

---

# 전세가 예측을 이용한 역전세 조기경보모형 개발

: ARIMA 및 머신러닝 기법을 활용하여

---



**중앙대학교 경제학부 S-kian**

주소영   박영균   김지수   이서현   전재인   원지우   한지훈

## 국문 초록

본 연구는 고전적 시계열 분석 기법의 틀에서 벗어난 머신러닝 기법을 활용하여 주택시장의 역전세 피해 조기경보모형(EWS)을 제작한다. 다양한 거시경제적 선행변수에 더불어 부동산을 소비하는 가계의 동태적 효용 최적화 모형을 이용하여 도출한 균형 부동산가격의 경제 변수들을 함께 고려하여 실제 데이터를 바탕으로 그 영향력을 분석한다. 선택된 독립변수와 종속변수인 전세가격지수를 활용하여 ARIMA 및 머신러닝 모델을 통해 예측을 실시하며, 성능이 가장 좋은 XGBoost의 예측 데이터로 신호접근법 기반 조기경보모형을 설계한다.

ARIMA 및 Linear Regression, SVM, Random Forest, XGBoost를 이용하여 예측한 결과 시계열 예측 모델인 ARIMA에 비해 머신러닝을 이용한 예측 성능이 높은 것으로 나타났다. 이는 신호접근법 기반 조기경보모형의 새로운 접근 가능성을 시사하고, 제작된 조기경보모형을 이용하여 역전세 피해가 발생하기 전 실질임금과 부동산 담보대출 이자율의 조정 등을 통해 부동산 시장 침체의 피해를 줄일 수 있으리라 기대한다.

키워드: 역전세, 동태적 최적화 모형, ARIMA, 머신러닝, 조기경보모형(EWS)

# 목 차

## I. 서 론

1. 연구 배경과 목적
2. 주택가격지수
3. 조기경보시스템(EWS)
4. 선행 연구 검토

## II. 본 론

1. 주택가격에 대한 선행변수 설정
  - (1) 내구재 소비 포함 가계선택변수의 동태적 최적화 모형
  - (2) 부동산 시장과 관련된 거시경제지표
2. 조기경보모형 제작
  - (1) 데이터 출처 및 설명
  - (2) ARIMA 및 머신러닝 기법 예측력 평가
  - (3) 조기경보모형의 개발

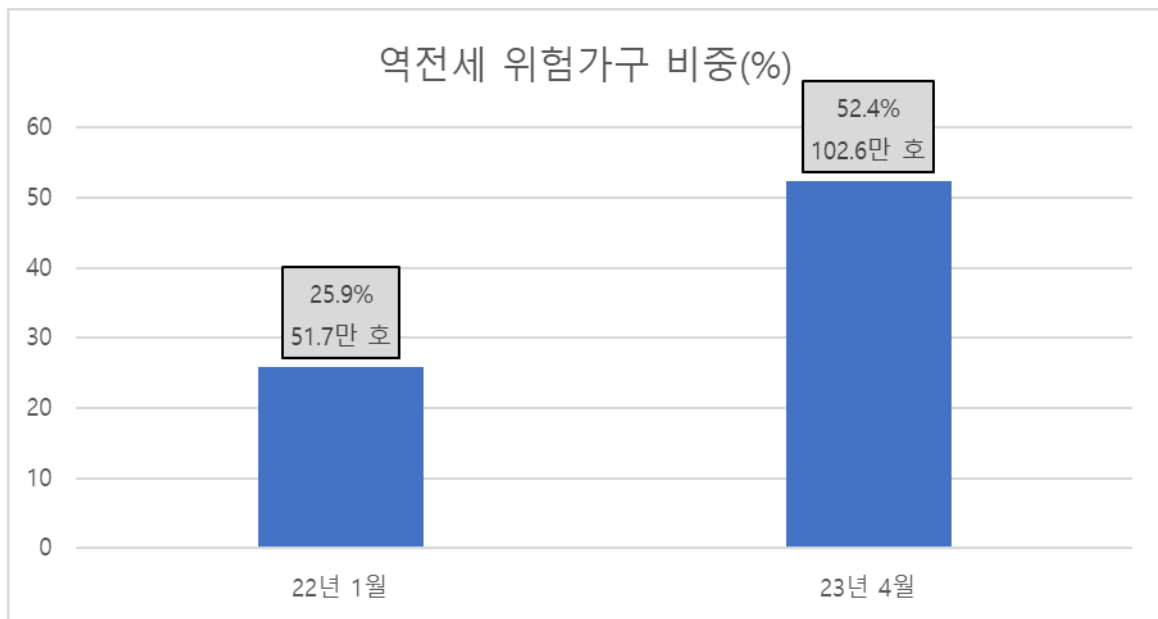
## III. 결 론

1. 부동산 시장 조기경보모형의 경제적 의의
2. 연구의 의의 및 한계

# I. 서론

## 1. 연구 배경과 목적

역전세란 전세가격이 하락하여 전세 계약이 만료된 시점에 임대인이 임차인에게 전세금을 반환하지 못하는 상황을 의미한다. 이러한 역전세 현상은 금리 인상의 영향이 크다. 전세대출금리가 상승하여 높은 이자 부담이 우려될 경우, 전세 거주를 포기하고 월세 거주를 택하는 사람들이 늘어난다.<sup>1</sup> 전세 수요 감소로 인해 전세 가격이 하락하고, 전세 가격이 하락하면 전세 계약이 만료된 임대인은 기존 계약을 갱신하려 하지 않고 보증금을 돌려받아 새로운 전세 계약을 하거나 월세 계약을 하고자 한다. 따라서 임대인은 기존의 높았던 전세 보증금은 돌려줘야 하고, 새로운 계약을 할 땐 이전보다 낮은 전세가로 계약을 해야 한다. 그러나 무리한 갭투자로 주택을 구매한 임대인은 여유 자금 부족으로 인해 만료된 전세 보증금을 돌려주지 못하는 역전세 상황에 직면한다. 전세가격 급감과 더불어 주택가격까지 하락할 경우, 단순히 새로운 전세가격으로 기존의 전세 보증금을 충당하지 못하는 역전세에 그치지 않을 수 있다. 주택을 경매에 넘기더라도 임차인의 보증금을 돌려주지 못하는 깡통전세 위험이 발생할 수 있다.<sup>2</sup> 이처럼 주택 시장의 불안정성으로 인한 주택 가격 급감은 임대인과 임차인에게 큰 경제적 손실을 안겨줄 수 있다.



각 시점의 잔존 전세계약(월세가 전혀 없는 순수 전세 기준) 중 기존 전세보다 최근 전세시세(주소, 건축년도 및 평형이 동일한 주택의 6개월 내 최근 전세가격)를 초과하는 주택을

<sup>1</sup> 이승연. 2022. '전세대출 금리 8%대 목전...'차라리 월세가 싸'. 파이낸셜뉴스. 6 December.

<sup>2</sup> 민병철. (2023). 깡통전세의 발생 확률 추정. 주택도시보증공사

역전세 위험가구라고 정의한다.<sup>3</sup> 다음 표는 2023년 5월 한국은행에서 발표한 역전세의 위험가구 비중 증가분이다. 22년 이후 전세가의 급격한 하락으로 역전세 위험가구가 급증하여 주택 시장에서의 불안감이 급증하였다. 만약 주택시장의 흐름을 파악할 수 있다면 이러한 주택시장 문제 발생 여부를 조기에 파악하고 관련 정책에 반영할 수 있다. 특히 아시아의 경우 개인 자산에서 부동산의 비중이 80%인 만큼 주택시장의 동향을 예측하는 것은 필수적이라 볼 수 있다.<sup>4</sup> 특히 한국의 가계는 보유자산 가운데 부동산이 가장 큰 비중을 차지하기 때문에 유동성 증가와 이자율 하락 부분을 중점으로 하여 부동산 가격의 급락에 관련한 연구는 활발히 이루어졌다. 그러나 주택담보대출 이자율 및 가계의 노동소득, 비내구재 소비량과 같은 미시적 요소를 고려하여 예측하는 연구는 변수에 중요성에 비해 활발하지 않다. 따라서 본 연구에서는 이러한 주택시장 동향 예측을 위해 주택가격지수를 종속변수로 하며, 다양한 거시적 경제변수에 더불어 미시적 변수도 고려한 데이터 모델을 설정하여 조기경보모형을 제작하였다.

## 2. 주택가격지수

주택가격지수(Housing Price Index)란 주택 시장에서의 가격 변동을 측정하는 지표로, 주택시장의 평균적인 가격변화를 측정하여 주택 시장의 전체적 동향을 판단하는 중요한 지표로 고려된다. 따라서 주택가격지수를 활용한 경제적 분석 활동은 전세사기 및 역전세와 같은 대한민국 주택시장의 문제를 해결하는데 그 의의가 있다.

주택가격지수는 기준시점에 대한 상대가격으로 표현된다. 우리나라의 경우 월간 기준시점은 2021.6=100, 주간 기준시점은 2021.6.28=100으로 한다.<sup>5</sup> 기하평균 산술 방식을 적용해 구하는데, 이는 최소공표단위지수와 이를 활용한 광역단위지수로 구분된다. 최소공표단위지수의 경우 아파트는 시군구별로, 단독주택과 연립주택은 시도 생활권역별로 지역을 구분한다. 즉 아파트의 경우, 단독주택과 연립주택보다 더욱 세분화된다. 산식은 다음과 같다.

1) 최소공표단위지수 산식 (아파트)

$$I_j = \prod_{i=1}^{n_j} x_i^{1/n_j}$$

단,  $I_j = j$  지역 지수,  $x_i = \frac{p_i^t}{p_i^0}$

$p_i^0$ : i 표본의 기준시점 조사가격

<sup>3</sup> 한국은행 2023년 5월 경제전망보고서

<sup>4</sup> 권영규. (2023). 부동산가격에 대한 결정요인 분석. 한국해양대학교 대학원

<sup>5</sup> 한국부동산원 부동산통계정보시스템(R-ONE)>전국주택가격동향조사>3. 지수 산정 및 편제

$P_i^t$ : i 표본의 t 시점(비교시점) 조사가격

$n_j$ : j 지역 표본 수

2) 광역단위지수 산식 (단독, 연립주택)

$$I_d = I_{d1}^{h_{d1}} \times I_{d2}^{h_{d2}}$$

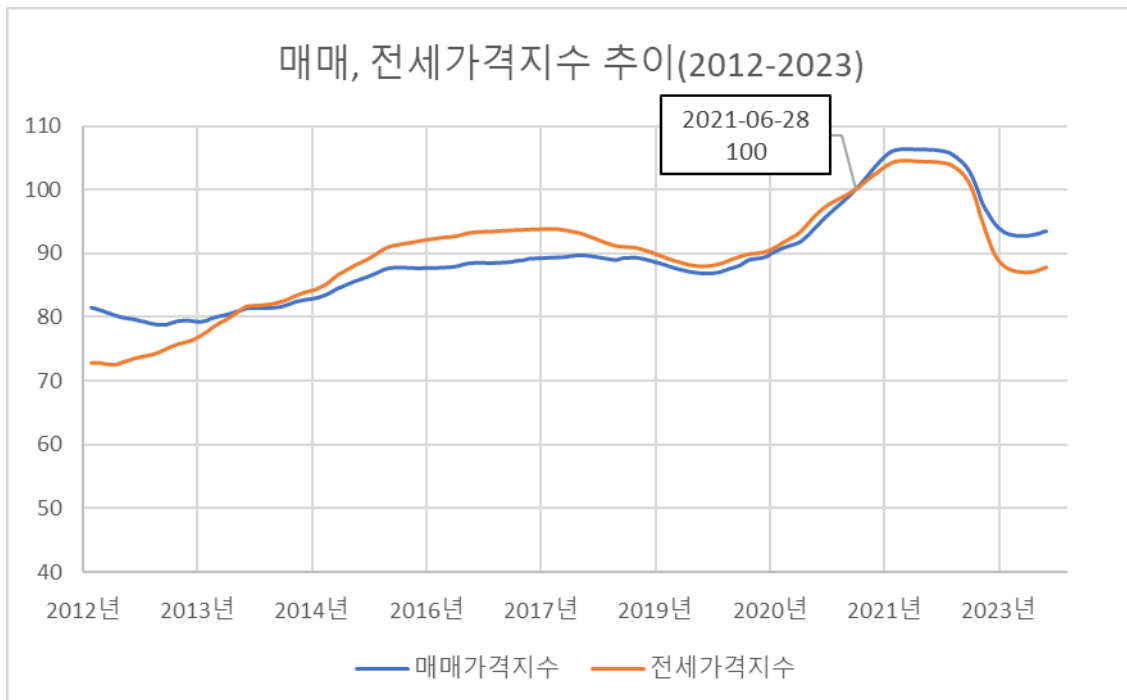
단,  $I_d$  = d 광역지역 지수

$I_{d1}, I_{d2}$ : d 광역지역에 포함된 최소 공표지역 지수들

$h_{d1}, h_{d2}$ : 최소 공표지역별 가중치

월간 주택가격지수 변동률(%): {(당월지수 - 전월지수) / 전월지수} \* 100

주간 주택가격지수 변동률(%): {(당주지수 - 전주지수) / 전주지수} \* 100



해당 산식으로 도출된 주택 매매가격 및 전세가격지수의 추이는 다음과 같다.

