계절조정 고도화 기법에 관한 연구

성병찬*, 김소정**

최근 경제 · 사회 전반에서 시계열의 변동성과 불규칙성이 확대됨에 따라, 안정적인 계절조정계열을 생산하기 위한 고도화 기법의 필요성이 증대되고 있다. 특히 코로나19 팬데믹과 같은 장기적 빅쇼크는 일시적 충격이 계절효과로 흡수되거나 이후 구간에 잔여 계절성을 남겨 단기 경기 신호를 왜곡할 위험을 내포한다. 또한 GDP와 같이 계층적으로 집계되는 통계의 경우, 직접법과 간접법은 각각 계절조정 품질과 합산 일관성 측면에서 한계를 가지므로, 두 방법의 장점을 결합하는 개선 방안이 요구된다.

본 연구는 이러한 문제의식에 기초하여 (1) 빅이벤트 하 계절조정 기법과 (2) 계층적 시계열의 계절조정 고도화 방법을 제안한다.

첫째, RegARIMA 사전조정 단계에서 주요국의 팬데믹 대응 사례를 반영한 네 가지 옵션(자동보정, AO, AO/LS/TC, SO)을 병렬적으로 적용하고, AICc 및 BIC 기준에 따라 최적 사양을 선택하는 알고리즘을 제시한다. 아울러 빅이벤트 이전 기간과의 품질진단 비교를 통해 재검토가 필요한 계열을 선별함으로써, 충격의 계절성 확산을 통제하고 계절 성분 식별의 정확성을 제고하고자 한다.

둘째, 계층적 자료에 대해 모든 노드에서 직접 계절조정을 수행하여 초기 적합치를 산출한 뒤, 이를 최적 결합(optimal combination) 방식으로 조정(reconciliation)함으로써 합산 일관성을 만족하는 계절조정계열을 도출하는 방법을 제안한다. 이 접근법은 직접법과 간접법을 동일한 모형 체계 내에서 일반화한 것으로, 특정 하위 계열의 변동이 상위 계열에 과도하게 전이되는 문제를 완화할 수 있다는 장점을 가진다.

실증분석에서는 한국 GDP 계열에 제안한 방법을 적용하여, 빅이벤트 이상치 처리 및 계층적 조정이 계절조정 결과에 미칠 수 있는 영향을 예시적으로 제시하였다. 또한 한국은행의 계절조정 계열과 비교함으로써, 제안 방법의 특징과 실무적 함의를 검토하였다.

※ 본 연구의 내용은 집필자들의 개인 의견으로 한국은행의 공식 견해를 나타내는 것은 아닙니다.

* 중앙대학교 응용통계학과 교수(bcseong@cau.ac.kr)
** 한국은행 경제통계1국 통계연구팀 과장(sojung.kim@bok.or.kr)

Ⅰ. 서론
1. 연구 배경
2. 주요 연구 결과 및 기대 효과

Ⅱ. 계절조정 이론
1. 계절조정 개념
2. 사전조정
3. X-11 분해 알고리즘
4. 직접법과 간접법
5. 계절조정의 품질평가

Ⅲ. 빅이벤트 하에서의 계절조정
1. 빅이벤트와 계절조정
2. 주요 국가 통계청들의 팬데믹 대응 방식
3. 빅이벤트 계절조정 방식
4. 제안 알고리즘
5. 실증 분석 결과

Ⅳ. 계층적 시계열 자료의 계절조정
1. 계층적 시계열
2. 계층적 시계열에서의 문제와 해결 방법
3. 최적 결합 적합
4. 실증 분석 결과

Ⅴ. 요약 및 제언
1. 주요 연구 결과
2. 후속 연구를 위한 제언

Ⅰ. 서론

1. 연구 배경

최근 계절조정 고도화 기법에 대한 필요성이 높아지고 있다. 무엇보다도 경제 · 사회 전반에 나타나는 시계열의 변동이 과거보다 더 심해지고 불규칙적이기 때문이다. 시계열의 높은 변동성과 불규칙성은 안정성을 요구하는 계절조정 시계열 생산을 어렵게 한다. 이의 원인으로는 경제 대공황, 기상 이변 및 이상 감염병에 의한 경제 · 사회의 빅쇼크 등을 들 수 있다. 과거에는 상대적으로 이러한 이상 현상들의 빈도가 적거나 그 영향이 크지 않았고, 장기적으로 지속되지 않아서 주기적 계절 변동에 미친 영향도 크지 않았다. 따라서, 계절조정 작업에 미치는 영향도 상대적으로 크지 않았다. 그러나, 예를 들면, 2020년 초부터 시작된 코로나 팬데믹은 어느 나라할 것 없이 경제 · 사회 · 문화 모든 부문에 그동안 없었던 예측하기 어려운 변화들을 만들었고, 이로 인해 과거에는 무난했던 계절조정 필터링 작업들의 난이도가 한층 높아지게 되었다.

계절조정 경제통계는 경제 상황의 단기 동향을 분석하고 판단하는 데 있어서 매우 중요한 역할을 한다. 따라서, 경제 데이터에 대한 계절조정은 정확하고 적절하게 이루어져야 한다. 특히, 최근 팬데믹과 같이 장기간 경제에 큰 영향을 미치는 이벤트가 발생한 경우, 그 효과가 규칙적으로 반복되는 계절 효과에 포함되지 않도록 주의하여 계절조정을 실시해야 한다. 이는 특정 이벤트가 계절적 패턴을 왜곡하여 잘못된 경제 신호를 줄 수 있기 때문이다. 그러나 이벤트가 진행 중인 기간에는 해당 이벤트의 경제적 영향을 정확하게 파악하는 것이 어렵기 때문에, 계절조정시 이러한 이벤트 효과를 신중하게 처리할 필요가 있다. 하지만 이러한 상황을 적절히 반영한 계절조정 방법론에 대한 연구는 아직 충분하지 않은 실정이다.

계절조정의 정교함을 제고하기 위해서 최신 시계열 분석 연구를 반영할 필요가 있다. 그 중 계층적 시계열(hierarchical time series) 자료에 대한 최신 분석 기법을 계절조정의 직접법(direct adjustment) 또는 간접법(indirect adjustment)에 반영할 필요가 있다. 흔히, 계절조정에서는 합계 시계열(aggregate series)의 하위 구성 시계열(lower level series) 계절변동이 일정할 경우에 직접법을, 하위 구성 시계열 계절변동에 차이가 있는 경우에 간접법을 사용한다. 그러나, 이러한 구분은 쉽지 않기 때문에 두 방법의 장점은 살리고 단점을 보완할 수 있는 개선된 계절조정 방법이 필요하다.

본 연구는 이러한 문제의식을 바탕으로 (1) 빅이벤트 하에서의 계절조정과 (2) 계층적 시계열 자료의 계절조정 고도화 방법에 대하여 연구하고자 한다.

2. 주요 연구 결과 및 기대 효과

2.1 빅이벤트 하에서의 계절조정

본 연구는 팬데믹 또는 경제 대공황 등의 경제 · 사회적 충격이 계절조정 계열의 추정을 왜곡함을 방지하고 계절성분 식별에 정확성을 제고할 수 있는 방법을 사전조정 RegARIMA 단계에서 제안한다. 제안된 4가지 사전조정 옵션은 각 국가들에서 실제 사용하는 방법들을 최적으로 조합하는 효과가 있으며 모수절약 원칙을 지키기 위하여 유의한 이상치 시점만을 사용한다는 특징이 있다.

제안된 방법을 활용하면 팬데믹으로 인한 큰 이상치를 적절한 종류의 이상치로 식별하고 해당 기간 동안의 계절성분 추정의 정확성을 바탕으로 계절조정의 정밀성을 제고할 수 있을 것이다. 또한 경제 산업별 상황에 따라서 상이할 수 있는 팬데믹 이상치 처리의 계속 여부를 판단할 수 있다. 마지막으로 팬데믹 데이터로써 기존 계절성을 강화하기 위해서 사용할 것인지 아니면 새로운 계절 패턴을 추정하기 위해서 사용할 것인지에 대한 판단 기준을 제시한다.

2.2 계층적 시계열 자료의 계절조정

본 연구에서는 개별 시계열 간의 합계 일관성의 관계를 유지하면서 계절조정 정확성을 높이는 최적 결합 적합 기법을 제안한다. 이 제안된 기법을 사용하면 직접법과 간접법에 대한 검정 또는 판단 필요없이 두 방법의 장점을 동시에 누릴 수 있다. 이는 간접법과 직접법의 관계를 동일한 모형 안에서 일반화하는 이론적 토대를 마련한 것이라 볼 수 있다. 또한, 각 계층 시계열에서 나타나는 계층 고유의 특성을 직접 반영하여 계절조정의 정확성을 제고할 뿐만 아니라, 간접법의 장점이라고 볼 수 있는 집계 일관성을 유지할 수 있는 장점도 있다. 특히, 빅이벤트 하에서의 개입분석과 통합된다면 계절조정의 정교함과 안정성을 제고할 수 있을 것으로 기대된다.

Ⅱ. 계절조정 이론[1)]

1. 계절조정 개념

시계열 분석 분야에서 계절성(seasonality)에 대한 100% 정확한 정의는 없지만, 일반적으로 계절성이란 매년 반복되는 변동(fluctuation)을 일컫는다. 지구의 공전에 의해서 발생하는 자연현상의 계절성 뿐만 아니라, 기후, 사회적 제도 등으로 인해 1년을 주기로 반복하여 나타나는 변동을 계절성으로 정의하는 것이다.

이러한 계절성의 특징을 파악하기 위해 가장 널리 사용되는 방법은 시계열 분해(decomposition) 기법이다. 이 기법은 시계열을 비관측 구성 성분(components) 또는 변동(fluctuations)의 합 또는 곱 등으로 가정하고 각 성분들을 추정한다. 이러한 성분들로는 추세(trend) 성분, 순환(cycle) 성분, 계절(seasonal) 성분, 달력(calender) 성분, 불규칙(irregular) 성분이 있다.

추세성분은 장기적 관점에서 증가하거나 감소하는 패턴으로 정의한다. 순환성분은 확률적 주기(non-fixed period)를 가지는 반복 변동으로 정의하며, 보통 2년 이상의 주기를 가진다. 계절성분은 계절적 요인(seasonal factors)에 의한 변동으로서 고정된 길이의 주기(fixed period)를 가진다. 예를 들면, 1년의 분기별, 월별 또는 1주일의 요일별로 나타나는 현상들은 계절성분의 예가 된다. 달력성분은 달력효과라고도 하며, 조업일수(trading-day)가 다름에 따라서 생기는 조업일수 효과 및 종교적 또는 문화적 관습의 차이로 발생하는 변동인 이동공휴일(moving holiday) 효과가 대표적이다. 윤년에 의한 효과도 달력성분에 포함된다. 불규칙성분은 예측할 수 없는 변동으로서 이상치, 수준이동(level shift) 및 일시적 변화(temporary change) 등을 포함한다. 특히 앞에서 소개된 5개의 비관측 구성 성분들 사이에서도 추세성분과 순환성분은 그 구별이 쉽지 않아서 합하여 분석되는 경우가 많다. 이 때 합해진 성분은 추세순환성분이라고 일컫는다. 〈그림 1〉은 분기 시계열 원계열에서 분해된 각 성분들의 예이다. 여기서, 계절조정 계열은 원계열에서 계절성분을 제거한 계열로 정의한다.

시계열 분해 기법의 토대가 되는 모형은 가법 모형과 승법 모형이다. 이것들을 혼합한 모형의 형태도 있을 수 있다. 가법 모형은 아래와 같이 비관측 성분들을 덧셈으로 조합한다.

$$X_t = T_t + S_t + C_t + I_t = A_t + S_t + C_t . \tag{2.1}$$

1) 본 내용은 여인권 외(2022)에서 발췌한 것임.

여기서, X_t는 원 시계열, T_t는 추세순환성분, S_t는 계절성분, C_t는 달력성분, I_t는 불규칙성분을 나타낸다. A_t는 계절조정계열(seasonally adjusted series)로서 가법모형에서는 $A_t = T_t + I_t$이다.

