

결정적 및 확률적 성분과 수준변수를 활용한

SMP 예측 성과 비교

Comparison of SMP Forecasting Performance: Deterministic-Stochastic Decomposition vs. Level Variables

성신여자대학교 경영학과 이현석

요약

성분 분해에 의한 가격 예측은 예측에 영향을 미치는 다양한 요인을 제시하기 때문에 예측 정확성을 높여 주기도 하며, 가격 변화에 대응하는 전략 수립도 용이하게 해준다. 그러나 수준변수에 비해 분석 단계가 많고 복잡하여 많은 연구가 이루어지지 않았다. 최근에는 컴퓨터 기술과 예측 모형의 발전으로 성분 분해를 이용한 가격 예측이 점차 증가하고 있다.

본 연구는 전력거래소의 계통한계가격(System Marginal Price; SMP)을 대상으로 수준변수와 성분 분해의 예측 성능을 비교하였다. 일중 평균 SMP를 스플라인 함수와 캘린더 효과 등을 바탕으로 결정적 성분과 확률적 성분으로 분해하였으며, 지수평활 모형을 사용하여 결정적 성분에 대한 예측을 수행하였다. SMP를 수준변수로 사용하는 경우에는 ARDL과 LSTM 모형 등을 사용하여 예측을 수행하였다. 성분 분해의 경우에는 AR에 점프 혹은 GARCH 계열 모형 등을 결합하여 확률적 성분을 예측하고, 이를 결정적 성분의 예측과 결합하여 예측을 수행하였다.

실증 결과는 성분 분해를 통한 SMP의 예측이 수준변수를 사용한 경우에 비해 높은 예측 성능을 갖고 있음을 보여주었다. DM 통계량 기준으로 확률적 성분을 AR과 점프를 결합해서 분석한 모형이 가장 높은 예측 성능을 보였으며, 그 다음으로 SMP와 에너지 가격의 공적분 관계를 이용한 ARDL 모형의 예측 성과가 좋았다. 확률적 성분을 GARCH 계열 모형으로 예측한 경우는 점프 모형으로 예측한 경우보다 성능이 낮게 나왔다. 수준변수의 경우에는 LSTM이 ARDL 모형에 비해 다소 예측 성능이 낮았다. 연구 결과는 수준변수와 성분 분해에 의한 예측의 정확도는 데이터의 특성을 반영한 적절한 분석 모형의 선택이 중요함을 보여주고 있다.

1. 도입

가격은 장기 추세와 단기적 변동성을 동시에 갖는 경우가 많다. 이를 반영하여 장기 추세와 단기적 변동성을 각각 결정적(deterministic) 성분과 확률적(stochastic) 성분으로 분해하여 자산 가격을 예측하려는 연구가 이루어져 왔다. 이러한 연구는 가격을 성분 분해하면 수

준변수를 직접 사용하는 것보다 예측 성능을 우수하게 할 수 있다는 믿음을 바탕으로 하고 있다.

장기 추세는 시간에 따른 구조적 변화를 나타내고, 단기적 변동성은 불확실성이나 외부 충격에 의해 발생한다. 예를 들어, 주식의 경우 기업의 장기 실적이나 산업 동향에 의해 장기적인 추세가 형성되지만, 단기적으로는 시장 심리, 뉴스, 정치적 사건 등으로 변동성이 발생한다. 환율 또한 장기적으로는 각 국가의 경제 상황, 무역수지, 금리 차이 등이 영향을 미치지만, 단기적으로는 정치적 불확실성, 중앙은행 개입, 국제적 사건 등이 변동성을 유발한다. 이러한 분석은 금융 자산(Choudhry(1994), Arias-Calluari, Najafi, Harre, Tang & Alonso-Marroquin(2022)) 외에도 원자재(Lee, List & Strazicich(2006)), 부동산(Fan, He, Hao & Wu (2024))과 같은 실물 자산 및 에너지 등의 가격으로도 확장할 수 있다.

가격을 결정적 및 확률적 성분으로 분해하면 가격 변화에 대한 대응 전략을 쉽게 구축할 수 있다. 예를 들어 장기 투자의 경우에는 자산 가격의 결정적 성분에 초점을 맞추어 투자 전략을 수립할 수 있으며, 위험 관리가 목표인 경우에는 자산 가격의 확률적 성분이 갖는 확률 분포에 초점을 맞출 수 있다. 또한 가격을 성분으로 분해하면 가격에 영향을 미치는 다양한 요인을 파악하고, 이를 장기와 단기 요인으로 나누어 보다 정교한 예측을 할 수 있다.

그러나 이러한 장점에도 불구하고, 성분 분해에 의한 예측에는 몇 가지 한계가 존재한다. 첫째, 두 성분을 분리하는 과정에서 분석 모형이 복잡해지고, 이로 인해 모형의 설정 및 해석이 어려워질 수 있다. 둘째, 확률적 성분은 본질적으로 불확실성을 내포하고 있어, 그 예측이 여전히 어렵다. 또한, 결정적 성분인 추세도 시간이 지남에 따라 변화할 수 있어, 모형의 안정성을 유지하기 어렵다. 이러한 문제들로 인해, 과거에는 성분 분해를 통한 가격 예측이 제한적으로 연구되었다. 그러나 최근에는 전통적 시계열 분석에 AI 모델, 베이지안 기법, 비정상 시계열 데이터 처리 기술의 발전이 결합되면서, 금융 및 실물 자산의 구조적 변동성과 불확실성을 더 정밀하게 설명하고 예측하는 방법론이 등장하고 있다.

본 연구는 가격 예측을 할 때, 수준변수를 그대로 사용하는 경우와 성분을 분해하는 경우를 각각 비교하여 그 성능을 측정하였다. 이를 위해 전통적 통계 모형과 딥러닝 모형 등 다양한 모형을 사용하였으며, 통계적 가정의 성립 여부, 모형의 설명력 등 모형의 유용성 측면도 고려하였다.

본 연구는 가격 성분의 구성 요소가 주식과 같은 자산 가격에 비해 비교적 단순한 일별 평균 계통한계가격(SMP; System Marginal Price)을 분석 대상으로 하였다. 전력거래소(KPX; Korea Power Exchange)는 다음 날의 1시간 단위 전력 수요를 예측한 후, 발전 변동비용이 낮은 발전사업자를 우선하여 전력 공급 입찰을 받는다. 입찰 마감 직후 다음 날 각 시간대의 전력 수요를 충족시키는 마지막 발전사업자가 결정되면, KPX는 이 발전사업자의 변동비용을 기준으로 해당 시간대의 모든 발전사업자에게 동일한 보상을 지급한다.

일반적으로 SMP는 전력 수요와 발전 비용의 두 가지 요인에 의해 결정된다. 전력 수요는

계절적 요인, 주중, 휴일, 일중 등의 캘린더 효과의 영향을 받고 있다. 또한 전기는 저장이용이하지 않기 때문에 수요와 생산의 차이가 클수록 SMP의 변동성은 매우 커진다. 전력 수요가 많더라도 LNG나 중유의 가격이 낮으면 SMP는 낮아진다. 예를 들어 여름에는 에어컨과 같은 냉방용품의 수요 증가로 전력 수요가 많지만 난방에 들어가는 석유 제품의 수요가 많지 않기 때문에 유가는 낮아진다. 이로 인해 여름은 겨울에 비해 SMP가 낮아지는 경우가 많다. 전력 수요와 발전 비용은 계절적 요인과 장기 추세라는 특성을 모두 갖고 있으므로 SMP에서 이들과 관련한 결정적 성분을 쉽게 분리할 수 있다.

본 연구는 SMP의 움직임에 수준변수와 성분 분해 측면에서 분석한 다음, 다양한 모형을 사용하여 예측 성과를 평가하려고 한다. 이는 다음과 같은 단계로 수행되었다.

첫째, SMP와 밀접한 관련을 가질 것으로 생각되는 에너지 가격, 기후, 캘린더 효과 등의 자료를 바탕으로 SMP와의 관련성을 분석한 후, 예측 모형에 포함시킬 것인지를 결정하였다.

둘째, SMP 수준변수를 결정적 성분과 확률적 성분으로 분해하는 작업을 수행하였다. 이를 통해 SMP에서 결정적 성분을 분리할 수 있기 때문에 확률적 성분에 대한 예측만 필요하므로 SMP의 예측 정확성이 상승할 수 있다. 그러나 결정적 성분 중에서도 '장기 추세'와 같은 요소는 예측해야 하는 부분으로 이를 예측하는 과정에서 불확실성이 개입할 수 있다.

셋째, SMP의 수준변수와 각 성분에 적합한 다양한 모형을 사용하여 예측 성과를 측정하였다. 자료의 시계열 안정성과 각 모형에 포함된 변수의 유의수준, 모형의 설명력 등을 검토하여 예측을 수행하였다.

넷째, 각 모형의 예측 결과는 평균절대오차(Mean Absolute Error, MAE), 평균 제곱 백분율 오차(Mean Squared Percentage Error, MSPE), 평균 절대 백분율 오차(Mean Absolute Percentage Error, MAPE)와 DM(Diebold-Mariano) 테스트 등에 의해 평가를 수행하였다.

컴퓨터 기술의 발전과 고급 통계모형, 인공지능모형 등의 개발로 인해 전기가격의 예측은 경제학이나 경영학보다는 공학 분야에서 많이 이루어져 왔다. 공학에서는 매우 많은 모형을 동시에 분석하여 예측 성능을 비교하거나 최신 통계 모형 등을 사용하여 전기가격 예측을 수행하고 있다. 그러나 이들은 주로 각 모형의 예측 성과에만 초점을 맞추고 있어 모형의 적합성 및 데이터의 특성에 대한 고려가 부족한 경우가 많다.

본 연구는 최신 통계 모형이나 머신 러닝 모형을 사용하기 보다는 데이터의 특성에 초점을 맞추어 SMP의 예측에 적합한 모형을 선택하였다. 선택한 모형의 과적합을 방지하고, 모형이 데이터를 잘 설명하는지 확인하기 위해 교차 검증(cross-validation) 검사 등을 사용하여 모형의 성능을 개선하였다.

본 연구는 다음과 같이 구성되었다.

제1장의 도입에 이어 제2장에서는 이론적 배경으로 가격의 성분 분해에 대한 기존 연구 및 SMP 예측 관련 기존 연구에 대해 살펴보았다. 제3장에서는 SMP 예측과 관련이 있는 데이터 및 이들의 통계적 특성을 언급하고 있으며, 제4장은 전기가격의 성분 분해 방법에 대해

설명하고 있다. 제5장은 예측에 사용된 다양한 통계적 모형을 소개하고 이들의 추정 및 예측 결과를 보여주고 있다. 제6장은 본 연구의 요약 및 결론을 나타내고 있다.

2. 이론적 배경

전기가격의 수준변수를 사용하여 예측하는 경우에는 회귀모형, VAR, 머신 러닝 및 딥러닝 기법 등이 이용되었다. Castelli, Groznik & Popovič(2020)은 머신 러닝 기법을 수준변수에 적용하여 전기가격을 예측하였다. 이들이 사용한 예측 모형은 기후 조건, 원유 가격, 탄소배출권 가격 등 다양한 변수가 포함되었다. Lago, Marcjasz, Schutter & Weron(2021)는 여러 전기가격 모형의 예측 성능을 비교한 후, 머신 러닝이 가장 높은 정확성을 갖고 있다고 주장하였다.

