creative commons

COMMONS DEED

저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:

저작자표시. 귀하는 원저작자를 표시하여야 합니다.

비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.

변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 이용허락규약(Legal Code)을 이해하기 쉽게 요약한 것입니다.

Disclaimer

dCollection

박사학위논문

농산물 도매시장 주별 가격예측 모델 및 출하 의사결정지원도구 개발 연구

2022년 8월

서울대학교 농업생명과학대학
농경제사회학부 지역정보전공
정민제

농산물 도매시장 주별 가격예측 모델 및 출하 의사결정도구지원 개발 연구

지도교수 최 영 찬

이 논문을 경제학박사학위논문으로 제출함

2022년 8월

서울대학교 농업생명과학대학

농경제사회학부 지역정보전공

정 민 제

정민제의 박사학위논문을 인준함

2022년 8월

위 원 장	김 홍 석	(인)
부 위 원 장	최 영 찬	(인)
위 원	나 명 환	(인)
위 원	박 흔 동	(인)
위 원	장 익 훈	(인)

<목 차>

제1장. 서론 ····· 1
제1절. 연구의 배경 ····· 1
1.1. 배경 및 목적 ····· 1

제2장. Hybrid FQMA를 활용한 주별 가락도매시장 토마토 가격 예측모델 연구 ····· 4
제1절. 연구의 배경 ····· 4
제2절. 예측 방법론 ····· 14
2.1. 예측 모델 프로세스 ····· 14
제3절. 분석 자료 ····· 20
제4절. 분석 결과 ····· 23
4.1. 변수 선정(Feature selection) 결과 ····· 23
4.2. 예측모델 평가(Forecasting evaluation) ····· 25

제3장. BMA와 PSO를 활용한 효율적인 농산물 출하 전략 수립을 위한 의사결정지원도구 개발 연구 · 32
제1절. 연구의 배경 ····· 32
제2절. 국내 농산물 유통 경로 및 A 연합사업단 ····· 36
2.1. 국내 농산물 유통 경로 ····· 36
2.2. A 연합사업단 ····· 38
제3절. 분석 자료 및 분석 모델 ····· 40
3.1. 분석 자료 ····· 40
3.2. 분석 모델 ····· 43

제4절. 분석 결과 ····· 44

제4장. 연구 결론 ····· 51

제1절. 제2장 연구 결론 ····· 51

1.1. 연구 결론 ····· 51

1.2. 한계점 및 후속 연구 ····· 52

제2절. 제3장 연구 결론 ····· 53

2.1. 연구 결론 ····· 53

2.2. 한계점 및 후속 연구 ····· 54

제5장. 참고문헌 및 부록(Appendix) ····· 56

참고문헌 ····· 56

부록(Appendix) ····· 68

부록 A. 국내외 연구 비교(도표) ····· 68

부록 B. H-FQMA 적용 통계 방법 ····· 70

부록 C. 주요 단일 모델 예측 성능 평가 ····· 79

부록 D. A 연합사업단 출하 효율성 분석 ····· 83

국문초록

서울대학교 농업생명과학대학
농경제사회학부 지역정보 전공
정민제

본 연구의 목적은 국내 최대 농산물 도매시장인 가락도매시장의 주별 토마토 가격 예측 모델 및 농산물 유통조직을 위한 의사결정지원도구를 개발하는데 있다. 국내 농산물 유통 조직(농협 등)의 농산물 시장 출하 환경에서는 주별 계약이 주를 이루고 있다. 또한, 가락도매시장 가격을 참조해 소매 조직과의 거래 가격이 형성되고 있다. 따라서 이들 유통 조직에 필요한 가격 예측 정보는 월 단위가 아닌 주 단위이며, 가락도매시장에 대한 가격 예측 모델 개발이 필요한 상황이다. 주별 도매시장 토마토 가격 예측 모델 개발을 위해 먼저, 국외 및 국내 농산물 가격 예측 통계 방법에 관한 선행연구를 진행하였다. 이를 통해 국외 연구에서 적용된 통계 방법 현황 및 국내 연구와의 격차를 분석하였다. 다음으로는 기존 FQMA(Factor quantile model averaging)에 변수 선정(Feature selection) 방법과 추가적인 통계 방법을 적용해 개선한 Hybrid FQMA(H-FQMA)를 제안한 후, 이에 대한 예측 성능을 평가하였다. H-FQMA는 예측 방법에서 예측 조합(Forecasts combination)에 해당되는 방법이다. 예측 조합은 국외 농산물 가격 예측 연구에서도 적용 빈도가 증가하고 있는 추세이다. H-FQMA의 예측 성능은 MAPE 기준 최고 11.2%를 기록하였으며, 적용된 벤치마킹 모델을 상회하는 예측 성능을 기록하였다.
농산물 유통 조직을 위한 주별 의사결정지원도구 개발에서는 농협 내 시장 출하 의사결정 조직 중 하나인 A 연합사업단 거래 자료와 도매시장 자료를 활용해 연구를 진행하였다. 본 연구의 의사결정지원도구의 목적은 시장 영향력을 보유한 농산물 유통 조직이 시장 상황 및 내부 출하 전략에 따라 가락도매시장 가격의 변동성을 분석할 수 있는 웹 기반(Web based)의 도구 개발에 있다. 상기 의사결정지원도구 개발을 위해 BMA(Bayesian model averaging)을 활용해 A 연합사업단과 주요 도매시장 토마토 반입량이 가락도매시장

토마토 가격에 미치는 영향력을 추정하였다. 다음으로는 PSO(Partical swarm optimization)를 적용해 주어진 환경 내 가락도매시장 가격이 높아지는 경우의 수를 추출하는 모델을 개발하였다. 이러한 결과를 종합해 웹 기반의 의사결정지원도구를 개발해 실무자의 활용 편의성을 제고하였다.

주요어 : FQMA, 예측 조합, 토마토, 농산물 가격, BMA, PSO

학 번 : 2016-30388

제1장. 서 론

제1절. 연구의 배경

1.1. 배경

그간 농산물 가격 예측에는 통계적 방법이 주로 활용되었다. 그러나 이와 관련된 문헌 연구가 부족한 상황이다. 일부 문헌 연구에서 데이터 마이닝(Data mining) 및 기계학습(Machine learning)에 관한 분석이 진행되었으나, 목표 변수, 예측 주기, 예측 성능 평가 지표 등 포괄적 관점에서 진행된 연구가 미흡하다. 따라서, 세부적인 분류 체계(목표 변인 성격, 예측 주기, 통계 방법, 평가 지표 등)에 따른 국내외 농산물 가격 예측 방법에 관한 문헌 연구가 필요하다.

또한, 제한적 정보를 활용한 주별 농산물 가격 예측 모델 개발 역시 필요하다. 국내 농산물 유통조직(농협 등)의 농산물 시장 출하는 주 단위 계약이 주로 이루어지고 있다. 지역 농협 연합 조직인 A 연합사업단은 이마트, 홈플러스 등 소매(Retail) 조직에 물량을 공급할 시, 주단위 계약을 통해 거래가 이루어진다. 거래 가격도 해당 시점의 가락도매시장 가격을 기준으로 형성되어, A 연합사업단과 같은 농산물 유통조직에게는 주별 가락도매시장의 가격 및 반입량 동향이 가장 중요한 정보이다. 그러나 국내 농산물 가격 예측은 주로 월 단위, 연 단위로 이루어지고 있으며, 한국농촌경제연구원(KREI) 역시 관측 정보를 월 단위로 제공하고 있어, 농산물 유통조직 입장에서는 예측 정보 활용에 제한적인 상황이다.

농협 내 A 연합사업단과 같은 대형 농산물 유통조직은 도소매 등 시장 출하 정보를 활용하고 있으며, 이러한 조직의 효율적인 출하 전략 수립을 위해 예측 모델과 이를 편리하게 활용할 수 있는 의사결정지원도구 개발이 필요하다. 농협 내 광역 단위로 존재하는 연합사업단의 주요 목적은 해당 광역 지역의 농협 및 농가로부터 확보된 계약 물량을 높은 가격에 공급하고, 이를 통해 참여 농협 및 농가의 이익을 극대화하는 것이다. 연합사업단은 광역 지역 내 농가 이익 보전에 기여 중이며, 더 나아가 수급 안정에도 영향을 미치고 있다. 따라서, 이들 농산물 유통 조직의 거래 자료를 활용해 예측 모델의 질을 향상시키고, 출하 의사결정의 효율성을 제고할 수 있는 의사결정지원도구 개발이 필요하다.

앞선 내용을 바탕으로 먼저 가락도매시장의 주별 토마토 가격 예측 모델을 개발하였다. 이를 위해 국내외 연구를 분석하고, 목표변수(Target variable), 예측 주기, 사용한 통계 방법, 예측 성능 평가 지표에 따라 연구를 정리하였다.

분석 결과를 바탕으로 가락도매시장 주별 토마토 가격 예측 모델 개발을 위해 최근 농산물 가격 예측 연구에서 활용 빈도가 높아지고 있는 예측 조합(Forecasts combination) 모델을 활용하여 범용적 성격을 지닌 예측 모델을 개발하였다. 예측 조합 방법으로 Model stacking 기반 FQMA(Factor quantile model averaging)를 개선한 Hybrid FQMA(H-FQMA)를 적용하였다. H-FQMA의 예측 성능 평가를 위해 세 개의 예측 조합 모델을 마련해 H-FQMA가 예측 성능을 비교 분석하였다. 예측 성능 평가를 위한 지표로는 RMSE(Root mean square deviation), MAPE(Mean avsolute percentage error), MZ(Mincer and Zarnowitz) 검증, SIS(Step indicator saturation) 등의 지표를 적용하였다.

다음으로 토마토 관련 농산물 유통조직의 출하 효율성 제고를 위한 출하 의사결정지원도구(Decision making support tool)를 개발하였다. 농산물 유통조직은 소매 조직과 계약을 체결할 때, 가락도매

시장 가격을 참조하여 거래 가격을 결정한다. 즉, A 연합사업단이 이마트, 홈플러스 등 소매 조직과 출하 계약 체결 시 해당 시점 가락도매시장 토마토 가격을 참조해 거래 가격이 형성된다. 따라서, 소매 및 도매 조직에 공급되는 물량을 가락도매시장 가격이 상승하도록 유도할 수 있다면 이익을 증대시킬 수 있을 것이다. 이를 위해서는 먼저, 농산물 유통조직 거래 자료를 활용한 의사결정지원도구 개발이 필요하다.

이를 위해 A 연합사업단의 토마토 시장 출하(도매 및 소매 출하량) 및 주요 도매시장의 토마토 반입량 정보를 활용해 가락도매시장 토마토 가격 모델을 추정한 후, 이를 토대로 최적화 모델을 적용해 가락도매시장 가격이 극대화되는 다수의 시나리오를 추출하여, 개발된 모델을 쉽게 활용할 수 있도록 웹 기반(Web based) 의사결정지원도구를 개발하였다.

제2장. Hybrid-FQMA를 활용한 주별 가락도매시장 토마토 가격 예측모델 연구

제1절. 연구의 배경

최근 국내 농산물 유통 시장에서는 토마토 수급 예측의 중요성이 높아지고 있다. 국내 토마토 생산량은 2019년 기준 358,580 톤(t)으로 증가세는 완만한 상태에 접어들었으나, 먹거리 형태가 서구화되면서 소비자 수요는 여전히 높은 상황이다(Kim et al., 2016). 웰빙과 다이어트에 관심이 집중되고 있는 만큼, 높은 포만감과 맛을 동시에 주는 토마토 수요가 증가하면서 토마토는 딸기와 함께 국내 시설원예 대표 품목으로 인식되고 있다(이동훈, 2018). 이에 따라, 본 연구에서는 Hybrid FQMA(Factor quantile model averaging, H-FQMA)를 활용해 국내 최대 농산물 도매시장인 가락도매시장의 토마토 주별 단위 가격(kg) 예측 모델을 개발하였다. H-FQMA는 변수 선정, 예측 조합, Quantile가 결합된 방법으로 농산물 가격의 높은 변동성에 대응하기 위한 모델이다. FQMA는 전기수요예측 분야에서 Maciejowska et al.(2016)에 의해 제안된 예측 조합(Forecasts combination) 모델로, 변동성이 높은 전기수요예측 분야에서 우수한 예측 성능을 기록하였다. 본 연구에서는 RMCFS(R-Monte carlo feature selection)와 BMA(Bayesian model averaging)와 같은 변수 선정(Feature selection) 방법을 FQMA에

추가하였다. 또한 단변량(Univariate) 시계열 방법만을 활용했던 기존 모델과는 다르게 일반 회귀, 기계학습 등의 통계 방법을 추가해 기존 농산물 가격 예측 연구에서 적용된 바 없는 새로운 예측 모델을 제안하고자 한다.

H-FQMA 개발을 통한 시사점으로는 먼저, 토마토 주별 가격 예측 정보가 필요한 농산물 유통 관계자를 위한 의사결정 지원이 가능하다. 농가로부터 원물을 매입해 소매 시장과 거래하는 농산물 유통 조직은 소매상과의 거래 시 주별 단위 계약이 주를 이루는데, 이때 거래 가격은 가락도매시장 가격을 참조해 형성된다. 따라서 거래가 형성되는 t시점 전 t-1시점에 가격을 전망할 수 있다면 출하 이익을 추정할 수 있고, 나아가 출하 의사결정을 효율적으로 수립할 수 있다. 제한된 자료를 활용해 변동성이 높은 변인을 예측하는 모델 개발 시, 모델의 예측 성능을 향상시킬 수 있다. 토마토 주별 가격 예측을 위해 국내에서 공식적으로 확보할 수 있는 자료는 도매시장 거래 자료가 있다. 이러한 제한된 자료만을 활용해야 할 경우 다수의 파생변수(시차, 이동평균 등)를 생성한 후 예측 변수를 선정하는 즉, 탐색적 성격의 연구가 필요하며, 적절한 변수 선정 방법이 요구된다. 토마토 가격은 변동성이 높은 농산물 중 하나다. 목표 변인의 변동성이 높을 때 단일 모델(Individual model) 활용은 예측 성능 위험도를 높아질 우려가 있다(Ramos-Pérez et al., 2019). 이를 보완하려 예측 연구에서는 예측 조합 방법으로 위험도를 감소하려는 시도가 진행되어 왔다(Wang et. al, 2019; Bates & Granger, 1969; Nowotarski et al., 2016; Sanchez, 2008).

H-FQMA 개발을 위해 먼저, 농산물 가격 예측 방법에 대한 국내외 연구를 수집 및 분석하여, 사용된 방법을 단일 모델, 하이브리드 모델, 예측 조합 모델 등 세 가지로 구분하였다. 이를 다시 일반 회귀(General regression), 시계열 분석(TSA), 데이터 마이닝(Data mining), 기계학습(Machine learning), 딥러닝(Deep learning) 등으로 재차 구분하였다. 예측 주기 및 성능 평가 방법에 대한 내용도 혼합

하여 정리하였다. 통계 방법에 관한 분류 체계는 기존 농산물 가격 예측 방법론 문헌 연구(e.g. Wang et al., 2020; Wihartiko, Fajar Delli, et al., 2021)에서 분류한 기준을 응용하였다.

표 1. 국외 농산물 가격 예측 연구 기반 통계 방법 분류

저자	목표 변수	예측 주기	방법			예측 성능 지표
			단일 모델	하이 브리드	예측 조합	
Zhang et al.(2018)	식량작물	월		C		비율
Ahumada & Conrnejo(2016)	식량작물	분기	B			단위, 비율, 기타
Kyriazi et al.(2019)	식량작물	분기			B	단위
Li et al.(2012)	축산물	월	A			단위
Li et al.(2010)	축산물	월	A			단위
Wang et al.(2019)	식량작물	일			B, C, D	단위, 비율
Kaur et al.(2014)	식량작물	월	C			단위
Nasira & Hemageetha(2012)	채소	월, 주	C			비율
Deepalakshmi et al.(2019)	채소	월	C			비율
Hasan et al.(2020)	채소	월	C			비율
Hemageetha & Nasira(2013)	채소	주		C		단위
Chen et al.(2021)	채소, 축산물	주	C			단위
LI et al.(2013)	축산물	주	C			단위, 비율
Shin et al.(2009)	채소	월		C		단위

A: 일반 회귀, B: 시계열, C: 데이터 마이닝 & 기계학습, D: 딥러닝

저자	목표 변수	예측 주기	방법			예측 성능 지표
			단일 모델	하이 브리드	예측 조합	
Yuan & Ling(2020)	채소, 축산물	주	C			단위
Zhang & Na(2018)	식량작물, 채소, 축산물	월		C		단위, 방향성
Zong & Zhu(2012)	식량작물, 축산물	주	B, C			비율
Jeong et al.(2017)	채소	월	C			비율
Sharifuddin et al.(2013)	축산물	월	B			단위
Wiles & Enke(2015)	식량작물	일	C			기타
Xiong et al.(2015)	식량작물	일		B, C		단위
Gargano & Timmermann(2014)	축산물	년, 분기, 월	A			단위
Onour & Sergi(2011)	식량작물, 축산물	월	A			단위
Luo et al.(2022)	식량작물	일		B		비율, 방향성
Degiannakis et al.(2022)	식량작물	일		B		비율, 방향성

A: 일반 회귀, B: 시계열, C: 데이터 마이닝 & 기계학습, D: 딥러닝

단일 모델은 단수의 통계 모델을 예측에 활용한 것을 의미한다. 먼저, OLS(Ordinary least squares), GLS(Generalized least squares) 등 회귀 방법을 적용한 농산물 가격 예측 연구는 목표 변인에 대한 예측 변수의 설명력이 요구될 시 활용된 사례가 많다(Li et al.(2010), Li et al.(2012), Gargano & Timmermann(2014), Onour & Sergi(2011)). 시계열 분석 방법에서는 ECM(Error correction model)이 주로 활용되었다. 변수 간 장기적 균형 즉, 공적분(Co-integration)이 존재하는 것으로 가정할 시 오차수정모델이 활용되어 왔다(Ahumada & Conrnejo, 2016; Sharifuddin et al.2013). 기계학습을 활용한 단일 모델은 ANN(Artificial neural network), SVR(Support vector regression), LSTM(Long short-term memory)이 주로 활용되었다(Nasira & Hemageetha, 2012; LI et al., 2013; Jeong et al., 2017). 세 가지 방법은 타 분야 예측 연구에서도 높은 빈도로 활용되고 있다. Chen et al.(2021), Yuan & Ling(2020)은 LSTM, SVR, ARIMA 등 5개의 기계학습 및 시계열 방법을 개별적으로 적용해 농산물 가격을 예측한 후 예측 성능을 비교하였다.

하이브리드 모델은 순차적으로(Sequentially) 복수의 방법을 결합하는 방법이다. 예를 들어, 변수 선정 과정에서 특정 변수 선정(Feature selection) 방법으로 예측변수를 선별한 후, 특정 예측 방법에 적용한다거나, 시계열 방법 또는 분해(Decomposition) 기법을 활용한 예측변수 조작으로 발생된 산출물을 예측 변수로 활용하는 등의 방법이 존재한다. Zhang et al.(2018)은 RBF(Radial based function) ANN의 매개변수 학습 과정에 GA(Genetic algorithm)과 SGD(Stochastic gradient descent)를 상황에 맞게 선별하는 방법을 적용하였다. Shin et al.(2009)은 주어진 예측 변수에 GA를 적용한 가중치 부여 방법을 통해 변수를 선정한 후, ANN과 CART(Classification and regression tree)에 적용하였다. Zhang & Na(2018)는 FIG(Fuzzy information granulation)을 활용해 예측 변

수를 분해하고, MEA로 초매개변수(Hyper-parameter)를 최적화한 SVR 모델을 적용한 예측 모델을 개발하였다. 시계열 방법 간 결합을 통해 예측모델을 개발한 연구도 존재한다. Xiong et al.(2015)은 VECM(Vector error correction model)에서 발생한 발생한 오차를 SVR의 예측변수로 활용해 일별 옥수수 선물 가격을 예측하는 모델을 연구하였다. 이 외에도 일별 농산물 가격 변동성(Volatility) 예측 연구에서 Hidden markov switching과 HAR(Heterogeneous autoregressive) 모델을 결합한 사례도 있다(Luo et al., 2022).

예측 조합은 단일 모델에서 발생 가능한 위험도에 대응하기 위한 방법으로, 과거 농산물 가격 예측 연구에서는 시도되지 않았으나, 최근 들어 예측 조합을 시도하는 연구가 나타나고 있다. 예측 조합 모델의 과제 중 하나는 각 모델에서 산출되는 결과에 가중치를 부여해 최적의 예측치를 산출하는 것으로, 주로 활용되는 방법이 모델에서 산출된 개별 결과값에 평균, 중앙값, 가중평균 등을 적용하는 것이다(Wang et. al, 2019; Blanc & Setzer, 2016; Hsiao & Wan, 2014). 이보다 고도화된 방법으로는 BA(Bayesian averaging), Boosting, Bagging 등의 방법이 여러 연구에서 활용되어 성능이 입증된 바 있다(Barrow & Crone, 2016; Graefe et al., 2015; Petropoulos et al., 2018; Bennett et al., 2014). 더불어, 여러 모델에서 나온 예측치를 다시 특정 모델의 예측변수로 적용하는 Model stacking도 존재한다. 관련 연구 사례로는, RF(Random forest), SVR, GARCH(Ganeralized ARCH model) 등에서 산출된 예측치를 다시 ANN에 투입한 연구를 꼽을 수 있다(Ramos-Pérez et al., 2019).

