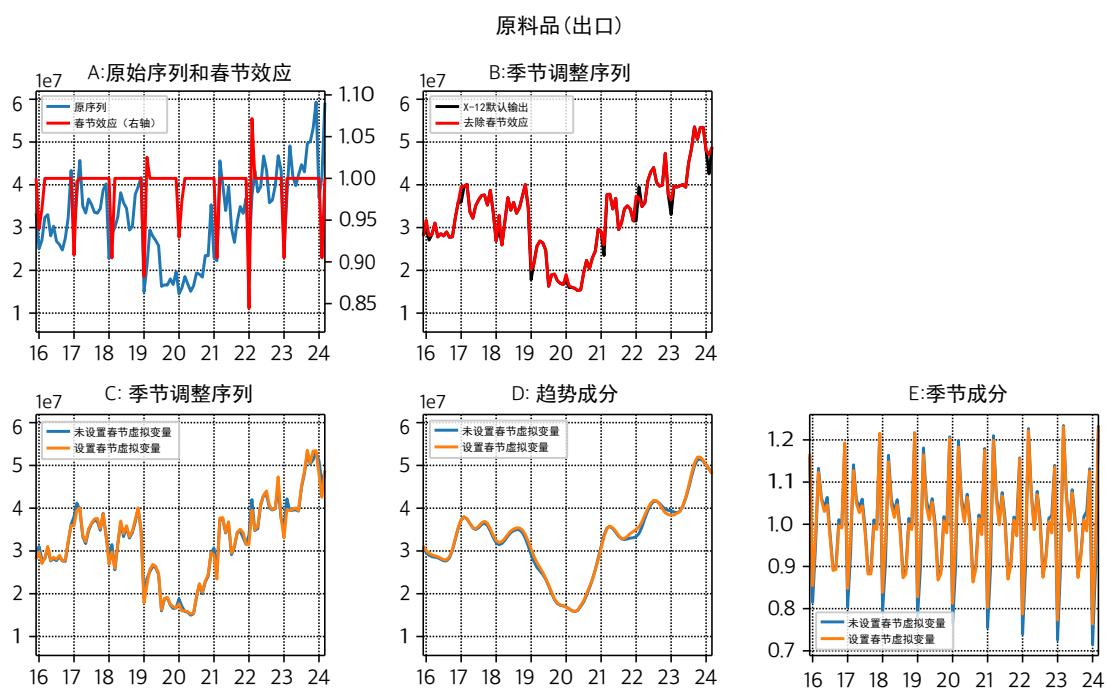


图 15: 原料 (出口)

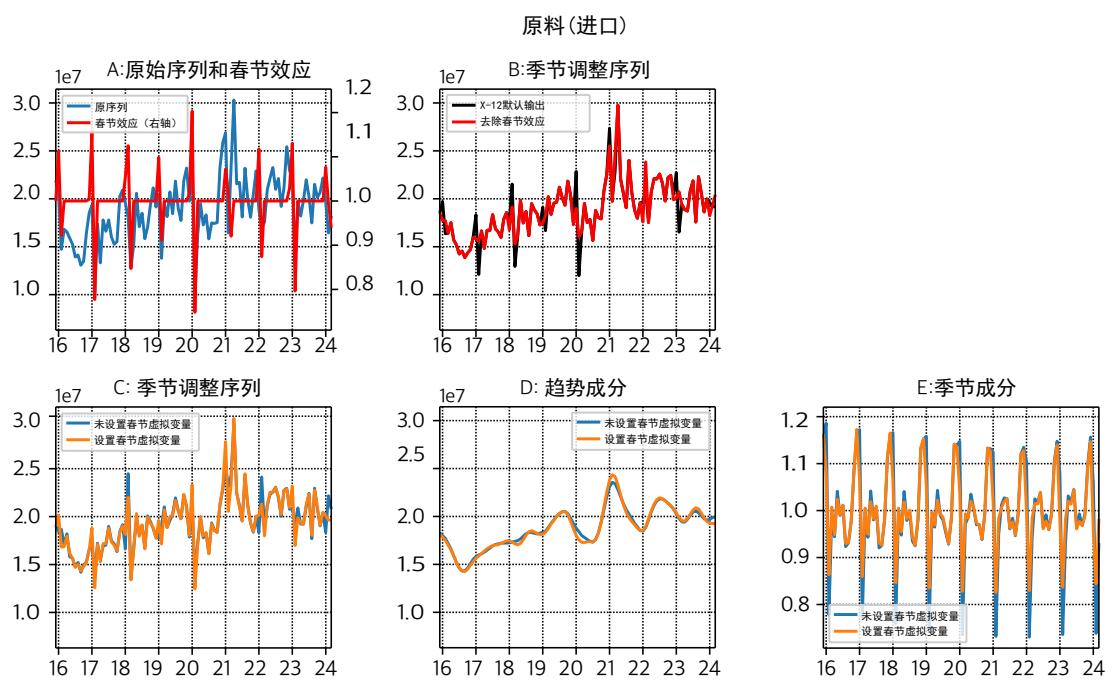


图 16: 原料(进口)

矿物燃料(出口)

A:原始序列和春节效应

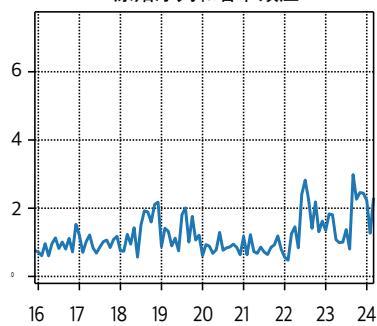


图 17: 矿物燃料 (出口)

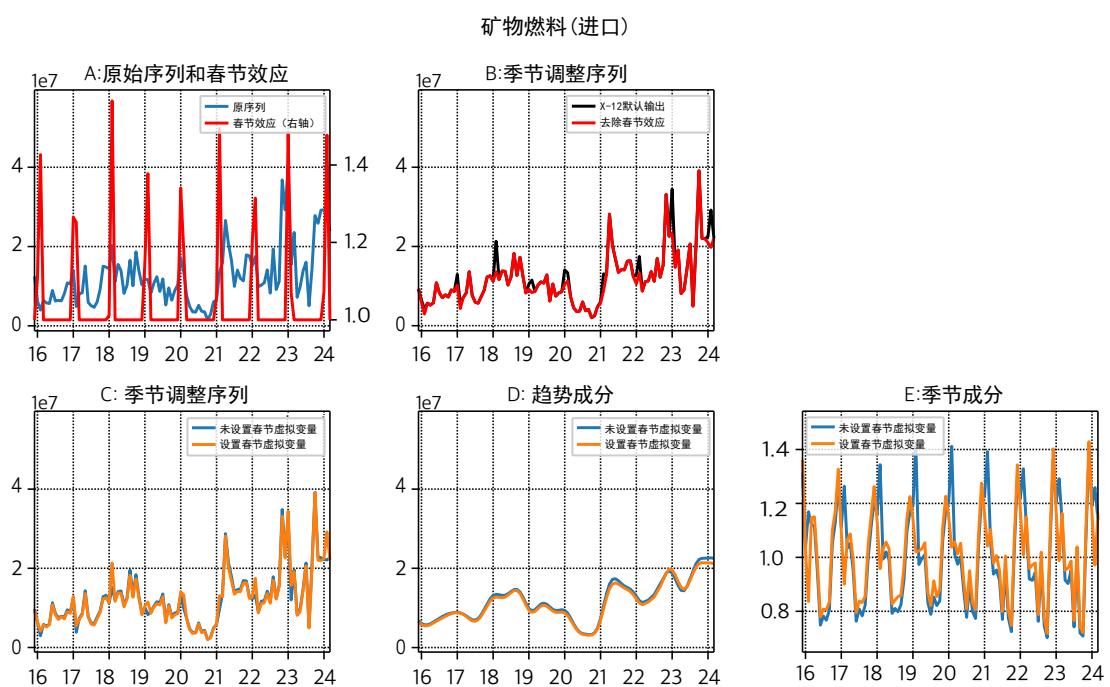


图 18: 矿物燃料(进口)

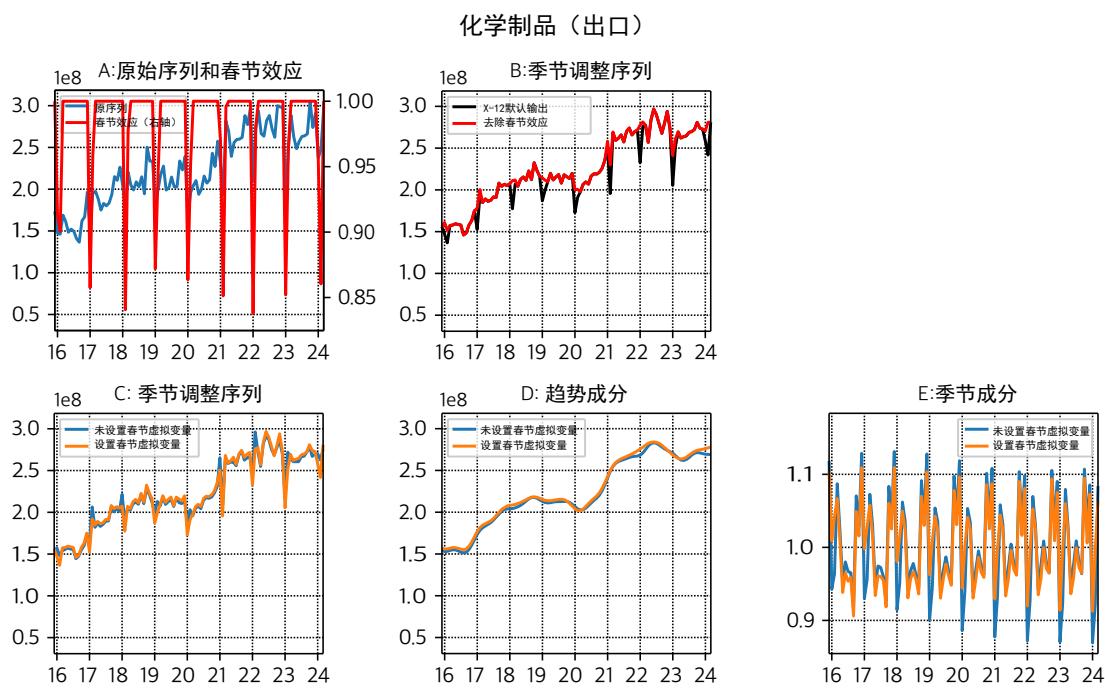


图 19: 化学制品(出口)

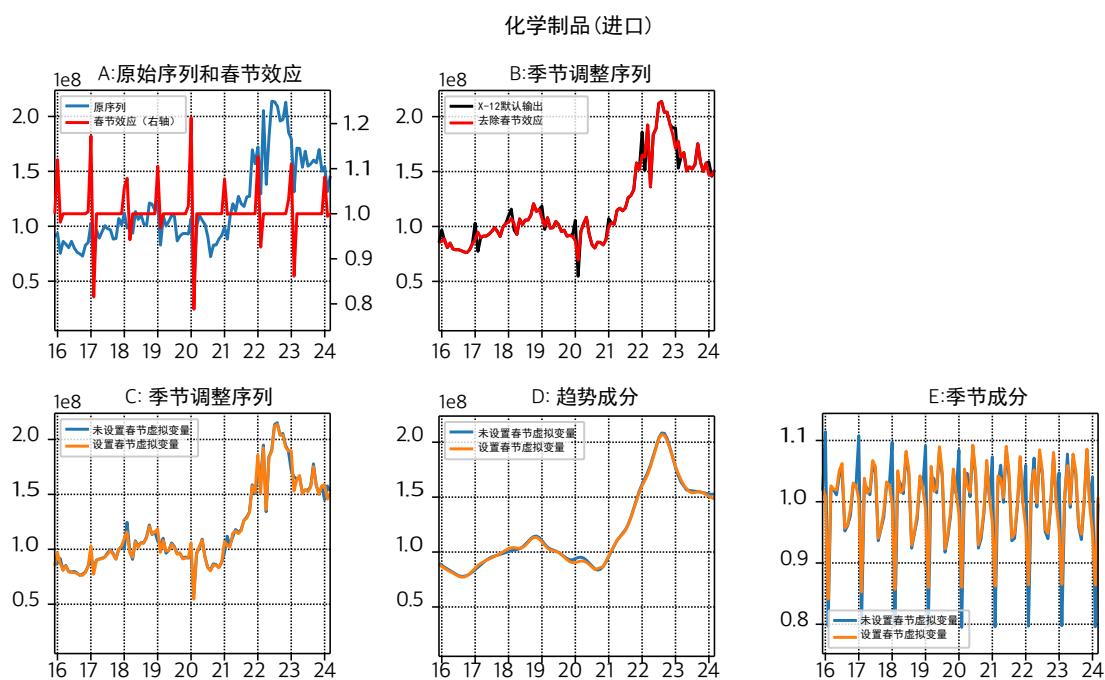


图 20: 化学制品(进口)

按原料分类的制成品(出口)

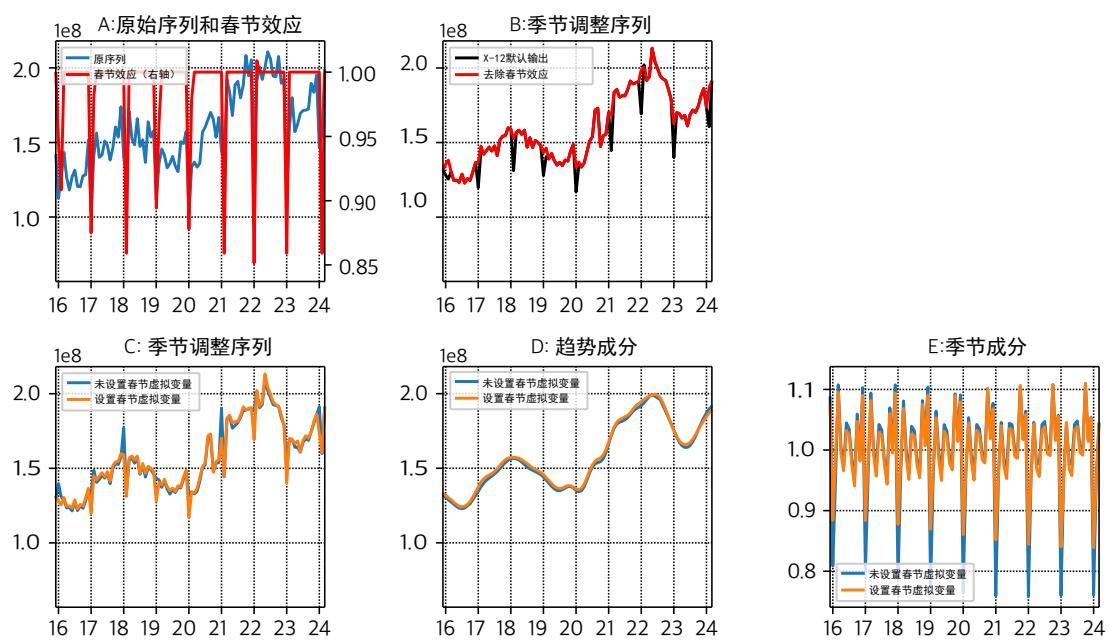


图 21: 按原料分类的制成品(出口)

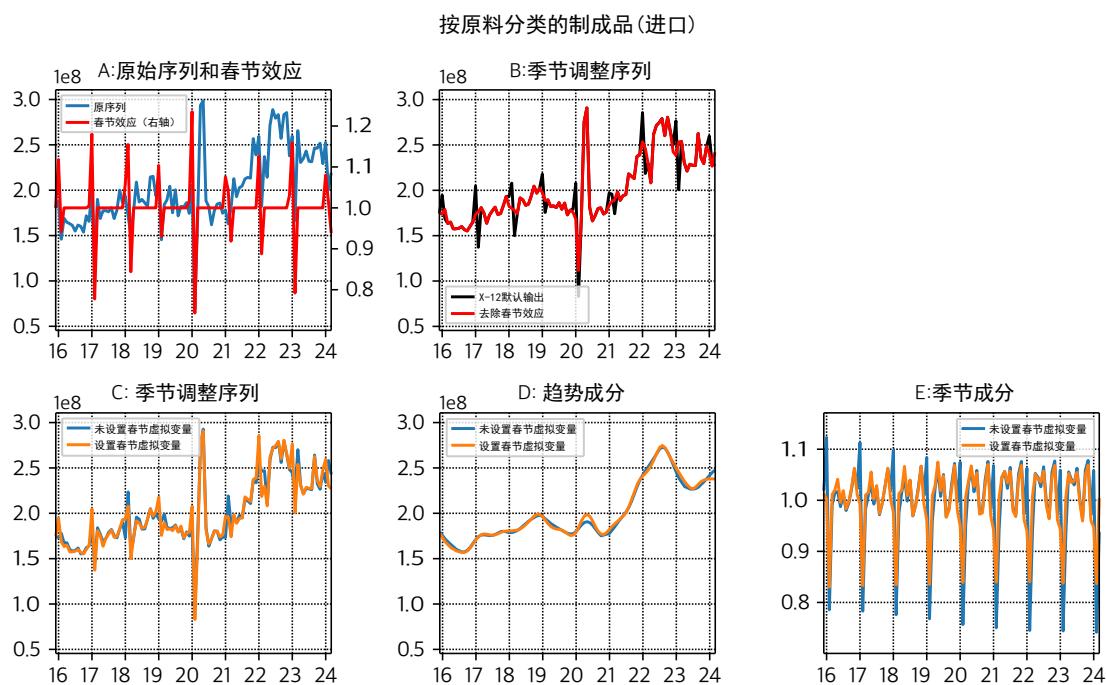


图 22: 按原料分类的制成品(进口)

一般机械(出口)

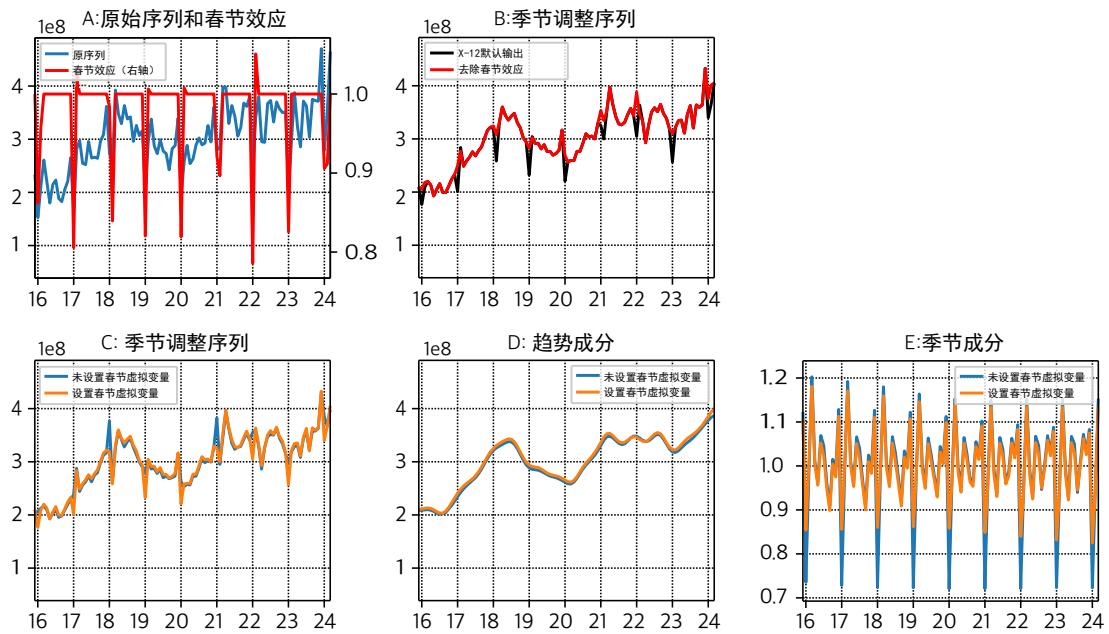


图 23: 一般机械(出口)

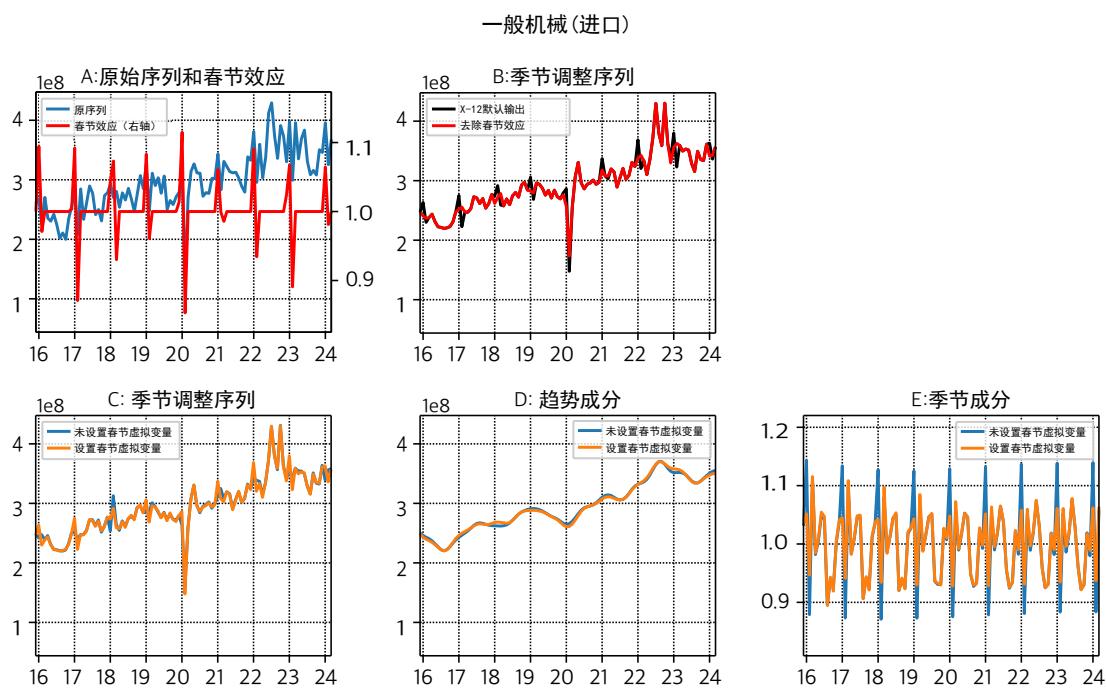


图 24: 一般机械(进口)