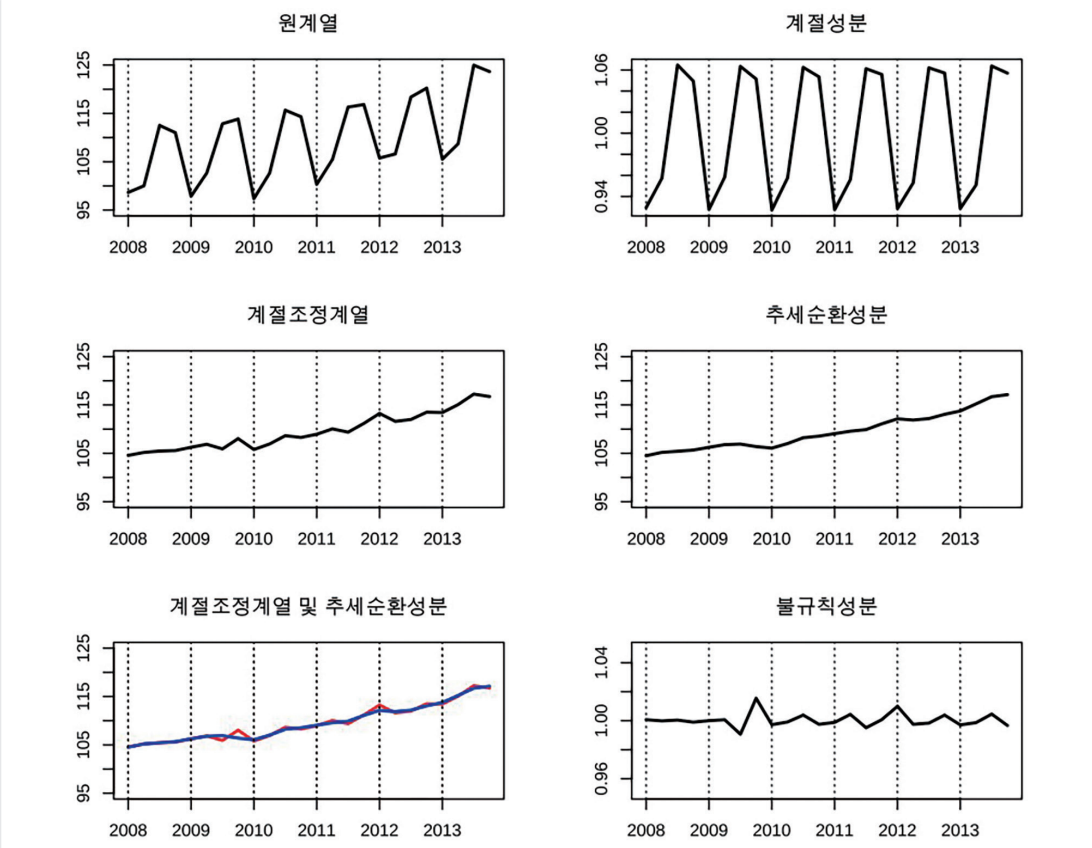
승법모형일 경우 다음과 같다.

$$X_t = T_t \cdot S_t \cdot C_t \cdot I_t = A_t \cdot S_t \cdot C_t \text{; 단, } A_t = T_t \cdot I_t \tag{2.2}$$

주로 원계열이 그 수준에 따라서 변동폭의 변화가 일정할 때는 가법모형이 선호되며, 그 수준에 따라서 원계열의 변동폭이 점차로 커지는 등 일정하지 않을 때는 승법모형이 사용된다. 분해법은 비관측 성분 모형 및 각 구성 성분들을 어떻게 정의 및 가정하고 추정하느냐에 따라서 다양하다. 다만, 적절한 가정을 통해서 승법모형은 분산안정화변환의 일종인 로그변환을 통해서 가법모형으로 변환될 수 있다.

〈그림 1〉

시계열 분해 기법과 계절조정의 예

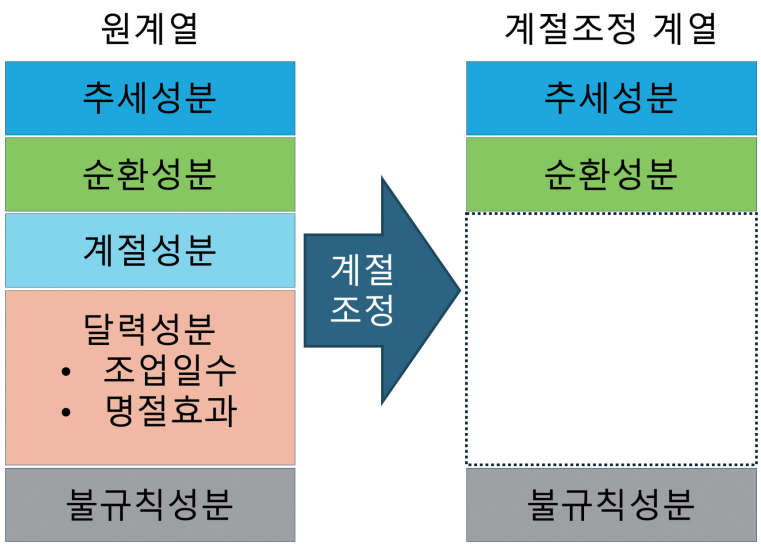


출처: International Monetary Fund (2018)

계절조정이란, 이렇게 경제통계 원계열로부터 주기적인 계절변동 및 달력효과를 제거하고 A_t 계절조정계열을 추출하는 작업이다. 〈그림 2〉에서처럼 원계열이 앞서 설명한 다섯 개의 성분으로 구성되어 있다고 가정하고 그 성분들을 추정한 이후, 계절 및 달력성분을 제거하는 작업이다. 이렇게 계절조정을 하게 되면, 서로 가까운 두 시간을 이용하여 시기적절하게 비교할 수 있고 계절을 구성하는 각 시간(예: 분기와 분기, 월과 월)을 공평하게 비교할 수 있게 된다.

〈그림 2〉

계절조정의 개요도



대표적인 계절조정 방법으로는, X-12-ARIMA와 TRAMO/SEATS를 들 수 있다. 전자는 미국 및 캐나다를 중심으로 발전해 온 방법으로서, 가정의 종류가 비교적 적고 비모수적(nonparametric) 접근 방법이다. 후자는 주로 유럽 등지에서 발전해 왔으며 많은 종류의 가정을 포함하는 모수적(parametric) 방법이다.

X-12-ARIMA 방법은 보다 정확하고 편리한 시계열 분해를 위해 사전조정(pre-adjustment) 단계를 거치며, 이 과정에서 ARIMA(autoregressive integrated moving average) 오차를 가정한 RegARIMA 회귀 모형을 사용한다. 이후 시계열 분해를 위하여 각종 이동평균 필터를 사용하는 X-11 알고리즘을 거친다. TRAMO/SEATS 방법은 사전조정을 위해서 RegARIMA와 유사한 TRAMO 시계열 회귀 모형을 사용하며, 이후 ARIMA 시계열에서 신호를 추출하는 SEATS(signal extraction in ARIMA Time Series) 알고리즘을 사용한다.

최근 이 두 가지 계절조정 방법은 융합되고 있으며 X-13-ARIMA-SEATS 방법으로 명명되고 있다. 이 방법은 사전조정계열을 위해서 RegARIMA 모형을 사용하고 시계열 분해를 위해서 X-11 알고리즘 또는 SEATS를 선택할 수 있다. 계절조정의 품질 진단을 위해서는 주로 X-12-ARIMA의 방법들을 사용한다.

계절조정을 위한 소프트웨어로는 그래픽 인터페이스를 강화한 Demetra 및 JDemetra가 있다. 이들은 Eurostat에 의해서 개발되었으며 X-12-ARIMA 및 TRAMO/SEATS를 위한 그래픽 및 메뉴 인터페이스를 장점으로 강조한다. JDemetra는 Demetra의 업그레이드 버전이며 X-13-ARIMA-SEATS를 사용할 수 있다. 한편 DOS 프로그램으로는 X-12-ARIMA가 있다. SAS에서도 계절조정을 위한 프로시저인 proc x12를 제공한다. 한국은행에서는 BOK-X-13-ARIMA를, 통계청에서는 X-12-ARIMA 계열인 KOSTAT-CIS 프로그램을 개발하였다. R 패키지로는 seasonal 및 RJDemetra가 있다.

2. 사전조정

사전조정(pre-adjustment)은 계절성분을 보다 정교하게 추출하기 위하여 사전에 달력 효과와 이상치를 제거하는 과정이다. 즉, 원계열을 종속변수로 하고 달력성분이나 이상치들로 구성된 설명변수를 이용하여 사전조정계열을 만드는 과정이다. 다음과 같은 모형을 고려할 수 있다.

$$X_t = \beta_1 Z_{1t} + \cdots + \beta_k Z_{kt} + O_t . \tag{2.3}$$

여기서, Z_{it}는 회귀계수 β_i를 가지는 설명변수들을 나타내며 O_t는 ARIMA 모형을 따르는 사전조정계열이다. 사전조정계열은 최종 계절조정 단계를 위하여 사전적으로 만드는 시계열로서, 원계열에서 달력성분과 이상치가 제거된 시계열이다. 특히, 달력성분은 이를테면 '불규칙적인' 계절성분이므로 최종 단계의 계절성분에 포함될 수 없다. 사전조정을 위한 설명변수는 대표적으로 달력과 외부적 요인(이상치)을 들 수 있다.

2.1. 달력

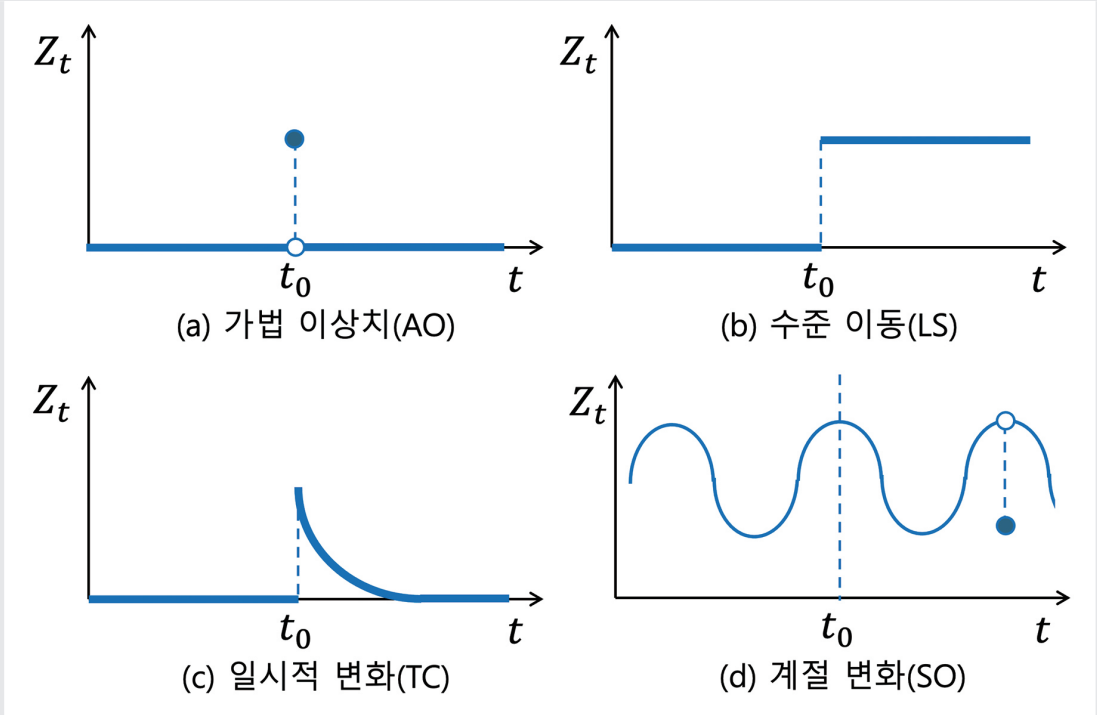
달력(calendar)에 의해서 나눠지는 각 월과 각 분기들은 서로 동등하게 비교하기 어렵다. 왜냐하면, 각 월의 길이가 서로 달라서 더 많은 조업일수(trading days)가 있는 월은 더 많은 생산력을 가지고 있기 때문이다. 같은 2월이라도 윤년일 경우 그 길이가 다르다. 각 월과 분기들은 평일과 공휴일의 수가 다르다. 또한, 공휴일은 이동 계절성(moving seasonality)이라 하여 그 날짜가 움직인다. 예를 들면, 부활절, 음력 설날 및 추석 등은 매년 그 날짜가 변화한다. 각 국가별로 고려하자면, 국가별로 공휴일이 다르다. 부활절은 동방 또는 서방 교회식에 따라 날짜가 다르며 이슬람 국가에서는 라마단 기간이 있다.

2.2. 외부적 요인

사전조정을 위한 설명변수로 외부적 요인에 의해서 발생되는 변화들을 고려해야 한다. 첫째 극단적이거나 이례적인 관측값들을 일컫는 이상치(outliers)이다. 예를 들면, 여름철 저온 현상으로 인하여 아이스크림 판매량이 갑자기 감소한다든지, 파업 또는 결빙일로 인하여 건설업 등에 큰 영향 또는 변화가 발생할 수 있다. 〈그림 3〉 (a)는 흔히 발생하는 이상치인 가법형태의 이상치(addive outlier, AO) 형태를 보여준다. 둘째, 영구적 변화를 일으키는 수준 이동(level shift, LS)이다. 예를 들면, 관세 및 세금의 변화로 매출에 영향을 주는 경우이다. 금융위기(financial crisis)에 의한 수준 이동도 고려할 수 있다. 〈그림 3〉 (b)는 기본적인 수준 이동의 형태이며, 이의 변형된 형태로 서서히 새로운 수준으로 이동하는 램프식도 고려할 수 있다. 셋째, 일시적 변화(transitory changes)이다. 예를 들면, 예상치 못한 폭풍으로 인한 목재의 벌채 또는 커피 가격의 상승과 같은 변화를 고려할 수 있다. 이런 변화는 〈그림 3〉 (c)처럼 시간이 지나면 원상태로 회복된다. 넷째, 계절적 이상치(seasonal outliers, SO)이다. 예를 들면, 빅이벤트의 영향으로 주기적으로 나타나던 계절적 패턴의 행태가 바뀔 수 있다. 즉, 여름이 겨울되고 겨울이 여름이 될 수 있다는 것이다. COVID-19 기간동안 각 국가의 중앙정부는 계절적 기념 행사 및 연말 행사 등 사람들의 집단적 모임을 금지했었다. 〈그림 3〉 (d)는 이러한 일시적 계절 이상치를 나타낸다.