### 3. 조기경보시스템(EWS)

조기경보모형 또는 조기경보시스템(EWS, Early Warning System)은 특정 지표 또는 자료에 대한 지속적인 모니터링을 통해 향후 발생할 위기에 대응하는 것을 목적으로 하는 시스템이다. 예를 들어, 외환시장의 경우 외환시장에 영향을 미친다고 판단되는 환율 변동, 정책금리 및 외환시장 개입을 모니터링하고 이를 종합하여 외환시장에 대한 충격의 규모를 예측한다. 이렇게 예측된 수치를 바탕으로 외환위기가 발생할 가능성을 추정하고, 이 수치가 일정 수준을 초과할 시 경보를 울리는 방식으로 조기경보체제가 운영된다.

EWS는 사회의 여러 분야에서 활용되고 있다. 가령, 각 지역에서는 농업기상재해 조기경보시스템을 구축하여 다가올 재난에 대비하고 있으며, 실시간 의약품 사용현황을 이용해 일부 감염병 확산에 대한 조기경보시스템을 연구한 사례도 있다.<sup>6</sup> 특히 경제 분야에서 한국 정부는 1997년 외환위기 이후 금융시스템 개혁이 필요해짐에 따라 외환위기 재발 방지를 위한 국가적인 조기경보시스템을 구축하였다.

한국은 외환시장, 국내 금융시장, 국내 금융기관, 석유, 원자재, 노동시장, 부동산시장의 총 7개 부문에 대한 조기경보시스템을 운영하고 있다. 각 부문에 대해 모형담당기관이 모형을 운영하고 분석하면 기획재정부는 7개 부문의 분석결과를 토대로 국내 경제위기의 발생 가능성을 최종적으로 종합하여 청와대에 보고한다.

이러한 조기경보시스템을 구축하는 방법론에는 주로 신호접근법이 사용된다.<sup>7</sup> 기존의 신호접근법은 선행지표를 활용해 위기 신호를 발생시키는 방식으로 위기 설명력이 높을 것으로 예상되는 적절한 선행지표를 선정해 임계값을 넘으면 신호가 발생한다고 판단한다. 이후 통계적으로 예측력이 높은 선행지표의 신호에 더 높은 가중치를 부여해 선행지표 신호를 가중합한 종합선행지수를 구성한다. 이러한 종합선행지수는 위기 전 상승하는 경향을 보임으로써 위기 신호 발생 사실을 인지시킨다.

이러한 조기경보모형은 위험요인을 적시에 평가하는 유용한 방법일 수 있으나 그 신뢰성에 대한 의문이 있다.<sup>8</sup> 기존 시계열 모형을 기반으로 한 신호접근법에서 조기경보모형을 구축하는 선행지표는 과거 데이터를 반영하지만 위기는 매번 같은 형태로 발생하지 않고 발전한다. 따라서 빠르게 변화하는 시장에 대한 예측 모형을 구축하는 경우 새로운 변수를 추가해야 할 필요성이 있다. 그러나 고전적 시계열 이론을 바탕으로 한 조기경보모형에 새로운 변수를 자유롭게 추가할 경우 엄격한 가설 검증을 거쳐야 하고, 가설 검증을 통과하지 못한다면 해당 지표가 위험을 예측하는 지표로 이용 가능한지에 대한 불안감이 생길 수 있다. 이러한 예측력 불신 문제를 회피하기 위해 기존의 검증된 소수의 선행지표로 위기를 예측할 경우, 해당 지표가 새로운 형태의 위기에 반응하지 않을 가능성이 존재한다. 결국 위기 예측 능력이 실제 위기와 동떨어질 수 있다.

부동산 시장은 다양한 경제 변수로부터 영향을 받아 시시각각 변화한다. 따라서 부동산 시장에 영향을 미치는 새롭고 다양한 변수를 반영하기 어려운 신호접근법은 부동산 시장을 예측하는 데 한계가 있을 수 있다. 반면 머신러닝은 고전적 시계열 모형에 비해 변수를 자유롭게

---

<sup>6</sup> 김동숙. (2016). 실시간 의약품 사용자료를 활용한 일부 감염병 유행 조기경보 시스템. 건강보험 심사평가원

<sup>7</sup> 이항용. (2015). 경제발전경험모듈화사업: 금융위기 조기경보시스템. 기획재정부

<sup>8</sup> 박원암. (2011). 2008년 위기 예측 가능했나?:신호접근법 분석. 對外經濟研究 15권 3호

추가 가능하다. 따라서 머신러닝을 활용한 조기경보모형을 구축할 경우 부동산 시장에 영향을 미치는 변수를 폭넓게 반영해 보다 현실에 부합하는 부동산 시장 예측이 가능하리라 예상된다.

## 4. 선행연구 검토

선행 연구에 따르면 주택은 일반 재화처럼 사용을 통한 효용이 존재할 뿐만 아니라, 재건축 등을 통하여 감가상각을 초과하는 자본이득을 기대할 수 있어 소비재와 투자자산의 속성을 동시에 지니고 있다.<sup>9</sup> (강희만 외, 2012) 주택은 공급 기간이 평균적으로 수년 이상 소요되기 때문에 주택공급은 상당히 비탄력적이다. 이로 인해 주택시장은 공급보다는 수요의 변화에 따라 가격이 지속해서 상승하거나 하락할 가능성이 있다.<sup>10</sup> (류지수 외, 2007) 따라서 본 연구에서는 주택이라는 소비재이자 자산을 소비하는 가계의 효용함수와 예산선을 설정하고, 주택의 공급은 경제적 외생변수로 판단하여 주택시장의 균형 가격을 추론하였다.

주택시장의 위험을 예측하는 기존 연구와 조기경보체계는 주로 고전적인 시계열 예측모형에 의존한다<sup>11</sup>(손정식 외, 2003). 그러나, 주택 정책을 비롯한 다양한 변수의 영향이 혼재하는 주택시장은 단순한 시계열 모형으로 예측하기에 적합하지 않을 수 있다. 로지스틱 회귀분석, SVM, RF, ANN의 네 모형의 성과를 비교한 결과, 기존에 사용되었던 로지스틱 회귀보다 머신러닝 및 인공지능망을 이용한 주택가격 예측의 성과가 높게 나타났다.<sup>12</sup> (박대현 외, 2021) 따라서 본 연구에서는, 시계열 이론 바탕 모델인 ARIMA 및 기본적인 머신러닝 모델과 앙상블 기법을 활용한 새로운 조기경보모형을 구축하여 그 성과를 비교하였다.

## II. 본론

### 1. 주택가격에 대한 선행변수 설정

#### (1) 내구재 소비 포함 가계선택변수의 동태적 최적화 모형

먼저 부동산을 통한 가계의 총효용과 예산제약식을 이용하여 부동산 시장의 균형가격을 도출한다. 부동산은 내구재적 특성을 가지며 투자자산이므로, 자산가격 결정이론(Asset Pricing Theory)을 활용하여 투자자산의 관점에서 부동산 기본가치를 평가한다.<sup>13</sup> 다시 말해, 합리적

---

<sup>9</sup> 강희만, 김정렬. (2012). 투기적 특성이 아파트가격에 미치는 영향 분석. 한국산업경제학회

<sup>10</sup> 류지수, 임규채, 기석도. (2007). 주택가격 및 토지 가격의 거품존재 가능성 검토. 한국산업경제학회

<sup>11</sup> 손정식, 김관영, 김용순. (2003). 부동산가격 예측 모형에 관한 연구. 한국주택학회

<sup>12</sup> 박대현, 김정한, 류두진. (2021). 기계학습 기반 주택시장의 조기경보체계. 한국부동산원

<sup>13</sup> 권영규. (2023). 부동산가격에 대한 결정요인 분석. 한국해양대학교 대학원



경제주체가 부동산이라는 투자자산을 통해 일생에 걸친 효용극대화를 위한 소비 선택 결정을 한다고 가정한다. 이때 미래의 불확실한 소비에 의한 효용과 현재의 확실한 소비에 의한 효용을 비교하여 선택함으로써 최적 소비경로를 결정하고, 이것을 만족하는 제약 조건하에서 자산균형가격을 도출한다.

장기적 동태 모형에서 가계가 유일하게 구매하는 내구재로 부동산을 가정한다. 가계는 비내구재 소비( $C_t$ )와 내구재인 부동산( $H_t$ ), 그리고 여가( $h_t$ )로 구성된 효용함수 식(1)을 설정한다. 이때 합리적 가계는 식(2)와 식(3)을 예산제약식으로 자신의 일생에 걸친 효용극대화를 추구한다. 이러한 가계의 효용극대화 조건으로부터 부동산 수요함수를 유도하였다.

$$V_0 = \sum_{t=0}^{\infty} \left( \frac{1}{1+\rho} \right)^t U(C_t, H_t, h_t) \quad \dots\dots\dots(1)$$

$$Z_{t-1}(g_t + y_t) + p_t(1 - \delta)H_{t-1} + L_t + w_t l_t + (1 + r_t)S_{t-1} = C_t + Z_t g_t + p_t H_t(1 + \mu) + (1 + R_{l,t})L_{t-1} + S_t \quad \dots\dots\dots(2)$$

$$L_t = \beta p_t H_t \quad \dots\dots\dots(3)$$

$$h_t + l_t = 1, h_t = 1 - l_t \quad \dots\dots\dots(4)$$

여기서  $V_0$ 는 한 가계가 일생에 걸친 총효용을 나타낸다.  $\rho$ 는 가계의 시간선호율,  $Z_t$ 는 주식 구매량,  $y_t$ 는 배당소득,  $g_t$ 는 주가,  $p_t$ 는 부동산의 실질가격,  $\delta$ 는 부동산의 감가상각율,  $H_t$ 는 부동산의 거래량,  $L_t$ 는 금융기관 차입,  $w_t$ 는 실질임금,  $l_t$ 는 노동시간,  $r_t$ 는 예금이자,  $S_t$ 는 가계 저축,  $C_t$ 는 비내구재 일반소비,  $R_{l,t}$ 은 차입이자율,  $\mu$ 는 부동산 유지비,  $\beta$ 는 부동산담보대출 비율을 나타낸다. 식(2)에서, 가계는 소유주식 $[Z_{t-1}(g_t + y_t)]$ , 소유 부동산 $[p_t(1 - \delta)H_{t-1}]$ , 금융기관 차입( $L_t$ ), 노동소득( $w_t l_t$ ), 그리고 지난 기 저축분 $[(1 + r_t)S_{t-1}]$ 등으로 조달한 자금을 소비( $C_t$ ), 주식구매 ( $Z_t g_t$ ), 부동산 구매 $[p_t H_t(1 + \mu)]$ , 금융기관 차입금 상환 $[(1 + R_{l,t})L_{t-1}]$ , 그리고 이번 기 저축( $S_t$ )에 사용한다. 식(3)에서 금융기관의 차입은 부동산담보대출 가능액( $\beta p_t H_t$ )의 범위에서 이루어진다는 것을 나타낸다. 식(4)에 대하여 가계가 매 기마다 가지는 시간의 단위를 1 이라 가정할 때, 노동시간과 여가시간을 합친 값이 1 이 되므로 여가시간인  $h_t$ 는  $(1 - l_t)$ 와 같다.  $(1 - l_t)$ 와  $L_t$ 를 식(1), (2)에 대입하면 최종적인 효용극대화 목적함수는 다음과 같다. 조세와 인플레이션은 없다고 가정한다.

$$\max V_0 = \sum_{t=0}^{\infty} \left( \frac{1}{1+\rho} \right)^t U(C_t, H_t, 1 - l_t) \quad \dots\dots\dots(5)$$

$$s. t. \quad Z_{t-1}(g_t + y_t) + p_t(1 - \delta)H_{t-1} + \beta p_t H_t + w_t l_t + (1 + r_t)S_{t-1} = C_t + Z_t g_t + p_t H_t(1 + \mu) + (1 + R_{l,t})\beta p_{t-1} H_{t-1} + S_t \quad \dots\dots\dots(6)$$

목적함수인 (5)식을 예산제약식 (6)하에서 극대화한다. 이때 선택변수는 비내구재 소비( $C_t$ ), 부동산 거래량( $H_t$ ), 노동시간( $l_t$ ), 저축량( $S_t$ ), 그리고 주식 구매량( $Z_t$ )이기 때문에 이들 변수에 의한 1 계균형조건(FOC)을 정리하면 다음과 같다.

$$\sum_{t=0}^{\infty} \left( \frac{1}{1+\rho} \right)^t U(C_t, H_t, 1-l_t) + \lambda_t [Z_{t-1}(g_t + y_t) - Z_t g_t - p_t H_t (1 + \mu - \beta) - H_{t-1} \{ \beta p_{t-1} (1 + R_{l,t}) - p_t (1 - \delta) \} + w_t l_t - C_t - S_t + (1 + r_t) S_{t-1}]$$

$$C_t: \left( \frac{1}{1+\rho} \right)^t U_c(C_t, H_t, 1-l_t) = \lambda_t \quad \dots\dots\dots(7)$$

$$H_t: \left( \frac{1}{1+\rho} \right)^t U_H(C_t, H_t, 1-l_t) = \lambda_t p_t (1 + \mu - \beta) + \lambda_{t+1} \{ \beta p_t (1 + R_{l,t+1}) - p_{t+1} (1 - \delta) \} \quad \dots\dots\dots(8)$$

$$l_t: \left( \frac{1}{1+\rho} \right)^t U_l(C_t, H_t, 1-l_t) = w_t \lambda_t \quad \dots\dots\dots(9)$$

$$S_t: \lambda_{t+1} = \frac{\lambda_t}{1+r_{t+1}} \quad \dots\dots\dots(10)$$

$$Z_t: g_t \lambda_t = \lambda_{t+1} (g_{t+1} + y_{t+1}) \quad \dots\dots\dots(11)$$