수준변수를 사용하는 경우 최근 신경망 모형을 사용한 연구가 증가하고 있다. Lehna, Scheller & Herwartz(2022)은 시계열 모형과 신경망모형의 예측 성능을 비교하고 있다. LSTM(Long Short-Term Memory, 장단기 메모리)이 대부분의 예측에서 높은 성과를 보였으며, 단기 예측에서는 VAR(Vector Autoregression) 모형이 유용하다고 주장하였다. 이들은 개선된 성과를 얻기 위해 날씨 데이터, 이산화탄소 배출 등의 외생적 변수를 고려하고, LSTM과 VAR 모형을 결합하는 방식을 제시하였다. Kapoor & Wichitaksorn(2023)은 LSTM, GARCH, Lasso, 확률적 변동성(SV; stochastic volatility) 등 다양한 모형과 기후, 선도 전기가격, 전력 수요 등 많은 특성 변수를 결합하여 일별 전기가격을 예측하였다. 그들은 GARCH와 SV와 같은 전통적 통계 모형이 LSTM, GRU(Gated Recurrent Unit; 게이트 순환 유닛), DNN(Deep Neural Network; 심층신경망)과 같은 딥러닝 알고리즘에 비해 우수한 예측 성능을 갖고 있음을 보여주었다.

전기가격에서 시계열 비정상성을 제거한 수준변수를 사용한 연구도 있다. Billé, Gianfreda, Grosso & Ravazzolo(2023)는 과거 8일 동안의 평균 전기가격을 사용하여 특정 요일이나 특정 시간에 발생하는 계절적 패턴, 일시적 충격이나 주기적 변동과 같은 시계열 데이터의 비정상성을 제거한 수준변수를 사용하여 전기가격을 예측하였다. Poggi, Persio & Ehrhardt(2023)은 Wavelet 변환을 활용하여 전기가격 시계열이 갖는 계절성 및 추세, 비정상적 가격 변동을 제거한 수준변수를 날씨, 경제적 요인 등의 외생변수와 결합하여 분석할 때 보다 높은 예측 성과를 얻을 수 있음을 보여주었다.

전기가격을 결정적 성분과 확률적 성분으로 분해한 연구로는 Escribano, Pena & Villaplana(2011)가 있다. 이들은 외부 충격의 영향을 받지 않고 일정한 주기가 반복되는 변동성을 결정적 성분으로 설정하였고, 주기적이지 않고, 외부 충격에 의해 변동될 수 있는 불확실한 변동성을 확률적 성분으로 설정하였다. Weron(2014)과 Mayer & Trück(2018)도 전기

가격을 예측하기 위해 연간 계절성을 포함하는 장기 추세와 요일 효과 등의 단기 계절성을 결정적 성분으로 나머지를 확률적 성분으로 구분한 후, 각각을 예측하여 결합하는 것이 모형의 해석 가능성과 예측 성능을 개선할 수 있다고 주장하였다. Shah, Bibi, Ali, Wang & Yue(2020)는 결정적 성분을 장기 추세, 연간 및 주별 효과, 일별 효과 등으로 보다 세분화하였으며, 예측 불가능한 단기 변동을 나타내는 확률적 성분으로 분해하였다.

Iftikhar, Turpo-Chaparro, Rodrigues and López-Gonzales(2023)는 결정적 및 확률적 성분의 분해를 위해 스플라인 함수를 사용하였다. 이들은 분해한 각 성분을 자기회귀(AR), 비모수 자기회귀, 자기회귀 이동평균(ARMA), 벡터 자기회귀 모형 등으로 예측한 후, 각 성분의 예측값을 결합하여 전기가격을 예측하고 이의 성과를 비교하였다.

전기가격의 결정적 성분의 추세 예측은 주로 회귀 모형과 스플라인 함수 등을 사용하여 이루어졌다. Weron(2014)은 계절적 요인과 캘린더 효과 등을 독립변수로 사용한 회귀 모형을 통해 결정적 성분의 장기 추세와 캘린더 효과 등의 패턴을 잘 포착할 수 있음을 보여주었다. Shah et al.(2020)은 결정적 성분을 선형 회귀와 스플라인 함수를 사용하여 예측을 하였으며, 모수적 방법인 선형 회귀가 결정적 성분 예측에 보다 적합하다고 주장하였다.

전기가격의 확률적 성분을 추정하기 위해 자기회귀(AR), 이동평균(MA), 점프 모형, GARCH 계열 모형 등 많은 모형이 사용되고 있다. Escribano et al. (2011)은 주중 및 주말, 계절성 등의 캘린더 효과를 통제한 후 확률적 성분이 갖는 변동성 클러스터와 전기가격의 급등락을 GARCH 모형과 점프 모형을 사용하여 분석하였다. Weron(2014)도 점프 모형을 사용하여 확률적 성분을 예측하였다. Shah et al.(2020)은 확률적 성분을 예측하기 위하여 자기회귀, 비모수적 자기회귀(NPAR; Nonparametric Autoregressive), 함수적 자기회귀 모형(FAR; Functional Autoregressive) 등을 사용하였으며, 이 중 FAR이 가장 높은 예측력을 갖고 있음을 보여주었다. 이는 Jan, Shah & Ali(2022)의 연구 결과와도 일치한다.

3. 분석 데이터

3.1 분석기간 및 데이터

본 연구의 예측 대상 변수는 일별 평균 SMP로서 단위는 kWh이다. SMP는 하루 중 매 시간마다 결정되므로 이들 가격의 일중 평균을 분석에 사용하였다. 일별 평균 SMP는 전력거래소 홈페이지에서 추출하였다.

본 연구는 시계열 회귀와 AR 계열 시계열 모형, 딥러닝 등에 의해 SMP를 예측하고 있다. SMP에 영향을 미치는 시계열 회귀 변수는 크게 에너지 가격과 기상 정보로 구분할 수 있다. Castelli et al.(2020)은 전기가격에 영향을 미치는 특성 변수로 원유 현물가격, 기온, 풍속, 상대습도, 일사량 등을 고려하였다. 본 연구에서는 이들이 제안한 변수를 바탕으로 에너지

가격과 기후 정보를 수집하였다.

본 연구에서는 에너지 가격을 각 주유소의 일별 평균 경유 가격으로 측정하였으며, 기후 관련 변수는 전국 평균 온도(°C)와 평균 풍속(m/s), 평균 상대습도(%rh), 일사합(MJ/m²) 등의 일별 자료로 측정하였다. 이들 자료는 유가정보 사이트인 오피넷(Opinet)과 기상청의 기상자료 개발 포털에서 추출하였다. 한편, 시계열 회귀를 위해 휴일 여부와 계절 여부를 나타내는 가변수를 사용하였다.

전체 자료의 표본기간은 2013년 2월 3일부터 2024년 6월 25일까지이며, 표본의 개수는 4163이다. 본 연구는 통계적 모형과 딥러닝 분석에서 전체 데이터 80%를 학습 데이터 혹은 표본내(in-sample)로 설정하였으며, 나머지 20%를 테스트 데이터 혹은 표본외(out-of-sample)로 설정하였다.

3.2 기초통계량

분석의 편의를 위해 일별 평균 SMP, 경유 가격, 평균 온도, 평균 풍속, 평균 상대습도, 일사합을 나타내는 변수의 이름을 AvgSMP, DieselPrice, AvgTMP, AvgWindSpeed, AvgHumidity, Solar_Radiation으로 각각 설정하였다. 한편 휴일과 계절 가변수는 각각 HolidayDummy, Spring, Summer, Autumn, Winter로 변수 이름을 설정하였다.

휴일과 계절 가변수를 제외한 변수의 기초통계량은 [표 1]과 같다. [표 1]은 SMP와 경유 가격 모두 매우 높은 변동성을 갖고 있음을 보여주고 있으며, 기후 관련 데이터도 계절적 특성을 반영하여 변동성이 높게 나타나고 있다.

[표 1] 주요 변수의 기초통계량

	평균	표준편차	최소	25%	50%	75%	최대
AvgSMP	114.18	45.33	37.55	81.15	97.72	141.65	282.65
DieselPrice	1441.59	222.51	1058.75	1272.03	1386.38	1609.91	2167.66
AvgTMP	13.29	9.45	-12.30	5.10	13.90	21.60	30.30
AvgWindSpeed	1.86	0.65	0.80	1.40	1.70	2.20	5.20
AvgHumidity	68.32	12.89	31.00	59.00	70.00	78.00	96.00
Solar_Radiation	14.15	6.43	0.96	9.23	13.50	18.93	29.96

[표 2]는 피어슨 상관계수를 보여주고 있다. AvgSMP와 가장 관련이 높은 변수는 DieselPrice로서 이는 SMP에 큰 영향을 미치는 LNG와 중유 발전비용에 경유 가격이 연동되어 있기 때문이다. 또한 [표 2]는 기온(AvgTMP)과 SMP가 음의 관계를 갖고 있음을 보여주고 있다. 이는 기온이 낮아지면 난방유에 대한 수요가 증가하여, SMP를 상승시키기 때문

로 해석할 수 있다. 난방유 수요가 많은 겨울(Winter)과 SMP의 양의 관계는 이를 뒷받침하고 있다. 또한 풍속이나 습도, 일사량 등은 한국의 계절적 특성과 매우 밀접한 관련을 갖고 있음을 보여주고 있다. 휴일과 SMP 사이에도 음의 상관관계가 있음을 보여주고 있어 휴일에는 전력 수요가 감소하고 있음을 알 수 있다.

[표 2] 피어슨 상관계수

굵은 숫자는 유의수준 0.05보다 작은 p 값을 갖는 경우를 나타낸다.

	AvgSMP	DieselPrice	AvgTMP	AvgWindSpeed	AvgHumidity	Solar_Radiation	HolidayDummy	Spring	Summer	Autumn
DieselPrice	0.762									
AvgTMP	-0.166	0.071								
AvgWindSpeed	0.008	-0.015	-0.224							
AvgHumidity	-0.093	0.035	0.562	-0.330						
Solar_Radiation	-0.048	0.054	0.426	-0.034	-0.331					
HolidayDummy	-0.099	0.001	0.004	-0.034	0.015	-0.002				
Spring	0.026	0.001	-0.030	0.187	-0.300	0.343	-0.004			
Summer	-0.120	0.037	0.678	-0.117	0.420	0.267	-0.018	-0.346		
Autumn	-0.027	0.041	0.085	-0.180	0.199	-0.201	0.014	-0.338	-0.324	
Winter	0.120	-0.080	-0.737	0.105	-0.312	-0.421	0.009	-0.342	-0.328	-0.321

3.3 변수의 시계열 안정성

가변수를 제외한 각 변수의 시계열 안정성(stationarity)을 검토하기 위하여 ADF(Augmented Dickey-Fuller) 테스트를 수행하였다. 이의 결과를 [표 3]에서 보여주고 있다. 안정성 테스트에서 두 변수 AvgSMP와 DieselPrice는 시계열 안정성을 갖지 못하고 있으며, 1차 차분을 통해 안정성에 도달하고 있다. 두 변수의 로그 변환도 안정성을 갖지 못하고 있다. 반면 기후 관련 변수는 모두 안정성을 갖는 것으로 나타났다.