농산물 가격예측 연구에서의 예측 조합 연구를 살펴보면, Kyriazi et al.(2019)은 다수의 시계열 방법에서 산출된 예측치를 조합해 과거 시점 예측 오차를 기반으로 예측 변수를 조정하는 'Adaptive learning'이라는 예측 기법을 개발하여 10개 유럽 국가의 농산물 가격 지수를 예측하였다. 이 과정에서 다수의 단변량 시계열 방법을

활용해 개별적인 예측 오차를 산출하였고, 이를 활용해 최종 예측 모델을 개발하였다. Wang et al.(2019)는 VMD(Variational mode decomposition), EMD(Empirical mode decomposition) 등의 분해 기법을 활용해 변수를 분해한 후, ARIMA, SVR, RNN(Recurrent neural network), GRU(Gated recurrent unit), LSTM 등 다섯 가지 모델을 활용해 개별 예측치를 산출하였다. 예측치를 조합하는 방법으로는 단순 평균이 아닌 최적화 기법 중 하나인 ABC 알고리즘을 활용해 단일 예측치를 생성하였다.

국내 농산물 가격 예측 연구는 시계열 방법이 주로 활용되는 경향을 보이고 있다. 명광식(2005)은 ARIMA 방법을 활용해 월별 육계 가격 예측 모델을 연구하였고, 남국현 & 최영찬(2015)는 ARDL 방법을 활용해 출하시기 양파 가격에 대한 월별 예측 모델을 연구하였다. 특히, 재배면적 및 단수를 예측한 후, 그에 대한 예측치를 출하시기 양파 가격 모델에 적용하였다.

최병옥 & 최익창(2016) 다변량 시계열 방법인 VAR(Vector autoregressive), 시계열 분산 방법 GARCH, 단변량 시계열 방법 ARIMA를 개별적으로 적용해 오이, 토마토, 호박 가격에 대한 월별 예측 모델을 연구하였다. 유도일(2016)은 BSTS(Bayesian structural time series)와 VAR을 활용해 월별 배추 가격을 예측하는 모델을 연구하였고, 이형용 외(2017)은 BVAR 및 ECM을 활용해 마늘 가격에 대한 월별 가격 모델을 개발하였다.

표 2. 국내 농산물 가격 예측 연구 기반 통계 방법 분류

저자	목표 변수	예측 주기	방법			예측 성능 지표
			단일 모델	하이 브리드	예측 조합	
명광식(2005)	축산물	월	B			단위
남국현&최영찬 (2015)	채소	월	A			비율
배경태&김창재 (2016)	채소	월		B, C		단위
최병옥&최익창 (2016)	채소	월	B			비율
유도일(2016)	채소	월	B			비율
윤성주 외(2016)	식량작물	일	A, C			비율
이형용 외(2017)	채소	일	B			단위, 비율
김종진 외(2017)	식량작물	분기	B			비율
채명수 외(2017)	채소	월	C			단위
장수희 외(2017)	채소	분기	B			단위
신성호 외(2018)	식량작물, 채소	일	C			단위
임진모 외(2018)	과수작물	월	C			단위
김종진 외(2018)	식량작물	월	B			단위
하지희 외(2019)	채소	월	B			비율

A: 일반 회귀, B: 시계열, C: 데이터 마이닝 & 기계학습, D: 딥러닝

그 외에도 수확기 쌀 가격 예측 모델, 비정형 데이터를 예측 변수로 생성해 시계열 모델을 활용한 연구 등이 수행되었고, 대부분의 연구는 ARIMA, GARCH, VECM과 같은 시계열 방법이 적용되었다(김종진 외, 2017; 장수희 외, 2017; 하지희 외, 2019).

기계학습을 활용한 연구에서는 LSTM과 같은 인공신경망 방법이 주로 활용되었다. 배경태 & 김창재(2016)는 OLS로 안동지역 오이에 미치는 영향 변수를 선별한 후, ANN을 이용해 가격을 예측하는 모델을 개발하였다. 윤성주 외(2016)는 옥수수, 쌀 등에 관한 국제 곡물 가격 지수를 예측하기 위해, 예측 변수를 수급요인, 외부시장요인으로 구분한 후, ANN을 적용하였다. 이 외에도 신성호 외(2018)는 LSTM을 활용해 대구, 광주, 서울, 대전, 부산 지역의 농산물(대파, 양파, 애호박 등) 가격예측 모델을 연구하였다.

국외 연구는 국내 연구 대비 적용된 방법론 범위가 다양하며, 최근에는 예측 조합 적용이 계속해서 시도되고 있다. 반면, 국내 연구는 시계열 방법을 활용한 단일 모델 연구가 주를 이루고 있다. ANN을 활용한 연구에서도 국외 연구는 최신 기술을 반영한 모델(학습과정에서의 최적화 방법, 초매개변수 학습법 등)을 개발하려는 움직임이 있는 반면, 국내 연구는 제한적인 범위 내에서 연구가 진행되고 있다. 또한, 국내 연구는 변동성이 높은 가격 예측에 유용한 예측 조합 모델에 대한 시도가 없는 상황이며, 이종 방법을 결합한 하이브리드 모델도 배경태 & 김창재(2016) 연구 외에는 찾아보기 힘들다.

국외 연구와는 달리 국내 농산물 가격 예측 연구는 월별 연구 또는 수확기 가격 전망에 집중되어 있다. 국내 농산물 가격 전망을 주도하는 한국농촌경제연구원은 월별 관측 정보를 제공하며, 따라서 대부분의 농산물 가격 예측 연구의 예측 주기가 월에 집중되어 있다. 수확기 가격을 예측하는 모델의 경우 활용할 수 있는 예측 변수의 범위가 확대되는 이점이 있다. 예를 들어, 주별 또는 월별 모델에서는 분기별 자료 또는 연도별 자료의 활용이 제한되는 반면, 수

확기 가격만을 예측하는 모델에서는 해당 자료 활용이 부분적으로 가능하다.

국내 연구 대비 국외 연구는 농산물 가격의 높은 변동성에 대응할 수 있는 방법 적용이 미진한 상황이다. 농산물 가격 예측의 정확성을 향상시키기 위해서는 예측 조합 및 Quantile 개념의 적용 등의 적용이 필요하다. 또한, 국내 농산물 유통 실무에서는 주별 단위 가격 예측이 요구됨에도 월별 예측 연구가 대부분이다. 이 시점에서 국내 농산물 가격 예측 방법의 범위를 확장할 수 있는 새로운 방법의 적용이 필요하다.

제2절. 예측 방법론

2.1. 예측 모델 프로세스

본 연구에서 제안하는 H-FQMA는 예측 조합 모델 중 Model stacking에 해당된다. Model stacking은 예측 조합 방법 중 하나로, 주로 변동성이 높은 주식시장 가격 예측이나 전기 수요 예측 연구에서 활용되고 있는 방법이다(Ramos-Pérez et al., 2019). Model stackingl은 개별 모델에서 발생한 예측치를 예측 변수(Predictor)로 재차 활용해 목표 변수(Target variable)를 다시 예측하는 방법이다.

예측 조합 연구의 주요 쟁점으로 개별 모델 간 불일치(Degree of disagreement)를 언급할 수 있다(Mendes-Moreir et al., 2012). 다수의 모델로 목표 변수에 대한 예측치를 생성할 때, 예측의 방향이 동종 성향인 모델(Homogeneous models)과, 이질 성향인 모델(Heterogeneous models)로 구분되며(Wang et al., 2019; essmann & Vob, 2017), Maciejowska et al.(2016)은 이 점을 고려해

Nowotarski & weron(2014)에 의해 제안된 QMA(Quantile model averaging)를 발전시켰다. 다수의 단변량 시계열 모델에서 발생한 예측치를 PCA(Principal component analysis)를 활용해 여러개의 성분으로 축약한 후, 이를 다시 Quantile regression에 적용해 전기수요량을 예측하는 예측 조합 모델을 사용하여 우수한 성능을 입증하였다. 본 연구는 Maciejowska et al.(2016)의 모델을 바탕으로 농산 기존 단변량 시계열 방법과 더불어 일반 회귀, 데이터 마이닝 & 기계학습 방법을 추가하였다. 단변량 시계열 방법 외 연구에 적용된 일반회귀, 데이터마이닝 & 기계학습 방법은 Lessmann & Voß(2017)의 연구를 참고하여 선정하였다. Lessmann & Voß(2017)은 그들의 중고차 가격 예측에서 예측 조합 모델을 적용하였으며, ANN, RF 등 범용적 성격을 가진 다수의 방법을 적용해 중고차 시장 가격을 예측하였다. 표 3.은 H-FQMA에 적용된 통계 방법 및 초매개변수 종류와 그에 대한 최적화 방법을 보여주고 있다.

표 3. H-FQMA에 적용된 통계 방법 및 초매개변수 정보

구분	방법	초매개변수	기타
단변량 시계열			
AR		-	Lag 1~8
ARX		-	Lag 1~8
TAR		-	Lag 1~8
NAR		• No. of layers & hidden nodes	Lag 1~8
SNAR		-	Lag 1~8
일반 회귀			
OLS		-	
Ridge		• Transformation of the target variable • Regularization parameter	Grid search
Lasso		〃	Grid search
GAM		• Optimizer methods • GCV penalty & Model type	Grid search
데이터 마이닝 & 기계학습			

Rprop+ ANN	• No. of layers & hidden nodes • Threshold & Learning rate	Pareto opt.
Rprop- ANN	〃	Pareto opt.
GD ANN	〃	Pareto opt.
Adam ANN	〃	Pareto opt.
AdaGrad ANN	〃	Pareto opt.
RBF SVR	• Cost & ϵ-insensitive loss • Sigma & Tolerance	Pareto opt.
Linear SVR	〃	Pareto opt.
Poly SVR	〃	Pareto opt.
CART	• Min. obs per leaf node • Pruning trees	Grid search
kNN	• No. of neighbors & Model type • Values by distance	Grid search
Random forest	• No. of trees & Node size • No. of max nodes	Pareto opt.
GBM	• No. of trees • No. of min nodes	Grid search

예측 변수 선정에는 BMA와 RMCFS를 적용하였다. 가락도매시장 토마토 가격은 다양한 변인(주요 도매시장 가격, 반입량 등)에 의해 영향을 받을 수 있다(Zhang et al., 2018; Baffes & Haniotis, 2016).

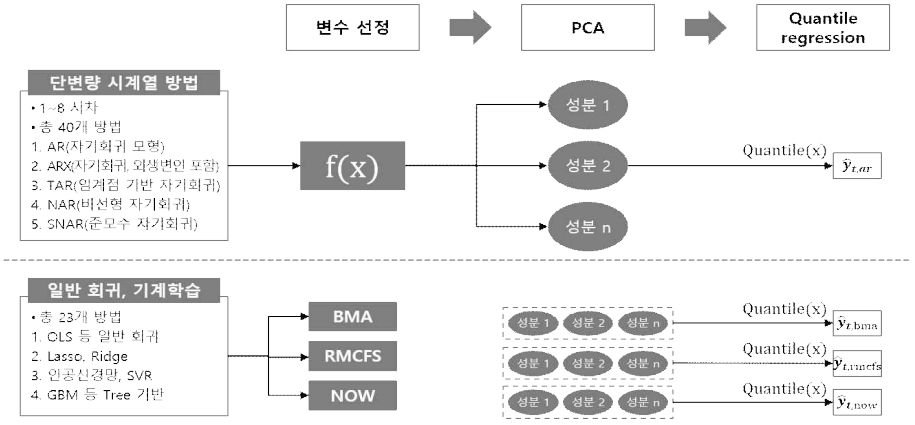

그림 1. H-FQMA의 진행 절차

주요 예측 변수 선정을 위해 BMA와 RMCFS를 활용하였다. BMA는 사전 분포(Prior distribution) 가정과 업데이트(Update) 등 베이지안 이론 기반 변수 선정 방법으로, 모델 불확실성(Model uncertainty)이 높은 탐색적 연구에 유용한 방법이다(Forte et al., 2018). RMCFS는 트리(Tree) 기반 변수 선정 방법으로, Draminski & Koronacki(2018)가 기존 MCFS를 고도화한 방법이다. 이를 통해 BMA를 통해 선정된 변수 집합, RMCFS를 통해 선정된 변수 집합, 마지막으로 목표 변수의 예측 시점인 t 시점 변수 집합 등 세 개의 변수 집합으로 예측 변수 후보군이 선정되었다.

데이터 마이닝 & 기계학습 방법에서 초매개변수 최적화는 예측 모델 개발에서 일반적으로 활용되는 그리드 검색(Grid search)과 파레토 최적화(Pareto optimization)를 활용한 방법을 활용하였다. 초매개변수(Hyper parameter) 학습은 데이터 마이닝 & 기계학습 방법의 예측 성능을 결정하는 중요한 요소로 본 연구에는 단변량 시계열 방법(Univariate time series), 일반 회귀 방법(General regression) 및 데이터 마이닝 & 기계학습 방법 등 총 63개 방법이 적용되었다.

H-FQMA의 분석을 위해 먼저, 주요 도매시장 가격 및 반입량 등의 시차 변수가 담긴 예측 변수 집합을 설정하고, 이들 중 식 (1)과과 같은 예측 변수 후보군을 선정하였다.

$$X = \begin{bmatrix} T_{week}, P_{garak,t:t-n}, P_{dae,t:t-n}, P_{buban,t:t-n}, P_{buum,t:t-n}, \\ Q_{garak,t:t-n}, Q_{dae,t:t-n}, Q_{buban,t:t-n}, Q_{buum,t:t-n}, \\ Q_{tratio,t:t-n}, P_{garakr,t:t-n}, Q_{garakr,t:t-n} \end{bmatrix} \quad (1)$$

가락도매시장 가격의 시차 변수 $P_{garak,t:t-n}$, 대구북부도매시장 가격 시차 변수 $P_{dae,t:t-n}$, 부산반여도매시장 가격 시차 변수 $P_{buban,t:t-n}$, 부산엄궁도매시장 가격 시차 변수 $P_{buum,t:t-n}$, 가락도매시장 반입량 시차 변수 $Q_{Qgarak,t:t-n}$, 대구북부도매시장 반입량 시차 변수 $Q_{dae,t:t-n}$, 부산반여도매시장 반입량 시차 변수 $Q_{buban,t:t-n}$, 부

산업궁도매시장 반입량 시차 변수 $Q_{buum,t:t-n}$, 전체 도매시장 반입량 중 가락도매시장 비율에 대한 시차 변수 $Q_{tratio,t:t-n}$, 가락도매시장 가격의 상대적 변화(해당 주차 가격 - 해당 연도 평균 가격)에 대한 시차 변수 $P_{garakr,t:t-n}$, 가락도매시장 반입량의 상대적 변화(해당 주차 반입량 - 해당 연도 평균 반입량)에 대한 시차 변수 $Q_{garakr,t:t-n}$ 등 이들 중 가락도매시장 가격에 영향을 미치는 주요 예측 변수를 선정하기 위해 BMA, RMCFS, NOW 등 세 가지 변수 집합을 선정하였다(식 (2)).

$$\begin{aligned} FS_{\text{bma}} &= \left[X_{1,\text{bma}}, X_{2,bma}, \ldots, X_{n,bma}\right] \\ FS_{\text{cfs}} &= \left[X_{1,\text{cfs}}, X_{2,\text{cfs}}, \ldots, X_{n,cfs}\right] \\ FS_{now} &= \left[X_{1,now}, X_{2,now}, \ldots, X_{n,now}\right] \end{aligned} \tag{2}$$

식 (2)에서 FS_{bma}는 BMA를 통해 선정된 주요 예측 변수 집합, FS_{cfs}는 RMCFS를 통해 선정된 주요 예측 변수 집합, FS_{now}는 t시점 예측 변수 집합을 의미한다. H-FQMA에서는 총 63개의 방법이 적용된다. 63개의 개별적인 방법은 아래와 식 (3)과 같이 정의된다.

$$m_i = \left[m_1, m_2, m_3, \ldots, m_n\right]_{i=1:63} \tag{3}$$

이 중 단변량 시계열 방법에는 가락도매시장 가격과 더불어 일부 외생 변수가 적용되고, 일반 회귀, 데이터 마이닝 & 기계학습에는 RMCFS와 BMA 등을 통해 생성된 세 개의 예측 변수 집합을 사용하여 주별 가락도매시장 토마토 가격 P_t가 예측된다.

$$
\begin{aligned}
P_{univ,m_i,t} &= \theta_0 + \sum_{i=1}^{8} \theta_1 P_{t-i} + \phi_t Q_{t-i} + \pi_t W_t + \epsilon_{univ,m_i,t} \\
P_{bma,m_i,t} &= f\left(FS_{bma,m_i,t};\, \hat{\theta}\right) + \epsilon_{bma,m_i,t} \\
P_{cfs,m_i,t} &= f\left(FS_{cfs,m_i,t};\, \hat{\theta}\right) + \epsilon_{cfs,m_i,t} \\
P_{now,m_i,t} &= f\left(FS_{now,m_i,t};\, \hat{\theta}\right) + \epsilon_{now,m_i,t}
\end{aligned} \tag{4}
$$

$P_{univ,m_i,t}$는 단변량 시계열 방법, $P_{bma,m_i,t}$는 BMA, $P_{cfs,m_i,t}$는 RMCFS, $P_{now,m_i,t}$는 예측 시점 변수로 생성된 가락도매시장 주별 가격 예측치를 의미한다. 단변량 시계열 방법의 경우, 외생 변수 Q_{t-i}, W_t가 추가된다. 생성된 예측치는 다시 주별 가락도매시장 토마토 가격 P_t를 예측할 수 있는 예측하는 예측 변수 X로 활용되었다.

$$
\hat{y}_{m_i,t} = \begin{pmatrix} \hat{y}_{m_1,1} & \cdots & \hat{y}_{m_n,1} \\ \vdots & \vdots & \vdots \\ \hat{y}_{m_1,n} & \cdots & \hat{y}_{m_n,n} \end{pmatrix} = X_i \tag{5}
$$

가락도매시장 주별 가격 P_t의 예측치인 $\hat{P}_{m_i,t}$는 방법 i에 따라 예측 방향이 군집화될 수 있다. 방법별 예측 방향성에 따라 n개의 성분($PC_1,\ldots,PC_n$)으로 축약될 수 있다. X_i에 SVD(Singular value decomposition)를 적용하면, $X = P \circ Q^T$와 같이 분해된다. 여기서 $P \circ$는 식 (6)과 같이 정의된다.

$$
\begin{aligned}
\Lambda &= P \circ \\
\Lambda &= P \circ = P \circ\ Q^T Q = XQ \\
&\text{여기서, } X = \Lambda Q^T \text{ and } Q^T Q = I
\end{aligned} \tag{6}
$$

F는 요인점수(Factor score)를 의미하며, 개별 방법 i에 의해 산출된 예측치가 n개의 성분 PC_n로 축약되었을 때, 각 성분 PC_i가 가지는 수치이다. 서로 다른 성분 PC_i와 PC_j는 상호 직교(Orthogonal) 관계이며($Cov(PC_i, PC_j) = 0$), 요인점수 Λ는 XQ로 정의되는데, X는 식 (5)에서 X_i를 의미하며, Q는 요인 적재량(Factor loading)을 의미한다. SVD를 통해 산출된 요인 적재량 Q가 X와 결합하면서 요인점수 Λ가 산출되고, 이는 축약된 각 성분 PC_i로 의미한다. PC_i는 새로운 예측 변수 X_{pc_i}로 나타내며, Quantile regression에 적용하여 H-FQMA를 통한 주별 가락도매시장 토마토 가격 예측치가 산출된다.

$$\hat{P}_{(q)} = f\left(X_{pc_i}; \hat{\theta}_{(q)}\right)$$
$$\hat{\theta}_{(q)} = \arg\min_{\theta_{(q)}} \sum_{n=1}^{N} \begin{Bmatrix} (q-1) \cdot (P_n - f(x_n;\theta_{(q)})) \text{ if } P_n < f\left(X_{pc_i};\hat{\theta}_{(q)}\right) \\ q \cdot (P_n - f(x_n;\theta_{(q)})) \end{Bmatrix} \quad (7)$$

식 (7)를 통해 추정된 모수 $\hat{\theta}_{(q)}$에 의해 주별 가락도매시장 토마토 가격 예측치가 결정된다. Quantile q는 $q \in (0,1)$로 정의되며, $\hat{P}_{(q)}$은 예측 변수 X_{pc}에 대한 y의 조건부 누적분포함수 $F^{-1}(q \mid X_{pc})$를 의미하며, H-FQMA로 예측된 값은 QMA, OMA, 단순평균모델 추정값과 비교하여 정확도를 분석하였다.

제3절. 분석 자료

연구에서 활용된 전체 데이터 기간은 2015년 13주차에서 2021년

13주차까지며, 여기서 2015년 1주차에서 2018년 10주차까지는 학습 데이터(Training set), 2018년 11주차에서 2019년 20주차를 검증 데이터(Validation set), 2019년 21주차에서 2021년 13주차까지를 테스트 데이터(Test set)로 구분하였다. 가락도매시장을 포함한 주요도매시장 가격은 kg당 가격(원) 즉, 단위가격으로 환산하였으며, 반입량도 kg 단위로 환산하였다. 가격 및 반입량 외에도, 석유 가격 예측 모델 연구를 진행한 Ye et al.(2005)의 연구를 참고하여 가격 및 반입량 평균과의 편차를 나타내는 상대적 변화(Relative change) 변수를 생성해 예측 변수 후보군으로 선정하였다. 각 기본 변수를 기반으로 시차를 1~10(1주~10주 전)까지 설정하였다. 그림 2.는 주별 가락도매시장 토마토 단위 가격 추이를 나타내고 있다. 2015년에서 2021년 13주차까지의 토마토 가격은 일정 수준을 유지하고 있는 것으로 보이며, 2020년 겨울에 급격한 가격 상승이 있었는데, 실제 큰 폭의 출하량 감소가 발생했었다(KREI, 2020).