식 (10)에 식 (11)을 대입하면 다음과 같은 결과를 얻을 수 있다.

$$\frac{\lambda_t}{\lambda_{t+1}} = \frac{g_t}{g_{t+1} + y_{t+1}} = \frac{1}{1 + R_{S,t+1}} \quad (\forall \frac{g_{t+1} + y_{t+1}}{g_t} = 1 + R_{S,t+1})$$

$$1 + r_{t+1} = \frac{g_{t+1} + y_{t+1}}{g_t} = 1 + R_{S,t+1} \quad \dots\dots\dots(12)$$

$(1 + R_{S,t+1} = \frac{g_{t+1} + y_{t+1}}{g_t})$ 은 t 기의 주식수익률을 의미한다. 즉, 식 (12)를 해석하면 합리적 가계가 효용 극대화를 이루는 경제적 균형 상태에서 은행의 저축성 예금이자와 주식 수익률이 같다는 결론에 이른다. 이러한 결과를 식 (7), (8), (9)에 대입하여 비율을 구하고 정리하면 다음과 같은 결과를 얻을 수 있다.

$$\frac{U_l(C_t, H_t, 1-l_t)}{U_c(C_t, H_t, 1-l_t)} = w_t$$

$$\frac{U_H(C_t, H_t, 1-l_t)}{U_c(C_t, H_t, 1-l_t)} = p_t [1 + \mu - \beta + \frac{1 + R_{l,t+1}}{1 + r_{t+1}} \beta - \frac{1 - \delta}{1 + r_{t+1}} p_{t+1}] \quad (\forall \frac{p_{t+1} - p_t}{p_t} = p_{t+1})$$

$(p_{t+1}: t \text{ 기에 예상하는 부동산 가격 상승률})$

가계 효용함수가 Cobb-Douglas 를 따른다고 가정하면, 다음과 같은 부동산가격( $p_t$ ) 균형식을 얻을 수 있다.

$$U(C_t, H_t, 1-l_t) = C_t^\alpha H_t^\beta (1-l_t)^{1-\alpha-\beta} \quad (0 < \alpha < 1, 0 < \beta < 1, \alpha + \beta < 1)$$

$$\frac{U_H(C_t, H_t, 1-l_t)}{U_c(C_t, H_t, 1-l_t)} = \frac{\beta C_t}{\alpha H_t} = \frac{\beta w_t (1-l_t)}{(1-\alpha-\beta) H_t}$$

$$p_t = \frac{\beta w_t (1-l_t)}{(1-\alpha-\beta) H_t} [1 + \mu - \beta + \frac{1 + R_{l,t+1}}{1 + r_{t+1}} \beta - \frac{1 - \delta}{1 + r_{t+1}} (1 + p_{t+1})]^{-1} \dots\dots\dots(13)$$

(13)의 균형가격에 의하면 부동산시장의 가격 결정에 양(+)의 영향을 미치는 변수는 실질임금( $w_t$ )과 t 기의 기대 부동산 상승률( $p_{t+1}$ )이고, 음(-)의 영향을 미치는 변수는 가계 근로시간( $l_t$ ), 대출이자율( $R_{l,t+1}$ ), 그리고 부동산 수, 즉 가계의 t 기 부동산 거래량( $H_t$ )이다.

균형가격에는 예금이자율(FOC 상태에서는 주식수익률과 같음)과 같은 변수도 포함되어 있는데, 해당 변수는 가격 결정에 음과 양의 영향력 모두 미치기 때문에 그 방향성을 단언할 수 없다.

그 이유는 두 가지 논리에서 설명할 수 있다. 예금이자율과 주식수익률 모두 가계의 비내구재 자산 소득 증가율을 나타낸다. 비내구재 자산 소득 증가율이 상승할 경우, 내구재인 주택의 소비를 선호하는 이는 증가한 자산으로 주택을 구매하려 하기 때문에 주택에 대한 수요가 증가하여 주택 가격이 상승한다. 한편 비내구재의 소비를 선호하는 이는 비내구재의 자산 증가를 위해 주식 또는 예금을 위한 자본 마련을 목적으로 많은 자산이 묶여 있는 주택을 팔 것이다. 이러한 소득의 증가율이 미치는 영향의 방향성은 개별 가계주체의 효용 파라미터( $\alpha, \beta$ )에 의해 달라진다. 이러한 파라미터는 가계마다 다르기 때문에 이 부분은 실증 분석을 통한 추가적인 논의가 필요하다.

가계의 선택변수를 통계적으로 검정하기 위해 GLS 회귀분석을 시행하였다. 시계열 자료에서  $t$  기의 오차항과  $t+1$  기의 오차항은 상관관계를 가지는 경우가 많다. 이를 자기상관이라 하는데, 이렇듯 오차항에 자기상관이 존재하거나, 이분산성이 존재하면 OLS 대신 GLS(generalized least squares)를 이용할 수 있다. GLS(generalized least squares)를 이용한 회귀분석에서는 오차항의 분산이 작은 관찰치에는 가중치가 크게 부여되고, 오차항의 분산이 큰 관찰치에는 가중치가 작게 부여된다.<sup>14</sup> 종속변수인 전세가격지수 또한 시계열 데이터이므로 자기상관을 가지기 때문에 GLS 모델을 통해 각 변수의 영향력을 분석하였다.

또한 코로나 19 라는 유례없는 상황이 야기한 주택가격의 급등을 통제하고자, 본 연구에서는 코로나 더미를 생성하였다. 2020 년 1 월 20 일 자 보건복지부 보도자료에 의하면, 질병관리본부는 기사가 발행된 당일에 코로나 19 와 관련하여 위기경보 수준을 '주의'로 상향시켰다.<sup>15</sup> 이에 따라 해당 날짜를 코로나 19 의 시작 시점으로 설정하였다. 그리고 코로나 19 의 종료 시점은 2023 년 8 월 31 일로 설정하였는데, 이는 2023 년 8 월 23 일 자 기사에 따르면, 그날이 코로나 19 가 독감과 같은 일반의료체계로 전환된 날이기 때문이다.<sup>16</sup> 코로나 변수의 영향이 없는 기간은 0, 변수의 영향이 있는 2020/01/20 ~ 2023/08/31 기간은 1 로 설정했다.

주택 전세가격지수 및 독립변수들은 2012 년 5 월 7 일부터 2023 년 7 월 31 일까지의 주간 데이터를 수집하여 분석하였다. 종속변수인 주간 전세가격지수 및 전세거래량은 한국부동산원,<sup>17</sup>

---

<sup>14</sup> 허강성. (2022). 부채계약가설의 적용은 경영성과에 따라 차이가 있는가?: OLS와 GLS 회귀모형 실증결과 비교분석. 국제회계연구, 106, 221-245.

<sup>15</sup>[https://ncov.kdca.go.kr/tcmBoardView.do?brdId=&brdGubun=&dataGubun=&ncvContSeq=352435&contSeq=352435&board\\_id=&gubun=ALL](https://ncov.kdca.go.kr/tcmBoardView.do?brdId=&brdGubun=&dataGubun=&ncvContSeq=352435&contSeq=352435&board_id=&gubun=ALL)

<sup>16</sup> <https://www.fnnews.com/news/202308230923274283>

<sup>17</sup> 한국부동산원 부동산통계정보시스템(R-ONE)

예상 부동산 가격 상승률을 반영하는 부동산 가격 전망지수, 주택담보대출 금리 및 저축성 예금금리는 한국은행의 데이터를 이용하였다.<sup>18</sup> 추가수익률을 반영한 코스피지수의 수익률은 한국거래소,<sup>19</sup> 실질임금은 고용노동부의 데이터를 반영하였다.<sup>20</sup> 가계주체의 노동시간 선택은 데이터를 찾을 수 없어 모델에서 제외되었다. 해당 데이터로 GLS 를 수행한 결과는 다음과 같다.

모형		베타	표준오차	t-value	유의확률	VIF
1	실질임금	0.0184	0.001	24.288	<0.001***	1.672
	주택담보대출금리	9.7100	1.586	6.122	<0.001***	16.975
	저축성예금금리	-9.5629	1.381	-6.925	0.001***	17.620
	전세거래량	0.0003	0.000	1.560	0.119	1.887
	부동산가격전망지수	0.1013	0.023	4.376	<0.001***	1.742
	코스피수익률	-0.1232	0.126	-0.979	0.328	1.003
	코로나더미	5.3427	0.896	5.966	<0.001***	2.077
2	실질임금	0.0199	0.001	26.532	<0.001***	1.659
	주택담보대출금리	-0.8274	0.465	-1.779	0.076*	1.932
	전세거래량	0.0001	0.000	0.580	0.562	1.851
	부동산가격전망지수	0.2146	0.017	12.611	<0.001***	1.413
	코스피수익률	-0.1522	0.131	-1.163	0.245	1.002
	코로나더미	4.9083	0.929	5.284	<0.001***	2.060
3	실질임금	0.0201	0.001	29.408	<0.001***	1.525
	주택담보대출금리	-0.6852	0.402	-1.704	0.089*	1.673
	부동산가격전망지수	0.2156	0.017	12.802	<0.001***	1.392
	코로나더미	5.2039	0.759	6.852	<0.001***	1.377

p-value < 0.1: \*, p-value < 0.05: \*\*, p-value < 0.01: \*\*\*

결과 표를 보면, 부동산 균형가격의 요소들을 모두 사용한 1 단계 모형은 많은 독립변수들이 통계적으로 유의하나, 주택담보 대출금리와 저축성 예금금리 데이터 간 다중공선성이 존재하여 예금금리의 데이터를 제외하였다. 2 단계 모형에서는 다중공선성이 통제되었으나 주택 전세거래량 및 코스피수익률 등 유의하지 않은 변수들이 존재한다. 해당 변수들을 제거한 3 단계 모형에서는 유의수준 10% 하에서 모두 통계적으로 유의하다. 코로나의 영향력을 통제한 모델에서, 장기동태모형을 이용하여 도출된 경제적 변수들의 영향이 실제 데이터와 유사함을 확인했다. 다만 전세거래량은 실제 주택 준공량 및 주택 관련 정책의 영향력이 존재하여, 이론적 예상과

<sup>18</sup> 한국은행 경제통계시스템

<sup>19</sup> 한국거래소 정보데이터시스템(KRX)

<sup>20</sup> 고용노동부 노동시장조사과

달리 실질적인 상관관계를 가지진 않았다. 따라서 본 연구에서는 실질임금 및 주택담보대출금리, 부동산가격전망지수, 그리고 코로나 통제더미를 예측 모델의 독립변수로 사용하였다.

## (2) 부동산 시장과 관련된 거시경제지표

모델의 예측력 증가를 위해, 미시적 경제변수에 더불어 부동산 시장 자체에 미치는 거시경제 변수들을 전세가 예측 모델 독립변수에 추가하였다. 통화는 기본적으로 주택가격과 양의 상관관계를 가진다. 유통되는 통화량이 증가하면 인플레이션율이 높아져 주택가격이 상승할 수 있다. 특히 M2의 설명력이 M1, Lf에 비해 설명력이 가장 뛰어난 것으로 도출되었다.<sup>21</sup>

산업생산지수는 공장, 광산 및 전력회사 등에서 생산한 재화를 월별로 발표하는 지수로 특히 제조업 분야에 직접적인 영향을 미친다. 주택 또한 원자재 생산에 크게 의존하므로 산업생산지수는 주택 공급량에 직결된다. 주택의 수급 비율의 부족은 주택 가격 상승으로 이어진다. 즉 산업생산지수는 부동산 가격과 음의 상관관계를 가진다. 지가는 주택 건설의 기반이 되는 토지의 교환 가격이므로 주택가격과 양의 상관관계를 가진다. 공시지가가 오르면 재산세와 종합부동산세도 함께 증가한다. 때문에 주택을 소유하고 있는 사람들은 주택 가격을 더 높은 가격에 판매하고자 하고, 이로 인하여 주택의 실거래가는 증가한다. 전세수급지수는 100을 기준으로 하여 이보다 낮으면 수요가 더 많음을, 이보다 높으면 공급이 더 많음을 의미한다. 전세수급지수는 2021년부터 100 이하로 떨어져 수위 우월로의 전환을 보이지 못하고 있다. 이러한 상황이 지속됨은 역전세나 강통전세의 위험을 예고하므로 주의 깊게 살펴볼 지표다.