[표 3] ADF 테스트 결과

ADF 테스트의 귀무가설은 '시계열 데이터에 단위근이 존재한다'이다. ADF 통계량의 p-값이 유의수준보다 크면 귀무가설을 기각할 수 없으므로 데이터는 안정성을 갖지 못할 가능성이 높다. 본 연구에서는 유의수준을 0.01로 설정하였다.

	ADF 통계량	p-값	안정성 여부
AvgSMP	-1.8069	0.3771	비안정성
log(AvgSMP)	-1.7686	0.3962	비안정성

diff(AvgSMP)	-13.9811	0.0000	안정성
DieselPrice	-2.5542	0.1028	비안정성
log(DieselPrice)	-2.5483	0.1042	비안정성
diff(DieselPrice)	-8.6860	0.0000	안정성
AvgTMP	-4.6115	0.0001	안정성
AvgWindSpeed	-8.0416	0.0000	안정성
AvgHumidity	-6.6782	0.0000	안정성
Solar_Radiation	-5.9888	0.0000	안정성

AvgSMP의 비안정성을 제거하지 않는 경우 예측 값의 신뢰구간이 넓어져 예측의 정확성이 떨어지며, 이로 인해 예측력이 떨어진다. 또한 AvgSMP를 종속변수로 사용하는 시계열 회귀 모형의 경우 오차항의 정상성(normality) 가정 위반, 회귀계수 추정치의 비효율 및 불일치, 허위회귀(spurious regression)의 발생 가능성이 커지게 된다. 따라서 AvgSMP를 사용하는 경우 이의 안정성 확보 과정이 반드시 필요하다.

4. 실증 분석

4.1 일별 평균 전기가격의 성분 분해

1) 전기가격의 성분 구성

본 연구에서는 전기가격 P_t 가 갖고 있는 추세(trend)와 캘린더 효과(calendar effect) 등을 제거하기 위하여 Shah et al.(2020)과 같이 P_t 에 로그를 취하고 이를 결정적 성분(deterministic component) D_t 와 확률적 성분(stochastic component) R_t 로 분해하였다.

$$\log P_t = D_t + R_t$$

여기서 $t = 1, \dots, T$ 이다. D_t 는 추세(trend) l_t 와 요일(weekday) w_t , 주중 휴일(weekday holiday) b_t , 연간(yearly) y_t 등의 효과로 구성하였다.

$$D_t = l_t + w_t + b_t + y_t$$

R_t 는 D_t 의 효과가 알려져 있다는 가정을 바탕으로 다음과 같이 추정하였다.

$$R_t = \log P_t - \hat{D}_t = \log P_t - (\hat{l}_t + \hat{w}_t + \hat{b}_t + \hat{y}_t)$$

여기서 $\hat{l}_t, \hat{w}_t, \hat{b}_t, \hat{y}_t$ 은 결정적 성분을 구성하는 l_t, w_t, b_t, y_t 의 추정값을 의미한다. 이를 바탕으로 1일 후의 전기가격 \hat{P}_t 은 다음과 같이 추정한다.

$$\hat{P}_t = \exp(\hat{D}_t + \hat{R}_t)$$

2) 결정적 성분의 분해 방법

전기가격의 결정적 성분 중 '추세'는 스플라인 함수(spline function)를 통하여 추정하였다. 스플라인 함수는 전체 표본을 몇 개의 매듭(knot)에 따라 구분한 다항식을 사용해 추세를 추정한다. 따라서 단기적으로 매우 많은 변동을 갖는 데이터라고 하더라도 스플라인 함수는 장기 추세의 적합을 나타낼 수 있다.

스플라인 함수를 데이터에 적합하는 과정에서 매듭의 수는 매우 중요하다. 매듭의 수에 따라 곡선의 형태와 예측 성능이 달라지므로 최적 매듭 수를 찾는 과정이 필요하다. 본 연구에서는 최적 매듭 수를 찾기 위해 교차 검증 검사 절차를 사용하였다.

'요일 효과'는 로그 변환된 전기가격 관측 개체($\log P_t$)에서 추정된 추세(\hat{l}_t)를 차감하여 추세를 제거한 시계열 자료 \hat{DT}_t 을 구한 다음 이의 각 요일 별 평균을 계산하여 추정하였다. 이는 다음과 같다.

$$\hat{w}_d = \frac{1}{N_d} \sum_{i=1}^{N_d} \hat{DT}_{d,i} = \frac{1}{N_d} \sum_{i=1}^{N_d} (\log P_{d,i} - \hat{l}_{d,i})$$

여기서 $\hat{DT}_d (d = 1, \dots, 7)$ 는 $\hat{DT}_t (t = 1, \dots, T)$ 중에서 특정 요일(d)에 해당하는 데이터 시리즈를 의미하며, N_d 는 각 요일별 관측개체의 수를, $\hat{DT}_{d,i}$ 는 \hat{DT}_d 의 각 관측개체를 의미한다. 따라서 요일 효과의 추정값(\hat{w}_d)은 7개가 존재한다. 이를 $t = 1, \dots, T$ 에 해당하는 각 관측개체의 요일에 할당하여 \hat{w}_t 시리즈를 구한다.

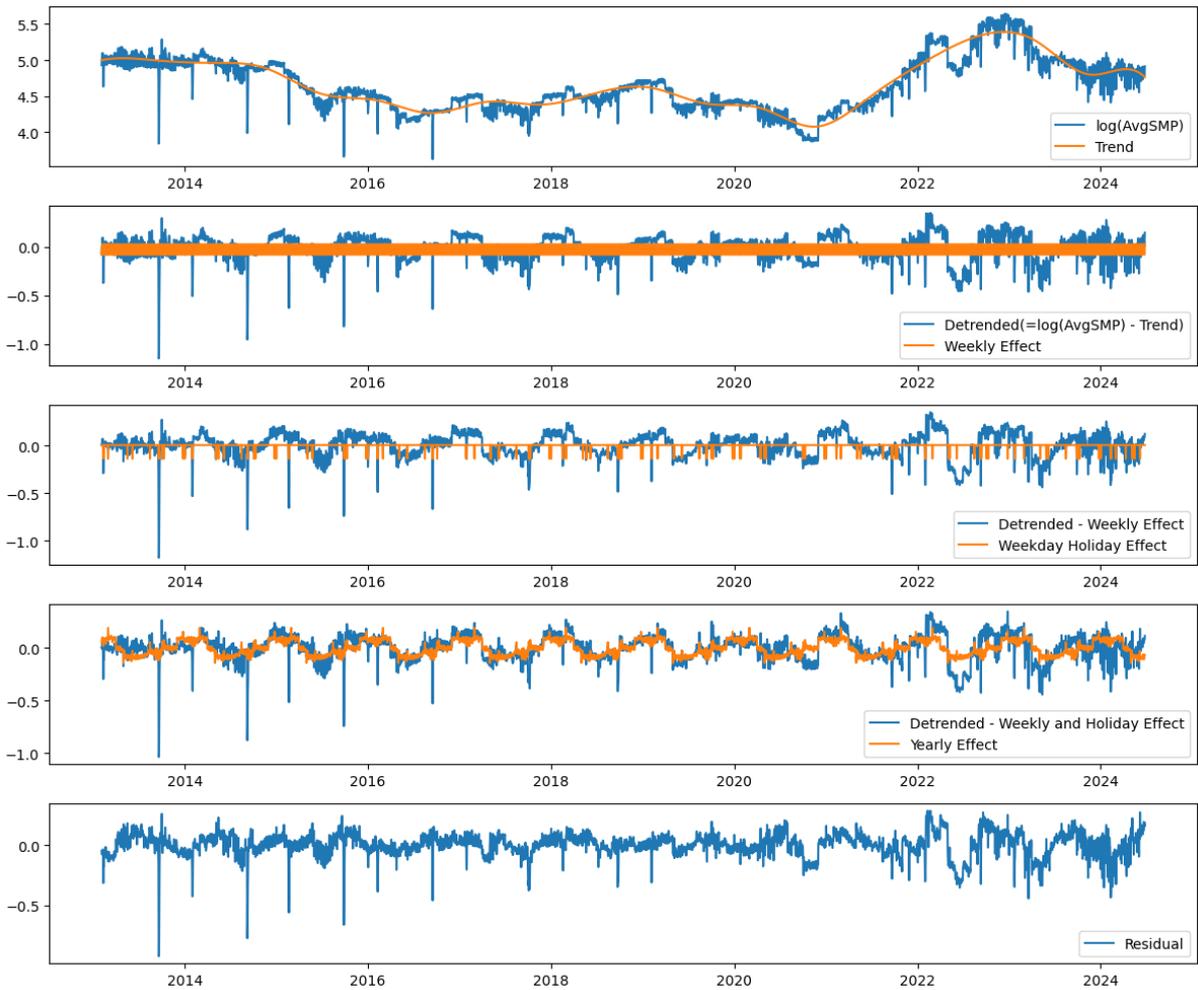
'주중 휴일 효과(\hat{b}_t)'는 토요일과 일요일을 제외한 공휴일과 대체휴일의 효과를 나타낸 것이다. 이는 시리즈 $\hat{DT}_t - \hat{w}_t$ 중에서 주중 휴일에 해당하면 1을, 그렇지 않으면 0을 할당한 다음 두 그룹의 평균으로 계산하였다.

'연간 효과'는 1월 1일부터 12월 31일까지의 각 날짜에 1부터 365를 할당한 후, 각 날짜의 $\hat{DT}_t - \hat{w}_t - \hat{b}_t$ 평균을 구하여 추정하였다. 연도별 연간 효과의 비교가 일관성을 갖도록 하기 위해 윤년의 2월 29일에 해당하는 관측개체는 제외하였다. 즉, 연간 효과는 각 날짜별로 365개가 존재한다.

3) 결정적 성분의 예측

결정적 성분 중 '요일 효과', '주중 휴일 효과', '연간 효과'와 같은 캘린더 효과는 표본내외 표본외 구분 없이 모두 동일한 것으로 가정한다. 이는 시간의 흐름에 따라 캘린더 효과가 변하지 않는다는 의미이다. 따라서 전체 표본에서 추정된 값과 동일한 캘린더 효과를 표본외의 관측개체에도 적용하였다.

[그림 1] $\log(\text{AvgSMP})$ 의 결정적 및 확률적 성분 분해



표본외의 '추세'는 이동 윈도우(rolling window)를 사용한 이중 지수평활 모형(double exponential smoothing)으로 예측하였다. 이중 지수평활 모형은 시계열 데이터의 추세를 예측하는 대표적인 방법이다. 이는 최근 데이터에 보다 많은 가중치를 부여하기 때문에 예측의 정확성이 높다. 이중 지수평활 모형의 '평활 상수(α)'와 '추세 평활 상수(β)', 이동 윈도우의 크기는 최적화 기법인 차분 진화(differential evolution) 알고리즘을 사용하여 추정하였다. 차

분 진화는 많은 반복 계산이 필요하지만 글로벌 최적해를 찾는 데 효과적이며, 비선형 함수와 이산적인 문제에 잘 적합하는 장점을 가지고 있다. 본 연구에서는 표본내 '추세' 데이터에 차분 진화를 적용하여 최적 모수와 최적 윈도우 크기를 구한 다음 이를 바탕으로 표본외 '추세'를 예측하였다.