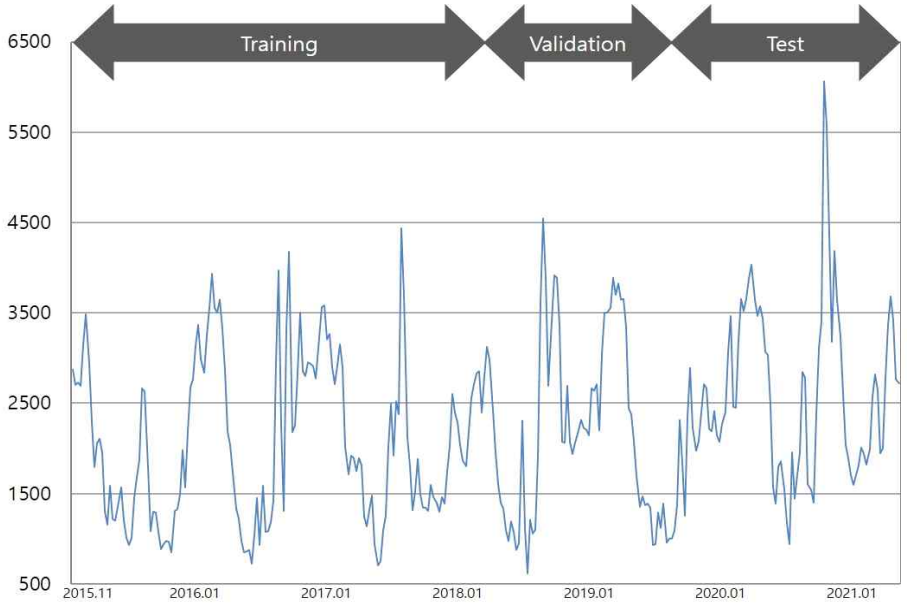

그림 2. 주별 가락도매시장 토마토 가격 추이

국내 도매시장 토마토 반입량에서 가락도매시장, 대구북부도매시장, 부산반여도매시장, 부산엄궁도매시장이 차지하는 비율이 2020년 기준 약 75% 수준이다. 따라서 상기 네 개 도매시장을 국내 주요 토마토 도매시장으로 간주하고, 이를 중심으로 예측 변수 집합을 생성하였다. 표 4.는 주요 도매시장 가격 및 반입량을 이용해 생성한 변수 집합을 나타내고 있다.

표 4. 예측 변수 후보군 개요

변수	설명
$week$	해당 연도 주차
P_{garak_t}	t주차 가락도매시장 평균 가격(kg당, 원)
P_{dae_t}	t주차 대구북부도매시장 평균 가격(kg당, 원)
P_{buban_t}	t주차 부산반여도매시장 평균 가격(kg당, 원)
P_{buum_t}	t주차 부산엄궁도매시장 평균 가격(kg당, 원)
Q_{garak_t}	t주차 가락도매시장 반입량(kg)
Q_{dae_t}	t주차 대구북부도매시장 반입량(kg)
Q_{buban_t}	t주차 부산반여도매시장 반입량(kg)
Q_{buum_t}	t주차 부산엄궁도매시장 반입량(kg)
Q_{tratio_t}	t주차 전체 도매시장에서 가락도매시장 반입량 비율
P_{garakr_t}	t주차 가락도매시장 가격 상대적 변화량(kg당, 원)
Q_{garakr_t}	t주차 가락도매시장 반입량 상대적 변화량(kg)
$P_{garak_{t-1:t-10}}$	$t-1$~$t-10$주차 가락도매시장 평균 가격(kg당, 원)
$P_{dae_{t-1:t-10}}$	$t-1$~$t-10$주차 대구북부도매시장 평균 가격(kg당, 원)
$P_{buban_{t-1:t-10}}$	$t-1$~$t-10$주차 부산반여도매시장 평균 가격(kg당, 원)
$P_{buum_{t-1:t-10}}$	$t-1$~$t-10$주차 부산엄궁도매시장 평균 가격(kg당, 원)

$Q_{garak_{t-1:t-10}}$	$t-1$~$t-10$주차 가락도매시장 반입량(kg)
$Q_{dae_{t-1:t-10}}$	$t-1$~$t-10$주차 대구북부도매시장 반입량(kg)
$Q_{buban_{t-1:t-10}}$	$t-1$~$t-10$주차 부산반여도매시장 반입량(kg)
$Q_{buum_{t-1:t-10}}$	$t-1$~$t-10$주차 부산엄궁도매시장 반입량(kg)
$Q_{tratio_{t-1:t-10}}$	$t-1$~$t-10$주차 전체 도매시장에서 가락도매시장 반입량 비율
$P_{garakr_{t-1:t-10}}$	$t-1$~$t-10$주차 가락도매시장 가격 상대적 변화량(kg당, 원)
$Q_{garakr_{t-1:t-10}}$	$t-1$~$t-10$주차 가락도매시장 반입량 상대적 변화량(kg)

제4절. 분석 결과

4.1. 변수 선정(Feature selection) 결과

BMA 활용에서는 계산 편의성이 높은 Zellner(1986)의 g-prior 분포를 사전 분포로 적용하였으며, 100,000회 반복을 통해 PIP(Posterior inclusion probability)가 높은 30개의 주요 변수를 선정하는 방식으로 진행하였다. PIP가 높을수록 목표 변수에 대한 영향력이 높음을 의미하며, PIP가 50 이상인 변수를 최종 예측 변수로 선정하였다(Forte et al., 2018). RMCFS를 활용한 변수 선정에서는, 개별 모델에서 150번 이상 중요한 변수로 포함된 변수만을 추출하도록 설정하였으며, RI(Relative importance, 변수중요도 점수)가 1.0 이상인 변수를 최종 예측 변수로 선정하였다(Draminski & Koronacki, 2018).

BMA를 활용한 변수 선정 결과에서는, 1주 전 가락도매시장 가격($P_{garak_{t-1}}$), 2주 전 부산엄궁도매시장 가격($P_{buum_{t-2}}$), 2주 전 가락도매

시장 가격($P_{garak_{t-2}}$), 3주 전 가락도매시장 가격($P_{garak_{t-3}}$), 9주 전 전체 도매시장에서 가락도매시장 반입량이 차지하는 비율($Q_{tratio_{t-9}}$), 9주 전 부산엄궁도매시장 물량($Q_{buum_{t-9}}$), 8주 전 전체 도매시장에서 가락도매시장 반입량이 차지하는 비율($Q_{tratio_{t-8}}$), 1주 전 가락도매시장 상대적 가격 변화($P_{garakr_{t-1}}$), 10주 전 가락도매시장 가격($P_{garak_{t-10}}$)이 선정되었다. BMA에 의해 선정된 주요 예측 변수 결과는 표 5에 나타나 있다.

표 5. BMA를 통해 선정된 주요 예측 변수 결과

변수	내용	PIP	부호
$P_{garak_{t-1}}$	$t-1$주차 가락도매시장 평균 가격(kg당, 원)	100	(+)
$P_{buum_{t-1}}$	$t-1$주차 부산엄궁도매시장 평균 가격(kg당, 원)	100	(+)
$P_{garak_{t-2}}$	$t-2$주차 가락도매시장 평균 가격(kg당, 원)	100	(−)
$P_{garak_{t-3}}$	$t-3$주차 가락도매시장 평균 가격(kg당, 원)	100	(+)
$Q_{tratio_{t-9}}$	$t-9$주차 전체 도매시장에서 가락도매시장 반입량 비율	87	(+)
$Q_{buum_{t-9}}$	$t-9$주차 부산엄궁도매시장 반입량(kg)	78	(−)
$Q_{tratio_{t-8}}$	$t-8$주차 전체 도매시장에서 가락도매시장 반입량 비율	71	(+)
$P_{garakr_{t-1}}$	$t-1$주차 가락도매시장 가격 상대적 변화량(kg당, 원)	59	(−)
$P_{garak_{t-10}}$	$t-10$주차 가락도매시장 평균 가격(kg당, 원)	53	(−)

RMCFS를 활용한 변수 선정 결과에서는, 1주 전 가락도매시장 가격($P_{garak_{t-1}}$), 1주 전 부산엄궁도매시장 반입량($Q_{buum_{t-1}}$), 1주 전 부산반여도매시장 반입량($Q_{buban_{t-1}}$), 2주 전 부산엄궁도매시장 반입량($Q_{buum_{t-2}}$), 2주 전 가락도매시장 가격($P_{garak_{t-2}}$), 5주 전 대구북부도매시장 반입량($Q_{dae_{t-5}}$), 6주 전 대구북부도매시장 반입량($Q_{dae_{t-6}}$), 4주 전 대구북부도매시장 반입량($Q_{dae_{t-4}}$), 2주 전 부산반여도매시장 반입량($Q_{buban_{t-2}}$)이 선정되었다. RMCFS를 통해 선정된 주요 예측 변수 정보는 표 6.에서 정리하였다.

표 6. RMCFS를 통해 선정된 주요 예측 변수 결과

변수	설명	RI
$P_{garak_{t-1}}$	$t-1$주차 가락도매시장 평균 가격(kg당, 원)	2.94
$Q_{buum_{t-1}}$	$t-1$주차 부산엄궁도매시장 반입량(kg)	2.17
$Q_{buban_{t-1}}$	$t-1$주차 부산반여도매시장 반입량(kg)	1.52
$Q_{buum_{t-2}}$	$t-2$주차 부산엄궁도매시장 반입량(kg)	1.47
$P_{garak_{t-2}}$	$t-2$주차 가락도매시장 평균 가격(kg당, 원)	1.42
$Q_{dae_{t-5}}$	$t-5$주차 대구북부도매시장 반입량(kg)	1.26
$Q_{dae_{t-6}}$	$t-6$주차 대구북부도매시장 반입량(kg)	1.10
$Q_{dae_{t-4}}$	$t-4$주차 대구북부도매시장 반입량(kg)	1.02
$Q_{buban_{t-2}}$	$t-2$주차 부산반여도매시장 반입량(kg)	1.01

4.2. 예측모델 평가(Forecasting evaluation)

본 연구에서 제안하는 H-FQMA의 예측 성능 평가를 위해 세 개의 예측 조합 모델을 벤치마킹 모델로 선정하였다. 따라서

H-FQMA를 포함한 총 4개 모델의 예측 성능이 비교되었으며, 4개 모델은 다시 변수 선정법(BMA, RMCFS, Now)에 따라 구분되어 총 16개 모델의 예측 성능이 비교되었다.

표 7.에서 H-FQMA를 포함한 16개 모델의 개요를 나타내고 있다. H-FQMA는 본 연구에서 제안하는 모델, Mean model은 개별 방법에서 발생한 예측치를 평균 처리한 것, OMA는 Model stacking과 OLS 방법을 혼합한 방식, QMA는 Model stacking과 Quantile regression을 혼합한 방식이다. 세부 분류에서 다시 AR은 단변량 시계열 방법만을 활용해 예측치를 생성한 예측 조합 모델, BM은 시계열 방법 외 BMA를 변수 선정에 활용한 예측 조합 모델, RF는 시계열 방법 외 R0MCFS를 변수 선정에 활용한 예측 조합 모델, NOW는 t−n 시차 변수를 제외한 t 시점 예측 변수만을 활용한 예측 조합 모델을 의미한다.

표 7. H-FQMA를 포함한 16개 예측 조합 모델 개요

모델		변수 선정	내용
H-FQMA	AR	-	단변량 시계열 + PCA + QRM
	BM	BMA	일반회귀/DM & ML(변수 선정: BMA) + PCA + QRM
	RF	RMCFS	일반회귀/DM & ML(변수 선정: RMCFS) + PCA + QRM
	NOW	-	일반회귀/DM & ML(예측 변수: NOW) + PCA + QRM
Mean model			단변량 시계열 + 평균 처리
	BM	BMA	일반회귀/DM & ML(변수 선정: BMA) + 평균 처리
	RF	RMCFS	일반회귀/DM & ML(변수 선정: RMCFS) + 평균 처리
	NOW	-	일반회귀/DM & ML(예측 변수: NOW) + 평균 처리

OMA	AR	-	단변량 시계열 + OLS
	BM	BMA	일반회귀/DM & ML(변수 선정: BMA) + OLS
	RF	RMCFS	일반회귀/DM & ML(변수 선정: RMCFS) + OLS
	NOW	-	일반회귀/DM & ML(변수 선정: NOW) + OLS
QMA	AR	-	단변량 시계열 + QRM
	BM	BMA	일반회귀/DM & ML(변수 선정: BMA) + QRM
	RF	RMCFS	일반회귀/DM & ML(변수 선정: RMCFS) + QRM
	NOW	-	일반회귀/DM & ML(변수 선정: NOW) + QRM

QRM: Quantile regression model
PCA: Principal component analysis
DM & ML: Data mining & machine learning

모델별 예측 성능 평가 지표로 RMSE(Root mean square deviation), MAPE(Mean absolute percentage error), R-squared, MZ 검증(Mincer-Zarnowitz test), SIS(Step indicator saturation) 등 다섯 가지 지표를 활용하였다. MZ 검증는 실제값에 대한 예측치의 편의(Bias) 정도를 통계적으로 추정하는 방법으로, $\sigma_{t+1} = \beta_0 + \beta_1 \hat{\sigma}_{t+1}$ 모델에서 $\beta_0 = 0, \beta_1 = 1$에 대한 가설 검증을 수행한다. 영가설이 기각되지 않으면 예측치는 실제값에 대한 편의가 없는 것으로 판단할 수 있다(Ahumada & Cornejo, 2016). SIS는 시계열 모델에서 구조변화(Structural break) 및 그에 따른 매개변수(parameter) 변화를 추정할 때 사용하는 방법으로, 예측 모델에서는 이상치에 대한 모델의 대응 정도를 평가할 수 있는 지표이다(Pretis et al., 2021). 예측 모델 연구에서의 활용 방법은 예측치를 기반으로 시계열 모델을 생

성한 후, 모델에서 선정된 이상치 및 실제 값의 이상치를 비교하는 방식이다. 본 연구에서는 실제값의 이상치 선정 기준으로 체비체프 부등식(Chebyshev inequality)을 활용했으며, 이에 따라 전일 대비 등락률이 41.92 이상 또는 -34.92 이하이면 이상치로 선정하였다.

표 8.은 본 연구에서 제안한 H-FQMA를 포함한 16개 모델의 RMSE, MAPE, R-Square 수치를 보여주고 있다. H-FQMA의 예측 성능이 다른 벤치마킹 모델 대비 우수한 것으로 나타났다. H-FQMA에서 AR 모델의 RMSE, MAPE, R-squared는 각각 429(Won), 13.2%, 0.814로 산출되었다. BMA 변수 선정을 활용한 BM 모델의 RMSE, MAPE, R-squared는 각각 356(Won), 12.5%, 0.872로 나타났다. RMCFS 변수 선정을 활용한 RF 모델의 RMSE, MAPE, R-squared는 각각 325(Won), 11.1%, 0.917로 16개 모델 중 가장 좋은 예측 성능을 보였다. t주차 예측 변수만을 활용한 NOW 모델의 RMSE, MAPE, R-squared는 각각 422(Won), 1389%, 0.818을 기록하였다.

단순 평균을 취한 Mean model은 다른 모델 대비 가장 낮은 예측 성능을 기록하였다. Mean model AR 모델의 RMSE, MAPE, R-squared는 각각 495(Won), 15.3%, 0.753, BM 모델은 531(Won), 16.2%, 0.742, RF 모델은 494(Won), 15.0%, 0.789, Now 모델은 545(Won), 16.5%, 0.761로 산출되었다. OMA AR 모델의 RMSE, MAPE, R-squared는 각각 431(Won), 13.6%, 0.809로 산출되었다. BM 모델의 RMSE, MAPE, R-squared는 각각 441(Won), 14.2%, 0.802, RF 모델은 423(Won), 13.6%, 0.817로, NOW 모델은 465(Won), 14.5%, 0.779을 기록하였다. QMA는 H-FQMA 다음으로 좋은 예측 성능을 기록하였다. QMA의 AR 모델 RMSE, MAPE, R-squared은 각각 448(Won), 12.9%, 0.796로 산출되었다. 특히 RF 모델에서 RMSE, MAPE, R-squared가 각각 392(Won), 12.66%, 0.845로 가장 좋은 성능을 기록하였다.

표 8. 16개 모델의 예측 성능

모델		RMSE(원)	MAPE(%)	R-Squared
H-FQMA	AR	429	13.2	0.814
	BM	356	12.5	0.872
	RF	325	11.1	0.917
	NOW	422	13.8	0.818
평균		**383.0**	**12.7**	**0.855**
Mean model	AR	495	15.3	0.753
	BM	531	16.2	0.742
	RF	494	15.0	0.789
	NOW	545	16.5	0.761
평균		**516.3**	**15.8**	**0.761**
OMA	AR	431	13.6	0.809
	BM	441	14.2	0.802
	RF	423	13.6	0.817
	NOW	465	14.5	0.779
평균		**455.3**	**14.3**	**0.802**
QMA	AR	448	13.2	0.796
	BM	445	13.1	0.799
	RF	392	12.6	0.845
	NOW	435	13.7	0.807
평균		**435.0**	**13.2**	**0.812**

그림 3. ~ 5.는 모델별 예측 성능을 그래프로 비교한 내용이다.

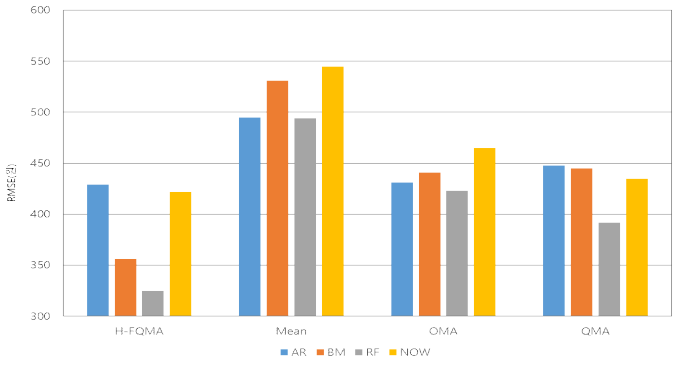

그림 3. 16개 모델의 예측 성능 도표(RMSE 기준)

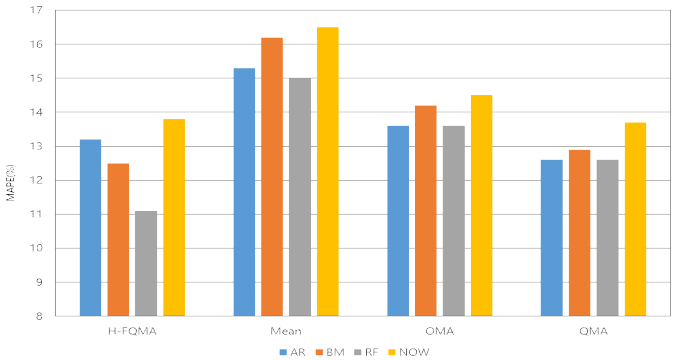

그림 4. 16개 모델의 예측 성능 도표(MAPE 기준)

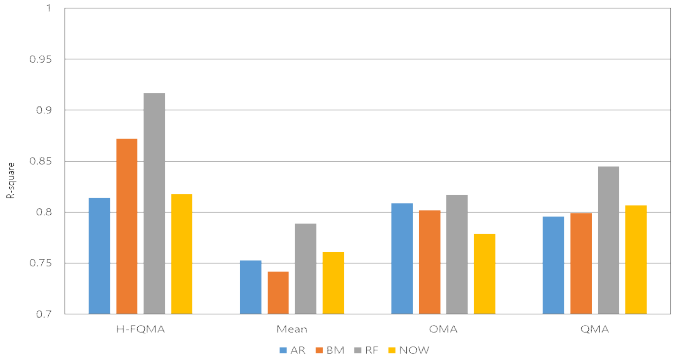

그림 5. 16개 모델의 예측 성능 도표(R-squared 기준)

표 9.는 MZ(Mincer-Zarnowitz) 검증과 예측값의 이상치 대응력을 평가하는 SIS 검증에 대한 결과를 보여주고 있다. MZ 검증에서 영가설이 기각되지 못하면, 실제값과 예측값 간 편의가 없는 것으로 해석할 수 있다. Mean model의 BM, RF, Now 외 영가설이 기각된 모델은 없다. SIS 검증을 위해서는 테스트 자료 기반으로 이상치 즉, 급등 및 급락으로 판단할 수 있는 기간이 선정되어야 한다. 급등 및 급락 기준은 체비체프 부등식을 활용하였으며, 테스트 자료에서 토마토 가격의 급등락으로 판정된 기간은 2019년 36주차, 2019년 39주차, 2020년 19주차, 2020년 26주차, 2020년 30주차, 2020년 32주

차, 2020년 35주차, 2020년 38주차로 나타났다. H-FQMA의 RF 모델에서는 총 8개의 급등락 기간 중 3개(2019년 36주차, 2020년 30주차, 2020년 38주차) 기간이 예측되었다. Mean model BM 모델 및 QMA RF 모델에서 각각 2개 기간(2020년 30주차, 2020년 38주차, 2019년 39주차, 2020년 19주차)을 예측하였다. 상기 3개 모델 외 나머지 모델의 가격 급등락 대응력 수준은 낮은 것으로 평가되었다.

