

Nowcasting Monthly Industrial Activity Deflators Using High-Dimensional Predictors*

Gyure Kim[†] Jihyun Kim[‡] Yerin Han[§] Heejoon Han[¶]

Abstract This study conducts nowcasting of monthly industrial activity deflators for the service and retail sales sectors using high-dimensional explanatory variables, and compares the performance of various linear and nonlinear forecasting methods. Linear approaches combine AR/ARX models with factor analysis and shrinkage estimation techniques, while nonlinear approaches include tree-based methods and neural network models. The results show that the AR + Ridge approach performs best for the service deflator, whereas the ARX + factor analysis with LASSO/Elastic Net yields the strongest performance for the retail sales deflator. In addition, incorporating a one-year lag improves forecasting accuracy for both deflators when newly released information is available. Overall, the findings highlight the importance of methodological structure and the timing of explanatory variables in nowcasting industrial activity deflators.

Keywords Industrial activity, deflator, nowcasting, big data, machine learning.

JEL Classification C32, C50, E31.

*This paper was supported by a 2025 research project commissioned by the Korea Development Institute (KDI) for the development of a data-driven economic diagnosis system. The authors are also grateful to the editor and the anonymous referees for their helpful comments and constructive suggestions.

[†]Department of Economics, Sungkyunkwan University, 25-2 Sungkyunkwan-ro, Jongno-gu, Seoul, Republic of Korea 03063. E-mail: kimkyokkr77@g.skku.edu.

[‡]Corresponding author. Department of Economics, Sungkyunkwan University, 25-2 Sungkyunkwan-ro, Jongno-gu, Seoul, Republic of Korea 03063. E-mail: kim.jihyun@skku.edu.

[§]Department of Global Economics, Sungkyunkwan University, 25-2 Sungkyunkwan-ro, Jongno-gu, Seoul, Republic of Korea 03063. E-mail: yerin.h0608@g.skku.edu.

[¶]Department of Economics, Sungkyunkwan University, 25-2 Sungkyunkwan-ro, Jongno-gu, Seoul, Republic of Korea 03063. E-mail: heejoonhan@skku.edu.

Received January 9, 2026, Revised February 10, 2026, Accepted February 20, 2026

고차원 설명변수를 활용한 월별 산업활동동향 디플레이터 나우캐스팅*

김겨레[†] 김지현[‡] 한예린[§] 한희준[¶]

Abstract 본 연구는 고차원 설명변수를 활용하여 서비스 및 소매판매 산업 활동동향 지표에 대한 디플레이터를 대상으로, 나우캐스팅(Nowcasting)을 수행하고, 다양한 선형 및 비선형 예측 방법의 성과를 비교·분석한다. 선형 방법에서는 AR/ARX 모형을 기반으로 요인분석과 축소추정 기법을 결합한 방법, 비선형 방법에서는 Tree 기반 방법, 인공신경망 기반 방법 등을 활용한다. 예측 결과, 서비스 디플레이터에서는 AR + Ridge 방식이, 소매판매 디플레이터에서는 ARX + 요인분석 + LASSO/ElasticNet 방식이 각각 가장 우수한 예측력을 보인다. 또한, 1년 시차항은 최신 정보가 추가로 반영되는 예측 시점에서 두 디플레이터 모두의 예측력을 개선하는 것으로 나타난다. 한편 비선형 머신러닝 방법은 설명변수의 시점 정보를 충분히 반영하지 못한 제약으로 인해 상대적으로 낮은 예측 성과를 보인다. 이러한 결과는 산업활동동향 디플레이터의 나우캐스팅에서 예측 시점별 그리고 예측 대상별 방법론 구조와 변수의 시점 차이가 예측 성능에 핵심적인 역할을 함을 시사한다.

Keywords 산업활동동향, 디플레이터, 나우캐스팅, 빅데이터, 머신러닝.

JEL Classification C32, C50, E31.

*이 논문은 KDI의 데이터 기반 경기진단 시스템 구축을 위한 용역연구(2025년)로 진행되었음을 밝힌다. 또한, 본 논문에 유익한 논평과 건설적인 제언을 해주신 익명의 심사위원들께 감사드립니다.

[†]성균관대학교 경제학과, 서울특별시 종로구 성균관로 25-2 03063. E-mail: kimkyokkr77@g.skku.edu.