주택가격과 소비자물가지수 또한 높은 양의 상관관계를 보인다. 이론적으로 보면 한 국가의 물가 상승은 자국 화폐 가치의 하락을 초래하기 때문에, 화폐를 보유하기보다는 실물자산인 주택을 보유하려는 욕구가 강해져 결과적으로 주택가격을 상승시키고 국내 생산품의 수출 경쟁력을 하락시키는 요인으로 작동한다.<sup>22</sup> 환율과 주택가격은 시기에 따라 상이한 상관관계를 가진다. 외환위기 이전에는 환율과 주택가격이 양의 상관관계를, 외환위기 이후에는 음의 상관관계를 가진다. 국제금융위기 이후에는 상관관계가 적었으나 음의 상관관계 기조는 유지하고 있는 것으로 나타났다.<sup>23</sup> 이는 정부의 환율 제도 변화에 관련이 있는 것으로 사료된다. 외환위기 이전엔 환율 상승에 따라 이자율이 하락하였으므로 거래 증가로 주택가격이 상승하였으나, 이후엔 지배적 환율 결정 이론이 신축가격모형으로 변화함에 따라 환율 상승이 이자율 상승으로

---

<sup>21</sup> 한준권. (2020). 주택가격과 사회 경제지표의 선후행성 관계 분석. 한밭대학교 산업대학원

<sup>22</sup> 모진성. (2018). 주요 경제변수와 정책변수가 주택가격에 미치는 영향에 관한 연구. 국립목포대학교 행정대학원

<sup>23</sup> 성용림, 유정석. (2013). 주택가격에 대한 환율의 동태적 파급효과 분석. 부동산학보, (54), 244-257.

이어져 반대의 결과를 낳았다.<sup>24</sup> 실질 GDP 또한 주택 시장과 상관관계를 가진다. 실질 GDP가 상승할 경우 일반적으로 수요층의 구매력 상승을 가져와 주택가격을 상승시킨다. 전국 지역에 대해 아파트 가격과 거시경제변수와의 회귀분석을 진행한 결과, GDP는 양의 방향으로 영향을 미치는 것으로 추정되었다.<sup>25</sup> 뉴스심리지수(NSI)는 뉴스 기사를 텍스트 데이터로 사용해 긍정 혹은 부정 기조를 분류한다. 자연어 처리 모형 중 감성분류모형을 통해 논조를 파악한다. 한국은행에 따르면 월별 뉴스심리지수는 소비자심리지수 및 기업경기실사지수 등 공식 경제심리지표에 1~2개월 선행하므로 경제적 예측에 유용하게 사용되는 지표이다. 또한 선행종합지수와 같은 실물경기지표에도 1~2개월 선행하는 등 유용성을 띠기도 했다.<sup>26</sup>

결론적으로 예측 모형의 종속변수는 전세가격지수, 독립변수는 실질임금, 주택담보대출금리, 부동산가격전망지수, 코로나 통제더미, M1, M2, Lf, 산업생산지수, 지가지수, 전세수급동향, 소비자물가지수, 원/달러 환율, 실질 GDP, 뉴스심리지수 및 주택시장의 공급량인 주택준공량을 사용하였다.

## 2. 조기경보모형 제작

### (1) 데이터 출처 및 설명

거시경제적 변수들의 지표 또한 2012년 5월 7일부터 2023년 7월 31일까지의 주간 데이터를 수집하였다. 지가지수 및 전세수급동향은 한국부동산원, 통화량인 M1, M2, Lf의 평잔, 산업생산지수 및 소비자물가지수, 실질 GDP, 원/달러 환율 및 뉴스심리지수는 한국은행, 주택공급 정책이 반영된 주택건설 준공실적은 통계청의 데이터를 활용하였다.<sup>27</sup>

종속변수인 전세가격지수를 그래프로 시각화하면 다음과 같다. 그래프를 보면 2012~2018년까지 전세가격지수는 꾸준히 증가하는 추세를 보였으나, 2018~2019년 후반까지 전세 매물 공급량의 초과로 하락하였다. 그리고 2020년부터 2022년 초까지 코로나 19의 영향으로 화폐공급의 증가 및 환율, 경제지수의 충격으로 전세가는 폭발적으로 상승하였다. 그러나 코로나 19의 경제적 영향력이 점차 감소하며, 부동산담보대출 이자율의 상승과 더불어 전세가는 다시 폭락하였다.

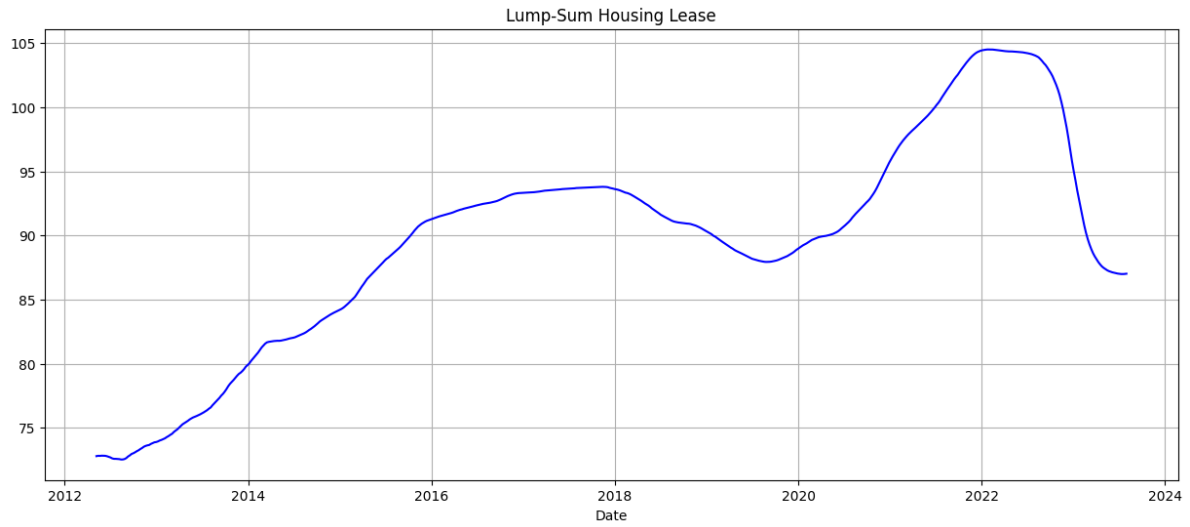
---

<sup>24</sup> 성용림, 유정석. (2013). 주택가격에 대한 거시경제적 요인의 영향 분석: 환율과 금리를 중심으로. 단국대학교

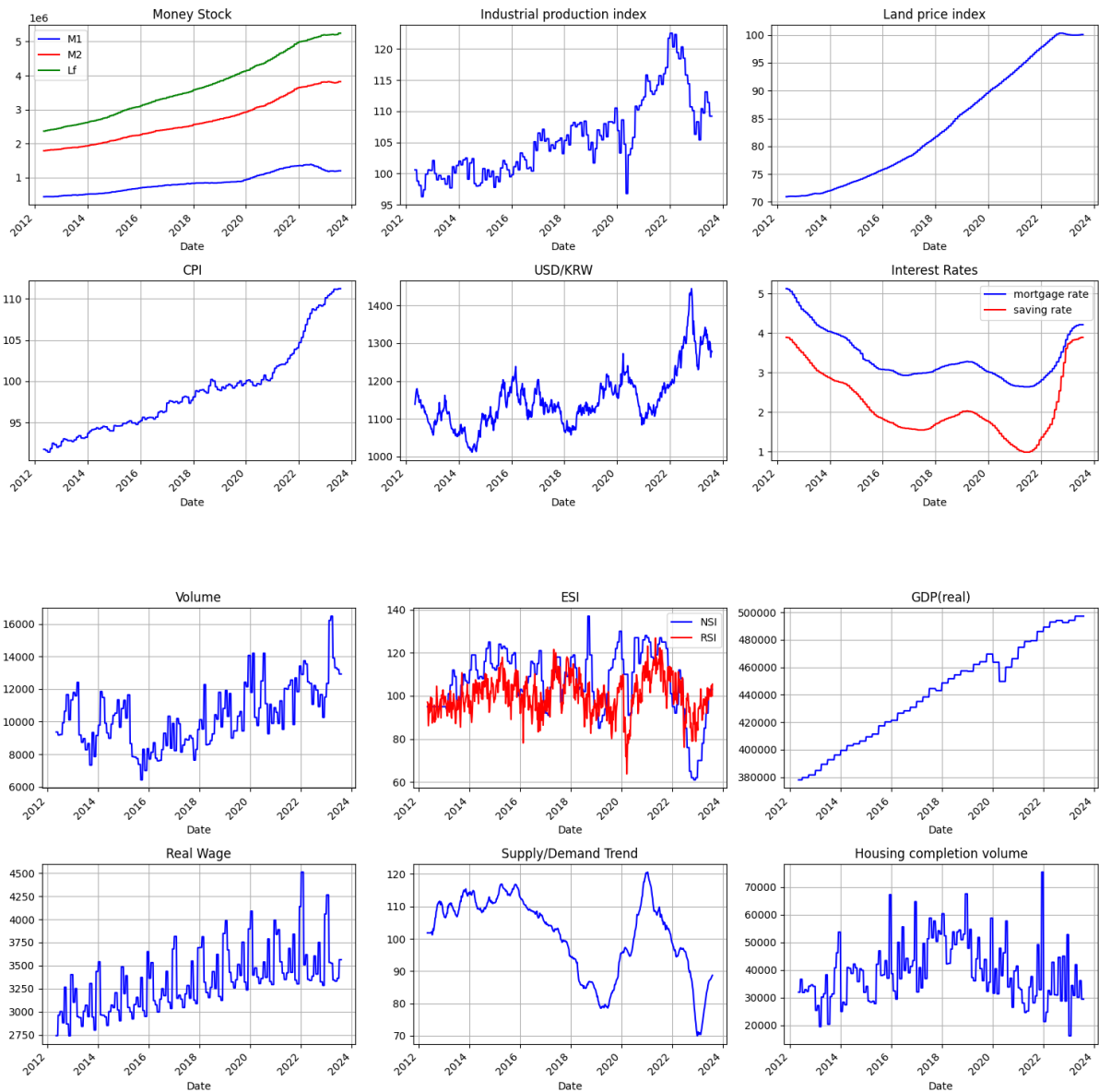
<sup>25</sup> 전건. (2010). 주택가격과 거시경제지표의 관계 분석. 한밭대학교 산업대학원

<sup>26</sup> 서범석, 이영환, 조형배. (2022). 기계학습을 이용한 뉴스심리지수(NSI)의 작성과 활용. 한국은행 경제통계국 통계연구반

<sup>27</sup> 통계청 국가통계포털(KOSIS)



다양한 독립변수들을 그래프로 시각화한 결과는 다음과 같다.

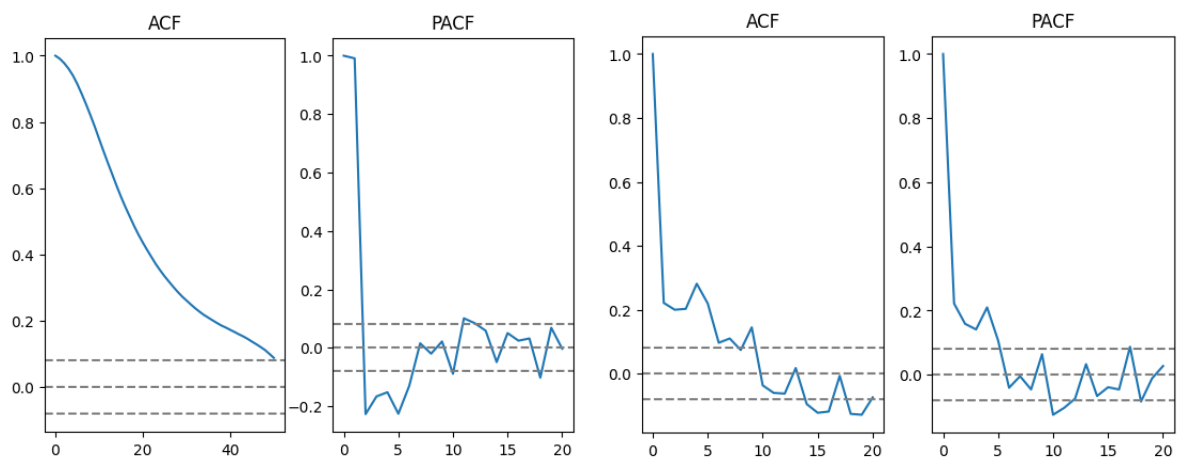


## (2) ARIMA 및 머신러닝 기법 예측력 평가

### - 전통적 시계열 분석: ARIMAX

ARIMA(자기회귀통합이동평균모형)는 Box 와 Jenkins 에 의해 제시된 시계열 예측분석 기법으로 자기회귀, 통합, 이동평균의 요소를 포함한다. 이 모형은 과거의 시계열 데이터를 사용해 간단한 예측을 할 수 있으며 정상성 시계열 데이터를 표현할 수 있다는 장점이 있어 사회과학 분야에서 널리 활용되었다. ARIMA 모형은 불안정한 시계열 데이터의 안정성을 차분을 통해 확보하며, 자기회귀(AR) 과정과 이동평균(MA) 함수를 사용하여 추정한다. 시계열 데이터는 규칙적인 패턴과 불규칙한 패턴의 조합으로 이루어져 있으며, 규칙적인 패턴 중 하나는 자기상관성을 통해 이전 값이 이후 값에 영향을 미치는 상황을 포함한다. ARIMAX 는 이러한 ARIMA 모형에 독립변수들을 추가적으로 고려하여 분석하는 모델이다.