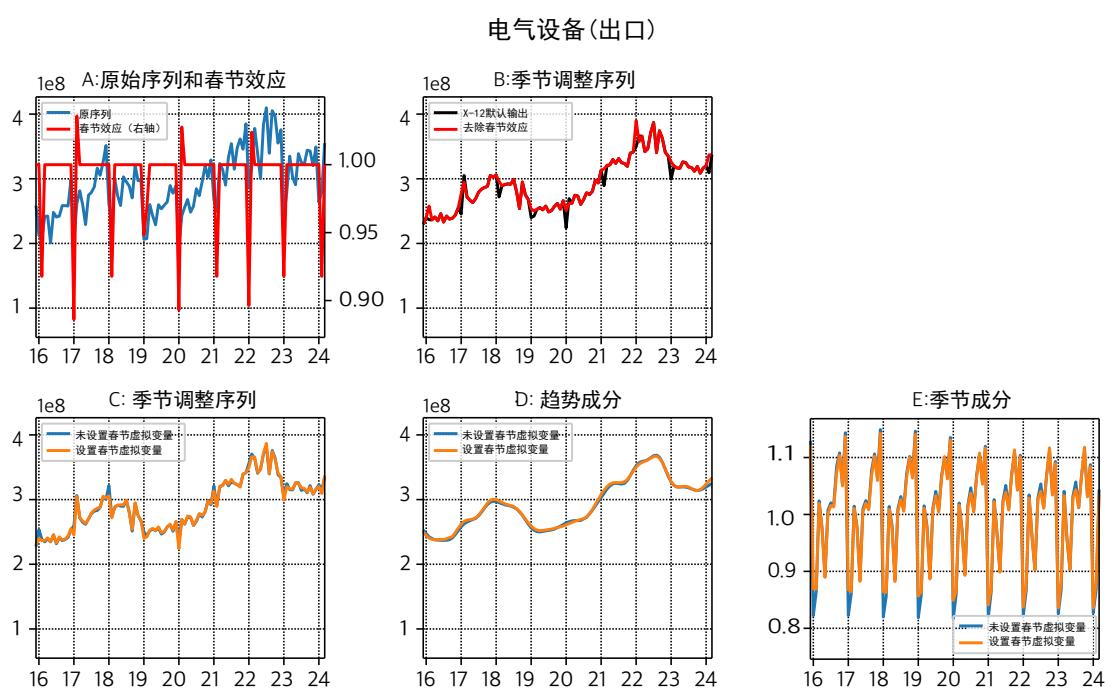


图 25: 电气设备(出口)

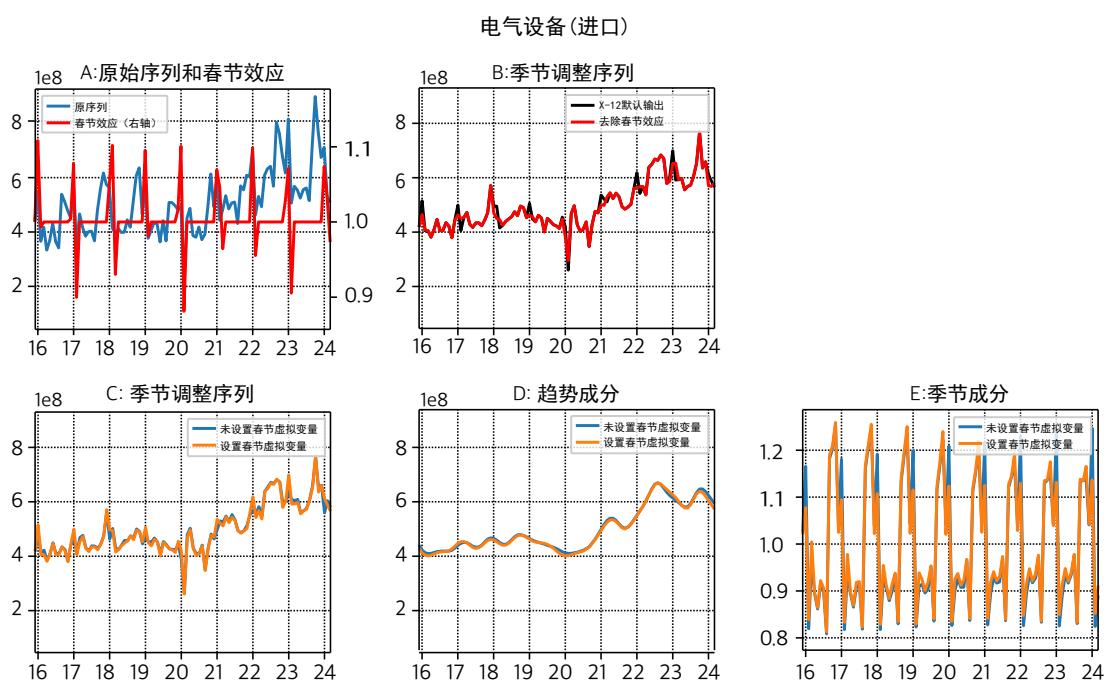


图 26: 电气设备 (进口)

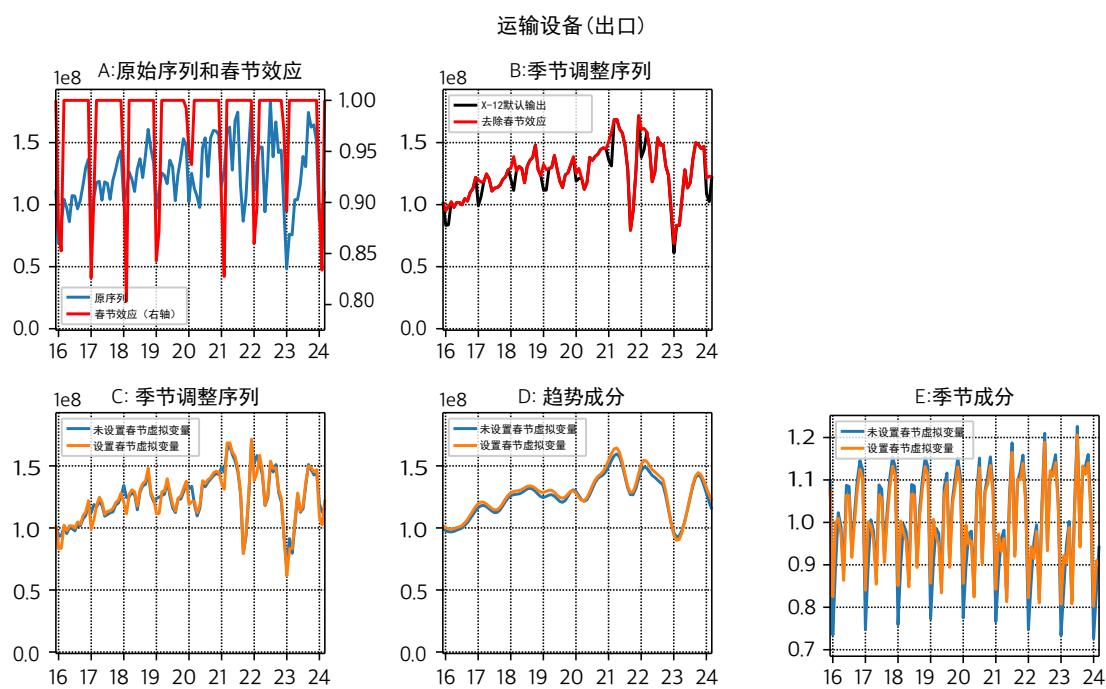


图 27: 运输设备(出口)

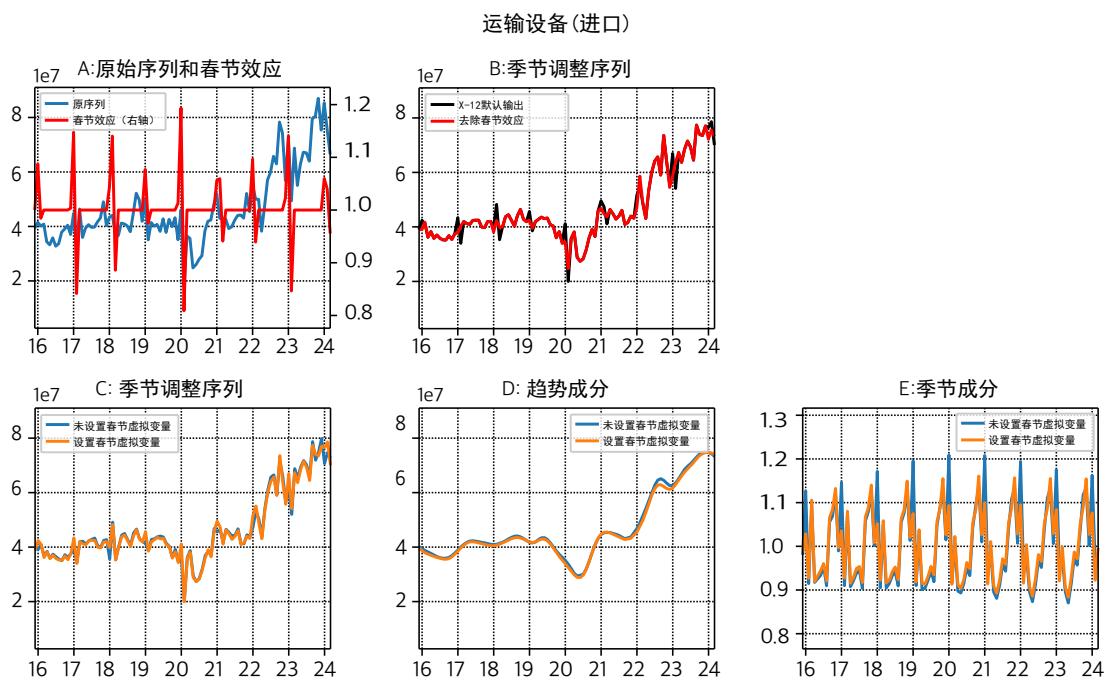


图 28: 运输设备(进口)

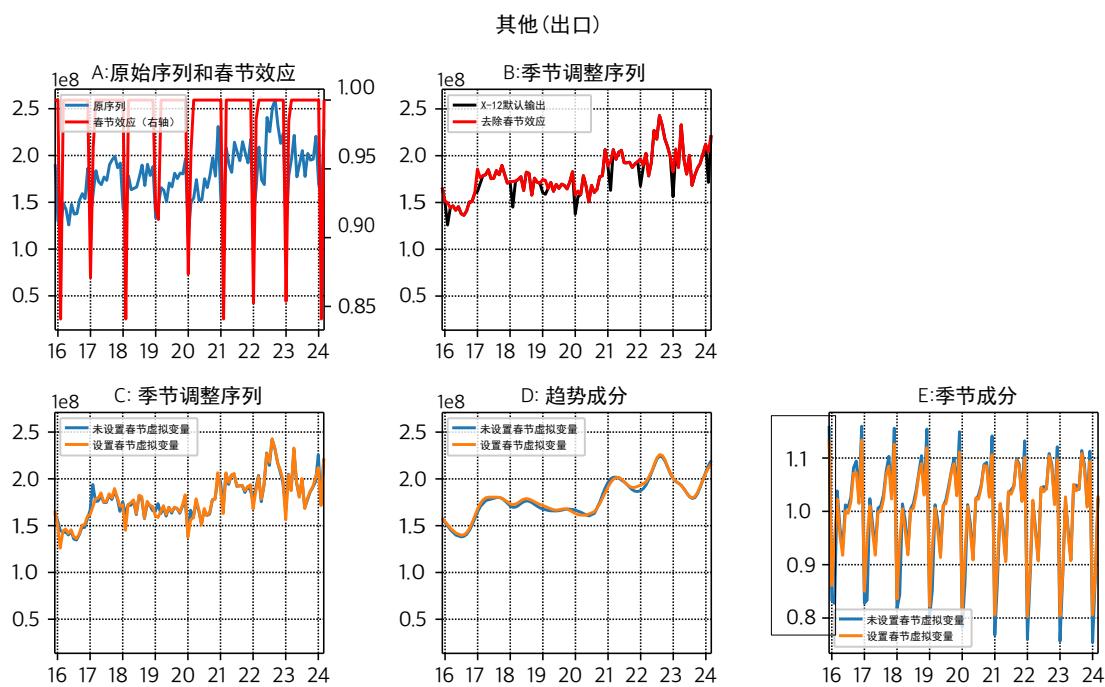


图 29: 其他(出口)

其他(进口)

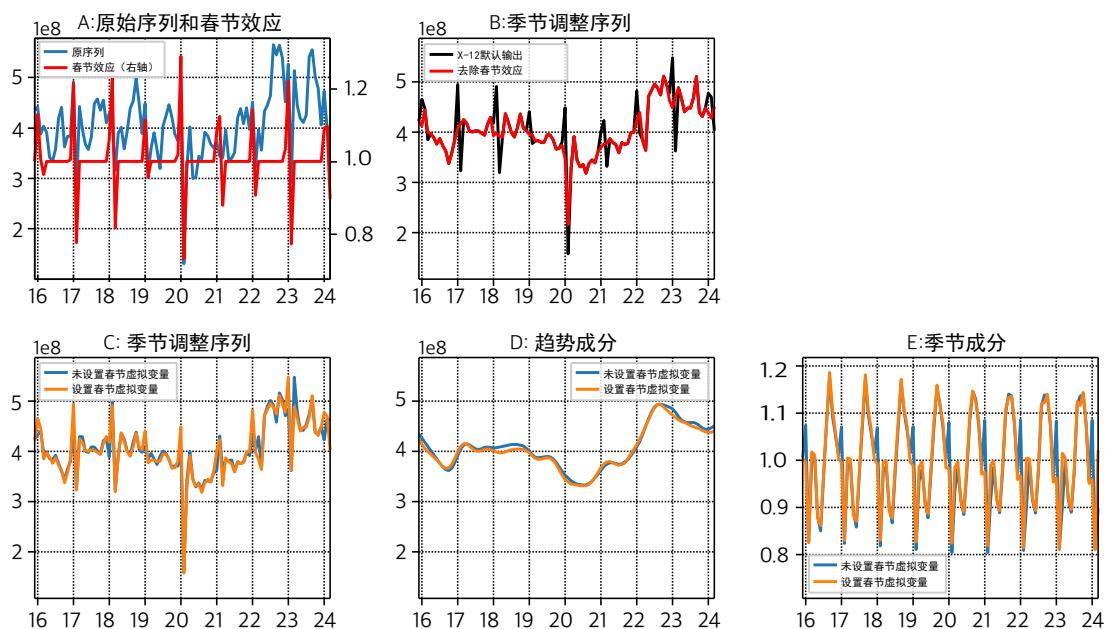


图 30: 其他 (进口)

4.7 小结

综合以上所有结果可以确认如下几点：

- 除矿物燃料（出口）以外的所有序列中，包含春节虚拟变量的模型的 AIC 均小于不包含的模型。
- 春节效应在出口与进口中表现出两种显著不同的模式。
- 在多数序列中，春节前的效应持续时间较春节后的效应更长。

此外，根据 X-12-ARIMA 与 X-13ARIMA-SEATS 的处理流程，可以预期如下趋势，并且这些趋势与本次估计结果基本一致。

未包含春节虚拟变量的情况

- 春节效应会被分散到趋势、季节成分和不规则成分之中。
- 春节效应会影响 ARIMA 部分的参数估计与预测。
- 季节调整序列中仍然会残留一定程度的春节相关变动。
- 由于春节效应的影响，不包含虚拟变量的季节成分相比包含的情况，整体振幅更大。

包含春节虚拟变量的情况

- 春节效应由 RegARIMA 模型进行估计。
- 春节效应不会影响 ARIMA 部分的参数估计与预测。
- 所估计出的春节效应最终被划分至不规则成分，对趋势和季节变动不产生影响。
- 默认设置下，季节调整序列是趋势与不规则成分之和，因此春节效应仍包含在季节调整序列中。
- 在默认设置下运行时，相较于不使用虚拟变量的情形，最终的季节调整序列中春节效应往往保留得更明显。

特别需要注意的是，在包含春节虚拟变量的情况下，默认处理流程下最终的季节调整序列中，春节效应反而更为突出。

各图的下方绘制了使用虚拟变量所得到的默认季节调整值与假设剔除春节效应后的季节调整值的对比，可以看到在多数序列中，默认的季节调整值中春节效应仍

非常明显。然而，简单地将剔除春节效应后的序列作为最终的季节调整值也存在问题。例如，对于所谓的星期效应而言，由于它基于格里高利历存在严格的周期性，因此只要在一定时间范围内进行时间聚合，该效应就会趋于零。因此，即使从季节调整序列中去除星期效应，其对季节调整值的累计值等影响也可以认为较小。

但另一方面，本研究中采用的春节虚拟变量在定义上并未施加使其具备周期性的约束。因此，即便对春节效应进行时间聚合，其结果也不会趋于零⁷。正因如此，如果简单地将剔除春节效应后的序列作为季节调整值，反而可能会对其水平造成高估或低估。另外，由于春节的影响非常显著且广泛，是否具有构造出“在春节不存在的情况下”的假想水平的实际意义也仍不明确。

综上所述，在经济分析中，更现实的应对方式或许是，在使用春节虚拟变量进行估计时，将注意力集中于趋势的变化而非季节调整值的波动。

5 修正历史检验（Revision History）

5.1 Revision History

本节将进行修正历史检验（Revision History）。修正历史是指当季节调整后的统计数据已经发布之后，由于新数据的追加而对过去数据重新进行季节调整时所产生的改动幅度。这在实务上对统计编制者而言尤为重要，且一般来说，季节调整后的历史值改动越小，通常被认为是越优秀的季节调整方法。

在使用 RegARIMA 模型时，若通过引入合适的回归变量提高模型的预测精度，则修正幅度会减小，从而提升季节调整的稳定性。X-12-ARIMA 与 X-13ARIMA-SEATS 中，为了评估此类稳定性，内置了计算以下统计量的功能：

$$R_t^{(T)} = \left| \frac{A_{t|T} - A_{t|t}}{A_{t|t}} \right|, \quad (t < T)$$

其中， $A_{t|s}$ 表示基于时点 s 为止的数据所计算的时点 t 的季节调整值或趋势成分。因此， $A_{t|t}$ 是在时点 t 实时计算出的数值，而 $A_{t|T}$ 是使用未来直到时点 T 的数据后，对时点 t 重新估计所得的值。两者之间的偏离率由 $R_t^{(T)}$ 表示。

⁷中国的农历（太阴太阳历）会通过在 19 年中插入 7 次闰月来修正与太阳年的偏差，因此虽然具有约 19 年的周期性，但农历的月日并不会完全回到同一公历日。

5.2 估计结果

图 31 至 图 54 展示了前节所用数据的偏离率计算结果及其图示。此处将 T 设定为 2022 年 12 月, t 设定为 2020 年 1 月至 2022 年 11 月之间。

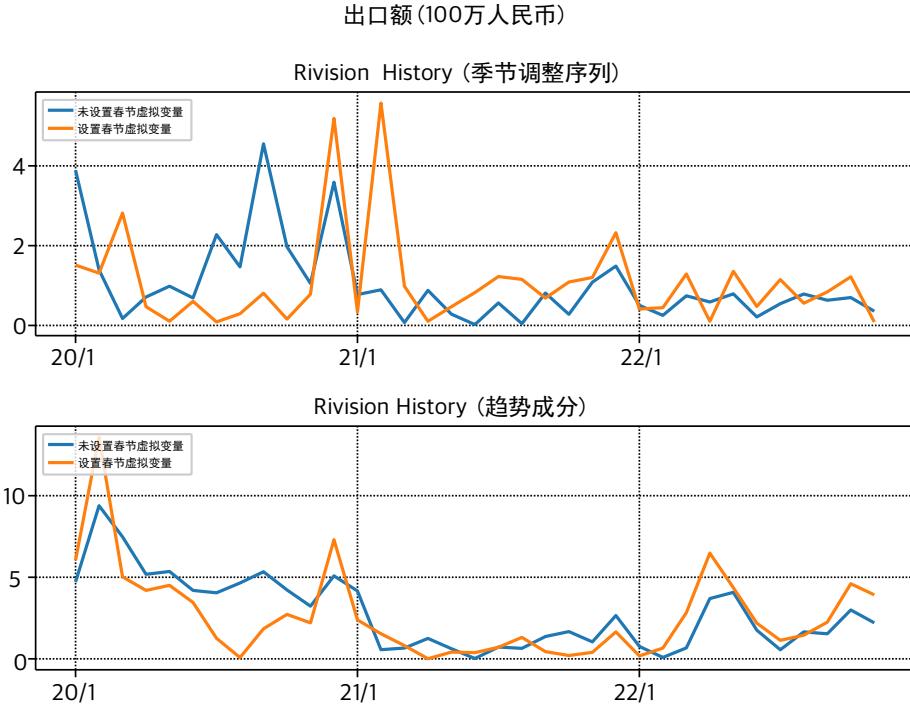


图 31: 出口额 (100 万人民币)

5.3 小结

偏离率 $R_t^{(T)}$ 在 RegARIMA 模型预测适当的情况下应接近于 0，根据 AIC 的比较结果可以推测，包含春节虚拟变量的模型应当具有更小的乖离率。然而，根据图 31 至图 54 的结果来看，无论是在季节调整序列还是趋势序列上，都未观察到明显的差异。

6 结论

6.1 结果总结

在本次估计中，我们构建了对应春节及其前后的三个虚拟变量，并据此进行了季节调整。在所分析的 24 个序列中，有 23 个序列至少有一个虚拟变量在统计上显著，且从 AIC 的角度来看，作为回归变量引入这些虚拟变量在大多数情况下都是合理的。