[‡]교신저자. 성균관대학교 경제학과 / 퀀트응용경제학과, 서울특별시 종로구 성균관로 25-2 03063. E-mail: kim.jihyun@skku.edu.

[§]성균관대학교 글로벌경제학과, 서울특별시 종로구 성균관로 25-2 03063. E-mail: yerin.h0608@g.skku.edu.

[¶]성균관대학교 경제학과 / 퀀트응용경제학과, 서울특별시 종로구 성균관로 25-2 03063. E-mail: heejoonhan@skku.edu.

1. 서론

산업활동동향 디플레이터는 산업활동동향 지표의 명목지수를 실질지수로 나누어 산출되는 물가 지표로서, 산업별 생산과 소비 흐름에 수반되는 가격 변동을 반영하며 산업 단위의 경기 및 물가 판단에 중요한 정보를 제공한다. 그러나 산업활동동향 지표는 통상 익월 말에 발표되기 때문에, 정책 결정이나 시장 대응이 요구되는 시점에는 당월 디플레이터를 즉시 관측할 수 없다는 한계를 가진다. 이러한 발표 시차는 산업별 경기 및 물가 판단에 불가피한 지연을 초래하며, 이를 보완하기 위해 산업활동동향 디플레이터에 대한 나우캐스팅(Nowcasting)이 요구된다. 이에 따라 본 연구는 산업활동동향 지표 중 서비스업생산지수와 소매판매액지수의 디플레이터를 대상으로, 나우캐스팅 환경에서 선형 및 비선형 예측 방법의 성과를 다양한 설명변수를 활용하여 비교함으로써, 정보 반영 방식과 예측 방법 설계가 예측 성능에 미치는 역할을 규명하는 것을 목적으로 한다.

산업활동동향 디플레이터는 기존 물가 예측 문헌의 주요 분석 대상인 소비자물가지수(CPI) 및 GDP 디플레이터와 구별되는 특성을 지니며, 이를 통해 두 물가 지표를 보완한다. CPI는 소비자 구입 품목의 가격 변동에 집중하여 생산 단계의 물가를 파악하는 데 한계가 있으며, GDP 디플레이터는 경제 전반의 종합적인 물가 수준을 나타내나 세부적인 산업별 경기를 파악하는 데는 한계가 있다. 반면, 산업활동동향 디플레이터는 산업별 생산 및 소비 단계의 가격 변동을 반영하여 산업 단위 물가 압력을 구체적으로 포착한다. 따라서 이를 바탕으로 산업별 수급 조절 대책이나 비용 부담 완화 조치의 시행 시기를 결정하는 등 정교한 정책 의사결정을 지원할 수 있다. 실제로, 김동석 (2003)은 산업활동동향 디플레이터를 산업별 물가 분석에 활용함으로써 거시경제 및 산업정책에 대해 논의한다.

이와 같이 산업활동동향 디플레이터는 CPI 및 GDP 디플레이터와 상이한 특성을 가지기 때문에 기존 물가 예측 문헌에서 설계된 예측 구조가 산업활동동향 디플레이터에서도 동일하게 작동할 것이라는 보장은 없다. 따라서 디플레이터 나우캐스팅에서는 해당 지표의 특성과 정보 구조를 고려한 별도의 검토가 필요하다.

한편 최근 거시경제 예측 문헌에서는 컴퓨팅 파워의 증가와 데이터 처리 기술의 발전에 힘입어 고차원 자료와 머신러닝 기법을 활용한 예측 방법이 활발히 논의되고 있다. Goulet Coulombe *et al.* (2022)과 Masini, Medeiros and Mendes (2023)은 머신러닝 기법이 거시경제 예측에서 유용함을 보이며, Medeiros *et al.* (2021)은 비선형 머신러닝 방법이 미국 CPI 예측 성능을 개선