ARIMAX 는 전통적인 시계열 예측분석 기법으로 종속변수 및 독립변수의 정상성을 가정하여 예측을 수행한다. 따라서 데이터의 정상성 확보를 위해 종속변수에 대한 차분을 실시하였다. 1 차 차분한 데이터의 ACF(자기상관함수), PACF(부분 자기상관함수)는 다음의 왼쪽 2 개, 2 차 차분한 데이터의 그래프는 다음의 오른쪽 2 개이다.



ACF 는  $t$  기의 종속변수  $y_t$ 와  $y_{t+k}$  사이의 관계를 측정한 것이며, 수식은 다음과 같다.

$$ACF(k) = \frac{cov(y_t, y_{t+k})}{var(y_t)} = \frac{\sum_{t=1}^{N-k} (y_t - \bar{y})(y_{t+k} - \bar{y})}{\sum_{t=1}^N (y_t - \bar{y})^2} \frac{N}{N-k}$$

$k$  가 커짐에 따라  $ACF(k)$ 의 값도 감소하는데, 다음의 그래프를 보면  $k$ 가 증가함에 따라 ACF는 천천히 감소한다. 즉, 1 차 차분한 데이터가 여전히 강한 자기상관성을 가진다는 뜻이다. 이는 데이터에 대한 2 차 차분이 필요할 수 있음을 시사한다. 따라서 2 차 차분을 진행한 결과 ACF가  $k$ 의 증가에 따라 빠르게 감소하는 추세를 보였다.

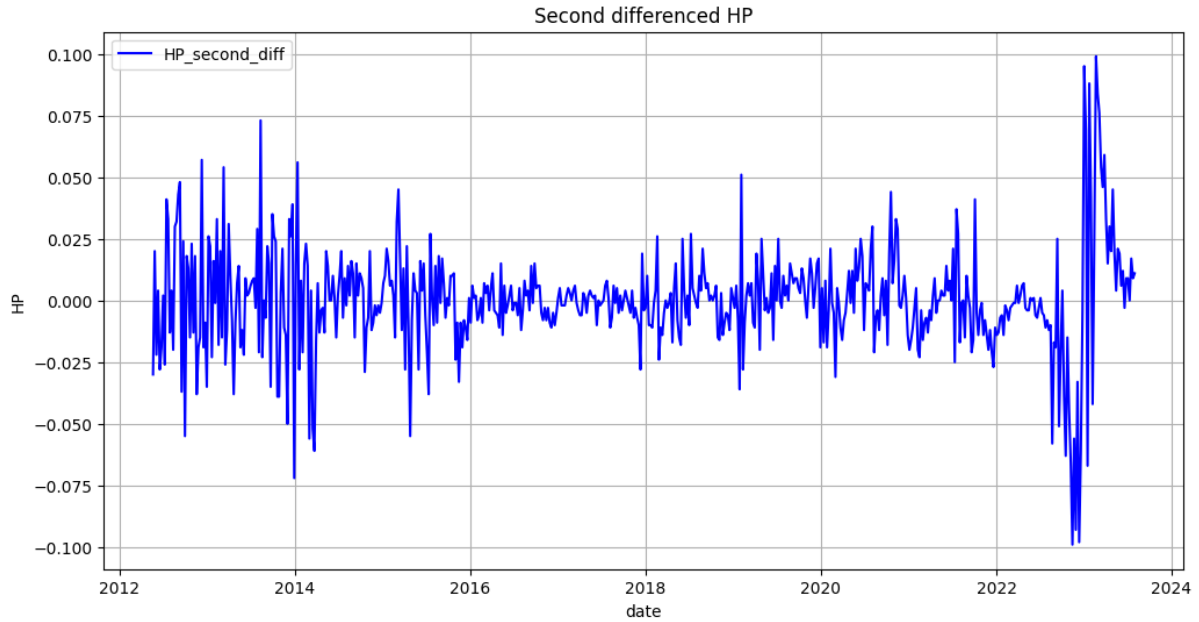
PACF는 동일하게  $y_t$ 와  $y_{t+k}$  사이의 관계를 측정하지만  $t$ 와  $t+k$ 기 사이 다른 값들의 영향력을 배제하고 측정한다. 수식은 다음과 같다.

$$PACF(k) = Corr(e_t, e_{t+k})$$

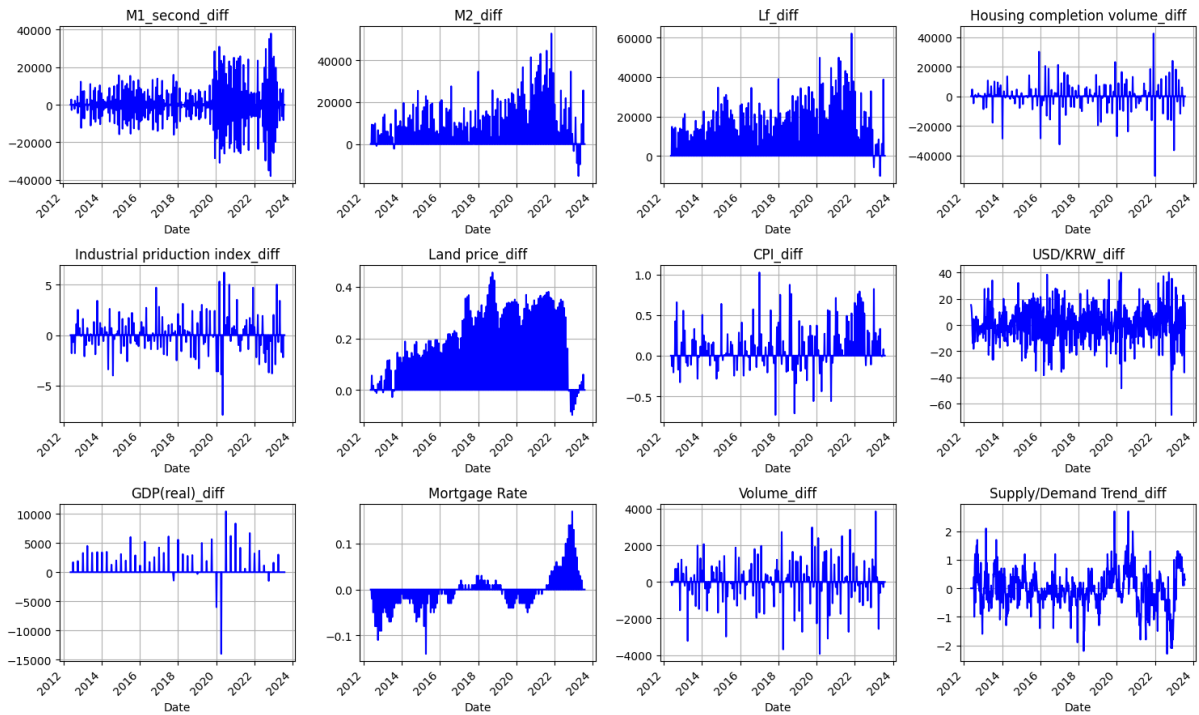


$$e_t = Y_t - (\beta_1 Y_{t-1} + \dots + \beta_{k-1} Y_{t-(k-1)})$$

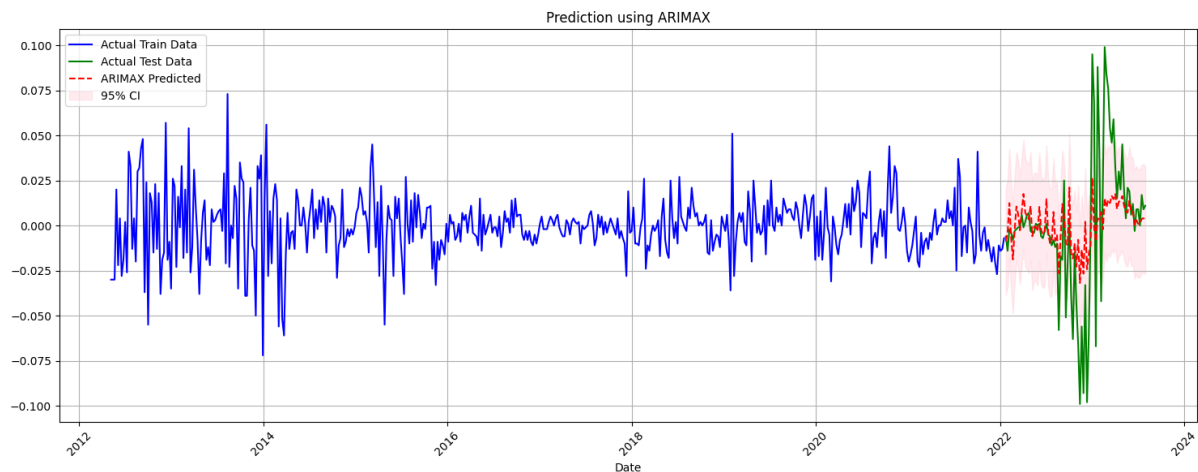
PACF 그래프 또한 2 차 차분 데이터에 대해 k의 증가에 따라 급감한다. 그러나 두 그래프 모두 0 까지 도달하는 데에 약간의 시차가 존재하기 때문에 ARIMA(1, 2, 1) 모델로 예측을 진행하였다. 전세가격지수에 대해 2 차 차분한 데이터의 추이는 다음과 같다.



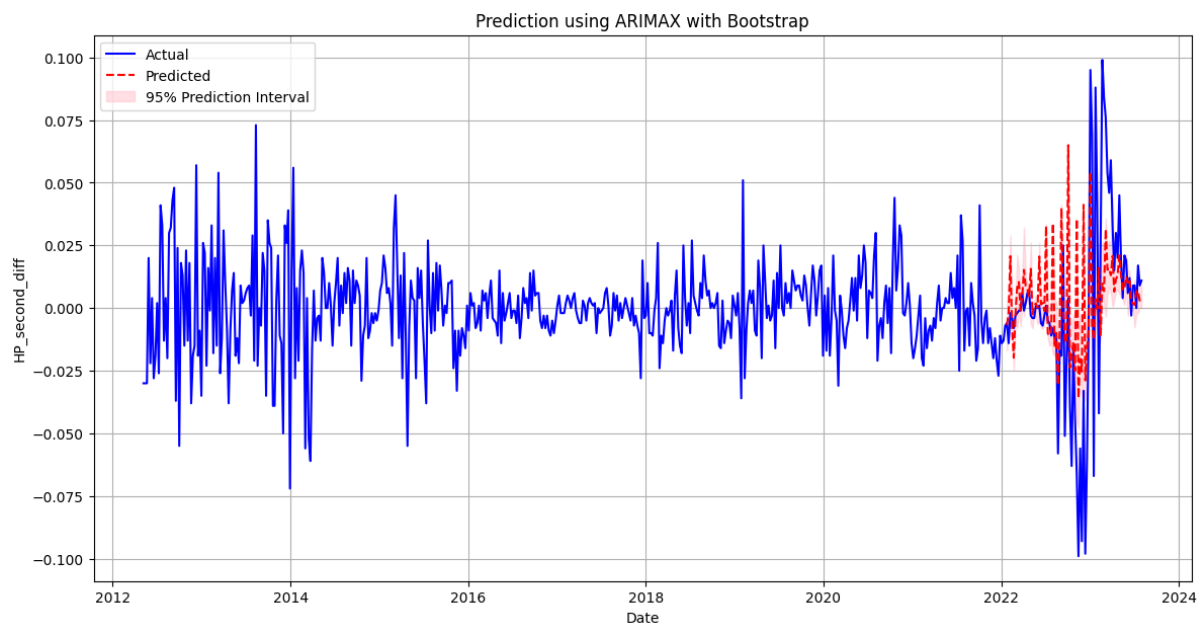
독립변수에 대해서도 시계열 정상성 확보를 위해, ADF test 결과 정상성을 가지지 않는 데이터에 대해 차분을 실시하였다. 차분한 데이터들을 그래프로 시각화한 모습은 다음과 같다.