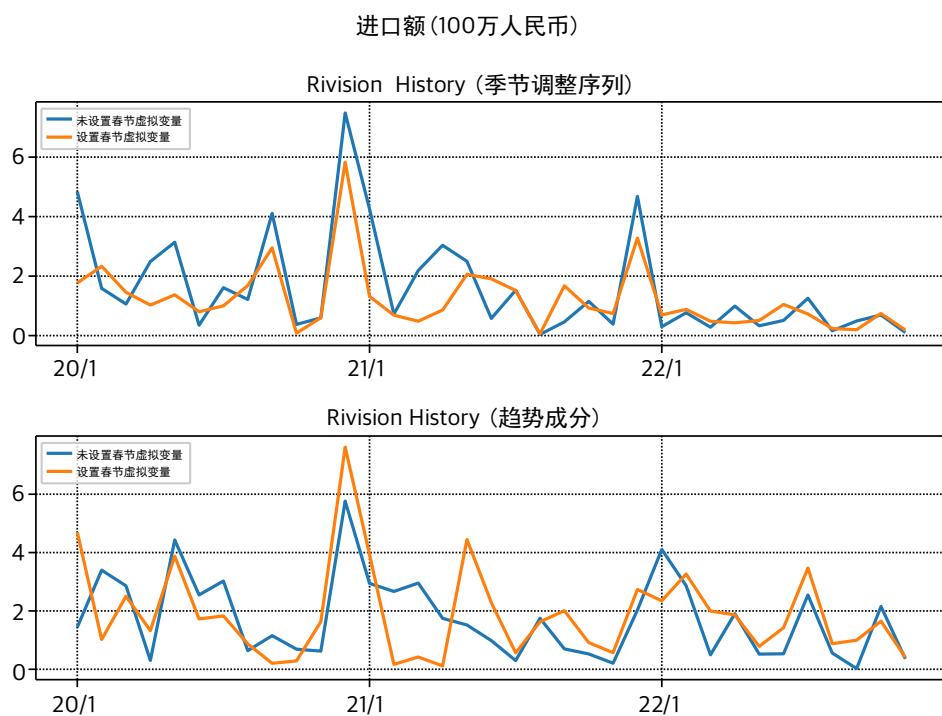
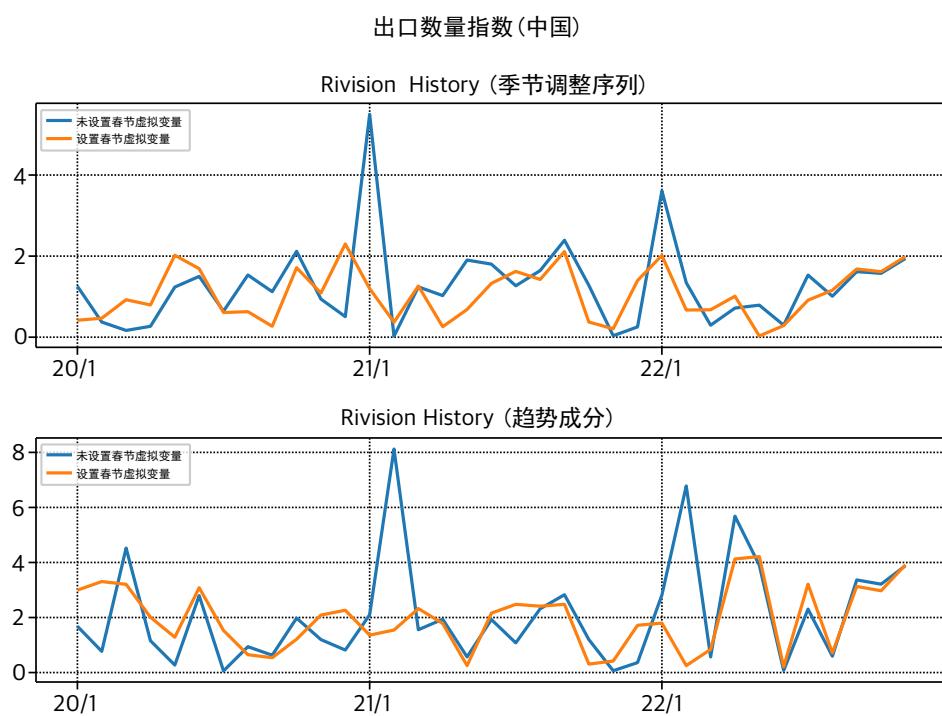


图 32: 进口额 (100 万人民币)



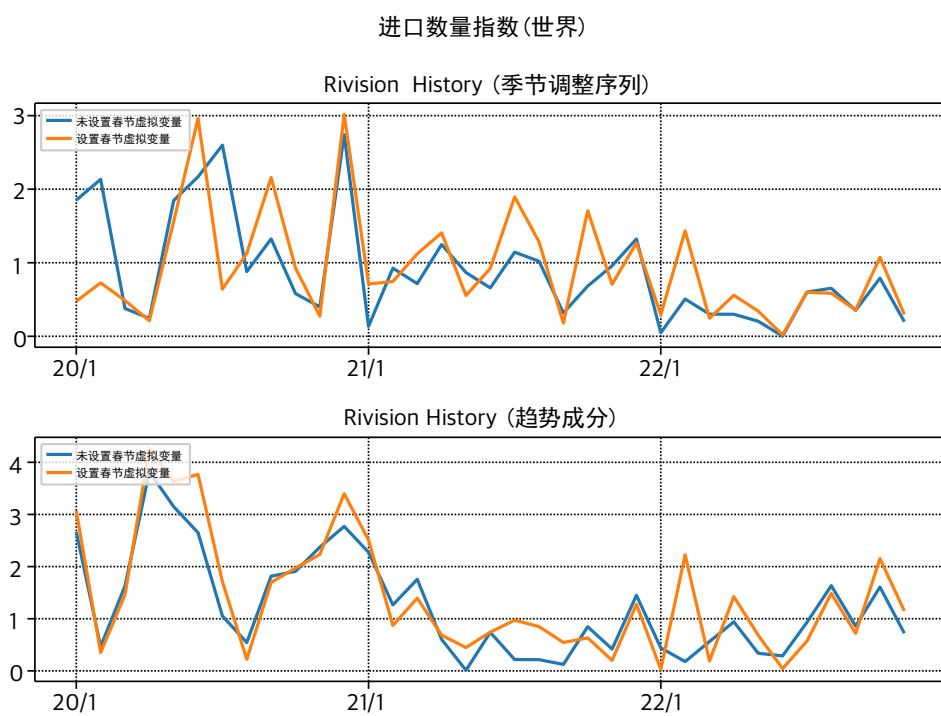


图 34: 进口数量指数 (世界)

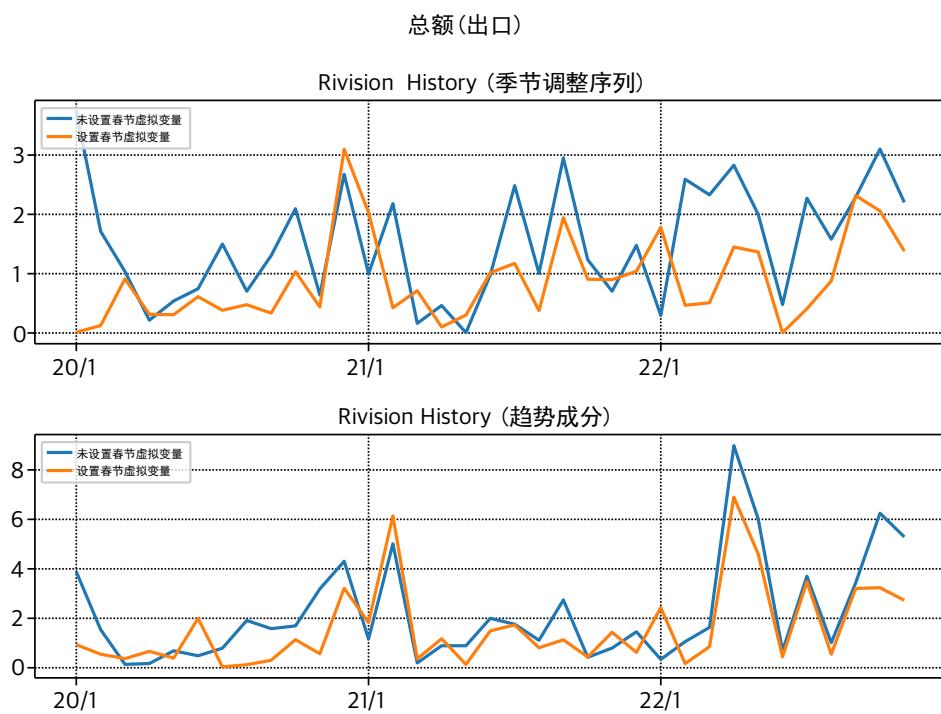


图 35: 总额(出口)

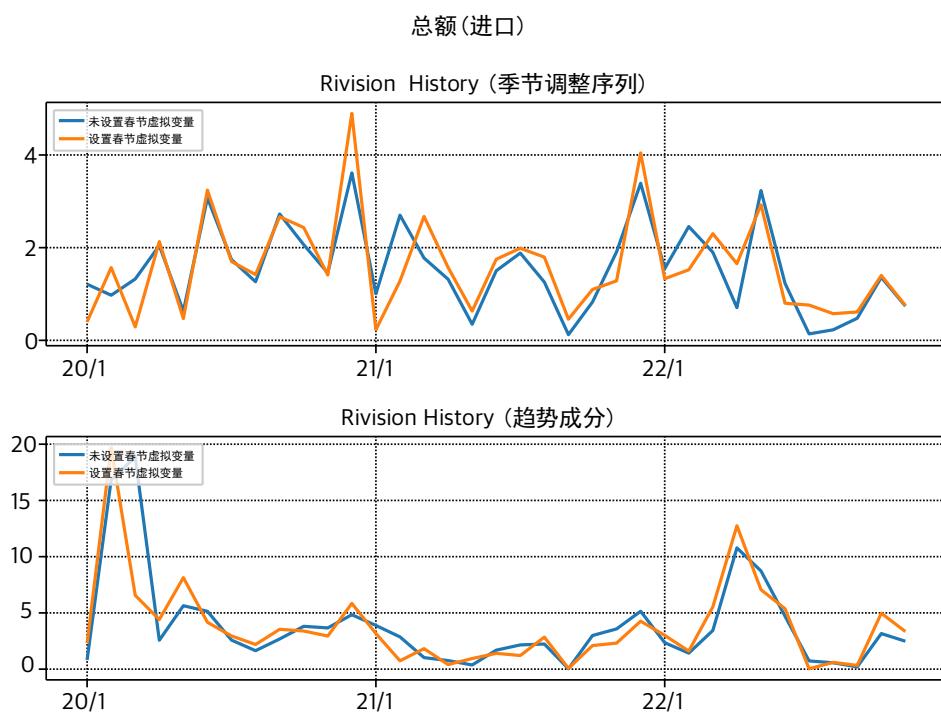


图 36: 总额(进口)

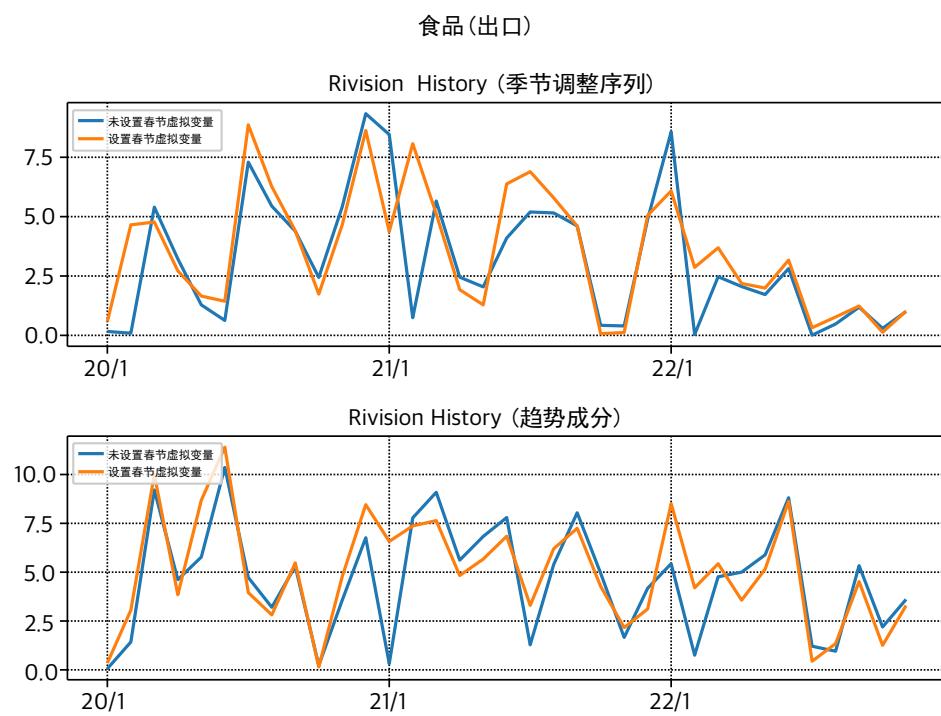


图 37: 食品(出口)

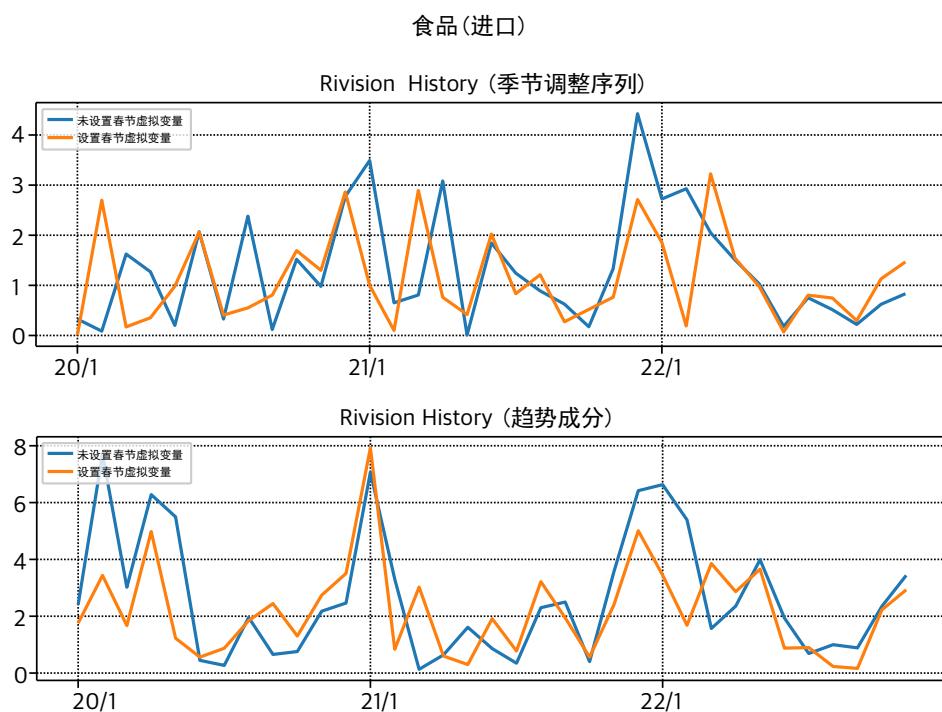


图 38: 食品(进口)



图 39: 原料(出口)

原料(进口)

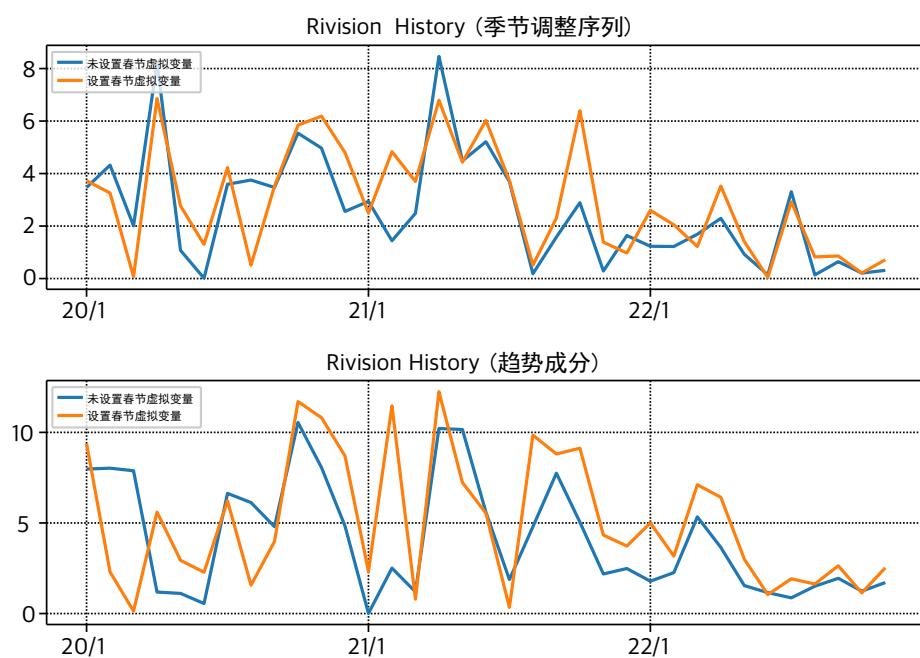


图 40: 原料(进口)

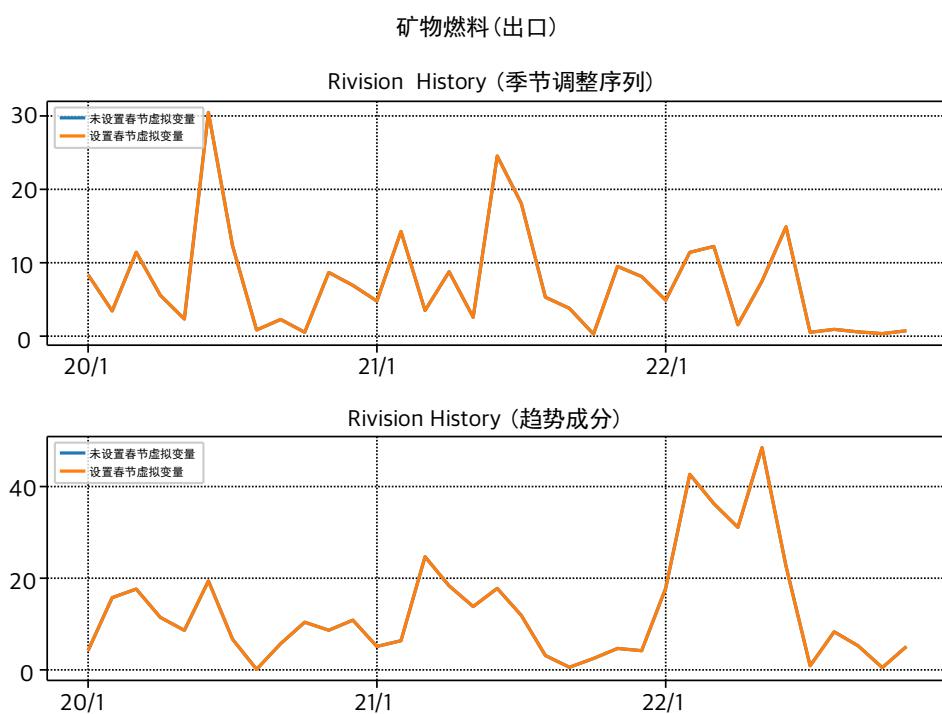


图 41: 矿物燃料(出口)

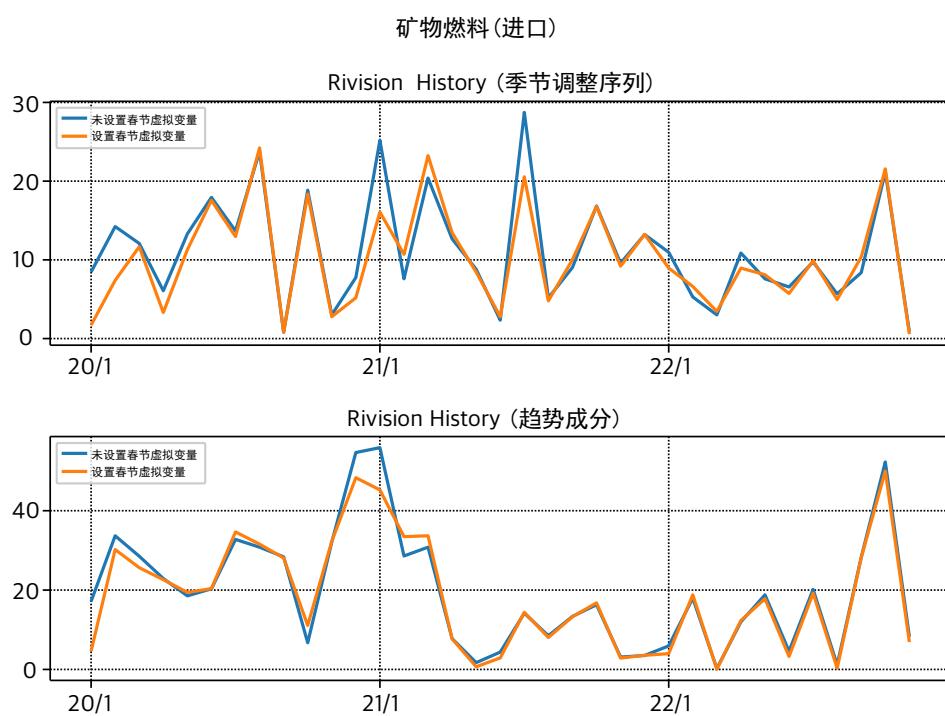


图 42: 矿物燃料(进口)