함을 제시한다. 한국의 경우에도 한희준 (2023, 2024)은 머신러닝 기반 접근이 한국 CPI 예측력을 향상시킬 수 있음을 보인다. 또한 Bernanke and Boivin (2003), Stock and Watson (1998, 2002a,b, 2006)은 요인분석(Factor analysis)을 통한 차원 축소가 고차원 자료 환경에서 효과적인 예측 도구임을 제시하며, 축소추정 기법(Ridge, LASSO, ElasticNet) 역시 단기 물가 예측에서 유용함이 보고된다. 또한, 이승윤·이은경·한진현 (2017), 이현창 외 (2022) 등은 머신러닝을 활용하여 GDP 나우캐스팅의 성과를 제고할 수 있음을 보인다. 다만 이러한 연구들은 예측 시점에 따라 활용 가능한 정보가 덜 제약된 환경을 가정하고 있고, 개별 산업 단위의 예측을 하지 않는다. 따라서 예측 시점별 정보 구조가 중요한 산업활동동향 디플레이터 나우캐스팅 환경에 기존 연구를 그대로 적용하기에는 한계가 있다.

이에 따라 본 연구는 이러한 기존 문헌을 확장하여, 산업활동동향 디플레이터 나우캐스팅에서는 단일한 예측 방법을 일률적으로 적용하기보다 디플레이터의 성격과 예측 시점에 따라 적합한 방법 구조를 선택하는 것이 중요함을 실증적으로 제시한다. 이를 위해 본 연구는 예측 시점에 따라 가용한 정보의 범위가 서로 다른 두 가지 나우캐스팅 환경(1차 및 2차)을 설정하고, 각 시점까지 실제로 활용 가능한 정보에 맞추어 설명변수를 구성한다. 추가로, AR/ARX 모형을 기본 구조로 하여 요인분석, 축소추정 기법, 1년 시차항 포함 여부, 자료 변환을 체계적으로 고려하여 선형 예측 방법을 설계하며, Random Forest, XGBoost, LSTM 등 비선형 머신러닝 방법도 활용하여 예측을 진행한다. 이러한 예측실험은 단순한 방법 간 예측력 비교를 넘어, 나우캐스팅 환경에서 설명변수의 시점 처리와 방법론 구조 선택이 예측 성능에 미치는 영향을 파악할 수 있게 한다.

분석 결과, 서비스 디플레이터에서는 두 나우캐스팅 시점 모두에서 밀집 구조를 갖는 AR + Ridge 방식이 가장 우수한 예측력을 보이며, 소매판매 디플레이터에서는 ARX 구조에 요인분석과 변수 선택을 결합한 LASSO 또는 ElasticNet 방식이 상대적으로 뛰어난 성과를 나타낸다. 또한 두 디플레이터 모두에서 활용 가능한 정보의 범위가 확대되어 최신 정보를 반영한 시점에는 1년 시차항을 포함하는 것이 예측력을 개선하는 것으로 나타난다. 반면 비선형 머신러닝 방법은 모든 설정에서 선형 방법 대비 낮은 예측 성과를 보이는데, 이는 방법 자체의 한계라기보다 예측 시점별로 상이한 정보 구조를 충분히 반영하지 못한 설계상의 제약에 기인한 결과로 해석된다.

이후 본 논문의 구성은 다음과 같다. 2장에서는 분석에 사용되는 자료와 예측 방법을 소개하고, 3장에서는 예측 절차를 설명한다. 4장에서는 예측 실험 결과를 제시하며, 마지막으로 5장에서는 이러한 결과를 바탕으로 산업활동동

향 디플레이터 나우캐스팅에 대한 시사점과 향후 연구 방향을 논의한다.

2. 자료 및 예측방법

2장에서는 자료의 구성과 전처리 방식, 산업활동동향 디플레이터 예측에 사용되는 벤치마크 모형, 그리고 본 연구에서 도입한 고차원 설명변수를 활용한 선형 및 비선형 예측 방법을 정리한다. 이를 통해 3장의 예측실험 설계와 4장의 결과 해석에 필요한 사항을 제시한다.

2.1. 자료 및 예측대상

본 연구에서 활용한 자료는 2004년 8월부터 2025년 6월까지의 월별 관측치로 구성되며, 총 251개 시점을 포함한다. 종속변수는 서비스생산지수 디플레이터(서비스 디플레이터)와 소매판매액지수 디플레이터(소매판매 디플레이터), 두 가지로, 각각 명목 지수를 실질 지수로 나누어 산출한다.