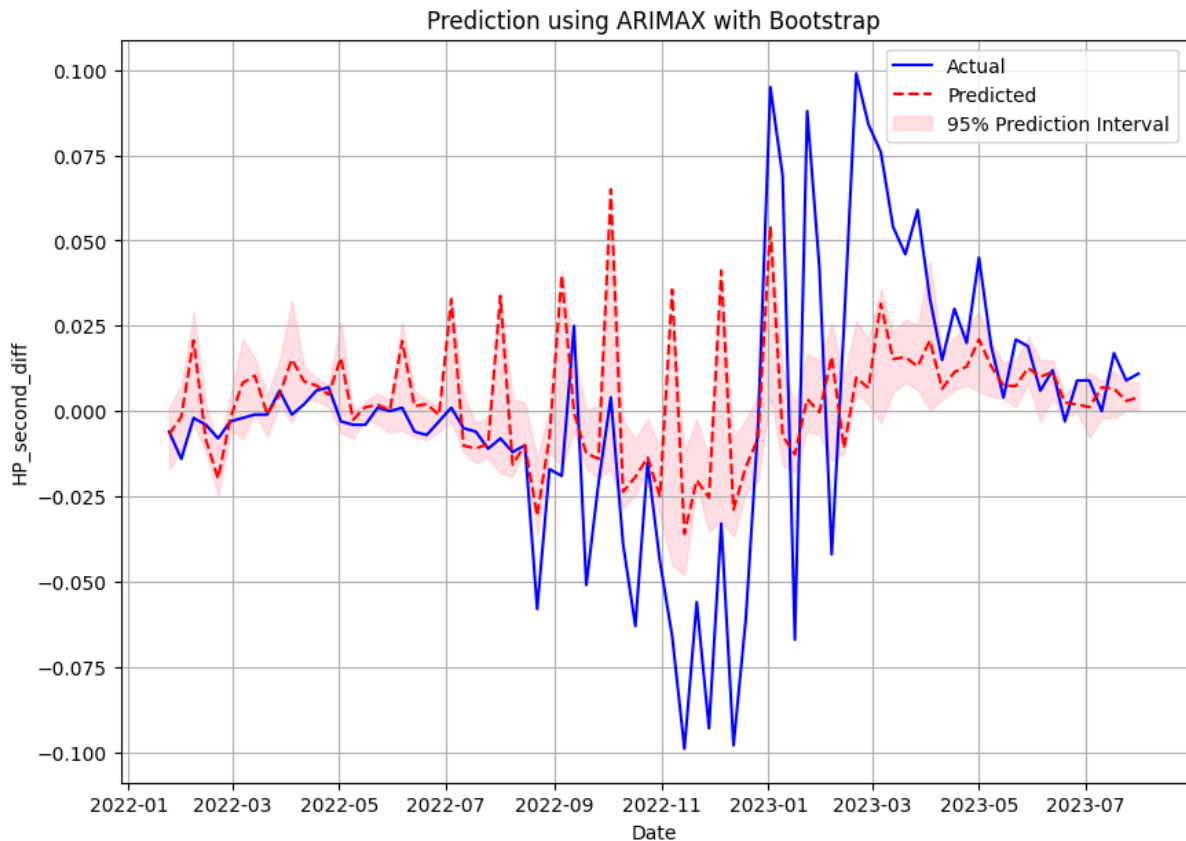


전세가격지수는 종속변수로, 실질임금, 주택담보대출금리, 부동산가격전망지수, 코로나 통제더미, M1, M2, Lf, 산업생산지수, 지가지수, 전세수급동향, 소비자물가지수, 원/달러 환율, 실질 GDP, 뉴스심리지수 및 주택준공량을 독립변수로 가지는 ARIMA (1, 2, 1) model 을 이용하여 전세가격지수를 예측한 결과는 다음과 같다.



그래프를 보면 예측 데이터가 어느 정도 기존 데이터의 추세를 따라가지만 극단적인 변화량까지 추적하지는 못함을 알 수 있다. 또한 예측 부분에 대한 신뢰구간이 넓어 예측의 정확도가 낮다. 따라서 예측 정확도 향상을 위해 Bootstrap 을 이용하여 1,000 개의 데이터셋을 추출하고 ARIMA 모델에 적합하여 예측을 시도하였다. Bootstrap 을 이용한 예측 데이터를 시각화한 그래프는 다음과 같다.





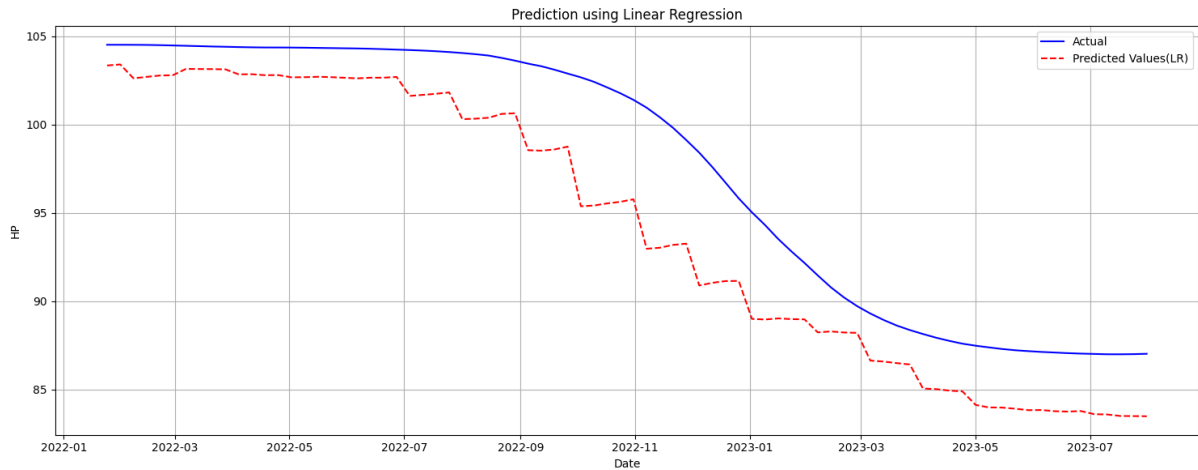
예측 데이터의 신뢰구간은 현저히 감소했으나, 데이터 자체의 예측 정확도는 변화하지 않았다. 즉 Bootstrap 을 포함한 ARIMA(1, 2, 1) 모델을 통한 예측은 보다 안정성을 가지지만, 부동산 시장의 불안정성과 수많은 사회적 변수의 영향력으로 인해 예측의 정확도는 높지 않다는 점을 시사한다. 예측의 정확도 향상을 위해 다양한 전통적 머신 러닝 예측 모델을 통한 전세가격지수 예측을 시도하였다.

#### - 선형 회귀(Linear Regression)

선형회귀모형(Linear Regression)은 가장 기본적인 회귀분석 방법 중 하나로 독립변수(예측 변수)와 종속변수(반응 변수) 사이에 선형 관계가 있다고 가정한다. 일반적인 다중 선형회귀모형은 아래와 같은 수식으로 표현된다.

$$y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \cdots + \beta_k x_{tk} + \varepsilon_t \quad (t = 1, 2, \dots, T)$$

여기서  $y_t$ 는  $t$  시점의 종속변수 값을,  $x_{ti}$ 는  $t$  시점에서  $i$ 번째 독립변수의 값을 나타낸다.  $\beta_0$ 는 절편(intercept)을,  $\beta_i$  ( $i = 1, \dots, k$ )는 각 독립변수의 계수(coefficient)를, 그리고  $\varepsilon_t$ 는 오차항을 나타낸다. 이러한 선형 회귀 모델을 일정한 양의 데이터로 학습시킨 뒤, 새로운 독립변수들을 입력하여 test를 진행하면 종속변수 값의 변동을 예측할 수 있다.



Linear Regression 을 통해 전세가를 예측한 결과는 다음과 같다. 전반적으로 기존 데이터의 추세를 잘 따라가고 예측 데이터의 정확도도 높음을 알 수 있다. 그러나 2023 년 5 월 데이터에 대해서 예측이 불안정할 수 있다.

#### - 서포트 벡터 머신(SVM)

서포트 벡터 머신(SVM)이란 주어진 데이터를 분류하는 경계면을 찾는 머신 러닝 기법이다. 두 집단으로 분류된 데이터에 대해 최대 마진을 가지는 분리 경계를 찾는다. 분리 경계에 가장 가까운 데이터들을 서포트 벡터라 하며, 서포트 벡터와 결정 경계 사이의 거리를 마진이라 부른다.<sup>28</sup> 선형으로 분리가 불가능할 경우 SVM 은 커널 함수를 이용해 저차원의 데이터를 고차원 공간으로 변환하여 선형 분리가 가능하도록 만든다. 이 과정에서 다양한 커널 함수를 이용할 수 있는데 선형함수, 다항함수, 가우시안 함수 등이 대표적이다.<sup>29</sup> SVM 은 다양한 데이터 세트에서 작동이 원활하다는 것이 최고 장점이다. 데이터 특성이 적더라도 복잡한 결정 경계를 만들어 과적합 문제가 발생하지 않을 수 있다. 뿐만 아니라 메모리 사용량이 크지 않고 예측 성공률이 높은 것 또한 장점 중 하나이다.<sup>30</sup>

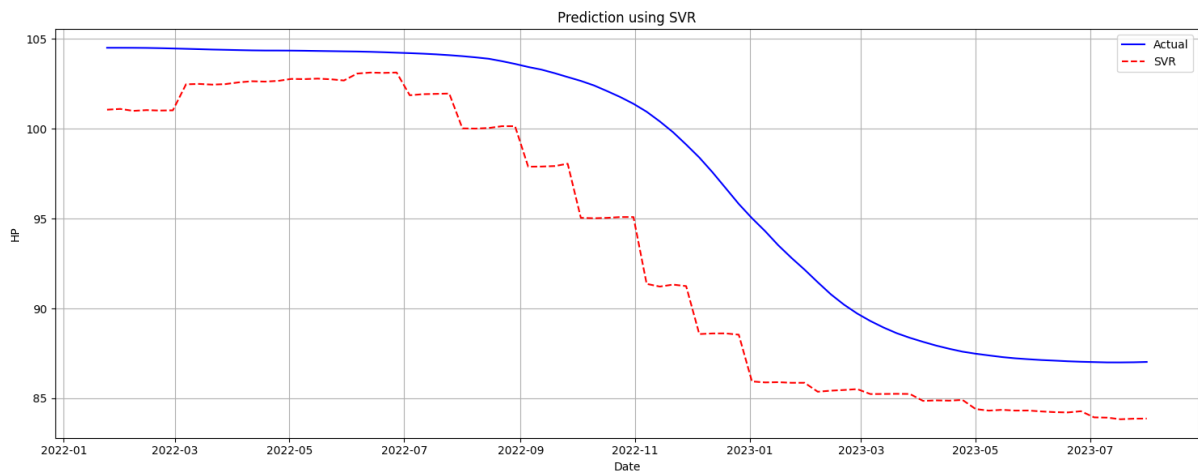
서포트 벡터 회귀(SVR)는 SVM 의 원리를 회귀분석으로 확장한 것으로 분류 대신 연속적인 값을 예측하기 위한 목적을 가진다. SVR 에서는 일정한 예측 오차 내에서 각 데이터와의 거리를

<sup>28</sup> 안현철, 김경재, (2009). 다양한 다분류 SVM을 적용한 기업채권평가. 한국경영정보학회

<sup>29</sup> 김선웅. (2020). SVM을 이용한 옵션투자전략의 수익성 분석. 중소기업융합학회

<sup>30</sup> 오윤태. (2019). 로지스틱 회귀와 서포트벡터머신의 머신러닝 성능 비교 실험. 국민대학교 일반 대학원

최소화하는 회귀식을 도출한다.<sup>31</sup> SVR 은 예측력이 뛰어나고, 벌칙항을 사용하여 과적합(overfitting) 문제를 발생시킬 가능성이 낮으며, 이상치에 둔감하다는 장점이 있다.<sup>32</sup>



SVR 을 통해 예측한 결과는 다음과 같다. Linear Regression 의 예측과 마찬가지로 전반적으로 데이터의 추세를 잘 따라간다. 그러나 2022 년 3 월까지의 데이터에 대해 예측의 방향성에서 조금씩 차이를 보인다.

#### - 결정 트리과 랜덤 포레스트

결정 트리(Decision Tree) 또는 의사 결정 트리는 일련의 분류 규칙을 거쳐 데이터를 분류하고 회귀하는 지도 학습 모델 중 하나이다. 각 질문과 예측값이 있는 부분을 노드(Node)라 하고, 예측값이 존재하는 바닥 부분을 리프 노드(Leaf Node)라 한다. 각 노드에서의 질문과 리프 노드의 값은 훈련 데이터에 따라 정해진 규칙을 적용해 구한다. 더 많은 질문을 통해 보다 복잡한 트리를 구성할 수 있는데 이를 더 깊은 트리라고 한다. 트리의 깊이가 얕으면 훈련 데이터로부터 충분한 특징을 얻어낼 수 없어 과소적합 문제가, 깊으면 각 리프 노드에 데이터가 거의 남지 않아 과적합 문제가 발생할 수 있다.<sup>33</sup>

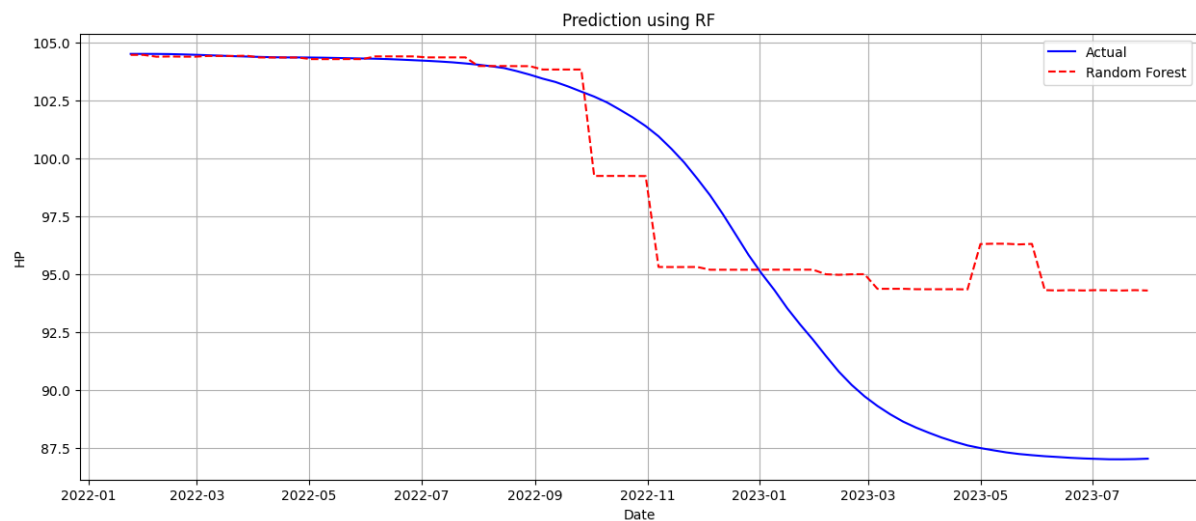
랜덤 포레스트(Random Forest)는 이런 결정 트리의 단점을 보완한 앙상블 학습의 일종으로 훈련 데이터에 랜덤성을 부여하여 여러 개의 결정 트리를 생성한 뒤 각각의 예측값의 평균을

<sup>31</sup> 이용상 외 5. (2013). 쓰기 자동채점 알고리즘의 성능 비료: 최대 엔트로피 기법과 서포트 벡터 회귀 기법. 한국교육과정평가원

<sup>32</sup> 김종수, 이성근. (2012). 헤도닉가격모형과 서포트 벡터 회귀분석모형을 이용한 공업용 부동산의 가격추정. 감정평가학 논집, 11(1), 71-89.

