

소비 관측 모형 개발 및 운영 방안 연구: 양파·마늘

이형용 · 한은수 · 최성천 · 정희주



소비 관측 모형 개발 및 운영 방안 연구: 양파·마늘

이형용 · 한은수 · 최성천 · 정희주



연구 담당

이형용 | 전문위원 | 연구 총괄, 제1, 5장 집필

한은수 | 전문연구원 | 제2~4장 집필

최성천 | 전문연구원 | 제3장 집필

정희주 | 연구원 | 제1장, 부록 집필

특별연구 E2026-02

소비 관측 모형 개발 및 운영 방안 연구: 양파·마늘

등 록 | 제6-0007호(1979. 5. 25.)

발 행 | 2026. 6.

발 행 인 | 한두봉

발 행 처 | 한국농촌경제연구원

우) 58321 전라남도 나주시 빛가람로 601

대표전화 1833-5500

인 쇄 처 | 더크리P&B (주)

I S B N | 979-11-6149-849-2 95520

※ 이 책에 실린 내용은 한국농촌경제연구원의 공식 견해와 반드시 일치하는 것은 아닙니다.

※ 이 책에 실린 내용은 출처를 명시하면 자유롭게 인용할 수 있습니다.

무단 전재하거나 복사하면 법에 저촉됩니다.

양파와 마늘은 우리 식생활에서 중요한 역할을 하는 대표적인 조미채소이므로 안정적인 수급 관리가 매우 중요하다. 이를 위해서는 생산과 공급뿐만 아니라 소비에 대한 이해가 필요하지만, 그동안 소비 부문에 대한 체계적인 정보는 상대적으로 부족하였다.

최근 다양한 부문의 농식품 소비 관련 자료가 축적되면서 소비 측면에서 농산물 수급을 살펴볼 수 있는 기반이 마련되고 있다. 이 연구는 양파와 마늘을 대상으로 농식품 소비정보 분석사업 자료를 활용하여 소비 관측 정보 생성을 위한 소비 부문별 예측모형을 구축 가능성을 시도하였다.

소비 관측은 아직 시작 단계로 볼 수 있으며, 자료 축적 기간 또한 충분하지 않은 상황이다. 따라서 이 연구는 완성된 관측 방법을 제시하기보다는 소비량 정보를 활용한 예측 가능성을 검토하고 향후 개선을 모색하기 위한 기초연구라는 점을 고려할 필요가 있다. 앞으로 소비 관련 자료가 계속 축적되고 분석 방법이 고도화된다면 농축산물 소비 정보는 농업관측정보의 중요한 축으로 활용될 수 있을 것으로 기대된다.

이 연구가 양파·마늘 소비에 대한 이해를 높이고 소비 정보를 활용한 농업관측의 고도화와 수급 관리 체계 개선을 위한 기초자료로 활용되기를 기대한다. 아울러 연구 수행에 힘쓴 연구진과 자료 구축 및 분석에 협조해 주신 관계자 여러분께 감사의 말씀을 전한다.

2026. 6.

한국농촌경제연구원장 한 두 봉

연구 목적

- 최근 농산물 소비가 가정 내 소비에 국한되지 않고 외식, 가공, 급식 등 다양한 부문에서 이루어지면서, 생산과 공급 중심의 기존 농업관측을 보완할 수 있는 소비 정보의 중요성이 높아지고 있음.
- 본 연구는 양파·마늘을 대상으로 ‘농식품 소비정보 분석사업’에서 축적된 실제 소비량 자료를 활용하여 소비 부문별 자료 특성에 적합한 예측모형을 개발하고, 모형의 지속적인 활용과 운영을 위한 방안을 마련하는 것을 목적으로 함.

연구 방법

- 본 연구는 수집된 양파·마늘 소비량 데이터의 특성에 맞추어 가구 소비량과 학교·군 급식 소비량은 시계열 분석을 수행하였으며, 산업 부문의 식품제조업, 외식업, 일반급식업은 패널 분석을 수행하였음.
- 가구 소비는 주별 시계열 자료를 활용하여 ARIMA, ARIMAX, ETS 모형을 적용하였으며, 식품제조업·외식업·일반급식업 부문은 월별·업체별 패널데이터를 통해 업체별 과거 평균 모형, FE 패널회귀모형, CatBoost 모형을 적용하였음. 학교·군 급식 부문은 행정자료 기반의 월별 시계열 자료로 계절성 단순 기법, ETS 지수평활, ARIMA, ARIMAX를 적용하였음. 본 연구에서 모형 예측 성능은 MAE, RMSE, MAPE를 중심으로 평가하였음.

연구 결과

- 가구 부문 소비 모형 추정 결과, 양파와 마늘 두 품목 모두에서 최근 52주 자료를 학습데이터로 이용하고 도매시장 반입량의 1~3주 시차 변수를 외생 변수로 포함한 선형보간 ARIMAX 모형이 비교 대상 모형 중 가장 낮은 예측 오차를 보였음. 이는 주별 소비량 예측에서 소비량 자체의 과거 흐름뿐만 아니라, 공급 측면의 물량 정보가 일정한 예측 정보를 제공할 가능성을 시사하며, 단기 소비량 예측에서 최근의 소비 패턴과 수급 여건을 반영하는 것이 중요할 수 있음을 보여줌.
- 산업 부문 양파·마늘 소비 모형 추정 결과, 일반 급식업과 외식업체에서는 양파·마늘 모두 원자료 기준 CatBoost 모형의 예측 성과가 가장 우수하였으며, 식품제조업에서는 양파는 원자료 기준 업체별 과거 평균 모형, 마늘은 원자료 기준 CatBoost 모형의 성과가 가장 우수한 것으로 나타났음.
- 학교·군 급식 부문에서는, 양파는 최근 소비량 수준과 계절 변동을 함께 반영하는 ETS 지수평활 모형의 성과가 가장 우수하였고, 마늘은 전년 동월 소비량을 활용하는 계절성 단순 기법의 성과가 가장 양호하였음. 이는 급식 소비량 예측에서 품목별 시계열 특성을 고려한 모형 선택이 필요함을 보여줌. 또한, 반입량 시차 변수를 포함하는 경우 양파는 일정 수준의 예측 성과를 보였으나, 마늘은 다른 결과를 보여 공급 측 물량 정보의 활용 가능성은 품목별로 다르게 적용될 필요가 있음을 보여줌. 급식 소비량 예측은 단일 모형을 일괄 적용하기보다 품목별 반복 패턴, 최근 수준 변화, 외생변수의 설명 가능성을 함께 비교하는 방식으로 접근할 필요가 있음.

결론 및 시사점

- 양파·마늘을 대상으로 실제 소비량 자료를 활용하여 소비 부문별 자료 특성에 적합한 예측모형을 추정한 결과, 소비 부문별 자료 구조와 소비 특성에 따라 상대적으로 우수한 예측 성능을 보이는 모형이 다르게 나타났으며, 단일 모형보다는 부문별 특성을 고려한 접근이 필요한 것으로 나타났음.
- 소비 모형은 단기 소비 흐름을 파악하고, 조사 공백 기간의 보조적 추정 및 기존 농업관측정보를 보완하는 정보로 활용될 수 있음. 또한, 본 연구는 양파·마늘을 대상으로 소비량 자료 기반 예측모형의 적용 가능성을 검토한 시범 연구로서 향후 소비 관측 품목 확대를 위한 기초 사례로 활용될 수 있음.
- 개발된 모형은 자료의 지속적인 갱신과 모형 재추정, 예측 성능 관리를 통해 운영할 필요가 있으며, 예측 결과는 부문별 소비 변화와 전체 소비 흐름을 파악하는 지표로 활용하되 기존 농업관측정보와 연계하여 종합적으로 해석하는 것이 바람직함.
- 다만, 활용 가능한 자료의 축적 기간이 충분하지 않은 만큼, 꾸준한 자료 축적과 결측치 처리 방법 개선, 모형의 지속적인 검증 및 고도화를 통해 소비 관측의 활용성을 계속 높여 나갈 필요가 있음.

제1장 서론

1. 연구의 필요성 및 목적	1
2. 선행연구 검토 및 차별성	7
3. 연구 내용 및 방법	13

제2장 가구 부문 소비 모형 추정

1. 분석 자료의 특성	23
2. 모형 구축 전략	27
3. 모형 추정 및 예측력 평가 결과	33

제3장 제조업·외식업·일반급식업 부문 소비 모형 추정

1. 분석 자료의 특성	41
2. 모형 구축 전략	47
3. 모형 추정 및 예측력 평가 결과	51

제4장 학교·군 급식 부문 소비 모형 추정

1. 분석 자료의 특성	57
2. 모형 구축 전략	60
3. 모형 추정 및 예측력 평가 결과	63

제5장 모형 운영 방안

1. 부문별 모형 분석 결과 요약	71
2. 모형 활용 및 운영 방안	78

3. 향후 과제	85
부 록	91
참고문헌	105

제1장

〈표 1-1〉 분석 방법 요약	22
------------------------	----

제2장

〈표 2-1〉 가구 소비량 자료의 관측 구조	25
〈표 2-2〉 품목별 가구 소비량의 기초통계량	25
〈표 2-3〉 결측 처리 방식별 예측모형 구성	30
〈표 2-4〉 결측 처리 방식별 예측모형 추정 결과	34
〈표 2-5〉 양파 가구 소비량 예측 정확도 평가 결과(N=37)	35
〈표 2-6〉 결측 처리 방식별 예측모형 추정 결과	38
〈표 2-7〉 마늘 가구 소비량 예측 정확도 평가 결과(N=37)	39

제3장

〈표 3-1〉 분석 대상 자료의 구성	42
〈표 3-2〉 품목·산업별 패널 구조와 소비량 관측 현황	43
〈표 3-3〉 품목·산업별 소비량 기초통계량	44
〈표 3-4〉 예측모형에 활용한 변수 구성	48
〈표 3-5〉 품목·산업별 학습 및 테스트 자료 구성	50
〈표 3-6〉 양파 산업별 소비량 예측 정확도 평가 결과	52
〈표 3-7〉 마늘 산업별 소비량 예측 정확도 평가 결과	54

제4장

〈표 4-1〉 품목별 학교·군 급식 소비량의 기초통계량	58
--------------------------------------	----

〈표 4-2〉 학교·군 급식 소비량 예측모형 구성	61
〈표 4-3〉 양파 학교·군 급식 소비량 예측모형 추정 결과	64
〈표 4-4〉 양파 학교·군 급식 소비량 예측 정확도 평가 결과(N=12)	65
〈표 4-5〉 마늘 학교·군 급식 소비량 예측모형 추정 결과	67
〈표 4-6〉 마늘 학교·군 급식 소비량 예측 정확도 평가 결과(N=12)	68

제5장

〈표 5-1〉 부문별 소비 예측모형 주요 추정 결과 요약	77
---------------------------------------	----

제1장

〈그림 1-1〉 양파·마늘 연간 생산 추이	2
〈그림 1-2〉 KREI-KASMO 채소류 수급 전망 모형의 구조 개념도	9
〈그림 1-3〉 일반적인 농식품 유통경로	17

제2장

〈그림 2-1〉 품목별 가구 소비량의 변화 추이	26
〈그림 2-2〉 양파 선형보간 ARIMAX 모형의 잔차 진단 결과	36
〈그림 2-3〉 마늘 선형보간 ARIMAX 모형의 잔차 진단 결과	40

제3장

〈그림 3-1〉 품목·산업별 소비량 분포	45
〈그림 3-2〉 품목·산업별 소비량 변화	46

제4장

〈그림 4-1〉 품목별 학교·군 급식 소비량의 월별 변화 추이	59
〈그림 4-2〉 양파 학교·군 급식 소비량 ETS 모형의 잔차 진단 결과	66
〈그림 4-3〉 마늘 학교·군 급식 소비량 ETS 모형의 잔차 진단 결과	69

제5장

〈그림 5-1〉 모형의 운영 절차	81
--------------------------	----

1

서론

1. 연구의 필요성 및 목적

1.1. 연구 배경 및 필요성

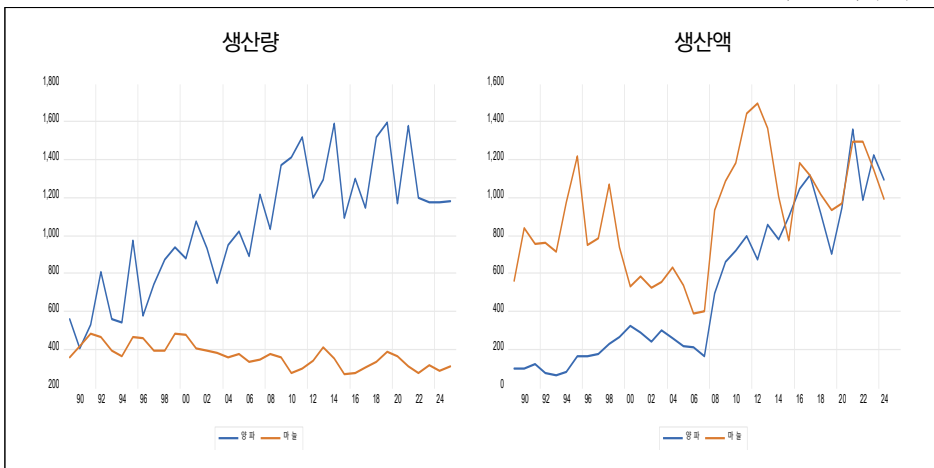
- 양파와 마늘은 한국 식생활 전반에서 거의 모든 조리의 기본 재료로 사용되는 핵심 농산물로 가정, 외식, 급식, 가공식품 전반에 걸쳐 소비 경로가 매우 넓은 품목 중 하나임.
- 양파와 마늘은 국내 채소 중에서도 생산 규모가 큰 품목으로 생산자의 주요 소득 작목으로 인식되고 있음.
 - 최근 5개년(2020~2024년산) 기준 양파와 마늘 생산량은 각각 125.6만 톤, 30.4만 톤, 공급량은 각각 137.6만 톤, 35.7만 톤이며, 생산이 감소해도 일정 수준의 수입량 유입으로 공급량이 보완되는 구조를 보이고 있음.
 - 양파와 마늘 생산액은 2024년 기준 각각 1조 942억 원, 9,910억 원으로

두 품목 생산액은 전체 재배업의 5.7%, 특히 조미채소 생산액의 절반(50.1%)을 차지하고 있음. 양파·마늘 생산액은 수급 여건에 따라 등락은 있었으나 장기적으로 증가하는 추세를 보이며, 주요 채소류 품목으로서의 중요성이 지속되고 있음.

- 양파와 마늘은 연 1회 생산되어 수확 이후 저장을 통해 연중 공급되는 구조이므로, 생산량 변화가 저장량 및 출하량 조절을 거쳐 가격 및 수급 불안정으로 연결될 수 있는 특성을 가짐.

〈그림 1-1〉 양파·마늘 연간 생산 추이

단위: 천 톤, 십억 원



자료: 국가데이터처(각 연도), 농작물생산조사; 농림축산식품부(각 연도), 농림업생산지수.

- 양파·마늘은 기상 여건, 재배면적 변화 등에 따라 생산량 변동이 크게 나타날 뿐만 아니라 저장 및 출하 시기 조절 과정에서도 가격 변동성이 확대될 가능성이 있음. 따라서 양파·마늘은 수급 불안시 언론이나 정책의 관심이 집중되어 정부의 개입이 반복되는 대표적인 정책 대상 품목임.

- 정부는 양파와 마늘의 중요성을 고려하여 배추, 무, 건고추 등과 함께 주요

수급 품목으로 지정해 소비지 물가 안정과 생산농가의 안정적 소득 확보를 위해 수급 안정 관련 다양한 정책을 수행하고 있음.

○ 양파·마늘처럼 생육기간이 비교적 긴 농산물의 경우 생산이 가진 내재적 불안정성으로 인해 특정 시기 수급 불안정 상황이 발생하게 되면 가격 변동성이 확대되고, 이는 결국 농가 소득 불안정과 소비자 부담 증가로 연결될 수 있는데, 농업관측정보는 이러한 영향을 줄이기 위한 노력의 하나로 이해할 수 있음(김관수 외, 2009).

- 농업관측사업은 주요 농축산물의 재배(의향)면적, 예상 생산량, 재고량, 수·출입량, 해외정보 등을 종합적으로 조사·분석¹⁾하여 향후 시장 상황을 전망함으로써 농축산물 수급 정책 수립과 시장 참여자의 의사결정을 지원하는 정책 정보 시스템으로 볼 수 있음.
- 현행 농업관측사업 대상 품목은 양파와 마늘을 포함한 36개 농축산물이며, 양파·마늘은 매월 수급 정보가 제공되고 있음.

○ 한편, 최근 식생활 및 식습관 변화에 따라 양파와 마늘 소비 구조가 가구 소비 중심에서 외식업, 제조업, 급식 등 다양한 소비 경로로 다변화되고 있음. 이에 기존의 공급 중심 접근만으로는 실제 소비 흐름을 충분히 설명하기 어려움.

- 양파와 마늘은 한국 사회에서의 전통적인 기초 식재료임에도 불구하고, 소비 방식의 변화가 뚜렷하게 나타나고 있으며, 이는 가격 및 수급 변동에 대한 시장 반응에도 영향을 미치고 있음.
- 그러나, 현재 농업관측정보는 재배면적, 작황, 출하량, 가격 전망 등 공급 중심 지표에 주로 의존하고 있어, 소비 변화가 양파·마늘 수급에 미치는 영향을 충분히 반영하는 데 한계가 있음.

¹⁾ 농업관측센터(<https://aglook.krei.re.kr/main/contents/officeIntro0301>), 검색일: 2026. 3. 27.

- 이러한 공급 및 가격 변동 특성과 함께 소비 구조가 다변화되면서 실제 소비 흐름을 직접 관측할 필요성이 확대되고 있음.
- 이에 정부는 최근 농산물 가격 변동성 완화 및 농가의 경영 안정을 위한 선제적 수급 관리 측면에서 소비 관측에 대한 필요성을 강조하고 있음.
 - 농식품부는 ‘한국형 소득·경영 안전망 구축 방안’, ‘농산물 유통구조 개선 방안’ 등을 통해 품목별 수급 안정을 위한 추진 과제로 선제적 수급 관리 체계 구축을 제시함(이형용 외, 2025: 2).
 - 정부는 농안법 개정을 통해 수요처별 소비량, 소비 실태 정보 등에 대한 조사·분석의 근거를 마련(2025. 8.)하고, 농축산물 소비 관측에 대한 운영 계획 및 방안을 마련하고 있음.
- 농축산물 소비 관측은 주요 농축산물에 대한 소비 관련 정보를 주기적으로 조사·수집·분석하여 소비 트렌드와 변화 등을 전망하여 그 결과를 합리적인 의사결정 지원을 위해 관련 수요자에게 제공하는 정보 기반 활동을 의미함.
- 농축산물 소비 정보 부족은 생산 및 유통 관계자의 의사결정과 정부의 선제적 수급 안정 정책 수립에 제약 요인으로 작용할 수 있어, 공급 정보 중심의 농업 관측 체계를 보완하고 주요 농축산물에 대해 소비 정보를 분석하는 소비 관측 도입 필요성이 증가하고 있음.
- 김상효 외(2016)는 기존 농업관측에 양질의 소비 관측 정보가 결합되어 제공되는 경우 농업관측정보만 제공되는 것보다 사회 후생을 추가로 증대(사회 후생 순손실 감소)시킴을 보였음.

- 농업관측정보는 사회 후생, 가격 변동성, 생산자 의사결정 측면에서 기여해 온 것으로 나타났는데(이용선 외, 2002; 김관수 외, 2009; 김배성 외, 2014; 김성훈 외, 2015; 서동주 외, 2022), 농업관측의 경제적 효과가 100%일 때 추가적인 소비 관측 효과는 39% 수준이며, 쌀, 배추, 양파, 한우, 사과 5개 품목 기준 소비 관측의 최대 사회 후생 효과는 467억 원 수준인 것으로 나타남(김상효 외, 2016: 52).

○ 기존 농축산물 소비 관련 정보는 농업관측센터 내에서 가구와 음식점을 중심으로 농축산물 구매 행태를 조사하는 형태로 진행되었음.

- 가구 조사의 경우 농업관측센터 자체 소비자 패널을 구축하여 ‘소비 의향 조사’나 ‘소비 실태 조사’ 형태로 진행되었음.
- 음식점 조사는 인구·사회적 변화와 식습관 변화를 고려해 서울 지역 330개 음식점 패널을 대상으로 농업관측 품목의 구매량과 구매 행태를 조사하였음.

○ 그러나, 기존 가구 및 음식점을 대상으로 한 조사는 조사 표본이 전체를 대표하기 어렵다는 점, 가구 내 소비 분석 시 실제 소비 실적이 아닌 구매 의향 데이터를 기반으로 삼았다는 점, 가공업체나 단체 급식, 유통업체 등 농산물이 소비되는 다각적인 경로를 종합적으로 반영하지 못했다는 점에서 한계가 있음(김상효 외 2016: 2-3).

○ 한국농촌경제연구원에서 수행하고 있는 ‘농식품 소비정보 분석사업’도 주요 농축산물에 대한 미래 정보 지표로 월별 소비 의향 정보를 지수 형태로 조사·제공하고 있긴 하지만, 조사 지수의 특성상 실제 소비와 차이가 발생할 가능성이 있으므로 보완적인 분석이나 연구가 병행될 필요가 있음이 지적되었음(이형용 외, 2025a).

- 이에 소비 관측 도입에 필요한 관측 방법들이 체계적, 점진적으로 마련될 필요가 있으며, 양파·마늘을 대상으로 다양한 소비 관측 기법이 개발·보완될 필요가 있음.
- 특히, 소비 관측이 아직 초기 단계에 있다는 점을 고려할 때, 실제 소비량 자료를 활용한 예측모형의 적용 가능성을 우선적으로 검토할 필요가 있음.

1.2. 연구 목적

- 본 연구는 양파·마늘을 대상으로 ‘농식품 소비정보 분석사업’에서 축적된 실제 소비량 자료를 활용하여 소비 부문별 자료 특성에 적합한 예측모형을 개발하고, 개발된 모형의 지속적인 활용·운명을 위한 방안을 제시하는 것을 목적으로 함.
- 단, 농축산물 소비 관측은 아직 초기 단계에 있으며 활용 가능한 자료의 축적 기간 또한 충분하지 않은 상황임.
- 따라서 본 연구는 특정 모형을 최종적인 표준모형으로 제시하기보다, 소비 자료의 특성에 적합한 예측모형을 개발하고 그 활용 가능성을 검토하는 과정에 있다고 볼 수 있음.
- 이러한 점에서 본 연구는 향후 소비 관측 체계 구축과 예측모형의 고도화를 위한 기초연구로서의 성격을 갖는다고 볼 수 있음.

2. 선행연구 검토 및 차별성

- 농산물 품목마다 생산, 유통, 소비 구조와 데이터 환경이 달라 기존 관측 체계를 그대로 적용하기 어려우므로 관측사업에서 새로운 품목이 도입될 때마다 품목 특성에 맞는 체계와 분석 방법을 설계해 왔음.
- 석현덕·민경택(2003), 신용광 외(2010), 허덕 외(2010)는 신규 관측 품목 도입 시 관측 방안과 기법에 대해 제시하였음.
 - 석현덕·민경택(2003)은 표고버섯 관측사업의 안정적인 정착을 목적으로 국내 표고버섯 산업의 전반적인 실태를 진단하는 한편, 향후 전개될 관측사업의 운영 프로세스와 추진 체계의 밑그림을 제안하였음.
 - 신용광 외(2010)는 농산버섯 시장의 안정적인 수급 관리 체계를 다지기 위해 산업 현황 분석부터 모형 구축에 이르는 종합적인 관측 프레임워크를 제안하였음.
 - 허덕 외(2010)는 오리 관측 분기보 발간 체계 수립을 목적으로 오리 산업 전반에 대한 다각적 검토를 수행하고, 예측력을 높이기 위한 오리 관측 모형 설계와 대표성 있는 표본농가 샘플링 작업을 병행함으로써 관측사업의 실행력을 담보할 수 있는 종합적인 방법론을 제안하였음.
- 앞서 제시한 품목별 관측 도입 방안이나 기법 연구들은 연구 대상이 생산이나 공급 단계에서의 정보를 기반으로 한 농업관측에 국한되어 있는 한계가 있음.
- 농축산물 소비 관측과 관련하여 김상효 외(2016), 이형용 외(2025b)는 소비 관측 도입 방안과 체계에 대해 제안하였음.
 - 김상효 외(2016)는 소비 관측 사업을 실질적으로 구동하기 위한 제도적·

기술적 요건을 정립하는 데 초점을 맞추었음. 이를 위해 활용 가능한 기존 소비 통계와 관측 방법론을 먼저 스크리닝하였으며, 실무적인 운영 체계 확립에 필수적인 법령 근거, 추진 조직 및 인력 수급 계획, 소요 재정 규모 등을 면밀하게 분석하여 구체적인 실행 로드맵을 제시하였음.

- 이형용 외(2025b)는 농축산물 소비 관측 체계 도입의 실효성을 담보하기 위한 도입 방향을 설정하고 주요 운영방안을 도출하였음. 이형용 외(2025b)는 농축산물 소비 관련 공공데이터, 현행 ‘농식품 소비정보 분석사업’ 데이터의 한계와 보완점을 제시하고, 농업관측사업과 민간 이커머스 서비스 업체의 농축산물 발주 체계를 분석하여 시의성과 신뢰성 있는 소비 관측 정보 생성을 위한 방안을 제시하였음.