〈그림 3〉

외부적 요인에 의한 변화



2.3. 사전조정 유형

위에서 설명된 2가지 사전조정을 위한 변수인 외부적 요인과 달력 효과는 조정 방식이 서로 다르다. 외부적 요인은 일시적 조정으로, 달력 효과는 영구적으로 조정된다. 즉, 최종 분해법 적용 이후, 일시적 조정은 계절조정계열에 다시 추가되지만 영구적 조정은 추가되지 않는다. 외부적 요인은 정교한 계절조정을 위한 사전단계로 일시적으로 조정하고, 최종 계열에는 실제 현실에서 예상할 수 없이 발생할 수 있기 때문에 다시 추가한다. 달력효과와 같은 영구적 조정은 그 목적이 표준적인 월 또는 분기를 만들어서 각 월 또는 분기를 공평하게 비교하기 위함이므로 최종 계열에 다시 추가할 필요가 없다. 〈표 1〉은 이러한 2가지 사전조정 유형을 정리한 것이다.

〈표 1〉

사전조정의 2가지 유형

사전조정 유형	일시적 조정	영구적 조정
종류	• 가법 이상치(AO) • 수준 이동(LS) • 일시적 변화(TC) • 계절 변화(SO)	• 조업일수(trading days) • 부활절 • 국경일
특징	• 외부적 요인(이상치) • 자동옵션(AO/LS)을 선호 • 최종 계절조정 계열에 다시 포함됨	• 달력효과 • 각 국가 및 특정 통계에 적합하도록 옵션 조정 • 최종 계절조정 계열에 포함되지 않음

외부적 요인은 계절조정 프로그램에서 자동 옵션으로 조정되는 것이 선호되며, 달력효과 등은 각 국가 및 특정 통계에 적합하도록 옵션을 잘 조정하는 것이 필요하다. 이러한 사전조정 설명변수들은 t-검정을 통하여 유의성이 판단되며 유의할 경우 사전조정에서 고려되어야 한다.

영구적 조정에 해당하는 조업일수는 설명변수 Z_t를 각 월의 조업일수 값으로 부여하여 조정된다. 즉, t월의 일수 − [일요일 수 + 그 이외 공휴일 수]로 계산될 수 있다. 예를 들어, Z_t = (20, 23, 21, 22, 22, 21, 23, ...)와 같은 형태로 주어진다. 각 월의 요일의 날수를 고려하기 위해서는 월요일부터 일요일까지의 날수를 나타내는 Z_{1t}, ... ,Z_{7t}를 도입하여야 한다. 예를 들어, t월의 월요일 수를 Z_{1t} = (4, 4, 5, 4, 4, 5, 4, 5, 4, ...)와 같이 만들 수 있다. 일반적으로 이러한 변수들은 해당 t월의 평균조업일수 또는 평균 요일수로 표준화되어 사용된다. 즉, 모든 월들이 같은 수의 조업일수 또는 요일들이 포함하도록 표준화 되어야 하는 것이다.

우리나라 경제통계의 계절조정 과정에서 가장 중요한 것 중 하나는 명절효과(음력 설날과 추석)를 조정하는 것이다. 기본적으로 명절효과는 Bell and Hillmer (1983) 등의 방법을 이용하여 더미변수를 설정하고 추정한다. 명절효과는 지속 형태와 기간에 따라 더미변수의 값이 변하는데 명절에 가까우면 가까울수록 효과가 더 커지는 Dagum형(Dagum, 1988), 지속기간동안 효과가 일정한 Bell형으로 구분된다. 명절 전후 효과의 지속 기간을 w, 해당 월에 명절이 미치는 영향 일수를 k라고 할 때 〈표 2〉는 Dagum형과 Bell형에 대한 명절효과를 정의한다(전백근, 2002). 단, 최종 명절효과 변수는 다음과 같이 계산된다.

$$\text{명절효과} = H(w,t) - \overline{H}(w,t); \quad \text{단, } \overline{H}(w,t)\text{은 } t\text{월에 대한 } H(w,t)\text{의 평균} \tag{2.4}$$

〈표 2〉

명절효과를 변수화하는 두 가지 방법

구분	Dagum	Bell
명절이 1월/9월 말인 경우	$H(w,t-1) = 0.5 + 0.5 \times \left(\sum_{i=k}^{w} i\right) / \left(\sum_{i=1}^{w} i\right)$ $H(w,t) = 0.5 \times \left(\sum_{i=1}^{k-1} i\right) / \left(\sum_{i=1}^{w} i\right)$	$H(w,t-1) = 0.5 + 0.5 \times \left(\frac{k}{w}\right)$ $H(w,t) = 0.5 \times \left(\frac{w-k}{w}\right)$
명절이 2월/10월 초인 경우	$H(w,t-1) = 0.5 \times \left(\sum_{i=1}^{k} i\right) / \left(\sum_{i=1}^{w} i\right)$ $H(w,t) = 0.5 \times \left(\sum_{i=k+1}^{w} i\right) / \left(\sum_{i=1}^{w} i\right) + 0.5$	$H(w,t-1) = 0.5 \times \left(\frac{k}{w}\right)$ $H(w,t) = 0.5 \times \left(\frac{w-k}{w}\right) + 0.5$

3. X-11 분해 알고리즘

사전조정계열을 추세-순환, 계절 및 불규칙 성분으로 분해하는 과정으로 본 연구에서 사용할 X-11 분해를 위한 필터와 알고리즘을 설명한다.

3.1. $n \times m$ 이동평균과 X-11의 필터

X-11 알고리즘의 중심 도구인 $n \times m$ 이동평균은, m개 항에 대한 이동평균을 연속적으로 n개 구간에서 계산하는 것을 의미한다. 예를 들어, 3×3 이동평균 $\overline{X}^{3\times3}$은 다음과 같이 정의한다.

$$\begin{aligned} X^{3\times3} &= \frac{1}{3}\left(\frac{X_{t-2}+X_{t-1}+X_t}{3} + \frac{X_{t-1}+X_t+X_{t+1}}{3} + \frac{X_t+X_{t+1}+X_{t+2}}{3}\right) \\ &= \frac{1}{9}X_{t-2} + \frac{2}{9}X_{t-1} + \frac{3}{9}X_t + \frac{2}{9}X_{t+1} + \frac{1}{9}X_{t+2} \end{aligned} \tag{2.5}$$

이러한 이동평균을 이용하여 X-11의 각종 필터들이 만들어진다.

최초 계열(raw series)인 사전조정계열 O_t에서 추세(순환)을 추정하기 위해서, 월별 자료에 대해서는 2×12 이동평균, 분기별 자료에는 2×4 이동평균을 사용한다. 이것들이 사전조정계열에서 계절성을 없애고 추세(순환)를 추정하는 단계를 위한 필터이다.

추세가 제거된 계열(SI-ratio, seasonal-irregular component 또는 detrended series)에서 계절

성분을 추정하는 필터로는 $3\times3S$, $3\times5S$ 또는 $3\times9S$를 사용한다. 여기서, S는 계절 주기를 의미한다. 예를 들어, 월별 자료에 대해서 $3\times3S$를 사용하고 추세가 제거된 계열을 $(SI)_t$로 나타낼 때 다음의 이동평균을 계산한다.

$$\frac{1}{3}\left((SI)_{t-12}+(SI)_t+(SI)_{t+12}\right) \tag{2.6}$$

3.2. X-11 알고리즘

X-11 분해 알고리즘은 다음과 같은 3단계의 과정을 거치며 각 시계열 성분들을 분해한다. 각 단계들은 특정 성분을 추정하기 위한 선형필터들로 구성된다.

- 1-단계: 초기 추정치 계산
- 2-단계: 최종 계절성분 추정
- 3-단계: 최종 추세 및 불규칙 성분 추정

이후 설명의 편리성을 위해서, 사용할 분해 모형을 가법모형, 계절조정의 대상이 되는 시계열을 월별 시계열, 사전조정계열 O_t를 다음과 같이 가정한다.

$$O_t = T_t + S_t + I_t = A_t + S_t \tag{2.7}$$

[1-단계]

추세성분의 초기치를 계산한다. 13개항을 이용하는 2×12 이동평균을 이용할 경우 다음과 같이 표현할 수 있다.

$$T_t^{(1)} = \frac{1}{24}O_{t-6}+\frac{1}{12}O_{t-5}+\cdots+\frac{1}{12}O_t+\cdots+\frac{1}{12}O_{t+5}+\frac{1}{24}O_{t+6} \tag{2.8}$$

이를 이용하여 추세가 제거된 계열 SI-ratio의 초기치를 계산할 수 있다.

$$(S_t+I_t)^{(1)} = SI_t^{(1)} = O_t - T_t^{(1)} \tag{2.9}$$

1-단계의 2번째 계산은 계절성분의 초기치 추정이다. 5개항을 사용하는 $3\times3S$의 계절 이동평균을 이용할 때 다음과 같이 표현할 수 있다.

$$\tilde{S}_t^{(1)} = \frac{1}{9}SI_{t-24}^{(1)}+\frac{2}{9}SI_{t-12}^{(1)}+\frac{3}{9}SI_t^{(1)}+\frac{2}{9}SI_{t+12}^{(1)}+\frac{1}{9}SI_{t+24}^{(1)} \tag{2.10}$$

이후 표준화 과정을 거치면 다음과 같이 계산된다. 즉, 전후 13개항의 2×12 이동평균을 이용하여 표준화한다.

$$S_t^{(1)} = \tilde{S}_t^{(1)} - \left(\frac{1}{24}\tilde{S}_{t-6}^{(1)} + \frac{1}{12}\tilde{S}_{t-5}^{(1)} + \cdots + \frac{1}{12}\tilde{S}_{t+5}^{(1)} + \frac{1}{24}\tilde{S}_{t+6}^{(1)} \right) \tag{2.11}$$

1-단계의 마지막 계산은, 1-단계 계절조정계열을 다음과 같이 계산하는 것이다.

$$A_t^{(1)} = O_t - S_t^{(1)} \left(= T_t^{(1)} + I_t^{(1)} \right) \tag{2.12}$$

[2-단계]

2-단계의 첫 번째 과정은, 1-단계의 $A_t^{(1)}$를 이용하여 2단계 추세성분 추정치 $T_t^{(2)}$를 계산하는 것이다. 이 때 경험적으로 유용하다고 알려져 있는 이동평균 형태인 Hendersen 필터를 사용한다. Hendersen 필터는 가중치 개수가 5~23개인 다양한 종류가 있으며 가중치의 값에 마이너스 값도 존재하는 특징이 있다. 계산된 추세성분 추정치 $T_t^{(2)}$는 다음과 같이 SI-ratio 계산에 사용된다.

$$(S_t + I_t)^{(2)} = SI_t^{(2)} = O_t - T_t^{(2)} \tag{2.13}$$

계절성분의 추정은, 1-단계와 유사한 과정으로 다음과 같이 계산된다. 즉, 계절 이동평균(흔히, 3×5S를 사용)으로 $\tilde{S}_t^{(2)}$를 계산하고 이를 표준화하여 2단계 계절성분 $S_t^{(2)}$를 추정한다.

$$\begin{aligned} \tilde{S}_t^{(2)} = & \frac{1}{15}SI_{t-36}^{(2)} + \frac{2}{15}SI_{t-24}^{(2)} + \frac{3}{15}SI_{t-12}^{(2)} \\ & + \frac{3}{15}SI_t^{(2)} + \frac{3}{15}SI_{t+12}^{(2)} + \frac{3}{15}SI_{t+24}^{(2)} + \frac{3}{15}SI_{t+36}^{(2)} \\ S_t^{(2)} = & \tilde{S}_t^{(2)} - \left(\frac{1}{24}\tilde{S}_{t-6}^{(2)} + \frac{1}{12}\tilde{S}_{t-5}^{(2)} + \cdots + \frac{1}{12}\tilde{S}_{t+5}^{(2)} + \frac{1}{24}\tilde{S}_{t+6}^{(2)} \right) \end{aligned} \tag{2.14}$$

이때 계절 필터의 선택은 보통 3×3S, 3×5S 또는 3×9S 중에서 사용한다. 불규칙성분 I_t에 비해서 계절성분 S_t의 변화가 클수록 더 짧은 필터가 선호되며, 반대의 경우는 더 긴 필터를 사용한다. 불규칙성분의 영향력을 축소하는 방법으로 계절 필터를 선택하는 것이다. 2-단계 최종 계절조정계열은 다음과 같이 계산할 수 있다.

$$A_t^{(2)} = O_t - S_t^{(2)} \tag{2.15}$$

[3-단계]

최종 추세 및 불규칙 성분을 추정하고 분해의 최종 결과를 구하는 단계이다. 우선, 최종 추세 추정치는 Hendersen 필터를 사용하여 다음과 같이 계산한다.

$$T_t^{(3)} = \sum_{i=-H}^{H} h_i A_{t+i}^{(2)} \text{ ; 단, } H\text{: 필터의 길이, } h_i\text{: 가중치} \tag{2.16}$$

최종 불규칙성분의 추정치는 다음과 같이 계산한다.

$$I_t^{(3)} = A_t^{(2)} - T_t^{(3)} \tag{2.17}$$

따라서, X-11 분해 알고리즘의 최종 결과는 다음과 같이 표현할 수 있다.

$$\tilde{O}_t = T_t^{(3)} + S_t^{(2)} + I_t^{(3)} \tag{2.18}$$

단, X-11 알고리즘의 반복 과정에서 SI-ratio의 극단값은 평균값으로 대체되기 때문에, 일반적으로 $\tilde{O}_t$는 O_t와 다르다.