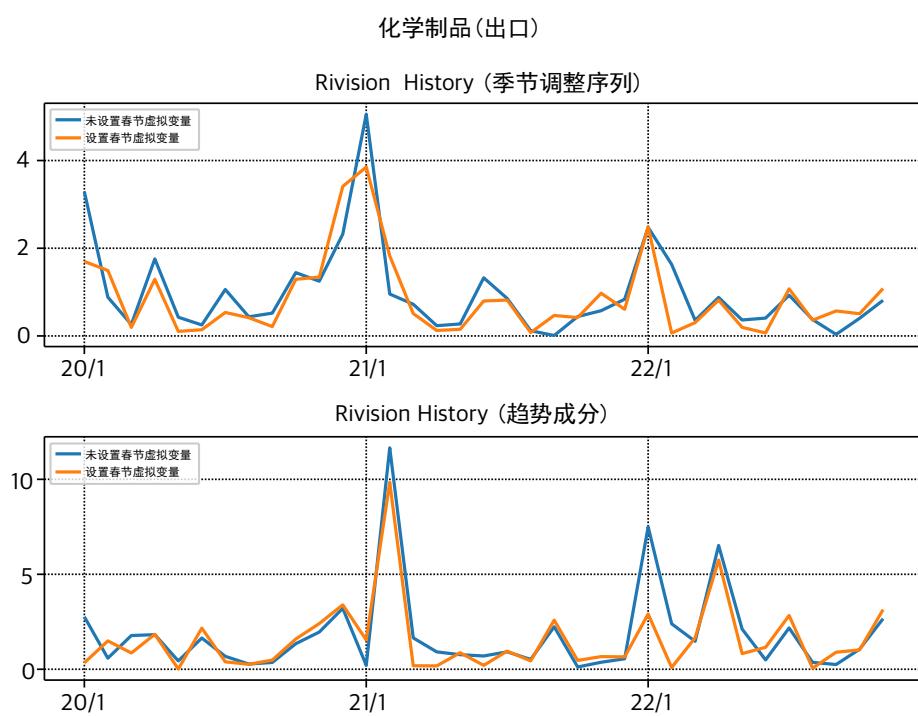
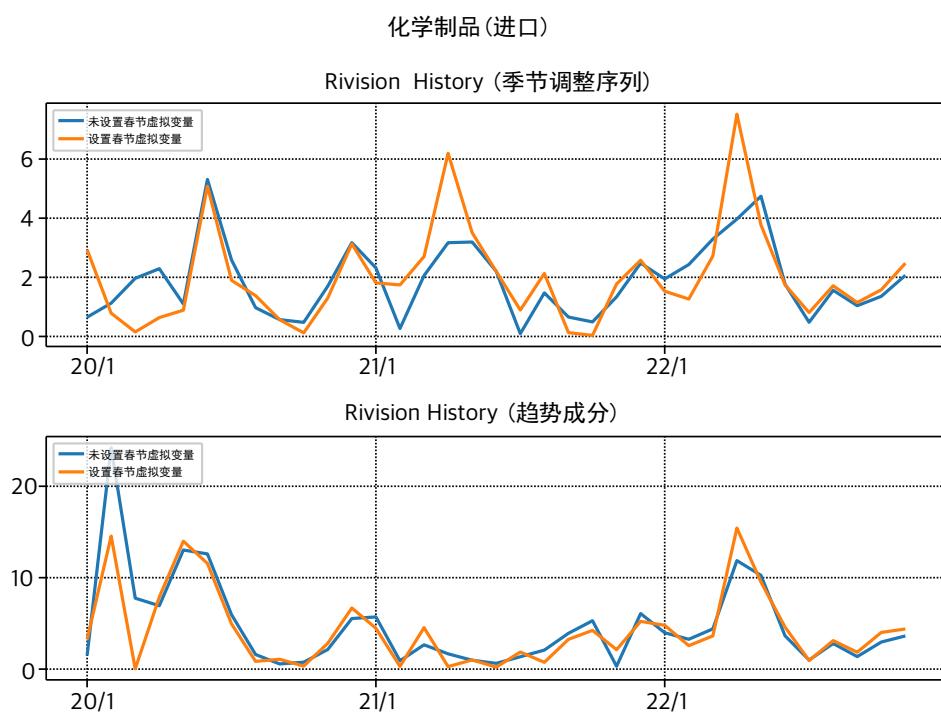


图 43: 化学制品(出口)



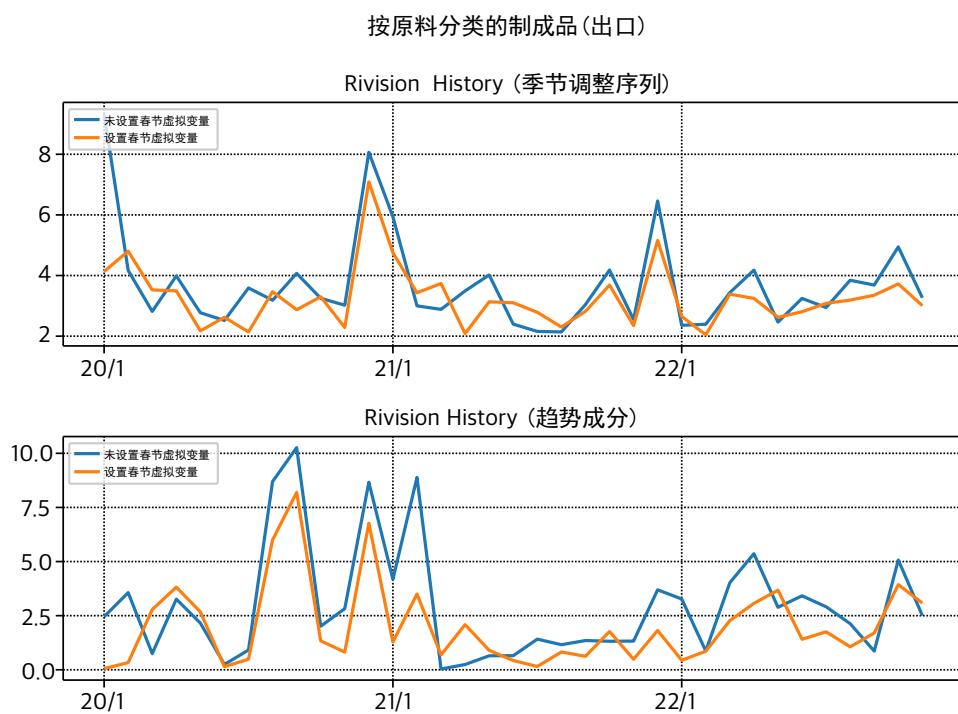


图 45: 按原料分类的制成品(出口)

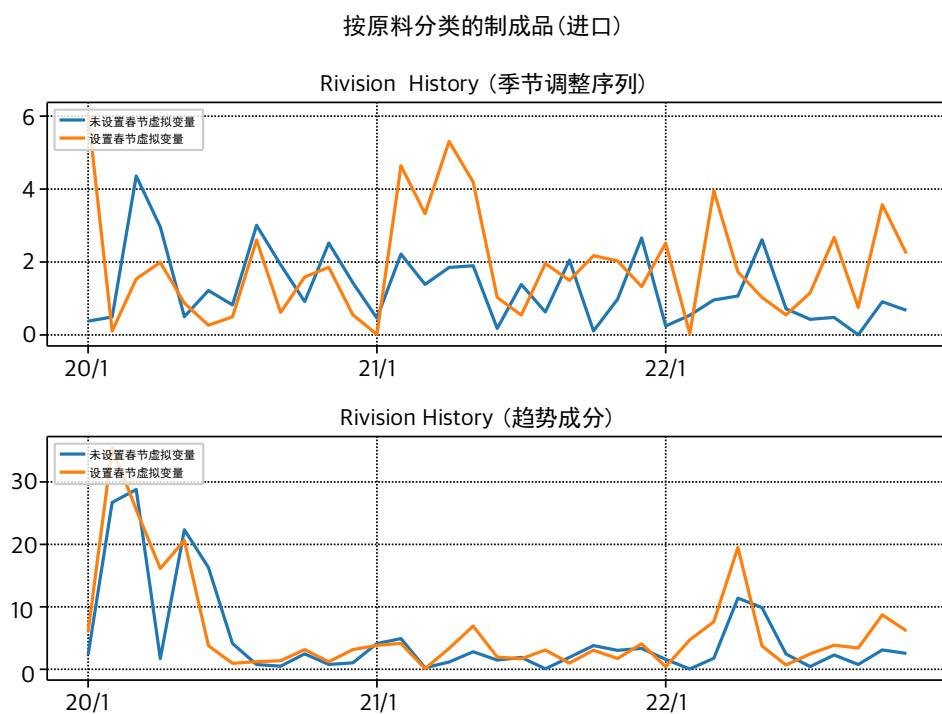


图 46: 按原料分类的制成品(进口)

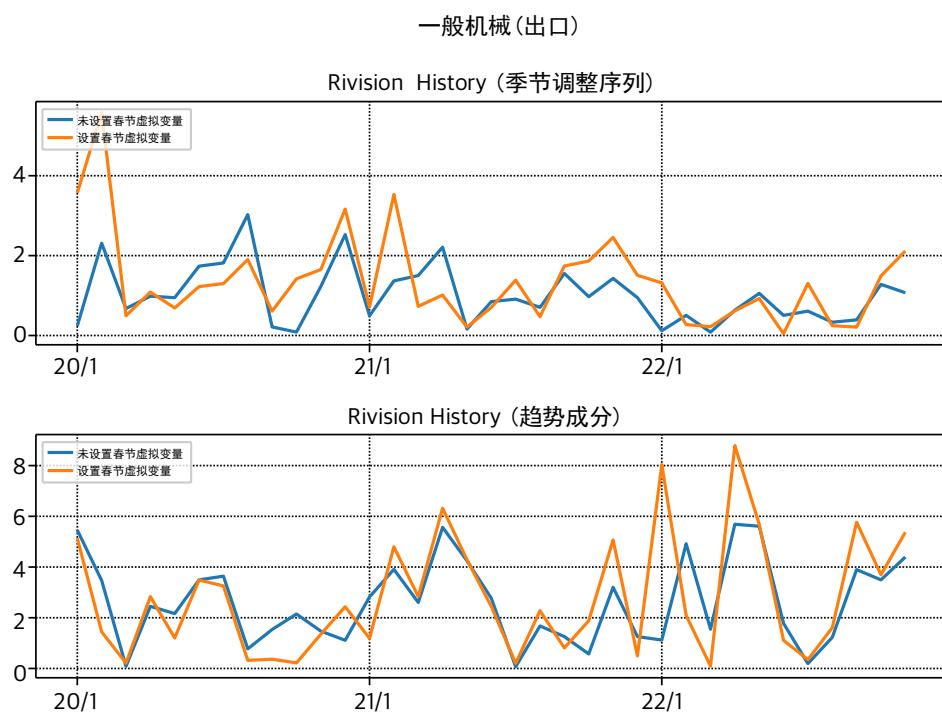
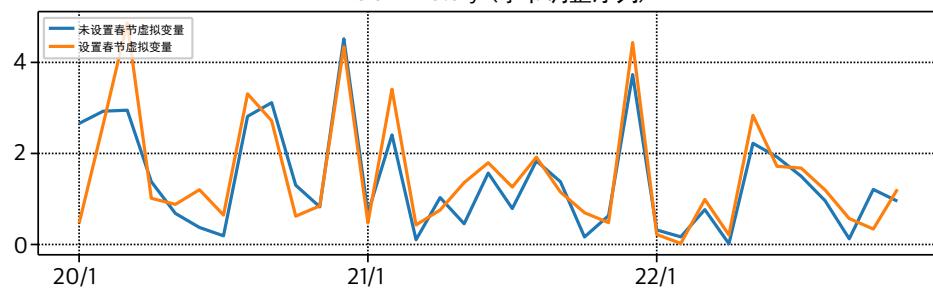


图 47: 一般机械(出口)

一般机械(进口)

Revision History (季节调整序列)



Revision History (趋势成分)

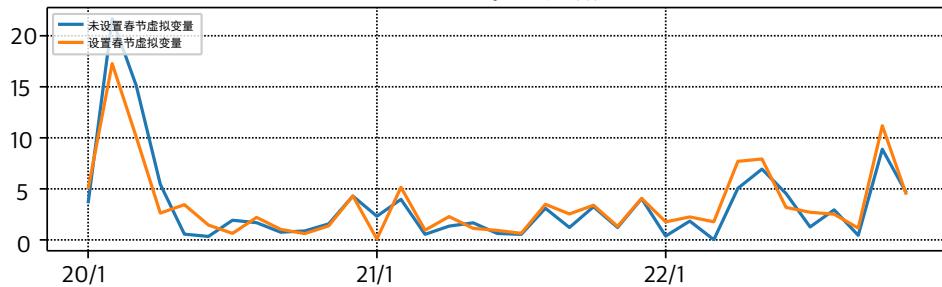
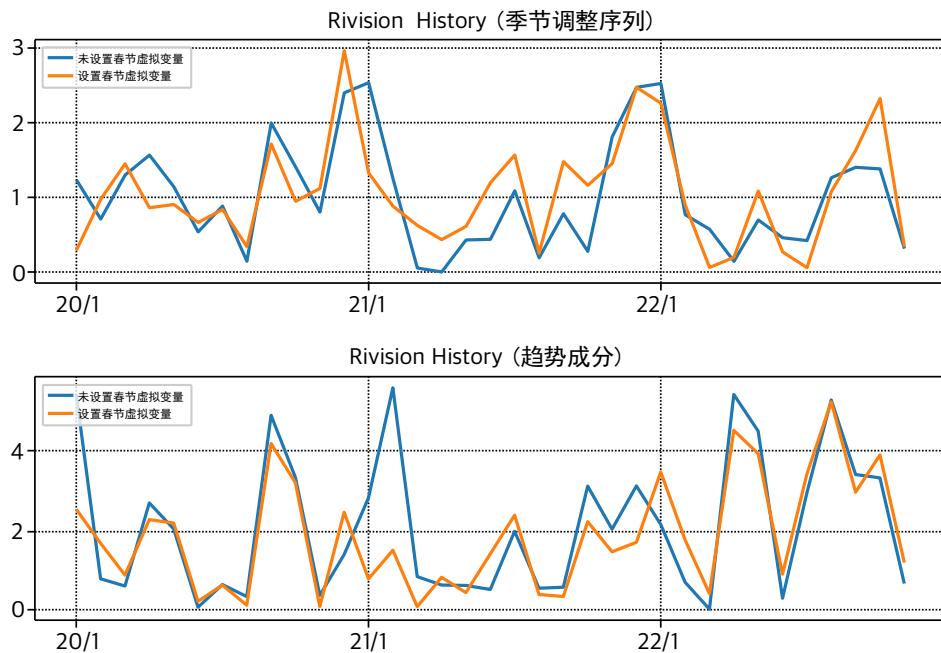


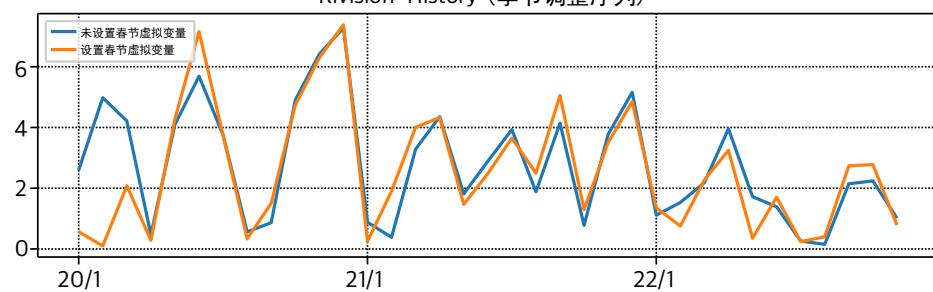
图 48: 一般机械(进口)

电器设备(出口)



电器设备(进口)

Revision History (季节调整序列)



Revision History (趋势成分)

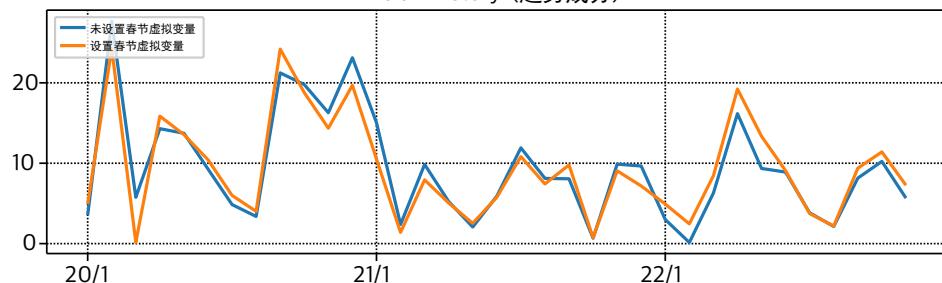


图 50: 电气设备(进口)

运输设备(出口)

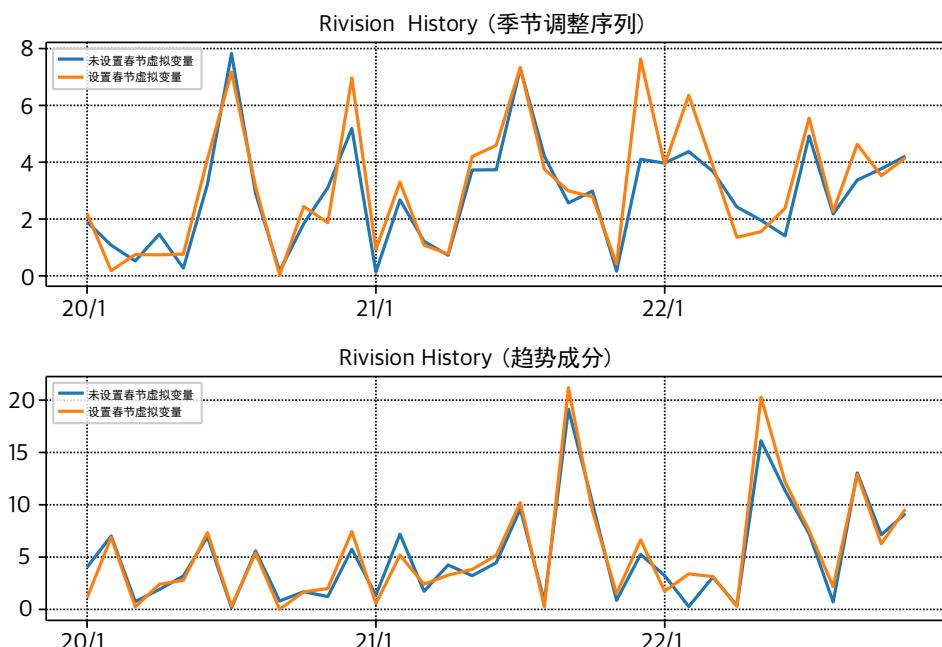


图 51: 运输设备 (出口)

运输设备(进口)

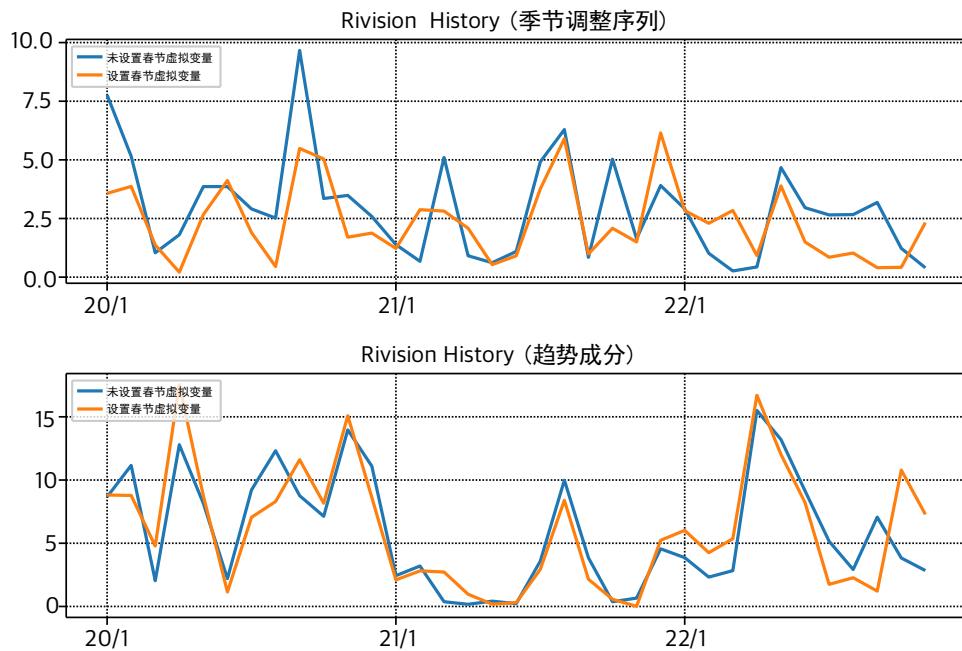
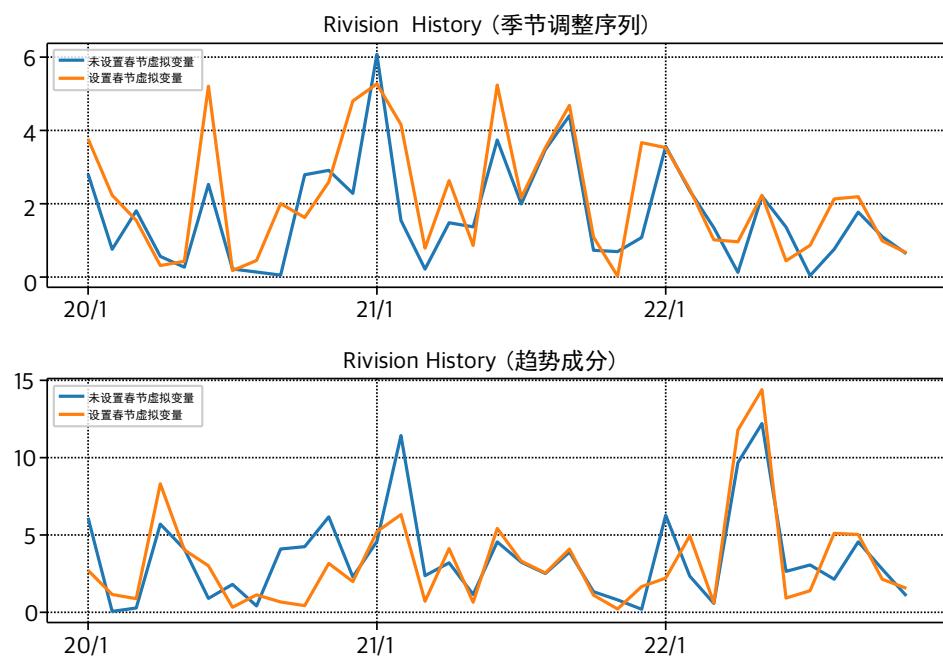