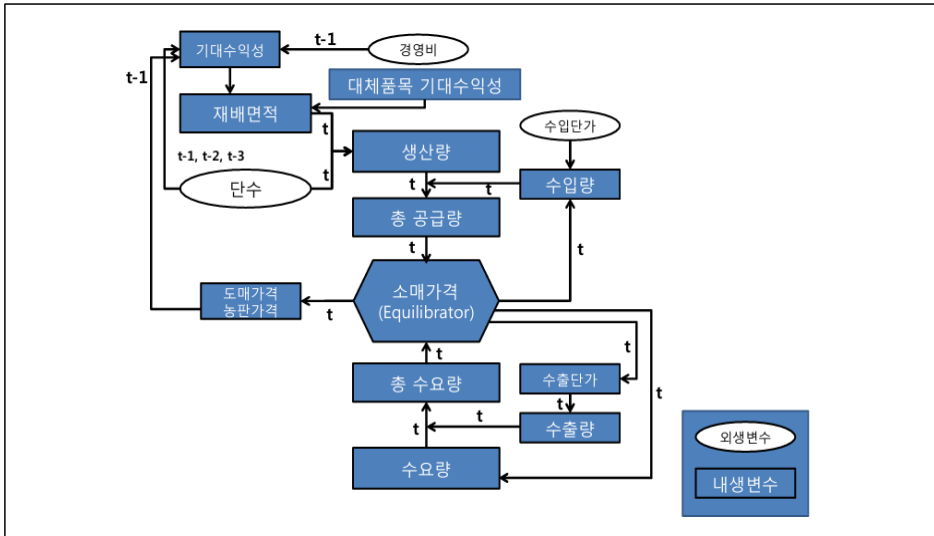
○ 김상호 외(2016)와 이형용 외(2025a)는 농축산물 수급 안정을 위한 소비 관측 정보 측면에서 접근하긴 하였으나 이들은 소비 관측의 전반적인 관측 체계와 관련한 연구로 세부 품목별로 접근한 연구는 아님.

○ 한편, 양파·마늘의 미래 수요를 다룬 연구로는 한국농촌경제연구원의 KREI-KASMO 관련 연구(서홍석 외, 2020)와 한석호 외(2010), 김인석 외(2013, 2014)의 연구가 있음.

○ 한국농촌경제연구원의 농업 부문 전망 모형인 KREI-KASMO에는 생산액이 높은 양파와 마늘에 대한 수급 추정 모형들이 포함되어 있으며, 품목별 모형에서의 연간 수요는 국내 생산량, 수출입량, 기초·기말 재고 증감분을 고려한 총 수요임.

○ KREI-KASMO의 경우 연간 모형으로 시의적절한 수급 변동 대응을 위해 단기의 수급 상황을 전망, 활용하기에는 적절치 않음.

〈그림 1-2〉 KREI-KASMO 채소류 수급 전망 모형의 구조 개념도



자료: 서흥석 외(2020: 27).

○ 한석호 외(2010), 김인석 외(2013, 2014)는 양파·마늘의 수급 불안정과 가격 변동성이 심화되면서 보다 유연한 수급량 조절을 위한 최소 2~3개월, 혹은 그 이상의 수급 상황을 전망하는 중기선행관측 모형을 개발하였음.

○ 중기선행관측 모형은 거시변수, 기후변수, 세계농산물 가격, 생산비용 등을 모형 내에서 추정·전망하지 않고, 외생변수로 취급하여 외부 추정·전망치를 이용하는 부분 균형 모형이며(김인석 외, 2013: 7), 수요와 공급이 균형가격에 의해 동시에 결정되는 연립 방정식 모형임.

- 모형은 재배면적, 단수, 생산량, 월별 출하량, 재고량, 수출입량, 도매단계 소비량 등을 추정하는 함수들로 구성되어 있으며, 모형에서 균형가격인 도매시장 가격은 월 출하량, 월초재고, 수입량의 합인 도매단계 공급량과 도매단계 소비량, 즉 소매업자 구입량, 수출량, 월말재고의 합인 도매단계 수요량의 항등식을 통해서 구해지는 구조임(김인석 외, 2013: 7).

- 중기선행관측 모형의 경우, 한석호 외(2010)는 수요 추정을 12개의 월별 축차적(recursive) 가격 신축성 함수 형태로 추정하였고, 김인석 외(2013, 2014)는 양파·마늘을 대상으로 매월 수요와 공급이 일치하는 12개(12개월)의 균형 조건이 성립하도록 반복(iteration) 계산을 통해 시장 균형가격(market-clearing price)을 도출하도록 구성하였음.
- 다만, 중기선행관측 모형에서의 수요는 최종 소비자가 실제로 소비한 물량이 아니라, 도매단계에서 소매업자의 구매 행위로 나타나는 파생수요에 가까움. 따라서 실제 최종 소비 단계에서 측정된 소비량으로 보기는 어려움.
- 기존 양파·마늘 수급 전망 관련 연구는 주로 가격과 소득 변수를 중심으로 한 수요함수 추정에 초점이 맞추어져 있는데, 소비는 생산, 수출입, 재고 변화를 이용한 수급 균형식을 통해 간접적으로 추정될 수는 있겠으나, 이러한 수급표(balance sheet) 방식은 소비를 직접 관측한 값이 아니라 잔차로 계산된 값(residual consumption)이므로 생산 및 재고 통계의 오차가 포함될 가능성, 소비 채널별 구조를 파악하기 어렵다는 등 다양한 한계가 있음(FAO, 2001).
 - 수급표의 방식(소비량=생산량+수입량-수출량-재고변동)은 식량수급표나 FAO food balance sheet 등에서 널리 사용되고 있으나, 공급 기반 소비는 엄밀히 가용량(availability)이지 실제 소비가 아님.
 - 실제 농산물 소비는 가구, 외식, 제조(식품가공), 급식 등 다양한 소비 채널을 통해 이루어지며, 이러한 소비 구조는 단일 수요함수로는 채널의 이질성을 반영하기 어렵다는 단점이 있음.
- 양파·마늘 품목별 소비행태나 소비량을 조사·추정한 연구로는 한국농촌경제연구원의 ‘식품소비행태조사’와 ‘농식품 소비정보 분석사업’이 있음.

○ ‘식품소비행태조사’는 빠르게 급변하는 식품 소비 트렌드를 포착하는 데 유용하며, 아울러 가구의 구조적 형태나 소득 계층, 개인의 식생활 습관 등 다각적인 변수를 연계하여 가구별 식품 소비 패턴을 심층적으로 규명할 수 있다는 점에서 활용 가치가 높음(이형용 외, 2025a: 43).

○ 단, ‘식품소비행태조사’는 소비자의 소비 행태 중심으로 조사되기 때문에 농축산물 품목 산업의 양적 정보의 흐름을 파악하기에는 어렵다는 단점이 있으며, 향후 미래에 대한 전망 정보를 담고 있지는 않음.

○ 반면, ‘농식품 소비정보 분석사업’은 양파와 마늘 소비를 물량 기준으로 조사하고 있으며, 향후 소비가 어떻게 될 것인지에 대한 미래 정보를 담고 있음.

○ ‘농식품 소비정보 분석사업’은 양파, 마늘, 배추, 무, 가공 김치, 쌀, 감자, 계란, 사과, 건고추 10개 품목에 대해 가구 소비자의 월별 소비 의향을 제공하고 있음.

- 농축산물 소비 의향 지수(Food Consumption Index: FCI)는 실제 구매·소비가 발생하기 이전 단계에서 소비자의 기대와 계획을 반영하므로 통계로 관측되는 소비량이나 매출액보다 선행적으로 변화의 조짐을 포착할 수 있는 신호임.

○ ‘농식품 소비정보 분석사업’은 향후 소비 의향을 3점 척도를 기반으로 한 소비 의향 지수(FCI)를 추정하고 있음.

- 품목별 소비 의향 지수는 $FCI = \sum_{i=1}^n (w_i \times FCI_i)$ 로 산출되며, 여기서 w_i 는 표본 설계 가중치, FCI_i 는 표본별 소비 의향 지수임.

- 소비 의향 지수는 3점 척도를 기반으로 가구 소비자의 최소 0에서 최대 200까지의 값으로 산출될 수 있으며($0 \leq FCI_i \leq 200$), 각 품목의 예상 소비량을 추정함($FCI_i = \text{감소} \times 0 + \text{동일} \times 100 + \text{증가} \times 200$).

○ ‘농식품 소비정보 분석사업’의 소비 의향 지수는 소비자의 기대와 심리를 신속하게 파악할 수 있다는 유용한 지표이지만, 소비자의 미묘한 의도 반영이나 물량 전환의 어려움, 응답자의 ‘의향’이 실제 소비로 이어지지 않을 가능성, 소비자 진술에 의존하는 방식에서 발생하는 편향이나 중립 응답의 왜곡 가능성, 품목 간 대체 효과 반영의 어려움 등의 약점 역시 존재함.

- 3점 척도는 소비자의 실제 의향의 세기(intensity)를 충분히 반영하지 못하므로 소비량 추정 오차 발생이 가능함. 척도를 활용한 지수는 어떤 방향이나 정도를 보여줄 뿐, 얼마나 소비량이 증가하거나 감소하는지(절대적 수량)는 직접적으로 알려주지 않음.
- 소비 의향은 심리적 기대일 뿐, 실제 행동은 가격 변화·소득·시기적 상황(명절, 기상, 물가 변동)에 따라 크게 달라짐. 특히, 농식품은 계절·가격 변동성이 크기 때문에, 의향과 실제 수요 간의 괴리가 발생하기 쉬움.
- 식품 항목 중 ‘건강·안전’ 관련 품목(사과, 김치 등)은 실제보다 “늘릴 것이다”라고 응답할 확률이 높아질 수 있으며, 응답자가 모호할 때 ‘유지’를 선택하는 경향(중심화 경향)이 존재하므로 실제 변화 의향을 과소평가할 위험이 있음.

○ 이처럼 실제 소비량 자료를 활용한 소비 관측 연구는 아직 초기 단계이며, 소비량 자료를 이용한 예측모형의 적용 가능성을 검토한 연구는 매우 제한적임.

- 본 연구는 ① 기존 공급 중심의 농업관측을 소비 영역까지 확장한다는 점, ② 단순히 소비 통계를 제공하는 것에서 벗어나 미래 정보 제공을 위한 관측형 소비 분석이라는 점, ③ 실제 소비량 자료를 활용한다는 점에서 기존 연구와 차별성이 있으며, 세부 품목에 대한 관측사업에서의 운용 및 정책 활용을 전제로 한 실행 가능한 기법 연구라는 점에서 의의가 있음.

3. 연구 내용 및 방법

3.1. 연구 내용 및 범위

- 본 연구는 농축산물 소비 관측 전반의 제도·조직 체계를 설계하는 것이 아니라, 특정 품목을 대상으로 소비 관측에 활용 가능한 방법론과 예측모형을 탐색하는 데 초점을 둠.
- 연구 대상은 주요 수급 관리 채소류인 양파와 마늘이며, 실제 소비량 자료를 활용하여 소비 관측에 적용 가능한 모형을 구축하는 데 초점을 둠.
- 본 연구는 다양한 예측기법을 적용하여 소비량 예측 가능성을 검토하고 모형별 예측 성능을 비교·평가함.
- 본 연구에서 활용한 소비 자료는 가구 패널 조사 자료, 산업 부문의 사업체 패널 자료(식품제조업, 외식업, 일반급식업²⁾), 학교·군 급식의 행정자료 등 부

²⁾ 본 보고서에서 일반급식업은 학교 및 군 급식을 제외한 기관 구내식당, 사업체 급식 등 일반 단체급식 형태의 급식업을 의미함.

문별로 자료 생성 방식과 축적 구조가 서로 상이하다는 특징을 가진. 이에 동일한 모형을 일괄적으로 적용하기보다는 자료 구조에 적합한 부문별 예측 전략을 적용할 필요가 있음. 또한, 양파·마늘 소비가 가구, 외식업, 제조업, 급식 등 소비 주체별로 소비 목적과 구매 방식, 가격 반응 구조 등이 서로 다르게 나타날 가능성이 있음.

- 만약 소비 부문별로 소비 결정 구조와 데이터 구조가 상이하면 단일 모형으로 전체 소비를 통합 추정하기 어려울뿐더러 부문별 특성이 충분히 반영되지 않을 가능성이 있음.
- 따라서 본 연구는 소비 부문별로 모형을 설정·추정하였으며, 그 결과를 장별로 구분하여 제시함. 먼저 2장에서는 양파·마늘 가구 소비량 예측모형을 추정하였으며, 3장에서는 식품제조업, 외식업, 일반급식업 부문의 모형을, 4장에서는 학교·군 급식 모형 추정 결과를 제시함. 마지막 5장에서는 모형 구축 결과를 종합적으로 요약·정리하고, 개발된 모형의 활용 방안과 모형 운영 방안, 향후 과제를 제시함.
- 한편, 본 연구에서 소비량은 최종 섭취 시점의 식품 섭취량을 의미하기보다는, 소비 주체가 양파·마늘을 소비 목적으로 취득한 물량을 의미함.
 - 가구 부문에서는 가구 소비자가 직접 구매한 물량뿐만 아니라 무상 수령 물량을 포함하며, 제조업·외식업·급식 부문에서는 원재료 사용 또는 구매를 위해 확보한 물량을 소비량으로 간주하였음.
 - 양파·마늘의 경우 구매 이후 저장, 조리, 가공 등을 거쳐 실제 섭취 시점이 달라질 수 있어 매 순간 섭취량을 직접 관측하기 어렵기 때문임.
 - 이에 본 연구에서의 소비량은 섭취 기준의 생리적 소비량이 아니라, 소비 경로에 진입한 시점에서 관측되는 구매·수령·사용 기준의 소비량으로 정의함.

○ 본 연구는 소비가 이루어지는 최종 단계에 가까운 소비 주체를 기준으로 소비 구조를 구분하며, 생산 이후의 유통 단계는 분석 범위에서 제외하였음.

- 본 연구에서의 소비량은 가구 및 산업 부문에서 소비 목적을 위해 확보된 물량을 기준으로 설정하였기에, 유통 과정에서의 중간재고 변동이나 저장에 따른 시차 등은 고려하지 않았음.

○ 본 연구는 변수 간 인과관계나 개별 요인이 예측 성과에 미치는 독립적인 효과를 규명하기보다는 표본 외 예측 성과를 바탕으로 소비 관측에 활용 가능한 모형을 탐색하는 데 초점을 두었음. 따라서 산업별 자료 특성과 예측 성과의 관계는 분석 결과에 근거하여 제한적으로 해석하였으며, 각 요인의 영향을 엄밀하게 식별하기 위한 별도의 분석은 연구 범위에 포함하지 않았음. 또한 본 연구에서 제안하는 운영 방안은 개발된 모형의 활용·운영에 대해 검토하였으며, 소비 관측의 전반적인 운영 체계에 대한 설계는 범위에서 제외하였음.

3.2. 연구 방법

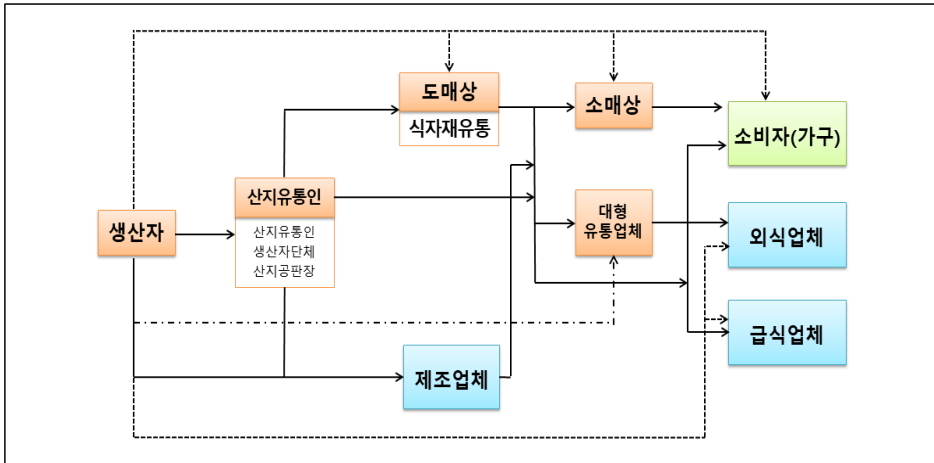
3.2.1. 활용 자료

○ 본 연구는 양파·마늘 소비 관측 모형 구축을 위해 ‘농식품 소비정보 분석사업’에서 축적된 소비량 데이터를 활용하였음.

○ ‘농식품 소비정보 분석사업’은 가구와 산업의 농식품 소비 동향과 트렌드를 분석하여 농식품 소비 분야의 변화를 가늠하고, 농식품 소비 관련 정보에 대한 수요자의 합리적인 의사결정과 관련 정책 수립을 돕는 기초자료를 제공하기 위해 수행되고 있음(이형용 외, 2025a).

- ‘농식품 소비정보 분석사업’은 농축산물 9개 품목(채소 5개, 식량작물 2개, 축산물 1개)과 가공식품 1개(김치) 총 10개 품목을 대상으로 하며, 여기에는 양파와 마늘이 포함되어 있음.
- ‘농식품 소비정보 분석사업’은 농축산물의 최종 소비 단계를 가정식, 외식, 급식, 가공품 구매 등 소비가 이루어지는 지점으로 구분하여 소비 실태와 소비량을 조사·분석하고 있으며, 조사는 가구 소비자, 외식업체, 급식업체, 식품제조업체를 대상으로 하고 있음.
- 이는 농축산물 유통 경로가 복잡하고 품목별로 유통 구조가 상이하여, 중간 유통 단계까지 고려하면 실제 소비량을 효과적으로 수집하기 어렵다는 점에 착안한 것임.
 - 농축산물은 생산 방식뿐만 아니라 유통 및 가공 방법과 유통되는 경로 혹은 주체의 취급 비중이 품목별로 매우 다양하고 복잡하기 때문에 경로별 취급량의 흐름을 파악하기는 현실적으로 어려움.
 - 생산, 출하, 유통 단계 정보만으로는 실제 소비 여부와 소비 주체별 사용 구조를 충분히 설명하기 어려움. 농축산물은 저장/가공/재유통 과정에서 시차와 중복 이동이 발생할 가능성이 있음.
 - 이에 소비가 실제 이루어지는 최종 수요처 기준 접근이 실제 소비 흐름에 보다 근접할 수 있으므로 각 소비 단계에서의 구매(무상 수령 포함), 사용 물량을 조사 대상으로 설정하고 있음.

〈그림 1-3〉 일반적인 농식품 유통경로



자료: 김상효 외(2016: 114).

○ ‘농식품 소비정보 분석사업’의 데이터 수집 방식은 다음과 같음.

- 가구 내 농식품 구매량과 구매 행태를 직접 조사하기 위해 가구 패널 조사를 수행함(2,000가구).
- 산업 부문은 ‘전국사업체조사’에 등록된 식품제조업 1,000개, 외식업 1,000개, 급식업(기관 구내식당업) 500개 내외 업체를 대상으로 조사함.
- 학교 급식(초·중·고)과 군 급식은 한국농수산식품유통공사(eaT)와 농협경제지주의 데이터를 제공 받아 분석함.

○ ‘농식품 소비정보 분석사업’의 소비 데이터는 가구 부문의 소비뿐만 아니라 관련 산업 소비량까지 월별로 동시에 조사된다는 점에서 특정 유통 채널의 소비가 조사되는 기존 소비 관련 데이터와 차별성이 있음.

○ 다만, ‘농식품 소비정보 분석사업’의 경우 보조금 사업 형태로 운영되는 관계로 사업 수행 기관이 선정되는 단계를 거치고 나면 이 사업으로 데이터를 조

사·수집하고 있는 기간이 4~12월이기 때문에 데이터 결측이 존재함.

- 가구 조사 부문은 1~4월, 제조업, 외식업 및 일반급식업 조사 부문은 1~3월 데이터가 수집되지 않음.

○ 또한, 해당 사업이 2020년부터 시작되었으나, 사업 초기 이슈 중심의 운영, 데이터 수집 체계 구축 시기 소요 등의 이유 때문에 데이터가 본격적으로 구축된 시점은 2023년 이후임. 따라서 월별 집계자료 형태로 발표되고 있는 소비 데이터는 비교적 시계열이 짧다는 한계가 있음.

○ 반면, 학교·군 급식 소비량 자료는 외부 조사 방식이 아니라 한국농수산식품유통공사(aT) 및 농협경제지주로부터 제공받은 행정자료 기반이라는 점에서 비교적 연속적인 월별 자료 구축이 가능함.

- 학교·군 급식 자료는 실제 급식 운영 과정에서 발생·구축된 자료라는 점에서 조사 기반 소비 자료와는 다른 특성을 가짐.
- ‘농식품 소비정보 분석사업’을 통해 축적된 학교 급식 소비량은 2019년 1월부터, 군 급식 소비량 자료는 2020년 1월부터 활용 가능함.

○ 이에 본 연구는 학교·군 급식 자료의 상대적으로 긴 시계열 특성과 소비 특성을 고려하여 일반급식 부문과 별도의 소비 부문으로 구분하여 분석함.

3.2.2. 분석 방법³⁾

○ 본 연구는 수집된 양파·마늘 소비량 데이터의 특성에 맞추어 가구 소비량과 학교·군 급식 소비량에 대해서는 시계열 분석을 수행하였으며, 식품제조업, 외식업, 일반급식업 등 산업 부문에 대해서는 패널 분석을 수행하였음.

³⁾ 본 연구에서 활용한 예측모형의 세부 이론과 추정 원리는 부록에 별도로 정리하였음.

가) 가구 부문

- 앞서 밝힌 바와 같이 ‘농식품 소비정보 분석사업’의 패널 조사자료는 매년 초 조사가 이루어지지 않은 결측 구간을 포함하고 있음. 이러한 자료 구조를 고려할 때, 월별 자료만을 이용하여 시계열 모형을 구축하는 경우 전체 관측치가 계절성 및 추세를 안정적으로 식별하는 데 제약이 따를 수 있음.
- 이에 본 연구에서는 기존 월별 집계자료를 가능한 범위에서 주별 자료로 전환하여 활용하였으며, 이러한 결측 구간을 전제로 예측 전략을 설계하고자 함.
- ‘농식품 소비정보 분석사업’의 가구 소비 조사는 매일 구매 품목을 체크(영수증 검증)하는 방식으로 조사되므로 주별 집계 가능한 구조임. 주별 집계는 월별 자료에 비해 시계열 관측 빈도를 높여 소비의 단기 변동과 계절적 패턴을 보다 세밀하게 반영할 수 있게 함.
- 다만, 2023년부터 2025년까지 1~4월 주별 자료가 존재하지 않는다는 점에서 주별 시계열 역시 구조적 결측을 포함하기 때문에, 모형의 추정과 예측 성과에 미치는 영향을 고려하여 결측을 직접 허용하는 예측 전략과 결측을 사전 보간 후 모형을 적용하는 전략을 비교·평가하고자 함.
 - Ahn et al.(2022)은 실제 결측이 존재하는 시계열에 대해 여러 결측 대체 방법을 적용한 뒤 각 방법으로 복원된 학습자료를 이용하여 예측모형을 구축하고 공통 데이터 셋에서 예측력을 비교하는 방식으로 결측 대체 방법의 효과를 평가하였음. 본 연구는 이러한 접근을 참고하여 결측값 복원 자체의 정확도보다 최종 예측 성과의 관점에서 결측 처리 전략과 예측모형의 조합을 평가하고자 함.
- 결측 구간이 존재하는 가구 소비량의 시계열 예측 전략은 결측을 허용한 직접

예측 전략과 결측 보간 후 예측 전략으로 구분하였음. 결측 허용 직접 예측 전략은 결측을 별도로 단순 대체하지 않고 적용 가능한 모형을 이용하여 예측을 수행한 것이며, 결측 보간은 선형보간(linear interpolation)을 이용하여 시계열을 먼저 보완한 후 모형 추정을 수행함.

- 가구 부문 양파·마늘 소비량 예측을 위한 후보 모형으로는 ARIMA, ARIMAX (ARIMA with exogenous variables), ETS(Error, Trend, Seasonality) 등을 고려할 수 있음. 이 접근은 보간 과정에서 발생할 수 있는 추가 왜곡을 최소화하고, 결측이 포함된 원시 시계열의 정보를 가능한 직접적으로 활용할 수 있다는 장점이 있음.

나) 식품제조업, 외식업, 일반급식업 부문

- 반면, 산업 부문(식품제조업, 외식업, 일반급식업) 자료는 조사 자체가 응답 기반의 월별 집계 중심으로 이루어지고 있으므로 가구 소비 자료와 같이 주별 시계열로 재구성하여 관측치 수를 확장하는 방식을 현실적으로 적용하기 어려운 측면이 있음. 이러한 점을 고려할 때 산업 부문 자료에 대해 단일 집계 시계열을 중심으로 모형을 구축하는 접근은 한계가 있음.

- 산업 부문(식품제조업, 외식업, 일반급식업) 자료는 2022년부터 2025년까지의 월별 자료로 구성되어 있으나, 매년 1~3월 자료가 존재하지 않는 구조적 결측을 포함하고 있음.

- 이에 본 연구는 산업 부문 자료에 대해 시계열 분석보다는 패널 분석을 중심으로 예측 전략을 설계하고자 함. 이는 시간 축이 길지 않고 구조적 결측이 존재하는 상황에서 사업체별 반복 관측정보를 활용하여 보다 많은 관측치를 확보하고 사업체 간 이질성과 시간 효과를 동시에 반영하기 위한 것임.

○ 구체적으로는 사업체-월 패널자료를 이용해 고정효과 패널회귀모형(fixed effects model), 업체별 과거 평균 모형과 함께 CatBoost와 같은 트리 기반 머신러닝 모형을 구축하고 예측력을 비교·평가함.

- 설명변수로는 사업체 규모, 과거 평균 소비량과 같은 개체별 특징과 함께 양파·마늘 생산/출하와 관련된 공급 측면의 정보들을 활용하였음. 이러한 접근은 구조적 결측이 존재하더라도 실제 관측된 월 자료를 중심으로 예측 모형을 구축할 수 있다는 장점이 있음.

- Martyushev et al.(2025)은 패널자료를 기반으로 패널회귀 분석과 CatBoost와 같은 트리 기반 머신러닝 비교를 통해, 다양한 변수 구성을 포함함으로써 트리 기반 머신러닝의 예측력이 개선될 수 있음을 보였음. 따라서 본 연구에서도 트리 기반 머신러닝 기법을 패널회귀모형과 함께 비교·검토하는 대안으로 활용함.