4. 직접법과 간접법

계층적(hierarchical) 또는 집계(aggregated) 시계열의 계절조정을 위해서 직접법과 간접법 중 어느 것을 사용할 지 선택하여야 한다. 즉 계층적 구조의 시계열들을 먼저 합하고 계절조정할 것인지? 계절조정한 이후 합할 것인지? 선택하여야 한다.

4.1. 직접법(direct method)

직접법은 계층 시계열이 합하여 만들어진 이후에 각 시계열에 대해 독립적으로 계절조정을 수행하는 방법이다. 이 방법을 통해서 추세-순환(trend-cycle) 성분의 추정을 개선할 수 있으며, 구성 시계열들이 유사한 계절 패턴을 보이고 추세-순환이 서로 높이 상관되어 있을 때 가장 좋은 결과를 제공한다고 알려져 있다.

4.2 간접법(indirect method)

간접법은 최하위(bottom) 시계열부터 먼저 계절조정을 수행한 뒤, 이를 합산(bottom-up)하는 방식이다. 지나치게 세분화된 하위 계열을 사용할 경우, 합산된 시계열에 잔여 계절성(residual seasonality)이 남을 수 있다는 위험이 있다. 구성 시계열들이 매우 상이한 계절 패턴을 보일 때 가장 좋은 결과를 제공한다고 알려져 있다.

5. 계절조정의 품질평가

계절조정의 품질 평가는 보통 다음의 4가지 측면을 포함한다.

- 원계열에 식별가능한 계절성이 존재하는가?
- 계절조정계열에 계절성이 남아 있는가? 즉, 잔존 계절성(residual seasonality) 유무를 검토
- 사전조정단계에서 포함 및 식별된 달력 효과 및 외부적 요인들의 유의성 검정
- 추정된 regARIMA 모형의 잔차에 대한 진단

계절조정에 대한 고급 진단으로는, 슬라이딩 스팬(sliding spans) 분석과 리비전 히스토리(revision history)를 포함한다. 이 진단들은 모두 계절조정 추정 절차에 추가적인 관측치가 포함될 때 계절조정 결과의 안정성(stability)을 살펴보는 것이다. 또한, 두 진단은 최소 길이의 시계열에 한해서 가능하며 그 결과는 계절조정에 사용되는 다양한 필터들의 길이에 의존한다.

5.1 계절성 검정

계절성 검정은 대상 시계열에 계절성이 존재하거나 잔존하는지를 F-검정한다. 원계열 및 계절조정계열에 대해서 계절성 검정이 수행된다. 원계열에 대해서는 다음의 3가지 검정이 이루어진다;

- 안정적 계절성의 존재성에 대한 F-검정
- 안정적 계절성의 존재성에 대한 Kruskall-Wallis 검정
- 이동 계절성(moving seasonality)에 대한 F-검정

추가적으로 5.2절에서 소개되는 M7 통계량을 사용할 수도 있다. 위의 3가지 검정을 통해서 식별가능한 계절성이 존재하지 않는다고 판단되면, 계절조정을 수행하지 않는다.

계절조정계열에 대한 계절성 검정은 주로 2가지 F-검정을 수행한다. 첫 번째는 계절조정계열 전체 길이에 대하여 계절성이 존재하는지 검정하는 것이다. 두 번째는 마지막 3년의 길이에 대해서 계절성이 존재하는지 검정한다. 두 번째 검정은, 가장 최근 자료에 초점을

맞추고 계절조정의 품질을 평가하는 것이다.

사전조정 단계에서 사용된 regARIMA 모형 또한 표준적 회귀진단 과정을 거쳐야 한다. 주로 정규성 검정과 자기상관성 검정을 위한 Ljung-Box의 Q-검정을 수행한다.

5.2 Q-통계량

또다른 유용한 계절조정에 대한 진단 통계량으로 M-통계량 및 이들을 요약하는 Q-통계량이 있다. 특히, Q-통계량은 계절조정 결과의 품질을 종합적으로 평가하기 위하여 불규칙변동의 정도 및 계절성의 안정성 등과 관련된 M-통계량들을 가중평균한 값이다. 이들 모두는 0~3 사이의 값을 가지며, 1보다 큰 값을 가질 때는 계절조정에 잠재적 문제가 존재함을 나타내고, 1보다 작거나 같은 값일 때 적절한 품질로 평가한다.

M-통계량들에 대한 정의는 아래와 같이 요약할 수 있다.

- M1: 원계열에 대한 불규칙성분의 분산 비율
- M2: 원계열의 변동 중 정상화 부분에 대한 불규칙성분의 상대적 기여도
- M3: 추세순환성분과 불규칙성분의 전월비 변화율 비교
- M4: 최종 불규칙성분의 자기상관에 대한 통계량
- M5: 불규칙성분보다 추세순환성분 변화율이 더 큰 월간격 통계량
- M6: 계절성분 연간변화율과 불규칙성분 연간변화율 비교
- M7: 이동계절성과 안정 계절성 사이의 관계를 측정
- M8~M11: 계절성분의 변동성 관련 통계

M-통계량의 주요 중요도는 아래와 같은 순서로 알려져 있다.

$$M7 > M1, M2 > M6 > M8, M9 > M3, M5$$

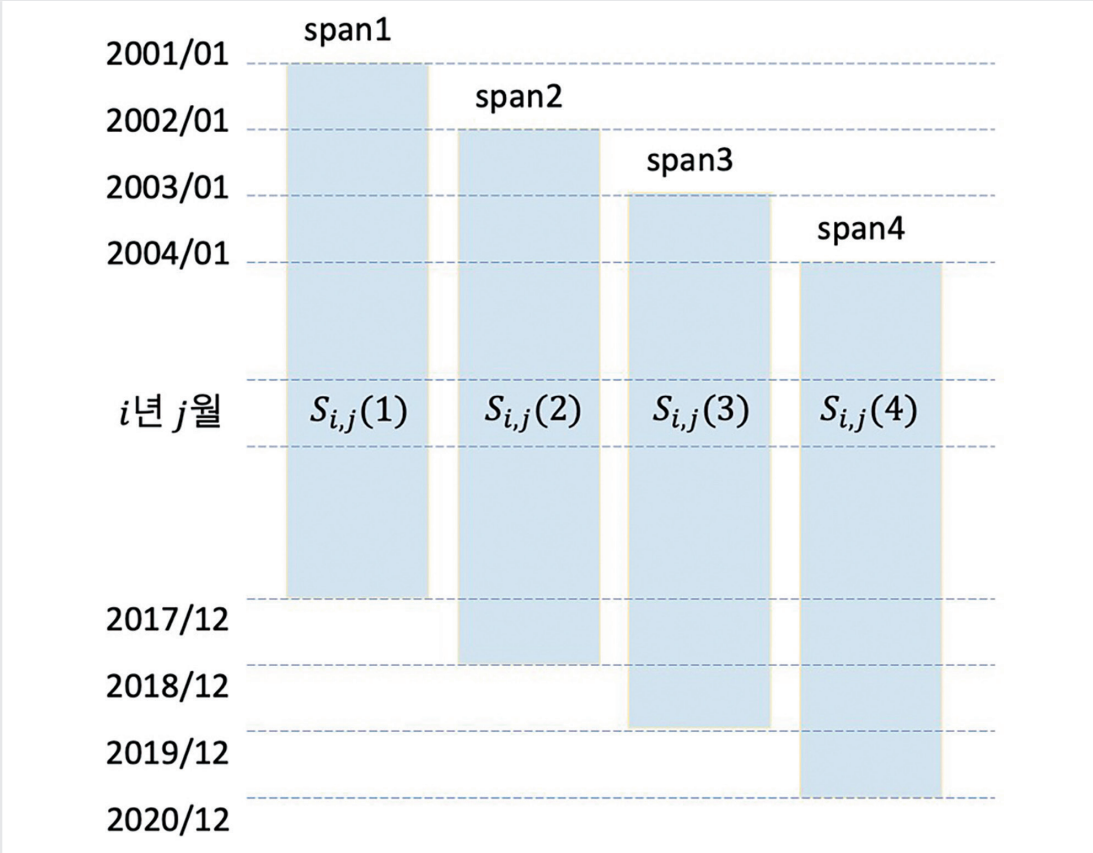
이 중에서, M7은 가장 중요한 M-통계량으로서 값이 클 경우 이동 계절성이 안정 계절성에 비해서 더 높음을 의미한다. 원계열의 계절성 존재 유무를 위한 검정에도 사용할 수 있다.

5.3 슬라이딩 스팬 분석

슬라이딩 스팬(sliding spans) 분석은, 〈그림 4〉와 같이 원계열로부터 만들어지는 (보통) 4개의 스팬에 의해서 계산된다. 스팬의 길이는 6~11년 사이에서 정해지며 이것은 원계열의 길이, 사용되는 계절 필터 및 계절 주기에 의해서 결정된다. 마지막 스팬은 원계열의 가장 최근 관측 기간을 포함하게 된다.

〈그림 4〉

슬라이딩 스팬 분석의 구조



출처: 이긍희(1998)

이제 각각의 스팬에서 계절조정작업을 수행한 후, 다음과 같이 2개 이상의 스팬에 따른 동일 시점(i년j월)에 대한 계절지수의 변동성과 전기비 증감률의 변동성을 측정한다.

$$sf(i,j) = \frac{\max_k S_{i,j}(k) - \min_k S_{i,j}(k)}{\min_k S_{i,j}(k)} \tag{2.19}$$

$$sa(i,j) = \max_k \frac{A_{i,j}(k)}{A_{i,j-1}(k)} - \min_k \frac{A_{i,j}(k)}{A_{i,j-1}(k)} \tag{2.20}$$

$S\%$와 $MM\%$를 다음과 같이 불안정한 계절조정 결과를 가지는 (i, j)의 개수가 정의할 때,

$$S\% = [sf(i,j) > 0.03]\text{인 } (i,j)\text{의 개수}$$
$$MM\% = [sa(i,j) > 0.03]\text{인 } (i,j)\text{의 개수} \tag{2.21}$$

$S\% < 25\%$ 그리고 $MM\% < 40\%$ 일 경우, 계절조정계열이 안정적이라고 판단할 수 있다.

5.4 리비전 히스토리 진단

리비전 히스토리 진단은 새로운 시계열 자료가 추가됨에 따른 계절조정계열이 얼마나 변화하는지를 측정한다. 다음과 같이 그 변동성을 MAPR(mean absolute percentage revision)로 정의한다.

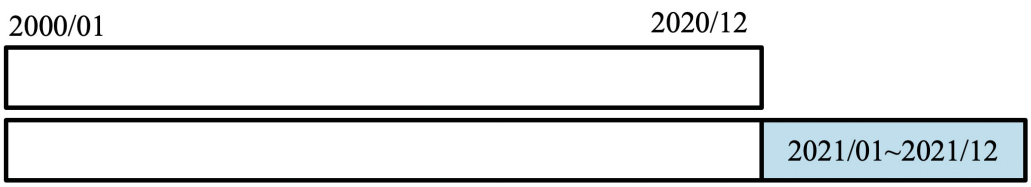
$$MAPR = \frac{100}{T - t_0 + 1} \sum_{t=t_0}^{T} \left| \frac{A_{t|T} - A_{t|t}}{A_{t|t}} \right| \tag{2.22}$$

MAPR은 그 값이 작을수록 계절조정의 안정성이 높은 것으로 평가할 수 있다. 단, $A_{t|n}$은 n 시점까지의 원계열을 이용한 t 시점의 계절조정계열을 나타낸다. t_0는 계절조정 프로그램에 의해서 자동선택되거나 적당한 시점을 부여할 수 있다. T는 원계열의 마지막 시점이다. 따라서, $A_{t|t}$는 t 시점에서의 동시적 조정법으로, $A_{t|T}$는 최종 계절조정계열로 볼 수 있다.

〈그림 5〉는 리비전 히스토리 진단의 구조의 예를 나타낸 것으로서, 이 경우에 리비전 히스토리의 시작시점 t_0는 2021/01이고, 원계열의 마지막 시점 T는 2021/12이다. 리비전 히스토리 진단은 두 개 이상의 계절조정방법을 비교할 때 (예: 직접법과 간접법의 비교), 편리하게 사용할 수 있다. 한 개의 방법에 대한 진단은 유용하지 않다.

〈그림 5〉

리비전 히스토리 진단의 구조 예



출처: 이긍희(1998)

Ⅲ. 빅이벤트 하에서의 계절조정

1. 빅이벤트와 계절조정

대규모 충격, 예를 들어 경기 침체 및 팬데믹은 계절조정 추정을 왜곡할 것으로 예상된다(Lucca and Wright, 2021; Dagum and Morry, 1985). 계절조정 절차가 대규모 충격의 영향을 계절성 변화와 정확하게 구분하지 못한다는 것이다. 이는 충격이 발생한 기간동안 뿐만 아니라 그 이후 기간의 계절성분 식별에도 영향을 미치게 된다. 팬데믹이나 대공황과 같은 일시적이지만 대규모의 경제 활동 및 물가 충격은, 이후의 시계열 자료에 허위 계절 패턴(spurious seasonal pattern), 즉 소위 충격의 메아리 효과(echo of the shocks)를 남길 수 있다.

거시경제 지표를 분석하는 여러 연구들에서 충격 이후의 시계열에서 잔여 계절성(residual seasonality)이 존재함을 보고하였다. Wright (2013)는 2008/2009년 금융 위기가 미국의 비농업 고용 데이터를 왜곡했음을 강조하였고, 이 효과가 처음에는 계절조정된 수치를 상반기에 약 10만 명 이상 상승시키고, 하반기에는 하락시켰음을 발견하였다. Lengermann et al. (2017)과 Rudebusch et al. (2015) 역시 계절조정된 GDP 성장률의 추정치가 잔여 계절성과 대공황의 영향으로 인해 편향되었음을 지적하였다.