표 9. MZ 및 SIS 검증 결과

모델		검증 방법	
		MZ 검증	SIS(1%)
H-FQMA	AR	0.712	S2019.40, S2020.07, S2020.18, **S2020.38**, S2020.42
	BM	0.854	S2019.40, S2019.47, S2020.18, **S2020.38**, S2020.42
	RF	0.932	**S2019.36**, **S2020.30**, S2020.33, **S2020.38**, S2020.42
	Now	0.760	S2019.37, S2020.03, S2020.17, S2020.37, S2020.42
Mean model	AR	0.490	S2019.40, S2020.07, S2020.18, S2020.39, S2020.42
	BM	0.004	S2019.37, S2019.47, S2020.20, **S2020.30**, **S2020.38**, S2020.42
	RF	0.000	S2019.40, S2019.47, S2020.20, S2020.36, S2020.41
	Now	0.000	S2019.37, S2020.20, S2020.37, S2020.42
OMA	AR	0.578	S2019.40, S2020.03, **S2020.19**, S2020.36, S2020.39, S2020.41
	BM	0.658	S2019.40, S2020.03, S2020.20, S2020.37, S2020.46, S2021.04
	RF	0.754	S2019.40, S2020.20, S2020.36, S2020.39, S2020.41
	Now	0.655	S2019.40, S2020.20, S2020.36, S2020.39, S2020.41
QMA	AR	0.722	S2019.40, S2020.03, S2020.18, S2020.20, S2020.36, S2020.39, S2020.41
	BM	0.756	S2019.40, S2020.03, S2020.20, **S2020.35**, S2020.39, S2020.42
	RF	0.790	**S2019.39**, **S2020.19**, S2020.36, S2020.41
	Now	0.640	S2019.40, S2020.07, S2020.17, S2020.20, S2020.37, S2020.42

제3장. 베이지안모델평균법과 PSO를 활용한 효율적인 농산물 출하 전략 수립을 위한 의사결정지원도구 개발 연구

제1절. 연구의 배경

본 연구의 목적은 토마토 관련 농산물 유통 조직의 효율적인 시장 출하 전략 수립을 위한 의사결정지원도구(Decision-making support tool)를 개발하는 것이다. 효율적인 시장 출하 의사결정은 농산물 유통 조직 이익에 영향을 미치는 가장 중요한 요인이다(김동환, 2010). 특히, 도매 및 소매 조직 모두에 대응하는 경우, 전략적인 물량 배분은 필수 역량 중 하나다.

국내 농산물 소매 가격은 도매 가격에 영향을 받는다(강태훈, 2009; 양한탁 외, 2018; 김기환 & 서병선, 2017). 농산물 유통 조직이 대형 마트(이마트, 홈플러스 등)와 같은 소매 조직에 물량을 공급할 시, 공급 가격은 가락도매시장[1] 가격을 참조해 체결된다. 소매 조직과는 주 단위 계약이 체결되는데, 해당 시점 가락도매시장 가격을 기준으로 계약이 결정되는 것으로 분석되었다. 즉, 소매 조직과

1) 국내 농산물 도매시장은 총 72개가 있는데, 이 중 토마토는 42개 시장에서 거래되고 있다. 이 중 물량이 집중되는 곳은 가락도매시장, 대구북부도매시장, 부산엄궁도매시장, 부산반여도매시장 등 4개 도매시장이 있으며, 이 중에서도 가락도매시장에서 가장 많은 물량이 거래되고 있다.

거래하는 농산물 유통 조직은 가락도매시장 가격이 높을수록 높은 이익을 얻게 되는 것이다.

도매시장 현황 및 조직 내부 물량 정보를 토대로 가락도매시장 가격 움직임을 추정한 후, 이를 통해 가락도매시장 가격 변동성에 대한 시뮬레이션(Simulation)이 가능한 도구가 개발된다면 농산물 유통 조직의 출하 의사결정 효율성을 제고할 수 있을 것이다. 해당 도구를 통해 농산물 유통 조직은 도소매 출하 물량을 적절하게 조절해 도매시장 가격 상승을 유도하고, 이를 통해 소매 거래 가격을 높이는 전략을 취할 수 있다.

의사결정지원도구 개발을 위해서는 두 가지가 필요한데 먼저 가락도매시장 및 주요 도매시장(대구북부도매시장, 부산엄궁도매시장, 부산반여도매시장 등) 가격 및 반입량과 농산물 유통 조직의 출하 정보를 이용해 가락도매시장 토마토 가격을 추정할 수 있는 모델 개발이 필요하다. 다음으로 가락도매시장 가격 추정 모델을 기반으로 가락도매시장 가격이 높아지는 다수의 시나리오를 선별할 수 있는 모델 개발이 필요하다.

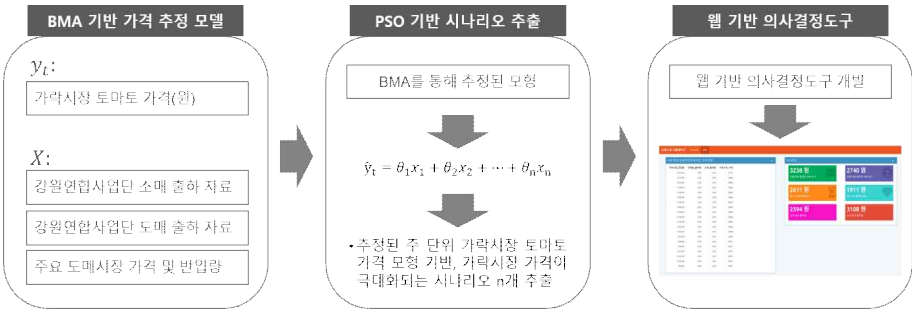

그림 1. 연구의 진행 절차

본 연구에서는 계약 농가 물량에 대한 주 단위 시장 출하 전략을 수립하는 농협 내부 조직인 A 연합사업단의 출하 정보와 주요 도매시장 정보를 활용해 주별 가락도매시장 토마토 가격 변동성을 시뮬

레이션할 수 있는 출하 의사결정지원도구를 개발하였다. 주별 가락도매시장 가격 추정 모델은 BMA(Bayesian model averaging)를 활용하였다. BMA는 반응 변수에 대해 다수의 설명 변수가 존재하는 경우, 즉 탐색적 연구에서 유용한 방법이다(Forte et al., 2018). 다음으로 BMA 결과를 기반으로 PSO(Particle swarm optimization)를 적용해 가락도매시장 가격이 높아지는 다수의 시나리오를 선별할 수 있는 모델을 개발하였다. PSO는 비선형 최적화 방법 중 하나로, 정의역 공간(Domain space, 본 연구에서는 A 연합사업단의 도소매 출하량, t 주차 가락도매시장 반입량 등)이 설정되면, 목적함수(Objective function, 본 연구에서는 BMA로 추정된 설명 변수의 추정량)를 최대 및 최소화하는 범위를 찾아내는데 유용한 방법이다(Zhu et al., 2020).

연구 시사점으로는 첫 째, 국내 처음으로 시도되는 농산물 유통 조직의 출하 의사결정지원도구 개발 연구라는 학문적 시사점을 꼽을 수 있다. 기존 국내 연구를 살펴보면 농산물 유통 조직의 출하 의사결정지원도구에 관한 연구는 찾기 힘든 상황이다. 해외 연구에서는 재고 및 수요 관리 측면에서 관련 연구가 진행되고 있으며, 풍부한 자료를 기반으로 System dynamics와 Agent based model이 주로 적용되고 있다(Behzadi et al., 2018). 하지만 대부분 연구가 재고와 수요 측면에 집중된 경향이 있으며, 시장 상황을 고려한 농산물 유통 조직의 출하 의사결정을 지원하는 연구는 찾기 힘든 상황이다. 둘 째, 본 연구의 실무적 시사점으로, 개발된 출하 의사결정지원도구를 활용해 관련 농산물 유통 조직은 내부 출하 전략에 따른 가락도매시장 토마토 가격 변동성을 확인할 수 있다. 농협과 같은 대형 농산물 유통 조직에서는 여전히 출하 담당자 직관에 의존한 의사결정이 이뤄지고 있다. 담당자 직관과 의사결정지원도구가 조합된다면 효율적인 출하의사결정이 가능할 것이다.

관련 연구에 대한 선행연구를 살펴보면, 농산물 유통 관련 의사결정지원도구 개발 연구는 시뮬레이션 연구가 주를 이루고 있으며,

System dynamics, Agent based model, LP(Linear programming) 등의 방법이 주로 활용되고 있다(Teimoury et al., 2013). System dynamics를 활용한 의사결정지원도구 개발 연구의 경우 농산물 재고 관리, 농산물 수요 예측에 따른 물량 배분에 관한 시뮬레이션 도구 개발이 대부분이다(Tako & Robinson, 2012). Kumar & Nigmatullin (2011)는 독점 환경에서 저장성이 높은 작물에 대한 시장 수요 및 공급 변동성이 농산물 공급망(Supply chain)에 미치는 영향을 System dynamics를 활용해 평가하였다. Ozbayrak et al.(2007)은 농산물 공급망 네트워크(Supply chain network)를 시뮬레이션하기 위해 System dynamics를 활용한 방법을 제안하였다. 연구를 통해 WIP(Work In Progress) 수준, 즉 주문 처리 수준, 고객 만족도 등 공급망 관련 성과 지표를 측정하였다. Teimoury et al.(2013)는 이란 내 채소 수입량을 조절하기 위한 시뮬레이션 모델을 제안하였다. 이를 위해 먼저 이란 농산물 유통 구조를 정의하였고, 경로 내 각 주체(agent)를 함수 형태로 정의하였다. 주체별 함수적 정의는 기존 연구로부터 알려진 부분과 System dynamics를 통한 교정(Calibration) 작업을 혼용해 진행하였다. Ferreira et al.(2016)은 오렌지 생산 단계에서 발생하는 박스, 비닐 등의 자재 비용과 농산물 유통 과정에서의 보관 및 운송 비용 요소를 정의한 후, 생산에서 소비지까지의 비용 발생에 관한 시뮬레이션 모델을 연구하였다.

Agent based model을 활용한 연구로는, Ge et al.(2015)는 밀, 옥수수 등 캐나다 곡물 시장 유통 과정에서 곡물 등급에 대한 잘못된 표기가 시장에 미치는 영향을 시뮬레이션하였다. 그들은 Agent based model을 활용해 시장에서 발생하는 전체 비용을 시뮬레이션할 수 있는 모델을 연구하였다. Ge et al.(2015)은 식량 안전 분야에서 일반적인 최적화 모델과 Agent based model 간 비교 분석하였고, 상황을 고려한 두 모델의 선별 적용의 필요성을 강조하였다.

LP를 활용한 연구로는, Flores et al.(2019)는 멕시코 농산물 시장

에서의 소규모 농가를 위한 최적화 모델을 연구하였다. 그들은 LP를 활용해 최적의 파종, 수확, 시장 출하 시기를 선정해 농가의 이익을 극대화(Maximized)하는 모델을 연구하였다. 이 외에도 다수 소매 조직과의 공급 효율성 제고를 위해 GA(Genetic algorithm)과 같은 비선형 최적화 방법을 활용한 시뮬레이션 도구 개발 연구도 존재한다(Wang et al., 2012).

농산물 유통 관련 의사결정지원도구 개발 연구 대부분이 재고 관리, 수요량에 대비한 공급량 조절에 초점이 맞춰져 있다. 국내 연구에서는 최적 경로 또는 물류 단지 내 동선 최적화에 대한 연구가 전부인 상황이다(서교 외, 2000, 박명규, 2015). 또한, 실제 연구된 모델을 실무자가 활용할 수 있도록 의사결정지원도구를 웹(Web) 기반 정보시스템 형태로 개발한 연구가 미흡한 상황이다. 따라서, 본 연구에서는 시장에 영향을 미치는 물량을 보유한 대형 농산물 유통 조직의 내부 출하 및 주요 도매시장 정보를 활용해 가락도매시장 토마토 가격 변동성을 확인할 수 있는 의사결정지원도구를 개발하였다. 또한, 이를 웹에서 실무자가 활용 가능하도록 조치하였다.

제2절. 국내 농산물 유통 경로 및 A 연합사업단

2.1. 국내 농산물 유통 경로

먼저 Teimoury et al.(2013)의 연구에서 정의된 농산물 유통 경로를 일부 수정해 국내 농산물 유통 경로를 정의하고자 한다(그림 2.). 그림 2.에서 생산자(Producers)를 살펴보면, 기상 상황 등 환경 요인

을 고려한 생육 조절을 통해 원물을 생산하는 주체를 뜻하며, 농가, 영농법인, 기업형 생산 조직 등이 포함된다. 도매 및 소매 단계만을 거쳤던 과거와는 달리, 최근에는 생산자가 최종 소비자에게 직접 농산물을 판매하는 비율이 증가하고 있다(김연중 외, 2010). 도매상(Wholesalers)은 최종 소비자가 아닌, 소매상, 식자재 업체, 식당 등 상업적 주체에게 농산물을 판매하는 상인을 의미한다. 소매상(Retailers)은 생산자, 도매상, 기타 중간 업체로부터 농산물을 매입해 최종 소비자에게 판매하는 주체를 뜻한다. 대형마트, 일반 청과마트가 여기에 해당된다. 소비자는(Consumers)는 생산자 및 소매상으로부터 농산물을 구입하는 최종 소비자로, 실질적인 농산물 수요에 영향을 미치는 주체다. 중간 업체(Middlemen)는 생산자로부터 원물을 수급하는 특수한 형태로, 계약농가로부터 원물을 수급한 후 선별 및 저장 과정을 통해 시장에 유통하는 산지유통센터(APC, Agricultural products center), 계약 농장의 원물을 직접 수거하는 밭떼기 업체, 필요한 만큼의 원물을 생산자로부터 수급한 후 거리에서 판매하는 포전거래 업체 등이 있다(전창곤 외, 2011). 이 외에도 식당, 식자재 업체 등의 식품 산업군(Food industries)이 존재하며, 외부적으로는 수출과 수입 관련 주체가 있다.

본 연구에서는 가장 많은 양의 농산물이 거래되고, 농산물 유통 분야에서 핵심 경로로 인지되는 도매와 소매를 중심으로 연구를 진행하였다. 여기서 A 연합사업단의 경우, 농가 및 중간 업체로부터 확보한 대규모 물량에 대한 시장 출하 전략을 수립하는 조직으로 중간 업체를 아우르는 조직으로 정의될 수 있다.

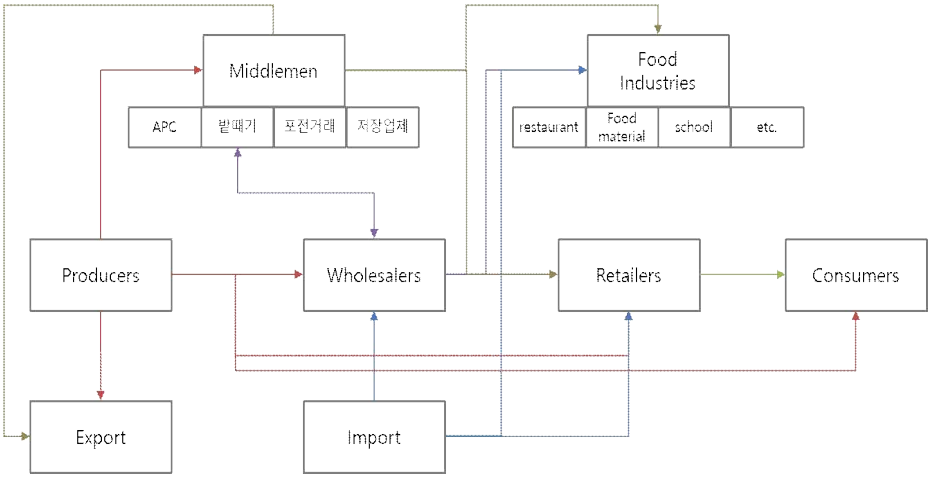

그림 2. 국내 농산물 유통 구조

2.2. A 연합사업단

A 연합사업단은 2001년부터 강원도 15개 시군과 33개 지역 농협이 참여해 설립한 농협 소속 국내 농산물 유통 조직이다. A 연합사업단은 농업인의 소득 증대와 소비지에서의 안정적인 농산물 공급을 위해 풋고추, 피망, 무, 배추를 연합 판매(이하 연합마케팅)하는 방식으로 사업을 시작하였다.

연합마케팅[2)]은 산지조직과 유통업체 및 물류센터, 도매시장 사이에서 사업단이 연결고리가 되어 품질 관리, 물류 관리, 통합브랜드 및 홍보, 가격 조정 역할을 수행하는 시스템이다. 즉, 연합마케팅은 농산물 유통 전반에 포함되는 사업으로 산지에 해당하는 '생산자'와 유통에 해당하는 '농산물산지유통센터(APC)', '소매업체' 및 '식자재',

2) 하절기(5~12월) 강원도 내 고랭지채소를 중심으로 상품화된 농산물을 기존 개별 농협 단위 단품 공급방식에서 벗어나 사업단에서 유통업체를 대상으로 일괄 납품해 유통업체의 희망 품목을 적정 시기에 납품하는 사업이다. 사업단이 농협에서 출하권을 위임받아 납품 계약, 수발주, 협상 및 정산까지 진행하는 거래 주체 역할을 수행하는 사업이다.

‘도매업체’의 주체가 포함된 농산물 유통 사업이다.

그림 3. A 연합사업단의 설립 목표

연합마케팅 사업의 주된 목적은 성출하기 과잉 생산에 의한 가격 폭락에 (산지폐기 및 분산출하 등으로) 탄력적으로 대응하기 위함이며, 농산물 유통 환경 변화에 대응해 대형유통업체 등으로 직거래를 활성화하기 위함이다.

그림 4. A 연합사업단 매출 추이(2001 ~ 2018년)

이 외에도 산지 농산물의 규모화, 규격화, 고품질화로 거래 교섭력 강화와 더불어, 강원도 내 우수 농산물의 안정적인 판로 구축 및 개별 농협의 마케팅 활동 지원으로 거래 편리성 제공 등의 목적이 있다.

연합마케팅 사업 실적은 2001년 114억으로 출발해 매년 증가해 왔으며, 2018년에는 2001년 대비 약 16배 상승한 2,345억 원에 달했다. 품목별 취급 비중은 고추류 10%, 고랭지 무 6%, 배추 34%, 토마토 19%로, 이들 품목이 전체 실적의 68%를 차지한다.

제3절. 분석 자료 및 분석 모델

3.1. 분석 자료

본 연구에서는 국내 토마토에 대한 농산물 도매시장 및 A 연합사업단 시장 출하 자료를 활용해 연구를 진행하였다. 농산물 도매시장 자료는 국가공공데이터포털(https://www.data.go.kr/)에서 개방형 API(Open API) 방식으로 제공(JSON 및 XML 형태)하고 있다. A 연합사업단의 출하 자료는 조직 내 전사적자원관리(ERP) 시스템에서 추출한 것이다. 자료 기간은 2017년 ~ 2020년까지의 토마토 관련 자료만 추출해 활용하였으며, 추출한 도매시장 자료는 513,955개, A 연합사업단 자료는 230,672개이다. 본 연구에서는 토마토 가격은 kg 당 단위 가격(원)으로, 반입량 및 출하량과 같은 시장 거래 물량은 kg 단위로 변환하였다. 또한, 본 연구의 목적은 주별 단위 의사결정 지원도구 개발에 있으므로, 자료를 주별 데이터로 변환(Aggregation)하여 활용하였다.

변환된 자료를 토대로 생성한 변수 집합은 표 1.과 같다. 국내 도

매시장에서 가락도매시장, 대구북부도매시장, 부산엄궁도매시장, 부산반여도매시장이 차지하는 토마토 물량 비율은 약 75%(2020년 기준) 수준이다. 이 점을 고려해 네 개 도매시장을 주요 도매시장으로 선정하였으며, 이를 기준으로 파생 변수를 생성하였다. 상대적 변화(Relative change)는 Ye et al.(2005)의 연구를 참고하여 생성한 변수로, $Volume_t - \frac{1}{N}\sum_{i=1}^{N} Volume_i$를 해당 변수에 적용해 생성하였다.

표 1. 기본 변인을 활용한 파생 변수 정보

변수	설명
P_{garak_t}	t주차 가락도매시장 토마토 가격(kg당, 원)
Q_{garak_t}	t주차 가락도매시장 토마토 반입량(kg)
Q_{garakr_t}	t주차 가락도매시장 토마토 반입량 상대적 변화량(kg)
Q_{WO_t}	t주차 A 연합사업단 도매 출하량 레벨(1~5)
Q_{RE_t}	t주차 A 연합사업단 소매 출하량 레벨(1~5)
$P_{garak_{t-1:t-4}}$	$t-1$~$t-4$주차 가락도매시장 토마토 가격(kg당, 원)
$Q_{garak_{t-1:t-4}}$	$t-1$~$t-4$주차 가락도매시장 토마토 반입량(kg)
$Q_{dae_{t-1:t-4}}$	$t-1$~$t-4$주차 대구북부도매시장 토마토 반입량(kg)
$Q_{buum_{t-1:t-4}}$	$t-1$~$t-4$주차 부산엄궁도매시장 토마토 반입량(kg)
$Q_{buban_{t-1:t-4}}$	$t-1$~$t-4$주차 부산반여도매시장 토마토 반입량(kg)
$P_{garakr_{t-1:t-4}}$	$t-1$~$t-4$주차 가락도매시장 토마토 반입량 상대적 변화량(kg)
$Q_{daer_{t-1:t-4}}$	$t-1$~$t-4$주차 대구북부도매시장 토마토 반입량 상대적 변화량(kg)
$Q_{buumr_{t-1:t-4}}$	$t-1$~$t-4$주차 부산엄궁도매시장 토마토 반입량 상대적 변화량(kg)
$Q_{bubanr_{t-1:t-4}}$	$t-1$~$t-4$주차 부산반여도매시장 토마토 반입량 상대적 변화량(kg)
$Q_{WO_{t-1:t-4}}$	$t-1$~$t-4$주차 A 연합사업단 도매 출하량 레벨(1~5)
$Q_{RE_{t-1:t-4}}$	$t-1$~$t-4$주차 A 연합사업단 소매 출하량 레벨(1~5)

표 1.에 있는 파생 변수 집합을 활용한 주별 가락도매시장 토마토 가격 추정 모델은 식 (1)과 같다.

$$\hat{P}_t = f(X; \hat{\theta})$$
$$X = \begin{bmatrix} P_{garak,t-n}, Q_{garak,t:t-n}, Q_{RE,t:t-n}, Q_{WO,t:t-n}, \\ Q_{dae,t-n}, Q_{buban,t-n}, Q_{buum,t-n}, P_{garakr,t-n}, \\ Q_{garakr,t-n}, Q_{daer,t-n}, Q_{buumr,t-n}, Q_{bubanr,t-n} \end{bmatrix} \quad (1)$$

식 (1)에서 $\hat{P}_t$는 t 주차 가락도매시장 토마토 가격 추정치, $f(\bullet)$은 가락도매시장 추정 방법을 의미하는데, 여기서는 BMA를 의미한다. X는 모델에 투입되는 설명 변수, $\hat{\theta}$은 모델에서 추정된 설명 변수별 추정량을 뜻한다. X를 살펴보면, 가락도매시장 가격 $t-1$ ~ $t-4$ 시차 변수는 $P_{garak,t-1:t-n}$, 가락도매시장 반입량 t ~ $t-4$의 시차 변수는 $Q_{garak,t:t-n}$, A 연합사업단 소매 출하량 레벨 t ~ $t-4$의 시차 변수는 $Q_{RE,t:t-n}$, A 연합사업단 도매 출하량 레벨 t ~ $t-4$의 시차 변수는 $Q_{WO,t:t-n}$, 대구북부도매시장 반입량 $t-1$ ~ $t-4$의 시차 변수는 $Q_{dae,t-n}$, 부산반여도매시장 반입량 $t-1$ ~ $t-4$의 시차 변수는 $Q_{buban,t-n}$, 은 부산엄궁도매시장 반입량 $t-1$ ~ $t-4$의 시차 변수는 $Q_{buumi,t-n}$, 은 가락도매시장 반입량의 상대적 변화량에 대한 $t-1$ ~ $t-4$의 시차 변수는 $Q_{garakr,t-n}$, 은 대구북부도매시장 반입량의 상대적 변화량에 대한 $t-1$ ~ $t-4$의 시차 변수는 $Q_{daer,t-n}$, 부산엄궁도매시장 반입량의 상대적 변화량에 대한 $t-1$ ~ $t-4$의 시차 변수는 $Q_{buumr,t-n}$, 부산반여도매시장 반입량의 상대적 변화량에 대한 $t-1$ ~ $t-4$의 시차 변수는 $Q_{buban,t-n}$이다.