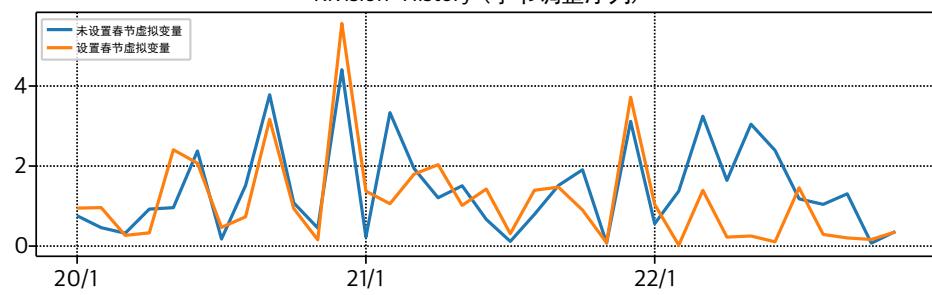
图 52: 运输设备(进口)

其他(出口)



其他(进口)

Revision History (季节调整序列)



Revision History (趋势成分)

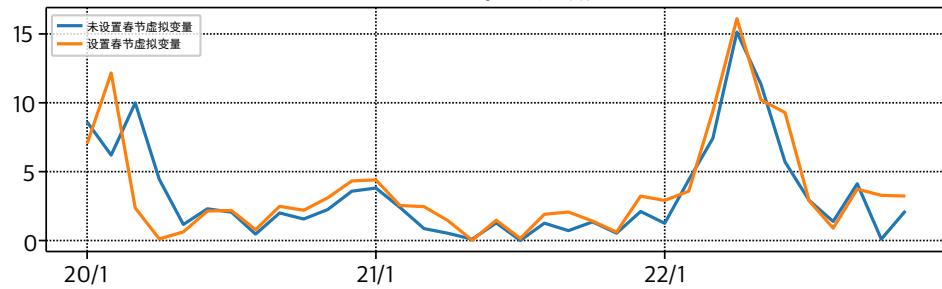


图 54: 其他(进口)

春节影响的表现形式大致可分为出口和进口两种类型：日本对中国的出口在春节期间出现下降，而中国对日本的进口则往往在春节前增长，并在春节期间出现回落。

但是，是否出现春节相关的剧烈波动因年份而异，使用本次估计中采用的简单虚拟变量模型，在某些年份中并未能很好地拟合整个序列。

春节的影响期间 τ 是通过 AIC 来选择的。整体来看，春节前的影响持续时间普遍比春节后更长，但不同序列之间差异较大，因此需要根据各个序列的特性进行个别调整。

此外，主要的长期假期之一是国庆节（10月1日开始），部分年份的中秋节（9月）可能移动至月底并与国庆节连休，这使得该假期具有较强的不规则性。虽然本稿未对其展开分析，但在以与春节相同的方法构建国庆节虚拟变量并进行估计后，结果显示多数情况下并不显著。对此，可能还需要对国庆节期间中国国内的行为模式或习惯进行更深入的调查分析。

6.2 季节调整所涉及的问题

当作为回归变量引入春节变量时，由于 X-12-ARIMA 与 X-13ARIMA-SEATS 的结构特性，在最终的季节调整序列中，春节效应所对应的峰值可能反而比未引入春节变量时更加显著，这一点也在本次估计结果中得到了确认。对此，如果简单地说明为“对春节效应进行了调整”，可能会导致使用者误解为已将春节带来的波动平滑化，因此在将春节调整机制引入官方统计中时，有必要对此保持谨慎。

虽然从季节调整值中剔除春节效应的序列在观察中长期趋势上更为清晰，但如果将季节调整理解为“仅剔除一年周期的变动，不进行其他加工处理”，那么将剔除春节效应后的序列视作季节调整值的做法在概念上或许并不恰当。另一方面，从统计建模的角度看，引入春节变量是合理且有益的，因此在采用春节变量的前提下，与其关注季节调整值的变化，将重点放在趋势成分的变动上可能是一种更现实的分析策略。

但是，关于引入春节变量是否能提升季节调整的稳定性（即减少对过去季节调整值的修正幅度），在本次的验证范围内尚未得出明确结论。

6.3 今后的研究方向

在日本的贸易统计中，部分序列目前仅作为参考值发布了季节调整序列，但近年来由于中国因素的影响日益加大，今后可能有必要引入对春节等特殊因素的适当处理。本次估计以贸易统计为对象，但对生产、观光、入境旅游等相关数据的验证也被认为是必要的。

关于虚拟变量的设定，也可考虑引入根据时期波动的变量，或是考虑农历周期性的变量等方式，这些仍需作为后续研究课题继续探索。

参考文献

- Bell, William R., and Scott C. Hillmer (1983). “Modeling Time Series with Calendar Variation”. In: *Journal of the American Statistical Association* 78.383, pp. 526–534.
- Lin, Jin-Lung, and Ta-Chung Liu (2003). “Modeling Lunar Calendar Holiday Effects in Taiwan”. In: *Taiwan Economic Forecast and Policy* 33.2, pp. 1–37.

非平稳季节性经济时间序列中的缺失值问题⁸

2024-5-22

2024-10-15(修订)

国友直人⁹

1. 前言

近年来，随着中国经济的扩张，其与日本经济及世界经济之间的联系与影响力不断增强。这种中国宏观经济的动向，尤其是近年来出现的大幅波动的走向，已成为日本经济学家们高度关注的话题。然而，负责中国官方统计的当局出于若干原因，有时采用与主要发达国家不同的宏观信息公布形式，使得掌握中国实体经济的实际动向变得困难。例如，近年来对于作为重要宏观经济指标的零售业数据，尽管其为月度数据，但中国并不单独公布1月与2月的数值，而是在之后仅公布“1月与2月的合计值”。这可以理解为中国统计当局在考虑春节等与主要发达国家不同的季节性变动因素后，为避免公布的数值在国内外引起误解而采取的做法。下图1展示的是由中国国家统计局公布的2002年至2023年的月度零售业数据。自2012年以后，1月与2月的官方数据均为缺失值）（不过，对于1月与2月的部分，现行系列的合计值会在事后公布。）

⁸本文是为统计咨询演习「可移动假日与经济指标的季节性」准备的备忘录。

⁹统计数理研究所

图 1：零售业销售数据（2002–2023 年）



另一方面，在日本等发达国家的官方统计人员和经济学家之间，对于主要的宏观经济数据，从原始序列中制作季节调整后的序列是一种普遍做法。在此过程中，常常使用由美国人口普查局开发的 X-12-ARIMA 方法或 X-13ARIMA-SEATS 方法进行季节调整。然而，当原始数据中存在缺失值时，如何恰当地处理其季节性问题，目前仍不明确。

本稿以中国若干重要宏观经济时间序列的分析应用为出发点，讨论非平稳季节性经济时间序列中的缺失值问题。在实际观测到的宏观经济原始时间序列中，趋势成分、循环成分、季节成分与不规则成分往往交织共存，可视为一种混合了非平稳性与平稳性的时间序列。然而，在这种情况下，传统统计时间序列分析的文献中，仍存在许多尚未得到充分讨论的统计问题。因此，本文尝试在此背景下探讨统计分析的方法，特别是利用 DECOMP 滤波与 SIML 滤波来处理缺失值问题，并对时间序列数据分析的结果进行解释与考察。

2. 非平稳时间序列中的缺失值处理

为简化起见，设时间序列的维度为 $p = 1$ ，观测数据 y_i 的观测期为 $i = 1, \dots, n$ ，但假定在总体集合 $\mathbf{I}_n = \{1, 2, \dots, n\}$ 中，有一部分 $i \in \mathbf{I}_c$ 的数据为缺失 (missing)。

$$y_i = x_i + v_i \quad (i \in \mathbf{I}_n) \quad (1)$$

其中， x_i 被视为非平稳时间序列，而 v_i 被视为平稳时间序列。在统计时间序列分析中所讨论的季节性 ARIMA 模型 (seasonal ARIMA model) 也可以看作是本模型的一个特殊情形。

在这种情况下，可以如下方式对问题进行形式化。例如，当存在 h 个缺失数据时，将缺失数据的集合记为 $\mathbf{I}_n^c = \{m, m+h-1\}$ ，而观测到的时间序列数据集合为 $\mathbf{I}_n = \{1, 2, \dots, m-1, m+h, m+h+1, \dots, n\}$ ，此时数据总数为 $n-h$ 个。记整体观测向量为 $\mathbf{y}(\mathbf{I}_n)$ ，缺失数据为 $y_i; (i \in \mathbf{I}_c)$ 。利用一个 $(n-h) \times 1$ 的向量 \mathbf{b}_i ，可定义： $\hat{y}_i = \mathbf{b}'_i \mathbf{y}(\mathbf{I}_n)$ 。因此， $\hat{y}_i = \mathbf{e}'_i \mathbf{y}; (i \in \mathbf{I}_n)$ ，其中 \mathbf{e}_i 为单位向量；而 $\hat{y}_i = \mathbf{b}'_i \mathbf{y}; (i \in \mathbf{I}_c)$ 。由此，通过一个 $n \times (n-h)$ 的变换矩阵，可以构成一个 $n \times 1$ 的向量：

$$\hat{\mathbf{y}} = \mathbf{B}\mathbf{y}(\mathbf{I}_n). \quad (2)$$

这种形式即使在缺失数据如实际应用中那样间隔分布（非连续）的情况下，只需调整下标即可使用，因此是一种一般化的表示方法。在这种设定下，若将向量 $\hat{\mathbf{y}}$ 视为

观测向量，并假定 $\mathbf{b}_i; (i \in \mathbf{I}_c)$ 已知，就可以进行相应的时间序列分解。

若观测数据所服从的联合分布已知，则条件分布 $y_i|\mathbf{Y}_{i-1} (i \in \mathbf{I}_n)$ 可表示为 $f(y_i|\mathbf{Y}_{i-1})$ 。当所有数据都可观测时，似然函数可写为

$$L_n(\mathbf{I}_n) = \prod_{i \in \mathbf{I}_n} f(y_i|\mathbf{Y}_{i-1}) \quad (3)$$

若观测数据中存在缺失值，则需要改为 $i \in \mathbf{I}_n \setminus \mathbf{I}_c$ 。当变换矩阵 \mathbf{B} 满足某些约束条件时，似然函数可近似为 $L_n(\mathbf{I}_n) = \prod_{i \in \mathbf{I}_n} f(\hat{y}_i|\hat{\mathbf{Y}}_{i-1})$ 。例如，若真实的状态服从 ARIMA 模型，且相对于数据总数而言缺失值并不多，则只要将变换设置为包含真实 ARIMA 模型的形式，即可实现一致（consistent）的估计。在这种情况下，若适当地令向量 $\mathbf{b}_i(\theta) (i \in \mathbf{I}_c)$ 依赖于参数 θ ，并假设状态变量 x_i 服从某种随机过程，则可以进行最大似然估计。

此外，在统计学中的缺失值问题中，若似然函数（例如高斯分布）已知，通常会使用 EM 算法进行估计。对于本研究所讨论的非平稳时间序列情形，包括其收敛性的合理性在内，都是一个值得深入探讨的问题，但本文暂不展开讨论。在实际的缺失值问题中，缺失可能是随机发生的，也可能有其他多种情况。即便如此，本文后半部分将介绍的 SIML 方法仍可以较为容易地扩展应用于这些情形。

(i) **Decomp 方法的情况**：Kitagawa (2020) 在假设高斯似然函数的前提下，引入状态变量并将模型表示为：

$$x_i = t_i + s_i + c_i , \quad (4)$$

其中： t_i ：趋势成分 (trend)， s_i ：季节成分 (seasonal component)， c_i ：循环成分 (cyclical component)，并假设：(a) $\Delta^k t_i = v_i^{(t)}$ （其中 k 为差分阶数）；(b) $(\sum_{j=0}^{p-1} c_{i-j} = v_i^{(s)})$ （其中 p 为季节周期）；(c) $c_i \sim AR(m)$ 服从以噪声 $v_i^{(v)}$ 为驱动的平稳时间序列；(d) 各噪声项均服从正态分布，模型的参数为 $v_i^{(t)}, v_i^{(s)}, v_i$ 的方差。在 Decompo 方法中，对这些参数进行最大似然估计。

根据 Kitagawa (2020)，对于缺测数据对应的时间点，由于没有观测值，不进行滤波处理，而是使用预测值来作为状态估计。通过最大化由预测误差得到的似然函数，可以实现状态变量的估计，并结合 smoothing（平滑化）可获得更为稳定的状态估计值。这一点在 Kitagawa (2020) 第 9 章中给出了示例。

在 DECOMP 的情况下，使用基于预测 (prediction) 误差的最大似然估计，但由于还进行平滑化 (smoothing)，其实际计算方法并非显而易见。例如，若先进行初始

状态估计，并通过迭代计算方式进行，则更加简便且实用。此外，对于星期变量等因素的处理也是可行的，但由于目前不存在现成程序，需要自行开发计算程序。

(ii) **SIML 的情况**：接下来介绍的 Sato 的 SIML-filtering 程序 (Kunitomo–Sato (2021), Sato–Kunitomo (2024)) 中，利用变换数据法对趋势、循环成分及季节成分进行状态估计，这种方法计算简单且实用，因此以下将更为详细地说明。

值得注意的是，当希望估计趋势与循环部分时，设 $x_i = t_i + c_i$ ，并令 v_i 中包含平稳的季节成分，则无需对季节成分或不规则波动成分假定特定分布等统计模型。只需根据最大频率参数 m 的选择，以 nonparametric (非参数) 方式仅估计周期长于一年的成分，然后再将其逆变换回时间域即可。因此，相较于通过插值补全缺测值的方法，此方法能得到更为稳健的趋势与循环部分的状态估计。

3.F-滤波与 B-滤波

设在第 i 期观测到的第 j 个变量为 y_{ji} ($i = 1, \dots, n; j = 1, \dots, p$)，假定 $\mathbf{y}_i = (y_{1i}, \dots, y_{pi})'$ 作为 $p \times 1$ 向量， $\mathbf{Y}_n = (\mathbf{y}_i')$ ($= (y_{ij})$) 作为 $n \times p$ 的观测矩阵，以及 \mathbf{y}_0 (或 \mathbf{y}_n) 作为 $p \times 1$ 的初始值向量，可以被观测到。此时，具有趋势与循环成分的非平稳状态变量 \mathbf{x}_i ($= (x_{ji})$) ($i = 0, 1, \dots, n$) 为 I(1) 过程，而平稳的不规则成分 $\mathbf{v}_i' = (v_{1i}, \dots, v_{pi})$ 为相互独立的序列。假设如下误差表达式成立：

$$\mathbf{y}_i = \mathbf{x}_i + \mathbf{v}_i \quad (i = 0, 1, \dots, n). \quad (5)$$

首先，我们考虑同时满足

$$\Delta \mathbf{x}_i = (1 - \mathcal{L}) \mathbf{x}_i = \mathbf{v}_i^{(x)} \quad (i = 1, \dots, n), \quad (6)$$

和 $\mathbf{E}(\mathbf{v}_i^{(x)}) = \mathbf{0}$, $\mathbf{E}(\mathbf{v}_i^{(x)} \mathbf{v}_i^{(x)'}) = \boldsymbol{\Sigma}_x$ 的随机变量序列 \mathbf{v}_i ($i = 0, 1, \dots, n$) 作为 i.i.d. 序列，使得 $\mathbf{E}(\mathbf{v}_i) = \mathbf{0}$, $\mathbf{E}(\mathbf{v}_i \mathbf{v}_i') = \boldsymbol{\Sigma}_v$ 成立的情形。此处进一步假设两个随机变量序列 $\Delta \mathbf{x}_i$ 与 \mathbf{v}_i 相互独立，并分别服从多元正态分布 $N_p(\mathbf{0}, \boldsymbol{\Sigma}_x)$, $N_p(\mathbf{0}, \boldsymbol{\Sigma}_v)$ 。若记 $n \times p$ 的观测矩阵为 $\mathbf{Y}_n = (\mathbf{y}_i')$ ，则在给定初始条件 \mathbf{y}_0 的情况下， $(\mathbf{y}_1', \dots, \mathbf{y}_n)'$ 的分布为

$$\text{vec}(\mathbf{Y}_n) \sim N_{n \times p} \left(\mathbf{1}_n \cdot \mathbf{y}_0', \mathbf{I}_n \otimes \boldsymbol{\Sigma}_v + \mathbf{C}_n \mathbf{C}_n' \otimes \boldsymbol{\Sigma}_x \right), \quad (7)$$

其中 $\mathbf{1}'_n = (1, \dots, 1)$,

$$\mathbf{C}_n = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ 1 & 1 & 0 & \cdots & 0 \\ 1 & 1 & 1 & \cdots & 0 \\ 1 & \cdots & 1 & 1 & 0 \\ 1 & \cdots & 1 & 1 & 1 \end{pmatrix}_{n \times n} . \quad (8)$$

接下来对 \mathbf{K}_n - 变换 (即从 \mathbf{Y}_n 到 $\mathbf{Z}_n (= (\mathbf{z}'_k))$ 的变换) 进行定义:

$$\mathbf{Z}_n = \mathbf{K}_n (\mathbf{Y}_n - \bar{\mathbf{Y}}_0) , \mathbf{K}_n = \mathbf{P}_n \mathbf{C}_n^{-1} , \quad (9)$$

其中

$$\mathbf{C}_n^{-1} = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & 0 & \cdots \\ 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & -1 & 1 \end{pmatrix}_{n \times n} , \quad (10)$$

并且其 (k, j) 元素 ($\mathbf{P}_n = (p_{kj}^{(n)})$) 定义为

$$p_{kj}^{(n)} = \sqrt{\frac{2}{n + \frac{1}{2}}} \cos \left[\frac{2\pi}{2n+1} (k - \frac{1}{2})(j - \frac{1}{2}) \right] \quad (11)$$

此时矩阵的谱变换 (参见 Kunitomo-Sato (2025) 第 2 章 Lemma A.1) 满足: $\mathbf{C}_n^{-1} \mathbf{C}'^{-1} = \mathbf{P}_n \mathbf{D}_n \mathbf{P}_n$, 其中 \mathbf{D}_n 为对角矩阵, 其第 k 个对角元为 $d_k = 2[1 - \cos(\pi(\frac{2k-1}{2n+1}))]$ ($k = 1, \dots, n$)。因此在下文中记作

$$a_{kn}^* (= d_k) = 4 \sin^2 \left[\frac{\pi}{2} \left(\frac{2k-1}{2n+1} \right) \right] \quad (k = 1, \dots, n) \quad (12)$$

在更一般的情况下, 对于式 (1) $\mathbf{y}_i = \mathbf{x}_i + \mathbf{v}_i$, 取状态变量的差分 $\Delta \mathbf{x}_i = \mathbf{v}_i^{(x)}$ 并假设噪声成分 \mathbf{v}_i ($i = 0, 1, \dots, n$) 及状态变量成分 $\mathbf{v}_i^{(x)}$ ($= \Delta \mathbf{x}_i$) 都是平稳随机过程, 可以表示为:

$$\mathbf{v}_i = \sum_{j=-\infty}^{\infty} \mathbf{C}_j^{(v)} \mathbf{e}_{i-j}^{(v)} \quad (13)$$

$$\mathbf{v}_i^{(x)} = \sum_{j=-\infty}^{\infty} \mathbf{C}_j^{(x)} \mathbf{e}_{i-j}^{(x)} . \quad (14)$$

其中, $\mathbf{e}_i^{(v)}$ 与 $\mathbf{e}_i^{(x)}$ 为 i.i.d. 随机变量序列, 且 $\mathbf{E}(\mathbf{e}_i^{(v)}) = \mathbf{E}(\mathbf{e}_i^{(x)}) = \mathbf{0}$, $\mathbf{E}(\mathbf{e}_i^{(v)} \mathbf{e}_i^{(v)'}) = \Sigma_e^{(v)}$ (半正定矩阵), $\mathbf{E}(\mathbf{e}_i^{(x)} \mathbf{e}_i^{(x)'}) = \Sigma_e^{(x)}$ (半正定矩阵)。为规范化, 设 $\mathbf{C}_0^{(v)} = \mathbf{C}_0^{(x)} = \mathbf{I}_p$, 并假定 $p \times p$ 系数矩阵 $\mathbf{C}_j^{(x)} (j = \dots, -1, 0, 1, \dots)$ 收敛, 即: $\sum_{j=-\infty}^{\infty} \|\mathbf{C}_j^{(v)}\| < \infty$, $\sum_{j=-\infty}^{\infty} \|\mathbf{C}_j^{(x)}\| < \infty$ 。其中 $\|\mathbf{C}_j^{(v)}\| = \max_{k,l=1,\dots,p} |c_{k,l}^{(v)}(j)|$ (矩阵 $\mathbf{C}_j^{(v)} = (c_{k,l}^{(v)}(j))$), $\|\mathbf{C}_j^{(x)}\| = \max_{k,l=1,\dots,p} |c_{k,l}^{(x)}(j)|$ (矩阵 $\mathbf{C}_j^{(x)} = (c_{k,l}^{(x)}(j))$)。