다) 학교·군 급식 부문

○ 학교 급식과 군 급식 소비 자료의 경우 앞선 가구 및 산업 부문 패널 조사 자료와 달리 결측 구간이 없는 비교적 긴 월별 시계열 자료라는 점에서 결측 처리 방식의 비교 없이 월별 시계열 자료에 대한 예측모형 간 성과를 비교하였음.

○ 학교·군 급식 소비량 예측모형에는 계절성 단순 기법(Seasonal Naive), ETS 지수평활, ARIMA, ARIMAX 기법을 적용하였음.

○ 한편, 학교·군 급식은 개인의 선택이나 가격에 의해 소비가 결정되는 시장 기반 소비와 달리 예산과 식단 계획에 따라 운영된다는 유사성, 전체 소비에서 차지하는 비중이 상대적으로 작고 시계열 변동성이 제한적이라는 점을 고려하여 본 연구에서는 두 부문을 통합하여 모형을 추정하였음.

〈표 1-1〉 분석 방법 요약

구분	가구 소비	산업 부문 소비 (식품제조업·외식업·일반급식업)	학교·군 급식 소비
자료 구조	가구-일자-품목 반복 관측 자료	사업체-월-품목 반복 관측 자료	월-품목 반복 관측 자료
현재 활용 자료	월별 집계자료	월별 집계자료	월별 집계자료
공통 제약	2023~2025년 매년 1~4월 결측	2022~2025년 매년 1~3월 결측	-
결측 성격	구조적 결측	구조적 결측	결측 없음
확장 가능성	주별 전환 가능	주별 전환 어려움	학교: 주별 전환 가능 군: 주별 전환 어려움
분석 방법	시계열 분석	패널 분석	시계열 분석
기본 전략	결측 허용/보간 후 모형 비교	원자료/로그 변환 소비량 모형 비교	월별 시계열 모형 비교
추정 모형	ARIMA, ARIMAX, ETS	FE 패널회귀, CatBoost, 단순 기준 모형	SNiave, ARIMA, ARIMAX, ETS 지수평활
연구 문제	어떤 결측처리-예측모형 조합의 예측력이 가장 우수한가?	어떤 패널 기반 예측모형의 예측력이 가장 우수한가?	어떤 시계열 예측모형의 예측력이 가장 우수한가?

자료: 저자 작성.

○ 본 연구에서는 소비량 예측 시 공급 측면의 정보를 반영하기 위해 ARIMAX 등 외생변수 활용이 가능한 모형에 양파·마늘의 도매시장 반입량 및 도매가격 시차 변수를 포함하였음.

○ 또한, 모형 적합성은 개별 계수의 통계적 유의성보다 예측오차(MAE, RMSE, MAPE)를 중심으로 평가하였음.

2

가구 부문 소비 모형 추정

1. 분석 자료의 특성

- 본 연구에서 활용한 가구 부문 소비 자료는 가구의 품목별 소비 내역을 일자별로 수집한 반복 관측 자료이며, 기본 관측 단위는 가구-일자-품목 단위임.
- 다만, 예측모형 구축에 활용 가능한 자료는 현재 원자료가 아닌 월별 집계자료 형태로 제공되고 있음. ‘농식품 소비정보 분석사업’의 데이터가 안정적으로 축적된 시기는 2023년부터이지만, 2023~2025년 매년 1~4월에는 조사가 이루어지지 않아 해당 기간의 소비량 자료가 결측 처리되어 있음.
- 이러한 자료 구조를 고려할 때, 월별 집계자료만을 이용하여 시계열 예측모형을 구축하는 데에는 구조적 제약이 따름.
 - 분석 기간이 2023~2025년으로 비교적 짧고, 매년 1~4월 자료가 부재하기 때문에 월별 기준의 유효 관측치 수가 충분하지 않으며, 계절성 및 추세를 안정적으로 식별하는 데에도 한계가 있음.

- 이에 본 연구에서는 가구 소비량 일자별 조사 원자료를 주별 자료로 집계하여 분석에 활용하였음. 주별 집계자료는 월별 자료보다 더 많은 시점의 관측치를 확보할 수 있고, 소비의 단기 변동과 시기별 변화 양상을 보다 세밀하게 포착할 수 있다는 장점이 있음.
- 한편, 주별 자료로 전환하더라도 2023~2025년 매년 1~4월 조사 미실시에 따른 구조적 결측은 근본적으로 해소되지 않음. 월별 조사 공백이 주별 단위로 세분화됨에 따라 해당 기간의 결측은 다수의 주차 단위 관측치 공백으로 나타남.
- 따라서 본 자료는 충분한 길이의 연속 시계열이라기보다 특정 시기에 반복적으로 자료가 부재한 구조적 결측을 포함한 불완전한 주별 시계열 자료로 이해할 필요가 있음. 본 연구의 분석 결과 역시 이러한 제한된 자료 조건에서 소비량 예측모형의 적용 가능성을 검토한 기초 분석 결과로 해석할 필요가 있음.
- <표 2-1>은 양파와 마늘 가구 소비량 자료의 연도별 관측 구조를 제시한 것임. 분석 자료 기준, 양파와 마늘은 각각 2023~2025년 전체 156개 주차로 구성되며, 이 중 소비량이 실제 관측된 주차는 107개, 결측으로 처리된 주차는 49개임.
 - 유효 관측치는 주로 4월 이후부터 확보되며, 연초 구간의 조사 공백으로 인해 전체 표본의 약 31.4%가 결측 처리 대상에 해당함.
 - 양파와 마늘은 같은 조사 체계에서 구축된 자료이므로 전체 표본 수, 유효 관측 수, 결측 처리 관측 수가 동일하게 나타남.

〈표 2-1〉 가구 소비량 자료의 관측 구조

단위: 개

전체 표본 (주차)	유효 관측 수	결측 처리 관측 수	결측 비중	비고
156	107	49	31.4%	4월 이후 소비량 관측

자료: 저자 작성.

- 품목별 가구 소비량의 기초통계량은 〈표 2-2〉와 같음. 양파 소비량은 유효 관측치 기준 평균 5,195.9톤, 중앙값 5,002.8톤, 표준편차 1,179.7톤으로 나타남. 마늘 소비량은 평균 1,376.6톤, 중앙값 1,241.3톤, 표준편차 554.2톤으로 나타남.

〈표 2-2〉 품목별 가구 소비량의 기초통계량

단위: 개, 톤

품목	유효 관측 수	평균	중앙값	표준편차	최솟값	최댓값
양파	107	5,195.9	5,002.8	1,179.7	2,748.8	9,349.4
마늘		1,376.6	1,241.3	554.2	525.2	3,328.1

자료: 저자 작성.

- 〈그림 2-1〉은 2023~2025년 양파와 마늘 가구 소비량의 주별 변화 추이를 나타낸 것임.
- 양파·마늘 가구 소비량은 유효 관측 구간을 기준으로 두 품목 모두 뚜렷한 장기 상승 또는 하락 추세가 지속적으로 나타나기보다는 일정 범위 내에서 단기적인 등락을 반복하는 양상을 보임. 이는 가구 소비량이 계절적 요인이나 일시적 수급 변동에 따라 단기적으로 반응하는 특성을 반영한 것으로 여겨짐.
- 이상 자료의 특성을 종합하면, 양파와 마늘 가구 소비량 자료는 ① 관측 기간의 단기성, ② 구조적 결측의 반복성, ③ 추세 불명확성이라는 세 가지 주요 특

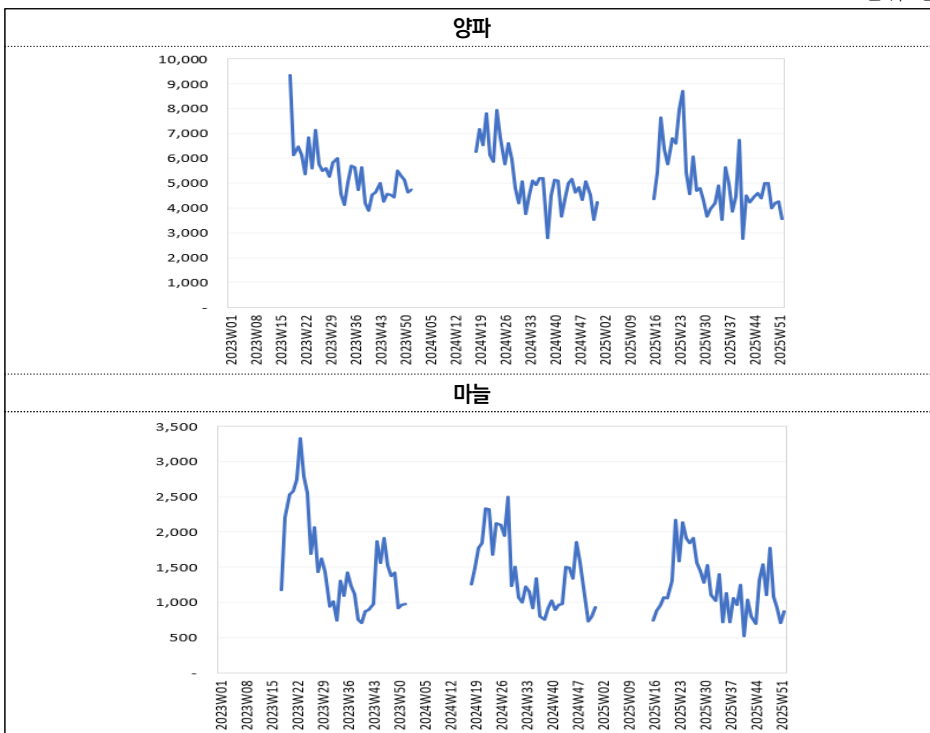
성을 지님. 따라서 일반적인 시계열 예측모형을 그대로 적용하기보다는 결측 처리 방식과 학습자료 구성 방식을 함께 고려한 예측 전략이 요구됨.

○ 이에 본 연구에서는 구조적 결측을 포함한 주별 소비량 자료에 적용 가능한 예측모형 구축 전략을 설정하고, 결측 처리 방식과 모형 조합별 예측 성과를 비교·평가하고자 함.

- 다만, 본 연구의 분석 결과는 특정 모형의 일반적 우월성을 확정하기 위한 것이 아니라, 현재 이용 가능한 자료 조건에서 상대적으로 안정적인 예측 성과를 보이는 모형 조합을 탐색하기 위한 것으로 해석함.

〈그림 2-1〉 품목별 가구 소비량의 변화 추이

단위: 톤



자료: 저자 작성.

2. 모형 구축 전략

- 앞서 제시한 분석 자료의 특성을 고려하여 본 연구는 결측 처리 방식과 예측 모형의 조합을 비교하는 방식으로 예측 전략을 설정하였음. 이는 구조적 결측이 포함된 주별 소비량 자료에서 특정 결측 처리 방식이나 단일 예측모형을 사전에 고정하기보다 자료 조건에 따라 상대적으로 안정적인 예측 성과를 보이는 결측 처리 방식-예측모형 조합을 확인하기 위한 것임.
- 본 연구의 예측 전략은 다음의 세 단계로 구성됨. 첫째, 소비량 자료의 결측 구조를 고려하여 보간 미적용 방식과 보간 적용 방식을 설정함. 둘째, 결측 처리 방식별로 적용 가능한 예측모형을 구성함. 셋째, 동일한 학습 구간과 테스트 구간을 기준으로 모형 조합별 예측 성과를 비교·평가함.
- 분석 과정에서 소비량 자료의 0값은 실제 소비량이 아닌 조사 미실시 또는 자료 미확보에 따른 결측값으로 처리함. 이후 결측값을 별도로 보간하지 않는 방식과 결측값을 보간한 후 예측모형을 적용하는 방식을 구분하여 분석함. 이러한 접근은 실제 결측값이 포함된 시계열 자료에서 결측 처리 방식별로 학습 자료를 구성하고, 예측 성과를 비교한 Ahn et al.(2022)의 연구와도 맥락을 같이함.
 - Ahn et al.(2022)은 결측값의 실제값을 직접 확인하기 어려운 실제 자료 환경에서 보간 방법의 효과를 최종 예측 성과를 통해 간접적으로 평가하는 방식을 제시하였음.
- 다만, 본 연구에서 결측 보간은 구조적 결측 문제를 근본적으로 해결하기 위한 절차가 아니라 제한된 자료 조건에서 예측모형을 적용하고 그 성과를 비교하기 위한 보완적 처리임을 유의할 필요가 있음.

- 특히, 1~4월 결측 구간에는 명절, 계절적 소비 증가, 저장 소비, 품목별 구매 주기 등 비선형적 소비 변동이 포함될 가능성이 있음. 따라서 선형보간으로 생성된 값은 실제 소비량을 정확히 복원한 값이라기보다, 결측값을 직접 허용하기 어려운 일부 예측모형을 비교하기 위해 구성한 분석용 자료로 이해할 필요가 있음.
- 따라서 이후의 예측력 평가 결과는 특정 보간 방법이 실제 소비량을 가장 정확하게 복원하였다는 의미가 아니라 주어진 자료 조건에서 어떤 결측 처리 방식과 예측모형의 조합이 상대적으로 낮은 예측오차를 나타내는지를 비교한 결과로 해석함.

2.1. 결측값 처리 방식

- 본 연구에서는 결측값 처리 방식을 보간 미적용 방식과 선형보간 적용 방식으로 구분하였음. 보간 미적용 방식은 결측값을 인위적으로 대체하지 않고, 원자료의 결측 구조를 가능한 범위에서 유지한 상태로 예측모형을 적용하는 방식임.
- 선형보간(linear interpolation) 적용 방식은 결측이 포함된 소비량 시계열을 사전에 보완한 후 예측모형을 적용하는 방식임. 선형보간은 결측 구간의 직전 관측값과 직후 관측값을 선형적으로 연결하여 결측값을 추정하는 방법임. 구조가 단순하고 적용이 용이하다는 장점이 있으나, 결측 구간이 길거나 실제 소비 흐름이 비선형으로 변동하는 경우 단기 변동과 계절적 소비 변화를 충분히 반영하지 못할 수 있음.

- 본 연구에서 선형보간을 적용한 것은 결측값이 포함된 주별 소비량 자료에 ETS(Error, Trend, Seasonality) 지수평활(exponential smoothing), ARIMA(autoregressive integrated moving average) 계열 모형을 동일한 자료 조건에서 적용하기 위함임. 특히, ETS 지수평활 모형은 결측값이 포함된 자료에 직접 허용하기 어렵기 때문에 보간 적용 자료를 별도로 구성하여 보간 미적용 자료와의 예측 성과를 비교하였음.

2.2. 모형의 설정

- 본 연구에서는 양파와 마늘 가구 소비량의 주별 예측을 위해 ETS 지수평활 모형과 ARIMA 계열 모형을 적용하였음. ETS 지수평활 모형은 시계열의 수준, 추세, 계절성을 반영하는 데 활용될 수 있으며, ARIMA 모형은 시계열의 자기상관 구조와 차분을 통해 과거 흐름을 반영할 수 있는 대표적인 정량 예측모형임.
- ETS 지수평활 모형은 과거 관측값의 가중평균을 이용하여 예측치를 산출하는 모형임. 시계열을 오차(E), 추세(T), 계절성(S)의 구성요소로 분해하고, 각 구성요소의 결합 방식(가법적·승법적)에 따라 다양한 형태로 설정될 수 있음. 본 연구에서는 선형보간을 적용한 자료에 ETS 모형을 적용하여 소비량 시계열의 단기적인 수준 변화와 반복적 변동을 반영하고자 하였음.
- ARIMA 모형은 자기회귀(autoregressive: AR) 과정과 이동평균(moving average: MA) 과정을 결합한 시계열 모형임. 차분(differencing)을 통해 비정상 시계열을 안정화한 후 과거 관측값과 과거 오차항의 정보를 이용하여 예측치를 산출함. 본 연구에서는 보간 미적용 자료와 선형보간 자료 모두에 ARIMA 모형을 적용하여 소비량 자체의 과거 흐름에 기반한 예측 성과를 평가함.

○ 아울러 외생변수를 포함한 ARIMAX 모델을 추가로 적용하였음. ARIMAX 모형은 ARIMA 구조에 설명변수를 추가한 확장 모형으로 본 연구에서는 도매시장 반입량의 1~3주 시차 변수를 외생변수로 포함하였음. 이는 소비량 자체의 과거 흐름만을 반영하는 단변량 모형과 공급 측면의 물량 정보를 추가로 고려한 모형 간의 예측 성과를 함께 비교하기 위한 설정임.

○ 결측값 처리 방식과 모형 적용 가능성을 고려하여 보간 미적용 자료에는 ARIMA 모형을 적용하였고, 선형보간 자료에는 ETS 지수평활, ARIMA, ARIMAX 모형을 각각 적용하였음. 이는 ETS 지수평활 모형이 결측값이 포함된 자료에 직접 적용되기 어렵다는 점을 고려한 설정임. Hyndman & Athanasopoulos(2021)는 결측값이 포함된 시계열에서 ARIMA 모형은 적용 가능하나, ETS 지수평활 모형은 결측값을 직접 처리하기 어렵다는 점을 설명한 바 있음.

○ 결측 처리 방식에 따른 적용 예측모형의 구성은 <표 2-3>과 같음.

<표 2-3> 결측 처리 방식별 예측모형 구성

결측 처리 방식	세부 처리 방식	적용 예측모형
보간 미적용	원자료 결측 구조 유지	ARIMA
보간 적용	선형보간(linear interpolation)	ETS, ARIMA, ARIMAX

자료: 저자 작성.

2.3. 학습·테스트 자료의 설정 및 예측력 평가 방법

○ 본 연구에서는 예측모형의 학습과 예측력 평가를 구분하기 위해 분석 자료를 학습 구간과 테스트 구간으로 나누어 활용하였음. 학습 구간은 예측모형의 구

조와 모수를 추정하는 데 사용되는 자료이며, 테스트 구간은 학습된 모형의 표본 외 예측 성과를 평가하기 위한 자료임.

○ 본 연구에서는 2023년부터 2024년까지의 관측자료를 초기 학습 구간(N=88)으로 설정하고, 2025년 자료를 테스트 구간으로 설정함. 단, 가구 소비량 자료는 매년 1~4월 조사 미실시에 따른 구조적 결측을 포함하고 있으므로 테스트 구간 전체 52개 주차 중 실제 소비량이 관측된 주차를 대상(N=37)⁴⁾으로 예측력 평가를 수행함.

- 개별 예측모형의 구조와 설명변수 구성은 2023~2024년의 초기 학습자료를 이용하여 결정하였음.

○ 예측은 rolling-origin 방식의 1주 선행 예측(one-step-ahead forecast)으로 수행하였으며, 학습자료의 구성 방식에 따라 고정 길이 이동 방식(fixed-length rolling window)과 확장 방식(expanding window)을 적용하였음.

- 고정 길이 이동 방식은 예측 시점 직전의 고정된 일정 기간 자료만을 활용하여 모형을 재추정하는 방식임. 확장 방식은 예측 시점이 이동함에 따라 이용 가능한 과거 자료를 순차적으로 누적하여 모형을 재추정하는 방식임.

- 구체적으로 초기 학습 구간에서 결정된 동일한 모형 구조를 유지한 상태에서 2025년의 각 예측 시점에 이용 가능한 자료를 학습창 방식에 따라 갱신하고 모수를 재추정하여 다음 1주 소비량을 예측하였음.

○ 주별 소비량 자료는 분석 기간이 비교적 짧고 구조적 결측이 반복되는 특성을 가지므로 예측 시점에서 어느 범위의 과거 자료를 활용하는지에 따라 예측 성

⁴⁾ 2025년에는 4월 15~30일 조사가 진행되었음.

과가 달라질 수 있음. 이에 본 연구에서는 특정 학습자료 범위를 사전에 고정하기보다 학습자료 구성 방식에 따른 예측 성과 차이를 비교하기 위해 세 가지 예측 시나리오를 설정하였음.

- 구체적으로 학습자료 구성 방식은 ① 최근 52주 자료를 이용하는 고정 길이 이동 방식, ② 최근 72주 자료를 이용하는 고정 길이 이동 방식, ③ 예측 시점 이전의 모든 자료를 누적으로 이용하는 확장 방식으로 구분하였음.
- 52주는 최근 1년의 소비 패턴을 반영하기 위한 기준이며, 72주는 보다 장기간의 소비 정보를 활용하는 경우 예측 성과가 어떻게 달라지는지를 비교하기 위한 대안적 설정임.

○ 예측 정확도 평가 지표로는 평균절대오차(mean absolute error: MAE), 평균제곱근오차(root mean squared error: RMSE), 평균절대백분율오차(mean absolute percentage error: MAPE)를 사용함.

- MAE는 실제값과 예측값 간 절대오차의 평균으로, 예측오차의 평균적인 크기를 직관적으로 파악하는 데 유용함.
- RMSE는 제곱오차 평균의 제곱근으로, 큰 예측오차에 상대적으로 더 큰 가중치를 부여하는 지표임. 따라서 일부 시점에서 큰 오차가 발생하는지 확인하는 데 유용함.
- MAPE는 실제값 대비 예측오차의 비율을 백분율로 나타낸 지표로, 예측 오차의 상대적 크기를 파악하는 데 활용됨. 다만, MAPE는 실제 소비량이 낮은 시점에서는 동일한 절대오차에도 상대적으로 크게 산출될 수 있으므로, 예측 성과를 평가할 때는 MAE와 RMSE를 함께 고려할 필요가 있음.
- 따라서 본 연구에서는 세 지표를 종합적으로 검토하여 모형의 예측 성과를 평가함.

○ 식(1)~(3)에서 y_t 는 t 시점의 실제 소비량, $\hat{y}_t$ 는 해당 시점의 예측값, n 은 예측력 평가에 활용된 관측치 수를 의미함. 세 지표는 모두 값이 작을수록 예측 오차가 작고 예측 성과가 우수한 것으로 해석함.

$$(식 1) \quad MAE = \frac{1}{n} \sum_{t=1}^n |y_t - \hat{y}_t|$$

$$(식 2) \quad RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2}$$

$$(식 3) \quad MAPE = \frac{100}{n} \sum_{t=1}^n \left| \frac{y_t - \hat{y}_t}{y_t} \right|$$

3. 모형 추정 및 예측력 평가 결과

3.1. 양파 가구 소비량

3.1.1. 모형 추정 결과

○ <표 2-4>는 2023~2024년 학습 구간을 이용하여 결측 처리 방식별 예측모형을 추정한 결과를 제시한 것임. 보간 미적용 자료에는 원자료의 결측 구조를 유지한 상태에서 ARIMA 모형을 적용하였으며, 선형보간 자료에는 ETS, ARIMA, ARIMAX 모형을 각각 적용하였음.

○ 추정 결과, 보간 미적용 자료에 적용한 ARIMA 모형은 ARIMA(4,1,1)로 나타났음. 선형보간 자료의 경우 ETS 모형은 ETS(M,N,N), ARIMA 모형은 ARIMA(1,1,1), ARIMAX 모형은 ARIMAX(2,0,1)로 각각 추정되었음.

- ARIMAX 모형에는 반입량의 1~3주 시차 변수를 외생변수로 포함하였음.
이는 소비량 자체의 과거 변동만을 반영하는 단변량 모형과 공급 측면의 물량 정보를 추가로 고려한 모형 간의 예측 성과를 비교하기 위한 설정임. 도매시장 반입량은 해당 시점에 도매시장으로 유입된 물량을 나타내므로 주별 소비량의 단기 변동을 설명하는 데 활용 가능한 외생 정보로 볼 수 있음.

○ <표 2-4>에 제시된 모형의 구조와 주요 설정은 이후 예측 과정에서도 동일하게 적용하였음. 다만 2025년 테스트 구간의 각 예측 시점마다 rolling 또는 expanding window 방식으로 학습자료를 구성하였으며, 해당 학습자료에 모형을 다시 적합하여 모수를 재추정한 후 1주 앞 예측값을 산출하였음.

○ 따라서 예측력 평가 결과는 <표 2-4>의 모형 구조를 바탕으로 하되, 예측 시점별 재추정 과정을 반영한 표본 외(out-of-sample) 예측 성과로 해석할 필요가 있음.

<표 2-4> 결측 처리 방식별 예측모형 추정 결과

결측 처리 방식		추정 구조	주요 설정
보간 미적용	원자료 결측 유지	ARIMA(4,1,1)	소비량의 과거 흐름을 이용한 단변량 모형
보간 적용	선형보간	ETS(M,N,N)	곱셈 오차(M), 추세 없음(N), 계절성 없음(N)
		ARIMA(1,1,1)	소비량의 과거 흐름을 이용한 단변량 모형
		ARIMAX(2,0,1)	반입량의 1~3주 시차변수 포함

자료: 저자 작성.

3.1.2. 예측 정확도 평가 결과

- <표 2-5>는 2025년 테스트 구간을 대상으로 결측 처리 방식, 예측모형, 학습창 설정별 예측 정확도를 비교한 결과를 제시한 것임. 예측 정확도는 MAE, RMSE, MAPE를 이용하여 평가하였으며, 평가 대상 관측치는 총 37개임.
- 분석 결과, 전체 모형 조합 중 최근 52주 자료를 이용하고 반입량의 1~3주 시차 변수를 포함한 선형보간 ARIMAX 모형의 예측오차가 가장 낮게 나타났음. 해당 모형의 MAE는 664.2, RMSE는 984.1, MAPE는 13.1로 모든 평가 지표에서 가장 낮은 값을 보였음.
- 학습자료 구성 방식별로 살펴보면, 과거 자료를 누적적으로 이용한 방식에서는 보간 미적용 ARIMA 모형의 예측오차가 가장 낮았음. 반면 최근 52주 및 최근 72주 자료를 이용한 이동 방식에서는 반입량 시차 변수를 포함한 선형보간 ARIMAX 모형이 상대적으로 우수한 예측 성과를 보였음.