3.2. 분석 모델

3.2.1. BMA(Bayesian model averaging)

BMA에 관한 이론적 배경은 부록 B에 설명하였다.

3.2.2. PSO(Particle swarm optimization)

PSO는 Kennedy & Eberhart(1995)에 의해 개발되었으며, 새떼 또는 물고기떼와 같은 생물학적 군집 행동을 근원으로 하고 있다(강석호 & 김승, 2008). 즉, PSO는 생물학적 군집의 행동을 확률론에 근거해 비선형 최적화 방법으로 변형한 것으로(Zhu et al., 2020), 농업 분야 예측 및 최적화 연구에서도 좋은 성과가 입증되었다(García, et al., 2016; Yu et al., 2016). PSO는 GA(Genetic algorithm)의 진화 구조와 유사한데, PSO에서 개별해는 particle, 해집단은 swarm으로 나타낼 수 있으며, 세때와 유사하게 각 particle은 최적 해를 찾기 위해 정의역 공간을 떠돌며 그들과 주변 군집 경험 정보를 활용해 최적 위치로 이동해 간다(강석호 & 김승, 2008).

D개의 차원(Dimension)을 가진 정의역 공간(Domain space) 안에 n개의 particle을 $X=(X_1,X_2,\ldots,X_n)$ 집단을 가정해 보자. 집단 내 각 particle i마다 위치 및 속도를 각각 $X_i=(x_{i1},x_{i2},\ldots,x_{iD})^T$과 $V_i=(v_{i1},v_{i2},\ldots,v_{iD})^T$로 정의된다. particle i의 최적 위치, 속도에 대한 전체 모집단은 각각 $P_i=(p_{i1},p_{i2},\ldots,p_{iD})^T$, $P_g=(p_{g1},p_{g2},\ldots,p_{gD})^T$로 표현된다. 최적화 과정에서 particle의 위치와 속도가 개선(Update)되는데, 식 (2), (3)과 같이 표현된다.

$$X_{id}=X_{id}+X_{id}(d=1,2,\ldots,D,\quad i=1,2,\ldots,n) \tag{2}$$

$$V_{id}^{k+1}=\omega v_{id}^{k}+c_1 r_{1,i}^{k}\left(P_{id}^{k}-X_{id}^{k}\right)+c_2 r_{2,i}^{k}\left(P_{gd}^{k}-X_{id}^{k}\right) \tag{3}$$

식 (3)에서 ω는 관성 가중치(Inertial weight), k는 최근 반복 횟수, c_1과 c_2는 가속 계수(Acceleration coefficients, >0), $r_{1,i}^k$와 $r_{2,i}^k$는 0 ~ 1 구간의 무작위 수치(Random values)를 의미한다.

본 연구에서는 BMA로 선정된 주요 설명 변수 중, $t-n$ 시점 시차 변수가 아닌 t 시점의 설명 변수(A 연합사업단 소매 및 도매 출하량, 가락도매시장 반입량)의 범위가 정의역 공간으로 정의될 수 있다. 실제 PSO 적용 시에는 선정된 주요 설명 변수를 활용해 제작된 목적함수의 크기(주별 가락도매시장 토마토 가격)를 최적화하는 방향으로 진행된다.

제4절. 분석 결과

먼저 BMA를 활용한 분석 및 변수 선정 결과는 표 2.와 같다. BMA에서는 PIP(Posterior inclusion probability)를 기준으로 목표 변수에 대한 설명 변수 중요도를 평가한다. PIP의 수치 구간은 0 ~ 100으로, 50 이상의 수치를 가지면 주요 영향력 변수로 선정하였다(Forte et al., 2018). PM(Posterior mean)은 모수 θ_i에 대한 사후 평균을 의미한다.

표 2. BMA를 활용한 주별 가락시장 토마토 가격 추정 모델 결과

번호	변수	설명	PM	PIP
1	Q_{garak_t}	t주차 가락도매시장 토마토 반입량(kg)	-0.0007861	100.0
2	$P_{garak_{t-1}}$	$t-1$주차 가락도매시장 토마토 가격(kg당, 원)	0.9178	100.0
3	$Q_{dae_{t-1}}$	$t-1$주차 대구북부도매시장 토마토 반입량(kg)	-0.002889	93.1

4	$Q_{garak_{t-1}}$	$t-1$주차 가락도매시장 토마토 반입량(kg)	0.0006663	87.0
5	$P_{garak_{t-2}}$	$t-2$주차 가락도매시장 토마토 가격(kg당, 원)	-0.2706	84.8
6	$P_{garak_{t-3}}$	$t-3$주차 가락도매시장 토마토 가격(kg당, 원)	0.1356	70.9
7	$Q_{RE_{t-2}}$	$t-2$주차 A 연합사업단 소매 출하량 레벨(1~5)	52.97	57.3
8	$Q_{buban_{t-2}}$	$t-2$주차 부산반여도매시장 토마토 반입량(kg)	0.001982	51.0
9	Q_{RE_t}	t주차 A 연합사업단 소매 출하량 레벨(1~5)	-43.24	50.8
10	Q_{WO_t}	t주차 A 연합사업단 도매 출하량 레벨(1~5)	-37.52	50.2

선정된 주요 설명 변수를 PIP 순으로 나열하면 t주차 가락도매시장 반입량(PIP=100.0, PM=-0.0007861), $t-1$주차 가락도매시장 가격(PIP=100.0, PM=0.9178), $t-1$주차 대구북부도매시장 반입량(PIP=93.1, PM=-0.002889), $t-1$주차 가락도매시장 반입량(PIP=87.0, PM=0.000666), $t-2$주차 가락도매시장 가격(PIP=84.8, PM=-0.2706), $t-3$주차 가락도매시장 가격(PIP=70.9, PM=0.1356), $t-2$주차 A 연합사업단 소매 출하량 레벨(PIP=57.3, PM=52.97), $t-2$주차 부산반여도매시장 반입량(PIP=51.0, PM=0,0001982), t주차 A 연합사업단 소매 출하량 레벨(PIP=50.8, PM=43.24), A 연합사업단 t주차 도매 출하량 레벨(PIP=50.2, PM=-37.5) 순이다.

표 3. BMA 모델 추정치 및 실제값 간 차이(2021.01 ~ 2021.13주차)

기간	실제값(A)	추정치(B)	차이 |(A)-(B)|	방향성
2021.01	1,819	1,850	31	
2021.02	1,991	1853	138	O
2021.03	2,570	2,272	298	O

2021.04	2,817	2,941	124	O
2021.05	2,649	2,942	293	X
2021.06	1,943	2,288	345	O
2021.07	2,003	2,399	396	O
2021.08	2,770	2,660	110	O
2021.09	3,369	3,420	51	O
2021.10	3,683	3,387	296	X
2021.11	3,417	3,420	3	X
2021.12	2,765	3,267	502	O
2021.13	2,717	2,972	255	O

t주차 가락도매시장 반입량과 $t-1$주차 대구북부도매시장 반입량은 목표 변수인 t주차 가락도매시장 가격에 부(-)의 영향을 미치는 것으로 나타났다. 반면 $t-2$주차 부산반여도매시장 반입량은 정(+)의 영향을 미치는 것으로 나타났다. 전국에서 토마토 물량이 수급되는 가락도매시장의 환경을 고려했을 때 경남 지역에 물량이 집중되면 가락도매시장 내 물량 수요가 높아지는 것으로도 해석할 수 있다. 가락도매시장 가격의 과거 시차 즉, $t-1 \sim t-3$주차 가격은 각각 정(+), 부(-), 정(+)의 영향을 미치는 것으로 나타났다. 마지막으로 A 연합사업단의 시장 출하량 관련 설명 변수를 살펴보면, $t-2$주차 소매 출하량 레벨은 t주차 가락도매시장 가격에 정(+)의 영향을 주는 반면, t주차 소매 및 도매 출하량 레벨은 부(-)이 영향을 미치는 것으로 나타났다. BMA에서 추정된 결과를 기반으로 PSO에 투입할 목적함수를 정리하면, 식 (4)와 같다.

$$P_{garak_t} = \max\left\{P_{garak_t} \middle| \begin{array}{c} -\theta_1 x_1 + \theta_2 x_2 - \theta_3 x_3 + \theta_4 x_4 - \theta_5 x_5 + \\ \theta_6 x_6 + \theta_7 x_7 + \theta_8 x_8 - \theta_9 x_9 - \theta_{10} x_{10} \end{array}\right\} \quad (4)$$

$\theta_1 x_1 \sim \theta_{10} x_{10}$은 선정된 주요 설명 변수의 모수 및 변수명을 축약한 것이다(표 2. 설명 변수 순서 및 PM 수치 참고). 주별 가락도매시장 토마토 가격이 높아지는 방향에 대한 PSO를 활용한 최적화

모델 관련 수식은 아래 식 (5)와 같다.

$$\max \quad P_{garak_t}$$

$$s.t. \quad \begin{pmatrix} P_{garak_{t-1,t-2,t-3}} > 0 \\ Q_{garak_{t,t-1}} > 0 \\ Q_{garak_{t,t-1}} > Q_{dae_{t-1}} \\ Q_{garak_{t,t-1}} > Q_{buban_{t-2}} \\ Q_{dae_{t-1}} > 0 \\ Q_{buban_{t-2}} > 0 \end{pmatrix} \tag{5}$$

제약 조건을 살펴보면, $t-1$ ~ $t-3$주차 가락도매시장 토마토 가격은 반드시 0보다 큰 수치를 가져야 하며, 이와 유사하게 t ~ $t-1$주차 가락도매시장 토마토 반입량 역시 0보다 큰 수치를 가져야 한다. 국내 시장 상황을 고려했을 때, 대구북부도매시장과 부산반여도매시장의 주별 토마토 반입량은 가락도매시장 주별 토마토 반입량 수치보다 낮아야 한다. 또한, 대구북부도매시장과 부산반여도매시장의 주별 토마토 반입량은 0보다 큰 수치를 가져야 한다. 식 (5)를 활용한 PSO의 수렴 절차는 아래와 같다.

Algorithm. Max P_{garak_t} using PSO pseudo-code

```
Inputs: f, C          ⇒ f is formula (11) above, C are control parameters
P ← initialization(C)      ⇒ set initial swarms such as position, velocity
B ← best(P, f)
i ← 0                                ⇒ number of Iterations of the method
while termination_criteria(P, f C, i) do
   for each particle x = (s, v, p, l) ∈ P do
       v ← volocity(s, v, p, l)                ⇒ calculate new velocity for x
       s ← s + v              ⇒ move particle to new position s(mutation)
       s ← confinement(s, C)
       if f(s) < f(p) then p ← s                    ⇒ update previous best
       end if
       x ← (s, v, p, l)                                  ⇒ update particle
       if f(s) < f(B) then B ← s                     ⇒ update best value
       end if all constraints were satisfied
    end for
    s ← s + v
end while
Ouput: B                                  ⇒ best solution for P_garak_t
```

PSO에 적용되는 정의역 집합(Domain set) 즉, D차원의 particle 집합은 BMA로 선정된 주요 설명 변수 중 t주차 관련 설명 변수 집합{Q_{garak_t}(t주차 가락도매시장 토마토 반입량), Q_{RE_t}(t주차 A 연합사업단 소매 출하량 레벨), Q_{WO_t}(t주차 A 연합사업단 도매 출하량 레벨)}이다. 선정된 10개의 주요 설명 변수 중 $t-n$주차 시차 변수는 이미 발생한 상황으로, t주차에서는 정보 확보가 가능하다. 그러나 t주차 변수 관련 정보는 예측은 가능하나 실제 정보에 대한 확보는 불가능하다. 이를 고려해 출하 담당자가 t주차 설명 변수에는 일정 범위의 수치를 부여할 수 있게 조치한 후, 그 외 이미 발생한 정보($t-n$주차 설명 변수)와의 조합을 통해 t주차 가락도매시장 토마토 가격이 가장 높아지는 시나리오 20개를 추출할 수 있는 의사결정지원도구를 개발하였다. 또한, 실무자 활용 편의성을 위해 웹 기반(Web based)으로 개발하였다.

그림 5. 웹 기반 출하 의사결정지원도구 시장 현황 입력 화면

그림 5.는 웹 기반 출하 의사결정지원도구에서 각 항목에 대한 현재 상황에 대한 입력 화면을 보여주고 있다. $t-n$주차의 과거 정보는 관련 정보가 데이터베이스(Database)를 통해 자동 입력되도록 조치하였으며(직접 입력도 가능), t주차 정보는 상한과 하한을 구분하여 입력할 수 있도록 구성하였다. A 연합사업단이 계획하고 있는 t주차 도매 및 소매 출하량 레벨, 실무 담당자가 예상하는 t주차 가락도매시장 반입량 범위를 입력하면 된다.

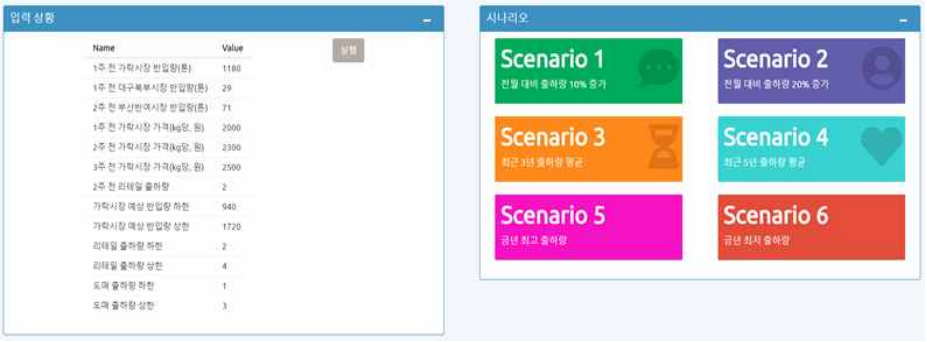

그림 6. 입력 항목 표출 및 6개의 기본 시나리오

그림 6.은 실무 담당자가 입력한 항목에 대해 확인할 수 있는 도표와 6개의 기본 출하 시나리오를 보여주고 있다. 6개의 기본 시나

리오는 실무 담당자의 입력 항목과는 무관하게 자동으로 산출되는 정보이며, 실무 담당자가 해당 시나리오를 변경할 수 있도록 구성하였다.

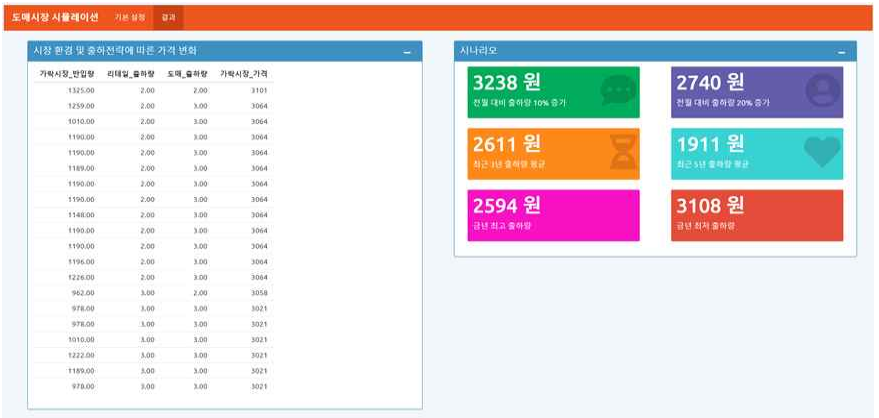

그림 7. 출하 의사결정지원도구 분석 결과

그림 7.은 실무 담당자가 입력한 정보 및 기본 6개의 시나리오에 대한 분석 결과이다. 담당자가 입력한 내용을 기반으로 t주차 가락도매시장 토마토 가격이 최대화되는 20개의 시나리오가 내림차순(가락도매시장 가격 기준)으로 표출되어 진다.

제4장. 연구 결론

제1절. 제2장 연구 결론

1.1. 연구 결론

제2장의 연구에서는 기존 FQMA를 개선한 H-FQMA를 활용해 주별 가락도매시장 토마토 가격 예측 모델을 개발하였다. H-FQMA에 RMCFS를 변수 선정에 적용한 RF 모델은 다른 15개 모델 대비 우수한 성능을 기록하였다. 특히, RF 모델은 이상치에 대한 대응력도 다른 모델 대비 우수한 것으로 나타났다. 그러나 RF 모델 역시 이상치에 대한 완벽한 대응은 어려운 상황이며, 이는 향후 보완이 필요한 부분이다.

그 외에도, t주차 예측 변수를 활용한 NOW 모델은 대체적으로 낮은 성능을 보였다. 또한, 단변량 시계열 방법만을 활용한 AR 모델은 변수 선정 방법이 적용된 BM, RF보다 낮은 성능을 기록하였다. 도매시장 토마토 가격은 여러 변인에 의해 영향을 받는다. 따라서, 자기회귀 과정만으로 도매시장 토마토 가격을 예측하는데 어려움이 따르며, 농산물 가격 예측 모델 개발 시 설명 변인을 투입하는 방법이 예측 성능 향상에 도움이 되는 것으로 해석 가능하다.

연구의 시사점으로는 먼저, 도매시장 거래 자료만을 활용한 제한적 환경에서, 통계 기법만으로 주별 가락도매시장 토마토 가격 모델의 예측 오차를 11% 수준(MAPE 기준)에 근접하게 만든 것을 언급할 수 있다. 국내 대부분의 농산물 가격 예측 오차는 월 기준 10 ~ 20% 수준임을 고려했을 때, 주 기준 오차가 11% 수준으로 기록된

것은 우수한 결과이다. 둘 째, 연구에서 예측 조합이라는 국내에서 시도되지 않은 방법론을 적용하는 등 국내 농산물 가격 예측 분야에 새로운 방향을 제시하였다. 또한, H-FQMA는 전기수요예측 연구에서 개발된 FQMA를 계승한 방법으로, 국외 농산물 가격 예측 분야에서도 이론적인 시사점이 존재한다. 마지막으로, 연구에서 제안한 H-FQMA는 토마토 뿐만 아니라 타 품목에도 적용될 수 있는 모델이다. 이와 관련해 라이브러리(Library)를 제작 중에 있으며, 이 또한 통계적 기술 관점에 있어 시사점을 제공할 수 있다.

1.2. 한계점 및 후속 연구

연구의 한계점으로는 첫 째, 주별 가락도매시장 토마토 가격에 영향을 미치는 예측 변수 선택에 대한 이론적 토대가 부족하다. 본문에서 언급했듯, 농산물 가격은 여러 요인에 의해 영향을 받으므로 특정 변인을 모델의 예측 변수로 고정하기 힘든 환경이다. 그러나 이러한 상황에도, 예측 모델의 객관적 설명력을 높이기 위해서는 이론적 토대를 기반으로 예측 변인을 선정할 필요가 있다. 단순 수요공급 논리에 따른 파생변수 생성이 아닌, 이론적인 기반 아래 파생변수가 생성되고, 최종 예측 변수가 선정될 필요가 있다. 둘 째, 월별 예측 모델 연구에 집중된 국내 농산물 가격 예측 연구를 고려했을 때, 주별 예측 모델 성능의 상대적 우위를 논하기 어려운 측면이 있다. 이전보다 높은 빈도의 주별 예측 모델 개발을 통해 모델의 성능을 객관적으로 평가할 수 있는 연구 환경 구축이 필요하다.