[그림 1]은 표본 기간 전체에 대한 $\log(\text{AvgSMP})$ 의 결정적 성분과 확률적 성분의 분해 결과를 보여주고 있다. 첫 번째부터 다섯 번째 그래프는 각각 장기 추세(Trend), 요일 효과(Weekly Effect), 주중 휴일 효과(Weekday Holiday Effect), 연간 효과(Yearly Effect), 확률적 성분을 나타내는 잔차(Residual)를 나타내고 있다. 잔차 $\hat{\epsilon}_t$ 에 대한 ADF 테스트 결과 ADF 통계량은 -6.1989로서 시계열 안정성을 갖는 것으로 나타나고 있다. 이는 결정적 성분과 확률적 성분의 분해가 성공했음을 의미한다.

5. 예측 성과 분석

5.1 SMP 수준변수에 의한 예측

1) SMP와 경유 가격의 수준변수를 사용한 예측

어떤 변수 사이에 공적분 관계가 있는 경우 보통 ECM(Error Correction Model), VECM(Vector Error Correction Model), ARDL(Autoregressive Distributed Lag) 등의 방법을 사용하여 두 변수의 관계를 추정한 후 예측할 수 있다. AvgSMP와 DieselPrice 사이의 공적분 관계는 Johansen 검정을 통해 확인하였다. 시차를 1로 설정한 두 변수 사이의 공적분 관계는 다음과 같이 나타났다.

[표 4] AvgSMP와 DieselPrice 사이의 공적분 관계

여기서 p-값은 MacKinnon-Haug-Michelis의 근삿값을 나타낸다.

공적분 관계 수	Eigenvalues	Trace 통계량	p-values
최대 0 개	0.041394	141.1888	0.01
최대 1 개	0.000175	0.582837	1.00

위의 결과는 AvgSMP와 DieselPrice 사이에 1개의 공적분 관계가 있음을 의미한다. 따라서 ECM이나 VECM, ARDL 등의 모형을 사용하여 분석하였다. 분석 결과 ECM 모형은 AvgSMP와 DieselPrice 사이에 장기적 균형 관계가 강하다는 점을 보여주었지만, DieselPrice의 단기 변화가 AvgSMP의 변화에 미치는 유의미한 영향을 찾을 수 없어 표본외 데이터의 예측에는 ECM 모형을 사용하기 어려웠다. 표본내 데이터에 대한 VECM의 추정 결과는 추정 계수의

p-값이 모두 유의수준을 갖지 못하는 것으로 나타났다. 이는 VECM 모형도 예측에 사용하기 어렵다는 것을 의미한다.

ARDL 모형은 공적분 관계가 있는 변수 사이의 장기 균형 및 단기 동태적 관계를 동시에 설명할 수 있는 유연하고 강력한 도구이다. 이는 다음과 같이 나타낼 수 있다.

$$y_t = \alpha_0 + \sum_{i=1}^p \alpha_i y_{t-i} + \sum_{j=1}^q \beta_j x_{t-j} + \epsilon_t$$

여기서 p 와 q 는 종속변수와 독립변수의 시차를 의미한다. 본 연구에서는 두 변수의 최적 시차를 찾기 위하여 다양한 시차 조합을 대입한 후, AIC(Akaike Information Criterion)를 최소로 하는 시차를 선택했다. 이를 통해 p 와 q 의 최적 시차는 각각 4와 0으로 나타났다.

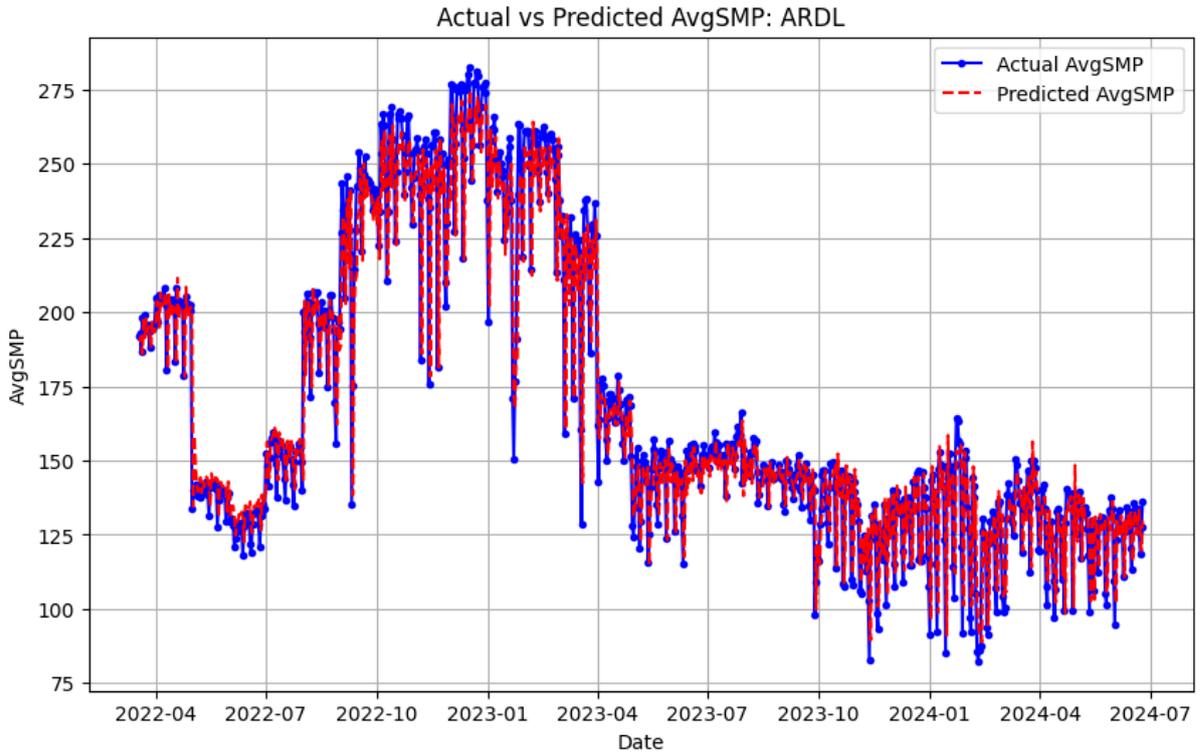
[표 5] AvgSMP와 DieselPrice의 ARDL 추정 계수: 표본내

	계수	t-값	p-값
$\hat{\alpha}_0$	-5.2343	-4.9266	0.0000
AvgSMP _{t-1}	0.9722	56.2324	0.0000
AvgSMP _{t-2}	-0.2674	-11.1335	0.0000
AvgSMP _{t-3}	0.1507	6.2742	0.0000
AvgSMP _{t-4}	0.0907	5.2449	0.0000
DieselPrice _t	0.0077	6.9140	0.0000

[표 5]는 ARDL의 표본내 추정 결과를 보여주고 있다. ARDL의 모든 계수는 매우 유의한 수준을 보이고 있다. AvgSMP_{t-1}의 계수가 1에 가까운 것은 AvgSMP가 높은 자기상관을 가지고 있음을 의미한다. AvgSMP_{t-2}가 음수인 것은 시차 2의 AvgSMP가 높으면 현재 시점의 AvgSMP를 낮추는 경향을 갖는다는 의미이다. 시차 3과 시차 4의 AvgSMP도 이와 마찬가지로 해석할 수 있다. 전체적으로 현재 시차에 가까운 AvgSMP가 현재의 AvgSMP에 보다 큰 영향을 미치고 있음을 확인할 수 있다. 한편, 현재 시점의 DieselPrice가 현재의 AvgSMP에 유의한 영향을 미치고 있음을 보여주고 있다.

이를 통해 ARDL 모형이 AvgSMP의 단기적 자기상관과 DieselPrice의 영향 등을 통계적으로 유의한 수준으로 보여주고 있으므로 이 모형을 예측에 사용할 수 있다. 모형에서 AvgSMP 시차 4까지 유의한 영향을 미치고 있으므로, 윈도우 크기를 4로 갖는 이동윈도우(rolling window) 방법으로 표본외 기간의 AvgSMP를 예측했다. 이에 대한 예측 성과를 [그림 2]에서 보여주고 있다.

[그림 2] 표본외 기간에 대한 AvgSMP의 예측 결과: ARDL 모형



표본외 기간에서 AvgSMP의 예측값은 실젯값에 비해 다소 작은 변동을 갖고 있는 모습을 보여주고 있으나 패턴은 매우 유사하게 나타나고 있다.

2) 딥러닝

본 연구는 심층 신경망(DNN, Deep Neural Network) 모형 중 LSTM(Long Short-Term Memory, 장단기 메모리)을 사용하여 시계열 예측을 수행하였다. LSTM은 금융 및 주식, 기상 데이터와 같은 시간 의존성이 높은 시계열 데이터의 장기 및 단기 정보 학습에 장점을 갖고 있다.

은닉층의 최적 수를 찾는 문제는 DNN 모형 설계에서 매우 중요하다. 은닉층의 수가 너무 많으면 과적합(overfitting) 문제가 발생하여 학습 데이터에는 잘 맞지만 검증 데이터에서는 성능이 떨어질 수 있다. 또한 은닉층의 수가 너무 적으면 데이터의 복잡한 구조를 제대로 반영하지 못하는 과소적합(underfitting)의 문제가 발생할 수 있다.

본 연구에서는 은닉층의 최적 수를 찾기 위하여 교차 검증 방법을 사용하였다. 이는 은닉층의 수를 여러 개 설정한 후, 각각에 대해 모형을 훈련한 후 성능을 평가하는 방법이다. 분석 결과, 은닉층이 1개일 때 평균 검증 손실(average validation loss)이 가장 작게 나타났다.

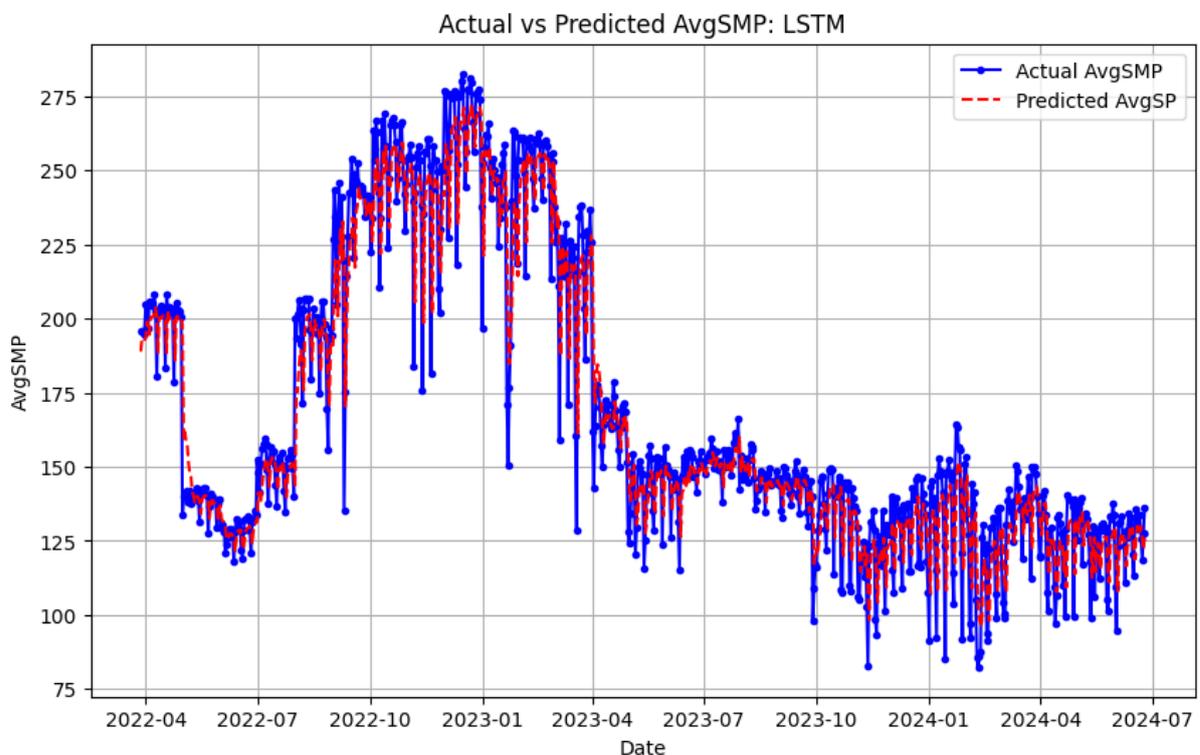
1개의 은닉층을 갖는 LSTM 모형에서 뉴런의 유닛(unit) 수와 이동 윈도우의 크기에 따라