在本文中, 噪声项中也可以包含季节性要素。在这种情况下, $\mathbf{s}_i = \sum_{j=-\infty}^{\infty} \mathbf{C}_{sj}^{(s)} \mathbf{e}_{i-sj}^{(s)} (s \geq 2)$ 并且假设 $\mathbf{e}_i^{(s)}$ 为独立同分布 (i.i.d.) 的随机变量序列, 满足 $\mathbf{E}(\mathbf{e}_i^{(s)}) = \mathbf{0}$, $\mathbf{E}(\mathbf{e}_i^{(s)} \mathbf{e}_i^{(s)'}) = \Sigma_e^{(s)}$ (其中, 系数矩阵 $\mathbf{C}_{sj}^{(s)}$ 的绝对值的和收敛, 即 $\sum_{j=-\infty}^{\infty} \|\mathbf{C}_j^{(s)}\| < \infty$)。一般而言, 也可以将非平稳的季节性成分纳入模型, 但本文对此不作讨论。

本文所介绍的 SIML 方法, 是一种通过变换将非平稳时间序列从时间域转换到频率域的方法。其核心思想是: 通过在频率域中对信号进行分解, 可以将时间序列的组成要素——趋势 (trend)、循环 (cycle)、季节性 (seasonality) 以及噪声 (noise) ——明确地区分开来, 并通过滤波 (filtering) 从观测序列中提取出所需的状态变量。在所使用的 \mathbf{K}_n - 变换中, \mathbf{P}_n 变换是一种取实数值的离散傅里叶变换 (Discrete Fourier Transform, DFT) 形式。结果上, 变量向量 $\mathbf{z}_k (k = 1, \dots, n)$ (属于 \mathbf{Z}_n) 已知在渐近意义下彼此无相关。(可参见 Anderson (1971) 第 7 章, Kunitomo and Sato (2025) 等文献。)

接下来介绍基于这种变换的两种状态变量滤波方法。

所谓前向滤波 (forward filtering), 是指基于 \mathbf{K}_n - 变换所得到的实值 (近似) 正交过程, 通过逆变换提取部分频率成分的方法。

设 $n \times p$ 矩阵为

$$\hat{\mathbf{X}}_n(\mathbf{Q}) = \mathbf{C}_n \mathbf{P}_n \mathbf{Q}_n \mathbf{P}_n \mathbf{C}_n^{-1} (\mathbf{Y}_n - \bar{\mathbf{Y}}_0) \quad (15)$$

以及

$$\mathbf{Z}_n = \mathbf{P}_n \mathbf{C}_n^{-1} (\mathbf{Y}_n - \bar{\mathbf{Y}}_0), \mathbf{Y}_n = \bar{\mathbf{Y}}_0 + \mathbf{X}_n^{(0)} + \mathbf{V}_n, \quad (16)$$

其中, 初始条件为 $\mathbf{X}_n^{(0)} = (\mathbf{x}_i^{(0)'}), \mathbf{V}_n = (\mathbf{v}_i')$ 为 $n \times p$ 矩阵, 并设 $\mathbf{x}_i^{(0)} = \mathbf{x}_i - \mathbf{x}_0 (i = 1, \dots, n)$ 。将噪声的初值取为零向量, 即 $\mathbf{y}_0 = \mathbf{x}_0$ 。

在这里, 随机过程 \mathbf{Z}_n 是原始时间序列 \mathbf{Y}_n 在频率域中的正交过程, \mathbf{Q}_n 是 $n \times n$ 的滤波矩阵。由于观测矩阵 \mathbf{Y}_n 是非平稳时间序列, 因此变换 \mathbf{K}_n (见式 (9)) 具有特殊结构。为了平滑 (或滤波) 趋势与循环成分, 这里特别加以说明。(关于其他变换的示例, 可参见国友・櫻井・佐藤 (2022)。)

利用 $m \times n$ 的选择矩阵 $\mathbf{J}_m = (\mathbf{I}_m, \mathbf{O})$, 考虑如下 $n \times p$ 的矩阵:

$$\hat{\mathbf{X}}_n(m) = \mathbf{C}_n \mathbf{P}_n \mathbf{Q}_n^{(m)} \mathbf{P}_n \mathbf{C}_n^{-1} (\mathbf{Y}_n - \bar{\mathbf{Y}}_0) . \quad (17)$$

其中 $n \times n$ 的矩阵 $\mathbf{Q}_n^{(m)}$ 满足 $\mathbf{Q}_n^{(m)} = \mathbf{J}'_m \mathbf{J}_m$ 。

这里, 对于 $n \times p$ 的潜在变量矩阵 \mathbf{X}_n , 只通过对 \mathbf{Z}_n 的逆变换提取低频成分, 并忽略高频成分。由此得到的趋势与循环成分可表示为

$$\mathbf{X}_n(m) = \mathbf{C}_n \mathbf{P}_n \mathbf{Q}_n^{(m)} \mathbf{P}_n \mathbf{C}_n^{-1} \mathbf{X}_n^{(0)} . \quad (18)$$

该状态估计量虽与原始的 \mathbf{X}_n 不同, 但式 (15) 与 (16) 中的状态变量 \mathbf{x}_i ($i = 1, \dots, n$) 不仅包含观测向量 \mathbf{y}_i ($i = 1, \dots, n$) 的趋势与循环成分, 还包含观测误差成分。因此, 将数据转换至频率域, 只提取趋势与循环成分的部分, 即为 SIML 方法的基本思想。趋势与循环成分对应于接近零频率的成分, 而观测误差成分经过变换 $(\mathbf{v}_i - \mathbf{v}_{i-1})$ 后, 其影响预计会变得极小、几乎可以忽略。

需要注意的是, SIML 方法虽然与传统的频率域分析或时间域分析有关, 但其视角略有不同。关于通过此处介绍的变换与逆变换对观测向量 \mathbf{y}_i 的数据分析方法进行统计上的正当化讨论, 可参见 Kunitomo and Sato (2025), 其中指出: 该方法可以通过对差分序列 $\mathbf{r}_i^{(n)} = \Delta \mathbf{y}_i$ ($= \mathbf{y}_i - \mathbf{y}_{i-1}$) 的频率分析来实现。

接下来介绍基于后向滤波 (backward-filtering) 的方法。

设 $n \times p$ 矩阵 $\mathbf{Y}_n^* = (\mathbf{y}'_{i-1})$, 并定义 $np \times 1$ 随机向量 $(\mathbf{y}'_0, \dots, \mathbf{y}'_{n-1})'$ ¹⁰。当初始条件为 \mathbf{y}_n 时, $np \times 1$ 随机向量 $(\mathbf{y}'_0, \dots, \mathbf{y}'_{n-1})' - \mathbf{1}_n \cdot \mathbf{y}'_n$ ($\mathbf{1}'_n = (1, \dots, 1)$) 的协方差矩阵为

$$\text{Cov}[\text{vec}(\mathbf{Y}_n)] = \mathbf{I}_n \otimes \boldsymbol{\Sigma}_v + \mathbf{C}_n \mathbf{C}'_n \otimes \boldsymbol{\Sigma}_x , \quad (19)$$

其中 \mathbf{C}_n 由式 (8) 给出。

现在, 将 \mathbf{K}_n^* - 变换应用于 \mathbf{Y}_n^* 到 \mathbf{Z}_n^* ($= (\mathbf{z}'_k')$) 的转换:

$$\mathbf{Z}_n^* = \mathbf{K}_n^* (\mathbf{Y}_n^* - \bar{\mathbf{Y}}_n^*) , \mathbf{K}_n^* = \mathbf{P}_n^* \mathbf{C}_n'^{-1} , \quad (20)$$

¹⁰当初始条件取为 \mathbf{y}_n 时, 需考虑随机向量 $(\mathbf{y}'_{n-1}, \dots, \mathbf{y}'_0)'$ 的联合分布, 其中 \mathbf{y}_i ($i = 0, 1, \dots, n$)。

其中 $\bar{\mathbf{Y}}_n^* = \mathbf{1}_n \mathbf{y}'_n$,

$$\mathbf{C}'^{-1} = \begin{pmatrix} 1 & -1 & \cdots & 0 & 0 \\ 0 & 1 & -1 & \cdots & 0 \\ 0 & 0 & 1 & -1 & \cdots \\ 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}_{n \times n}, \quad (21)$$

并令其 (k, j) -元素 $(\mathbf{P}_n^* = (p_{kj}^{(n)}))$ 定义为:

$$p_{kj}^{*(n)} = \sqrt{\frac{2}{n + \frac{1}{2}}} \sin \left[\frac{2\pi}{2n+1} \left(k - \frac{1}{2} \right) j \right]. \quad (22)$$

此时, 根据谱分解, 有: $\mathbf{C}'^{-1} \mathbf{C}_n^{-1} = \mathbf{P}_n^{*\prime} \mathbf{D}_n \mathbf{P}_n^*$ 其中 \mathbf{D}_n 的对角元素为: $d_k = 2[1 - \cos(\pi(\frac{2k-1}{2n+1}))]$ ($k = 1, \dots, n$)。 (参见 Sato and Kunitomo (2024) 的引理 A-1。)

上述两种变换的特征值相同: $a_{kn}^* (= d_k) = 4 \sin^2 \left[\frac{\pi}{2} \left(\frac{2k-1}{2n+1} \right) \right]$ ($k = 1, \dots, n$)

接下来, 通过对变换后的随机过程进行逆变换, 可构造 $n \times p$ 矩阵:

$$\hat{\mathbf{X}}_n^*(\mathbf{Q}_n) = \mathbf{C}' \mathbf{P}_n^{*\prime} \mathbf{Q}_n \mathbf{P}_n^* \mathbf{C}'^{-1} (\mathbf{Y}_n^* - \bar{\mathbf{Y}}_n^*) \quad (23)$$

以及

$$\mathbf{Z}_n^* = \mathbf{P}_n^* \mathbf{C}'^{-1} (\mathbf{Y}_n^* - \bar{\mathbf{Y}}_n^*), \quad \mathbf{Y}_n^* = \bar{\mathbf{Y}}_n^* + \mathbf{X}_n^* + \mathbf{V}_n^*. \quad (24)$$

其中 $\mathbf{X}_n^* = (\mathbf{x}_{i-1}^{*'})$ 、 $\mathbf{V}_n^* = (\mathbf{v}_{i-1}^{*'})$ 为 $n \times p$ 矩阵, 并设 $\mathbf{x}_{i-1}^* = \mathbf{x}_{i-1} - \mathbf{x}_n$ ($i = 1, \dots, n$)。由原过程 \mathbf{Y}_n^* 变换得到的随机过程 \mathbf{Z}_n^* , 对应于在频率域中的正交分解, 其中 \mathbf{Q}_n ($n \times n$) 为滤波矩阵。由于原随机过程 \mathbf{Y}_n^* 是非平稳时间序列, 因此采用了特殊形式的变换 \mathbf{K}_n^* 。例如, 对于趋势变换, 可取 $n \times p$ 矩阵:

$$\hat{\mathbf{X}}_n^*(m) = \mathbf{C}' \mathbf{P}_n^{*\prime} \mathbf{Q}_n^{(m)} \mathbf{P}_n^* \mathbf{C}'^{-1} (\mathbf{Y}_n^* - \bar{\mathbf{Y}}_n^*) \quad (25)$$

并设 $\mathbf{Q}_n^{(m)} = \mathbf{J}'_m \mathbf{J}_m$ 。

在对潜在变量 $n \times p$ 矩阵 \mathbf{X}_n^* 的估计中, 由于仅使用了 \mathbf{Z}_n^* 的低频成分, 因此在本记号下, 所估计的状态变量为:

$$\mathbf{X}_n^*(m) = \mathbf{C}' \mathbf{P}_n^{*\prime} \mathbf{Q}_n^{(m)} \mathbf{P}_n^* \mathbf{C}'^{-1} \mathbf{X}_n^{*(0)}, \quad (26)$$

其中 $\mathbf{X}_n^{*(0)} = (\mathbf{x}_i^{*(0)'})$ 为潜在变量矩阵, $\mathbf{x}_i^{*(0)} = \mathbf{x}_i^* - \mathbf{x}_n^*$ ($i = 0, \dots, n-1$)。该矩阵对应于潜在变量的前 m 个低频成分。

4. 相关问题

(i) **趋势与循环成分的选择**：许多经济学家关注的并不是季节成分的变化，而是从原始序列中剔除季节性影响后得到的季节调整序列的波动。在此需要注意的是，在美国人口普查局开发的季节调整方法（如 X-13ARIMA-SEATS）中，季节调整后的序列 $y_t^* (t = 1, \dots, T)$ 是通过从原始序列 y_t 中减去估计的季节成分 \hat{S}_t 得到的：

$$y_t^* = y_t - \hat{S}_t \quad (t = 1, \dots, T) \quad (27)$$

¹¹ 因此，季节调整后的序列中仍然包含不规则波动成分。当观测数据中存在缺失值时，X-12-ARIMA 或 X-13ARIMA-SEATS 等方法中的季节成分 S_t 的估计方法会变得复杂，目前尚不明确。

相比之下，DECOMP 方法在假设正态性等条件下，可以相对简单地处理这一问题，正如北川（2020）所说明的那样。

另一方面，如果研究者并不关注短期或暂时性的波动，则应当聚焦于趋势与循环成分的变化，这在 SIML-filtering 方法中对应于参数 m 的选择问题。一般而言，将 m 设定为对应 1 年以上周期的成分是较为合理的做法。SIML 方法还能结合较大波动或变动点分析（参见 Sato-Kunitomo (2024)、Kunitomo-Sato (2025)）。

(ii) **SIML 季节调整值**：vspace.1cm

国友・樱井・佐藤（2022）提出了一种称为 Type-II 平滑（Type-II smoothing）的方法。利用这种方法，在存在缺失值的情况下，无需依赖特定分布即可较为容易地构造季节调整值。

这里按照国友・樱井・佐藤（2022）的记号作部分调整，设总观测数为 (n)，若数据为 L 年期，则 $n = 4L$ あるいは（季度数据）或 $n = 12L$ （月度数据）。当将季节性状态估计纳入时，需要更复杂的变换。定义一个 $q \times n$ 矩阵 \mathbf{F}_q ($0 < q < n$)，如下式构造出经变换的季节频率成分矩阵：

$$\mathbf{S}_s = \mathbf{F}_q \mathbf{P}_n \mathbf{C}_n^{-1} (\mathbf{Y}_n - \bar{\mathbf{Y}}_0) \quad (28)$$

其中 \mathbf{F}_q 对应不同数据类型：季度数据用 \mathbf{F}_q^Q ，月度数据用 \mathbf{F}_q^M 。定义

$$\mathbf{Q}_n = \mathbf{I}_n - \mathbf{F}_q' \mathbf{F}_q \quad (29)$$

¹¹ 关于 X-12-ARIMA、X-13ARIMA-SEATS 方法的说明，可参考 <https://stat-expert.ism.ac.jp/wp/wp-content/uploads/2023/03/SSE-DP-2023-1.pdf>。

若 $\mathbf{F}_q \mathbf{F}'_q = \mathbf{I}_q$, 则 \mathbf{Q}_n 为投影算子。此时, 去除季节性的部分可以通过下式估计:

$$\hat{\mathbf{X}}_{Tn*} = \mathbf{C}_n \mathbf{P}_n [\mathbf{I}_n - \mathbf{F}'_q \mathbf{F}_q] \mathbf{P}_n \mathbf{C}_n^{-1} (\mathbf{Y}_n - \bar{\mathbf{Y}}_0)。$$

因此观测序列可分解为: $\mathbf{Y}_n = \mathbf{S}_s + \hat{\mathbf{X}}_n^*$ 。

对于季度数据的情况, $s = 4$ 对应频率 $\lambda_s = 1/4, 1/2$, 其中 $1/2$ 表示 2 季周期, $1/4$ 表示 4 季周期。但在季度数据中区分 12 个月与 6 个月周期的波动是困难的。因此, 设 $m_1 = [2L/s]$, 构造以下 $(n - 2h - 1) \times n$ 和 $(n - 3h - 2) \times (n - 2h - 1)$ 的选择矩阵:

$$\mathbf{J}_1^Q = \begin{bmatrix} \mathbf{I}_{m_1-(h+1)} & \mathbf{O} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} & \mathbf{I}_{n-m_1-h} \end{bmatrix}, \quad \mathbf{J}_2^Q = [\mathbf{I}_{n-3h-2}, \mathbf{O}], \quad (30)$$

并由此构成 $q \times n$ 矩阵:

$$\mathbf{F}_q^Q = \mathbf{J}_2^Q \mathbf{J}_1^Q \quad (31)$$

其中 $h > 0$ 为小的正整数, $q = n - 3h - 2$ 。对于月度数据 ($s = 12$), 区分 12 的周期波动和 6、4、3、2.4、2 个月周期的波动很困难。因此, 设 $m_l = l[2L/s]$, 构造以下 $(n - l(2h + 1)) \times (n - (l - 1)(2h + 1))$ ($(l = 1, \dots, 5)$) 的矩阵以及 $(n - 5(2h + 1) - (h + 1)) \times (n - 5(2h + 1))$ 的矩阵:

$$\mathbf{J}_l^M = \begin{bmatrix} \mathbf{I}_{m_l-(l-1)(2h+1)-(h+1)} & \mathbf{O} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} & \mathbf{I}_{n-m_l-h} \end{bmatrix}, \quad \mathbf{J}_6^M = [\mathbf{I}_{n-11h-6}, \mathbf{O}], \quad (32)$$

并定义:

$$\mathbf{F}_q^M = \prod_{j=1}^6 \mathbf{J}_{7-j}^M \quad (33)$$

此时 $q = T - 11h - 6$ 。

在此季节调整值的构造中, 未使用任何特定分布假设或移动平均操作, 实际上是一种几乎不依赖于真实成分的非参数方法¹²。

(iii) **未来值的预测**: 从经济判断的角度来看, 当非平稳季节性时间序列中存在缺失值时, 研究未来值的预测方法是必要的。此时, 需要构造去除季节性后的趋势与循环成分的未来值, 或其增长率的预测值。一种简单的方法是: 用 ARIMA 等模型拟合“(过去的数据) – (不规则成分) – (季节成分)”部分, 从而进行趋势预测。

¹²实际计算可使用佐藤整尚先生开发的 x12siml 程序实现。季节周期的带宽由参数 sorder 选择。详细文档参见 <http://www.kunitomo-lab.sakura.ne.jp/x12simldoc92.pdf>

5. 应用实例

本文以中国的宏观零售销售数据为例，进行了统计分析。该数据涉及诸如过去市场的快速扩张、新冠疫情的影响、以及其后增长率放缓等宏观经济话题。在月度数据中，存在一种特殊的缺失结构：自某时点起，1月与2月的单月数据缺失，但1-2月的合计数据可用。

在此，我们作为新计算程序开发的前期准备，尝试比较几种不同的缺失值代入（补全）方法。由于无法直接使用1月与2月的数据，但可以利用两者的合计值，因此可以考虑多种代入策略。

首先，在不使用虚拟变量（dummy variables）的情况下，比较不同代入方法下的AIC（赤池信息量准则），结果表明：采用 $y_{1月} = y_{12月}$, $y_{2月} = y_{3月}$ 的代入方式，AIC最小。这意味着，中国零售销售额在12月与1月之间具有强烈的联动关系，反映了时间序列的内在季节性结构。然而，当存在合计数据 y_{12} 时，对于季节调整后的序列，并不一定保证 $\hat{y}_{1月} + \hat{y}_{2月} = \hat{y}_{12}$ 成立。例如，当合计数据 $y_{1月} + y_{2月} = y_{12}$ 已公布时，可以设定： $\hat{y}_{1月} = w_1 y_{12月} + w_2 y_{1月} + w_3 y_{3月}$ ($w_k \geq 0$), $\hat{y}_{2月} = \hat{y}_{12} - \hat{y}_{1月}$ ，此时存在多种可能。若假设 $\sum_{k=1}^3 w_k = 1$ ，则该模型中自由参数为两个。

进一步地，利用国友·樱井·佐藤（2022）提出的程序来估计趋势与循环部分后发现，不同代入法之间的差异相当小。这表明，即使部分月度数据缺失，也有可能构建出在中国宏观经济分析中依然有效的月度时序数据。

从原理上讲，在使用北川（2020）的DECOMP方法时，可以在一定约束条件下进行极大似然估计。但最终重要的是趋势与循环部分的估计值。因此，有必要分析季节性与不规则波动在结果中的相对重要性。

作为示例，下面展示若干滤波计算结果。图2显示了对1990年至2023年原始序列进行缺失值填补后得到的时间序列。从图中可以明显看出，在过去30年间，零售业销售呈现出趋势性增长，同时季节性波动的幅度也不断扩大，即方差非均一性显著。在DECOMP方法中，估计季节性时对这种方差不均一性的标准化处理至关重要。因此，我们对原始序列进行了对数变换，设趋势次数为2、季节性次数为1，并使用RS-DECOMP¹³进行季节调整。结果如图3所示。通过对数变换后，虽然估

¹³ 可在线访问统计数理研究所网站。使用示例及说明参见：<https://stat-expert.ism.ac.jp/wp/wp-content/uploads/2023/03/SSE-DP-2023-1.pdf>

计的季节性成分随时间略有减小，但也可以考虑使用 Box-Cox 变换。值得注意的是，估计出的季节性自 2010 年前后开始出现变化，这很可能与此后中国官方停止公布 1-2 月分月数据有关。此外，当采用不同的月度数据插补方法时，估计出的季节性模式也会发生变化，但这属于由数据特征所决定的自然结果。

最后，展示基于 S-SIML 季节调整法¹⁴ 的分析结果。我们采用了两种插补方案：

- 方案 1：1 月数据 = 前年 12 月数据，2 月数据 = 3 月数据；
- 方案 2：1 月数据 = 2 月数据 = (1-2 月合计) /2。

使用佐藤 SIML 季节调整法对补全后的数据进行分析后发现：季节性在频率成分中得到了良好估计，而趋势与循环成分的估计值对插补方法的差异并不敏感。这表明，通过补全后的月度时间序列，可以构造去除季节周期的季节调整序列，且其估计结果与 RS-DECOMP 方法所得结论相似。

References

- [1] Kitagawa, G. (2020), *Introduction to Time Series Modeling with Applications in R*, Second Edition, CRC Press. (北川源四郎「R による時系列モデリング入門」岩波書店。)
- [2] Kunitomo, N. and S. Sato (2021), “A Robust-filtering Method for Noisy Non-Stationary Multivariate Time Series with Econometric Applications,” Japanese Journal of Statistics and Data Science, 4, 373-410. <https://doi.org/10.1007/s42081-020-00102-y>.
- [3] Kunitomo, N. and S. Sato (2023), “Frequency Regression and Smoothing for Noisy Nonstationary Multivariate Time Series,” SSE-DP-2023-2, <https://stat-expert.ism.ac.jp/wp/wp-content/uploads/2023/03/SSE-DP-2023-2.pdf>
- [4] Kunitomo, N. and S. Sato (2025), *The SIML Filtering Method for Noisy Non-stationary Economic Time Series*, Springer. forthcoming.