연구를 기반으로 다음과 같은 후속 연구를 고려할 수 있다. 첫 째, 농산물 도매시장 가격에 영향을 미치는 예측 변인 선정에 있어 이론적 토대를 마련하기 위해 도매시장 간 관계 분석을 진행할 필요가 있다. 2020년에 'Journal of the American Statistical Association'에 게제된 Ando & Bai(2020)가 제안한 'Panel Quantile Model'을 기반으로 국내 도매시장 간 동행성(Co-movement)을 모델 관점에서 추정할 수 있다. 이러한 연구를 통해 국내 농산물 도매시장 간 관계

규명이 필요하다. 둘 째, 농산물 가격 예측 연구에서 최근 진행된 세 개의 예측 조합 방법과 H-FQMA의 예측 성능을 비교하는 연구를 생각해 볼 수 있다. 기존 예측 조합 연구는 오차 기반(Error based), 최적화 알고리즘(ABC algorithm) 등의 방법을 활용하였다. 기존 세 개 방법과 본 연구에서 제안한 H-FQMA를 비교한다면 의미 있는 연구가 될 것이다. 마지막으로, 예측 조합의 대상을 개별 통계적 방법이 아닌 사람의 의견을 적용하는 연구를 고려할 수 있다. Atalla et al.(2016)는 석유 가격에 대한 다수의 전문가 예측을 변수로 활용하였다. 이를 통해 석유 가격 변동성(Volatility)을 시계열 모델로 추정하고, 다음 분기 석유 가격에 대한 예측 모델을 개발하였다. 국내에서도 농산물 가격에 대한 전문가 집단을 구성해 다음 주기(분기, 월, 주) 예측치를 조사한 후, 이를 활용해 예측 모델을 개발하는 방법을 고려할 필요가 있다.

제2절. 제3장 연구 결론

2.1. 연구 결론

BMA를 활용해 주별 가락도매시장 토마토 가격을 추정한 결과, A 연합사업단의 시장 출하는 가락도매시장 토마토 가격에 영향을 미치는 것으로 나타났다. 즉, A 연합사업단의 시장에 미치는 영향력이 존재하는 것으로 해석 가능하다.

이 결과를 토대로 PSO를 적용해 주별 가락도매시장 토마토 가격이 최대화되는 시나리오 20개를 추출하는 모델을 개발하였고, 이를 활용해 웹 기반 출하 의사결정지원도구를 추가로 개발하였다. 개발된 출하 의사결정지원도구는 다음과 같이 활용될 수 있다. A 연합

사업단과 같은 공급자 입장에서는 구매자의 힘(Buyer's power)이 강한 소매 조직(이마트, 홈플러스 등) 있을 수도 있고, A 연합사업단이 우위에 있는 즉, 공급자의 힘(Supplier's power)이 강한 유통 조직도 존재할 수 있다. 구매자의 힘이 강한 소매 조직 대상으로 물량을 조절하는 것은 어려운 부분이 있다. 그 이유로는 구매자가 원하는 시점과 수요량에 항시 대응해야 하는 위치이기 때문이다. 그러나 공급자의 힘이 강한 소매 또는 도매 조직에게는 자체적인 물량 조절이 가능하다. 이러한 점을 고려했을 때, 출하 담당자는 물량 조절이 가능한 도매 및 소매 물량을 의사결정지원도구를 통해 입력하여 가락도매시장 가격 변화를 확인한 후, 적절한 출하 전략을 구상할 수 있다.

연구의 시사점으로는 그간 농산물 유통 연구에서 개별 유통 조직의 출하량이 시장에 미치는 영향력을 추정한 연구는 없었다. 개별 농산물 유통조직이 시장 가격에 미치는 영향력에 대해 모델 관점에서 접근한 부분은 농산물 유통 연구 분야에서 이론적 가치가 존재한다. 실무적 관점에서도 단순 모델 개발에 그치지 않고 웹 기반 의사결정지원도구를 제공하는 등 출하 담당자의 현장 활용성을 제공한 점에서 실무적 시사점이 존재한다.

2.2. 한계점 및 후속 연구

BMA를 통해 도출된 주별 가락도매시장 토마토 가격에 영향을 미치는 주요 변인에 대한 이론적 토대가 부족하다. 또한, 주별 예측 모델을 개발하는데 따른 제약으로 인해 A 연합사업단 내부 자료 활용에도 제약이 발생하였다. 모델 개발 시 다양한 시장 출하처를 예측 변인으로 고려하려 했으나 주 단위에서는 출하 이력이 없는 출하처가 많아 다양한 변인을 모델에 투입하지 못하였다. 전국 연합사업단 자료를 활용하지 못한 것도 한계점으로 작용한다. 연합사업단은 광역 단위로 설립되었다. 모든 연합사업단의 자료를 활용할 수 있다면, 이를 기반으로 농협 출하 물량이 시장에 미치는 영향력을

편의가 낮은 방향으로 추정할 수 있을 것이다.

후속 연구로는, 국외 연구에서는 농산물 유통 관련 의사결정지원 도구 연구에서 System dynamics가 주로 활용되었다. 국내 도매시장은 다양한 변인이 상호 영향을 미치는 환경으로, 복잡계 연구의 대표적 방법인 System dynamics와 같은 모델이 적합할 수도 있다. 정성적 및 정량적 분석을 조합하여 복잡계 연구의 대표 모델인 시스템 다이내믹스를 활용해 국내 도매시장을 둘러싼 각 변인의 유기적인 관계를 시뮬레이션 할 수 있는 연구가 수행된다면 시사점을 제공할 수 있을 것이다.

제5장. 참고문헌 및 부록(Appendix)

참고문헌

(1) 강석호, & 김승. (2008). Particle 2-Swarm Optimization for Robust Search. Journal of information and operations management (경영정보논총), 18.

(2) 강태훈. (2009). 농산물 가격전달의 비대칭성에 관한 실증 분석. 농촌경제, 32(5), 63-81.

(3) 김기환, & 서병선. (2017). 사과의 산지도매가격과 소매가격 간 동태적 인과성 및 비대칭적 가격 조정 연구. 농업경제연구, 58(1), 97-116.

(4) 김동환. (2010). 소매업체와 농산물 도매시장의 연계성 강화 방안: 청과물을 중심으로. 유통연구, 15(5), 273-285.

(5) 김연중, 박기환, 서대석, & 한혜성. (2010). 주요 농산물의 가치사슬 분석과 성과제고 방안. 서울: 한국농촌경제연구원.

(6) 김종진, 김종인, & 조남욱. (2017). 수확기 쌀 가격 예측 모델 개선에 관한 연구. 원 장 김 창 길, 1.

(7) 김종진, & 형남원. (2018). 동태요인모델을 이용한 국제곡물 가격 예측. 농촌경제, 41(4), 29-51.

(8) 남국현, & 최영찬. (2015). 양파 출하시기 도매가격 예측모델 연구. 농촌지도와 개발, 22(4), 423-434.

(9) 남민호, 김현태. (2014), 베이지안 모델평균법을 이용한 경상수지 벤치마크 추정. 한국은행 경제연구원, 20(3), 39-72

(10) 명광식. (2005). Box-Jenkins 모델을 이용한 육계가격 예측. 농

촌경제 28 (2), 73-83.

(11) 박명규. (2015). 시뮬레이션을 활용한 농식품 유통물류 동선최적화 설계방안 비교연구. 한국시뮬레이션학회 논문지, 24(4), 163-170.

(12) 배경태, & 김창재. (2016). 인공신경망의 은닉층 최적화를 통한 농산물 가격예측 모델. 한국정보기술학회논문지, 14(12), 161-169.

(13) 서교, 이정재, 김한중, & 나준엽. (2000). 농산물 유통을 위한 최적 경로 선정 모델의 개발. 한국농공학회 학술대회초록집, 2000, 233-238.

(14) 신성호, 이미경, & 송사광. (2018). LSTM 네트워크를 활용한 농산물 가격 예측 모델. 한국콘텐츠학회논문지, 18(11), 416-429.

(15) 양한탁, 이연정, & 윤성민. (2018). 농산물 도매-소매단계 가격전달의 비대칭성 분석: 대형마트와 재래시장의 비교를 중심으로. Journal of The Korean Data Analysis Society, 20(2), 759-771.

(16) 유도일. (2016). 기상변인을 고려한 채소가격 예측모델 개발. 농업경제연구, 57(1), 1-24.

(17) 윤성주, 이춘수, & 양승룡. (2016). 인공 신경망을 이용한 국제곡물가격예측 모델 개발. 농업경제연구, 57(2), 83-101.

(18) 이동훈. (2018). 스마트 팜 투자 사업의 실물옵션 분석:-딸기·토마토 농가를 중심으로. 입법과 정책, 10(2), 275-303.

(19) 이형용, 여민수, & 홍승지. (2017). 마늘 도매가격 시계열 예측모델 비교. FTA 이행에 따른 낙농· 유가공 산업의 경제적 영향 분석, 55.

(20) 임진모, 김월용, 변우진, & 신승중. (2018). 인공지능을 이용한 과일 가격 예측 모델 연구. 문화기술의 융합, 4(2), 197-204.

(21) 장수희, 전희주, 조인호, & 김동환. (2017). 비정형 농업기상자료를 활용한 배추 도매가격 예측모델 연구. 한국데이터정보과학회지, 28(3), 617-624.

(22) 전창곤, 최병옥, & 김동훈. (2011). 산지유통인의 법인화와 계약재배 참여방안 연구. 한국농촌경제연구원 정책연구보고서, 1-135.

(23) 채명수, 김상식, 이채석, 이요한, & 정성관. (2017). 다중 뉴럴 네트워크 기반의 기상정보를 반영한 농산물 가격 예측. 한국통신학회 학술대회논문집, 878-879.

(24) 최병옥, & 최익창. (2007). 시계열 분석방법을 이용한 과채류 월별가격 예측. 농촌경제 30 (1), 129-148.

(25) 최수진, & 김한호. (2017). 옥수수와 밀 선물가격 결정요인 비교분석 베이지안모델평균을 이용하여. 농촌경제, 40(2), 99-123.

(26) 하지희, 서상택, & 김선웅. (2019). 양파와 마늘가격 예측모델의 예측력 비교. 한국식품유통학회 하계학술발표논문집, 2019, 559-572.

(27) Abdi, H., & Williams, L. J. (2010). Principal component analysis. Wiley interdisciplinary reviews: computational statistics, 2(4), 433-459.

(28) Ahumada, H., & Cornejo, M. (2016). Forecasting food prices: The case of corn, soybeans and wheat. International Journal of Forecasting, 32(3), 838-848.

(29) Amidan, B. G., Ferryman, T. A., & Cooley, S. K. (2005, March). Data outlier detection using the Chebyshev theorem. In 2005 IEEE Aerospace Conference (pp. 3814-3819). IEEE.

(30) Ando, T., & Bai, J. (2020). Quantile co-movement in financial markets: A panel quantile model with

unobserved heterogeneity. Journal of the American Statistical Association, 115(529), 266-279.

(31) Atalla, T., Joutz, F., & Pierru, A. (2016). Does disagreement among oil price forecasters reflect volatility? Evidence from the ECB surveys. International Journal of Forecasting, 32(4), 1178-1192.

(32) Behzadi, G., O'Sullivan, M. J., Olsen, T. L., & Zhang, A. (2018). Agribusiness supply chain risk management: A review of quantitative decision models. Omega, 79, 21-42.

(33) Bates, J. M., & Granger, C. W. (1969). The combination of forecasts. Journal of the Operational Research Society, 20(4), 451-468.

(34) Blanc, S. M., & Setzer, T. (2016). When to choose the simple average in forecast combination. Journal of Business Research, 69(10), 3951-3962.

(35) Barrow, D. K., & Crone, S. F. (2016). A comparison of AdaBoost algorithms for time series forecast combination. International Journal of Forecasting, 32(4), 1103-1119.

(36) Baffes, J., & Haniotis, T. (2016). What explains agricultural price movements?. Journal of Agricultural Economics, 67(3), 706-721.

(37) Bennett, C. J., Stewart, R. A., & Lu, J. W. (2014). Forecasting low voltage distribution network demand profiles using a pattern recognition based expert system. Energy, 67, 200-212.

(38) Cao, R., Hart, J., & Saavedra, Á. (2003). Nonparametric maximum likelihood estimators for AR and MA time series. Journal of Statistical Computation and Simulation,

73(5), 347-360.

(39) Chen, H., Chong, T. T. L., & Bai, J. (2012). Theory and applications of TAR model with two threshold variables. Econometric Reviews, 31(2), 142-170.

(40) Dąbrowski, J., & Adamczyk, T. (2010). Application of GAM additive non-linear models to estimate real estate market value. Geomatics and Environmental Engineering, 4(2), 55-62.

(41) Dramiński, M., & Koronacki, J. (2018). rmcfs: an R package for Monte Carlo feature selection and interdependency discovery. Journal of Statistical Software, 85(1), 1-28.

(42) Fernandez, C., Ley, E., & Steel, M. F. (2001). Benchmark priors for Bayesian model averaging. Journal of Econometrics, 100(2), 381-427.

(43) Ferreira, J. O., Batalha, M. O., & Domingos, J. C. (2016). Integrated planning model for citrus agribusiness system using systems dynamics. Computers and Electronics in Agriculture, 126, 1-11.

(44) Flores, H., Villalobos, J. R., Ahumada, O., Uchanski, M., Meneses, C., & Sanchez, O. (2019). Use of supply chain planning tools for efficiently placing small farmers into high-value, vegetable markets. Computers and electronics in agriculture, 157, 205-217.

(45) Forte, A., Garcia Donato, G., & Steel, M. (2018). Methods and tools for Bayesian variable selection and model averaging in normal linear regression. International Statistical Review, 86(2), 237-258.

(46) García Nieto, P. J., García-Gonzalo, E., Arbat, G., Duran-Ros, M., Ramirez de Cartagena, F., &

Puig-Bargués, J. (2016). A new predictive model for the filtered volume and outlet parameters in micro-irrigation sand filters fed with effluents using the hybrid PSO-SVM-based approach. Computers and Electronics in Agriculture, 125, 74-80.

(47) Gargano, A., &Timmermann, A. (2014). Forecasting commodity price indexes using macroeconomic and financial predictors. International Journal of Forecasting , 30 (3), 825-843.

(48) Ge, H., Gray, R., & Nolan, J. (2015). Agricultural supply chain optimization and complexity: A comparison of analytic vs simulated solutions and policies. International Journal of Production Economics, 159, 208-220.

(49) Ge, H., Nolan, J., & Gray, R. (2015). Identifying strategies to mitigate handling risks in the Canadian grain supply chain. Canadian Journal of Agricultural Economics/Revue canadienne d'agroeconomie, 63(1), 101-128.

(50) Graefe, A., Küchenhoff, H., Stierle, V., & Riedl, B. (2015). Limitations of Ensemble Bayesian Model Averaging for forecasting social science problems. International Journal of Forecasting, 31(3), 943-951.

(51) Gibson, D., & Nur, D. (2011). Threshold autoregressive models in finance: a comparative approach.

(52) Hsiao, C., & Wan, S. K. (2014). Is there an optimal forecast combination?. Journal of Econometrics, 178, 294-309.

(53) Kamilaris, A., & Prenafeta-Boldú, F. X. (2018). Deep learning in agriculture: A survey. Computers and electronics in agriculture, 147, 70-90.

(54) Kaur, M., Gulati, H., & Kundra, H. (2014). Data mining in

Agriculture on crop price prediction: Techniques and Applications. International Journal of Computer Applications, 99(12), 1-3.

(55) Kim, K. H., Kim, Y. S., Hong, M. S., & Yook, H. S. (2016). Quality characteristics of meringue cookies added with tomato powder. Journal of the Korean Society of Food Science and Nutrition, 45(3), 366-371.

(56) Kumar, S., & Nigmatullin, A. (2011). A system dynamics analysis of food supply chains - Case study with non-perishable products. Simulation Modelling Practice and Theory, 19(10), 2151-2168.

(57) Kyriazi, F., Thomakos, D. D., & Guerard, J. B. (2019). Adaptive learning forecasting, with applications in forecasting agricultural prices. International Journal of Forecasting, 35(4), 1356-1369.

(58) Lessmann, S., & Voß, S. (2017). Car resale price forecasting: The impact of regression method, private information, and heterogeneity on forecast accuracy. International Journal of Forecasting, 33(4), 864-877.

(59) Lever, J., Krzywinski, M., & Altman, N. (2017). Points of significance: Principal component analysis. Nature methods, 14(7), 641-643.

(60) LI, G. Q., XU, S. W., LI, Z. M., SUN, Y. G., & DONG, X. X. (2012). Using quantile regression approach to analyze price movements of agricultural products in china. Journal of Integrative Agriculture, 11(4), 674-683.

(61) Li, Z., Li, G., & Wang, Y. (2010). Construction of short-term forecast model of eggs market price. Agriculture and Agricultural Science Procedia, 1, 396-401.

(62) Luo, J., Klein, T., Ji, Q., & Hou, C. (2022). Forecasting realized volatility of agricultural commodity futures with infinite Hidden Markov HAR models. International Journal of Forecasting, 38(1), 51-73.

(63) Maciejowska, K., Nowotarski, J., & Weron, R. (2016). Probabilistic forecasting of electricity spot prices using Factor Quantile Regression Averaging. International Journal of Forecasting, 32(3), 957-965.

(64) Melkumova, L. E., & Shatskikh, S. Y. (2017). Comparing Ridge and LASSO estimators for data analysis. Procedia engineering, 201, 746-755.

(65) Mendes-Moreira, J., Soares, C., Jorge, A. M., & Sousa, J. F. D. (2012). Ensemble approaches for regression: A survey. Acm computing surveys (csur), 45(1), 1-40.

(66) Nasira, G. M., & Hemageetha, N. (2012, March). Vegetable price prediction using data mining classification technique. In International Conference on Pattern Recognition, Informatics and Medical Engineering (PRIME-2012) (pp. 99-102). IEEE.

(67) Nowotarski, J., Liu, B., Weron, R., & Hong, T. (2016). Improving short term load forecast accuracy via combining sister forecasts. Energy, 98, 40-49.

(68) Nowotarski, J., & Weron, R. (2014). Merging quantile regression with forecast averaging to obtain more accurate interval forecasts of Nord Pool spot prices. In 11th International Conference on the European Energy Market (EEM14) (pp. 1-5). IEEE.

(69) Onour, I., & Sergi, B. S. (2011). Modeling and Forecasting Volatility in the Global Food Commodity Prices

(Modelování a Prognózování Volatility Globálních cen Potravinářských Komodit). Agricultural Economics-Czech, 57(3), 132-139.

(70) Ordiano, J. Á. G., Gröll, L., Mikut, R., & Hagenmeyer, V. (2020). Probabilistic energy forecasting using the nearest neighbors quantile filter and quantile regression. International journal of forecasting, 36(2), 310-323.

(71) Oshiro, T. M., Perez, P. S., & Baranauskas, J. A. (2012). How many trees in a random forest?. In International workshop on machine learning and data mining in pattern recognition (pp. 154-168). Springer, Berlin, Heidelberg.

(72) Özbayrak, M., Papadopoulou, T. C., & Akgun, M. (2007). Systems dynamics modelling of a manufacturing supply chain system. Simulation Modelling Practice and Theory, 15(10), 1338-1355.

(73) Petropoulos, F., Hyndman, R. J., & Bergmeir, C. (2018). Exploring the sources of uncertainty: Why does bagging for time series forecasting work?. European Journal of Operational Research, 268(2), 545-554.

(74) Ramos-Pérez, E., Alonso-González, P. J., & Núñez-Velázquez, J. J. (2019). Forecasting volatility with a stacked model based on a hybridized Artificial Neural Network. Expert Systems with Applications, 129, 1-9.

(75) Sabu, K. M., & Kumar, T. M. (2020). Predictive analytics in Agriculture: Forecasting prices of Arecanuts in Kerala. Procedia Computer Science, 171, 699-708.

(76) Sánchez, I. (2008). Adaptive combination of forecasts with application to wind energy. International Journal of

Forecasting, 24(4), 679-693.

(77) Shah, A., Chauhan, Y., & Chaudhury, B. (2021). Principal component analysis based construction and evaluation of cryptocurrency index. Expert Systems with Applications, 163, 113796.

(78) Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. Statistics and computing, 14(3), 199-222.

(79) Sterman, J., 2000. Business dynamics: systems thinking and modeling for a complex, world. Irwin/McGraw-Hill.

(80) Tako, A. A., & Robinson, S. (2012). The application of discrete event simulation and system dynamics in the logistics and supply chain context. Decision support systems, 52(4), 802-815.

(81) Teimoury, E., Nedaei, H., Ansari, S., & Sabbaghi, M. (2013). A multi-objective analysis for import quota policy making in a perishable fruit and vegetable supply chain: A system dynamics approach. Computers and electronics in agriculture, 93, 37-45.

(82) Wang, J., Wang, Z., Li, X., & Zhou, H. (2019). Artificial bee colony-based combination approach to forecasting agricultural commodity prices. International Journal of Forecasting.

(83) Wang, L., Feng, J., Sui, X., Chu, X., & Mu, W. (2020). Agricultural product price forecasting methods: research advances and trend. British Food Journal.

(84) Wang, P., Liu, B., & Hong, T. (2016). Electric load forecasting with recency effect: A big data approach. International Journal of Forecasting, 32(3), 585-597.

(85) Wihartiko, F. D., Nurdiati, S., Buono, A., & Santosa, E. Agricultural Price Prediction Models: A Systematic Literature Review.

(86) Xiong, T., Li, C., Bao, Y., Hu, Z., & Zhang, L. (2015). A combination method for interval forecasting of agricultural commodity futures prices. Knowledge-Based Systems, 77, 92-102.

(87) Yang, W., Chan, F. T., & Kumar, V. (2012). Optimizing replenishment polices using genetic algorithm for single-warehouse multi-retailer system. Expert Systems with Applications, 39(3), 3081-3086.

(88) Ye, M., Zyren, J., & Shore, J. (2005). A monthly crude oil spot price forecasting model using relative inventories. International Journal of Forecasting, 21(3), 491-501.

(89) Yu, H., Chen, Y., Hassan, S. G., & Li, D. (2016). Prediction of the temperature in a Chinese solar greenhouse based on LSSVM optimized by improved PSO. Computers and Electronics in Agriculture, 122, 94-102.