모형의 성능이 달라진다. 본 연구에서는 Grid Search 방법을 이용하여 유닛 수와 윈도우 크기의 최적 조합을 찾았다. 딥러닝 모형은 입력 데이터의 패턴을 학습하여 예측을 수행하므로 입력 데이터의 범위가 크게 되면 모형 학습이 어려워진다. 이러한 문제를 해결하기 위해 모든 특성의 값을 일정 범위로 변환하는 데이터 정규화(data normalization)가 필요하다. 본 연구는 데이터 정규화 방법 중 모든 특성 값을 [0, 1]의 범위로 대응시키는 최소-최대 정규화 (Min-Max normalization)를 사용하였다.

모형의 학습을 위한 프로세스를 위해 손실함수는 평균 제곱 오차(Mean Squared Error, MSE)를 사용하였으며, MSE를 빠르게 줄여주기 위한 최적화 알고리즘으로는 Adam(Adaptive Moment Estimation)을 사용하였다. 모형의 성능을 평가하는 지표로는 전체 데이터에서 제대로 예측한 비율을 측정하는 '정확도(accuracy)'를 사용하였다. 전체 훈련 데이터의 반복 학습 횟수(epoch)는 20회로 설정하였으며, 모형이 한 번의 학습 단계에서 처리하는 표본의 수 (batch size)는 32로 설정하였다.

[그림 3]은 이상의 설정을 바탕으로 추정한 테스트 데이터에 대한 AvgSMP의 예측 결과를 보여주고 있다. 그림에서는 AvgSMP의 예측값이 실젯값의 추세와 비슷한 패턴을 보이고 있으며, 변동성은 실제보다 작게 나타나고 있음을 보여주고 있다.

[그림 3] 표본외 기간에 대한 AvgSMP의 예측 결과: LSTM 모형



5.2 SMP의 확률적 성분 예측

1) 시계열 회귀

SMP의 결정적 성분은 캘린더 효과와 장기 추세 등을 포함하고 있다. 이 경우 [표 3]에서 보여주는 것처럼 SMP는 시계열 안정성을 갖지 못하기 때문에 회귀 분석에서 이를 고려하여야 한다. 본 연구에서는 AvgSMP 중에서 시계열 안정성을 갖는 확률적 성분 Residual을 종속변수로 하는 회귀모형을 설정하였다. 독립변수로는 온도, 상대습도, 풍속, 일사량 등의 기후 관련 변수를 사용하였다. DieselPrice는 시계열 안정성을 갖지 못하므로 이는 차분을 사용하여 독립변수에 포함시켰다. 종속변수인 Residual을 구하기 위해 결정적 성분인 추세와 캘린더 효과를 제외했으므로, 독립변수에서 시간과 관련한 변수는 제외하였다.

회귀분석 관련 문제를 파악하기 위해 다중공선성의 존재 여부에 대해서는 VIF(Variance Inflation Factor)를, 오차항의 이분산과 자기상관에 대해서는 White 테스트와 Ljung-Box 테스트를 사용하였다. 표본내 자료의 회귀분석에 사용된 모형으로는 OLS(Ordinary Least Squares)와 LASSO(Least Absolute Shrinkage and Selection Operator), GLS(Generalized Least Squares), WLS(Weighted Least Squares) 등을 사용하였으며, 오차항의 시계열 상관을 고려하기 위하여 MA(Moving Average)를 결합하였다¹.

다양한 모형과 독립변수에 따른 회귀 분석 결과는 거의 모두 결정계수 R^2 가 매우 낮게 나타났다. 또한 대부분의 모형에서 오차항이 높은 이분산이나 시계열 상관을 갖고 있음을 보여주었다. 이를 해결하기 위해 오차항에 여러 차수를 갖는 MA를 적용했지만 이 경우에는 회귀계수가 유의하지 않게 되는 문제가 발생하였다.

본 연구에서는 오차항의 이분산과 자기상관 문제를 극복하기 위해 먼저 GLS를 사용하였으나 오차항의 자기상관은 여전히 존재하였다. 따라서 이분산성 문제에 대해서는 WLS를 수행하고 오차항에 대해서는 시차 9를 갖는 MA를 적용하였다. WLS에서 가중치 행렬은 오차항 공분산 행렬의 역행렬로 MA의 시차는 AIC(Akaike Information Criterion), BIC(Bayesian Information Criterion)를 최소로 하는 값을 찾아 선택하였다. 다중공선성 문제를 피하기 위해 독립변수에서는 평균 기온을 나타내는 AvgTMP를 제거하였다.

WLS 모형은 오차항의 이분산과 자기상관, 다중공선성 문제를 나타내지는 않았지만 추정계수가 모두 0에 가까워 확률적 성분의 예측을 수행하기에는 부적합한 것으로 나타났다. 이러한 결과가 나타난 이유는 AvgSMP의 결정적 성분에 독립변수의 영향이 이미 상당부분 포함되어 있기 때문으로 해석할 수 있다. 따라서 WLS가 데이터의 변동성을 제대로 설명하지 못하여 예측값이 평균을 중심으로 크게 벗어나지 않은 것으로 판단한다.

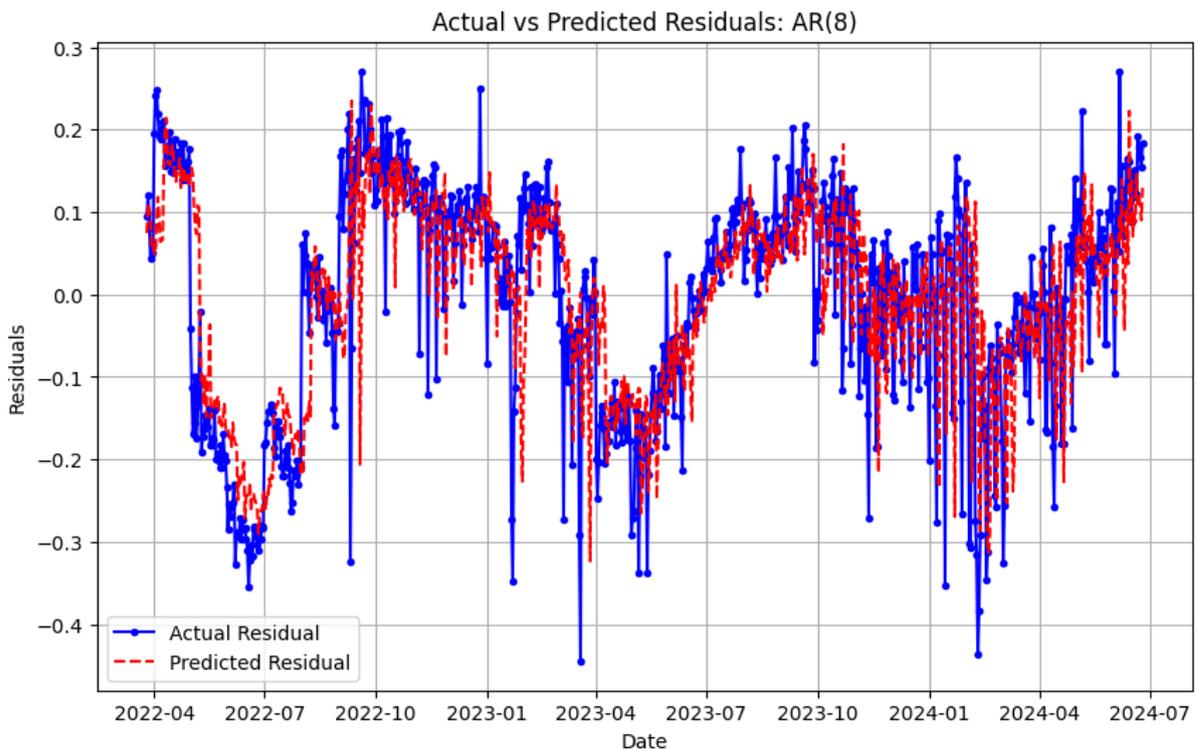
¹ 지면을 아끼기 위해 시계열 회귀 분석 표는 생략하였다.

2) AR 모형

본 연구에서 사용하고 있는 SMP는 하루 중 시간별 SMP의 평균으로 계산되었다. 이처럼 연속적인 값이 아니라 단일 대푯값을 사용하는 경우에는 함수적 AR이 적합하지 않다. 함수적 AR은 일중 시간이나 분과 같이 연속적 시간 함수를 갖는 고빈도 데이터에 보다 적합하기 때문이다. 한편 확률적 성분 Residual의 자기상관 구조와 비선형 추세를 파악하기 위하여 Hurst 지수(Hurst Exponent)를 측정하였다. 측정 결과 Hurst 지수는 0.174를 보여주고 있어 Residual이 평균-회귀 경향을 갖고 있음을 보여주고 있다. 따라서 비모수적 AR 대신 AR 모형만 사용해도 확률적 성분을 효과적으로 설명할 수 있을 것으로 판단한다.

AR(Autoregressive) 모형에서 최적의 시차를 찾는 방법으로는 첫째, AIC와 BIC를 최소로 하거나, 둘째, PCF(Partial Autocorrelation Function) 함수의 그래프가 급격히 감소하는 시점을 찾거나, 셋째, 데이터 셋을 여러 개의 폴드로 나누고 MSE와 같은 예측 성능 지표가 가장 좋은 시차를 선택하는 교차 검증 등이 있다.

[그림 4] 표본외 기간에 대한 확률적 성분의 예측 결과: AR(8) 모형



본 연구에서는 확률적 성분의 추정값(\hat{r}_t)에 대한 AR 모형의 최적 시차를 찾기 위하여 교차 검증 방법을 사용하였다. 이를 위해 데이터를 시간에 따라 다섯 개의 폴드로 설정하고 각

폴드에서 1부터 15시차까지의 MSE를 조사하였다. 그 결과 최적 시차는 8로 나타났다. 따라서 \hat{R}_t 은 다음과 같이 설정할 수 있다.

$$\hat{R}_t = \sum_{i=1}^8 \varphi_i \hat{R}_{t-i} + \varepsilon_t$$

표본외 기간의 데이터를 대상으로 AR(8) 모형으로 확률적 성분에 대한 예측을 수행했다. 표본외 기간 데이터를 사용하므로 최초 예측값은 2022년 3월 26일부터 시작한다. 이에 대한 그래프를 [그림 4]에서 보여주고 있다.