<https://jasp.ism.ac.jp/RS-Decomp/>

¹⁴由佐藤・国友（2024）提出的季节调整方法。

- [5] Sato, S. and Kunitomo, N. (2024) “Forward and Backward Smoothing for Noisy Nonstationary Time Series,” SSE-DP-2024-5, <https://stat-expert.ism.ac.jp/wp/wp-content/uploads/2023/03/SSE-DP-2023-2.pdf>
- [6] 国友直人・櫻井智章・佐藤整尚 (2022), “経済時系列の状態推定とマクロ指標,” 統計研究彙報, 79, 1-20, 統計局.
- [7] 佐藤整尚・国友直人 (2024), “SarSIML 季節調整,” 未定稿.
- [8] Anderson, T.W. (1971), *The Statistical Time Series Analysis*, Wiley.

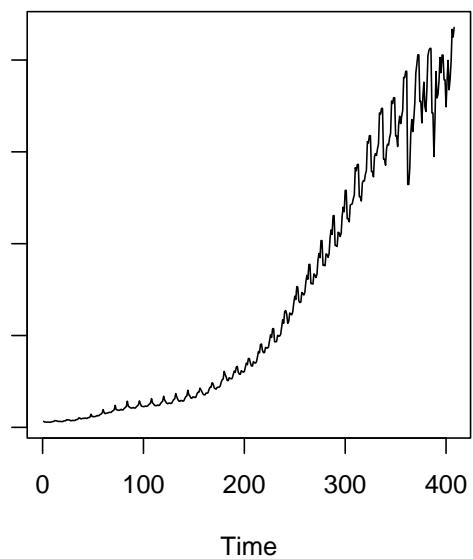


图 2：插补后的零售业月度序列（1990 年–2023 年，原始序列）

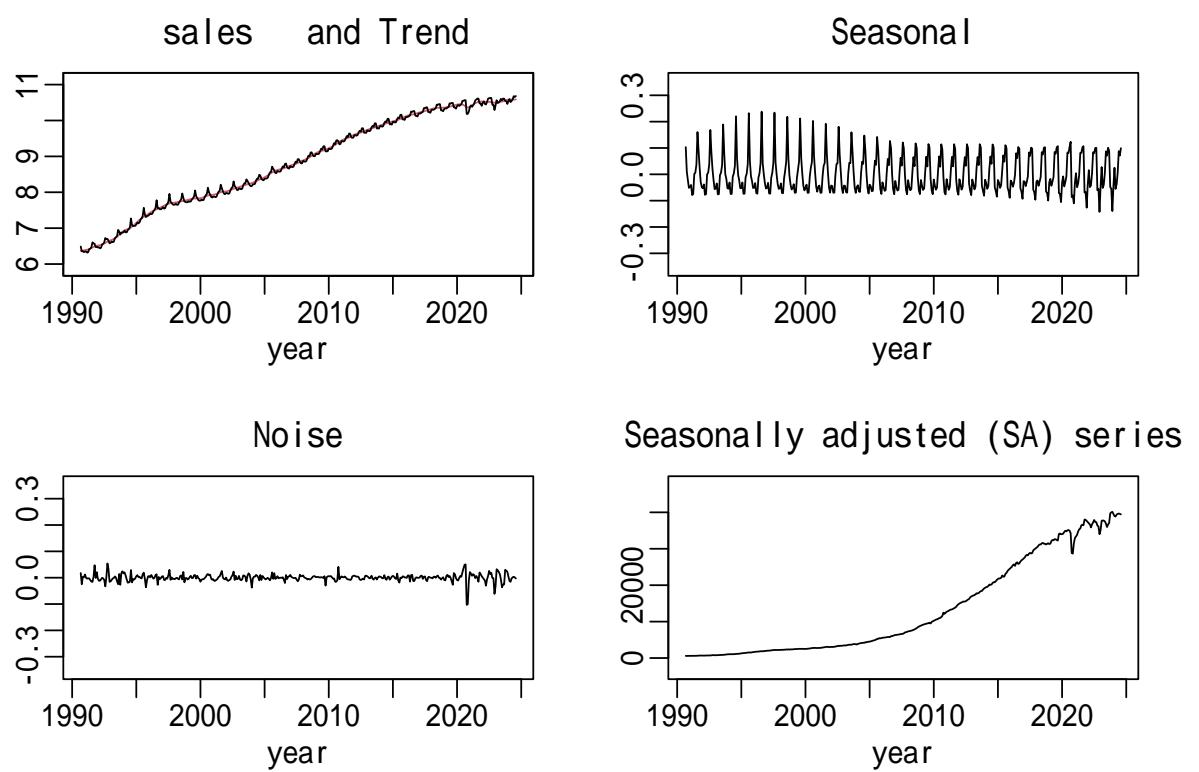


图 3: 基于 RS-DECOMP 的成分分解 (对数变换)

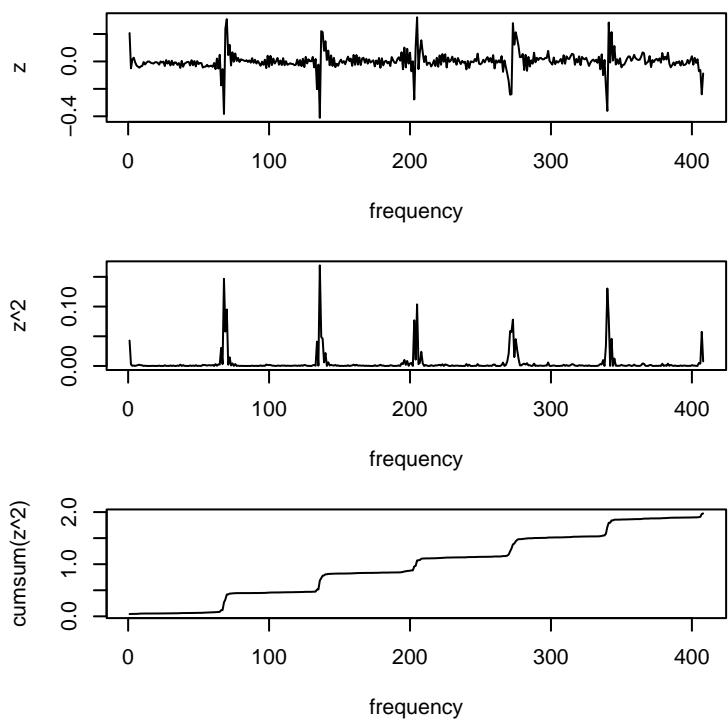


图 4：基于 SIML 的成分分解（原始序列）

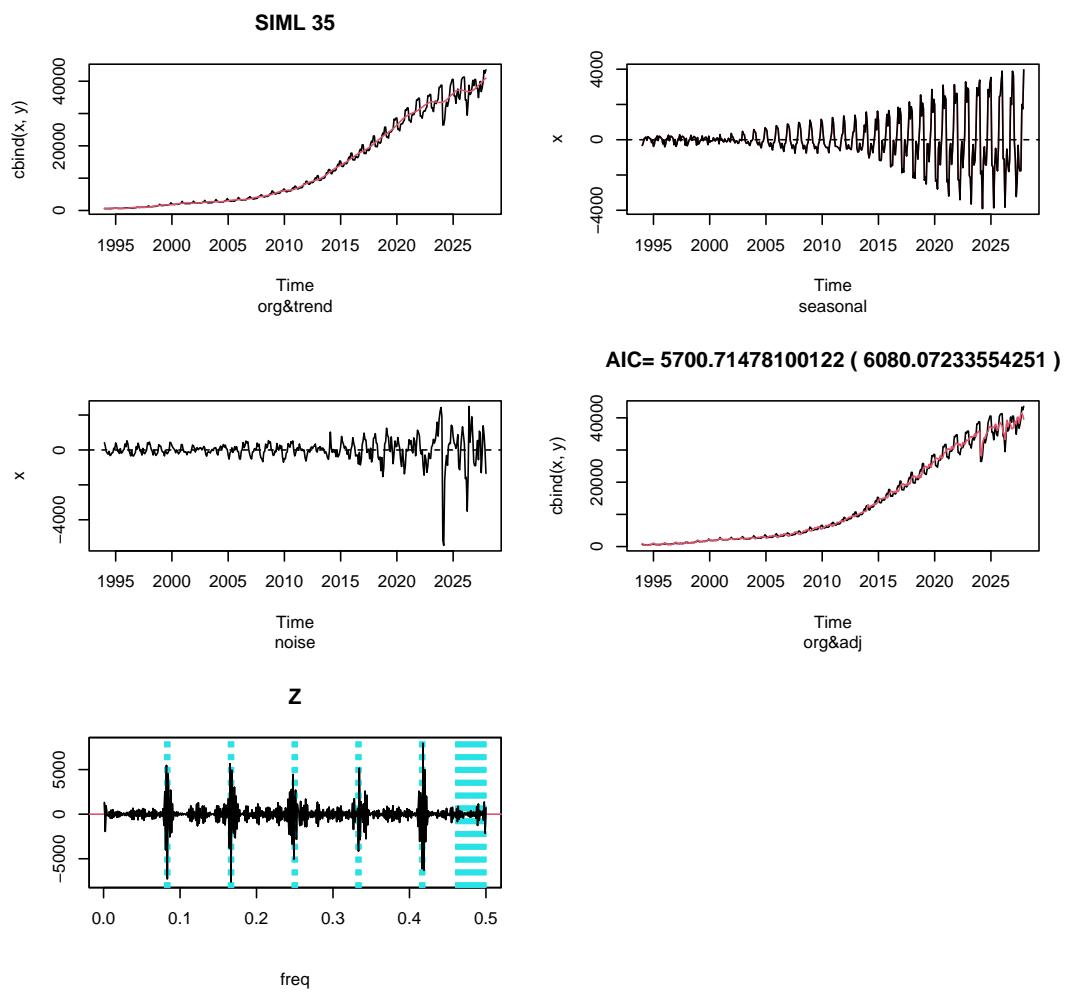


图 5: S-SIML 季节调整 1 (1 月 =12 月, 2 月 =3 月, 原始序列)

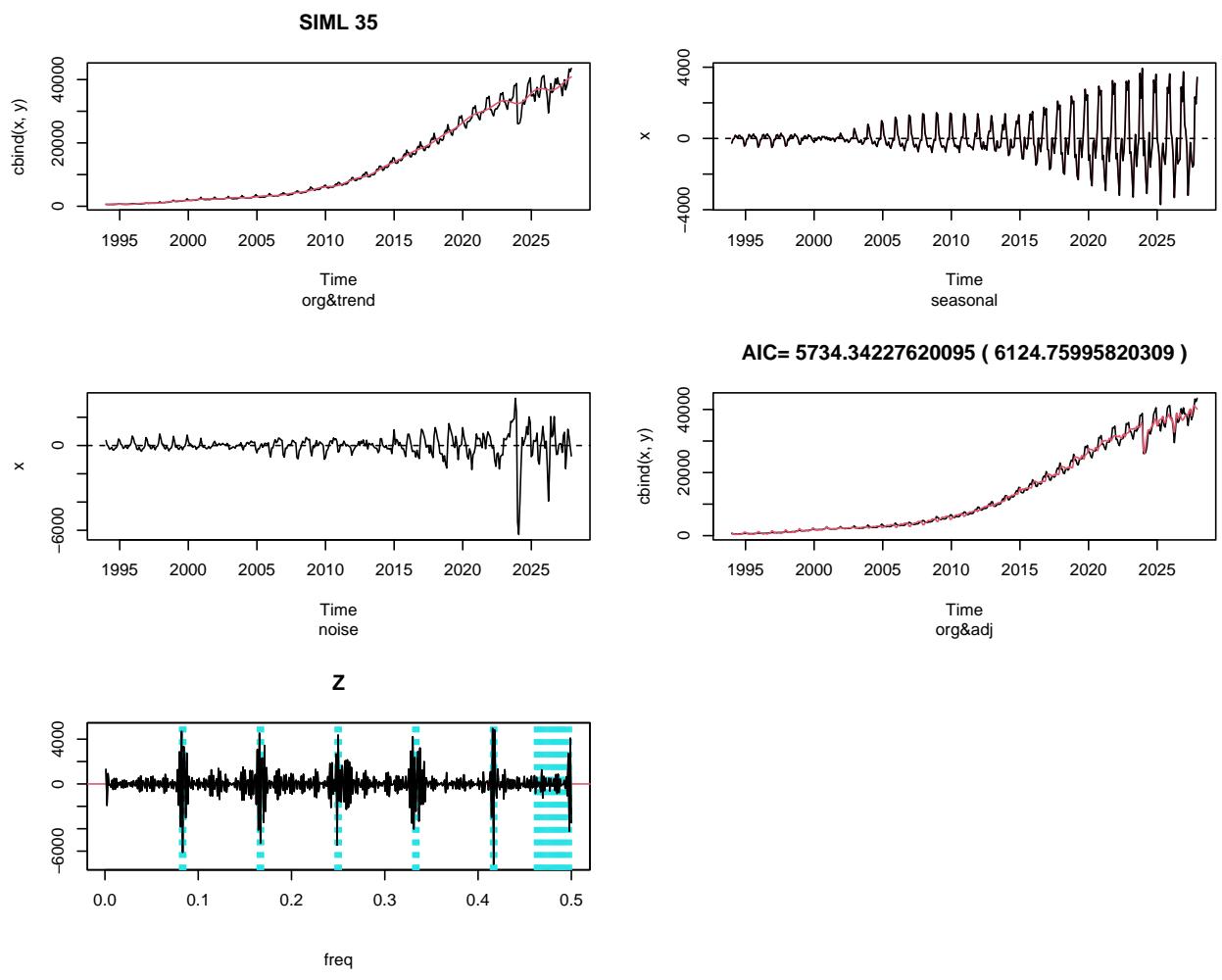


图 6: S-SIML 季节调整 2 (average, 原始序列)

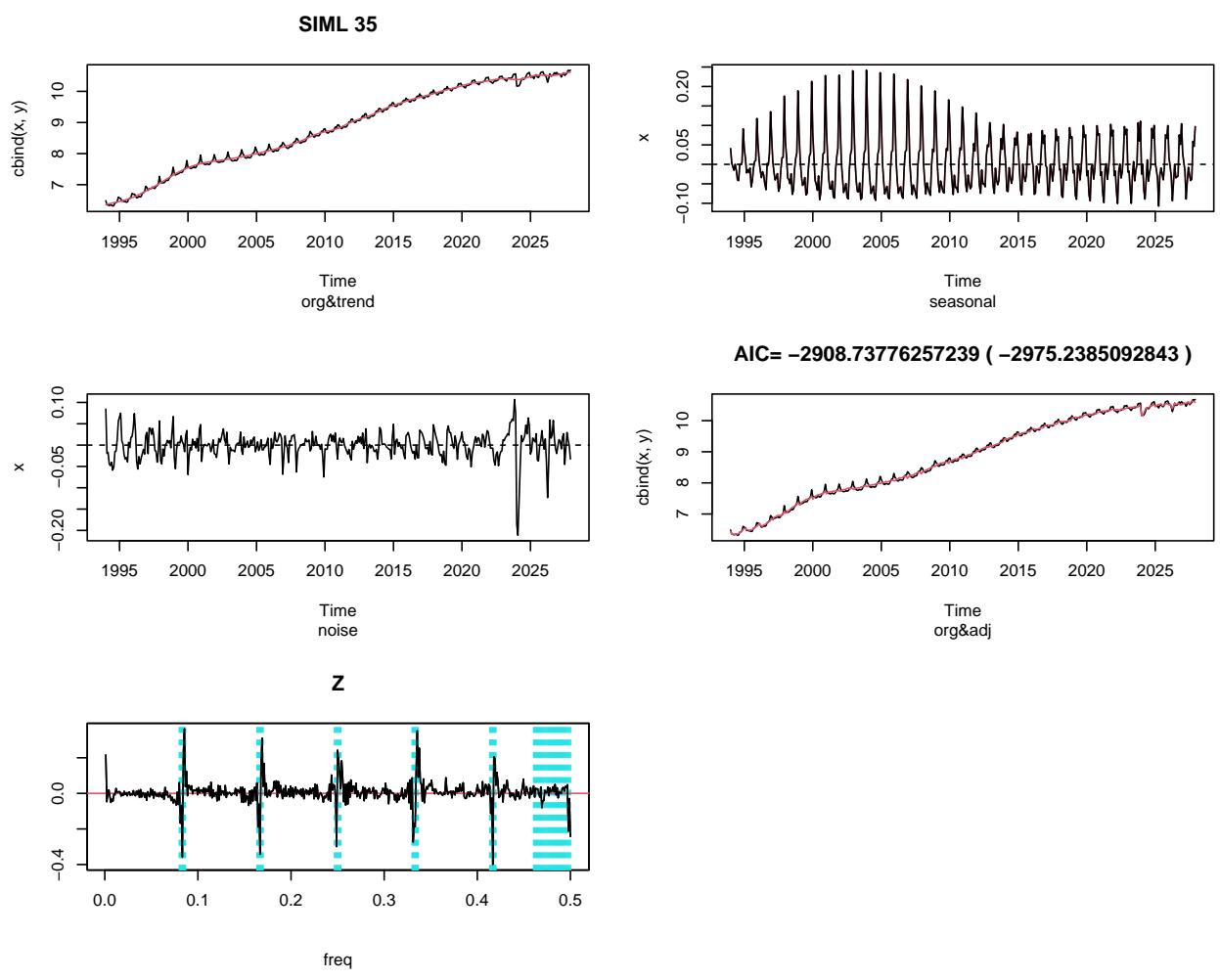


图 7: S-SIML 季节调整 2 (average, 对数变换序列)

SarSIML 季节调整 2024-10-17

2024-10-19(修订)

佐藤整尚¹⁵

国友直人¹⁶

1. 前言

随着时间的推移，人们已经能够观测到大量的经济时间序列。特别是对于主要的宏观经济官方数据而言，由于数据收集和编制的原因，这些数据以不同的频率和不同的时间节点被测量，包括日度序列、月度序列、季度序列以及年度序列等，并由政府的各个部门分别公布。像消费、投资、政府支出、进出口等主要的宏观经济时间序列，都是基于各自的基础统计调查而编制的，因此每个序列都有其特有的性质，并不一定经过相互协调后才对外发布。然而，从理解经济整体动向、进行数据分析并据此进行政策判断的角度来看，这种分散的编制与发布体系并不是理想的状况。对于反映宏观经济整体趋势的时间序列来说，为了尽早掌握最新的经济状况，应尽快编制并公布准确的数据。但在编制国家层面的统计数据时，必须先收集大量基础调查数据，这往往需要较长时间。例如，日本的GDP及其主要分项指标，其基础数据会陆续公布，因此存在时间差。因此，GDP通常经历一次速报（初步统计）、二次速报（修正版）和确报（最终版）等多个阶段。在投资等项目中，由于在GDP一次速报公布后才获得更精确的基础数据，结果导致季度数据的初次公布值会在一段时间后被修订。这种因数据修订所产生的误差，可视为在真实状态变量估计过程中出现的观测误差。但在实践中，人们也常常发现：最新公布的季度数据，有时与用于编制它的月度基础数据在表面上存在矛盾。此外，在日本的官方统计体系中，各时间序列的调查由不同的部门负责，因此数据公布的责任主体各不相同，这也使得问题更加复杂。尤其是在包括日本在内的许多国家中，官方统计普遍使用基于移动平均的季节调整方法，如X-12-ARIMA、X-13ARIMA-SEATS。这些季节调整程序的内部计算中，采用了一种单变量的时间序列回归模型Reg-ARIMA模型。在日本的实际操作中，各政府部门对所负责的每个统计序列分别独立地应用并运行Reg-ARIMA模型。因此，即使在宏观时间序列之间存在多种经济关系，由于不同系列包含不同的

¹⁵东京大学

¹⁶统计数理研究所

构成要素，需要进行成分分解（趋势、季节性、不规则成分等），但多个时间序列之间并不一定实现一致的季节调整。

本稿作为改善这种现状的一项尝试，从统计时间序列分析的角度出发，介绍一种基于 SIML 滤波法（Spectral Inference via Matrix Linearization）的季节调整方法，即由佐藤开发的 S-SIML 季节调整法（S-SIML seasonal adjustment），并说明其在 R 程序中的实现方式。SIML 滤波法是由 Kunitomo–Sato (2021)、Kunitomo–Sato (2025) 等人提出并发展的统计时间序列分析方法，可用于估计趋势–循环成分、季节成分、不规则成分，甚至包括变动点（结构性突变）等状态变量。其日文版介绍可参见国友・櫻井・佐藤 (2022)。在 SIML 方法中，通过频率分解（spectral decomposition），可以直接从具有非平稳性与季节性的季度或月度时间序列中，获得趋势–循环成分、季节成分以及季节调整后的序列等估计结果。

2. S-SIML 季节调整程序的使用方法

我们开发了 S-SIML 季节调整的 R 程序“sarsiml”，下面说明其使用方法。该程序可在免费软件 R 上运行。它是由佐藤开发的更大型 R 程序 x12siml（参见 Sato, 2023）的简化版。x12siml 可用于分析具有季节性与非平稳性的时间序列经济数据。¹⁷

(i) 将程序文件放在 R 的启动目录下，在 R 控制台中输入：

```
source("sarsiml_ver1.1.R")
```

即可完成安装。

(也可以直接读取 R 源文件运行。)