(90) Yuan, C. Z., & Ling, S. K. (2020, August). Long Short-Term Memory Model Based Agriculture Commodity Price Prediction Application. In Proceedings of the 2020 2nd International Conference on Information Technology and Computer Communications (pp. 43-49).

(91) Zhang, D., Zang, G., Li, J., Ma, K., & Liu, H. (2018). Prediction of soybean price in China using QR-RBF neural network model. Computers and Electronics in Agriculture, 154, 10-17.

(92) Zhu, B., Feng, Y., Gong, D., Jiang, S., Zhao, L., & Cui, N. (2020). Hybrid particle swarm optimization with extreme

learning machine for daily reference evapotranspiration prediction from limited climatic data. Computers and Electronics in Agriculture, 173, 105430.

부록(Appendix)

부록 A. 국내외 연구 비교(도표)

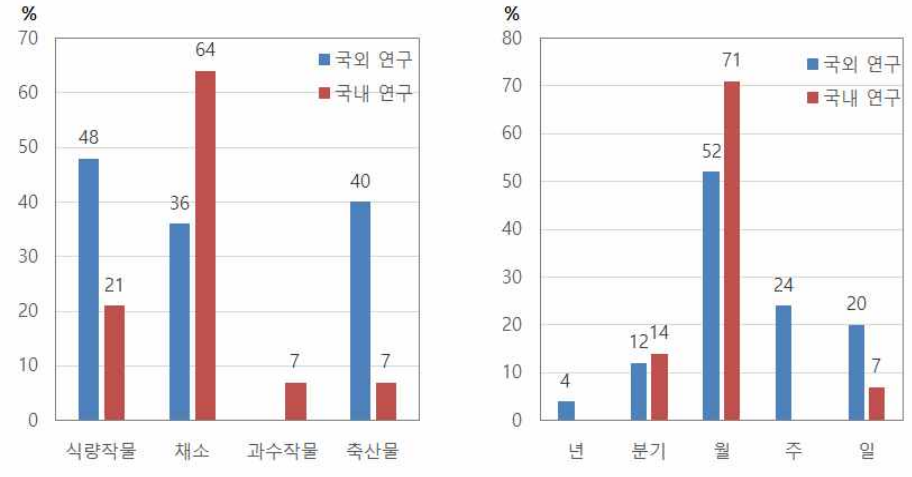

그림 A1. 목표 변인 비율(좌) 및 예측 주기 비율 (우)에 대한 국내외 연구 비교

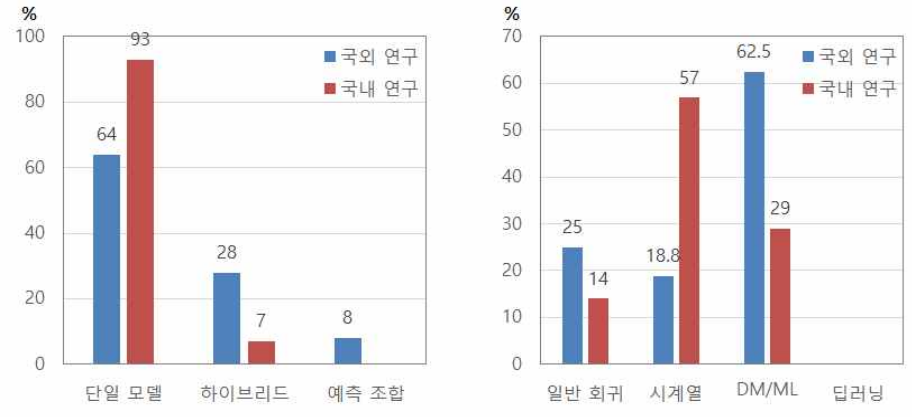

그림 A2. 예측 방법 비율(좌) 및 단일 모델 내 통계 방법 비율(우)에 대한 국내외 연구 비교

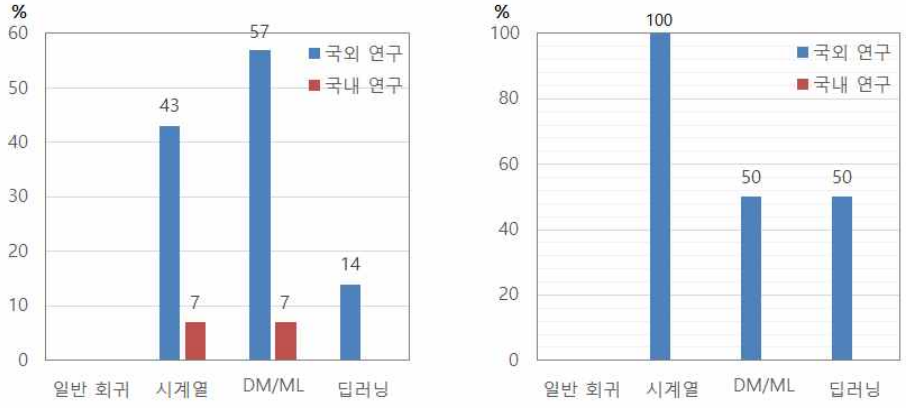

그림 A3. 하이브리드 모델 내 통계 방법 비율(좌) 및 예측 조합 내 통계 방법 비율(우)에 대한 국내외 연구 비교

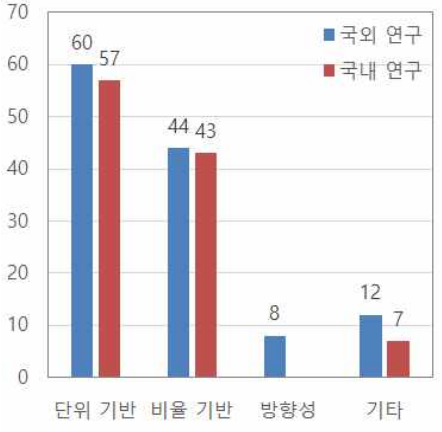

그림 A4. 예측 성능 지표 비율에 대한 국내외 연구 비교

부록 B. H-FQMA 적용 통계 방법

B.1. 단변량 시계열 모델(Univariate time series models)

AR 모델은 시계열 방법에서 가장 기본이 되는 방법으로, 아래와 같이 표현될 수 있다.

$$P_t = \theta_0 + \sum_{i=1}^{8} \theta_1 P_{t-i} + \phi_t Q_{t-1} + \pi_t MW_t + \epsilon_{ar,t} \qquad \text{(B1)}$$

P_t는 주별 가락도매시장 토마토 가격, θ_0은 절편. ϕ_t는 외생 변수(Exogenous variable) $t-1$주차 가락도매시장 토마토 반입량에 대한 계수(Coefficient), π_t는 월/주 변수의 계수를 의미한다. 시차(Lag)는 총 1~8로 설정했으며, 외생 변수인 가락도매시장 토마토 반입량(Q_{t-1})이 포함될 경우 ARX 모델이 되며, LS(Least squares)로 추정하였다. 정상성(Stationary)을 검증한 후 차분을 결정하였다.

AR 모델에서 TAR(Threshold autoregressive)은 Tong(1983)에 의해 개발된 방법으로(Chen et al., 2012), AR 모델에 국면전환(Regime-switching) 개념이 적용되어(Gibson & Nur, 2011), 임계점을 기준으로 두 개의 국면(Regime)이 구분되며, 각 국면에 따라 추정 계수가 산출되는 방법이다(Maciejowska et al., 2016).

비선형(Non-linear) AR 모델 종류인 SNAR(Semiparametric autoregressive)과 ANN 계열인 NAR(Non-linear autoregressive)도 포함하였다. SNAR은 Cao et al.(2003)에 의해 제안된 방법으로 비모수 시계열 방법의 일종이다.

B.2. 일반 회귀 모델(General regression models)

먼저 OLS는 고전 회귀 모델에서 가장 보편적으로 활용되는 방법이다. 기본 가정{선형성(Linearity), $E(\epsilon_i)=0, var(\epsilon_i)=\sigma^2, Cov(\epsilon_i,\epsilon_j)=0$

등}을 만족하면 BLUE(Best linear unbiased estimator)에 가까운 추정치를 얻을 수 있다.

GAM(Generalized additive model)은 비선형 관계를 모델화하는데 유용하며(Dąbrowski & Adamczyk, 2010), 본 연구에서는 토마토 가격 및 반입량의 비선형 관계를 고려해 GAM을 회귀 모델로 선택하였다.

모델에 투입되는 예측 변수의 수가 증가할수록 다중공선성(Multicollinearity) 발생 가능성이 높아지며, 이는 모델의 예측 성능을 저하할 수 있다(Melkumova & Shatskikh, 2017). 특히 탐색적 연구에서 예측 변수 간 높은 상관관계가 존재할 경우 추정량에 편의가 발생할 수 있는데, 이러한 상황에서 유용한 방법이 Ridge와 Lasso regression이다.

B.3. 데이터 마이닝 & 기계학습(Data mining & Machine learning models)

ANN은 입력층(Input layer)과, 은닉층(Hidden layer), 출력층(Output layer)으로 된 네트워크 구조로 식 (B2)로 나타낸다. 이를 통해 손실함수(Loss function)를 최소화하는 매개변수(Parameter)를 찾는 방법이다(Wang et al., 2019).

$$O(X_t) = f_2\left(\sum_{j=1}^{m} w_j f_1\left(\sum_{i=1}^{n} v_{ji} x_{ji}\right)\right) \tag{B2}$$

X_t는 인공신경망 입력층에 투입되는 예측 변수를 의미하며, w_j와 v_{ji}는 은닉층에서 학습되는 가중치(Weights)를 나타낸다. f_2와 f_1는 활성화 함수(Activation function)로 시그모이드(Sigmoid), 선형(Linear), RELU 등 다양한 함수가 사용되며, $O(X_t)$는 입력층에 투입된 예측 변수를 기반으로 생성되는 출력물로, 여기서는 주별 가락도매시장 토마토 가격 예측치가 된다. ANN은 역전파오류(Backpropagation)를 사용해 w_j와 v_{ji}를 학습하며, 이 과정에서 미분

을 사용한다. 미분에 의한 최적화 과정에 따라 출력물 $O(X_t)$가 달라질 수 있으며, 본 연구에서는 최적화 방법 중 보편적으로 활용되는 Rprop+, Rprop−, GD(Gradient decent), Adam, AdaGrad 등 다섯 가지 방법을 적용하였다.

SVR은 ANN과 더불어 대표적인 기계학습 방법으로, 커널 트릭(Kernel trick)으로 원 데이터를 고차원 공간으로 선형변환하는 과정을 통해 목표 변수를 잘 표현할 수 있게 하는 모델이다(Smola & Schölkopf, 2004). 커널을 통해 데이터를 변환하고 목표 변수를 잘 표현하는 서포트 벡터(Support vector)를 찾은 후, 실제 데이터와의 차이를 최소화한다. 최상의 주별 가락도매시장 토마토 가격 예측치를 얻기 위해 서포트 벡터와의 거리 $\| w \|$를 최소화해야 한다.

$$\begin{aligned} &Minimize\ \frac{1}{2} \| w \|^2 \\ &subject\ to\ \begin{Bmatrix} P_i - (w, x_i) - b \le \epsilon \\ (w, x_i) + b - P_i \le \epsilon \end{Bmatrix} \end{aligned} \tag{B3}$$

Slack variable을 추가해 식 (B3)과 같이 최적화를 하게 되고, 식 (B4), (B5)와 같이 최적해가 도출된다..

$$\begin{aligned} &Minimize\ \frac{1}{2} \| w \|^2 + C\sum_{i=1}^{l}(\xi_i + \xi_i^*) \\ &subject\ to\ \begin{Bmatrix} P_i - (w, x_i) - b \le \epsilon + \xi_i \\ (w, x_i) + b - P_i \le \epsilon + \xi_i^* \\ \xi_i, \xi_i^* \ge 0 \end{Bmatrix} \end{aligned} \tag{B4}$$

$$L = \frac{1}{2} \| w \|^2 + C\sum_{i=1}^{l}(\xi_i + \xi_i^*) - \sum_{i=1}^{l}\lambda_i(\epsilon + \xi_i - P_i + (w, x_i) + b) -$$
$$\sum_{i=1}^{l}\lambda^*(\epsilon + \xi_i^* + P_i - (w, x_i) - b) - \sum_{i=1}^{l}(\eta_i\xi_i + \eta_i^*\xi_i^*)$$
$$\frac{\partial L}{\partial b} = \sum_{i=1}^{l}(\lambda_i^* - \lambda_i) = 0 \qquad \text{(B5)}$$
$$\frac{\partial L}{\partial w} = w - \sum_{i=1}^{l}(\lambda_i^* - \lambda_i)x_i = 0$$
$$\frac{\partial L}{\partial \xi_i^*} = C - \lambda_i^* - \eta_i^* = 0$$

본 연구에서는 커널 함수로 RBF(Radial basis function), 선형함수, 다항함수 등 세 가지 방법을 사용하였다.

이 외에도 의사결정나무(Decision tree) 모델도 사용하였다. 의사결정나무 모델은 계산 효율성이 높아 Model stacking과 같은 앙상블 모델에 보편적으로 활용되는 모델이다(Oshiro et al., 2012). 또한, 복원 샘플링을 기반으로 다수의 나무 모델을 추정하는 즉, Bagging 기법을 적용한 방법도 존재하는데, RF와 GBM(Gradient boosting model)이 이러한 유형의 대표 방법이다. 그 외, Lessmann & Voß(2017)의 앙상블 모델에서도 적용된 바 있는 kNN(k-Nearest neighbor)도 추가로 본 연구에 적용하였다.

B.4. 변수 선정(Feature selection) 방법

BMA는 k개의 설명변수 중 $2k$개의 모델을 구성한 후, 각 모델의 PMP(Posterior model probability)를 가중치로 활용한 모수의 PM(Posterior mean)을 산출하는 방법이다(남민호 & 김현태, 2014). 일반적으로 BMA은 각 모델의 우도 함수(Likelihood function)를 기반으로 PMP를 산출하기 때문에 각 모델에 대한 단순 평균법보다 최종 모수의 편의(Bias) 발생 가능성이 낮다(남민호 & 김현태, 2014). BMA에서는 PMP(Posterior model probability)를 먼저 산출해야 하는데, 식 (B6)와 같다.

$$p(M_i|D) = \frac{f(D|M_i)p(M_i)}{P(D)} = \frac{f(D|M_i)p(M_i)}{\sum_{i=1}^{2k} f(D|M_i)p(M_i)} \quad \text{(B6)}$$

D는 모델에서 활용한 자료, k는 설명 변수의 수, M_i는 특정 설명 변수 조합으로 구성된 모델 i를 의미한다. 좌변에서 분자에 위치한 $p(M_i)$는 각 모델의 사전 확률(Prior probability)를 의미하는데, 모델 M_i가 좋은 모델일 가능성을 담고 있다. 베이지안 기본 원리에 따라 사전 확률 $p(M_i)$는 모델의 수렴이 진행될수록 개선(Update)되는 특성을 지니고 있다(Forte et al, 2018). $f(D|M_i)$는 한계우도함수(Marginal likelihood function)로 식 (B7)과 같다.

$$p(D|M_i) = \int f(D|\theta_i, M_i)p(\theta_i|M_i)\, d\theta_i \quad \text{(B7)}$$

여기서 다시 모델 M_i에 포함된 모수(Parameter) θ_i의 PPD는 식 (B8)과 같이 정리된다.

$$p(\theta_i|D, M_i) = \frac{f(D|\theta_i, M_i)p(\theta_i|M_i)}{f(D|M_i)} \quad \text{(B8)}$$

위 과정을 통해 산출된 추정치를 기반으로 BMA의 핵심 중 하나인 모수 θ_i PM(Posterior mean)과 PV(Posterior variance) 산출할 수 있으며, 식 (B9), (B10)으로 정리된다. PM은 각 모델에서 산출된 모수 추정치의 가중평균(Weighted mean)으로, 가중치 역할은 PMP(Posterior model probability) $P(M_i|D)$가 활용된다. 사전 확률에 대한 설정은 대부분의 BMA에서 활용하는 Zellner's g 분포를 활

용하였다(최수진 & 김한호, 2017).

$$E(\theta|D)=\sum_{i=1}^{2^k}\hat{\theta}_i p(M_i|D) \tag{B9}$$

$$Var(\theta|D)=\sum_{i=1}^{2^l}p(M_i|D)\,Var(\theta|D,M_i)+ \sum_{i=1}^{2^l}p(M_i|D)\left[E(\theta|D,M_i)-E(\theta|D)\right]^2 \tag{B10}$$

마지막으로 변수 선택에 대한 문제가 있는데, PM 및 PV 외 특정 변수가 특정 모델에 포함될 확률의 합인 PIP(Posterior inclusion probability)를 활용하였다. PIP가 높은 변수는 반응변수에 대해 높은 영향력을 지니고 있음을 의미한다(남민호 & 김현태, 2014). PIP를 기준으로 선정된 각 모수 θ_i의 점 추정치(Point estimator)를 기반으로 주별 토마토 가락도매시장 가격에 대한 함수를 구성하였다.

RMCFS는 의사결정나무의 대표 알고리즘 중 하나인 C5 알고리즘 기반의 변수 선정 방법으로, 2005년 개발된 첫 알고리즘이 분류(Classification) 문제에 적합한 알고리즘이었다면, Draminski & Koronacki(2018)가 이를 연속형 타겟 변수(Continuous target variable)에도 적합한 형태로 개선하였다. 몬테카를로(Monte Carlo) 기반의 샘플링 기법이 적용된 만큼, 본 연구와 같이 샘플 n의 수 대비 예측 변수 p의 수가 많은 상황에서 적절한 방법이다(Draminski & Koronacki, 2018). 그림 B1.은 변수 선정 방법 RMCFS의 기본 구조를 나타내고 있다.

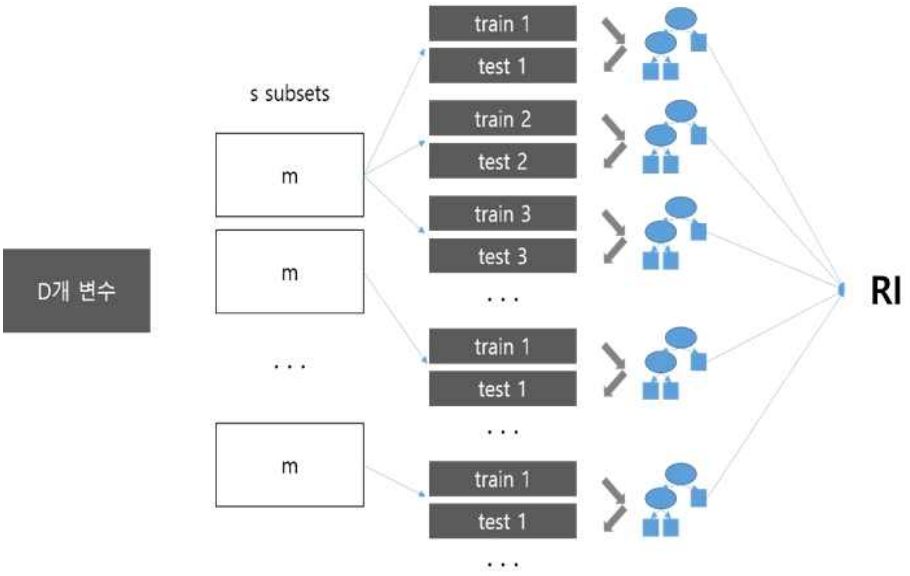

그림 B1. RMCFS의 기본 구조

d개의 예측 변수 후보군을 기준으로 s개로 분할(Subset)된 m이 생성되고, 이에 대해 무작위(Random)로 후보 예측 변수가 분할된 s개셋으로 할당된다. 분할된 m에 무작위로 할당된 각 예측 변수 조합은 다시 무작위로 생성된 학습데이터(Training set) 조합에 할당되고, 사용자가 지정한 분할(Split) 기준에 따라 모델이 구성된다. 즉, s(분할된 예측 변수 조합)$*t$(나무 모델)가 산출되며, 각 모델의 결과를 종합한 변수 중요도는 식 (B11)과 같이 산출된다.

$$RI_{gk} = \sum_{\tau}^{s \cdot t} wAcc_{\tau}^{\mu} \sum_{\eta_{g_k}(\tau)} GR\left(\eta_{g_k}(\tau)\right)\left(\frac{no.\eta_{g_k}(\tau)}{no.(\tau)}\right)^{v} \qquad \text{(B11)}$$

s(분할된 예측 변수 조합)$*t$(나무 모델)에서 발생한 결과를 기반으로 특정 예측 변수 중요도 RI_{gk}가 산출된다. τ번 모델의 η_{g_k}는 n

번 노드(Node) 분기점을 형성하는 변수로 중요한 예측 변수로 간주할 수 있다. $wAcc_{\tau}^{\mu}$는 τ번 모델의 가중치가 적용된 정확도(MAPE, RMSE 등)를 의미하며, $GR(\eta_{g_k}(\tau))$는 η_{g_k}의 Gain ratio를 의미한다. 산출된 중요도를 기반으로 다시 중요 변수를 구분하는 임계점(Threshold)이 필요한데, Critical angle, k-means, Max RI based Permutations, z-score based Permutations, Contrast attributes 등 5가지 방법 중 하나를 선택하면 된다. 여기서는 각 모델에서 최고 중요도($Max\,RI$)로 선정된 경우가 n번 이상을 만족하는 경우에만 주요 영향력 변수로 선정하는 Max RI based Permutations을 적용하였다.