<sup>33</sup> <https://www.kaggle.com/code/dansbecker/how-models-work>

최종 예측값으로 이용하는 기법이다. 랜덤 포레스트는 일반적으로 단일 결정 트리보다 정확하게 예측하며 성능의 변동폭이 적다.<sup>34</sup>



RandomForest(트리의 수=100)를 통해 전세가 예측을 실시한 결과는 다음과 같다. 마찬가지로 추세를 잘 따라가며 특히 2022년 후반까지는 예측의 정확도가 매우 높다. 그러나 2023년 이후 데이터의 경우 전세가 감소폭에 대한 예측은 정확도가 떨어진다.

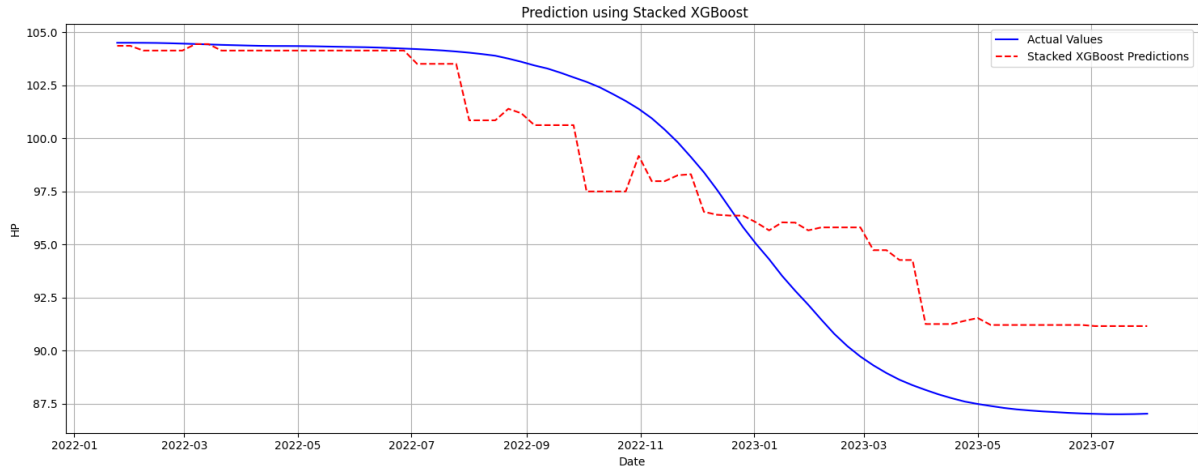
#### -XGBoost

3개의 머신러닝을 통해 예측을 실시한 결과, 모두 기존 데이터의 추세를 잘 따라가지만 각 모델마다 일부 구간에 대해 전세가 변동성을 예측함에 있어서 불안정성을 보일 수 있음을 관찰하였다. 따라서 데이터의 예측력 향상을 위해 SVM과 Linear Regression 모델에 대해 XGBoost를 통한 앙상블을 진행하였다. XGBoost는 예측력이 낮은 분류기들을 순차적으로 개선하여 상대적으로 예측력이 좋은 분류기를 만드는 트리모형에 그래디언트 부스팅 기법을 적용한 방법이다.<sup>35</sup> 그래디언트 부스팅 기법이란 이전 트리의 잔차를 기반으로 다음 트리를 훈련하면서 잔차를 줄여나가는 방법으로, 우수한 예측 능력으로 여러 분야의 예측에서 사용되는 기법이다.<sup>36</sup>

<sup>34</sup> <https://www.kaggle.com/code/dansbecker/random-forests>

<sup>35</sup> 하대우 외 2. (2019). XGBoost 모형을 활용한 코스피 200 주가지수 등락 예측에 관한 연구. 한국데이터정보과학회.

<sup>36</sup> 김상환. (2022). Xgboost 모형의 주가예측성과에 대한 실증연구. 충북대학교 사회과학연구소



XG Boost 를 이용하여 3 개의 예측 모델을 개선한 예측 결과는 다음과 같다. 마찬가지로 기존 데이터의 추세를 잘 따라가며, 각 예측 모델에서의 장점은 가져오고 단점은 보완한 추세를 보인다. RF 처럼 22 년도 후반까지의 예측 정확도가 매우 높고, 하락폭이 큰 부분은 다른 데이터의 예측 정확도가 반영되어 더욱 성능이 향상되었다고 볼 수 있다. 각 모델에 대한 예측력을 평가한 값은 다음과 같다.

예측 모델	ARIMAX	Linear Regression	SVR	Random Forest	XGBoost
RMSE	13.7734	3.7621	4.7291	4.2197	3.0448

평균제곱근오차 (RMSE : Root Mean Squared Error)는 예측 모델의 성능을 평가하는 대표적인 지표이다. 기존의 데이터와 예측 데이터 간 표준오차를 구하는 방식으로 계산된다.

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m (H(x_i) - y_i)^2}$$

RMSE 의 값이 낮을수록 예측 데이터와 기존 데이터 간 차이가 적으므로 예측력이 좋다고 볼 수 있다. 해당 표를 보면 ARIMAX 의 예측 성능이 가장 낮고,  $SVR < Random\ Forest < Linear\ Regression < XGBoost$  순으로 모델의 예측 성능이 좋았다. 따라서 본 연구에서는 SVR 과 Linear Regression 의 앙상블 모델인 XGBoost 로 신호접근법 기반 조기경보모형을 제작하였다.

### (3) 조기경보모형의 개발

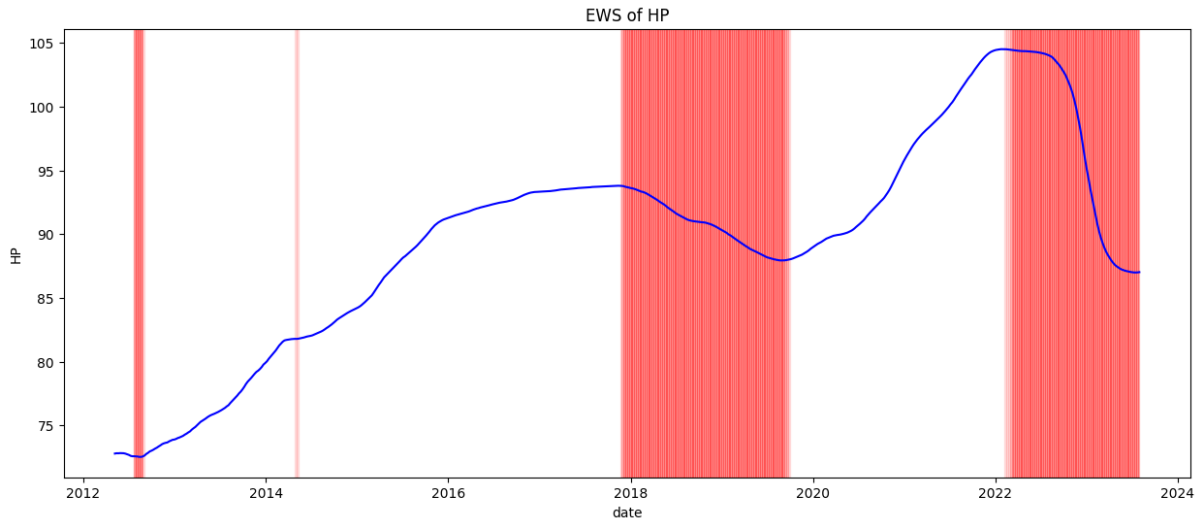
이동평균법 기반 신호접근법을 활용하여 조기경보모형을 제작하였다. 이동평균법은 동적 데이터의 특정 방향을 따라 이동하면서 평균을 구한다. 주로 주가 차트 등 시계열 데이터의 추세를 파악하는데 쓰인다. EWS 모델의 안정성을 위해 분기 단위로(약 12 주) 전세가의 하락이 영향을 미친다고 가정하였다.

$$mean_{12}(p_t) = \frac{1}{12} \sum_{k=t-11}^{k=t} p_k$$

$$std_{12}(P_t) = \sqrt{\frac{1}{12-1} \sum_{k=t-11}^{k=t} (P_k - mean_{12}(P_t))^2}$$

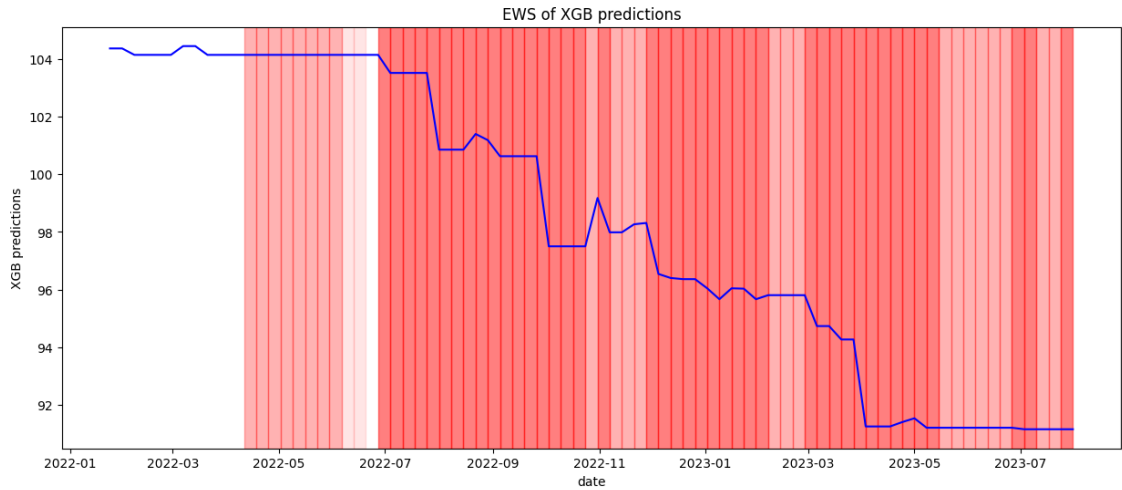
$$signal(P_t) = \begin{cases} 3, & \text{if } P_t < mean_{12}(P_t) - 0.8 \times std_{12}(P_t) \\ 2, & \text{if } mean_{12}(P_t) - 0.8 \times std_{12}(P_t) \leq P_t < mean_{12}(P_t) \\ 1, & \text{if } mean_{12}(P_t) \leq P_t < mean_{12}(P_t) + 0.8 \times std_{12}(P_t) \\ 0, & \text{if } P_t \geq mean_{12}(P_t) + 0.8 \times std_{12}(P_t) \end{cases}$$

새로운 EWS의 위험신호는 다음과 같이 정의된다. 이동평균법을 기반으로 구한 12주 전 ~ 현재 전세가 예측 데이터의 평균 및 분산을 이용하여, 하락세의 예상 정도에 따라 0 ~ 3 단계의 위험신호를 보낸다. 위험 단계가 높아질수록 진한 바탕으로 배경을 채워서 시각화 하였다. 기존 전세가격지수 데이터로 계산한 신호는 다음과 같다.