가격과 관련된 사례에서는 Peneva and Sadee (2019) 및 Peneva (2014)는 미국 소비자 물가 상승률 지수(CPI)가 계절조정이 되었음에도 불구하고 연초에 비해 연말에 더 높아지는 경향이 있음을 보여주었다. 이들은 이를 잔여 계절성에 기인한 것으로 설명하였다.

2. 주요 국가 통계청들의 팬데믹 대응 방식

특히 팬데믹과 같이 경제 외부적 요인에서 비롯된 사건은 계절성 처리에 추가적인 복잡성을 유발한다. 주요 국가 통계청들(National Statistical Offices, NSOs)과 학계는 COVID-19 팬데믹 동안 발생한 혼란을 어떻게 처리할 것인지에 대해 의견이 일치하지 않았다.

예를 들어, 미국 노동 통계국(bureau of labor statistics, BLS)은 팬데믹 이전 및 전체를 각각 따로 계절조정 후 결합하는 2단계 방법을 사용하였으며 이를 통하여 계절조정의 수정(revision) 절차를 최소화하였다. 특히, 계절조정의 사전조정 단계에서 다양한 이상치(AO, LS, TC)를 모두 활용하였다.

유럽연합 EU의 통계기관 Eurostat는 팬데믹 충격을 가법 이상치(AO)로 모델링할 것을 권고

하였으며, AO 중심의 개입모형(intervention analysis) 사용을 보편화하고 필요시 LS와 TC를 병행하였다. 한편으로는 계절인자 계산에서 팬데믹 기간을 제외하여 그 충격이 계절성으로 확산되는 것을 방지하는 노력을 하였으며, 일부 국가에서는 팬데믹 발생 이전인 2019년 계절인자를 이후 연도에 고정 적용하기도 하였다.

호주 통계청(Australian bureau of statistics, ABS)은 의사통제(pseudo control, counter−factual) 기법을 도입하였다. 그들은 개입분석으로 빅이벤트가 부재한 상황 하의 정상 경로를 추정하여 팬데믹 기간의 실제치와 차이를 일시적 변화(TC)로 간주하여 제거하였다. 이런 방법이 AO의 단순 투입보다 충격 제거에 효과적이라고 보고하였다.

Bógalo et al. (2022)은 각국 NSOs가 COVID−19 충격을 처리하는 노력을 두 가지 주요 범주로 분류하였다; 첫번째는 가법 이상치 개입을 사용한 동시 계절조정(concurrent seasonal adjustment)이며, 두번째는 팬데믹 이전 해의 추정 계절성분을 투영(projection)하여 계절조정하는 방식이다. 그들은 팬데믹 이전 해의 계절 요인을 투영하는 방식이 AO 개입 처리보다 나쁜 선택이라는 결론을 내렸다.

Arend et al. (2024)에 의하면, 실질적으로 NSO들은 개입 모형을 사용할 때 더 세밀한 방식으로 조정을 시행할 수 있다고 보고하였다. 일부 NSO는 가법 이상치 대신 수준 이동(LS) 또는 일시적 변화(TC)와 같은 다른 이상치 처리 방식을 제안하였다. 또한, 이상치로 처리하는 기간에 대해서도 각국의 NSO들은 다른 판단을 하였다고 보고한다.

3. 빅이벤트 계절조정 방식

각 국가들의 빅이벤트와 계절조정 왜곡에 대한 공통된 대응 방식은, 대응은 반드시 필요하되 정해진 정답이 없다는 점이다. 이를 원문으로 표현하면 다음과 같다;

Seasonal adjustment is as much an art as it is a science.

본 절에서는 빅이벤트 하의 계절조정 방식을 다음의 3가지로 정리한다.

3.1 계절조정 프로그램의 자동 보정 방식에 맡기는 방식

이 방법은 주로 AO 및 LS 중심의 사전조정을 사용한다. 계절조정 프로그램(X−11 or SEATS)의 디폴트 기능 이외의 조정을 하지 않는다는 측면에서 매년 이루어지는 계절조정에 일관성과 편리함을 제고할 수 있으나, 빅이벤트 충격과 계절을 적절하게 분리하지 못할

때가 자주 발생할 수 있다.

3.2 빅이벤트 기간 동안은 모두 AO로 처리하는 방식

이 방법은 편리하지만 과적합(over-fitting)의 위험이 있다. 특히 계절성 변화(SO) 또는 단절이 존재할 경우 계절성분 추정을 왜곡할 수 있다. 따라서, AO 조정을 먼저 실시한 이후 나중에 LS 및 TC 조정을 통한 정교화 단계를 거치는 것이 바람직하다.

3.3 개입분석을 통한 빅이벤트 효과 추정을 사전조정에 반영하는 방식

개입분석으로 의사통제를 먼저 추정하고 사전조정 단계에서 이 결과를 설명변수로 반영하는 방법이다. 이를 위하여 빅이벤트의 영향을 외생변수 및 개입분석을 통하여 추정하여야 하기 때문에, 개입에 영향을 받지 않는 대상변수를 우선 찾아야 한다. 그러나 실제로는 대상변수를 발견하기가 쉽지 않고, 빅이벤트의 효과가 추정되었다고 하더라도 빅이벤트의 기간이 길고 분명하지 않을 경우 추정된 효과의 편차 또는 변동성이 커진다는 단점이 있다.

위와 같은 대응 방식에도 불구하고, 빅이벤트 계절조정은 자동 방식 및 모든 보정 방식에 있어서 사후 검토 과정이 반드시 요구된다. 빅이벤트 상황에서의 계절조정은 문헌에서도 지적되었듯이, 경제 전반에 대한 해석을 왜곡시켜 통계 이용자를 오도할 위험이 크기 때문이다.

4. 제안 알고리즘

계절조정 작업은 그 계열에 대한 해석이 큰 파급력을 가질 뿐만 아니라 매년 대규모 시계열에 대해서 실시된다는 점 그리고 수십년 간 그 메커니즘이 일관되게 정해져 왔다는 점을 유의하여야 한다. 따라서, 특별한 상황의 계절조정 방식을 고안한다고 하더라도 인위적이고 수동적인 프로세스를 가미하기는 쉽지 않다. 이런 관점에서 본 연구에서 제안하는 빅이벤트 하에서의 계절조정 알고리즘은, 다음의 3가지 사항을 중요하게 고려하였다.

(i) 각 국가 통계청들이 권고하는 방식들을 충분히 고려하여야 한다.
(ii) 대규모 시계열에 대한 계절조정 자동화가 가능하여야 한다.
(iii) 산출된 계절조정 계열에 대한 계절조정 품질진단이 가능하여야 한다.

본 연구에서는 위의 사항들을 고려하여 다음의 4단계를 통한 빅이벤트 계절조정 알고리즘을 제안한다.

[1-단계]

다음 4가지 옵션의 사전조정(RegARIMA)을 이용한 계절조정을 각각 실시한다.

- 자동보정(AO/LS) 방식
- 빅이벤트 기간에 대한 유의한 AO 반영
- 빅이벤트 기간에 대한 유의한 AO/LS/TC 반영
- 빅이벤트 기간에 대한 유의한 SO 반영

[2-단계]

1-단계에서 추정한 4가지 계절조정 모형에 대해 정보량 기준인 AICc(corrected Akaike Information Criterion)와 BIC(Bayesian Information Criterion)를 각각 계산하고, 이를 바탕으로 최적의 계절모형을 선택한다.

[3-단계]

2-단계에서 선택된 최적 모형을 빅이벤트 이후 기간을 포함하는 전체 기간에 적용한 계절조정 결과와 빅이벤트 이전 기간까지 적용한 결과를 상대적으로 비교하여 계절조정 품질 비교를 실시한다. 2장 5절에서 언급된 다양한 계절조정의 품질 진단 통계량을 사용할 수 있으나, 대표적으로 다음과 같은 진단 통계량의 변화 비율을 계산할 수 있다.

- Q-통계량의 변화비율을 통하여 불규칙 변동의 정도 및 계절성의 안정성 등의 평가한다.
- 슬라이딩 스팬 분석 변화 비율을 통하여 계절지수의 변동성을 평가한다.
- 리비전 히스토리 변화 비율을 통하여 계절조정의 변동성을 평가한다.

[4-단계]

다음 사항을 종합적으로 고려하여 빅이벤트 기간의 계절조정 결과를 재검토한다.

- 3-단계 분석에서 빅이벤트 기간의 포함으로 인해 계절조정 품질이 저하된 것으로 평가된 경우
- 빅이벤트 기간에 도입된 이상치 유형이 SO(Seasonal Outlier)로 식별된 경우
- 시각적 도구(예: season plot)를 활용하여 빅이벤트 기간의 계절적 움직임 및 변화를 점검할 필요가 있는 경우

5. 실증 분석 결과

본 절에서는 앞서 제안한 빅이벤트 하에서의 계절조정 알고리즘을 한국의 실질 GDP(생산측면) 48개 계열에 적용하여, 팬데믹과 같은 대규모 충격이 계절조정 결과(계절요인, 불규칙 성분, 리비전 등)에 미치는 영향을 점검하고 제안 알고리즘의 실무적 유효성을 검증한다. 분석 설계는 다음과 같다.

- 대상 계열: 한국 실질 GDP 관련 48개 계열(산업별 부가가치 및 순생산물세, 시장가격 GDP 포함)
- 자료 주기: 분기(quarterly) 자료
- 전체 분석기간: 1980년 1분기 ~ 2024년 4분기
- 빅이벤트 정의: COVID-19 팬데믹
- 코로나 발생기간(빅이벤트 기간): 2020년 1분기 ~ 2023년 4분기
- 코로나 이전기간(정상기간): 1980년 1분기 ~ 2019년 4분기
- 사용 프로그램: R package 'seasonal' version 1.10.0 (RegARIMA 기반 사전조정 및 X-13 계절조정 절차를 R 환경에서 구현)

〈표 3〉은 실질 GDP 48개 계열에 대해 다음 정보를 계열별로 정리한 결과이다.

(1) 코로나 이상점(최적 후보군): 4개 후보 중 AICc · BIC 기준으로 최종 선택된 사양

(2) 이상점 종류 및 시점: 팬데믹 기간(2020Q1~2023Q4)에서 유의하게 포함된 AO/LS/TC/SO의 분기

(3) 코로나 이전 대비 품질진단: 팬데믹 기간 포함으로 인해 계절조정 품질이 얼마나 변했는지를 다음의 3개 지표로 계산

① 통계량 변화비율: 불규칙 변동의 정도, 계절성 안정성 변화 등을 반영

② 슬라이딩 스팬(sliding spans) 변화: 계절지수의 스팬별 변동성(안정성 저하 여부)

③ 리비전 히스토리(revision history) 변화: 계절조정치의 개정 변동성(리비전 안정성)

〈표 3〉에서 별표(*)는 코로나 이전기간 대비 품질이 악화된 것으로 판단되는 경우(권고기준 초과)를 의미하며, NA는 진단 통계량이 해당 계열/설정에서 산출되지 않은 경우를 의미한다. 즉, 이 표는 단순히 식별된 이상점의 개수를 제시할 뿐만 아니라, 어떤 유형의 충격(AO/LS/TC/SO)이 산업별로 더 적합했는지, 그리고 그럼에도 불구하고(또는 그 결과로) 계절조정 품질이 악화되었는지를 동시에 보여주는 요약표이다.

〈표 3〉의 “코로나 이상점” 열을 기준으로 48개 계열을 분류하면 다음과 같다.

- 자동보정: 25개 계열 (52.1%)
- AO: 5개 계열 (10.4%)
- AO/LS/TC: 13개 계열 (27.1%)
- SO: 5개 계열 (10.4%)

이를 해석하면, 과반수(52.1%)의 계열에서는 정보량 기준상 자동보정이 가장 적합한 것으로 나타났다. 이는 해당 산업이 팬데믹의 영향을 받지 않았다는 의미가 아니라, 추가적인 개입변수를 명시적으로 도입하지 않거나 기본 탐지 범위 내에서 처리하더라도 통계적으로 가장 균형적인 적합도를 보였다는 뜻이다. 반면, 나머지 47.9%의 계열은 팬데믹 기간에 대해 명시적인 개입이 필요한 것으로 나타났다. 특히 AO/LS/TC 유형의 비중(27.1%)이 가장 높게 나타났는데, 이는 팬데믹 충격이 산업별로 단순한 일회성 이상치(AO)에 그치지 않고, 수준 이동(LS)이나 감쇠 구조를 갖는 일시적 변화(TC)를 동반한 경우가 적지 않음을 시사한다. 또한 SO 유형(10.4%)이 선택된 계열이 존재한다는 점도 주목할 필요가 있다. 이는 일부 산업에서 팬데믹이 단순한 수준 충격을 넘어 계절 패턴 자체를 왜곡하거나 변화시켰을 가능성, 즉 잔여 계절성이나 계절 구조 변화를 유발했을 가능성을 보여준다.

실증분석 결과, 코로나 이전 기간(1980Q1~2019Q4)과 비교하여 팬데믹 기간을 포함할 경우 품질진단 값이 1을 초과하여(빅이벤트로 인한 상대적 품질 저하가 관측되어) 재검토가 필요한 계열은 〈표 3〉의 3번, 6번, 21번, 27번, 31번으로 확인되었다. 이들 계열은 빅이벤트 구간을 포함함에 따라 불규칙 변동의 확대, 계절지수의 안정성 저하, 또는 리비전 변동성 확대가 상대적으로 크게 나타날 가능성이 있으므로, 4-단계 사후 점검을 우선적으로 적용할 필요가 있다.