표 1은 각 디플레이터의 기초 통계 및 단위근 검정(ADF: Augmented Dickey-Fuller) 검정통계량을 제시하며, 그림 1은 2004년 8월부터 2025년 6월까지의 서비스 및 소매판매 디플레이터의 추이를 제시한다. ADF 검정의 귀무가설은 해당 변수가 단위근을 가진다는 것이다. 표 1에 따르면, 서비스 및 소매판매 디플레이터는 모두 귀무가설을 기각하지 못하는 것으로 나타난다. 이는 두 시계열이 단위근을 갖는 것으로 해석된다. 시계열에 단위근이 존재할 경우, 데이터 변환을 통해 정상성을 확보한 뒤 이를 예측에 사용하는 것이 적절하다. 이에 따라 본 연구에서는 종속변수를 로그차분($\log y_t - \log y_{t-1}$)하여 정상시계열로 변환한 후 예측을 수행한다.

본 연구에서는 95개 변수를 설명변수로 활용한다. 2.2절에서 설명할 벤치마크 모형(ARX 모형)은 서비스 CPI와 상품 CPI를 설명변수에 포함하고, 종속변수의 시차항을 이용하여 예측하는 구조를 가진다. 본 연구에서도 이러한 설정을 반영하여 서비스 및 상품 CPI와 종속변수의 시차항을 설명변수로 포함하여 예측을 수행하였다. 각 예측 방법에 활용되는 구체적인 변수 구성은 2.2절에서 제시한다.

모든 설명변수는 시계열 정상성을 확보한 후 사용하며, 변수의 변환 방식은 McCracken and Ng (2016)와 한희준 (2023)을 참고하여 결정한다. 각 변수의 변환 방식(Tcode)은 부록의 표 8에 제시하며, Tcode의 의미는 부록에서 설명한다. 또한 서비스 및 상품 CPI를 제외한 93개 설명변수는 익익월 중순에 최신화가 완료되기 때문에, 설명변수를 각 방법에 적용할 때 예측 시점 기준으로 실제로 이용 가능한 정보만을 포함하도록 구성한다.

통계량	서비스 디플레이터	소매판매 디플레이터
평균	0.940	0.973
중간값	0.933	0.983
최소값	0.776	0.806
최대값	1.132	1.165
표준편차	0.099	0.093
단위근 검정통계량 (ADF 검정)		
절편 포함	0.964	-0.144
절편 및 선형 추세 포함	-1.214	-1.831

표 1: 서비스생산지수 및 소매판매액지수 디플레이터 기초 통계. 이 표는 본 연구의 예측 대상 변수들에 대한 기초통계량 및 단위근 검정 결과를 제시한다. 단위근 검정 결과, 모든 변수는 단위근을 포함하고 있는 것으로 나타난다.

Table 1: DESCRIPTIVE STATISTICS OF DEFLATORS FOR THE SERVICE PRODUCTION INDEX AND RETAIL SALES INDEX. This table presents descriptive statistics and unit root test results for the target variables of this study. The unit root test results indicate that each variable contains a unit root.



그림 1: 서비스 및 소매판매 디플레이터. 이 그림은 2004년 8월부터 2025년 6월까지의 예측 대상 변수별 추이를 보여준다. 점선은 서비스 디플레이터, 실선은 소매판매 디플레이터를 나타낸다.

Figure 1: SERVICE AND RETAIL DEFLATORS. This figure shows the trends of the target variables from August 2004 to June 2025. The dashed line represents the service deflator, and the solid line represents the retail sales deflator.

단, 비선형 머신러닝 방법은 계산 비용과 적합 과정의 제약으로 인해 예측 시점별로 최신 서비스·상품 CPI 정보를 반영하거나 종속변수의 시차항을 개별적으로 포함하는 구조를 구현하는 데 제한이 있다. 따라서 비선형방법은 예측 시점 기준으로 변수의 시차를 세분화하여 처리하지 못하고, 모든 변수에 동일한 시차를 적용한다. 이로 인해 일부 비선형방법의 성과가 상대적으로 낮게 나타날 수 있으나, 이는 방법 자체의 한계가 아닌 예측 시점에서 활용 가능한 정보를 정교하게 반영하지 못한 데서 발생한 제약으로 해석해야 한다.