<표 2-5> 양파 가구 소비량 예측 정확도 평가 결과(N=37)

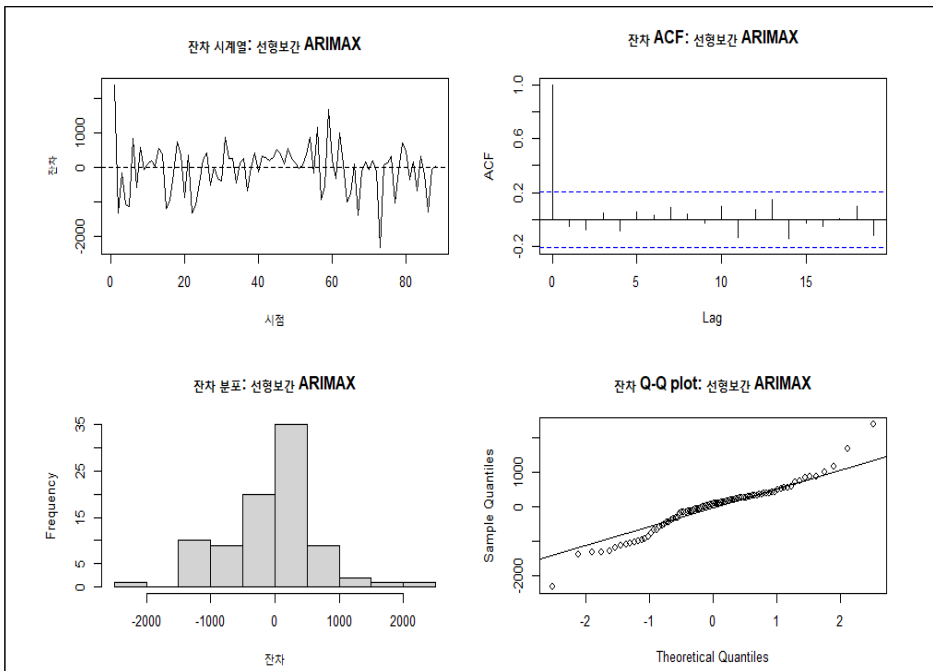
학습창	결측 처리 방식		추정 구조	MAE	RMSE	MAPE
과거 자료 누적 이용 (Expanding)	보간 미적용	원자료 결측 유지	ARIMA(4,1,1)	814.5	1,035.5	16.98
	보간 적용	선형보간	ETS(M,N,N)	836.2	1,161.0	17.27
			ARIMA(1,1,1)	821.2	1,105.4	17.28
			ARIMAX(2,0,1)	831.3	1,095.6	17.19
최근 52주 자료 이용 (Rolling_52)	보간 미적용	원자료 결측 유지	ARIMA(4,1,1)	956.7	1,297.6	20.08
	보간 적용	선형보간	ETS(M,N,N)	846.4	1,181.4	17.56
			ARIMA(1,1,1)	901.0	1,254.0	18.82
			ARIMAX(2,0,1)	664.2	984.1	13.10
최근 72주 자료 이용 (Rolling_72)	보간 미적용	원자료 결측 유지	ARIMA(4,1,1)	992.3	1,249.5	21.12
	보간 적용	선형보간	ETS(M,N,N)	838.4	1,166.0	17.33
			ARIMA(1,1,1)	879.9	1,222.4	18.50
			ARIMAX(2,0,1)	802.8	1,067.9	16.54

자료: 저자 작성.

○ ARIMA 계열 모델을 기준으로 비교한 결과, 선형보간의 적용 효과는 학습창에 따라 일관되게 나타나지 않았음. 반면 최근 52주와 72주 자료에서는 동일한 선형보간 ARIMA보다 ARIMAX의 예측오차가 낮게 나타났으며, 특히 최근 52주 ARIMAX 모형의 성과가 가장 우수하였음. 따라서 최종 예측 성과는 선형보간 자체보다는 최근 소비 흐름과 외생변수의 시차 정보를 함께 반영한 결과로 해석할 수 있음.

○ 예측 정확도 평가에서 가장 낮은 예측오차를 보인 선형보간 ARIMAX 모형에 대해 잔차 진단을 수행하였음. <그림 2-2>는 2023~2024년 학습 구간에서 추정한 해당 모형의 잔차 시계열, 잔차 자기상관함수(ACF), 잔차 분포 및 Q-Q plot을 제시한 것임.

<그림 2-2> 양파 선형보간 ARIMAX 모형의 잔차 진단 결과



자료: 저자 작성.

- 잔차 시계열을 보면 잔차는 대체로 0을 중심으로 분포하였으나, 일부 시점에 서는 비교적 큰 잔차가 관찰되었음. 이는 특정 시점의 소비량 변동이 모형에 의해 충분히 반영되지 않을 가능성을 시사함.
- 잔차 자기상관함수(ACF)에서는 대부분의 자기상관 계수가 신뢰구간 내에 위치하였음. Ljung-Box 검정 결과에서도 p-value가 0.926으로 나타났음. 따라서 5% 유의수준에서 잔차 자기상관이 통계적으로 유의하지 않은 것으로 확인되어, 자기상관 측면에서는 모형 적합성에 큰 문제가 없는 것으로 판단됨.
- 다만, 잔차 분포와 Q-Q plot에서는 일부 꼬리 부분의 이탈이 확인되었음. 이는 이상치 또는 급격한 소비량 변동의 영향이 일부 존재할 가능성을 의미하며, 해당 모형이 특정 시점의 급변까지 충분히 설명하는 데에는 한계가 있을 수 있음을 시사함.

3.2. 마늘 가구 소비량

3.2.1. 모형 추정 결과

- 마늘 가구 소비량에 대해서도 양파와 동일하게 결측 처리 방식별 예측모형을 구성하였음. 보간 미적용 자료에는 ARIMA 모형을 적용하였으며, 선형보간 자료에는 ETS, ARIMA, ARIMAX 모형을 각각 적용하였음.
- 2023~2024년 학습 구간을 이용하여 추정한 결과, 보간 미적용 ARIMA 모형은 $ARIMA(1,1,2)$ 로 나타났음. 선형보간 자료의 경우 ETS 모형은 $ETS(M,N,N)$, ARIMA 모형은 $ARIMA(2,0,2)$, ARIMAX 모형은 $ARIMAX(4,0,0)$ 으로 각각 추정되었음.

- ARIMAX 모형에는 반입량의 1~3주 시차 변수를 외생변수로 포함하였음.
이는 마늘 소비량 자체의 과거 변동만을 반영하는 단변량 모형과 공급 측면
의 물량 정보를 추가로 고려한 모형 간 예측 성과를 비교하기 위한 설정임.

〈표 2-6〉 결측 처리 방식별 예측모형 추정 결과

결측 처리 방식		추정 구조	주요 설정
보간 미적용	원자료 결측 유지	ARIMA(1,1,2)	소비량의 과거 흐름을 이용한 단변량 모형
보간 적용	선형보간	ETS(M,N,N)	곱셈 오차(M), 추세 없음(N), 계절성 없음(N)
		ARIMA(2,0,2)	소비량의 과거 흐름을 이용한 단변량 모형
		ARIMAX(4,0,0)	반입량의 1~3주 시차변수 포함

자료: 저자 작성.

3.2.2. 예측 정확도 평가 결과

○ 〈표 2-7〉은 2025년 테스트 구간을 대상으로 마늘 소비량 예측모형의 정확도를 비교한 결과임. 예측 정확도는 MAE, RMSE, MAPE를 이용하여 평가하였으며, 평가 대상 관측치는 총 37개임.

○ 분석 결과, 전체 모형 조합 중 최근 52주 자료를 이용하고 반입량의 1~3주 시차 변수를 포함한 선형보간 ARIMAX 모형의 예측오차가 가장 낮게 나타났음. 해당 모형의 MAE는 243.6, RMSE는 311.4, MAPE는 23.2로 나타났음.

○ 학습자료 구성 방식별로 살펴보면, 과거 자료를 누적적으로 이용한 방식에서는 선형보간 ETS 모형의 RMSE가 가장 낮았음. 반면, 최근 52주 자료를 이용한 방식에서는 선형보간 ARIMAX 모형, 최근 72주 자료를 이용한 방식에서는 보간 미적용 ARIMA 모형의 예측오차가 상대적으로 낮게 나타났음.

○ 마늘에서도 양파와 마찬가지로 최근 52주 자료를 활용한 선형보간 ARIMAX 모형의 예측오차가 가장 낮게 나타났으나, MAPE는 23.2%로 양파보다 높은

수준을 보였음. 이는 마늘 소비량 규모가 양파보다 작아 동일한 수준의 절대 오차도 상대적 오차로 크게 산출될 수 있으며, 주별 소비량의 상대적 변동성 차이도 영향을 미쳤을 가능성이 있음.

○ ARIMA 계열 모형을 기준으로 비교하면, 선형보간의 효과와 ARIMAX의 상대적 우수성은 학습창에 따라 일관되게 나타나지 않았음. 다만 최근 52주 자료에서는 동일한 선형보간 자료를 이용한 ARIMAX 모형이 ARIMA 모형보다 낮은 예측오차를 보였음. 따라서 최근 52주 선형보간 ARIMAX 모형의 우수한 성과는 반입량 시차 정보와 최근 소비 흐름을 함께 반영한 조건에서 나타난 결과로 해석할 수 있음.

〈표 2-7〉 마늘 가구 소비량 예측 정확도 평가 결과(N=37)

학습창	결측 처리 방식		추정 구조	MAE	RMSE	MAPE
과거 자료 누적 이용 (Expanding)	보간 미적용	원자료 결측 유지	ARIMA(1,1,2)	288.1	368.5	25.86
	보간 적용	선형보간	ETS(M,N,N)	288.8	347.9	26.49
			ARIMA(2,0,2)	273.8	358.6	24.09
			ARIMAX(4,0,0)	282.6	377.7	24.96
최근 52주 자료 이용 (Rolling_52)	보간 미적용	원자료 결측 유지	ARIMA(1,1,2)	269.7	340.4	25.03
	보간 적용	선형보간	ETS(M,N,N)	295.3	356.0	27.04
			ARIMA(2,0,2)	265.0	329.1	24.85
			ARIMAX(4,0,0)	243.6	311.4	23.15
최근 72주 자료 이용 (Rolling_72)	보간 미적용	원자료 결측 유지	ARIMA(1,1,2)	258.2	325.4	24.47
	보간 적용	선형보간	ETS(M,N,N)	292.3	352.2	26.72
			ARIMA(2,0,2)	271.2	330.9	24.81
			ARIMAX(4,0,0)	266.1	326.5	24.33

자료: 저자 작성.

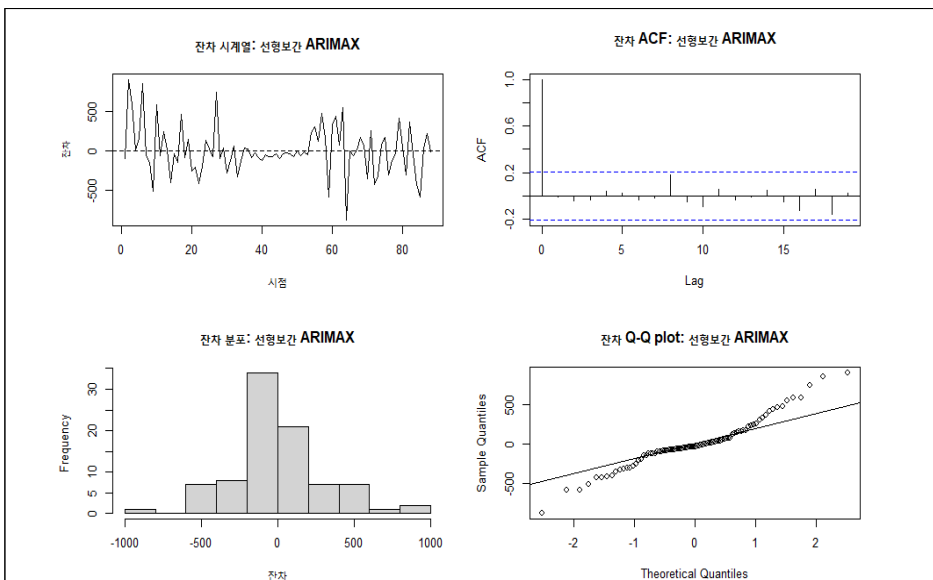
○ 예측 정확도 평가에서 가장 낮은 예측오차를 보인 선형보간 ARIMAX 모형에 대해 잔차 진단을 수행하였음. 〈그림 2-3〉은 2023~2024년 학습 구간에서 추정한 해당 모형의 잔차 시계열, 잔차 자기상관함수(ACF), 잔차 분포 및 Q-Q plot을 제시한 것임.

○ 잔차 시계열을 보면 잔차는 대체로 0을 중심으로 분포하였으나, 일부 시점에
서는 비교적 큰 양(+) 또는 음(-)의 잔차가 관찰되었음. 이는 마늘 소비량의 단
기 변동 중 일부가 모형에 충분히 반영되지 않았을 가능성을 시사함.

○ 잔차 자기상관함수(ACF)를 살펴보면 대부분의 자기상관 계수가 신뢰구간 내
에 위치하였음. Ljung-Box 검정 결과에서도 p-value가 0.897로 나타나 5%
유의수준에서 잔차 자기상관은 통계적으로 유의하지 않은 것으로 확인됨. 따
라서 자기상관 측면에서는 모형 적합성에 큰 문제가 확인되지 않았음.

○ 다만, 잔차 분포와 Q-Q plot에서는 일부 꼬리 부분의 이탈이 관찰되었음. 이
는 잔차가 완전한 정규분포를 따른다고 보기는 어렵고, 이상치 또는 급격한
소비량 변동의 영향이 일부 존재할 가능성을 의미함. 따라서 해당 모형은 특
정 시점의 급변까지 충분히 설명하는 데에는 한계가 있을 수 있음.

〈그림 2-3〉 마늘 선형보간 ARIMAX 모형의 잔차 진단 결과



자료: 저자 작성.

3

제조업·외식업·일반급식업 부문 소비 모형 추정

1. 분석 자료의 특성

1.1. 분석 대상 및 자료 구조

- 본 장에서는 양파와 마늘의 식품제조업, 외식업 및 일반급식업 소비량 자료를 활용하여 산업별 소비량 예측 가능성을 검토하였음. 분석 자료는 업체-월 단위의 반복 관측자료로 구성되어 있으며, 각 업체의 월별 소비량을 예측 대상 변수로 설정하였음.
- 분석 대상은 양파와 마늘 각각에 대해 일반급식업, 식품제조업, 외식업으로 구분되며, 총 6개의 품목·산업별 자료를 활용하였음. 분석 기간은 2023년 4월부터 2025년 12월까지이며, 품목·산업별로 동일한 기간을 적용하였음.
- 본 자료는 모든 업체가 모든 월에 소비량을 보고한 균형패널(balanced pan-

el)이 아니라, 업체별 응답 여부와 소비량 관측 여부가 월별로 상이한 불균형 패널(unbalanced panel) 구조를 가짐. 이하에서는 품목·산업별 패널 규모, 소비량 관측 현황, 결측률 및 소비량 분포 특성을 중심으로 분석 자료의 구조를 검토함.

〈표 3-1〉 분석 대상 자료의 구성

품목	산업	관측 단위	분석기간	예측 대상
양파, 마늘	일반급식업 식품제조업 외식업	업체-월	2023. 4. ~ 2025. 12.	월별 소비량

자료: 저자 작성.

1.2. 품목·산업별 패널 구조와 소비량 관측 현황

- 〈표 3-2〉는 품목·산업별 패널 구조와 소비량 관측 현황을 정리한 것임. 분석 대상 업체 수는 품목과 산업에 따라 차이가 있음. 양파는 일반급식업 733개, 식품제조업 962개, 외식업 1,253개 업체로 구성되었으며, 마늘은 일반급식업 730개, 식품제조업 910개, 외식업 1,033개 업체로 구성됨.
- 전체 업체-월 관측치 수는 양파 외식업이 30,228개로 가장 많았고, 마늘 일반급식업이 16,851개로 가장 적었음. 소비량 관측치 수는 양파 외식업이 24,281개로 가장 많았으며, 마늘 외식업 19,712개, 양파 제조업 16,341개, 마늘 식품제조업 15,229개, 양파 일반급식업 12,967개, 마늘 일반급식업 12,543개로 나타났음.
- 소비량 결측률은 품목·산업별로 약 19.7~27.8% 수준임. 양파 외식업의 결측률은 19.7%로 상대적으로 낮은 반면, 마늘 식품제조업은 27.8%로 가장 높

있음. 이는 분석 자료가 일정 수준의 소비량 결측을 포함한 불균형 패널 구조임을 확인시켜 줌.

- 산업 부문 자료의 결측은 구조적 결측과 조사 기간 중 업체별 응답 누락이 반영된 결과임. 결측률이 높은 경우 모형 추정에 활용되는 유효 관측치가 감소하고, 지속해서 응답한 업체의 특성이 상대적으로 크게 반영될 수 있으므로 산업별 예측성과는 결측 비율과 불균형 패널 구조를 고려하여 해석할 필요가 있음.

〈표 3-2〉 품목·산업별 패널 구조와 소비량 관측 현황

단위: 개, %

품목	산업	업체 수	전체 관측치 수	소비량 관측치 수	소비량 결측치 수	소비량 결측률
양파	일반급식업	733	17,076	12,967	4,109	24.1
	식품제조업	962	22,405	16,341	6,064	27.1
	외식업	1,253	30,228	24,281	5,947	19.7
마늘	일반급식업	730	16,851	12,543	4,308	25.6
	식품제조업	910	21,089	15,229	5,860	27.8
	외식업	1,033	24,885	19,712	5,173	20.8

자료: 저자 작성.

1.3. 품목·산업별 소비량 분포와 월별 변동 특성

- 〈표 3-3〉은 품목·산업별 소비량의 기초통계량을 제시한 것임. 모든 품목·산업에서 평균이 중앙값보다 높고 최댓값이 매우 크게 나타났음. 이는 대부분의 업체-월 소비량은 낮거나 중간 수준에 분포하지만, 일부 업체 또는 특정 시점에서 매우 큰 소비량이 관측되어 전체 평균을 높이는 구조를 의미함. 따라서 본 자료는 일부 큰 소비량 관측치의 영향을 받는 비대칭적 분포 특성을 갖는 것으로 판단됨.

- 산업별 분포 특성을 살펴보면, 식품제조업 자료에서 평균과 중앙값 간 격차가 가장 두드러졌음. 양파 식품제조업체의 평균 소비량은 35,750kg인 반면 중앙값은 4,336kg에 불과하였고, 표준편차는 185,111kg으로 크게 나타났음. 마늘 식품제조업도 평균 20,241kg, 중앙값 2,294kg, 표준편차 109,969kg으로 유사한 패턴을 보였음.
- 외식업에서도 양파의 평균 소비량이 57,599kg, 중앙값 14,368kg으로 나타났으며, 마늘은 평균 16,026kg, 중앙값 4,183kg으로 확인되었음. 일반급식업의 경우 상대적으로 분포 편차가 작았으나, 양파의 최댓값이 200,375kg, 마늘이 51,064kg에 달하는 등 일부 대규모 관측치가 포함되어 있음.
- 이상의 분포 특성은 이후 모형 구축 시 원자료 기준 분석과 함께 로그 변환 기준 보조 분석을 병행해야 하는 근거를 제공함.

〈표 3-3〉 품목·산업별 소비량 기초통계량

단위: 개, kg

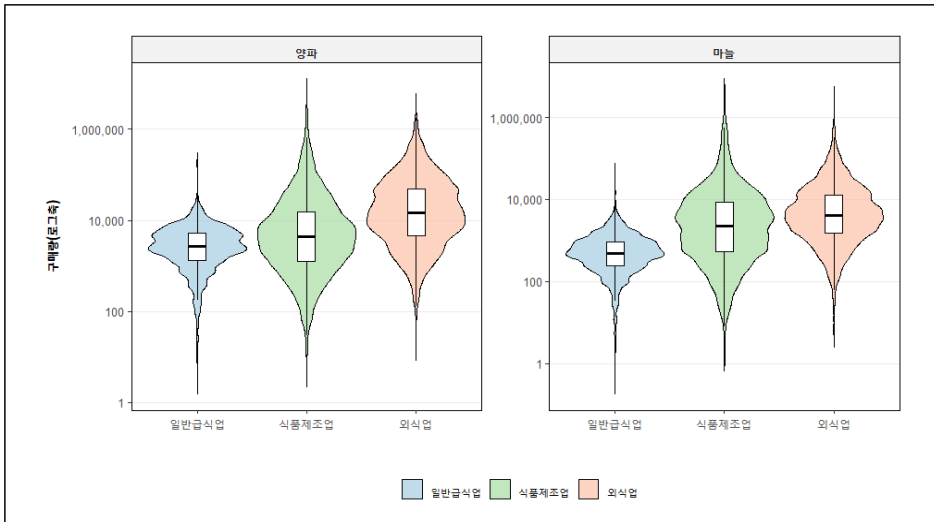
품목	산업	관측치 수	평균	중앙값	표준편차	최솟값	최댓값
양파	일반급식업	12,967	4,135	2,681	7,741	2.3	200,375
	식품제조업	16,341	35,750	4,336	185,111	4.3	6,236,566
	외식업	24,281	57,599	14,368	149,539	15	3,224,612
마늘	일반급식업	12,543	750	495	1,000	0.3	51,064
	식품제조업	15,229	20,241	2,294	109,969	1.4	3,892,111
	외식업	19,712	16,026	4,183	50,309	4.4	3,163,675

자료: 저자 작성.

- 〈그림 3-1〉은 품목·산업별 소비량 분포를 바이올린 플롯(violin plot)과 상자 그림(box plot)으로 나타낸 것임. 소비량 분포 폭이 매우 크기 때문에 로그 척도를 적용하여 표시하였음.

- 그림에서 확인할 수 있듯이, 품목·산업별 소비량은 중앙값 주변에 관측치가 집중되어 있는 반면, 일부 관측치는 높은 소비량 구간까지 분포하는 양상을 보임. 이는 앞서 확인한 평균과 중앙값의 차이, 큰 표준편차 및 최댓값과 일관된 결과임.

〈그림 3-1〉 품목·산업별 소비량 분포



주: 소비량의 분포 폭이 매우 크기 때문에 그림에서는 로그값 기준으로 표시하였음. 바이올린 플롯은 소비량의 분포 형태를, 상자 그림은 중앙값과 사분위 범위를 나타냄.

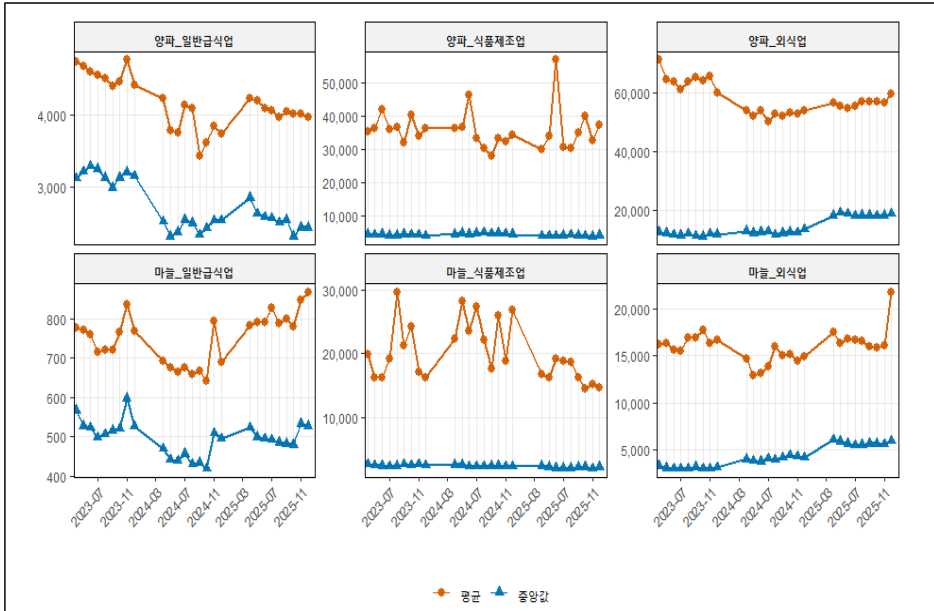
자료: 저자 작성.

- 〈그림 3-2〉는 품목·산업별 월별 소비량 변화를 평균과 중앙값 기준으로 나타낸 것임. 평균은 해당 월에 소비량이 관측된 업체들의 평균 소비량을 의미하며, 중앙값은 대규모 관측치의 영향을 상대적으로 덜 받는 대푯값으로서 분포의 보완적 파악을 위해 함께 제시하였음.

- 품목·산업별 소비량은 분석 기간 동안 일정한 월별 변동을 보였으며, 일부 품목·산업에서는 특정 월에 평균 소비량이 상대적으로 크게 상승하는 구간이 나타났다. 또한, 대부분의 품목·산업에서는 평균이 중앙값보다 높은 수준에서

변동하여, 월별 소비량 변화가 일부 대규모 구매 관측치의 영향을 함께 받고 있음을 시사함.

〈그림 3-2〉 품목·산업별 소비량 변화



자료: 저자 작성.

○ 이상의 결과를 종합하면, 품목·산업별 소비량 자료는 업체-월 단위의 불균형 패널 구조를 가지며, 소비량 결측과 업체별 구매 규모 차이가 함께 존재하는 것으로 나타났음. 또한, 일부 대규모 구매 관측치의 영향으로 평균이 중앙값보다 높게 나타나, 소비량 자료는 비대칭적 분포 특성을 보이는 것으로 판단됨.

○ 따라서 이후 예측모형 구축에서는 업체별 과거 구매 수준, 월별 변동성 및 소비량 분포의 비대칭성을 함께 고려할 필요가 있음. 본 연구에서는 원자료 기준 예측모형을 기본으로 설정하되, 로그 변환 소비량 기준의 보조 분석을 병행하였음.