한편, 계절변화(Seasonal Outlier, SO) 유형의 이상치가 식별된 계열은 7번, 21번, 30번, 32번, 44번으로 나타났다. SO는 빅이벤트가 단순한 수준 충격(AO/LS/TC)을 넘어 특정 계절 패턴 자체를 일시적으로 왜곡하거나 변화시켰을 가능성을 시사한다. 따라서 해당 계열에 대해서는 season plot 등 시각적 도구를 활용한 계절 패턴 점검과 함께, AO/LS/TC 중심 모형과 SO를 포함한 모형을 비교하여 최종 계절조정 결과를 재검토할 필요가 있다.

〈표 3〉

계열별 코로나 이상점 및 품질진단 결과

계열번호(실질)	코로나 이상점	이상점 종류 및 시점	코로나 이전 대비 품질진단
1: 농림어업	자동보정		1.09*, 31.25*, 1.00
2: 광공업	AO/LS/TC	TC2020.2	1.07*, 0.00, 0.90
3: 광업	자동보정		1.25*, 2.17*, 1.12*
4: 제조업	AO/LS/TC	TC2020.2	1.07*, 0.00, 0.95
5: 음식료품제조업	자동보정		1.38*, 0.00, 1.08*
6: 섬유및가죽제품제조업	AO/LS/TC	TC2020.2	1.16*, 1.03*, 1.16*
7: 목재종이인쇄및복제업	SO	SO2023.1	0.96, 55.56*, 1.22*
8: 코크스및석유정제품제조업	AO/LS/TC	LS2020.1, LS2021.1, AO2021.4, LS2022.4	1.50*, 3.50*, NA
9: 화학물질및화학제품제조업	AO	AO2020.2	1.42*, 0.00, 1.25*
10: 비금속광물제품제조업	자동보정		1.53*, 0.07, 1.42*
11: 1차금속제조업	AO	AO2020.2, AO2022.3, AO2022.4	1.08*, 0.00, 1.16*
12: 금속가공제품제조업	자동보정		0.81, 0.62, 1.07*
13: 컴퓨터전자 및 광학기기제조업	자동보정		0.96, 62.50*, 1.09*
14: 전기장비제조업	자동보정		0.98, 0.33, 1.05*
15: 기계및장비제조업	자동보정		1.13*, 0.00, 1.33*
16: 운송장비제조업	AO/LS/TC	TC2020.2	1.10*, 31.25*, 0.89
17: 기타제조업및산업용 장비수리업	자동보정		1.05*, 0.00, 1.04*
18: 전기가스및수도사업	AO	AO2020.1, AO2020.2	0.91, 0.00, 0.89
19: 전기업	AO/LS/TC	LS2023.2, LS2020.1	1.11*, 0.75, 1.10*
20: 가스증기 및 공기조절공급업	자동보정		0.00, 2.00*, 1.01*
21: 수도하수및폐기물 처리원료재생업	SO	SO2023.1, SO2023.3, SO2023.4	1.02*, 1.25*, 1.09*
22: 건설업	자동보정		1.21*, 0.00, 0.97
23: 건물건설및건축보수업	자동보정		0.95, 0.00, 0.95
24: 주거용건물건설업	자동보정		0.89, 1.07*, 0.95
25: 비주거용건물건설업	자동보정		0.00, 62.50*, 0.93
26: 건축보수업	자동보정		0.92, 2.00*, 0.93
27: 토목건설업	자동보정		1.03*, 93.75*, 1.01*
28: 서비스업	AO/LS/TC	LS2020.1, AO2020.1, AO2022.1	1.03*, 1.04*, NA
29: 도소매및숙박음식업	AO	AO2020.1, AO2020.2, AO2020.3, AO2020.4, AO2021.1, AO2021.2, AO2021.3, AO2022.1, AO2022.4	1.23*, 0.00, 1.07*

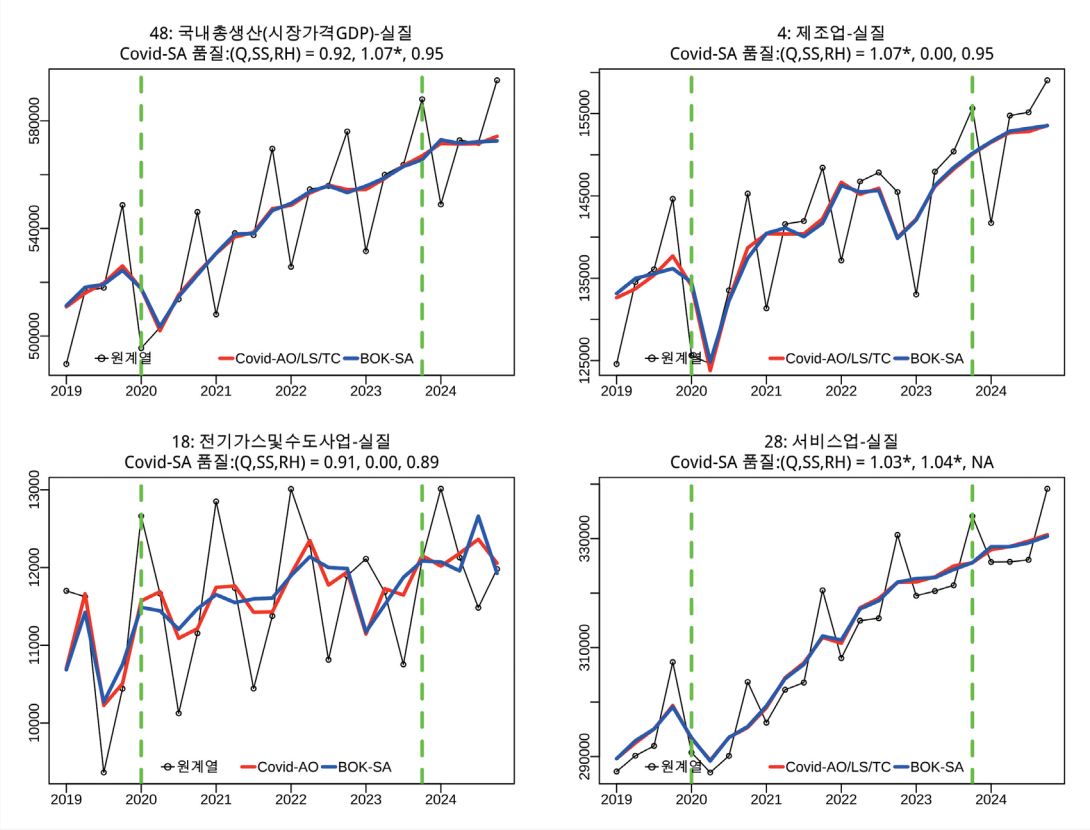
계열번호(실질)	코로나 이상점	이상점 종류 및 시점	코로나 이전 대비 품질진단
30: 도소매업	SO	LS2023.1, SO2023.1, SO2023.2, SO2023.3, SO2023.4	0.84, 0.71, NA
31: 숙박및음식점업	AO/LS/TC	LS2020.1, TC2020.4, LS2021.2, AO2021.4, LS2022.2	1.33*, 31.25*, 1.29*
32: 운수업	SO	AO2020.1, LS2020.1, SO2023.1	0.83, 0.00, 1.03*
33: 금융및보험업	AO/LS/TC	LS2020.2, TC2023.4	1.13*, 0.94, 0.88
34: 부동산업	자동보정		0.88, 0.00, 0.90
35: 정보통신업	자동보정		0.85, 0.00, 0.96
36: 통신업	자동보정		0.91, 0.00, 0.87
37: 출판방송영상정보서비스	자동보정		0.95, 0.00, 1.41*
38: 사업서비스업	자동보정		1.09*, 0.00, 0.98
39: 전문과학및기술 관련서비스업	자동보정		0.94, 0.00, 0.97
40: 사업지원서비스업	자동보정		1.04*, 0.00, 0.95
41: 공공행정국방및사회보장	자동보정		0.95, 0.00, 0.97
42: 교육서비스업	AO/LS/TC	LS2020.1, TC2020.3	0.90, 0.00, 1.06*
43: 의료보건업 및 사회복지서비스업	자동보정		1.05*, 0.00, 0.97
44: 문화및기타서비스업	SO	AO2020.1, LS2020.2, SO2023.1	1.26*, 0.00, 1.08*
45: 예술스포츠 및 여가관련서비스업	AO/LS/TC	TC2020.1, LS2020.2	1.21*, 0.00, 1.15*
46: 기타서비스업	AO/LS/TC	LS2020.1	0.00, 0.00, 0.93
47: 순생산물세	AO	AO2021.2	0.93, 0.00, 0.95
48: 국내총생산(시장가격GDP)	AO/LS/TC	LS2020.1,TC2020.2	0.92, 1.07*,0.95

팬데믹 전후의 계절조정 결과가 실제로 어떻게 달라질 수 있는지를 직관적으로 확인하기 위해, 국내총생산(시장가격GDP), 제조업, 전기가스및수도사업, 서비스업의 최근 구간 시계열을 시각화하였다. 〈그림 6〉에서는 원계열(검정)과 함께, 본 연구의 빅이벤트 반영 계절조정 결과(적색) 및 한국은행(BOK) 계절조정 계열(청색)을 함께 제시하였다. 또한 각 그림에는 팬데믹 기간에 추정된 이상점(AO/LS/TC/SO)의 시점을 표시하여, 이상점 처리 방식이 계절조정 계열의 단기 경로 및 수준에 영향을 줄 수 있음을 확인할 수 있도록 하였다.

다만, BOK 계절조정 결과는 적용 표본 구간이 상이(예: BOK는 1960~2024, 본 연구는 1980~2024)하고, 계절조정 프로그램 또한 서로 다름(본 연구: R seasonal 기반 X-13, BOK: 한국형 X-13)으로 인해, 두 결과를 정량적으로 1:1 직접 비교하여 우열을 판단하는 것은 적절하지 않다. 따라서 본 비교는 "동일 계열이라도 빅이벤트 구간의 이상점(특히 LS/TC 또는 SO) 반영 여부에 따라 SA 경로가 달라질 수 있다"는 점을 참고적 · 보조적으로 확인하는 목적에 한정한다.

〈그림 6〉

계열별 팬데믹 이상점 반영 SA 비교



Ⅳ. 계층적 시계열 자료의 계절조정

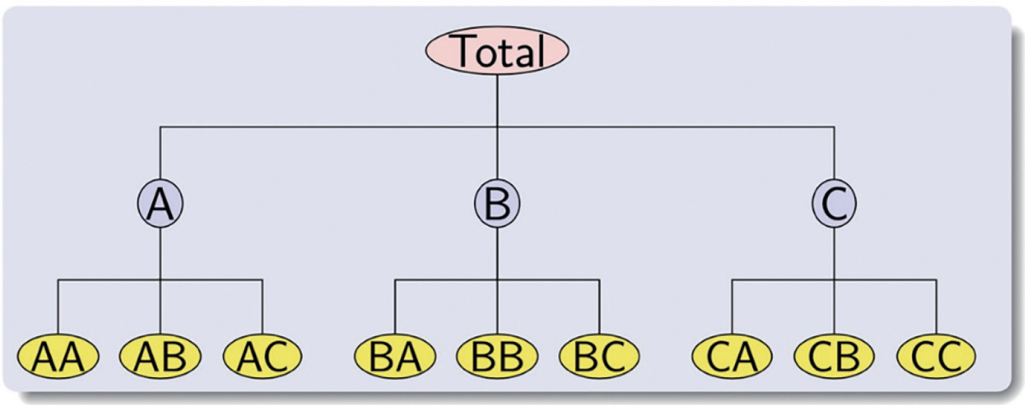
1. 계층적 시계열

다양한 분야에서 다양한 형태로 시계열 자료는 계층적으로 집계된다. 예를 들어, 지역별 관광 수요는 국가 전체의 관광 수요로 집계되며, 개별 제품 단위의 매출은 모든 제품의 총 매출로 집계된다. 또한, 한 나라의 국내총생산(GDP)은 개별 구성 요소들의 집계로 구성되며, 분기별로 측정된 시계열 자료는 반년 또는 1년으로 집계될 수 있다. 〈그림 7〉은 이러한 계층적 시계열 자료의 예시이다. 2개의 알파벳으로 구성된 최하위 수준의 시계열은 1개의 알파벳으로 구성된 상위 수준의 시계열로 합계되고, 이것들은 최상위 시계열인 Total로 합계 또는 집계된다. 이 때, 각 수준의 각 시계열들을 노드(node)라고 표현한다.

계층적 시계열 자료 분석 및 예측에서는 개별 시계열 간의 관계를 유지하면서 일치성(coherence)을 달성하는 것이 중요하다. 예를 들어, 3개의 시계열 X, Y, $Z = X + Y$에 대하여 적합 및 예측 모형이 각각 독립적으로 작동하거나 다른 조직의 데이터에 의존하면 계층적 구조의 제약을 충족하지 못하는 경우($\hat{Z} \neq \hat{X} + \hat{Y}$)가 발생하여 의사결정에서 심각한 문제가 발생하거나 적합 및 예측의 효율성이 저하되게 된다.

〈그림 7〉

계층적 시계열 자료의 예시



출처: Hyndman and Athanasopoulos(2018)

2. 계층적 시계열에서의 문제와 해결 방법

계층적 시계열 자료 구조에서의 문제점은 다음과 같다.