2.2. 예측방법

본 절에서는 산업활동동향 디플레이터 예측에 사용되는 예측방법을 설명한다. 디플레이터 나우캐스팅은 두 가지 시점에서 수행된다. 예컨대 5월 디플레이터는 일반적으로 6월 말에 공표되므로, 이를 앞당겨 추정하기 위해 (1) 5월 중순과 (2) 6월 초에 각각 예측을 실시한다. 본 절에서는 이를 각각 1차 나우캐스트와 2차 나우캐스트로 칭한다. 또한 각 방법의 구조는 예측 시점별로 활용 가능한 정보의 차이를 반영하여 제시한다.

2.2.1. 벤치마크 모형

Stock and Watson (1998, 2002a,b)과 같은 기존 물가 예측 연구에서 AR/ARX 모형은 벤치마크 모형으로 활용되어 왔고, AR 기반의 모형은 단기 물가 예측에서 우수한 성능을 보여주는 것으로 알려져 있다. 이와 같은 선행연구를 반영하여 본 연구는 ARX 모형을 벤치마크 모형으로 설정한다.

ARX 모형에서 활용되는 설명변수는 서비스 CPI 및 상품 CPI이다. 서비스 디플레이터 예측에는 서비스 CPI, 소매판매 디플레이터 예측에는 상품 CPI를 사용한다. CPI는 디플레이터와 마찬가지로 물가 흐름에 대한 정보를 제공할 뿐 아니라 산업활동동향 지표보다 발표 시점이 빠르기 때문에 디플레이터 예측력을 향상시키는 데 유용하다.

CPI(x)는 동월 말에 발표되므로, 1차 나우캐스트 시점 t 에서는 x_{t-1} 까지의 정보만 이용할 수 있다. 또한, 산업활동동향 지표가 발표되는 시기를 고려했을 때, 1차 나우캐스트 시점에는 2개월 전 디플레이터(y_{t-2})까지 이용할 수 있다. 이에 따라 1차 나우캐스트에 적용되는 ARX 모형은 (1)과 같다.

$$y_t = \alpha_0 + \alpha_1 y_{t-2} + \alpha_2 x_{t-1} + \varepsilon_t, \quad (1)$$

위 식에서 y_t 는 로그 차분한 디플레이터이고, x_t 은 로그 차분한 전월의 서비스

혹은 상품 CPI를 의미한다. 다음으로, 2차 나우캐스트는 동월 CPI x_t 와 직전월 디플레이터 y_{t-1} 가 발표된 이후 이루어지므로, (2)와 같은 모형을 사용한다.

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 x_t + \varepsilon_t. \quad (2)$$

2.2.2. 선형방법

본 연구의 선형방법은 AR/ARX 구조에 고차원 설명변수, 요인(Factor), 1년 시차항 등 다양한 확장 요소를 결합한 형태이다. 1차와 2차 나우캐스트에 해당하는 선형방법의 기본 구조는 각각 (3) 및 (4)와 같이 표현할 수 있다.

$$y_t = \alpha_0 + \alpha_1 y_{t-2} + \alpha_2 x_{t-1} + Z_{t-2}^\top \beta + \varepsilon_t, \quad (3)$$

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 x_t + Z_{t-2}^\top \beta + \varepsilon_t. \quad (4)$$

여기서 x_{t-1} , x_t 는 서비스 혹은 상품 CPI, Z_{t-2} 는 고차원 설명변수 벡터를 의미한다. Z_{t-2} 는 해당 변수들이 익익월 중순에 최신화된다는 점을 반영하여, 예측 시점에서 실제로 이용 가능한 최신 정보가 $t-2$ 시점이라는 사실을 고려해 구성한 것이다. 추가로, 본 연구는 물가의 연간 흐름을 고려하기 위해 1년 시차항 (y_{t-12} , x_{t-12})을 포함한 (5) 및 (6)과 같은 방식을 고려한다.

$$y_t = \alpha_0 + \alpha_1 y_{t-2} + \alpha_2 x_{t-1} + \alpha_3 y_{t-12} + \alpha_4 x_{t-12} + Z_{t-2}^\top \beta + \varepsilon_t, \quad (5)$$

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 x_t + \alpha_3 y_{t-12} + \alpha_4 x_{t-12} + Z_{t-2}^\top \beta + \varepsilon_t. \quad (6)$$

선형방법은 설명변수 x_{t-1} , x_t , Z_{t-2} 를 어떻게 반영하는지에 따라 다음과 같은 두 가지 기준으로 구분된다.