2. 모형 구축 전략

2.1. 분석 표본 및 예측변수의 구성

- 분석 표본은 양파와 마늘 각각에 대해 일반급식업, 식품제조업, 외식업으로 구분된 6개의 품목·산업별 불균형 패널자료로 구성되며, 관측 단위는 업체-월임.
- 예측 대상 변수는 원자료 소비량과 로그 변환 소비량으로 구분하였음. 원자료 기준 분석은 예측 결과를 실제 소비량 단위로 직접 해석할 수 있다는 장점이 있음. 한편, 로그 변환 기준 분석은 소비량 자료에서 확인된 평균과 중앙값 간 차이, 일부 대규모 구매 관측치의 영향 등 분포의 비대칭성을 완화하기 위한 보완적 접근으로 병행하였음.
- 로그 변환 모형에서 산출된 예측값은 원자료 소비량 단위로 환산하여 예측 정확도를 평가하였음. 이때 로그 예측값에 단순 지수변환을 적용하면 원자료 단위의 평균 예측값이 과소 산출될 가능성이 있음.
- 이에 본 연구에서는 학습자료에서 산출된 로그 모형 잔차의 분산을 반영하기 위해 역변환 보정을 적용함.
 - 구체적으로 로그 예측값을 $\hat{y}_t$, 학습자료의 로그 잔차 분산을 $\hat{\sigma}^2$ 라 할 때, 원자료 단위 예측값은 $\hat{Y}_t = \exp(\hat{y}_t + \hat{\sigma}^2/2)$ 로 산출하였음.
- 설명변수는 6개 품목·산업 자료에 공통 적용 가능한 변수로 구성하였으며, 업체별 과거 평균 소비량, 종사자 수, 전월 도매가격, 월 더미를 활용하였음. 각

변수의 구성 및 활용 방식은 <표 3-4>와 같음.

- 업체별 과거 평균 소비량은 각 예측 시점 이전까지 해당 업체에서 관측된 소비량의 평균으로 산출하였으며, 업체별 기본 구매 수준과 과거 구매 이력을 반영함. 다만, 관측 이력이 충분하지 않은 경우에는 해당 시점까지의 전체 평균을 보완적으로 활용함.
- 종사자 수는 업체 규모에 따른 소비량 차이를 반영하기 위해 포함하였음.
- 전월 도매가격은 예측 시점에서 이용 가능한 가격 여건을 반영하기 위해 포함하였고, 월 더미는 계절적 구매 패턴의 가능성을 고려하여 포함하였음.

○ 본 연구에서 활용한 설명변수는 예측 시점에서 실무적으로 확보 가능한 정보를 중심으로 구성하였음.

<표 3-4> 예측모형에 활용한 변수 구성

구분	변수	변수 구성 및 활용
예측 대상 변수	월별 소비량	업체·품목별 소비량, 원자료 및 로그변환 기준 예측모형의 종속변수로 활용
	로그 변환 소비량	
설명변수	업체별 과거 평균 소비량	예측 시점 이전까지의 업체별 평균 소비량, 업체별 기본 구매 수준과 과거 구매 이력 반영
	종사자 수	업체 규모에 따른 소비량 차이를 반영
	전월 도매가격	예측 시점에서 이용 가능한 가격 여건을 반영
	월 더미	월에 따라 소비량 수준이 달라질 가능성을 반영

자료: 저자 작성.

2.2. 모형의 설정

○ 본 연구에서는 품목·산업별 업체-월 소비량 예측을 위해 단순 기준 모형, 패널모형, 머신러닝 모형을 비교모형으로 설정하였음. 이는 서로 다른 예측 접

근법을 불균형 패널자료에 적용하고, 품목·산업별 자료 특성에 따라 상대적으로 안정적인 예측 성과를 보이는 모형을 검토하기 위한 것임.

○ 단순 기준 모형으로는 업체별 과거 평균 모형을 설정하였음. 이 모형은 예측 시점 이전까지 관측된 해당 업체의 과거 평균 소비량을 이용하여 향후 소비량을 예측값으로 사용하는 방식임. 별도의 모수 추정을 요구하지 않으면서도 업체별 구매 수준의 지속성을 반영할 수 있다는 점에서 불균형 패널자료의 기준 모형으로 적합함.

○ 패널회귀모형으로는 관측되지 않는 업체 고유 특성이 소비량에 영향을 미치고, 이러한 특성이 설명변수와 상관될 가능성을 고려하여 본 연구에서는 고정효과 모형(Fixed Effects model)을 적용하였음. 고정효과 모형은 시간에 따라 변하지 않는 업체별 특성을 통제된 상태에서 업체 내부의 시점 변화와 소비량 간의 관계를 추정하는 모형임.

- 확률효과 모형(Random Effects model)은 업체별 개체효과와 설명변수 간 상관관계가 없음을 전제로 함. 그러나 본 연구의 자료에서는 이러한 무상관 가정을 전제하기 어렵다고 판단하여 고정효과모형을 중심으로 분석하였음.

○ 머신러닝 모형으로는 CatBoost 모형을 적용하였음. CatBoost는 의사결정나무 기반의 gradient boosting 앙상블 모형으로, 순차적으로 생성되는 의사결정나무가 이전 단계의 예측오차를 보완하는 방식으로 학습되며, 변수 간 비선형 관계와 상호작용 효과를 반영하는 데 유용함(Martyushev et al., 2025).

- 본 연구는 업체 ID, 월 정보, 업체별 과거 평균 소비량, 종사자 수, 전월 도매가격을 입력변수로 활용하였음. 다만, CatBoost는 고정효과 모형과 달리 업체별 개체효과를 명시적으로 추정하지 않으므로, 본 연구에서는 계수 해석보다 예측 가능성 검토를 목적으로 한 비선형 예측모형으로 활용함.

2.3. 학습 및 테스트 자료의 설정 및 예측력 평가 방법

- 표본 외 예측력을 평가하기 위해 분석 자료를 학습 구간과 테스트 구간으로 구분하였음. 학습 구간(2023년 4월부터 2024년 12월)과 테스트 구간(2025년 4월부터 2025년 12월)으로 설정하였음. 2025년 1~3월은 소비량 자료가 구조적으로 결측되어 평가 대상에서 제외하였음.
- <표 3-5>는 품목·산업별 소비량 관측치 수를 학습 구간과 테스트 구간으로 구분하여 제시한 것임. 학습 구간 관측치 수는 모형 추정에 활용된 자료 규모를, 테스트 구간의 평가 가능 관측치 수는 실제 소비량이 관측되어 예측오차 산출에 활용된 업체-월 자료를 의미함.

<표 3-5> 품목·산업별 학습 및 테스트 자료 구성

품목	산업	관측치 수	학습 구간	테스트 구간
양파	일반급식업	12,967	8,422	4,545
	식품제조업	16,341	11,018	5,323
	외식업	24,281	15,916	8,365
마늘	일반급식업	12,543	8,056	4,487
	식품제조업	15,229	10,258	4,971
	외식업	19,712	12,926	6,786

자료: 저자 작성.

- 예측 방식은 확장형 학습창을 적용한 1개월 선행 예측 방식으로 설정하였음. 2025년 4월 예측은 2023년 4월부터 2024년 12월까지의 초기 학습자료를 이용하였으며, 이후에는 초기 학습 구간에서 결정된 동일한 모형 구조를 유지한 상태에서 매월 직전 테스트 월의 신규 관측치를 학습자료에 순차적으로 추가하여 모수를 재추정하고 다음 1개월의 값을 예측하였음.
 - 이 방식은 예측 시점에서 이용 가능한 정보만을 활용한다는 점에서 실무적 예측 절차와 부합하며, 제한된 패널 기간에서 가용 정보를 점진적으로 활

용할 수 있다는 장점이 있음.

- 예측 정확도는 테스트 구간 중 실제 소비량과 예측값이 모두 존재하는 업체-월 기준으로 평가하였음. 소비량 결측이 존재하는 업체-월은 예측오차 산출 대상에서 제외하였으며, 평가 지표로는 RMSE와 MAE를 사용하였음.
- 산업 부문 소비 자료는 업체별 소비 규모 차이가 크고 소비량이 매우 작은 관측치가 다수 존재하므로, 본 장에서는 실제값이 작은 관측치에서 오차가 과대 평가될 수 있는 MAPE는 평가 지표에서 제외하였음.

3. 모형 추정 및 예측력 평가 결과

3.1. 양파 산업별 소비량 예측 결과

- <표 3-6>은 양파 산업별 소비량 예측모형의 예측 정확도 평가 결과를 제시한 것임. 원자료 기준 분석과 로그 변환 기준 분석으로 구분하여 RMSE와 MAE를 평가하였음.
- 일반급식업에서는 원자료 기준 CatBoost 모형의 MAE가 1,541로 가장 낮았으며, 업체별 과거 평균 모형(1,652)과 고정효과 모형(1,664)이 유사한 수준으로 뒤를 이었음. 다만 모형 간 예측오차 차이가 크지 않아, 일반급식업에서는 복잡한 모형 적용에 따른 예측력 개선 효과가 제한적이었음.
- 식품제조업의 경우 원자료 기준 업체별 과거 평균 모형의 MAE가 23,765로 가장 낮았으며, 고정효과 모형(24,266)과 CatBoost(24,332)가 근접한 수준으

로 뒤를 이었음. 로그 변환 기준에서는 고정효과 모형의 MAE가 28,501로 로그 변환을 적용한 모형 중 가장 낮았으나, 원자료 기준 모형들보다는 예측오차가 오히려 높게 나타나 로그 변환에 따른 예측력 개선 효과는 확인되지 않음.

○ 외식업에서는 원자료 기준 CatBoost 모형의 MAE가 26,904로 가장 낮았으며, 고정효과 모형(28,239)과 과거 평균 모형(31,313)이 뒤를 이었음. 로그 변환 기준에서는 고정효과 모형의 MAE가 30,157로 로그 변환 모형 중 가장 낮았으나 원자료 기준 CatBoost보다는 높은 수준이었으며, 로그 변환 기준 CatBoost(41,072)와 업체별 과거 평균 모형(64,010)은 원자료 기준 대비 예측 오차가 크게 증가하였음.

〈표 3-6〉 양파 산업별 소비량 예측 정확도 평가 결과

산업	분석구분	모형	RMSE	MAE
일반급식업	원자료	업체별 과거 평균 모형	4,062	1,652
		고정효과 모형	4,090	1,664
		CatBoost	4,071	1,541
	로그 변환	업체별 과거 평균 모형	4,134	1,822
		고정효과 모형	4,289	1,708
		CatBoost	5,662	1,605
식품제조업	원자료	업체별 과거 평균 모형	156,820	23,765
		고정효과 모형	156,210	24,266
		CatBoost	158,680	24,332
	로그 변환	업체별 과거 평균 모형	235,739	43,970
		고정효과 모형	188,751	28,501
		CatBoost	169,200	29,071
외식업	원자료	업체별 과거 평균 모형	91,690	31,313
		고정효과 모형	81,191	28,239
		CatBoost	74,882	26,904
	로그 변환	업체별 과거 평균 모형	173,239	64,010
		고정효과 모형	80,012	30,157
		CatBoost	197,908	41,072

자료: 저자 작성.

3.2. 마늘 산업별 소비량 예측 결과

- <표 3-7>은 마늘 산업별 소비량 예측모형의 예측 정확도 평가 결과를 제시한 것임. 양파와 동일하게 원자료 소비량 기준 분석과 로그 변환 소비량 기준 분석을 구분하여 평가하였음.
- 일반급식업에서는 원자료 기준 CatBoost 모형의 MAE가 287로 가장 낮았으며, 고정효과 모형(310)과 업체별 과거 평균 모형(330) 순으로 나타났음. 로그 변환 기준 분석에서도 CatBoost 모형의 MAE(301)가 가장 낮았으나, 원자료 기준 대비 다소 높게 나타나 일반급식업에서는 로그 변환의 추가적 개선 효과가 제한적이었음.
- 식품제조업에서는 원자료 기준 CatBoost 모형의 MAE가 11,767로 전체 모형 중 가장 낮았으며, 업체별 과거 평균 모형(12,071)과 고정효과 모형(14,062)이 뒤를 이었음. 로그 변환 기준에서는 CatBoost 모형의 MAE가 12,380으로 로그 변환 모형 중 가장 낮았으며, 원자료 기준 CatBoost 모형보다는 오차가 높게 나타났음. 로그 변환 기준 업체별 과거 평균 모형(26,114)과 고정효과 모형(13,797)은 원자료 기준 대비 예측오차가 확대되어 마늘 식품제조업에서도 로그 변환에 따른 예측력 개선 효과는 확인되지 않았음.
- 외식업에서는 원자료 기준 CatBoost 모형의 MAE가 9,021로 가장 낮았으며, 고정효과 모형(9,938)과 업체별 과거 평균 모형(10,659)이 뒤를 이었음. 로그 변환 기준에서는 CatBoost 모형의 MAE가 9,119로 로그 변환 모형 중 가장 낮아 원자료 기준과 근접한 수준을 보였으나, 여전히 원자료 기준 CatBoost보다는 높았음. 로그 변환 기준 고정효과 모형(14,012)과 업체별 과거 평균 모형(21,274)은 원자료 기준 대비 오차가 크게 확대되었음.

○ 한편, RMSE의 경우 해당 지표가 상대적으로 큰 오차에 민감하게 반응하므로, 양파와 마늘 두 품목 모두 원자료 기준에서 대규모 소비 관측치의 예측 성과가 결과에 보다 크게 반영될 수 있음을 보여주고 있음.

○ 양파와 마늘에 대한 산업별 우수 모형의 차이는 소비 규모와 변동성, 업체 간 이질성 및 결측 구조 등 자료 특성이 복합적으로 반영된 결과로 볼 수 있음. 일반급식업과 외식업에서는 업체별 특성과 변수 간 비선형 관계를 반영하는 CatBoost가 상대적으로 우수하였으며, 양파 식품제조업에서는 업체별 소비 수준의 지속성을 반영한 과거 평균 모형이 가장 우수하게 나타났음. 다만, 이는 자료 특성과 예측 성과를 연계한 해석으로 개별 요인의 영향을 직접 식별한 결과는 아님.

〈표 3-7〉 마늘 산업별 소비량 예측 정확도 평가 결과

산업	분석구분	모형	RMSE	MAE
일반급식업	원자료	업체별 과거 평균모형	699	330
		고정효과 모형	645	310
		CatBoost	567	287
	로그 변환	업체별 과거 평균모형	727	369
		고정효과 모형	857	361
		CatBoost	636	301
식품제조업	원자료	업체별 과거 평균모형	57,899	12,071
		고정효과 모형	58,819	14,062
		CatBoost	51,910	11,767
	로그 변환	업체별 과거 평균모형	106,194	26,114
		고정효과 모형	65,371	13,797
		CatBoost	53,534	12,380
외식업	원자료	업체별 과거 평균모형	48,926	10,659
		고정효과 모형	47,108	9,938
		CatBoost	44,100	9,021
	로그 변환	업체별 과거 평균모형	69,196	21,274
		고정효과 모형	78,944	14,012
		CatBoost	44,473	9,119

자료: 저자 작성.

○ 종합하면, 마늘과 양파는 원자료 기준 CatBoost 모형이 상대적으로 낮은 MAE를 기록하였으며, 로그 변환 기준 모형은 일부 사례에서 원자료 기준에 근접한 수준을 보였으나 이를 상회하는 개선 효과는 나타나지 않았음.

4

학교·군 급식 부문 소비 모형 추정

1. 분석 자료의 특성

- 양파와 마늘 급식 소비량 자료는 aT와 농협에서 제공하는 월별 학교 및 군 급식 소비량 자료를 활용하였음. 본 연구에서는 학교 급식 소비량과 군 급식 소비량을 합산한 품목별 급식 소비량을 예측 대상으로 설정하였음.
- 원자료는 2019년 1월부터 2025년 12월까지의 월별 자료로 구성되어 있으나, 군 급식 소비량은 2020년부터 관측됨. 따라서 학교 급식과 군 급식을 합산한 소비량을 일관되게 분석하기 위해 2020년 1월 이후 자료를 분석 대상으로 설정하였음. 이에 따라 양파와 마늘 모두 2020년 1월부터 2025년 12월까지 총 72개 월별 관측치가 활용되었음.
- 앞서 검토한 가구 소비량 자료가 조사 미실시에 따른 구조적 결측을 포함한 주별 자료인 것과 달리 급식 소비량 자료는 월별 자료가 비교적 연속적으로

확보된다는 점에서 자료 구조상 차이가 있음. 따라서 급식 소비량 자료는 결측 처리보다는 월별 시계열의 변동 양상과 예측 가능성을 검토하는 것이 보다 중요한 분석 과제라 할 수 있음.

○ 단, 학교 급식과 군 급식은 모두 급식 운영 과정에서 발생하는 소비량이라는 공통점을 가지지만, 소비 발생 구조는 서로 다를 수 있음. 학교 급식은 학기와 방학, 개학 시기, 급식 일수의 영향을 크게 받는 반면, 군 급식은 상대적으로 연중 지속적인 급식 운영 구조를 가질 가능성이 존재함. 본 연구에서는 자료의 특성과 분석의 안정성을 고려하여 두 시계열을 합산하여 분석하였으나, 해석 시 두 급식 부문의 계절성이 완전히 동일하다고 보기는 어렵다는 점에 유의할 필요가 있음.

○ <표 4-1>은 양파와 마늘 학교·군 급식 소비량의 기초통계량을 제시한 것임. 양파 학교·군 급식 소비량은 평균 2,161.2톤, 중앙값 2,227.0톤, 표준편차 372.9톤으로 나타났으며, 마늘은 평균 407.3톤, 중앙값 406.8톤, 표준편차 62.3톤으로 나타났음.

<표 4-1> 품목별 학교·군 급식 소비량의 기초통계량

단위: 개, 톤

품목	유효 관측 수	평균	중앙값	표준편차	최솟값	최댓값
양파	72	2,161.2	2,227.0	372.9	1,200.0	3,441.7
마늘		407.3	406.8	62.3	256.0	611.0

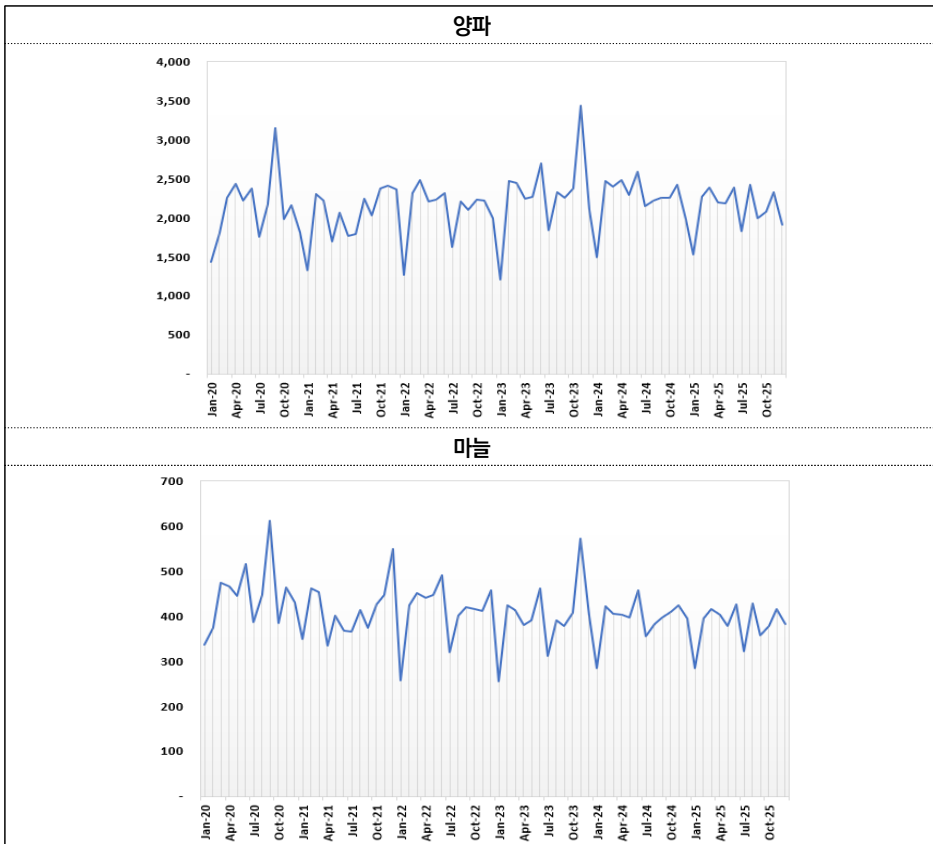
자료: 저자 작성.

○ <그림 4-1>은 2020~2025년 양파와 마늘 학교·군 급식 소비량의 월별 변화 추이를 나타낸 것임. 양파와 마늘 모두 분석 기간 동안 뚜렷한 장기 상승 또는 하락 추세가 지속적으로 나타나기보다는 일정 범위 내에서 월별 등락을 반복하는 양상을 보였음.

○ 이러한 변동은 급식 운영 일정, 방학 및 개학 시기, 월별 급식 일수, 식단 구성 등 급식 수요의 월별 변화와 관련될 수 있음. 따라서 급식 소비량 자료는 장기 추세보다는 월별 변동성과 반복적 패턴을 고려하여 예측모형을 설정할 필요가 있음.

〈그림 4-1〉 품목별 학교·군 급식 소비량의 월별 변화 추이

단위: 톤



자료: 저자 작성.

2. 모형 구축 전략

- 앞서 제시한 분석 자료의 특성을 고려하여 본 연구는 양파와 마늘 학교·군 급식 소비량에 대해 월별 시계열 예측모형을 구축하였음. 급식 소비량 자료는 2020년 이후 월별 관측치가 비교적 연속적으로 확보되어 있어 가구 소비량 자료와 달리 별도의 결측 보간 절차를 적용하지 않았음.
- 이에 급식 소비량 분석에서는 결측 처리 방식의 비교보다는 월별 시계열 자료에 대한 예측모형 간 성과를 비교하는 방식으로 예측 전략을 설정하였음. 구체적으로 월별 반복 패턴을 반영하는 계절성 단순 기법(seasonal naive method), ETS 지수평활 모형, ARIMA 모형을 비교 대상으로 설정하였음.
- 본 연구의 예측 전략은 다음의 세 단계로 구성됨. 첫째, 월별 급식 소비량 자료에 적용 가능한 예측모형을 설정함. 둘째, 2020~2024년 자료를 학습 구간으로, 2025년 자료를 테스트 구간으로 구분함. 셋째, 테스트 구간에 대해 순차적으로 1개월 앞 예측을 수행하고 예측 정확도를 비교·평가함.

2.1. 모형의 설정

- 본 연구에서는 양파와 마늘 학교·군 급식 소비량의 월별 예측을 위해 계절성 단순 기법(SNaive), ETS 지수평활, ARIMA, ARIMAX 모형을 적용하였음. 이들 모형은 월별 급식 소비량의 반복적 변동, 최근 수준 변화, 자기상관 구조 및 공급 측면의 정보를 반영하기 위한 비교 모형으로 설정되었음.
- 계절성 단순 기법은 전년 동월의 관측값을 해당 월의 예측값으로 사용하는 방

법임. 급식 소비량은 방학, 개학, 월별 급식일수, 식단 구성 등 급식 운영 여건에 따라 월별 반복성이 나타날 수 있으므로 계절성 단순 기법을 기준 모형으로 설정하였음.

- ETS 지수평활 모형은 급식 소비량의 최근 수준 변화와 월별 반복 변동을 반영하기 위해 적용하였음. ARIMA 모형은 급식 소비량 자체의 과거 흐름과 자기상관 구조를 반영하기 위한 모형으로 포함하였음.
- ARIAMX 모형은 ARIMA 구조에 외생변수를 추가한 확장 모형으로 본 연구에서는 도매시장 반입량의 시차 변수를 외생변수로 포함하였음. 이는 급식 소비량 자체의 과거 흐름만을 반영하는 단변량 모형과 공급 측면의 물량 정보를 추가로 고려한 모형 간 예측 성과를 비교하기 위한 설정임.
- 급식 소비량 예측에 적용한 모형 구성은 <표 4-2>와 같음.

<표 4-2> 학교·군 급식 소비량 예측모형 구성

구분	예측모형	적용 목적
단순 기준 모형	계절성 단순 기법(SNaive)	전년 동월 소비량을 기준으로 월별 반복 패턴 반영
시계열 모형	ETS 지수평활	최근 소비량 수준과 계절 변동 반영
	ARIMA	과거 소비량의 자기상관 구조 반영
	ARIMAX	반입량 시차 변수를 이용한 공급 측면 정보 반영

자료: 저자 작성.

2.2. 학습·테스트 자료의 설정 및 예측력 평가 방법

- 본 연구에서는 예측모형의 학습과 예측력 평가를 구분하기 위해 분석 자료를 학습 구간과 테스트 구간으로 나누어 활용하였음. 학습 구간은 예측모형의 구조를 식별하고 모수를 추정하는 데 사용되는 자료이며, 테스트 구간은 설정된 모형의 표본 외 예측 성과를 평가하기 위한 자료임.
- 양파와 마늘 학교·군 급식 소비량 자료는 2020년 1월부터 2025년 12월까지 총 72개 월별 관측치로 구성됨. 본 연구에서는 2020년 1월~2024년 12월 자료를 학습 구간으로 설정하고, 2025년 1~12월 자료를 테스트 구간으로 설정하였음. 이에 따라 학습 구간은 60개월, 테스트 구간은 12개월로 구성됨.
- 먼저 초기 학습 구간 자료를 이용하여 각 모형의 구조를 결정하였음. 즉, ETS 지수평활 모형의 구성요소와 ARIMA 계열 모형의 차수는 2020~2024년 학습자료를 기준으로 설정하였음. 이후 테스트 구간 예측 과정에서는 초기 학습 구간에서 결정된 동일한 모형 구조를 유지한 상태에서 예측 시점별 학습자료를 이용하여 모수를 재추정하였음.
- 테스트 구간의 예측은 고정 길이 이동 창 방식을 적용한 1개월 선행 예측 방식으로 수행하였음. 이는 2025년 각 예측 시점마다 고정 길이 이동 창 방식에 의해 구성된 자료를 이용하여, 초기 학습 구간에서 결정된 모형의 구조를 유지한 상태에서 모수를 재추정하고 다음 1개월의 값을 예측하는 방식임.
- 학습자료 구성 방식은 최근 60개월, 48개월, 36개월 자료를 각각 이용하는 고정 길이 이동 창 방식을 설정하였음. 고정 길이 이동 창 방식은 예측 시점이 이동함에 따라 학습자료의 종료 시점도 함께 이동하되, 학습자료의 길이는 사전에 설정한 기간으로 고정하는 방식임.

- 예측 정확도 평가 지표는 평균절대오차(MAE), 평균제곱근오차(RMSE), 평균 절대백분율오차(MAPE)를 사용하였음. 세 지표는 모두 값이 작을수록 예측오차가 작고 예측 성과가 우수한 것으로 해석함.