(1) 모든 노드에서의 시계열을 적합(예측)할 때, 그 결과가 원자료의 집계 방식과 동일하게 합산 일관성(add-up consistency)을 유지하려면 어떻게 해야 하는가

(2) 서로 연결된 시계열 간의 관계를 활용하여 적합(예측)의 정확성을 높일 수 있는가?

위와 같은 문제점들은 시계열 예측 분야에서 활발하게 연구되어 왔다. 관련 연구를 예를 들면, Hyndman et al. (2011), Athanasopoulos et al. (2017), Wickramasuriya et al. (2019) 등이 있다. 이들이 제안한 중요 해결책들을 요약하면 공통적으로 다음과 같은 알고리즘을 가진다.

- 먼저, 모든 집계 수준(aggregated levels)의 시계열에 대해 각각 모형을 적합한다.
- 그 다음, 다양한 최적화(optimization) 기법을 이용하여 적합 결과를 조정(reconciliation)하여, 원래의 집계 구조와 일치하도록 한다.

즉, 관련 연구들의 방법은 원래의 적합치에 가까우면서 동시에 합산 일관성을 만족하는 조정된 적합치(reconciled fits)를 찾는 작업으로 요약될 수 있다.

계층적 시계열 자료를 위한 표기법은 다음을 따르기로 한다. 우선, y_t는 시점 t에서 각 시계열(aggregate)의 관측치(스칼라)를 나타내며, $y_{X,t}$는 시점 t에서 노드 X의 관측치이다. $\underline{b}_t$는 시점 t에서 최하위(bottom) 수준 K에서의 모든 시계열 벡터를 나타낸다. 수준(level)은 최상위 수준과 K개의 하위 수준(0, 1, 2, ⋯, K)이 있다고 가정한다. 또한, 각 수준의 시계열의 개수를 합하면 n이 되어 다음이 성립한다; $n = n_0 + n_1 + \cdots + n_k$.

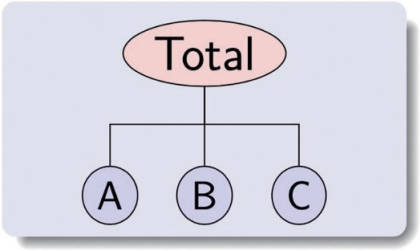
모든 집계 제약(aggregation constraints)을 가진 시계열 집합은 다음과 같이 표현할 수 있다;

$$\underline{y}_t = \mathrm{S}\,\underline{b}_t\,. \tag{4.1}$$

여기서, $\underline{y}_t$는 시점 t에서의 모든 시계열을 모은 벡터이며, S는 집계 제약 조건을 의미하는 합산 행렬(summing matrix)이다.

〈그림 8〉

K = 1 인 계층적 시계열



예를 들어, 계층적 시계열의 구조가 〈그림 8〉와 같은 경우, 수식 (4.1)은 다음과 같이 구체적으로 표현될 수 있다;

$$\underline{y}_t = \begin{pmatrix} y_t \\ y_{A,t} \\ y_{B,t} \\ y_{C,t} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \mathrm{S}\,\underline{b}_t . \tag{4.2}$$

3. 최적 결합 적합

계층적 시계열에서의 문제를 해결하기 위해서는 먼저 $\hat{\underline{y}}_t$과 같은 초기 적합치(initial fits) 벡터가 필요하다; $\hat{\underline{y}}_t$는 $\underline{y}_t$와 동일한 순서로 적층(stacked)되어 있다고 가정한다. 계절조정 문제에서 초기 적합치란, 계층적 시계열에 대하여 직접법 또는 간접법 등으로 만들어진 최초의 계절조정 계열이며, 일반적으로 이들은 4장 1절에서 언급된 계층적 시계열 자료 구조에서의 문제점을 가지는 계열들이다. 직접법에 의할 경우 계절조정 계열들은 집계 제약을 만족하지 않을 것이며, 간접법에 의할 경우 집계 제약은 만족하지만 하위 수준에서 낮은 품질을 가진 계절조정 결과가 상위 수준에 그대로 전달될 가능성이 크게 된다.

최적 결합 적합을 유도하기 위한 모형은 다음으로 가정한다;

$$\hat{\underline{y}}_t = \mathrm{S}\,\underline{\beta} + \underline{\epsilon}\,,\ \ \underline{\epsilon}_t \sim \mathrm{WN}(\underline{0}, \mathrm{W}) . \tag{4.3}$$

위의 수식은, 계층적 시계열에서의 문제점을 해결할 수 있는 미지의 벡터 $\underline{\beta}$가 존재하고 이들이 합산행렬 S에 의하여 그리고 백색잡음(white noise)이 더해져 초기 적합치 벡터 $\hat{\underline{y}}_t$가 생성되었다고 고려하는 것이다. 이러한 내재적 모형 가정은 합산 일관성을 만족하는 조정된

적합치 $\tilde{\underline{y}}_t$를 GLS(generalized least squares) 추정법에 의하여 다음의 형태로 유도할 수 있도록 한다;

$$\tilde{\underline{y}}_t = \mathrm{S}\,\hat{\underline{\beta}} = \mathrm{SP}\,\hat{\underline{y}}\,. \tag{4.4}$$

즉, 추정된 $\underline{\beta}$는 다음과 같이 표현될 수 있다;

$$\hat{\underline{\beta}} = (\mathrm{S}'\mathrm{W}^{-1}\mathrm{S})^{-1}\mathrm{S}'\mathrm{W}^{-1}\,\hat{\underline{y}}_t = \mathrm{P}\,\hat{\underline{y}}_t\,. \tag{4.5}$$

이러한 적합을 흔히 최적 결합 적합(optimal combination fits)이라고 한다. 여기서, P는 초기 적합치 $\hat{\underline{y}}_t$로부터 최하위 계열 적합치를 추출하여 결합하는 행렬로 이해할 수 있으며, S는 최하위 적합치들을 합산해 전체 구조를 재구성하는 역할을 하게 된다.

특수한 경우로서, $\mathrm{P} = [\,\mathrm{O} \mid \mathrm{I}_{n_K}\,]$ 이면 간접법(bottom-up)에 의한 계절조정을 의미하게 되며, $\mathrm{P} = [\,\underline{p} \mid \mathrm{O}\,]$ 이면, Top-down 방식의 계절조정을 의미한다; 여기서, $\underline{p} = [\,p_1, p_2, \cdots, p_{n_K}\,]'$는 상위 집계치의 적합을 최하위 시계열로 배분하는 역할을 한다. 이러한 특수 경우에 대한 문제점으로, 두 경우(bottom-up/top-down) 모두에서 일부 낮은 품질 시계열이 전체 적합치에 과도한 영향을 줄 수 있다는 것이다. 즉, 최적 결합 적합은 하위 수준이든 상위 수준이든 특정 적합치의 영향을 완화시킬 수 있다.

계층적 시계열을 위한 예측 이론에서는, 수식 (4.4)-(4.5)와 같은 최적 결합 적합의 장점으로 다음의 3가지를 강조한다.

- 초기 적합치 $\hat{\underline{y}}_t$가 불편(unbiased)이고 $\mathrm{SPS} = \mathrm{S}$가 성립한다면, $\tilde{\underline{y}}_t$ 또한 불편 추정량 성질을 가진다.
- 계층 내 모든 시계열의 정보를 균형있게 활용할 수 있다.
- 분산-공분산 구조를 고려하여 이론적으로 최소 평균제곱오차(mean squared errors)를 달성할 수 있다.

이제 남은 문제는 모형 (4.3)식의 오차항 공분산 행렬 $\mathrm{W} = \mathrm{cov}\,(\underline{\epsilon})$를 추정하는 것이다. 본 계절조정 연구에서는 예측 분야에서 흔히 사용하는 해결책을 차용하기로 한다.

첫번째 방법은, Hyndman et al. (2011)이 제시한 단위 행렬 가정이다. 즉, $\hat{\mathrm{W}} = \mathrm{I}_n$를 사용한다. 이 방법은 GLS 추정법을 단순히 OLS(ordinary least squares)로 단순화하여 사용하는 방법이다. 계산의 간편성 및 추정의 안정성 측면에 기여할 수 있다. 그러나, 계층적 이질성을 전혀 고려하지 못하는 단점이 있다.

두번째 방법은, Athanasopoulos et al. (2017)의 방법으로 흔히 구조적(structural) 가정이라고 일컫는다. $\hat{\mathrm{W}} = \mathrm{diag}\left(\mathrm{S}\,\underline{1}_{n_K}\right)$로 표현되며, 이는 각 노드의 시계열을 계산하기 위해서 하위

수준에서 합해진 시계열의 개수를 분산으로 간주하는 방법이다. 합계 시계열의 분산 또는 불확실성은 구성 시계열의 개수에 비례한다고 가정하는 것이다. 여기서, $\underline{1}_{n_K}$는 최하위(bottom) 수준 시계열 차원의 1-벡터를 나타낸다.

세번째 방법은 Hyndman et al. (2016)에 의한 것으로서 계절조정의 사전조정 회귀 RegARIMA 단계에서 추정된 불규칙 성분의 공분산 행렬($\widehat{\mathrm{W}}_*$)의 대각 원소만 사용하는 방법이다; $\widehat{\mathrm{W}} = \mathrm{I}_n \odot \widehat{\mathrm{W}}_*$. 첫번째 방법의 확장이라는 의미에서 WLS(weighted least squares) 방법이라고도 표현한다. 여기서, $\odot$는 Hadamard 곱의 기호로서 행렬의 원소별 곱을 의미한다.

네번째 방법은, 수식 (4.5)의 GLS 추정 형태를 그대로 이용하는 것이며 $\widehat{\mathrm{W}} = \widehat{\mathrm{W}}_*$로 표현할 수 있다. 이는 Wickramasuriya et al. (2019)에 의하여 제안되었으며 추정이 불안정하여 $\widehat{\mathrm{W}}$의 역행렬이 존재하지 않는 경우가 발생할 수 있다. 그러나, 계층적 이질성 및 계층간 상호 관계를 모두 고려하는 일반적인 형태이다.

이제, 본 연구에서 제안하는 계층적 계절조정의 최적 결합 알고리즘을 다음의 3단계로 표현한다.

[1-단계]

직접 계절조정을 실시하는 단계이다. 계층의 모든 수준에 대해 직접법(direct method)으로 계절조정을 수행하는 것이다.

[2-단계]

공분산 행렬 $\mathrm{W} = \mathrm{cov}\,(\underline{\epsilon})$을 위에서 정의된 4가지 방법으로 추정하여 $\widehat{\mathrm{W}}$를 구하는 단계이다.

[3-단계]

최적 결합(optimal combination)을 적합하는 단계로서, 1-단계의 계절조정 시계열들을 2-단계에서 추정한 $\widehat{\mathrm{W}}$를 활용하여 일관되게(reconcile) 결합하는 단계이다.

4. 실증 분석 결과

본 절에서는 제안한 계층적 계절조정의 최적 결합(optimal combination) 방법을 실제 국가계정 자료에 적용하여, 어떠한 형태의 조정(reconciliation)이 발생하는지를 확인하고자 한다. 이를 위해 명목 GDP 계층 자료를 예제로 제시한다. 명목 GDP는 실질 GDP와 달리 합산 일관성(add-up consistency)을 만족하는 시계열 자료라는 점에서 계층적 조정 효과를 분석하기에

적합하다. 다만, 본 분석의 목적은 계층적 계절조정 결과의 정답을 판별하는 데 있지 않다. 오히려 (i) 합산 일관성을 강제하는 과정에서 (ii) 각 노드의 계절조정 계열이 초기 적합치로부터 어느 정도, 그리고 어떤 방향으로 조정되는지를 실증적으로 확인하는 데 목적이 있다.

계층의 모든 수준(최상위 합계 및 산업 대분류 노드)에 대해 직접법(direct method)으로 계절조정하여 초기 적합치 $\hat{y}_t$를 생성한 뒤, 3절의 수식 (4.4)−(4.5)에 따라 최적 결합 적합 $\tilde{y}_t$를 산출하였다.

본 절에서는 공분산 행렬 $W = \mathrm{cov}(\epsilon)$의 추정 방식 중 구조적(structural) 가정에 기반한 경우(2번 방법)를 대표적으로 제시하며, 그림에서는 이를 OPT−2로 표기하였다. 이 방식은 각 노드가 하위 계열의 합으로 구성된다는 계층 구조를 명시적으로 활용함으로써, 계층 전체의 정보를 반영하는 동시에 과도한 불안정성을 완화하는 실무적 절충안으로 해석할 수 있다.

최적 결합 결과 $\tilde{y}_t$는 구성상 모든 시점에서 합산 제약을 만족한다. 즉, 하위 노드들의 계절조정계열을 합산하면 상위 노드의 계절조정계열과 정확히 일치한다. 따라서 본 장의 계층적 계절조정은 각 계열의 개별적 계절조정 품질에 더해, 부문 합계와 총합 간의 일치성을 동시에 확보하고자 하는 상황(예: 산업별 기여도 분석, 부문별 동학을 합산해 총량을 해석하는 경우)에 실무적으로 유의한 함의를 가진다.