이를 바탕으로 확률적 성분의 추정값 \hat{R}_t 에 대한 자기상관함수(Autocorrelation Function, ACF) 및 부분자기상관함수(Partial Autocorrelation Function, PACF)의 그래프를 검토하였다. 그 결과 두 함수 모두 시차 1에서 강한 상관관계를 보이고, 그 이후 시차에서는 매우 약한 상관관계를 보여주었다. 이는 AR(8) 모형이 데이터를 잘 설명하고 있어 AR(8) 모형의 잔차는 백색 잡음에 가깝다는 의미이다. 이 결과는 AR(8) 모형이 데이터를 잘 설명하고 있으며, 추가적인 모형 설정이 필요하지 않을 수 있음을 시사한다. 그러나 모형의 예측력을 증가시키기 위하여 AR 모형에 점프와 GARCH 계열을 결합하여 분석하였다.

3) AR 모형과 점프 모형의 결합

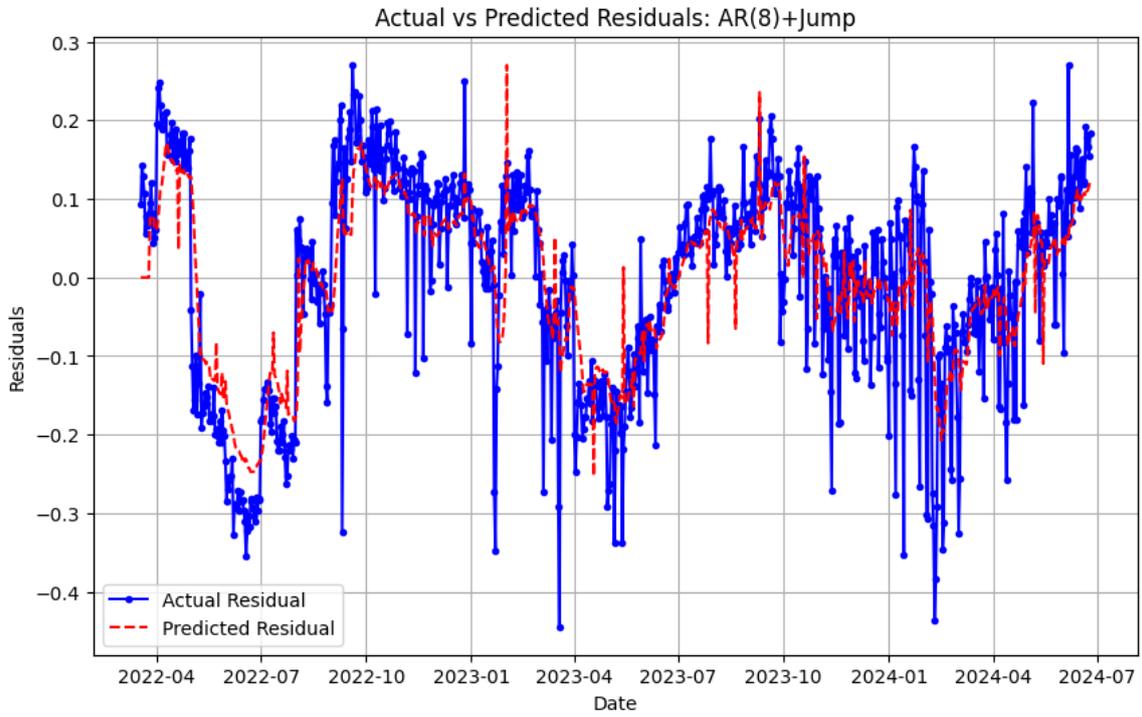
AvgSMP의 확률적 성분 \hat{R}_t 에 대한 AR(8)의 잔차 ε_t 가 점프 속성을 갖고 있는지를 평가하였다.

점프의 비대칭성을 고려하지 않고 일정한 확률에 따라 값이 급격히 변하는 점프 모형은 다음과 같이 나타낼 수 있다.

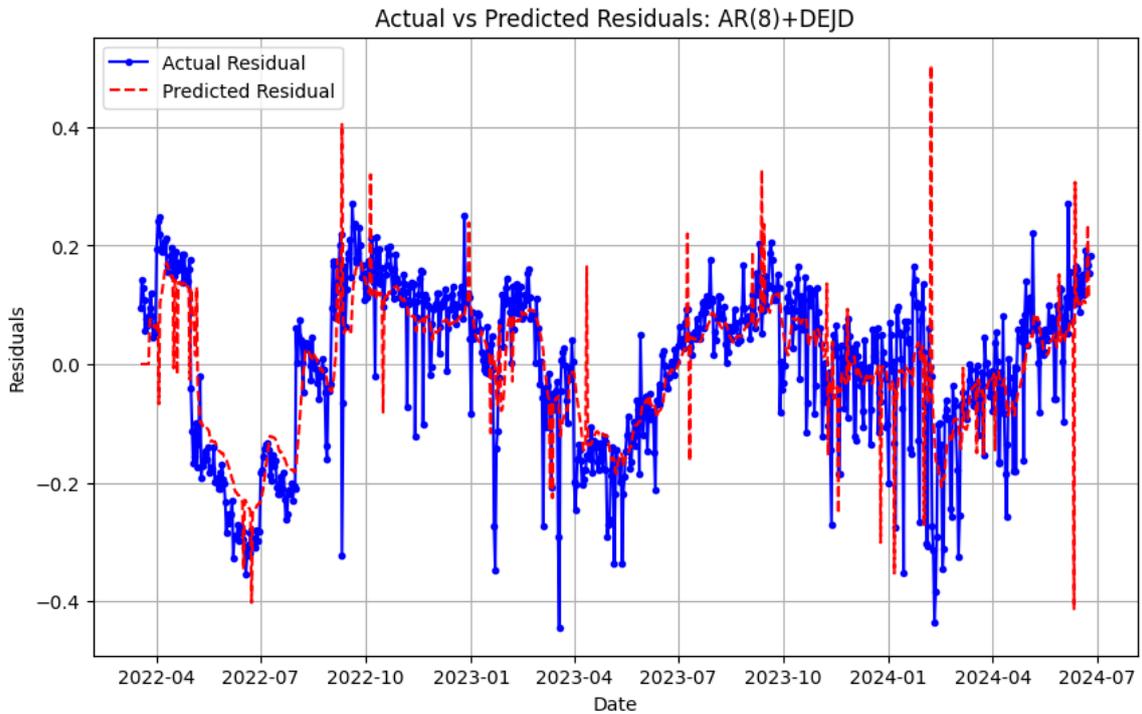
$$\hat{R}_t = \sum_{i=1}^8 \varphi_i \hat{R}_{t-i} + J_t + \eta_t, \quad J_t = \begin{cases} \mu_J + \sigma_J z_t, & \text{확률 } \lambda \\ 0, & \text{확률 } 1 - \lambda \end{cases}$$

여기서 J_t 는 시점 t 의 점프 크기이며, μ_J 는 점프의 평균 크기이다. σ_J 는 점프의 표준편차이며, z_t 는 표준 정규분포를 따르는 확률변수이다. λ 는 점프가 발생할 확률이다.

[그림 5] 표본외 기간에 대한 확률적 성분의 예측 결과: (AR(8)+점프) 모형



[그림 6] 표본외 기간에 대한 확률적 성분의 예측 결과: (AR(8)+DEJD) 모형



한편, 점프가 비대칭성을 갖는 경우를 나타내는 이중지수 점프 모형(Double Exponential Jump Diffusion Model, DEJD)은 다음과 같다.

$$\hat{R}_t = \sum_{i=1}^8 \varphi_i \hat{R}_{t-i} + J_t + \eta_t, \quad J_t = \begin{cases} \mu_j^+ + \sigma_j^+ \cdot E_t, & \text{확률 } \frac{\lambda}{2} \\ -(\mu_j^- + \sigma_j^- \cdot E_t), & \text{확률 } \frac{\lambda}{2} \\ 0, & \text{확률 } 1 - \lambda \end{cases}$$

여기서 μ_j^+ 와 μ_j^- 는 각각 양의 점프와 음의 점프의 평균 크기를 나타낸다. 또한 σ_j^+ 와 σ_j^- 는 각각 양의 점프와 음의 점프의 표준편차를 나타낸다. E_t 는 평균과 표준편차가 1인 지수 분포를 따르는 확률변수이다.

[그림 5]와 [그림 6]은 점프 모형과 DEJD 모형에 의한 예측 결과를 보여주고 있다. 두 모형 모두 확률적 성분의 급격한 변동을 잘 나타내고 있다. 점프 모형은 상향 부분의 점프 예측이 많은 반면, DEJD는 상향보다는 하향 부분 점프에 대한 예측 빈도가 많다. 또한 점프의 크기도 DEJD가 점프 모형보다 큰 경우의 빈도가 많다.

4) AR 모형과 GARCH 계열 모형의 결합

확률적 성분에 대해 AR(8) 모형을 적용한 후의 잔차 ε_t 가 GARCH(1,1) 효과를 갖는지 검토하였다. 이를 위해 먼저 Engle의 ARCH(Autoregressive Conditional Heteroskedasticity) 테스트를 실시하여 변동성 클러스터링(clustering)을 검토하였다.

[표 6] AR(8) 모형의 잔차에 대한 Engle의 ARCH 테스트 결과

LM 통계량	LM p-값	F 통계량	F의 p-값
238.0059	0.0000	25.1845	0.0000

[표 6]은 Engle의 ARCH 테스트 결과를 보여주고 있다. 여기서 LM 통계량과 F 통계량은 매우 높은 값을 보여주고 있으며, 각각의 p-값도 0에 가깝게 나타났다. AR(8) 모형의 잔차는 강한 ARCH 효과를 보여주고 있으므로 GARCH 모형을 사용할 수 있음을 알 수 있다. Escribano et al.(2011)와 마찬가지로 GARCH(1,1)을 사용하여 전기가격을 예측하였다. 여기서 z_t 는 평균이 0이고 분산이 1인 표준 정규분포를 따르는 것으로 가정한다.

$$\varepsilon_t = \mu + \sigma_t z_t, \quad \sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