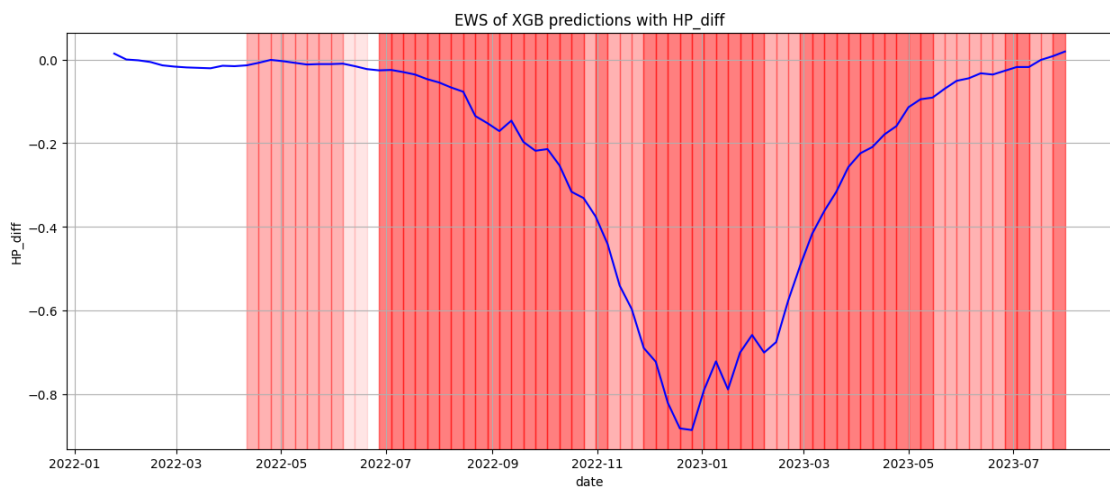


전세가격이 명확하게 하락할수록 높은 수준(3 단계)의 신호를 보내고, 예상 하락률이 낮은 경우에는 낮은 수준(1 단계)의 신호를 보냄을 확인하였다. 이제 XGBoost를 이용해 예측한 값에 해당 EWS 산식을 대입하여 위험신호를 계산하였다. 해당 데이터를 예측값과 함께 시각화한 결과는 다음과 같다. 전반적으로 예측값의 하락을 잘 반영했으나, 일부 구간에서 신호가 불안정할 수 있다.

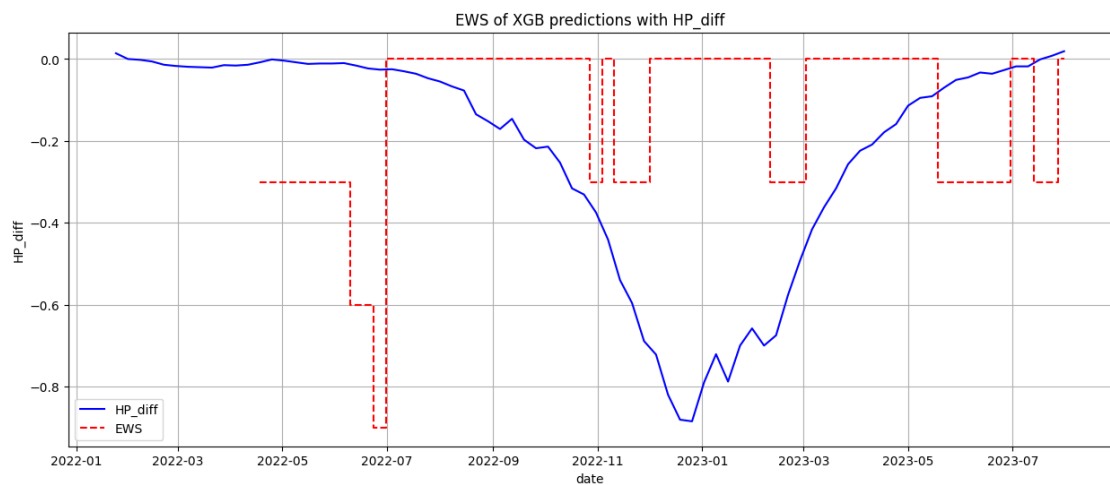




전세가 변동률의 실제 데이터와 함께 시각화한 결과는 다음과 같다. 변동률의 값이 0에서 멀어질수록 높은 수준의 조기경보가 울린다. 머신러닝을 이용한 모델의 불안정성으로 일부 구간에서 오차를 보일 수 있으나, 전반적으로 변동률의 추세가 잘 반영되었다고 볼 수 있다.



해당 조기경보모형의 단계를 전세가격지수 변동률에 맞춰 스케일 조정한 뒤 선으로 시각화한 결과는 다음과 같다.



### Ⅲ. 결론

#### 1. 부동산 시장 조기경보모형의 경제적 의의

부동산 가격의 변동은 경제 전반에 파급되며 많은 이들의 삶에 직접적인 영향을 미친다. 특히 전월세 가격의 안정은 서민의 주거생활 안정과 밀접한 연관성이 있다. 우리나라의 전세라는 독특한 주거 점유형태는 추가적인 주택시장 불안정을 야기할 수 있다.<sup>37</sup> 최근 전세가격 급락과 함께 사회 문제로 대두되었던 역전세, 갭통전세와 이에 따른 전세 사기 또한 주거생활 안정성을 위협하는 요소로 작용하였다. 주택이 단순히 거주 장소로만 여겨지지 않고, 하나의 투자 자산으로 활용되는 사회에서 부동산 시장 둔화는 소비 침체로 이어져 경제 전반으로 그 영향력이 파급될 수 있다. 뿐만 아니라 부동산 가격의 불안정성은 물가와 경제 전체의 불안정을 초래할 수 있다.

따라서 부동산 가격변동의 파급력과 불안정성에 대응해 부동산 정책을 선제적으로 수행하기 위해서는 부동산 가격변동을 예측하는 조기경보모형이 필요하다. 또한 예측을 수행하는 조기경보 모형은 부동산 시장의 복합적인 요소를 반영하여 다양한 변수를 활용 가능해야 한다. 머신러닝을 활용한 부동산 조기경보모형을 통해 전세가격을 예측할 경우 비교적 자유롭게 부동산 시장에 영향을 미치는 변수를 추가함으로써, 보다 현실에 부합하는 전세가격 변동 예측이 가능하다. 이러한 예측을 기반으로 전세 가격의 급격한 변동을 예측하고, 경제적 조정 과정을 거침으로써 급격한 가격 변동으로 인해 경제에 파급될 위험을 축소할 수 있다.

#### 2. 연구의 의의 및 한계

해당 연구를 통해 거시경제적 선행변수 뿐만 아니라 가계의 주택 수요량에 영향을 미치는 다양한 경제적 변수까지 반영한 머신러닝 기법의 예측 유용성을 확인할 수 있었다. 따라서 이렇게 제작한 조기경보모형으로 분석한 전세가의 급락이 예측될 경우 실질임금 및 주택담보대출금리와 같이, 회귀분석을 통해 명확한 상관관계를 가진다고 판단한 경제적 변수를 조절함으로써 많은 이들이 역전세의 피해를 입기 전에 그 위험을 완화할 수 있다. 또한 머신러닝을 이용한 예측 방식을 기반으로 조기경보모형의 단계를 가변적으로 확장하여 주택가격의 과도한 상승에 대한 조기경보모형 제작 등 모형의 응용 가능성을 시사한다.

반면 전세가격지수 및 각종 선행변수에 대한 일별 데이터 등 충분히 많은 양의 데이터를 수집하지 못해 예측력이 감소했을 수 있다는 점에서 한계가 있다. 본 연구에서는 총 586개의 데이터를 사용하여 분석했기 때문에, 기본적인 머신러닝 모델만을 적용할 수 있었다. 만약 충분히 많은 데이터를 수집할 수 있다면 RNN, LSTM 등 우수한 예측력을 가진 딥러닝 모델을 적용하여 EWS의 성능을 발전시킬 수 있으리라 기대한다.

---

<sup>37</sup> 서승환. (2004). 부동산시장 조기경보체계 구축연구. 건설교통부

## [참고문헌]

이승연. 2022. '전세대출 금리 8%대 목전...'차라리 월세가 싸". *파이낸셜뉴스*. 6 December.

<https://www.fnnews.com/news/202212061159374776>

민병철. (2023). 깡통전세의 발생 확률 추정. 주택도시보증공사. 8(1), pp.7-20

한국은행 2023 년 5 월 경제전망보고서

권영규. (2023). 부동산가격에 대한 결정요인 분석. 한국해양대학교 대학원

한국부동산원 부동산통계정보시스템(R-ONE) <https://www.reb.or.kr/r-one/main.do>

김동숙. (2016). 실시간 의약품 사용자료를 활용한 일부 감염병 유행 조기경보 시스템.

건강보험심사평가원. 2016 10(5):13

이항용. (2015). 경제발전경험모듈화사업: 금융위기 조기경보시스템. 기획재정부

박원암. (2011). 2008 년 위기 예측 가능했나?:신호접근법 분석. *對外經濟研究* 제 15 권 제 3 호

강희만, 김정렬. (2012). 투기적 특성이 아파트가격에 미치는 영향 분석. *한국산업경제학회*.

25(4):2841-2858

류지수, 임규채, 기석도. (2007). 주택가격 및 토지 가격의 거품존재 가능성 검정.

*한국산업경제학회*. 20(6):2245-2264

손정식, 김관영, 김용순. (2003). 부동산가격 예측 모형에 관한 연구. *한국주택학회*. 49-76

박대현, 김정환, 류두진. (2021). 기계학습 기반 주택시장의 조기경보체계. *한국부동산원*. 7(1):29-45

허강성. (2022). 부채계약가설의 적용은 경영성과에 따라 차이가 있는가?: OLS와 GLS 회귀모형 실증결과 비교분석. *국제회계연구*. 106, 221-245.

질병관리청. 『검역단계에서 해외유입 신종코로나바이러스 확진환자 확인 감염병 위기경보를 '주의'단계로 상향, 대응』

[https://ncov.kdca.go.kr/tcmBoardView.do?brdId=&brdGubun=&dataGubun=&ncvContSeq=352435&contSeq=352435&board\\_id=&gubun=ALL](https://ncov.kdca.go.kr/tcmBoardView.do?brdId=&brdGubun=&dataGubun=&ncvContSeq=352435&contSeq=352435&board_id=&gubun=ALL)

강중모. 2023. '31 일부터 코로나 19 감염병 4 급 하향..'일반의료체계' 전환'. *파이낸셜뉴스*. 23

August. <https://www.fnnews.com/news/202308230923274283>

한국은행 경제통계시스템 <https://ecos.bok.or.kr/#/SearchStat>

한국거래소 정보데이터시스템(KRX)

<http://data.krx.co.kr/contents/MDC/MDI/mdiLoader/index.cmd?menuId=MDC0201010105>

고용노동부 노동시장조사과

[https://www.moel.go.kr/news/enews/report/enewsView.do?news\\_seq=15745](https://www.moel.go.kr/news/enews/report/enewsView.do?news_seq=15745)

한준권. (2020). 주택가격과 사회 경제지표의 선후행성 관계 분석. 한밭대학교 산업대학원

모진성. (2018). 주요 경제변수와 정책변수가 주택가격에 미치는 영향에 관한 연구.

국립목포대학교 행정대학원

성용림, 유정석. (2013). 주택가격에 대한 환율의 동태적 파급효과 분석. 부동산학보, (54), 244-257.

성용림, 유정석. (2013). 주택가격에 대한 거시경제적 요인의 영향 분석: 환율과 금리를 중심으로. 단국대학교

전건. (2010). 주택가격과 거시경제지표의 관계 분석. 한밭대학교 산업대학원

서범석, 이영환, 조형배. (2022). 기계학습을 이용한 뉴스심리지수(NSI)의 작성과 활용. 한국은행 경제통계국 통계연구반. 68-90

통계청 국가통계포털(KOSIS)

[https://kosis.kr/statHtml/statHtml.do?orgId=116&tblId=DT\\_MLTM\\_5372&vw\\_cd=MT\\_ZTITLE&list\\_id=116\\_11626\\_003&seqNo=&lang\\_mode=ko&language=kor&obj\\_var\\_id=&itm\\_id=&conn\\_path=MT\\_ZTITLE](https://kosis.kr/statHtml/statHtml.do?orgId=116&tblId=DT_MLTM_5372&vw_cd=MT_ZTITLE&list_id=116_11626_003&seqNo=&lang_mode=ko&language=kor&obj_var_id=&itm_id=&conn_path=MT_ZTITLE)

안현철, 김경재, (2009). 다양한 다분류 SVM 을 적용한 기업채권평가. 한국경영정보학회. Vol. 19, Issue 2, p. 157-178

김선웅. (2020). SVM 을 이용한 옵션투자전략의 수익성 분석. 중소기업융합학회. Vol. 10, Issue 4, p. 46-54

오윤태. (2019). 로지스틱 회귀와 서포트벡터머신의 머신러닝 성능 비교 실험. 국민대학교 일반대학원

이용상 외 5. (2013). 쓰기 자동채점 알고리즘의 성능 비료: 최대 엔트로피 기법과 서포트 벡터 회귀 기법. 한국교육과정평가원. Vol.16, no.3, 통권 32 호 pp. 147-165

김종수, 이성근. (2012). 헤도닉가격모형과 서포트 벡터 회귀분석모형을 이용한 공업용 부동산의 가격추정. 감정평가학 논집, 11(1), 71-89.

<https://www.kaggle.com/code/dansbecker/how-models-work>

<https://www.kaggle.com/code/dansbecker/random-forests>

하대우 외 2. (2019). XGBoost 모형을 활용한 코스피 200 주가지수 등락 예측에 관한 연구.

한국데이터정보과학회. 30(3):655-669

김상환. (2022). Xgboost 모형의 주가예측성과에 대한 실증연구. 충북대학교 사회과학연구소. Vol. 39, Issue 1, p. 29-55

서승환. (2004). 부동산시장 조기경보체계 구축연구. 건설교통부