3. 모형 추정 및 예측력 평가 결과

3.1. 양파 학교·군 급식 소비량

3.1.1. 모형 추정 결과

- <표 4-3>은 양파 학교·군 급식 소비량을 대상으로 2020년 1월~2024년 12월의 학습 구간을 이용하여 예측모형 구조를 식별한 결과를 제시한 것임. 분석에는 계절성 단순 기법, ETS 지수평활, 계절 ARIMA·ARIMAX를 적용함.
- 추정 결과, ETS 지수평활 모형은 ETS(M,N,M)로 식별되었음. 이는 양파 학교·군 급식 소비량 예측에서 추세 성분보다는 수준과 계절 변동을 중심으로 반영하는 구조가 선택되었음을 의미함.
- 계절 ARIMA 모형은 SARIMA(0,0,0)(1,1,0)[12]로 식별되었음. 이는 비계절 자기회귀 및 이동평균항은 포함되지 않았으나, 월별 자료의 계절적 변동을 반영하기 위해 계절 차분과 계절 자기회귀항이 포함된 구조임.
- 계절 ARIMAX 모형은 SARIMAX(0,0,0)(0,1,1)[12] 구조에 반입량의 1~2개월 시차 변수를 포함한 형태로 설정되었음. 이는 양파 학교·군 급식 소비량 자체의 계절적 변동 구조에 더하여 반입량의 시차 정보를 추가적으로 고려하기 위한 설정임.

〈표 4-3〉 양파 학교·군 급식 소비량 예측모형 추정 결과

구분	추정 구조	주요 설정
계절성 단순 기법	Seasonal Naive(period=12)	전년 동월 소비량을 예측값으로 활용
ETS 지수평활 모형	ETS(M,N,M)	급셈 오차(M), 추세 없음(N), 급셈 계절성(M)
계절 ARIMA 모형	SARIMA(0,0,0)(1,1,0)[12]	계절 차분 및 계절 자기회귀 구조 반영
계절 ARIMAX 모형	SARIMAX(0,0,0)(0,1,1)[12]	반입량 1~2개월 시차 변수 포함

자료: 저자 작성.

3.1.2. 예측 정확도 평가 결과

- 〈표 4-4〉는 2025년 1~12월 테스트 구간을 대상으로 양파 학교·군 급식 소비량의 예측 정확도를 평가한 결과를 제시한 것임. 예측 정확도는 MAE, RMSE, MAPE를 기준으로 비교하였으며, 학습자료 구성 방식은 최근 60개월, 48개월, 36개월 자료를 이용하는 고정 길이 이동 창 방식으로 구분하였음.
- 분석 결과, 전체 모형 중 최근 48개월 자료를 이용한 ETS 지수평활 모형의 예측 성과가 가장 우수한 것으로 나타났음. 해당 모형의 MAE는 127.6, RMSE는 162.3, MAPE는 6.19로 모든 평가 지표에서 가장 낮은 값을 보였음.
- 모형별로는 ETS 지수평활 모형과 SARIMAX 모형이 상대적으로 낮은 예측오차를 보였음. 특히 최근 60개월 자료를 이용한 SARIMAX 모형은 MAE 137.0, RMSE 171.1, MAPE 6.68로 전체 모형 중 두 번째로 낮은 예측오차를 나타냈음. 반면, SARIMA 모형은 모든 학습창에서 상대적으로 예측오차가 크게 나타났음.
- 학습창 설정 측면에서는 ETS 지수평활 모형의 경우 최근 48개월 자료를 이용했을 때 예측오차가 가장 낮았으며, SARIMAX 모형은 최근 60개월 자료를 이용했을 때 가장 낮은 예측오차를 보였음. 이는 양파 학교·군 급식 소비량 예

측에서 모형별로 적정 학습창 길이가 다르게 나타날 수 있음을 시사함.

〈표 4-4〉 양파 학교·군 급식 소비량 예측 정확도 평가 결과(N=12)

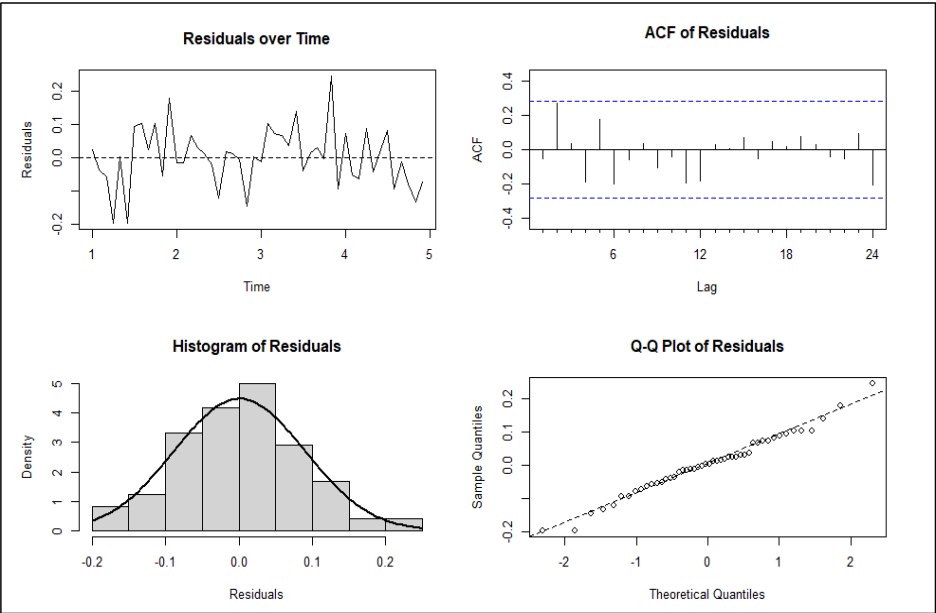
학습창	구분	추정 구조	MAE	RMSE	MAPE
계절성 단순 기법		Seasonal Naive(period=12)	165.3	191.3	7.83
최근 60개월 자료 이용 (Rolling_60)	ETS 지수평활 모형	ETS(M,N,M)	154.6	194.7	7.43
	계절 ARIMA 모형	SARIMA(0,0,0)(1,1,0)[12]	209.0	250.2	9.89
	계절 ARIMAX 모형	SARIMAX(0,0,0)(0,1,1)[12]	137.0	171.1	6.68
최근 48개월 자료 이용 (Rolling_48)	ETS 지수평활 모형	ETS(M,N,M)	127.6	162.3	6.19
	계절 ARIMA 모형	SARIMA(0,0,0)(1,1,0)[12]	212.2	262.2	9.96
	계절 ARIMAX 모형	SARIMAX(0,0,0)(0,1,1)[12]	147.3	176.4	7.11
최근 36개월 자료 이용 (Rolling_36)	ETS 지수평활 모형	ETS(M,N,M)	150.7	177.8	7.11
	계절 ARIMA 모형	SARIMA(0,0,0)(1,1,0)[12]	232.2	333.5	10.9
	계절 ARIMAX 모형	SARIMAX(0,0,0)(0,1,1)[12]	175.9	219.3	8.40

주: 계절성 단순 기법은 전년 동월 값을 예측값으로 사용하므로 학습창 길이와 관계없이 동일한 결과가 산출됨.
자료: 저자 작성.

- 예측 정확도 평가에서 가장 낮은 예측오차를 보인 최근 48개월 이동 창 기반 ETS 지수평활 모형에 대해 잔차 진단을 수행하였음. 〈그림 4-2〉는 해당 모형의 잔차 시계열, 잔차 자기상관함수(ACF), 잔차 분포 및 Q-Q plot을 제시한 것임.
- 잔차는 대체로 0을 중심으로 분포하였으며 특정 방향으로 치우치는 양상은 뚜렷하게 나타나지 않았음. 다만, 일부 시점에서는 상대적으로 큰 잔차가 관찰되어 특정 월의 소비량 변동이 모형에 충분히 반영되지 않았을 가능성이 있음.
- 잔차 자기상관함수(ACF)에서는 대부분의 자기상관 계수가 신뢰구간 내에 위치하였음. Ljung-Box 검정 결과에서도 p-value가 0.355로 나타나 5% 유의수준에서 잔차 자기상관은 통계적으로 유의하지 않는 것으로 확인되었음.
- 잔차 분포와 Q-Q plot에서는 대부분의 관측치가 기준선 주변에 위치하였음.

나, 일부 꼬리 구간에서 이탈이 관찰되었음. 이는 특정 시점의 급식 소비량 변동이 모형에서 충분히 설명되지 않았을 가능성을 시사함.

〈그림 4-2〉 양파 학교·군 급식 소비량 ETS 모형의 잔차 진단 결과



자료: 저자 작성.

3.2. 마늘 학교·군 급식 소비량

3.2.1. 모형 추정 결과

- <표 4-5>는 마늘 학교·군 급식 소비량을 대상으로 2020년 1월~2024년 12월의 학습 구간을 이용하여 모형의 구조를 식별한 결과를 제시한 것임. 분석에는 양파와 동일하게 계절성 단순 기법, ETS 지수평활, 계절 ARIMA·ARIMAX 모형을 적용하였음.
- 추정 결과, ETS 지수평활 모형은 ETS(M,N,A)로 식별되었음. 이는 오차항은 곱셈형, 추세 성분은 포함하지 않으며, 계절성은 덧셈형으로 반영하는 구조임. 따라서 마늘 학교·군 급식 소비량 예측에서는 장기 추세보다는 월별 계절변동을 중심으로 한 구조가 선택된 것으로 볼 수 있음.
- 계절 ARIMA 모형은 ARIMA(0,0,0)(1,0,0)[12]로 식별되었음. 이는 비계절 자기회귀 및 이동평균항은 포함하지 않고, 월별 자료의 계절 자기회귀 구조를 반영하는 형태임.
- 계절 ARIMAX 모형은 SARIMAX(0,0,0)(1,0,0)[12] 구조에 반입량 1~2개월 시차 변수를 포함한 형태로 설정되었음. 이는 마늘 학교·군 급식 소비량 자체의 계절적 자기상관 구조에 더하여 반입량의 시차 정보를 함께 고려하고 있음.

<표 4-5> 마늘 학교·군 급식 소비량 예측모형 추정 결과

구분	추정 구조	주요 설정
계절성 단순 기법	Seasonal Naive(period=12)	전년 동월 소비량을 예측값으로 활용
ETS 지수평활 모형	ETS(M,N,A)	곱셈 오차(M), 추세 없음(N), 덧셈 계절성(A)
계절 ARIMA 모형	SARIMA(0,0,0)(1,0,0)[12]	계절 자기회귀 구조 반영
계절 ARIMAX 모형	SARIMAX(0,0,0)(1,0,0)[12]	반입량 1~2개월 시차 변수 포함

자료: 저자 작성.

3.2.2. 예측 정확도 평가 결과

- <표 4-6>은 2025년 1~12월 테스트 구간을 대상으로 마늘 학교·군 급식 소비량의 예측 정확도 평가 결과를 제시한 것임.
- 분석 결과, 계절성 단순 기법의 예측 성과가 가장 우수한 것으로 나타났음. 해당 모형의 MAE는 21.5, RMSE는 26.2, MAPE는 5.59로 나타나 전체 모형 중 가장 낮은 예측오차를 보였음. 이는 마늘 학교·군 급식 소비량 예측에서 전년 동월 소비량에 기반한 월별 반복 패턴이 중요한 예측 정보를 제공하고 있음을 시사함.
- 그 외 모형에서는 최근 48개월 자료를 이용한 ETS 지수평활 모형의 예측오차가 상대적으로 낮게 나타났음. 해당 모형의 MAE는 23.1, RMSE 26.6, MAPE 5.97로 계절성 단순 기법과 유사한 수준의 예측 성과를 보였음. 반면, SARIMA와 SARIMAX 모형은 학습창 설정에 따라 성과 차이가 있었으나, 전반적으로 계절성 단순 기법 및 ETS 지수평활 모형보다 예측오차가 크게 나타났음.

<표 4-6> 마늘 학교·군 급식 소비량 예측 정확도 평가 결과(N=12)

학습창	구분	추정 구조	MAE	RMSE	MAPE
계절성 단순 기법		Seasonal Naive(period=12)	21.5	26.2	5.59
최근 60개월 자료 이용 (Rolling_60)	ETS 지수평활 모형	ETS(M,N,A)	23.9	31.9	6.22
	계절 ARIMA 모형	SARIMA(0,0,0)(1,0,0)[12]	26.9	35.6	7.83
	계절 ARIMAX 모형	SARIMAX(0,0,0)(1,0,0)[12]	30.0	39.8	8.70
최근 48개월 자료 이용 (Rolling_48)	ETS 지수평활 모형	ETS(M,N,A)	23.1	26.6	5.97
	계절 ARIMA 모형	SARIMA(0,0,0)(1,0,0)[12]	24.5	30.1	6.94
	계절 ARIMAX 모형	SARIMAX(0,0,0)(1,0,0)[12]	27.1	36.2	7.82
최근 36개월 자료 이용 (Rolling_36)	ETS 지수평활 모형	ETS(M,N,A)	26.2	31.6	6.79
	계절 ARIMA 모형	SARIMA(0,0,0)(1,0,0)[12]	22.7	28.0	6.40
	계절 ARIMAX 모형	SARIMAX(0,0,0)(1,0,0)[12]	27.8	37.2	7.94

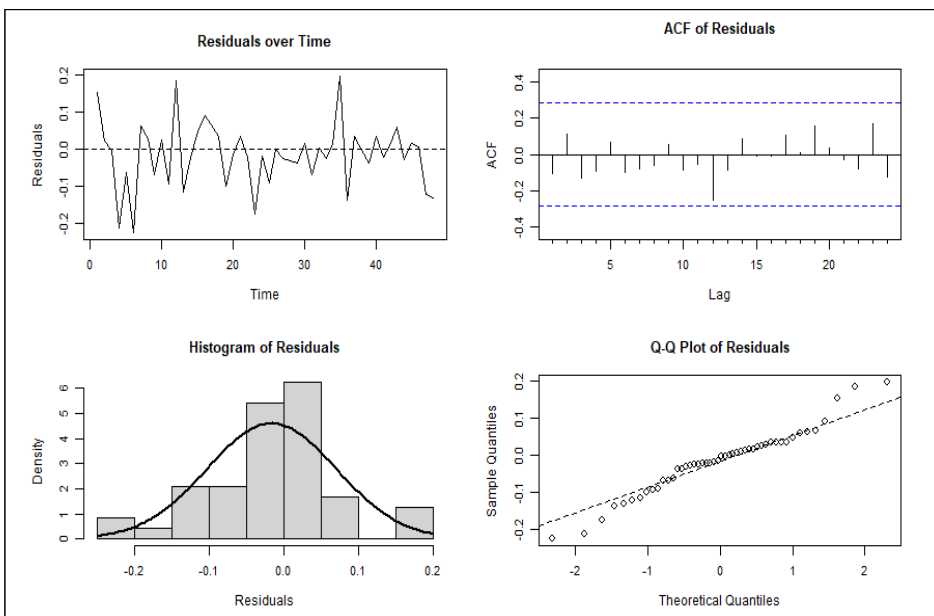
주: 계절성 단순 기법은 전년 동월 값을 예측값으로 사용하므로 학습창 길이와 관계없이 동일한 결과가 산출됨.
자료: 저자 작성.

○ 예측 정확도 평가에서는 계절성 단순 기법이 가장 낮은 예측오차를 보였음.
 다만 계절성 단순 기법은 별도의 모수 추정 없이 전년 동월 값을 예측값으로 사용하는 모형임. 이에 모형 적합성 검토를 위한 잔차 진단은 추정 모형 중 예측 성과가 가장 양호한 최근 48개월 이동 창 기반 ETS 지수평활 모형을 대상으로 수행하였음.

○ <그림 4-3>은 마늘 학교·군 급식 소비량 ETS 지수평활 모형의 잔차 시계열, 잔차 자기상관함수(ACF), 잔차 분포 및 Q-Q plot을 제시한 것임.

○ 잔차는 대체로 0을 중심으로 분포하였으며, 특정 방향으로 지속적인 편이가 나타나지 않았음. 다만, 일부 시점에서 상대적으로 큰 잔차가 관찰되어 특정 월의 소비량 변동이 모형에 충분히 반영되지 않았을 가능성이 있음.

<그림 4-3> 마늘 학교·군 급식 소비량 ETS 모형의 잔차 진단 결과



자료: 저자 작성.

- 잔차 자기상관함수(ACF)에서는 대부분의 자기상관 계수가 신뢰구간 내에 위치하여 잔차의 자기상관 측면에서는 뚜렷한 구조적 문제가 확인되지 않았음.
- 잔차 분포와 Q-Q plot에서는 대부분의 관측치가 기준선 주변에 위치하였으나, 일부 꼬리 구간에서 이탈이 관찰되었음. 이는 일부 월의 급식 소비량 변동이 잔차 분포에 영향을 미쳤을 가능성을 시사함.

5

모형 운영 방안

1. 부문별 모형 분석 결과 요약

- 본 연구는 양파·마늘을 대상으로 소비량 자료를 활용하여 소비 부문별 자료 특성에 적합한 예측모형을 개발하고자 하였으며, 양파·마늘 소비가 이루어지는 가구, 식품제조업, 외식업, 일반급식업, 학교 급식, 군 급식 등 소비 부문별 자료를 활용하여 구축된 모형을 비교하였음.
- 본 연구에서 활용한 자료는 가구 패널 조사자료, 산업 부문의 사업체 패널자료, 학교·군 급식의 행정자료 등 부문별로 자료 생성 방식과 축적 구조가 상이하다는 특징을 가짐. 이에 동일한 모형을 일괄적으로 적용하기보다 자료 구조와 소비 특성에 적합한 부문별 예측 전략을 적용하였음.
- 또한, 본 연구에서 활용한 소비 자료는 대부분 2023년 이후 본격적으로 축적되기 시작한 자료로 시계열 길이가 비교적 짧고, 조사사업 운영 일정에 따라

매년 일부 기간의 자료가 반복적으로 누락되는 구조적 결측을 포함하고 있음. 따라서, 본 연구는 제한된 자료 조건과 구조적 결측 문제를 고려하여 소비 부문별로 차별화된 분석 전략을 적용하였음.

1.1. 가구 부문

- 본 연구의 2장에서는 가구 소비량 자료를 대상으로 구조적 결측이 포함된 단기 시계열 자료에 적합한 예측 전략과 예측 성과를 검토하였음.
- 가구 소비 자료는 반복 조사 기반 자료라는 특성을 고려하여 기존 월별 집계 자료를 가능한 범위에서 주별 자료로 전환하여 활용하였으며, 이를 통해 제한된 시계열 내에서 보다 많은 관측치를 확보하고자 하였음. 또한, 매년 반복적으로 발생하는 구조적 결측 문제를 고려하여 결측을 직접 허용하는 방식과 선형보간 후 예측모형을 적용하는 방식을 비교하였음.
- 가구의 양파·마늘 소비량의 경우 2023년부터 2024년까지의 주별 자료를 학습 구간($N=88$)으로 설정하고, 2025년 자료를 테스트 구간으로 설정하였으며, 예측 방식은 rolling 1-step-ahead forecast 방식으로 설정하였음.
- 예측모형으로는 ETS, ARIMA, ARIMAX 모형 등을 적용하였으며, ARIMAX 모형에는 반입량의 1~3주 시차 변수를 외생변수로 포함하였음.
- 분석 결과, 양파와 마늘 두 품목 모두에서 최근 52주 자료를 이용하고 반입량의 1~3주 시차 변수를 포함한 선형보간 ARIMAX 모형이 비교 대상 모형 중 가장 낮은 예측오차를 보였음.

- 이는 주별 소비량 예측에서 소비량 자체의 과거 흐름뿐만 아니라, 공급 측면의 물량 정보가 일정한 예측 정보를 제공할 가능성을 시사함. 특히, 전체 과거 자료를 누적으로 활용하는 방식보다 최근 일정 기간의 자료를 활용한 모형에서 예측오차가 낮게 나타난 점은, 단기 소비량 예측에서 최근의 소비 패턴과 수급 여건을 반영하는 것이 중요할 수 있음을 보여줌.
- 다만, 본 분석은 2023~2025년의 짧은 시계열 자료와 매년 1~4월 조사 미실시에 따른 구조적 결측을 포함하고 있어, 모형의 안정성을 충분히 검증하기에는 한계가 있음. 또한, 일부 시점에서는 실제 소비량의 급변이 모형에 충분히 반영되지 않아 예측오차가 확대되는 양상이 확인되었음.
- 따라서 향후 단순히 모형 간 예측오차를 비교하는 수준을 넘어, 예측오차가 크게 나타난 시점의 원인을 구체적으로 검토할 필요가 있음. 예를 들어, 명절, 계절 요인, 가격 변동, 저장·비축 행태, 학교 급식 및 외식 수요, 조사 응답 변동성 등 소비량 변동에 영향을 미칠 수 있는 요인을 추가 설명변수로 반영할 필요가 있음.

1.2. 식품제조업·외식업·일반급식업

- 3장에서는 식품제조업, 외식업, 일반급식업 등 산업 부문의 소비 자료를 대상으로 패널 기반 예측 전략을 적용하였음.
- 산업 부문 자료는 사업체 패널자료 형태로 구축되어 있으나 월별 집계 자료 중심으로 구축되어 있다는 특성상 가구 부문과 같은 주별 시계열 확장이 어려운 한계가 존재하였음. 이에 본 연구는 단일 시계열 접근보다 사업체별 반복 관측정보를 활용한 패널 기반 접근을 중심으로 예측모형을 구축하였음.

- 모형 추정에 앞서 분석 자료를 살펴본 결과 업체-월 단위의 불균형 패널 구조를 가지며, 소비량 결측, 업체별 구매 규모 차이, 일부 대규모 구매 관측치의 영향을 받는 비대칭적 분포 특성이 함께 확인되었음. 이에 본 연구에서는 업체별 과거 평균 모형, 고정효과 패널회귀모형, CatBoost 모형을 적용하고, 원자료 기준 분석과 로그 변환 기준 분석을 병행하여 품목·산업별 예측 성과를 비교하였음.
- 분석 결과, 일반급식업과 외식업은 양파·마늘 모두 원자료 기준 CatBoost 모형의 예측 성과가 가장 우수하였으며, 식품제조업에서는 양파는 원자료 기준 업체별 과거 평균 모형, 마늘은 원자료 기준 CatBoost 모형의 성과가 가장 우수한 것으로 나타났다.
- 모형별로 보면, 업체별 과거 평균 모형이 구조가 단순함에도 불구하고 업체별 구매 수준의 지속성을 반영한다는 점에서 유용한 모형으로 기능하였음. 고정효과 모형은 업체별 고유한 구매 수준 차이를 반영할 수 있다는 점에서 의의가 있으나, 원자료 기준 분석을 상회하는 예측 성과는 확인되지 않았음. CatBoost 모형은 변수 간 비선형 관계와 상호작용을 반영할 수 있는 방법으로써 전반적으로 낮은 MAE를 보였음.
- 다만, 예측 성과를 해석할 때는 자료의 제약을 함께 고려할 필요가 있음. 본 자료는 분석 기간이 짧고, 업체별 관측 이력이 불균형하며, 구매량 결측과 업체 간 구매 규모 차이가 존재함. 또한, 본 연구의 예측에 활용한 설명변수 구성에 있어서도 개별 산업의 수요 변화를 충분히 반영하는 데에는 제약이 있었음. 따라서 본 연구의 분석 결과는 특정 모형의 우수성을 확정하기보다는, 현재 이용 가능한 산업 패널자료에서 적용 가능한 예측 전략을 비교한 결과로 해석하는 것이 적절함.

1.3. 학교·군 급식업

- 4장에서는 학교·군 급식 소비 자료를 대상으로 시계열 모형을 적용하였음. 학교·군 급식 자료는 조사 기반 자료가 아닌 행정자료 기반으로 구축되어 상대적으로 긴 월별 시계열 확보가 가능하였으며, 가구 및 산업 부문과 달리 구조적 결측이 존재하지 않는다는 특징을 가짐.
- 학교·군 급식 소비량 분석에서는 결측 처리 방식의 비교보다는 월별 시계열 자료에 대한 모형 간 성과를 비교하는 방식으로 예측 전략을 설정하였음. 가구 소비량 자료와 달리 결측 처리보다는 월별 반복성과 단기 변동을 어떻게 반영할 것인가를 예측모형 구축의 핵심 과제로 설정함.
- 학교·군 급식 소비량 자료는 2020년 1월부터 2025년 12월까지 총 72개 월별 관측치로 구성되는데, 본 연구에서는 2020년 1월~2024년 12월 자료를 학습 구간으로 설정하고, 2025년 1~12월 자료를 테스트 구간으로 설정함.
- 모형은 계절성 단순 기법(Seasonal Naive), ETS 지수평활, SARIMA, SARIMAX 등을 적용하여 예측 성과를 비교하였음.
- 분석 결과, 양파와 마늘 모두 복잡한 자기상관 구조를 반영한 ARIMA 계열 모형보다 월별 반복 패턴을 직접적으로 반영하는 모형의 예측 성과가 상대적으로 양호하였음. 이는 학교·군 급식 소비량이 장기 추세보다는 방학, 개학, 급식 운영 일정 등 월별 수요 구조의 영향을 크게 받을 가능성을 시사함.
- 품목별로 양파는 최근 소비량 수준과 계절 변동을 함께 반영하는 ETS 지수평활 모형의 성과가 가장 우수하였고, 마늘은 전년 동월 소비량을 활용하는 계

절성 단순 기법의 성과가 가장 양호하였음. 이는 급식 소비량 예측에서 품목별 시계열 특성을 고려한 모형 선택이 필요함을 보여줌.

○ 반입량 시차 변수를 포함한 SARIMAX 모형은 양파에서는 일정 수준의 예측 성과를 보였으나, 마늘에서는 Seasonal Naive 기법 및 ETS 지수평활 모형보다 상대적으로 높은 예측오차를 보였음. 따라서 공급 측 물량 정보의 활용 가능성은 품목별로 다르게 평가될 필요가 있으며, 향후에는 반입량 외에도 급식 일수, 방학 및 개학 일정, 공휴일, 식단 구성 등 급식 수요와 직접적으로 관련된 변수를 함께 검토할 필요가 있음.