〈그림 9〉은 대표적인 산업 대분류 노드인 농림어업, 광공업, 건설업, 서비스업에 대해 최근 2024년 분기 경로를 중심으로, 본 연구의 계층적 최적 결합 계절조정 결과(OPT−2, 녹색 실선)와 한국은행 계절조정 계열(BOK−SA, 적색 점선)을 함께 나타낸 것이다. 다만, BOK−SA은 적용 표본 구간, 사전조정 방식 및 계절조정 시스템(한국형 X−13 등)이 본 연구와 상이할 수 있으므로, 본 비교는 정량 비교가 아니라 "계층적 조정이 실제로 어떤 방향의 재배분을 유발할 수 있는지"를 보여주는 참고적 비교로 제한한다.

〈그림 9〉에서 공통적으로 관찰되는 특징은 다음과 같다.

(i) 분기별 변동의 재배분(quarterly redistribution)

OPT−2는 계층 합산 제약을 만족시키는 과정에서, 일부 산업에서 특정 분기의 수준을 낮추고 다른 분기의 수준을 높이는 방식으로 분기 경로를 미세 조정할 수 있다. 예컨대 농림어업에서는 1분기의 상대적으로 큰 수준이 완화되고 3분기 저점이 다소 상향되는 형태가 관찰되어, 분기 변동의 진폭이 상대적으로 완만해진다.

(ii) 급격한 말단 분기 변화의 완화 가능성

건설업의 경우 BOK−SA에서는 4분기 수준이 상대적으로 큰 폭으로 하락하는 모습이 나타나나, OPT−2는 하락 폭을 일부 완화하는 경로를 보인다. 이는 계층적 조정이 특정 노드의 단기 급변을 그대로 반영하기보다, 계층 전체 정보와 가중구조에 의해 변동

을 분산시키는 방향으로 작동할 수 있음을 시사한다.

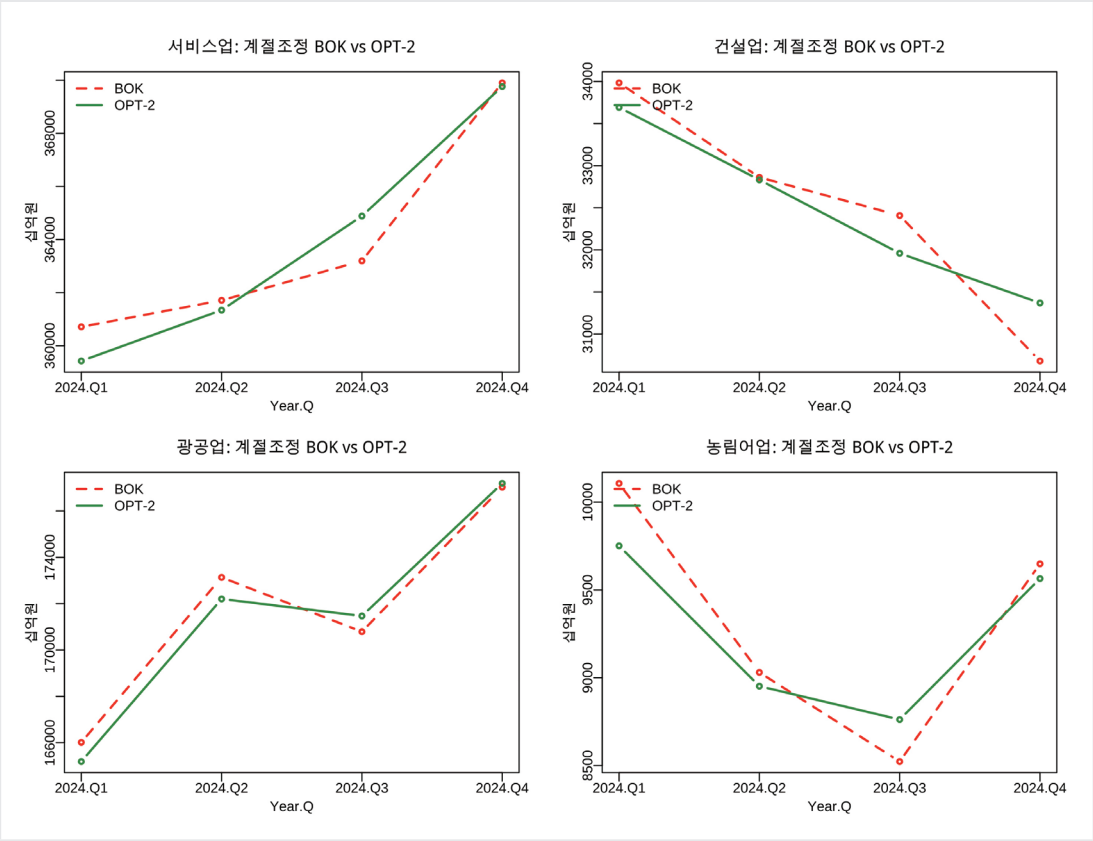
(iii) 산업별로 조정 방향과 크기가 상이

광공업과 서비스업에서도 분기별로 두 계열의 차이가 존재하나, 그 차이의 방향(어느 분기가 상향/하향되는지)과 크기는 산업별로 다르게 나타난다. 이는 계층적 결합이 일괄적인 평활화(smoothing)가 아니라, 계층 구조 및 노드별 변동성(가중치 구조)에 따라 조정이 차등적으로 발생한다는 점을 보여준다.

요약하면, 계층적 최적 결합 계절조정은 (i) 합산 일관성을 보장하는 동시에, (ii) 일부 노드에서 분기 경로를 소폭 재배분함으로써 단기 변동의 해석(예: 분기 성장률, 전기 대비 변화)에 영향을 미칠 수 있다. 따라서 계층 구조 하에서 부문별 계절조정 결과를 동시에 활용해야 하는 분석에서는, 각 계열을 개별적으로 계절조정한 결과를 단순 병렬적으로 사용하는 방식보다, 보다 일관된(coherent) 계절조정 체계를 제공한다는 점에서 활용 가치가 크다.

〈그림 9〉

계절조정계열(BOK-SA vs OPT-2) 비교(명목, 2024년)



Ⅴ. 요약 및 제언

1. 주요 연구 결과

최근 시계열 자료의 높은 변동성과 불규칙성은 안정성이 요구되는 계절조정 메커니즘에 보다 정교한 방법론을 요구하고 있다. 본 연구는 이러한 문제의식에 기초하여 두 가지 계절조정 고도화 기법을 제안하였다.

첫째, 빅이벤트 하에서의 계절조정 기법이다. 팬데믹이나 경제 대공황과 같은 경제 · 사회적 충격은 계절조정 계열의 추정을 왜곡하고 계절 성분 식별의 정확성을 저하시킬 수 있다. 이에 본 연구는 사전조정(RegARIMA) 단계에서 적용 가능한 네 가지 옵션을 제안하였다. 제안된 옵션은 각국에서 실제 활용되는 방법들을 효율적으로 결합한 형태이며, 모수 절약 원칙에 따라 통계적으로 유의한 이상치 시점만을 반영하도록 설계되었다. 또한 단순히 사전조정 옵션을 적용하는 데 그치지 않고, 빅이벤트 이전 기간의 모형 및 계절조정 품질진단과 비교하여 재검토가 필요한 계열을 선별하는 절차를 포함하였다. 특히 활용 방법이 명확히 정립되어 있지 않으나 중요한 의미를 갖는 계절적 이상치(Seasonal Outlier, SO)를 체계적으로 고려하였다는 점도 본 연구의 특징이다.

둘째, 계층적 시계열 자료에 대한 계절조정 기법이다. 대부분의 주요 거시경제 변수는 계층적 구조를 가지므로, 개별 시계열 간 합산 일관성을 유지하면서도 계절조정의 정확성을 확보하는 방법이 요구된다. 본 연구는 이를 위해 최적 결합(optimal combination) 적합 기법을 제안하였다. 이를 통해 직접법과 간접법의 장점을 동시에 활용할 수 있으며, 특정 하위 계열이 상위 계열의 계절조정 결과에 미치는 부정적 영향을 완화할 수 있다. 아울러 직접법과 간접법의 관계를 하나의 통합된 모형 체계 안에서 일반화할 수 있는 이론적 토대를 마련하였다.

2. 후속 연구를 위한 제언

본 연구에서 제안한 계절조정 고도화 기법은 추가적인 후속 연구를 통해 더욱 정교화될 수 있다.

첫째, 국내총생산(GDP), 실업률, 산업생산지수 등 다양한 실제 거시경제 자료에 적용하여

실증적 유효성을 지속적으로 검증할 필요가 있다. 이를 통해 이론 단계에서 충분히 고려되지 못한 요소를 보완하고, 각 시계열의 특성에 따른 개선 방향을 도출할 수 있을 것이다.

둘째, 제안된 알고리즘이 즉시 활용 가능하도록 독립적인 프로그램 형태로 개발될 필요가 있다. 나아가 빅이벤트 대응 기법과 계층적 계절조정 기법을 통합한 단일 시스템으로 구현한다면, 빅이벤트 하의 계층적 시계열에 대한 일관된 계절조정이 가능해질 것이다.

셋째, 최근의 빅이벤트는 과거보다 충격의 강도가 크고 구조적 변화의 성격을 띠는 경우가 많아 계절적 이상치(SO)의 빈번한 발생을 초래하고 있다. 이에 대한 탐지 방법과 모형화 전략을 체계적으로 정립하는 연구가 필요하다.

넷째, 본 연구에서 활용한 계절조정 품질진단 통계량을 보다 다양화하고, 진단 절차의 매뉴얼화 및 구체적인 기각 기준 설정에 대한 연구가 요구된다. 이는 실무 적용의 객관성과 재현 가능성을 높이는 데 기여할 것이다.

마지막으로, 본 연구의 계층적 계절조정은 계층 구조를 주로 공간적(산업 · 부문) 개념으로 한정하였다. 향후에는 이를 시간적 계층 개념까지 확장하는 연구가 필요하다. 예를 들어 월별 자료를 분기 자료로, 분기 자료를 반기 및 연간 합계 자료로 연결하는 구조를 고려할 수 있다. 이와 같은 시공간적 계층 구조를 반영한다면, 계절조정의 정합성과 정교함을 한층 더 제고할 수 있을 것으로 기대된다.

참고문헌 및 사이트

- 여인권, 성병찬, 김시현 (2022). 가계동향조사 계절조정계열 연구. 통계청.
- 이긍희 (1998). 한국경제시계열의 계절조정방법(X−12−ARIMA법을 중심으로). 경제분석, 4, 205−242.
- 전백근 (2002). 산업생산통계의 계절변동조정방법. 통계분석연구, 7(1), 1−48.
- Arend, T., Botev, J., Guidetti, E. Mourougane, A. and Park, M. (2024). Seasonal adjustment of CPIs during the COVID−19 pandemic and beyond. OECE Statistics Working Papers 2024/04.
- Athanasopoulos, G., Hyndman, R.J., Kourentzes, N. and Petropoulos, F. (2017). Forecasting with temporal hierarchies. European Journal of Operational Research, 262(1), 60−74.
- Bell, W.R. and Hillmer, S.C. (1983). Modeling time series with calender variation. Journal of the American Statistical Association, 78, 526−534.
- Bógalo, J., Llada, M., Poncela, P. and Senra, E. (2022). Seasonality in COVID−19 times. Economics Letters, 211(110206).
- Dagum, E.B. (1988). The X−11 ARIMA/88 Seasonal Adjustment Method− Foundations and User's Manual, Statistics Canada, Ottawa.
- Dagum, E.B. and Morry, M. (1985). Seasonal adjustment of labour force series during recession and non−recession periods. Statistics Canada Survey Methodology, 11(2), 133−144.
- Hyndman, R.J., Ahmed, R.A., Athanasopoulos, G. and Shang, H.L. (2011). Optimal combination forecasts for hierarchical time series. Computational Statistics and Data Analysis, 55(9), 2579−2589.
- Hyndman, R.J. and Athanasopoulos, G. (2018). Forecasting: principles and practice, 2nd Ed, OTexts.
- Hyndman, R.J., Lee, A. and Wang, E. (2016). Fast computation of reconciled forecasts for hierarchical and grouped time series. Computational Statistics and Data Analysis, 97, 16−32.
- International Monetary Fund. (2018). Quarterly National Accounts Manual: 2017 Edition. Washington, DC: International Monetary Fund.
- Lengermann, P., Morin, N., Paciorek, A., Pinto, E. and Sahm, C. (2017). Another look at residual seasonality in GDP. FEDS Notes, No. 2031, July.

- Lucca, D. and Wright, J. (2021). Reasonable seasonals? Seasonal echoes in economic data after COVID-19. Liberty Street Economics 20210325, Federal Reserve Bank of New York.
- Peneva, E. (2014). Residual seasonality in core consumer price inflation. FEDS Notes, Washington: Board of Governors of the Federal Reserve System.
- Peneva, E. and Sadee, N. (2019). Residual seasonality in core consumer price inflation: An update. FEDS Notes 2019−02−12−2, Board of Governors of the Federal Reserve System (U.S.).
- Rudebusch, G.D., Wilson, D.J. and Pyle, B.D. (2015). Residual seasonality and monetary policy. FRBSF Economic Letters, 2015−27.
- Sax, C. and Eddelbuettel, D. (2018). Seasonal adjustment by X−13ARIMA−SEATS in R. Journal of Statistical Software, 87(11), 1−17.
- Wickramasuriya, S.L., Athanasopoulos, G., and Hyndman, R.J. (2019). Optimal forecast reconciliation for hierarchical and grouped time series through trace minimization. Journal of the American Statistical Association, 114(526), 804–819.
- Wright, J.H. (2013). Unseasonal seasonals? Brookings Papers on Economic Activity, 44(2), 65−126.