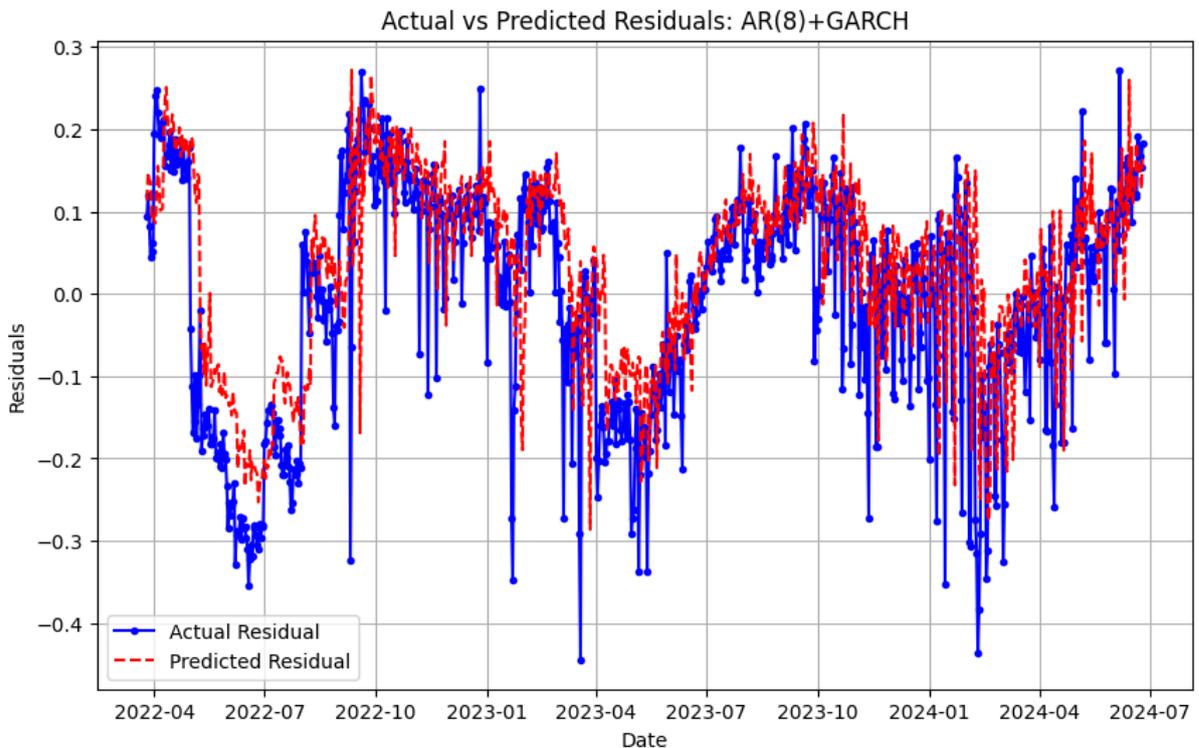
AR(8) 모형의 잔차 ε_t 에 대한 GARCH(1,1)의 적합 결과는 [표 7]과 같다. 여기서 $\hat{\mu}$ 은 ε_t 의 평균으로서 0에 가까운 값을 보여주고 있다. p-값도 통계적으로 유의미한 값을 보이지 않아 0에서 크게 벗어나지 않고 있음을 확인시켜주고 있다.

변동성을 나타내는 σ_t^2 의 추정 결과에서 조건부 분산 모형의 상수인 $\hat{\omega}$ 은 p-값이 0.198로 유의미한 값을 보여주지 않고 있다. 잔차의 제곱이 현재 변동성에 미치는 영향을 나타내는 $\hat{\alpha}_1$ 의 p-값은 0.0291로서 다소 유의한 결과를 보여주고 있다. 이전 시점의 변동성이 현재 변동성에 미치는 영향을 나타내는 $\hat{\beta}_1$ 은 매우 유의한 통계적 결과를 보여주고 있으며, 추정 값도 크므로 변동성의 지속 시간이 길게 유지되는 경향이 강하다는 것을 알 수 있다.

[표 7] AR(8) 모형의 잔차에 대한 GARCH(1,1) 적합 결과

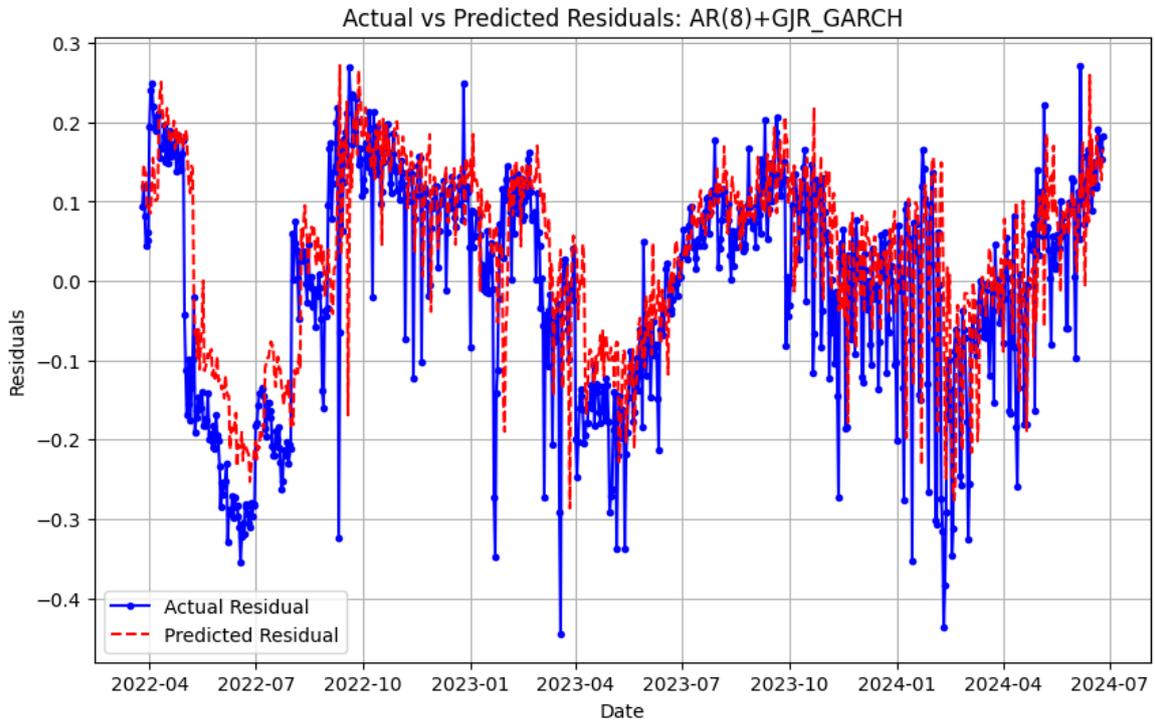
모수	계수	표준오차	t-값	p-값
$\hat{\mu}$	-0.0004	0.0009	-0.420	0.674
$\hat{\omega}$	0.00017	0.00013	1.2870	0.1980
$\hat{\alpha}_1$	0.18620	0.08535	2.1820	0.0291
$\hat{\beta}_1$	0.79790	0.11200	7.1000	0.0000

[그림 7] 표본외 기간에 대한 확률적 성분의 예측 결과: (AR(8)+GARCH(1,1)) 모형

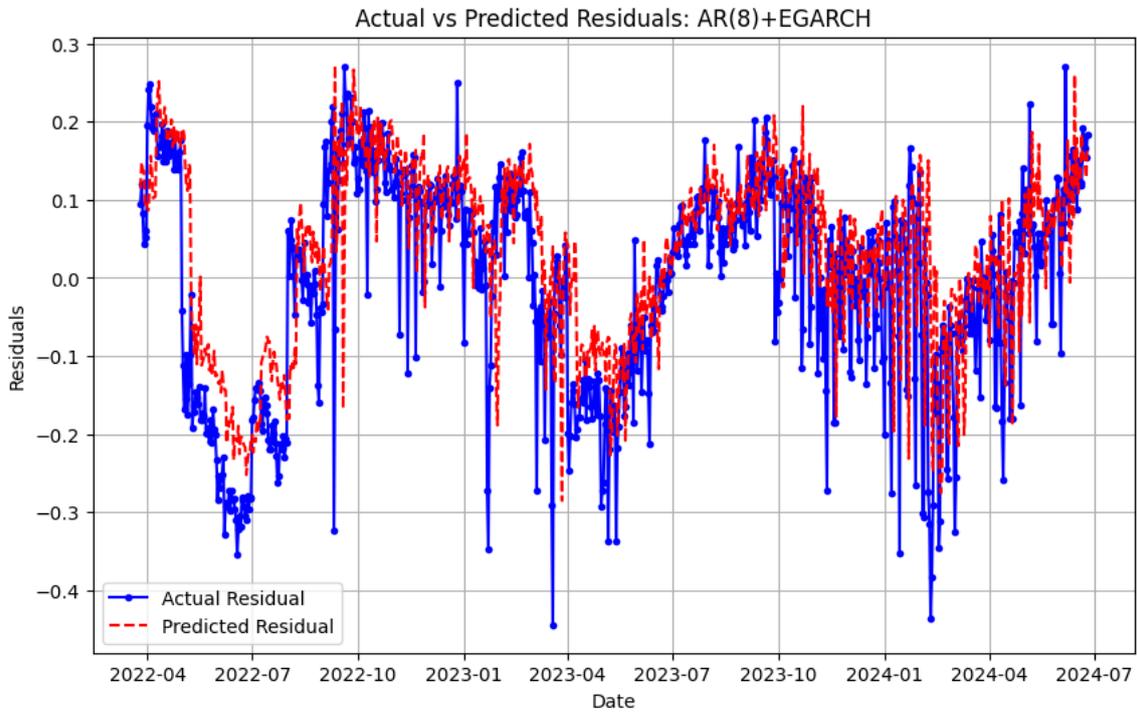


이를 바탕으로 테스트 기간의 확률적 성분을 예측한 결과는 [그림 7]과 같다.

[그림 8] 표본외 기간에 대한 확률적 성분의 예측 결과: (AR(8)+GJR_GARCH(1,1)) 모형



[그림 9] 표본외 기간에 대한 확률적 성분의 예측 결과: (AR(8)+EGARCH(1,1)) 모형



한편, AR(8) 모형의 잔차가 비대칭적 변동성을 갖는 경우를 고려하여 GJR-GARCH와 EGARCH도 분석 모형에 포함시켰다. GJR-GARCH는 음의 충격이 양의 충격보다 변동성을 더 크게 증가시키는 경우에 적합하고, EGARCH는 음과 양의 충격이 비대칭적인 경우에 비대칭성을 잘 포착할 수 있다. 두 모형의 수학적 표현은 생략하였다.

두 모형을 사용한 분석 결과를 [그림 8]과 [그림 9]에서 각각 보여주고 있다. 두 그래프는 GARCH(1,1)과 대체로 비슷한 모습을 보이고 있다.

5.3 예측 추정 모형의 성과 비교

1) 예측 성과 평가 방법

표본외 기간에 대한 각 모형의 예측 성과를 평가하기 위하여 평균절대오차(Mean Absolute Error, MAE), 평균 제곱 백분율 오차(Mean Squared Percentage Error, MSPE), 평균 절대 백분율 오차(Mean Absolute Percentage Error, MAPE) 등의 지표를 사용하였으며, 두 예측 성과의 차이를 보기 위하여 DM(Diebold-Mariano) 테스트를 수행하였다.

MAE와 MSPE, MAPE의 수학적 모형은 각각 다음과 같다. 여기서 y_i 는 표본외 기간의 i 시점 실제 AvgSMP를 나타내며, \hat{y}_i 는 i 시점의 예측 AvgSMP를 나타낸다. n 은 표본외 기간 개체의 수이다.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

$$MSPE = \frac{1}{n} \sum_{i=1}^n \left(\frac{y_i - \hat{y}_i}{y_i} \right)^2$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100$$

한편, DM 통계량은 다음과 같이 나타낼 수 있다.

$$DM \text{ 통계량} = \frac{\bar{d}}{\sqrt{\{\sum_{i=1}^n (d_i - \bar{d})^2\}/n}}$$

여기서

$$\bar{d} = \frac{1}{n} \sum_{i=1}^n d_i, \quad d_i = L(e_{1,i}) - L(e_{2,i})$$

$L(\cdot)$ 은 손실 함수이며, 여기서는 다음과 같은 평균 제곱 오차(Mean Squared Error, MSE)를 사용하였다.