○ 이상의 결과를 종합하면, 급식 소비량 예측은 단일 모형을 일괄 적용하기보다 품목별 반복 패턴, 최근 수준 변화, 외생변수의 설명 가능성을 함께 비교하는 방식으로 접근하는 것이 바람직함. 현재 단계에서는 계절성 단순 기법과 ETS 지수평활 모형이 실무적으로 활용 가능한 기준 모형 역할을 할 수 있으며, 향후 자료가 더 축적되는 경우 외생변수를 포함한 확장 모형의 성과를 추가로 검토할 필요가 있음.

〈표 5-1〉 부문별 소비 예측모형 주요 추정 결과 요약

구분	자료 구조	학습 설정	테스트 설정	적용 모형	최적 모형	예측 성과 특징
가구	주별 시계열	■ 학습 구간: 2023년 4월 ~ 2024년 12월(N=88) ■ fixed-length rolling window (최근 52주, 72주), expanding window	■ 테스트 구간: 2025년 (N=37) ■ rolling 1-step-ahead forecast	ETS, ARIMA, ARIMAX	■ 선형모간 ARIMAX ■ 양파: ARIMAX(2,0,1) ■ 마늘: ARIMAX(4,0,0)	■ 외생변수 포함(반입량) 시 예측력 개선 ■ 과거 자료 누락 방식보다 최근 자료 이동 방식이 우수
식품제조업	사업체-월 (불균형)패널	■ 학습 구간: 2023년 4월 ~ 2024년 12월	■ 테스트 구간: 2025년 4월 ~ 12월 ■ expanding window 기반의 1개월 앞 예측 방식	업체별 과거 평균 모형, 고정효과모형, CatBoost	■ 양파: 원자로 과거 평균 모형 ■ 마늘: 원자로 CatBoost	■ 산업별 구매 구조와 자료 분포 특성이 예측모형의 성과에 영향
					■ 양파·마늘: 원자로 CatBoost	
					■ 양파·마늘: 원자로 CatBoost	
학교·군 급식	월별 시계열	■ 학습 구간: 2020년 1월 ~ 2024년 12월(N=60) ■ fixed-length rolling window(최근 60개월, 48개월, 36개월)	■ 테스트 구간: 2025년 1월 ~ 12월(N=12) ■ rolling 1-step-ahead forecast	Seasonal Naive, ETS 지수평활, SARIMA, SARIMAX	■ 양파: ETS(M, N, M) ■ 마늘: Seasonal Naive	■ 월별 반복 패턴을 직접 반영하는 모형 성과 우수 ■ 품목별 예측 성과 우수 모형 차이 ■ 공급 측 물량 정보의 활용 가능성은 품목별로 다르게 평가

자료: 저자 작성.

2. 모형 활용 및 운영 방안

2.1. 모형 활용 방안

2.1.1. 소비 실태 파악을 위한 활용

- 본 연구에서 검토한 모형은 양파·마늘 소비량의 단기적인 변화 방향을 파악하는 데 활용될 수 있을 것임. 특히, 가구, 식품제조업, 외식업, 급식 등 주요 소비 부문별 소비 변화를 예측함으로써 최근 소비 흐름이 확대되는지 또는 둔화되는지를 사전에 판단하는 참고자료로 활용될 수 있음.
- 관측사업 운영 과정에서는 실제 조사자료가 집계되기 이전 단계에서 모형 기반 소비량 추정치(잠정치)를 활용하여 단기 소비 흐름에 대한 시의성 있는 관측정보 제공 가능성을 검토할 수 있음.
- 또한, 소비 자료는 생산량이나 출하량과 달리 최종 소비 단계에서 발생하는 정보를 반영한다는 특징이 있음. 따라서 예측모형은 공급 측면 자료만으로는 파악하기 어려운 소비 변화 양상을 제공할 수 있으며, 향후 소비 흐름에 대한 이해를 높이는 데 기여할 수 있을 것임.

2.1.2. 농업관측과 연계한 활용

- 본 연구에서 구축한 예측모형의 결과는 기존 농업관측정보를 보완하는 참고 지표로 활용될 수 있을 것임.
- 농업관측은 생산, 재고, 출하, 가격 등 공급 측면 정보를 중심으로 운영되고

있으며, 이를 바탕으로 수급 상황을 판단하고 있음. 하지만, 실제로는 공급 여건뿐 아니라 소비 변화 역시 가격 및 수급 상황에 영향을 미칠 수 있음.

- 소비 모형은 공급 중심 정보만으로 설명하기 어려운 시장 상황을 이해하는 데 보조적인 역할을 수행할 수 있음. 소비 예측 결과를 기존 농업관측정보와 함께 활용하는 경우 수급 상황에 대한 보다 입체적인 해석과 정교한 수급 대응이 가능할 것으로 기대됨.
- 또한, 기존 중기선행관측모형 등 공급 중심 수급 전망 결과와 소비 예측 결과를 함께 검토하는 경우 공급 단계에서 예상되는 수요와 실제 최종 소비 흐름 간의 차이를 비교·분석할 수 있음. 이를 통해 소비 변화가 시장에 미치는 영향을 보다 종합적으로 검토할 수 있을 것으로 여겨짐.
- 본 연구에서는 도매시장 반입량이나 도매가격의 시차 변수를 사용하였는데, 실제 소비 관측 운영 시점에서 해당 변수를 안정적으로 확보하기 위해서는 기존 농업관측 중기 예측과의 연계도 이루어질 수 있을 것임.

2.1.3. 조사자료 공백 보완을 위한 활용

- ‘농식품 소비정보 분석사업’ 자료는 조사 운영 일정에 따라 일부 기간의 자료가 제공되지 않는 구조적 특성이 있음. 이로 인해 현재 소비 자료는 매년 일정 기간의 조사 공백이 발생하며, 소비 흐름을 연속적으로 파악하는 데 한계가 있음.
- 본 연구에서 검토한 모형은 이러한 조사 공백 기간 동안의 소비 흐름을 추정하는 보조적 수단으로 활용될 수 있음. 예측 결과는 실제 조사자료를 대체하

기보다 조사 공백 기간의 소비 변화 양상을 참고적으로 파악하기 위한 정보로 활용함으로써 소비 정보의 연속성을 보완하는 데 기여할 수 있을 것으로 기대됨.

- 향후 소비 자료가 지속적으로 축적되고 모형의 예측 성능이 검증된다면, 조사 공백 기간에 대한 보완적 추정정보를 제공하여 소비 관측 체계의 연속성과 활용성을 높이는 데 활용할 수 있을 것으로 기대됨.

2.1.4. 소비 관측 품목 확대를 위한 운영 기반 마련

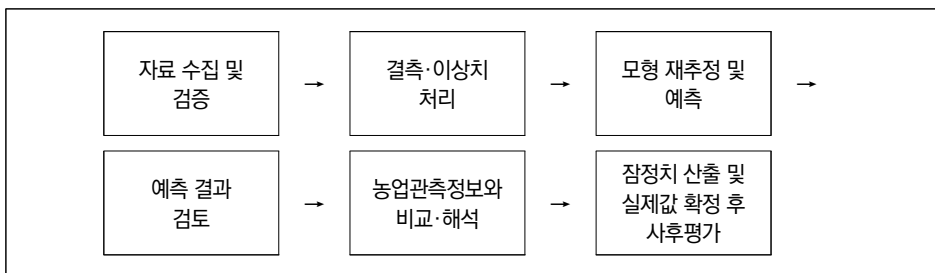
- 본 연구는 양파·마늘을 대상으로 소비량 자료를 활용한 예측모형의 적용 가능성을 검토했으나, 의의가 특정 품목의 예측 성과에만 있는 것은 아닐 것임.
- 본 연구의 자료 분석 절차와 모형 적용 경험은 향후 다른 농축산물 품목으로 소비 관측을 확대하는 과정에서 참고자료로 활용될 수 있음. 특히, 소비 부문별 자료 구조를 고려하여 적합한 모형을 적용한 접근 방식은 품목이 달라지더라도 활용 가능성이 있으며, 향후 소비 데이터가 축적되는 품목에 대해서도 단계적으로 적용을 검토할 수 있을 것으로 기대됨.
- 품목별 소비 자료의 축적 수준 또한 상이하기 때문에 현 단계에서 모든 품목에 대해 동일한 방법을 적용하기보다는 자료 축적 수준과 품목 특성을 고려하여 단계적으로 적용 범위를 확대할 필요가 있음. 이러한 측면에서 본 연구는 양파·마늘을 대상으로 소비 예측모형의 적용 가능성을 검토한 시범 연구로서의 의미가 있음.

2.2. 모형 운영 방안

2.2.1. 모형의 운영 절차

- 소비 모형은 소비 데이터를 생산·관리하는 기관을 중심으로 운영하되, 자료의 갱신과 모형의 재추정, 예측 결과의 검토 및 관리가 일관되게 이루어질 수 있는 체계가 요구됨.
- 본 연구를 통해 개발된 모형이 실제 관측에 활용되기 위해서는 자료의 수집부터 예측 결과의 검토와 사후평가에 이르는 정기적인 운영 절차가 마련될 필요가 있음.
- 모형 운영은 크게 ① 자료의 수집 및 품질 점검, ② 결측치와 이상치 처리, ③ 부문별 모형 재추정 및 예측, ④ 예측 결과의 적정성 검토, ⑤ 농업관측정보와 연계한 해석, ⑥ 실제 소비량 확정 이후 예측 성과 평가의 단계로 구성할 수 있을 것임.

〈그림 5-1〉 모형의 운영 절차



자료: 저자 작성.

- 우선 가구, 산업, 학교·군 급식 부문별 양파·마늘 소비량 자료와 반입량 등 외 생변수를 정기적으로 갱신하고, 자료의 누락 여부와 비교 시점 대비 급격한

변동을 점검할 필요가 있음. 모형에서 활용되는 공급 측 정보는 농업관측에서 수집·생산되는 최신 정보와 연계하여 점검이 가능함.

○ 자료 처리 시에는, 조사자료에 기반한 가구 및 산업 부문은 조사의 공백, 패널 이탈, 무응답 및 이상치가 예측 결과에 영향을 줄 수 있으므로 자료 처리 기준을 일관되게 적용할 필요가 있음.

○ 자료에 대한 점검이 완료된 이후에는 본 연구를 통해 예측성고가 높은 부문별 모형을 적용하여 단기 소비량을 예측함.

- 가구 부문은 최근 52주 자료를 이용한 이동 학습 구간 방식으로 모형을 재추정하고, 식품제조업·외식업·일반급식업은 새롭게 확보된 업체별 자료를 기존 패널자료에 추가하여 예측을 수행할 수 있음.

- 학교·군 급식 부문은 월별 자료가 입수되는 시점에 맞추어 시계열을 갱신하는 방식으로 운영할 수 있음.

- 도매시장 반입량이나 가격을 외생변수로 활용하는 경우 농업관측정보가 제공하는 향후 출하량이나 가격 전망치를 활용할 수도 있을 것임.

○ 모형에서 산출된 예측 결과는 농업관측 담당자와 함께 생산·출하·재고·가격 등 관련 관측정보를 종합적으로 검토하여 결과의 적정성을 확인할 필요가 있으며, 모형에서 산출된 결과를 바로 확정적인 소비 전망치로 활용하기보다 전월 및 전년 동기와 비교하여 소비 변화의 방향과 정도를 판단함.

○ 만약 급격한 소비량 변화가 예측되는 경우는 명절, 방학, 외식 수요, 가공업체의 일시적 대량 구매, 가격 및 반입량 변화 등 관련 요인을 함께 검토하여 결과의 적정성을 판단할 필요가 있으며, 이때 양파·마늘 소비 및 유통 전문가를 통

신원 혹은 자문위원으로 위촉하여 시장 정보에 대한 점검을 수행할 수 있음
(이형용 외, 2025b: 115).

2.2.2. 모형의 갱신 및 성능 관리

- 양파·마늘 소비 구조와 자료의 특성은 시간에 따라 변화할 수 있으므로 본 연구를 통해 최초로 선정된 모형을 고정적으로 적용하기보다는 신규 자료가 축적됨에 따라 모형을 반복적으로 재추정하고 예측 성과를 점검하는 체계가 필요할 것임.
- 모형의 일상적인 갱신은 부문별 자료 입수 주기에 맞추어 수행하되, 모형 간 성능 비교와 모형 교체 여부에 대한 검토는 일정 기간의 자료가 추가로 축적된 이후 주기적으로 수행될 필요 있음.
- 모형의 성능 평가는 본 연구에서 활용한 MAE, RMSE, MAPE 등을 중심으로 실시할 수 있으며, 최근 기간에 대한 예측오차가 지속 확대되는 경우 변수 구성, 학습 기간, 결측 처리 방식 및 후보 모형을 재검토할 필요가 있음.
- 특히, 본 연구에서 가장 우수한 예측 성과를 보인 모형이 향후에도 항상 최적의 성능을 유지한다고 보기 어려우므로, ARIMA나 과거 평균 모형 등 구조가 단순한 기준 모형과 예측 성과를 지속적으로 비교할 필요도 있음.
- 상대적으로 복잡한 모형의 예측오차가 지속해서 기준 모형보다 크게 나타나는 경우는 모형을 재설정하거나 기준 모형을 일시적으로 활용·참고하는 방식도 고려할 수 있음.

- 이와 함께 모형의 추정 시점, 활용 자료, 결측치 처리 방식, 주요 모수, 예측 결과 및 사후 예측오차를 정기적으로 기록하여 모형 운영의 일관성과 재현성을 확보할 필요가 있음.

2.2.3. 모형 결과의 활용

- 모형 예측의 결과는 부문별 소비 변화를 우선으로 파악하되, 필요에 따라 부문별 전망치를 종합한 총 소비 흐름을 함께 검토하는 방식으로 활용할 수 있음. 소비 총량은 양파·마늘이 소비되는 각 부문의 소비를 각각 추정한 후 이를 합산하여 산출함($\hat{Y}_t = \sum_i \hat{y}_{it}$).
- 부문별 소비량은 소비 주체별 변화 양상을 파악하는 데 유용하며, 총 소비 흐름은 품목의 전반적인 소비 동향을 종합적으로 이해하기 위한 참고자료로 활용할 수 있을 것임.
- 다만, 앞서 분석 과정에서 밝힌 바와 같이 본 연구의 활용 데이터는 축적 기간에 한계가 있으므로 현 단계에서는 모형 결과를 단일 수준 전망치로 제시하기 보다 소비의 변화 방향을 보여주는 지표로 제공하는 것이 적절할 것임. 예를 들어, 부문별 양파·마늘 소비량의 전월 대비 및 전년 동기 대비 증감률과 함께 소비 흐름을 ‘증가’, ‘보합’, ‘감소’ 등으로 구분하여 제시할 수 있을 것임. 이때 모형에 의해 산출된 값은 실제 조사치와 구분하여 ‘잠정 추정치’ 또는 ‘모형 전망치’임을 명시할 필요가 있음.
- 소비 모형의 결과는 기존 농업관측정보를 보완·연계하는 방식 등으로 활용하는 방법도 고려해볼 필요 있으며, 향후 소비 관측 체계가 구축되는 경우 별도의 관측보를 통해 활용하는 방안도 검토할 수 있음.

- 본 연구의 소비 모형은 시장가격을 직접 전망하거나 수급 균형을 구조적으로 추정하기 위한 모형이 아니므로, 예측 결과를 가격 전망이나 정책 판단의 근거로 직접 활용하는 데에는 주의가 필요함. 사업의 초기 단계에서는 모형의 결과를 기존 농업관측정보를 보완하는 소비 측면의 정보로 활용하고, 생산·출하·재고·가격 등 다른 관측정보와 함께 종합적으로 해석하는 것이 바람직함.
- 다만, 향후 소비 자료의 지속적인 축적과 모형의 검증을 통해 활용 범위를 점진적으로 확대해 나갈 필요가 있음.

3. 향후 과제

3.1. 소비 데이터의 지속적인 축적 및 관리

- 본 연구는 ‘농식품 소비정보 분석사업’을 통해 구축된 실제 소비량 자료를 양파·마늘 소비 모형에의 활용 가능성을 검토하였다는 점에서 의의가 있음.
- 그러나, 본 연구에서 활용한 자료는 대부분 2023년 이후 본격적으로 구축되기 시작했으므로 활용 가능한 시계열 길이가 상대적으로 짧다는 제약이 있음. 예측모형의 경우 장기간의 자료를 활용할수록 계절성, 추세, 구조 변화 등을 보다 안정적으로 식별할 수 있으나, 현재의 소비 자료는 이러한 특성을 충분히 반영하기에는 한계가 있는 것으로 판단됨.
- 본 연구는 주별 자료 전환 등을 통해 관측치 수를 확대하여 모형 구축에 활용

하긴 했으나, 이는 자료의 관측 빈도를 높인 것이며 장기 시계열의 축적을 대체하는 것은 아님.

- 또한, 소비 자료의 경우 가구, 식품제조업, 외식업, 일반급식업, 학교 및 군 급식 등 소비 부문별로 자료의 수집 방식과 축적 구조가 상이함. 따라서 동일한 기준으로 자료를 통합하거나 동일한 예측 방식을 일률적으로 적용하는 데에는 한계가 존재함.
- 따라서 ‘농식품 소비정보 분석사업’의 지속적인 운영을 통해 장기 시계열 자료를 축적하고, 소비 부문별 자료의 안정성과 대표성을 지속 확보할 필요가 있음.
- 소비량 예측의 정확도를 높이기 위해서는 자료 구축 및 관리 체계를 지속적으로 개선할 필요가 있음. 특히, 산업 부문은 일부 업체의 미구매와 조사 특성에 따른 결측 발생 가능성이 있으므로 결측 구조를 체계적으로 관리하고 개체별 소비량을 지속 축적할 필요가 있음. 또한, 소비 부문별 수요 특성을 반영할 수 있는 설명변수를 꾸준히 확보하고, 이상치 관리와 자료 정합성 점검 등을 통해 소비 자료의 품질을 지속적으로 개선할 필요가 있음.

3.2. 결측치 처리 방법의 개선

- 본 연구에서 활용한 소비량 자료는 조사사업 운영 일정에 따라 매년 일정 기간 자료가 반복적으로 누락되는 구조적 결측을 포함하고 있음. 특히, 가구 부문과 산업 부문의 표본조사 자료는 매년 같은 시기에 결측이 발생하고 있어 양파·마늘 소비의 계절성 및 단기 변동 구조를 안정적으로 추정하는 데 제약이 있을 수 있음.

- 본 연구에서는 이러한 자료의 제약을 고려하여 선형보간을 통해 일부 결측 구간을 보완하고 모형을 추정하였음. 단, 선형보간은 구현이 간단하고 연속적인 시계열을 구성할 수 있다는 장점이 있으나, 소비량 변동성이 크거나 결측이 연속되는 경우 실제 급등락과 계절적 변화를 평활화할 가능성이 있음. 따라서 본 연구에서는 선형보간을 결측값의 정확한 복원 방법이 아니라 연속적인 시계열이 요구되는 모형에 적용하기 위한 보조적 처리 방식으로 활용하였음.
- 특히, 가구 소비량 자료는 전체 156주 중 49주(31.4%)가 결측이며, 결측이 매년 1~4월에 반복적으로 발생함. 결측 비율이 높아질수록 미적용 자료에서는 모형 추정에 활용되는 유효 관측치가 감소하는 반면, 보간 자료에서는 보간값의 비중이 증가하여 실제 변동이 평활화될 가능성이 커짐. 이에 따라 보간 적용 여부에 따른 예측성과는 현재의 결측 비율과 발생 패턴에 영향을 받은 조건부 결과로 해석할 필요가 있음.
- 향후 자료가 추가로 축적되면 Cubic spline, 다중대체법, 계절성 분해 기반 보간(Seasonal decomposition), 상태공간모형 및 Kalman smoothing 등 다양한 결측치 처리 방법을 비교하여 소비 자료의 특성에 적합한 방법을 검토할 필요가 있음. 이는 소비 관측 모형의 신뢰성과 안정성을 높이는 데 기여할 수 있을 것임.

3.3. 모형의 검증 및 고도화

- 본 연구는 소비 부문별 자료 특성을 고려하여 다양한 예측모형을 적용하고 예측 성능을 비교하였음. 그 결과 소비 부문별 상대적으로 우수한 성능을 보이는 모형이 상이하며, 단일 모형을 일괄적으로 적용하는 방식보다는 부문별 특

성에 적합한 예측 전략을 적용하는 것이 효과적인 것으로 나타남.

- 다만, 본 연구 결과를 특정 모형의 우수성을 최종적으로 입증한 결과로 해석하기에는 무리가 있음. 활용 가능한 자료의 기간이 제한적이고 구조적 결측이 존재하는 만큼, 예측 성능 역시 자료의 특성에 영향을 받을 가능성이 존재하기 때문임. 따라서 본 연구는 소비 관측을 위한 최종모형이나 표준모형을 제시한다기보다, 소비량 자료를 활용한 모형의 적용 가능성을 검토한 탐색적 연구로 이해하는 것이 적절할 것임.
- 앞서 운영 방안에서 제시하였듯이 추정된 모형 결과에 대해서는 향후 자료가 추가로 축적됨에 따라 동일한 모형에 대한 반복적인 검증이 요구될 것임. 또한, 이상치 처리 및 민감도 분석을 통해 모형의 안정성을 지속 검토할 필요가 있음.
- 활용 변수 측면에서는 소비 부문별 특성을 반영할 수 있는 외생변수의 확대 역시 검토할 필요가 있음. 현재는 도매시장 반입량이나 도매가격을 중심으로 분석하였으나, 실제 양파·마늘 소비는 가격과 공급 상황 외에도 기상 여건, 외식 경기, 소득 및 물가 수준, 명절 및 계절 요인, 소비자 할인 행사 등 복합적인 영향을 받을 가능성이 있으므로 앞으로는 다양한 외생 요인을 활용함으로써 예측력을 높일 수 있을 것으로 기대됨.
- 아울러 최근 활용이 확대되고 있는 머신러닝 기법에 대한 검토도 지속될 필요가 있음. 본 연구에서는 CatBoost를 중심으로 적용 가능성을 검토하였으나, 향후 다양한 머신러닝 및 하이브리드 모형을 활용하여 예측 성능을 비교하고 소비 데이터에 적합한 활용 방안을 검토할 수 있을 것임.

- 학교 급식과 군 급식 부문의 경우, 두 소비 모두 집단급식이라는 공통적인 특성이 있으나, 운영 방식과 계절성 측면에서 차이가 존재할 수 있음. 따라서, 향후 학교 및 군 급식에 대한 고도화가 요구되는 경우 개별 부문에 대한 별도 분석도 검토될 필요는 있음.

3.4. 소비 관측의 발전 방향

- 본 연구는 완성된 소비 관측 체계를 제시하기보다, 실제 소비량 자료를 활용한 소비 관측의 가능성을 검토하기 위한 기초연구로 수행되었음. 따라서 본 연구의 결과는 소비 관측의 종착점이라기보다 향후 소비 관측 체계 구축을 위한 출발점으로 이해할 필요가 있음.
- 기존 농업관측정보는 오랜 기간 생산, 재고, 가격 중심의 자료 축적과 모형 개선을 통해 발전해 왔음. 소비 관측 역시 단기간에 완성되기보다는 지속적인 자료 축적과 반복적인 모형 검증을 통해 점진적으로 발전해 나갈 필요가 있음. 특히, ‘농식품 소비정보 분석사업’을 통해 구축되는 자료가 장기간 축적되면 소비 변화의 계절성, 장기 추세 및 소비행태 변화 등을 보다 안정적으로 분석할 수 있을 것으로 기대됨.
- 향후 소비 관측은 단순히 모형의 정확도를 높이는 것에 그치지 않고, 소비 관련 정보가 농업관측 체계 내에서 어떠한 역할을 수행할 수 있는지에 대한 검토와 함께 발전할 필요가 있음. 이를 위해서는 소비 자료의 축적, 자료 품질 개선, 결측치 처리 기법의 고도화, 개발된 모형의 지속적 검증이 병행되어야 하며, 이를 통해 소비 정보가 농업관측을 보완하는 유용한 정보로 활용될 수 있을 것으로 기대됨.

○ 또한, 장기적으로는 소비량 자료뿐만 아니라 농업관측센터에서 활용하고 있는 생산량, 출하량, 도매시장 반입량 및 가격 등 공급 측면의 자료를 함께 활용하여 공급과 수요를 동시에 고려하는 모형으로 확장할 필요도 있음. 이를 통해 소비 관측 결과의 활용성을 높이고 농업관측의 정확성과 종합성을 제고하는 데 기여할 수 있을 것으로 기대됨.

1. Naive and Seasonal Naive Method

- Naive 방법과 Seasonal Naive 방법은 시계열 예측에서 가장 기본적으로 활용되는 기준 모형(baseline model)임(Hyndman & Athanasopoulos, 2021).
- 두 방법 모두 복잡한 모수 추정이나 구조 설정 없이 과거 관측값을 이용하여 미래값을 예측한다는 공통점이 있음.
- Naive 방법은 가장 최근 관측값이 미래에도 유지된다고 가정하는 방식임. 즉, 다음 시점의 소비량은 직전 시점의 소비량과 동일하다고 가정함. Naive 방법은 시계열의 추세나 계절성을 별도로 고려하지 않으며, 가장 최근의 소비 수준이 단기적으로 지속된다는 단순한 가정에 기반함. 이러한 특성 때문에 Naive는 계산이 간단하고 자료 구조에 대한 강한 가정을 요구하지 않는다는 장점이 있음.
- Naive는 다음과 같이 표현할 수 있음(Hyndman & Athanasopoulos, 2021).
 - $\hat{y}_{t+1|t} = y_t, \hat{y}_{t+h|t} = y_t$
 - 여기서 y_t 는 t 시점의 소비량, $\hat{y}_{t+h|t}$ 는 t 시점 정보를 이용한 h 기 이후의 예측값을 의미함.
 - 즉, 미래 모든 시점의 소비량이 가장 최근 관측값과 동일하다고 가정하는 방식임.