B.5. PCA(Principal component analysis)

PCA는 탐색적 데이터 분석 및 예측모델 개발에 주로 사용되는 다변량 분석을 위한 통계적 방법으로(Shah et al., 2021), 관측 데이터 내 상관관계가 존재하는 변수 집합을 직교 변환(Orthogonal transformation)을 통해 상관관계가 없는 주성분(PC, Principal component)으로 변환시켜 차원을 축소한다(Abdi & Williams, 2010). 원 데이터에서 선형 관계를 형성하던 변수 집합은 저차원 공간인 주성분에 기하학적으로 투영되고(Geometrically projecting), 축약된 주성분은 원 데이터에 대한 최상의 요약이 가능해 진다(Lever et al., 2017). PCA의 차원 축소는 분산을 기준으로 이루어지며 모든 주성분은 서로 직교한다.

B.6. Quantile regression

Quantile regression은 분위 τ별 별도의 파라미터를 추정하는 회귀분석 모델에 해당한다(Ordiano et a., 2020). 식 (B12)과 같이 표현되며, $\hat{y}_{(q)}$는 τ별 추정된 추정치를, $\hat{\theta}$는 τ별 추정량의 의미한다. 본 연구에서는 농산물 가격의 높은 변동성을 고려해 Quantile regression을 H-FQMA의 마지막 단계에서 최종 예측치를 산출하는

방법으로 활용하였다.

$$\hat{y}_{(q)} = f\left(x;\hat{\theta}_{(q)}\right) \tag{B12}$$

부록 C. 주요 단일 모델 예측 성능 평가

H-FQMA의 예측 성능 평가를 보강하기 위해 단일 모델과의 성능 비교를 추가하였다. 본 분석에서는 RMCFS, BMA를 활용한 변수 선정 절차를 생략하고, 기본 시차 변수만을 활용해 모델을 구성하였다. 이를 통해 H-FQMA의 RMCFS, BMA를 통한 변수 선정 결과의 타당성을 검증하고, 더 나아가 단일 모델과의 성능을 비교 분석하는 것이 목적이다. 적용 방법으로는 기계학습 대표 방법인 ANN, SVR과 대표적인 단변량 시계열 모델 3종 AR(1), AR(2), AR(3)을 적용하였다. ANN 및 SVR에 적용된 예측 변수는 아래 식 (C1)과 같다.

$$X = \begin{bmatrix} P_{garak,t-1}, P_{garak,t-2}, P_{garak,t-3}, Q_{garak,t-1}, \\ Q_{garak,t-2}, Q_{dae,t-1}, Q_{buban,t-1}, Q_{buum,t-1}, \\ Q_{tratio,t-1}, Q_{tratio,t-2} \end{bmatrix} \quad \text{(C1)}$$

식 (C1)에서 1 ~ 3주 전 가락도매시장 가격은 각각 $P_{garak,t-1}$, $P_{garak,t-2}$, $P_{garak,t-3}$, $Q_{garak,t-1}$, 1 ~ 2주 전 가락도매시장 반입량은 $Q_{garak,t-2}$, 1주 전 대구북부도매시장 반입량은 $Q_{dae,t-1}$, 1주 전 부산반여도매시장 반입량은 $Q_{buban,t-1}$, 1주 전 부산엄궁도매시장 반입량은 $Q_{buum,t-1}$, 1 ~ 2주 전 전체 도매시장 반입량 중 가락도매시장 반입량 비율에 대한 시차 변수는 각각 $Q_{tratio,t-1}$, $Q_{tratio,t-2}$이다.

먼저, 주어진 예측 변수를 기반으로 선형 회귀를 통해 변수의 유의성 및 영향력을 확인하였다. 분산팽창요인(Variance inflation factors) 검증을 통해 적용된 예측 변수의 활용 가능(다중공선성(Multicollinearity) 존재 여부) 여부를 확인하였다. 분산팽창요인 검증 결과에서는, 다중공선성을 유발할 수 있는 높은 위험도를 지닌 변수는 없는 것으로 나타났다(표 C1. 참고).

표 C1. 분산팽창요인을 활용한 다중공선성 검증 결과

Garakp(t-1)	Garakp(t-2)	Garakp(t-3)	Garaki(t-1)	Garaki(t-2)
8.38	6.84	3.62	3.76	2.96
Daei(t-1)	Bubani(t-1)	Buumi(t-1)	Tratio(t-1)	Tratio(t-2)
2.35	7.65	9.76	1.94	1.75

FGLS(Feasible generalized squares)를 활용해 모델을 구성하였고, 추정한 결과는 표 C2.와 같다.

표 C2. FGLS를 활용한 추정 결과

변수	Estmate	t	p-value
Garakp(t-1)	0.913	9.422	0.0000
Garakp(t-2)	-0.448	-4.727	0.0000
Garakp(t-3)	0.272	4.339	0.0000
Garaki(t-1)	0.172	1.660	0.0986
Garaki(t-2)	-0.027	-0.294	0.7693
Daei(t-1)	-1.817	-4.113	0.0000
Bubani(t-1)	-22.04	-0.149	0.8819
Buumi(t-1)	165.721	0.976	0.3305
Tratio(t-1)	-12.86	-1.654	0.0998
Tratio(t-2)	16.95	1.785	0.0759

유의성에 여부와 무관하게 10개 예측 변수를 준비한 5개 방법에 적용한 결과는 표 C3과 그림 C1. ~ C3.에 나타나 있다.

표 C3. 5개 모델의 예측 성능 결과(RMSE, MAPE, R-squared)

모델	RMSE(원)	MAPE(%)	R-Squared
ANN	515	15.9	0.740
SVR	634	16.8	0.727
AR(1)	509	15.9	0.743
AR(2)	502	15.7	0.745
AR(3)	503	15.5	0.745

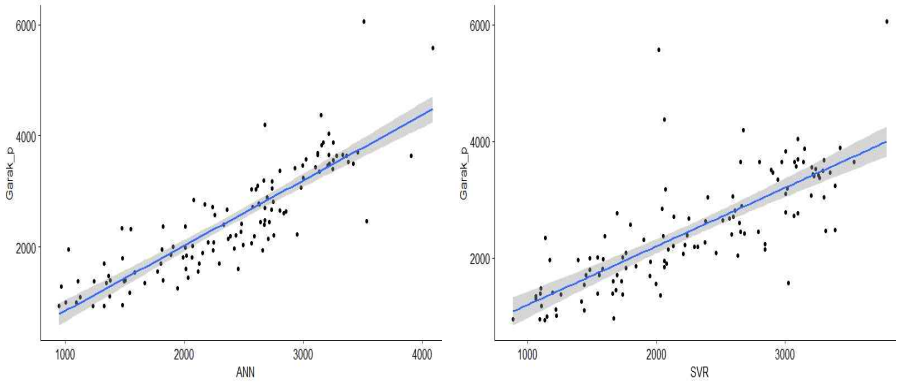

그림 C1. 예측치 및 실제값의 선형 관계(ANN 및 SVR)

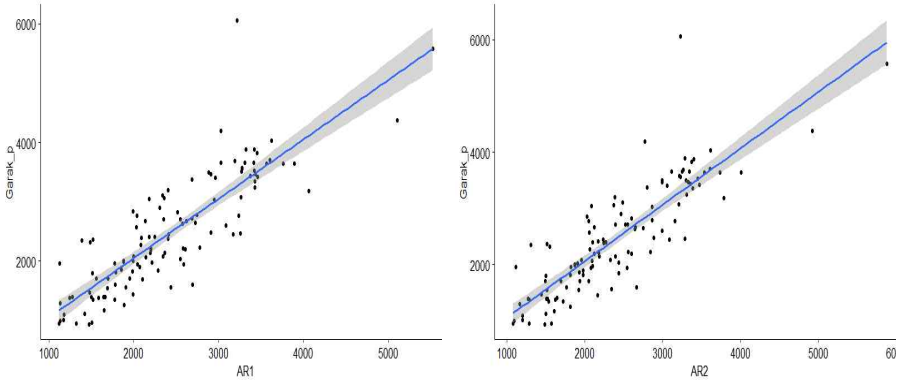

그림 C2. 예측치 및 실제값의 선형 관계(AR(1) 및 AR(2))

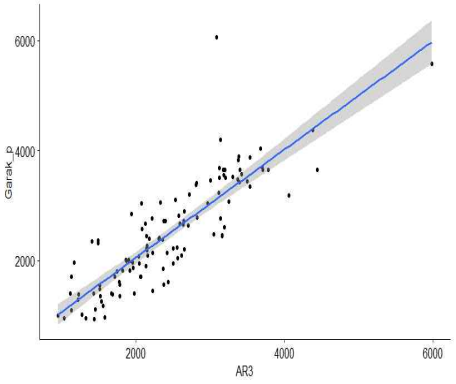

그림 C3. 예측치 및 실제값의 선형 관계(AR(3))

SVR 외 네 개 모델의 예측 성능은 비슷한 수준으로 나타났다. 앞서 제2장에서 언급된 H-FQMA에서 발생한 네 개 모델(AR, BM, RF, NOW)의 평균 예측 성능은 RMSE가 383, MAPE가 12.7%, R-Squared 0,917로 나타났었다.

그 외, ANN과 SVR에 적용된 초매개변수는 표 C4. ~ C5.와 같다. ANN의 최적화 방법은 Adam을, SVR의 커널(Kernel)은 RBF를 적용하였다.

표 C4. ANN에 적용된 초매개변수 수치

Optimi-zation	Layer 1	Layer 2	Learning rate	Threshold	Loss function
Adam	13	22	0.2	0.037	SSE

표 C5. SVR에 적용된 초매개변수 수치

Kernel	Model type	Sigma	Cost	Epsilon	Tolerance
RBF	nu-regression	0.085	35	0.234	0.0065

Appendix D. A 연합사업단 출하 효율성 분석

BMA를 통해 추정된 주별 가락도매시장 토마토 가격 모델을 토대로 A 연합사업단의 출하 전략 효율성을 분석하였다.

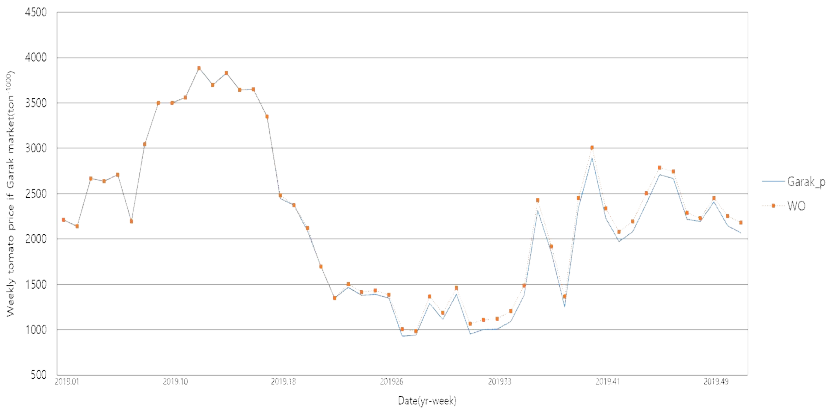

그림 D1. 실제 토마토 가격(Garak_p) 및 A 연합사업단의 도매 출하량 조절을 통한 가락도매시장 가격 상승 최대치(WO)(2019년 기준)

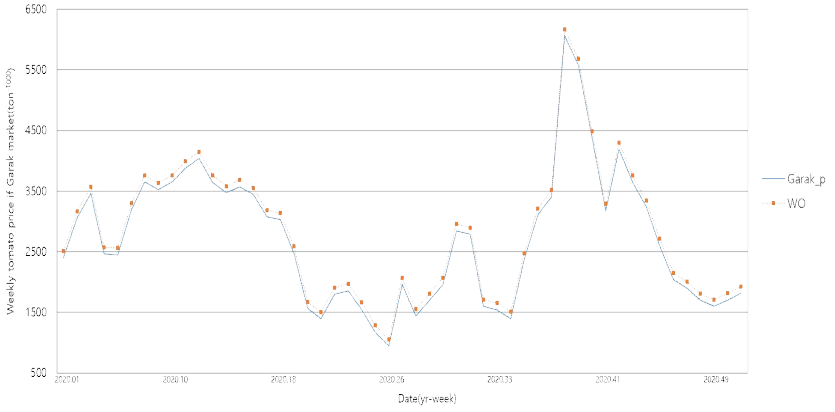

그림 D2. 실제 토마토 가격(Garak_p) 및 A 연합사업단의 도매 출하량 조절을 통한 가락도매시장 가격 상승 최대치(WO)(2020년 기준)

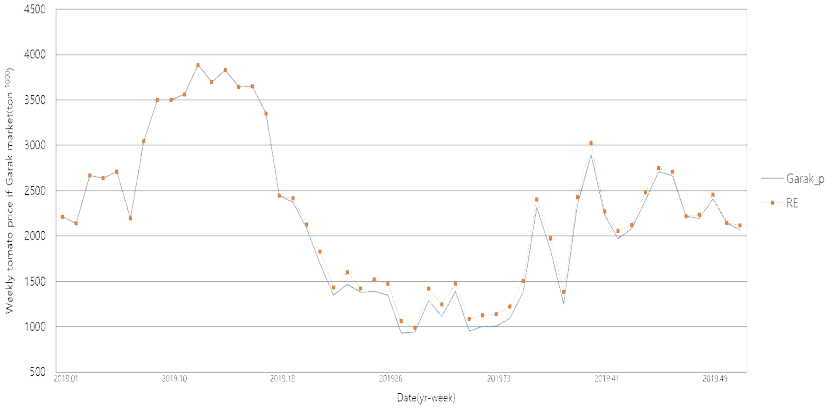

그림 D3. 실제 토마토 가격(Garak_p) 및 A 연합사업단의 소매 출하량 조절을 통한 가락도매시장 가격 상승 최대치(RE)(2019년 기준)

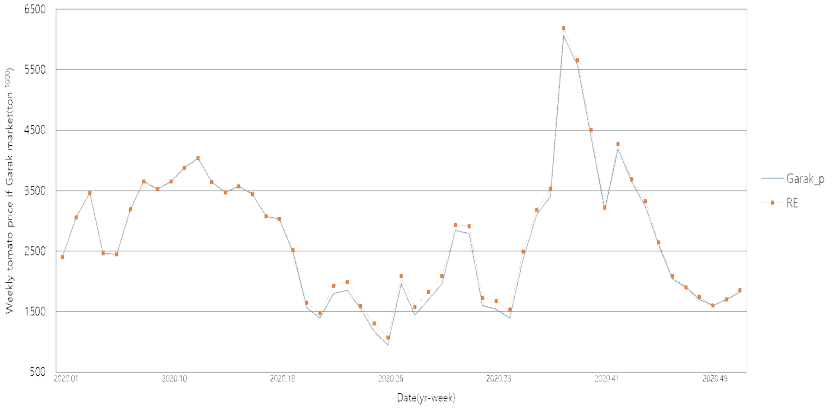

그림 D4. 실제 토마토 가격(Garak_p) 및 A 연합사업단의 소매 출하량 조절을 통한 가락도매시장 가격 상승 최대치(RE)(2020년 기준)

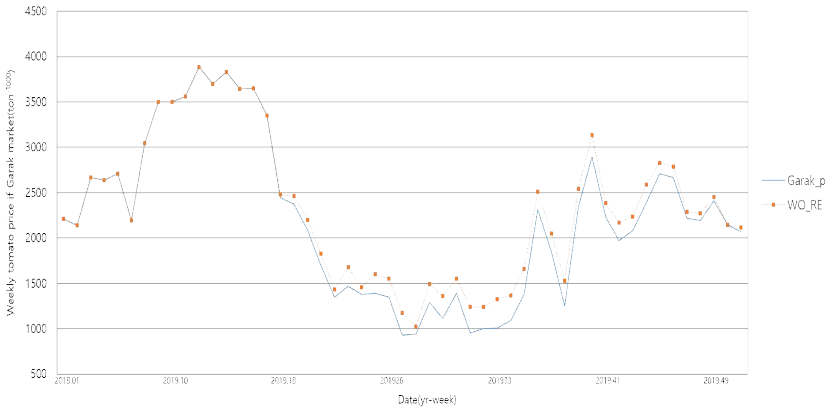

그림 D5. 실제 토마토 가격(Garak_p) 및 A 연합사업단의 도소매 출하량 조절을 통한 가락도매시장 가격 상승 최대치(WO_RE)(2020년 기준)

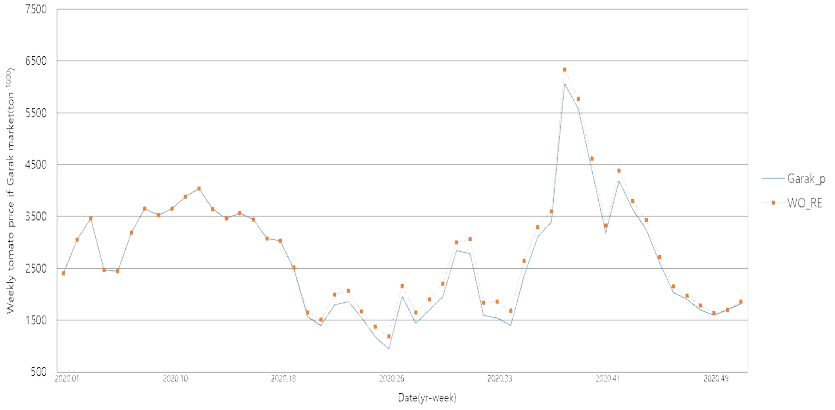

그림 D6. 실제 토마토 가격(Garak_p) 및 A 연합사업단의 도소매 출하량 조절을 통한 가락도매시장 가격 상승 최대치(WO_RE)(2020년 기준)

그림 D1. ~ D6.는 2019 ~ 2020년 주별 가락도매시장 토마토 실제 가격과 A 연합사업단의 출하량(도매, 소매, 도매 및 소매) 조절 가능 범위 내에서 가락도매시장 가격이 가장 높아지는 수치를 나타낸 도표이다. 도표에서 'WO'는 A 연합사업단의 도매시장 출하량, 'RE'는 A 연합사업단 소매 출하량, 'WO_RE'는 도매 및 소매 출하량을

합한 것을 의미한다. A 연합사업단의 시장 출하량이 확대되는 3 ~ 9월 기간에 출하량 조절을 통해 가락도매시장 가격을 상승시킬 수 있는 범위가 확대되는 것을 확인할 수 있다. 특히 그림 D5 ~ D6에 나타난 것처럼, 도매 및 소매 출하량 모두를 조절할 수 있을 경우, 가락도매시장 가격 상승의 변화폭은 넓어진다.

마지막으로 표 D1.은 시장 출하량이 확대되는 3 ~ 9월 기간 가락도매시장 토마토 실제 가격 평균과 A 연합사업단 출하량 조절을 통한 가락도매시장 가격의 최대치 평균을 정리한 자료이다. 'Actual'은 해당 연도 3 ~ 9월 기간의 실제 가락도매시장 토마토 가격 평균, '도매'는 A 연합사업단 도매 출하량 조절을 통해 형성 가능한 가락도매시장 토마토 가격 최대치 평균, '소매'는 A 연합사업단 소매 출하량 조절을 통해 형성 가능한 가락도매시장 토마토 가격 최대치 평균, '도소매'는 A 연합사업단 도매 및 소매 출하량 조절을 통해 형성 가능한 가락도매시장 토마토 가격 최대치 평균을 의미한다. 2019년의 경우, A 연합사업단 출하량 조절을 통해 가락도매시장 토마토 가격 상승폭은 도매<소매<도매/소매 순으로 나타났던 반면, 2020년에는 소매<도매<도매/소매 순으로 나타났다.

표 D1. 가락도매시장 실제 평균 가격 및 A 연합사업단의 출하량 조절을 통한 가락도매시장 가격 상승 최대치 평균(3 ~ 9월 기준)

연도	실제값	도매	소매	도소매
2019	1,711	1,785	1,804	1,896
2020	2,559	2,670	2,657	2,749

Abstract

Seoul National University, CALS
AGRICULTURAL & RESOURCE ECONOMICS
Minje Jeong

The purpose of this study is to develop a forecasting model for weekly tomato price of Garak wholesales market, which is the largest agricultural wholesale market in Korea, and a decision-making support tool for agricultural products distribution organization. In the agricultural market shipments of domestic agricultural product distribution organizations(e.g. NH), weekly contracts are predominant. in addition, transaction prices with retailers are being formed with reference to Garak wholesales market. Therefore, forecasting information required for these distribution organizations is weekly, not monthly, and it is necessary to develop a forecasting model for weekly tomato price of Garak wholesales market. In order to develop a forecasting model in the weekly Garak wholesales market, first, a literature review is conducted on the statistical methods for forecasting prices of agricultural products in foreign and domestic markets. Through the above review, the status of statistical methods applied in foreign studies and gaps with domestic studies were analyzed. Next, Hybrid FQMA improved by applying a feature selection method and additional statistical methods to the existing FQMA(factor quantile model averaging) was proposed, and then the forecasting performance was evaluated. Hybrid FQMA is a method corresponding to a forecasts combination in a prediction method. The frequency of application of forecasts combination is also increasing in overseas researches for forecasting agricultural product price. The forecasting performance of Hybrid FQMA recorded the highest 11.2% based on MAPE, and exceeded other benchmarking models.
In the development of weekly decision-making support tool for

an agricultural product distribution organization, research is conducted using internal data of the Gangwon Association, one of decision-making organizations for the market shipment within the NH, and data from the wholesale market. The purpose of this study's decision support tool is to develop a web-based tool that enables an agricultural product distribution organization with market influence to analyze the price movement of Garak wholesales market, according to market conditions and internal shipping strategies. For the development of the decision support tool, BMA(Bayesian model averaging) was used to estimate the influence of tomato imports from the Gangwon Association and major wholesale markets on tomato price in Garak wholesale market. Next, a model that extracts the number of cases maximizing price of Garak market in a given environment is developed by applying PSO(Partical Swarm Optimization). Using these results, a web-based decision support tool was developed to enhance the practical use of the research.

Keyword : FQMA, Forecasts combination, Agricultural products price, BMA(Bayesian model averaging), PSO(Particle swarm optimization)

Student ID : 2016-30388