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

2) 예측 성과 결과

본 연구는 AvgSMP에 대한 표본외(혹은 테스트 데이터) 예측 성과를 다양한 모형을 통해 비교하고 있다. 예측을 위한 모형은 첫째, 표본내(혹은 훈련 데이터) AvgSMP의 수준변수를 그대로 사용하거나 둘째, AvgSMP의 성분을 분해한 후 확률적 성분을 대상으로 분석하는 두 가지 경우로 나눌 수 있다. [표 8]은 두 그룹에 대한 예측 모형 분석 결과를 보여주고 있다.

[표 8] 모형의 예측 추정 성과

여기서 수준변수는 종속변수 혹은 분석 대상 변수로 AvgSMP를 사용한 경우를 나타내며, 확률적 성분은 AvgSMP의 확률적 성분을 사용한 경우를 나타낸다. ARDL(Autoregressive Distributed Lag)은 2개 변수 AvgSMP와 DieselPrice를 바탕으로 설정하였다. LSTM은 'Long Short-Term Memory'을, AR8은 시차 8을 갖는 AR(Autoregressive) 모형을 의미한다. AR8 이후의 기호는 확률적 성분에 대한 시차 8을 갖는 AR 모형의 전차에 대해 각각 점프(Jump), 이중지수 점프(DEJD), GARCH(1,1), GJR-GARCH(1,1), EGARCH(1,1)을 사용한 예측 결과를 나타낸다.

구분	Model	MAE	MSPE	MAPE	DM 통계량	DM p-값
수준변수	ARDL	10.87994	0.01096	6.89839	-	-
	LSTM	11.42062	0.00981	7.06649	2.91634	0.00354
성분 분해	AR8	11.58533	0.01160	7.34978	1.64386	0.10020
	AR8+Jump	10.04346	0.00813	6.33519	-2.00584	0.04487
	AR8+DEJD	10.88386	0.01049	6.86673	0.00830	0.99338
	AR8+GARCH	12.68366	0.01441	8.22611	4.02583	0.00006
	AR8+GJR_GARCH	12.67082	0.01439	8.21772	3.99868	0.00006
	AR8+EGARCH	12.73761	0.01450	8.26133	4.13947	0.00003

분석 결과는 확률적 성분의 AR과 점프의 결합 모형이 가장 탁월한 것으로 나타났다. 점프의 경우에는 단일 점프나 이중지수 점프 좋은 성과를 보였다. 그러나 확률적 성분의 모든 모

형이 AvgSMP 수준변수에 비해 우수한 성능을 갖지는 않았다.

수준변수를 사용한 ARDL과 LSTM 모형은 확률적 성분 모형을 사용한 GARCH 계열 모형에 비해 우수한 예측 성능을 보여주고 있다. ARDL 모형의 높은 성능은 AvgSMP의 자기회귀와 DieselPrice 변수를 모두 고려할 필요가 있음을 시사하고 있다. 이는 에너지 가격이 일별 전기가격의 예측에 필요한 것임을 의미한다. 수준변수 AvgSMP를 딥러닝으로 예측한 결과도 좋은 성과를 보여주고 있다. 그러나 딥러닝을 수행하기 위해서는 은닉층의 수와 뉴런의 유닛 수, 윈도우의 크기, 최적화 알고리즘, 손실함수 등 고려해야 할 요소가 매우 많다. 이러한 요소는 데이터의 형태와 연구자에 따라 달라지므로 딥러닝을 통한 예측은 많은 주의를 기울여 평가해야 한다.

모든 모형 중 확률적 성분에 대한 AR8+Jump 모형이 모든 지표와 DM 통계량에서 가장 우수한 성과를 나타내고 있다. 이는 확률적 성분이 자기회귀를 갖고 있으며, 잔차는 점프 속성을 갖고 있음을 의미한다. AR8+Jump는 비대칭적 점프를 가정한 AR8+DEJD보다 모든 지표에서 다소 우수한 예측 성과를 보이고 있어 점프가 대칭에 가깝다고 해석할 수 있다.

확률적 성분을 분석한 AR8 모형의 잔차에 GARCH 계열 모형을 적용한 경우는 이를 적용하지 않은 AR8 모형에 비해 예측 성과가 낮게 나타났다. 이는 AR8 모형 잔차의 이분산성이 크지 않다는 의미로 해석할 수 있다. GARCH 계열 중에서는 GJR-GARCH(1,1) 모형이 GARCH(1,1)이나 EGARCH(1,1)에 비해 예측 성능이 다소 높은 것으로 나타났다. 이는 AR8 모형 잔차의 이분산성은 크지 않지만 음의 충격이 양의 충격보다 변동성을 증가시키는 효과를 갖고 있음을 의미하는 것이다.

확률적 성분을 종속변수로, DieselPrice의 차분과 기후 관련 변수를 사용한 회귀모형의 예측 결과는 매우 낮은 성과를 보이고 있다. 이는 결정적 성분의 장기 추세가 이들 변수의 영향을 이미 어느 정도 반영하고 있기 때문인 것으로 판단한다.

확률적 성분을 대상으로 예측을 수행하는 경우 예측 결과를 장기추세와 캘린더 효과를 합하여 예측 AvgSMP를 구하므로, 추정 과정에서 여러 단계가 필요하다. 따라서 오차의 개입 가능성이 커질 수 있다. 그럼에도 불구하고 확률적 성분을 사용한 예측 결과가 나쁘지 않은 것은 결정적 성분을 제외함으로써 예측 대상의 값이 변동할 수 있는 범위를 축소하는 것이 예측 성능 제고에 효과적일 수 있음을 의미한다.

6. 요약 및 결론

본 연구는 한국의 도매 전기가격인 SMP의 일별 평균 가격을 예측하고 있다. 이를 위해 SMP에 영향을 미칠 수 있는 여러 변수를 찾았으며, 다양한 모형을 적용하여 예측 성능을 평가하였다.

분석에 사용된 변수는 경유 가격, 전국 평균 온도, 평균 풍속, 평균 상대습도, 일사합 등이

었으며, 요일 및 휴일, 연간 효과 등의 캘린더 효과도 고려하였다. 변수들 사이에는 매우 높은 상관관계를 갖는 경우도 많고, 시계열 안정성을 갖지 못하는 경우도 있어 예측 모형에 적용하기 위해서는 변수의 신중한 선택이 필요함을 확인하였다.

분석은 일별 평균 SMP 수준변수를 그대로 사용하거나 혹은 수준변수를 결정적 및 확률적 성분으로 분해하는 두 가지 접근 방법으로 수행되었다. 수준변수를 사용한 예측은 ECM과 VECM, ARDL 모형, LSTM을 사용한 딥러닝 등으로 하였으며, 이 중에서 ARDL 모형과 LSTM 모형이 신뢰할 수 있는 예측 성과를 보여주었다.

수준변수의 확률적 성분을 분석하기 위해 다양한 시계열 회귀를 시도하였다. 이에는 OLS, LASSO, GLS, WLS 등이 있으며, 이중 WLS만이 통계적 가정을 충족시켰다. 그러나 WLS는 전체 예측 모형 중 가장 낮은 예측 성능을 보여주었다. 이는 분석에 사용한 독립변수가 확률적 성분에 대한 설명력이 매우 낮기 때문이다. 따라서 전기가격의 예측에서 시계열 회귀모형은 전체적으로 그 유용성이 높지 않음을 발견하였다.

확률적 성분을 자기회귀나 점프, GARCH 계열 모형 등을 고려하여 분석한 결과 자기회귀를 수행한 후의 잔차에 점프를 고려하는 것이 가장 좋은 성과를 보이고 있음을 확인하였다. 이는 확률적 성분에 의한 분석이 일부 모형에서는 매우 우수한 예측 성능을 가질 수 있음을 의미한다.

분석을 통해 전기가격의 예측을 위해 수준변수나 확률적 성분 어느 것을 사용하더라도 실제 예측 모형에 따라 그 결과가 매우 달라질 수 있음을 확인하였다. 그러나 확률적 성분에 의한 분석이 많은 실행 과정과 시간을 필요로 하지만 잡음의 크기를 줄일 수 있다는 점에서 예측 성능을 제고할 수 있는 보다 바람직한 접근 방법이라고 판단한다.

7. 참고문헌

Arias-Calluari, K., Najafi M. N., Harre M. S., Tang Y. and Alonso-Marroquin F., 2022, 'Testing Stationarity of the Detrended Price Return in Stock Markets', *Physical A*, 587, 126487.

Billé, A. G., Gianfreda, A., Grosso, F. D. and Ravazzolo, F., 2023, 'Forecasting Electricity Prices with Expert, Linear, and Nonlinear Models', *International Journal of Forecasting*, 39, 570-586

Castelli, M., Groznic, A. and Popovič, A., 2020, 'Forecasting Electricity Prices: A Machine Learning Approach', *Algorithms*, 13, 119.

Choudhry, T., 1994, 'Stochastic Trends and Stock Prices: an International Inquiry', *Applied Financial Economics*, 4, 383-390.

- Escritano A., Pena J. I. and Villaplana P., 2011, 'Modelling Electricity Prices; International Evidence', *Oxford Bulletin of Economics and Statistics*, 73(5), 622-650
- Fan W., He Y., Hao L. and Wu F., 2024, 'Do High House Prices Promote the Development of China's Real Economy? Empirical Evidence Based on the Decomposition of Real Estate Price', *Plos One*, 19(1), 0295311.
- Iftikhar H., Turpo-Chaparro J. E., Rodrigues P. C. and López-Gonzales J. L., 2023, 'Forecasting Day-Ahead Electricity Prices for the Italian Electricity Market Using a New Decomposition—Combination Technique', *Energies*, 16, 6669.
- Jan F., Shah I. and Ali S., 2022, 'Short-Term Electricity Prices Forecasting Using Functional Time Series Analysis', *Energies*, 15, 3423.
- Kapoor, G. and Wichitaksorn N., 2023, 'Electricity Price Forecasting in New Zealand: A Comparative Analysis of Statistical and Machine Learning Models with Feature Selection', *Applied Energy*, 347, 121446.
- Lago, J., Marcjasz, G., Schutter, B. D. and Weron, R., 2021, 'Forecasting day-ahead electricity prices: A review of state-of-the-art algorithms, best practices and an open-access benchmark', *Applied Energy*, 293, 116983.
- Lee, J, List J. A. and Strazicich M. C., 2006, 'Non-Renewable Resource Prices: Deterministic or Stochastic Trends?', *Journal of Environmental Economics and Management*, 51, 354-370.
- Lehna, M., Scheller, F. and Herwartz H., 2022, 'Forecasting day-ahead electricity prices: A comparison of time series and neural network models taking external regressors into account', *Energy Economics*, 106, 105742.
- Mayer, K. and Trück, S., 2018, 'Electricity markets around the world', *Journal of Commodity Markets*, 9, 77-100.
- Poggi A., Persio L. D. and Ehrhardt M., 2023, 'Electricity Price Forecasting via Statistical and Deep Learning Approaches: The German Case', *Applied Math*, 3, 316-342.
- Shah I., Bibi H., Ali S., Wang L. and Yue Z., 2020, 'Forecasting One-Day-Ahead Electricity Prices for Italian Electricity Market Using Parametric and Nonparametric Approaches', *IEEE Access* 8(1), 123104-123113.
- Weron, R., 2014, 'Electricity price forecasting: A review of the state-of-the-art with a look into the future', *International Journal of Forecasting*, 30, 1030-1081.