○ 반면, Seasonal Naive 방법은 계절성이 존재하는 시계열을 대상으로 동일 계절의 과거 관측값이 반복된다고 가정하는 방식임.

○ 일반 Naive 방법이 직전 시점의 값을 사용하는 것과 달리, Snaive는 이전 동일 계절의 값을 미래 예측값으로 사용함. 예를 들어, 월별 자료에서는 전년 동월의 소비량을 미래 소비량 예측값으로 활용함. 따라서 계절성이 존재하는 시계열에서는 일반 Naive 방법보다 현실적인 기준 예측값을 제공할 가능성이 있음.

○ Seasonal Naive Method는 다음과 같이 표현할 수 있음.

- $\hat{y}_{t+h|t} = y_{t+h-m(k+1)}$
- 여기서 m 은 계절 주기(seasonal period)를 의미하며, 월별 자료의 경우 일반적으로 $m=12$, 주별 자료의 경우 $m=52$ 로 설정 가능함.
- k 는 $k = [(h-1)/m]$ 로 정의되며, 동일 계절 위치를 맞추기 위한 정수값임.

2. ARIMA

○ ARIMA(Auto-Regressive Integrated Moving Average) 모형은 시계열 자료의 자기상관 구조를 이용하여 미래 값을 예측하는 대표적인 시계열 분석 방법임(Box & Jenkins, 1970).

- ARIMA 모형은 과거 소비량의 변화 패턴과 오차 구조를 이용하여 현재 및 미래 소비량을 설명하는 방식으로 구성됨.
- 가장 널리 활용되는 전통적인 시계열 예측모형 중 하나이며, 비교적 짧은 시계열에서도 안정적인 예측 성능을 보이는 것으로 알려져 있음.

○ ARIMA 모형은 자기회귀(AR), 차분(I), 이동평균(MA)의 세 요소를 결합한 구조를 가짐.

- 자기회귀(auto-regressive) 과정은 현재 소비량이 과거 소비량에 의해 영향을 받는 구조를 의미함.
- 이동평균(moving average) 과정은 과거 예측오차가 현재 소비량에 영향을 미치는 구조를 반영함.
- 차분(differencing)은 평균이나 추세가 시간에 따라 변하는 비정상(non-stationary) 시계열을 정상(stationary) 시계열로 변환하기 위해 사용됨.

○ 농산물 소비량 시계열은 계절성, 추세, 가격 충격 등에 의해 평균과 분산이 일정하지 않을 가능성이 존재함.

- 양파·마늘 소비량은 가격 급등락, 계절적 소비 변화, 외식 수요 변화 등의 영향을 받을 수 있음.
- 이러한 시계열은 단순 평균 회귀 과정만으로 설명하기 어려우므로, 시계열의 자기상관 구조를 반영한 모형 접근이 필요함.
- ARIMA 모형은 과거 소비량 패턴 자체를 이용하여 미래 소비량을 추정하므로 소비량의 동태적 변화를 반영할 수 있다는 장점이 있음.

○ ARIMA 모형은 시계열의 자기상관 구조를 반영하기 위해 자기회귀(AR)와 이동평균(MA) 요소를 결합한 모형으로, 필요 시 차분을 통해 정상성을 확보함 (Box & Jenkins, 1970; Box et al., 2016).

○ 일반적인 $ARIMA(p, d, q)$ 모형은 다음과 같이 표현할 수 있음.

$$- \phi(B)(1-B)^d y_t = c + \theta(B)\varepsilon_t$$

- 여기서 $\phi(B)$ 는 자기회귀항(AR), $\theta(B)$ 는 이동평균항(MA)를 의미하며, B 는 후행연산자(backshift operator)로 $By_t = y_{t-1}$ 를 의미함.
- d 는 차분 차수(differencing order), ε_t 는 백색잡음(white noise) 오차항임.

○ 자기회귀항과 이동평균항은 각각 다음과 같이 표현할 수 있음.

- $\phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$
- $\theta(B) = 1 + \theta_1 B + \dots + \theta_q B^q$
- 여기서 p 는 자기회귀 차수, q 는 이동평균 차수를 의미함.
- 자기회귀항은 과거 소비량이 현재 소비량에 미치는 영향을 반영하며, 이동평균항은 과거 충격이나 예측오차의 영향을 설명함.

○ ARIMA 모형은 과거 소비량의 정보만을 활용하여 현재 소비량을 설명하는 순수 시계열 모형임.

3. ARIMAX

○ ARIMAX(Auto-Regressive Integrated Moving Average with Exogenous Variables) 모형은 ARIMA 모형에 외생변수(exogenous variables)를 추가한 확장형 시계열 모형임(Pankratz, 1991).

- 기본적인 ARIMA 모형이 과거 소비량의 자기상관 구조만을 이용하여 미래 값을 설명하는 반면, ARIMAX 모형은 가격이나 계절 변수 등 외부 요인이 소비량에 미치는 영향을 함께 반영할 수 있음.

- 즉, 시계열 내부의 동태적 구조와 외부 경제 변수의 영향을 동시에 고려할 수 있다는 점에서 순수 시계열 모형보다 확장된 형태의 접근으로 볼 수 있음.

○ ARIMAX 모형은 소비량의 시계열적 지속성과 외부 변수의 영향을 동시에 반영할 수 있다는 장점이 있음.

- 소비량은 일반적으로 이전 시점의 소비 수준에 영향을 받는 경향이 존재함.
- 동시에 가격 상승, 공급 부족, 계절 변화 등 외생적 충격이 발생하는 경우 소비량이 추가적으로 변동할 가능성이 있음.
- ARIMAX 모형은 이러한 두 가지 요소를 하나의 모형 내에서 함께 설명할 수 있다는 특징이 있음.

○ 일반적인 ARIMAX 모형은 다음과 같음.

- $\phi(B)(1-B)^d y_t = c + \beta' x_t + \theta(B) \varepsilon_t$
- 여기서 y_t 는 소비량, x_t 는 외생변수 벡터, β 는 외생변수의 계수 벡터이며, $\phi(B)$ 와 $\theta(B)$ 는 각각 자기회귀항과 이동평균항, ε_t 는 백색잡음 오차항임.
- 즉, 현재 소비량은 과거 소비량의 구조와 외부 변수의 영향을 동시에 받아 결정되는 구조로 이해할 수 있음.

○ ARIMAX 모형은 외생변수를 활용함으로써 ARIMA 모형 대비 설명력과 예측력을 개선할 가능성이 있음.

- 특히, 가격 변동성이 큰 농산물의 경우 소비량 변화가 단순 자기상관 구조만으로 설명되지 않을 가능성이 존재함.

- 따라서 외생변수를 함께 고려하는 ARIMAX 모형은 소비량 변동 원인을 보다 현실적으로 설명할 수 있다는 장점이 있음.

○ 다만, ARIMAX 모형은 외생변수의 품질과 예측 가능성에 따라 예측 성능이 달라질 수 있다는 한계도 존재함.

- 외생변수 자체의 변동성이 크거나 미래값을 적절히 확보하기 어려운 경우 예측오차가 확대될 가능성이 있음.
- 또한, 외생변수 간 다중공선성이나 구조 변화가 발생하는 경우 계수 안정성이 저하될 가능성도 존재함.

4. ETS

○ 본 연구는 소비량 시계열의 수준(level), 추세(trend), 계절성(seasonality)을 반영하기 위한 방법으로 ETS(Error-Trend-Seasonality) 모형을 활용하였음.

- 본 연구에서 ETS 모형은 소비량 시계열의 추세 및 계절성을 반영하기 위한 비교모형으로 활용하였음.

○ 양파·마늘 소비량은 월별 소비 패턴과 계절적 변동 가능성이 존재하므로, ETS 모형은 ARIMA와 함께 소비량 예측 성능을 비교할 수 있는 유용한 시계열 모형으로 판단됨.

- ETS 모형은 시계열의 구조적 구성요소를 직접 반영한다는 점에서, 자기상관 구조 중심의 ARIMA 모형과 상호보완적인 특성을 가짐.

- 또한, ETS는 비교적 짧은 시계열에서도 안정적인 예측 성능을 보이는 것으로 알려져 있어(Hyndman & Athanasopoulos, 2021), 본 연구와 같이 소비 자료 축적 기간이 제한적인 경우에도 활용 가능성이 존재함.

○ ETS 모형은 지수평활법(exponential smoothing)을 상태공간(state space) 형태로 일반화한 시계열 모형으로, 시계열의 구조적 구성요소를 이용하여 미래 값을 예측하는 방법임(Hyndman et al., 2002; Hyndman & Athanasopoulos, 2021).

○ 전통적인 ARIMA 모형은 과거 관측값과 오차의 자기상관 구조를 기반으로 시계열을 설명하는 반면, ETS 모형은 시계열을 수준(level), 추세(trend), 계절성(seasonality)의 구성요소로 분해하여 설명한다는 차이가 있음. 따라서 ARIMA 모형이 자기상관 구조 중심의 확률적 접근이라면, ETS 모형은 시계열의 구조적 패턴을 반영하는 구성요소 기반(component-based) 접근으로 볼 수 있음.

○ ETS 모형은 오차(error), 추세(trend), 계절성(seasonality)의 형태를 각각 가법(additive) 또는 승법(multiplicative) 구조로 설정할 수 있으며, 일반적으로 ETS(E, T, S) 형태로 표현됨.

- 예를 들어, ETS(A, A, N)는 가법 오차(additive error), 가법 추세(additive trend), 비계절(non-seasonal) 구조를 의미함(Hyndman et al., 2008).

○ 가장 기본적인 단순 지수평활(simple exponential smoothing)은 다음과 같이 표현됨.

$$\hat{y}_{t+1|t} = \ell_t, \ell_t = \alpha y_t + (1 - \alpha)\ell_{t-1}$$

- ℓ_t 는 시점 t 의 수준(level), α 는 평활계수(smoothing parameter)이며 $0 < \alpha < 1$ 의 값을 가짐. 현재 수준은 실제 관측값 y_t 와 이전 수준 ℓ_{t-1} 의 가중평균으로 계산됨.

○ 추세를 포함한 Holt의 선형 추세 모형(Holt's linear trend method)은 다음과 같이 확장됨(Holt, 1957).

- 수준 방정식: $\ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)(\ell_{t-1} + b_{t-1})$
- 추세 방정식: $b_t = \beta(\ell_t - \ell_{t-1}) + (1-\beta)b_{t-1}$
- 계절 방정식: $s_t = \gamma(y_t - \ell_{t-1} - b_{t-1}) + (1-\gamma)s_{t-m}$
- 예측 방정식: $\hat{y}_{t+h|t} = \ell_t + hb_t + s_{t+h-m(k+1)}$
- 여기서 s_t 는 계절성(seasonality), m 은 계절 주기(seasonal period), γ 는 계절 평활계수임.

5. CatBoost

○ CatBoost(Prokhorenkova et al., 2018)는 Gradient Boosting Decision Tree(GBDT) 기반의 머신러닝 알고리즘으로, 범주형 변수 처리와 과적합 완화를 위해 개발된 예측모형임.

○ CatBoost(Categorical Boosting)는 의사결정나무(decision tree)를 기반 학습기(base learner)로 사용하는 그래디언트 부스팅 계열의 머신러닝 알고리즘임.

- 일반적인 그래디언트 부스팅 모형은 손실함수를 최소화하는 방향으로 의사 결정나무를 순차적으로 학습하며, 이전 단계에서 설명하지 못한 오차를 다음 단계 모형이 보완하는 방식으로 예측 성능을 향상시킴. CatBoost는 이러한 그래디언트 부스팅 구조를 기반으로 하면서 범주형 변수 처리와 과적합 완화 측면을 개선한 알고리즘임(Prokhorenkova et al., 2018).
- CatBoost의 주요 특징 중 하나는 범주형 변수를 별도의 원-핫 인코딩(one-hot encoding) 없이 처리할 수 있다는 점임. 일반적으로 범주형 변수는 수치형 변수로 변환하는 전처리 과정이 필요하지만, CatBoost는 범주형 변수의 정보를 학습 과정에서 직접 활용할 수 있음.⁵⁾
- 이를 위해 CatBoost는 Ordered Target Statistics 방식을 활용함. 이는 범주형 변수의 범주별 평균 목표값을 순차적으로 계산하여 활용하는 방법으로, 범주형 변수 처리 과정에서 발생할 수 있는 정보 누출(target leakage) 문제를 완화하는 데 도움을 줌.
- 또한, CatBoost는 Ordered Boosting 기법을 적용하여 예측 시점에 사용할 수 없는 정보가 학습 과정에 반영되는 문제(prediction shift)를 줄이고 과적합 가능성을 완화함.
- CatBoost는 수치형 변수의 결측치를 별도의 사전 대체 과정 없이 처리할 수 있으며, 결측 여부 자체를 정보로 활용할 수 있다는 장점이 있음. 이러한 특성은 관측 누락이 존재하는 소비 데이터를 적용할 때 유용할 수 있음.
- CatBoost는 반복적으로 의사결정나무를 학습하면서 이전 단계의 예측오차

⁵⁾ CatBoost documentation(<https://catboost.ai/docs/en/>), 검색일: 2026. 3. 5.

를 보완하는 방식으로 예측력을 향상시킴. 또한, 비선형 관계와 변수 간 상호 작용을 효과적으로 반영할 수 있어 다양한 분야의 예측 문제에 활용됨.

○ CatBoost의 기본 구조는 다음과 같은 부스팅 모형으로 표현할 수 있음. 예측 함수 $F(x)$ 는 여러 개의 회귀나무 $h_m(x)$ 의 가중합으로 구성됨.

$$- \hat{y}_i = F_M(x_i) = \sum_{m=1}^M \nu h_m(x_i)$$

○ 여기서 x_i 는 i 번째 관측치의 설명변수 벡터, $\hat{y}_i$ 는 예측값, $h_m(x_i)$ 는 m 번째 의사결정나무, M 은 나무의 개수, ν 는 학습률(learning)을 의미함. 각 단계에서 CatBoost는 손실함수 $L(y_i, F(x_i))$ 를 줄이는 방향으로 새로운 나무를 추가함.

$$- F_m(x) = F_{m-1}(x) + \nu h_m(x)$$

○ 일반적인 그래디언트 부스팅에서는 다음 단계의 나무가 손실함수의 음의 그라디언트, 즉 pseudo-residual을 설명하도록 학습됨.

$$- g_i^{(m)} = - \left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)} \right]_{F=F_{m-1}}$$

○ 따라서 m 번째 나무는 다음 문제를 근사적으로 해결함.

$$- h_m = \arg \min_h \sum_{i=1}^n (g_i^{(m)} - h(x_i))^2$$

○ 회귀 문제에서는 보통 제곱오차 손실함수 또는 MAE, RMSE 등을 기준으로

모형을 평가할 수 있으며, 소비량 예측에서는 실제 소비량 y_i 와 예측 소비량 $\hat{y}_i$ 간의 차이를 최소화하는 방향으로 모형을 학습함.

- CatBoost의 중요한 이론적 특징 중 하나는 ordered boosting임. 기존 그래디언트 부스팅에서는 모든 학습자료를 사용하여 각 단계의 잔차 또는 그래디언트를 계산하기 때문에, 동일한 자료가 모형 학습과 잔차 계산에 반복적으로 사용되면서 예측 편의(prediction shift)가 발생할 수 있음.
- Prokhorenkova et al.(2018)은 이러한 현상이 목표변수 누출과 유사한 문제를 만들 수 있음을 지적하고, 이를 완화하기 위해 순열 기반의 ordered boosting을 제안하였음. Ordered boosting은 임의의 순열에 따라 각 관측치의 예측값 또는 범주형 변수 변환값을 계산할 때 해당 관측치보다 앞선 자료만 사용하도록 하여, 학습 과정에서 미래 정보 또는 자기 자신의 목표값이 반영되는 문제를 줄임.
- 범주형 변수 $x_{i,k}$ 에 대한 목표통계량은 일반적으로 다음과 같이 표현할 수 있음.

$$- TS(x_{i,k}) = \frac{\sum_{j=1}^n 1(x_{j,k} = x_{i,k})y_j + aP}{\sum_{j=1}^n 1(x_{j,k} = x_{i,k}) + a}$$

- 여기서 $x_{i,k}$ 는 i 번째 관측치의 k 번째 범주형 변수값, y_j 는 목표변수, P 는 사전값(prior), a 는 사전값에 부여되는 가중치임.
- 그러나 위와 같은 단순 목표통계량은 i 번째 관측치의 목표값 y_j 가 자기 자신의 설명변수 변환에 포함될 수 있어 목표변수 누출 문제가 발생할 수 있음. CatBoost는 이를 방지하기 위해 순열 σ 를 설정하고, i 번째 관측치에 대한 목표통계량을 해당 관측치보다 순서상 앞선 관측치만 이용하여 계산함.

$$- TS_{\sigma}(x_{i,k}) = \frac{\sum_{j:\sigma(j) < \sigma(i)} 1(x_{j,k} = x_{i,k})y_j + aP}{\sum_{j:\sigma(j) < \sigma(i)} 1(x_{j,k} = x_{i,k}) + a}$$

- 이러한 ordered target statistics는 범주형 변수의 정보를 활용하면서도 목표변수 누출을 줄이는 방식임. 이 때문에 CatBoost는 업종, 사업체 유형, 지역, 국산 사용 여부, 조사 월 등 범주형 변수가 포함된 자료에서 특히 유용하게 활용될 수 있음.
- 또 다른 특징은 대칭 트리(symmetric tree 또는 oblivious tree)를 사용한다는 점임. 일반적인 의사결정나무는 각 분기마다 서로 다른 변수를 기준으로 분할할 수 있으나, CatBoost는 동일한 깊이의 모든 노드에서 같은 분할 기준을 적용하는 대칭 구조의 나무를 사용함. 이 방식은 모형 구조를 단순화하여 예측 속도를 높이고, 과적합을 완화하며, 계산 효율성을 개선하는 데 기여함. 공식 문서 및 관련 자료에서는 CatBoost가 ordered boosting, 범주형 변수 처리, GPU 학습 및 빠른 예측 구조 등을 주요 장점으로 제시하고 있음.

6. 고정효과 패널모형(Fixed Effects Model)

- 본 연구의 산업 부문 자료는 동일 사업체의 양파·마늘 구매량이 여러 시점에 걸쳐 반복적으로 조사된 구조로, 사업체-월(panel) 형태의 자료로 볼 수 있음. 패널자료는 시계열 자료와 횡단면 자료의 특성을 동시에 포함하고 있어 개체별 특성과 시간 변화 정보를 함께 반영할 수 있다는 장점이 있음.
- 산업 부문 소비 자료는 사업체별 소비 구조와 구매 특성이 서로 다를 가능성

이 존재함. 예를 들어, 외식업체와 제조업체는 양파·마늘 사용 목적과 구매 규모가 상이할 수 있음. 또한, 동일 업종 내에서도 사업체 규모, 메뉴 구성, 원재료 사용 비중 등에 따라 소비 수준이 달라질 가능성이 존재함. 이러한 사업체 간 이질성(heterogeneity)을 고려하지 않는 경우 추정 결과의 편의가 발생할 가능성이 있음.

○ 본 연구에서는 업체별 고유 특성이 소비량에 영향을 미칠 수 있다고 판단하여 고정효과모형을 적용하였음. 고정효과모형은 관측되지 않는 업체별 특성을 통제된 상태에서 설명변수의 영향을 추정할 수 있다는 장점이 있음.

○ 고정효과모형(FE)은 개체별로 고정되어 있으나 관측되지 않는 특성이 설명변수와 상관되어 있다고 가정하는 패널회귀모형임(Wooldridge, 2010). 양파·마늘을 사용하는 사업체마다 고유한 구매 특성이나 운영 방식이 존재할 수 있으며, 이러한 특성이 설명변수와 연관될 가능성이 존재함. 예를 들어, 대형 외식업체는 지속해서 높은 구매량을 유지할 가능성이 있으며, 동시에 특정 업종 특성과도 연관될 수 있음. FE 모형은 이러한 사업체별 고정 특성을 통제함으로써 설명변수의 순수한 영향을 추정하고자 함.

○ 일반적인 고정효과모형은 다음과 같이 표현할 수 있음.

$$- y_{it} = \alpha_i + \beta X_{it} + \varepsilon_{it}$$

- 여기서 y_{it} 는 사업체 i 의 t 시점 소비량, X_{it} 는 설명변수 벡터, ε_{it} 는 오차항이며, α_i 는 사업체별 고정효과(individual fixed effect)로 시간에 따라 변하지 않는 사업체 고유 특성을 나타냄.

○ FE 모형은 사업체별 불변 특성을 제거하기 위해 개체 내 변화(within varia-

tion)를 이용하여 추정함. 즉, 동일 사업체 내에서 시간에 따른 소비량 변화를 중심으로 설명변수의 영향을 추정함.

- 따라서 사업체별 고유 특성이 설명변수와 상관되어 있더라도 추정 편의를 줄일 수 있다는 장점이 있음. 특히, 산업 부문 소비 자료는 사업체 간 이질성이 클 가능성이 있으므로, 개체별 고정 특성을 통제하는 접근이 필요할 수 있음.

참고문헌

- 국가데이터처(각 연도), 농작물생산조사.
- 김관수·안동환·성재훈(2009), “농업관측정보의 사회적 후생효과 추정,” 농촌경제, 32(5): 1-16.
- 김배성·고성보·현창석·고광훈·김충현(2014), 농업관측사업의 평가, N2014-1, 제주대학교·한국농촌경제연구원.
- 김상효·이계임·허성운(2016), 농산물 소비관측 도입 및 운영방안 연구, C2016-33, 한국농촌경제연구원.
- 김성훈·홍승지·명수환·이도경·유주영(2015), 농업관측사업 평가 및 개선 방안, (사)한국농업경제학회·한국농촌경제연구원.
- 김인석·김라이·변승연·김화영(2013), 견고추·마늘 중기선행관측 모형 개발, N2013-34, 한국농촌경제연구원.
- 김인석·정가연·변아름(2014), 양파·대파 중기선행관측 모형 개발, N2014-34, 한국농촌경제연구원.
- 농림축산식품부(각 연도), 농림업생산지수.
- 서동주·김준기·최성천·양지아(2022), 농업관측사업 효과 분석, E22-2021, 한국농촌경제연구원.
- 서홍석·김충현·김준호(2020), 농업부문 전망모형 KREI-KASMO 2020 운용·개발 연구, 한국농촌경제연구원.
- 석현덕·민정택(2003), 표고버섯 산업의 동향과 관측사업 구상, P65, 한국농촌경제연구원.
- 신용광·한승용·심민희·반현정(2010), 농산버섯 산업 현황과 관측 기법 연구, D290, 한국농촌경제연구원.
- 이용선·김연중·김윤희(2002), “농업관측의 경제적 효과 분석,” 농촌경제, 25(3): 1-15.
- 이형용·노호영·이형우·정세미·윤성주·김다정·윤성욱·정채은·정희주(2025a), 2025년 농식품 소비정보 분석사업, 한국농촌경제연구원.
- 이형용·노호영·정세미·박지원·정채은(2025b), 선제적 수급 관리를 위한 소비 관측 기초 연구, P2024-11, 한국농촌경제연구원.
- 한국농촌경제연구원(각 연도), 식품소비행태조사.
- 한석호·이정민·박미성·박영구·장석진(2010), 중기선행관측 기본모형 개발연구, M102, 한국농촌경제연구원.
- 허덕·이형우·이정민·경준형(2010), 오리산업 현황과 관측기법 연구, D288, 한국농촌경제연구원.

- Ahn, H., K. Sun & K. P. Kim(2022), "Comparison of missing data imputation methods in time series forecasting," *Computer, Materials & Continua*, 70(1): 767-778.
- Box, George E. P. & Gwilym M. Jenkins(1970), *Time Series Analysis: Forecasting and Control*, San Francisco: Holden-Day.
- Box, George E. P., Gwilym M. Jenkins, Gregory C. Reinsel & Greta M. Ljung(2016), *Time Series Analysis: Forecasting and Control*, 5th ed. Hoboken, NJ: Wiley.
- FAO(2001), *FOOD BALANCE SHEETS A handbook*, FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS Rome.
- Holt, Charles C.(1957), "Forecasting Trends and Seasonal by Exponentially Weighted Averages", *Carnegie Institute of Technology*.
- Hyndman, R. J., A. B. Koehler, J. K. Ord & R. D. Snyder(2008), *Forecasting with Exponential Smoothing: The State Space Approach*, Springer.
- Hyndman, Rob J. & George Athanasopoulos(2021), *Forecasting: Principles and Practice*, 3rd ed, Melbourne: OTexts.
- Hyndman, Rob J., Yeasmin Khandakar, Anne B. Koehler & Ralph D. Snyder(2002), "A State Space Framework for Automatic Forecasting Using Exponential Smoothing Methods", *International Journal of Forecasting* 18(3): 439-454.
- Martyushev, N. V., K. Spitsin, R. V. Klyuev, L. Spitsina, V. Y. Konyukhov, T. A. Oparina & A. E. Boltrushевич(2025), "Predicting firm's performance based on panel data: Using hybrid methods to improve forecast accuracy", *Mathematics*, 13: 1-33.
- Pankratz, Alan(1991), *Forecasting with Dynamic Regression Models*, New York: Wiley.
- Prokhorenkova, Liudmila, Gleb Gusev, Aleksandr Vorobev, Anna Veronika Dorogush & Andrey Gulin(2018), "CatBoost: Unbiased Boosting with Categorical Features", *Advances in Neural Information Processing Systems* 31: 6638-6648.
- Wooldridge, Jeffrey M.(2010), *Econometric Analysis of Cross Section and Panel Data*, 2nd ed. Cambridge, MA: MIT Press.

〈온라인 자료〉

농업관측센터(<https://aglook.krei.re.kr/main/contents/officeIntro0301>), 검색일:
2026. 3. 27.

CatBoost documentation(<https://catboost.ai/docs/en/>), 검색일: 2026. 3. 5.

KREI

www.krei.re.kr



소비 관측 모형 개발 및 운영 방안
연구: 양파·마늘

한국농촌경제연구원

전라남도 나주시 빛가람로 601 T.1833-5500 F.061) 820-2211