(ii) 若为季度数据：

```
sarsiml(quarterlydataname)
```

若为月度数据：

```
sarsimll(monthlydataname, frequency=12)
```

运行后会输出 5 张图表（原始序列与趋势–循环成分（图 1），季节成分（图 2），不规则成分（图 3），原始序列与季节调整值（图 4），以及 Z 变换（图 5））。

(iii) 数据可通过 R 命令读取后使用。

(iv) 趋势参数 m 与季节波动参数 (sorder) 默认自动选择（‘sorder=1’， m 对应 2 年周期）。根据分析目的，也可参考国友・櫻井・佐藤 (2022) 或 x12siml 手册手动设定。

¹⁷<https://github.com/sato-labo/x12siml>

由于非平稳经济时间序列包含趋势、循环、季节与不规则成分等多种特征，应灵活调整参数。例如：

```
x12siml(monthlydataname, frequency=12, trend=10,sorder=2)
```

若设 ‘sorder=0‘，则不估计季节成分。部分时间序列在默认设定下可能无法充分去除季节性，此时可调整 ‘sorder‘，或添加选项：

```
smooth=T
```

以消除在频率 $\lambda = 1/2$ 附近更广范围内季节成分。

(v) 提取估计结果时，可先将结果赋值后分别提取各成分：
res <- x12siml(shouhi)

```
res$trend
```

```
res$adj
```

```
res$Z
```

```
res$seasonal
```

```
res$noise
```

这些对象均可再利用。

(vi) 图中显示的 AIC 是基于估计噪声的伪 AIC (pseudo AIC)，由假设噪声服从高斯分布的似然计算得到。其中一个 AIC 基于数据总数 n ，另一个基于趋势-循环成分的样本数 m 。

3. 实例分析

以下以日本内阁府发布的季度实际最终消费序列（1994–2019）** 为例，其分析结果汇总于图 1。左上图：黑线为原始序列，红线为趋势-循环成分。右上图：估计得到的季节成分。左中图：估计的不规则（噪声）成分。进入 2000 年代后，东日本大地震与雷曼冲击等重大事件的宏观影响主要反映在从原序列中减去估计季节性后得到的不规则波动中。右中图：原序列与季节调整序列。请注意，季节调整序列是通过从原序列中减去估计的季节成分得到的。

与美国人口普查局的 X-12-ARIMA、X-13ARIMA-SEATS 不同，S-SIML 直接从频率成分估计季节性，因此在求季节调整序列时不需要 Reg-ARIMA 模型。左下图为通过 K_{n-} 变换得到的 Z 序列。在具有季节性的季度宏观经济时间序列中，季节性主要出现在频率 $\lambda = 0.25, 0.5$ 附近。

SIML 17

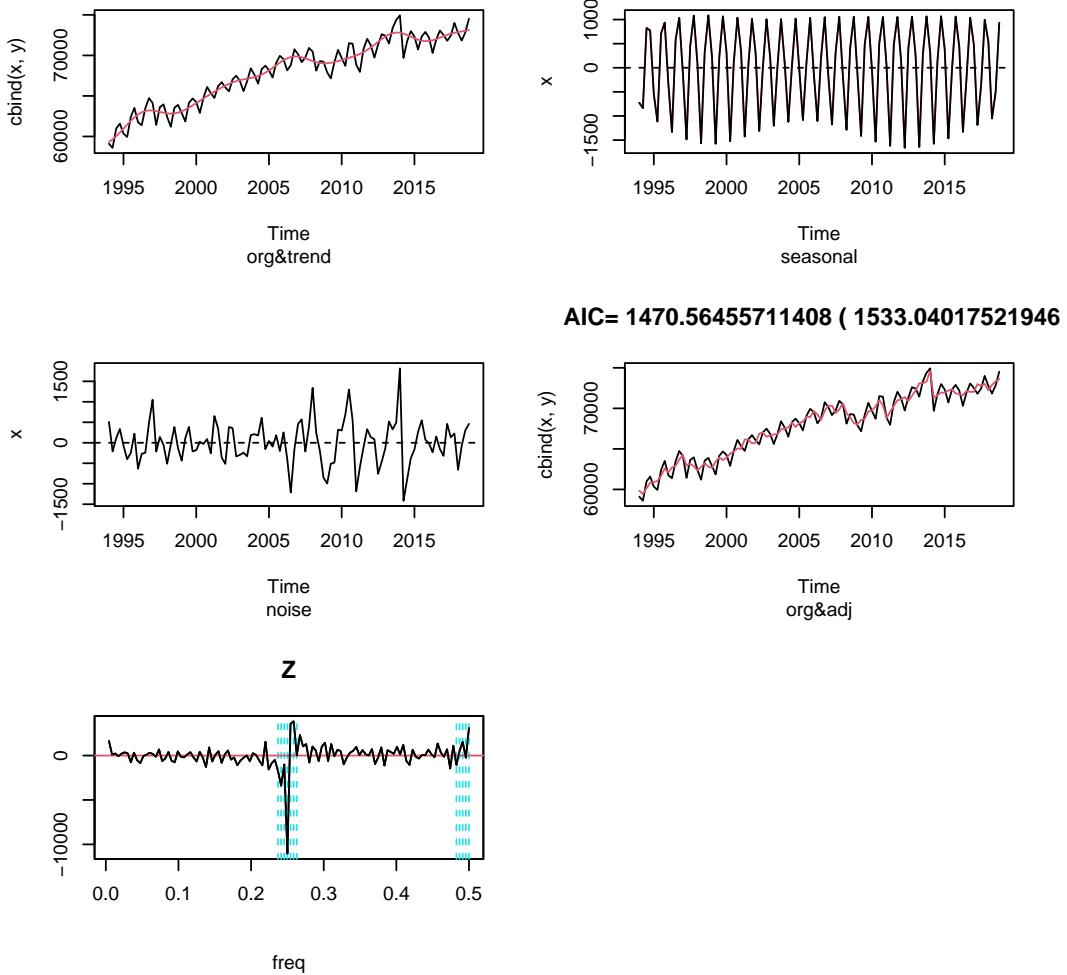


图 1: S-SIML 季节调整 (实际消费)

作为月度数据的例子，下面展示国友・櫻井・佐藤（2022）所分析的日本家计调查消费数据（2002–2019）的季节调整结果。与季度数据相同，原序列、季节成分、不规则成分及季节调整值汇总在图 2 中。通过频率域的 $K_n -$ 变换得到的直交成分 Z ，对应于季节性频率 $\lambda = 1/12, 2/12, 3/12, 4/12, 5/12, 6/12$ 附近的波动。由于月度经济时间序列作为离散序列，其季节成分会在多个频率上同时出现，这也是以往关于季节性问题的讨论长期存在混乱的主要原因之一。

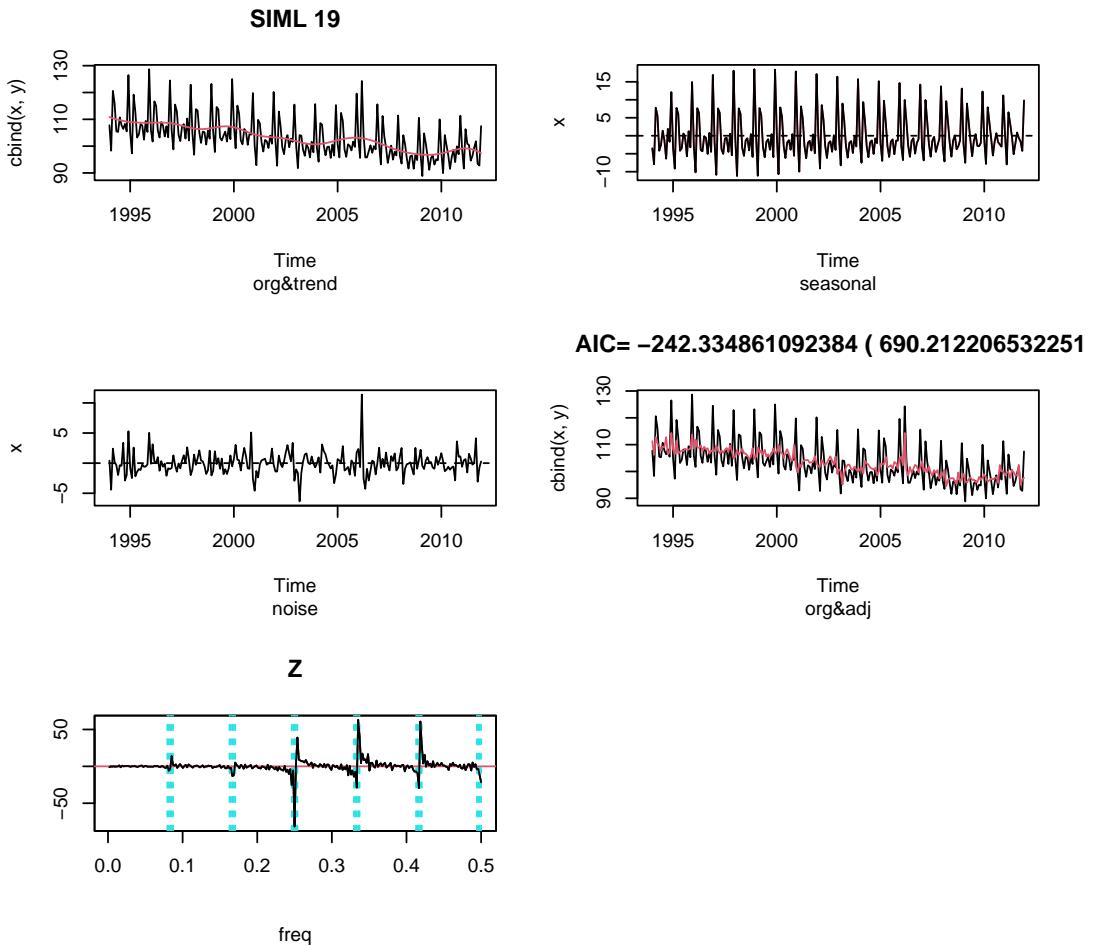


图 2: S-SIML 季节调整 (家庭消费)

4. 结语

本稿通过对包含噪声和季节性的非平稳时间序列进行频率分解，说明了如何从非平稳的季度或月度时间序列中，直接提取趋势—循环成分、季节成分以及季节调整序列的方法。详细的理论与方法可参考 Kunitomo-Sato (2021)、Kunitomo-Sato (2025) 以及国友·樱井·佐藤 (2022) 等文献。利用频率分析进行季节调整的程序在实践中并不常见，然而从“去除季节性”这一概念的本质来看，通过频率分解来实现季节调整是一种自然的思路。不过，由于大多数经济时间序列明显包含非平稳成分，基于这一前提的谱分解 (spectral decomposition) 方法长期以来并未受到足够关注。

在此，作为本文方法的背景说明，我们提及中国宏观经济数据的特殊情况。在中

国的官方统计中，对于部分重要的宏观经济指标，即使原始数据为月度数据，仍存在不公布1月数据、而仅公布1月与2月合计值的长期惯例。这种做法可以理解为中国统计当局在考虑春节等不同于其他主要发达国家的季节性波动因素时，为避免对内外造成误解而采取的公表策略。在美国、日本和欧洲诸国的官方统计中，为了理解宏观经济形势，通常会制作并公开经过季节调整的系列数据。这些国家主要采用美国人口普查局（U.S. Census Bureau）开发的X-12-ARIMA与X-13ARIMA-SEATS季节调整程序。然而，这些方法默认基于格里高利历（阳历）的一年季节循环来定义季节性。相比之下，中国受农历（阴历）等因素影响较大，因此这实际上提出了一个根本性的问题：“什么是季节性？”“所谓季节调整究竟意味着什么？”

本文所介绍的佐藤SIML季节调整法在非平稳季节性经济时间序列中，即使存在缺测值，也能通过简便的方法进行有效适用。正如国友（2024）所示，该方法在实证上表现出较高的灵活性与稳健性。希望这篇短文能够为经济时间序列中的**季节性与季节调整方法这一统计学问题的进一步研究与发展提供新的契机与启示。

References

- [1] 国友直人 (2024), 「非定常季節経済時系列における欠測値問題」, 未定稿。
- [2] Kunitomo, N. and S. Sato (2021), "A Robust-filtering Method for Noisy Non-Stationary Time Series," Japanese Journal of Statistics and Data Science (JJSD, Springer), 4, 373-410, Open Access.
- [3] Kunitomo, N. and S. Sato (2025), *The SIML Filtering Method for Noisy Non-stationary Economic Time Series*, Springer, forthcoming.
- [4] 国友直人・櫻井智章・佐藤整尚 (2022), “経済時系列の状態推定とマクロ指標,” 統計研究彙報, 79, 1-20, 統計局.
- [5] 佐藤整尚 (2023), “x12SIMLを使った季節調整の実際,” 未定稿.
(日本語 version)
<http://www.kunitomo-lab.sakura.ne.jp/x12simldoc92.pdf>
(English version)
[http://www.kunitomo-lab.sakura.ne.jp/x12simldoc92\(kuni2023-2-2\).pdf](http://www.kunitomo-lab.sakura.ne.jp/x12simldoc92(kuni2023-2-2).pdf)

<S-SIML 季节调整程序>

以下所述的 S-SIML 季节调整程序是基于本文所介绍的方法构建的，可通过以下网页下载：

<https://github.com/sato-labo/x12siml>

关于季节性与季节调整的讨论¹⁸

2024年12月3日

国友直人¹⁹

1. 引言

近年来，随着中国经济的扩张，其与日本经济、全球经济之间的联系日益紧密，其影响力也在不断增强。中国宏观经济的动向，尤其是近期的大幅波动，也成为日本经济学界高度关注的话题之一。然而，中国负责发布官方统计的机构出于若干原因，在宏观经济信息的公开方式上与主要发达国家有所不同，这使得掌握中国实体经济动态存在一定困难。例如，关于一些重要的宏观经济指标，即便是月度数据，也长期存在不公布1月单月数据，而是在事后统一公布1月与2月合计值的做法。与此同时，包括日本在内的众多国家的经济政策制定者与经济学家，普遍依据从原始数据生成的季节调整后系列，来判断宏观经济的实际状况。近年来，各国负责统计制定与发布的机构多采用由美国人口普查局开发的X-12-ARIMA法与X-13ARIMA-SEATS法作为季节调整工具。然而，面对如非格里历（阳历）而是阴历的日历因素，或原始数据中存在缺失值的情况，应采用何种选项以妥善处理季节性，至今并无定论。

本文将对2024年应内阁府委托而实施的统计咨询专题研讨“移动节假日与经济指标的季节性”中所探讨的问题进行补充说明。由于相关议题较多，本文仅选取与咨询请求宗旨相符的几个要点加以阐述。

2. 对非平稳宏观经济时间序列的加法分解

宏观经济时间序列中，非平稳性与季节性往往是两项显著特征。因此，基于传统平稳过程理论的统计分析方法，往往不适用于此类数据。在传统的宏观经济数据分析中，常将观测数据建模为加法形式：

$$(\text{原始序列}) = (\text{趋势} \cdot \text{循环成分}) + (\text{季节成分}) + (\text{不规则波动})$$

其中“不规则波动”还包括星期效应、突变点、闰年效应等要素，应予以留意。而在分析中国的宏观经济时间序列时，进一步将“季节成分”细分为：

¹⁸本稿是根据2024年内阁府政策统括官（经济财政分析）下属参事官（海外担当）的委托，在“移动节假日与经济指标的季节性”统计咨询项目中探讨的内容，由国友直人负责任地进行了补充整理。

¹⁹统计数理研究所

(中国宏观序列中的季节性) = (阳历季节性) + (阴历季节性) 如赵宇与国友 (2024) 所指出，春节等节日的季节性具有由政府指定的制度性特征，并非完全基于天文历法，因此需要特别注意。

3. 关于 X-12-ARIMA 与 X-13ARIMA-SEATS 的使用

在 X-12-ARIMA 方法中，为了估计季节性，会使用 RegARIMA 模型，首先从原始序列中剔除被视为非季节性的要素，如闰年效应、星期效应、春节效应等，然后通过复杂的移动平均（即 X-11 滤波器）估计季节性成分。这种估计出的季节性可理解为“阳历季节性”。例如，借助虚拟变量就可以对这种“阳历季节性”进行建模与估计。高冈 (2024) 以中国贸易统计为例进行分析，所得出的季节调整后的序列形式如下：

(季节调整后系列) = (原始序列) - (阳历季节性) = (趋势 · 循环) + (阴历季节性) + (不规则波动) 因此，对于经济学家而言，真正关注的宏观经济状态，可以通过 (趋势 · 循环) 的状态估计值，或 (季节调整后系列) - (阴历季节性) 的状态估计值加以把握。

4. 关于 S-SIML 的使用

在 X-12-ARIMA 模型中，季节性的定义并不明确，其估计是通过重复的复杂移动平均操作实现的。该方法未考虑阴历季节性，因此可以认为仅对应“阳历季节性”。S-SIML 季节调整方法则明确地基于频域分解对趋势 · 循环和阳历季节性进行定义，从而使季节性的概念更加清晰。它通过包含季节周期波动与季节差分操作的状态变量模型，对季节因子进行状态估计。因此，S-SIML 方法得到的季节调整序列：

(季节调整后系列) = (原始序列) - (阳历季节性) 此序列依然包含阴历季节性成分。

此外，如国友 (2024) 所讨论，S-SIML 方法对于含有缺失值的非平稳季节性时间序列也较易处理。阴历季节性的估计可借助频率分解与虚拟变量实现。由于阳历年约为 365 天，而阴历年约为 354 天，其年周期频率略有差异。因此，调整季节阶数 (sorder) 扩大设定可能是实际可行的方式之一。但考虑到中国的节假日安排每年由官方决定，使用虚拟变量法也是一种有前景的替代方案，有待进一步研究。无论采用哪种方法，对于经济学家而言，合理的宏观经济判断依据仍是：(趋势 · 循环) 或 (季节调整后系列 - 阴历季节性) 的状态估计值。

5. 结语

中国的宏观经济数据为经济统计研究提出了若干前所未有的且颇具挑战性的统计问题。

在美国、日本与欧洲诸多国家，季节调整后的时间序列被广泛用于了解宏观经济状态与政策制定，并被媒体定期报道。但在实际操作中，这些数据多使用 X-12-ARIMA、X-13ARIMA-SEATS 等经典季节调整程序处理。这些方法默认采用格里历（阳历）定义年度周期与季节性。然而在中国，由于阴历及制度性节假日安排的广泛影响，引发了一个更为根本的问题：“什么是季节性？”、“什么是季节调整？”

本文希望能作为推动宏观经济数据中季节性问题研究与解决的一次契机，为后续的研究与实践奠定基础。

References

- [1] 趙宇・国友国友直人 (2024), 「中国のカレンダーと経済活動：趙ダミー」, 未定稿。
- [2] 高岡慎 (2024), 「中国の貿易統計と季節調整」, 未定稿。」
- [3] 国友直人 (2024), 「非定常季節経済時系列における欠測値問題」, 未定稿。
- [4] 佐藤整尚・国友直人 (2024), 「SarSIML 季節調整」, 未定稿。
- [5] Kunitomo, N. and S. Sato (2025), *The SIML Filtering Method for Noisy Non-stationary Economic Time Series*, Springer, forthcoming.
- [6] 国友直人・櫻井智章・佐藤整尚 (2022), “経済時系列の状態推定とマクロ指標,” 統計研究彙報, 79, 1-20, 統計局.

结语

作为“统计专家培养项目”的一环，我们策划了2024年度的统计咨询专题研究讨论“移动节假日与经济指标的季节性”。在该演习中，我们听取了日本内阁府中负责中国经济分析的人员关于分析中国宏观经济数据所面临课题的说明。尽管参与“统计数理研究所·统计专家培养项目”的成员中并没有专门研究中国经济的专家，但我们还是能够理解，在统计的时间序列分析中，确实存在一些引人关注的问题。例如，对于趋势等非平稳性显著的宏观经济指标而言，存在着季节性不一定符合阳历的、显著的离散型经济时间序列中，如何处理缺失值及季节性等并非显而易见的问题。这是一个具有统计挑战性的问题。该问题在现有文献中几乎未被提及，但在中国经济影响力日益增强的国际背景下，它被视为经济统计中的一个重要课题。本次演习从统计学的角度出发，成为探讨并深入讨论该问题的一个宝贵机会。换句话说，我们通过这次统计咨询演习，得以思考围绕经济时间序列与实际季节调整之间新的、具体性问题。

正如本报告中所述，X-11、X-12-ARIMA、X-13ARIMA-SEATS 等现有的季节调整方法，基本上是在美国或欧洲的季节性假设基础上开发出来的。在本次演习中，我们首先探讨了中国历法相关的问题，包括公历（阳历）与农历（阴历）、春节效应等，并进一步研究了使用 X-12-ARIMA、X-13ARIMA-SEATS 方法时可能遇到的问题。此外，我们还探讨了在含有季节性的非平稳时间序列中，当存在缺失值时的滤波问题等。

本报告从较为自由的学术视角出发，说明了从中国宏观经济数据的处理，到经济活动中季节性与季节调整所面临的课题，并进一步提及了有关季节性的一些基础性问题。虽然 X-12-ARIMA 这种在官方统计中被普遍使用的季节调整方法，在统计科学角度来看，其机械化的使用其实并不理想。而中国的宏观经济统计正好将这一问题凸显出来。此外，本文也从统计学的角度，探讨了替代方案。我们希望本文的讨论能对今后日本理解中国宏观经济动向有所帮助。同时，本报告中介绍的 SarSIML 季节调整法（或称为 S-SIML 季节调整），其实也可以广泛应用于官方统计工作中。在这方面，我们也期望本报告的内容能为改进官方统计、判断国内外经济形势、乃至政策决策提供准确的信息，作出贡献。

这次统计咨询演习，成为了一个思考围绕经济统计的基本统计问题的机会。借此机会，谨向给予我们协助的内阁府相关人士、参与统计专家培养项目的各位成员，

以及所有相关人员表示衷心的感谢。

2025 年 1 月

国友直人