1. FA (Factor-Augmented) 기법 / 축소추정 기법 (Shrinkage Method)
2. AR / ARX : 혹은 항을 포함시킬지 여부

본 연구의 선형방법은 첫번째 기준에 따라 1) FA 기법, 2) 축소추정 기법, 3) 혼합형 기법(Hybrid method)의 세 가지 그룹으로 구분되며, 각 그룹은 AR과 ARX의 두 형태 및 1년 시차항을 포함한 형태를 모두 고려한다.

FA 기법 Stock and Watson (1998, 2002a,b)은 광범위한 거시경제 변수들로부터 적은 수의 요인을 추출하여 AR 구조에 결합하는 예측 방법을 제시하였으며, 해당 방법이 단순 AR 모형에 비해 물가의 예측력을 개선한다는 점을

보였다. 이와 같은 선행연구를 기반으로 본 연구는 1차 및 2차 나우캐스트에 대해 각각 (7) 및 (8)과 같은 방식을 고려한다.

$$y_t = \alpha_0 + \alpha_1 y_{t-2} + \alpha_2 x_{t-1} + Z_{F,t-2}^\top \beta + \varepsilon_t, \quad (7)$$

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 x_t + Z_{F,t-2}^\top \beta + \varepsilon_t. \quad (8)$$

$Z_{F,t-2}$ 는 고차원 설명변수 Z_{t-2} 로부터 주성분분석(PCA: Principal Component Analysis)을 통해 추출한 요인 벡터를 의미한다.¹

이때 요인은 고차원 변수들이 가지는 공통된 흐름을 요약한 지표로, 모든 변수가 요인 산출에 반영되기 때문에 요인 기반 방법은 밀집(Dense) 구조를 갖는다. 이는 많은 변수의 정보를 누락하지 않고 요약해 반영한다는 점에서 Stock and Watson (2002a,b), McCracken and Ng (2016) 등 경제 예측 문헌에서 널리 활용된다.

축소추정 기법 고차원 자료를 활용한 회귀분석에서는 Ridge, LASSO, ElasticNet(Hoerl and Kennard, 1970; Tibshirani, 1996; Zou and Hastie, 2005)과 같은 축소추정(regularization) 기법이 널리 사용된다. 이러한 기법들은 계수에 제약(penalty)을 부여하여 추정의 안정성을 높이고, 다중공선성이나 고차원 자료에서 발생하는 불안정한 추정을 완화한다는 장점이 있다. 하지만 각 축소추정 기법은 제약의 형태가 다르기 때문에, 구조적 속성이 다르다.

먼저, Ridge는 L^2 제약을 통해 계수를 크기가 작은 값으로 수축시키는 방식이지만, 일반적으로 변수 선택을 수행하지 않는다. 즉, Ridge는 FA 기법과 같이 변수 선택 없이 고차원 변수 전체를 모두 활용하므로 밀집 구조를 갖는다. 반면 LASSO와 ElasticNet은 L^1 제약을 통해 계수 중 일부를 0으로 만들어 변수 선택을 수행하므로, 희소(Sparse) 구조를 갖는 변수선택 기법이다. 이러한 방식은 고차원 자료에서 예측력 있는 변수를 선택할 수 있다는 장점이 있다. 이와 같은 축소추정 기법을 적용하기 위해, 본 연구는 (9) 및 (10)과 같은 형태의

¹요인의 수 결정은 요인 기반 방법의 예측 성능에 영향을 줄 수 있다. 본 연구에서는 요인의 수를 1개부터 10개까지 변화시키며 예측 성능을 비교하였고, FA 방법에서 요인을 3개 사용할 때 예측오차가 가장 낮게 나타났다. 이에 따라 본 연구에서는 모형 간 비교의 일관성과 안정성을 확보하기 위해 요인의 수를 3개로 고정하여 분석을 진행하였다. 이러한 설정은 경제 예측 연구에서 요인의 수를 소수로 제한하는 기존 연구(Stock and Watson, 2002a,b, 2006)와 일관된다.

방식을 고려한다.

$$y_t = \alpha_0 + \alpha_1 y_{t-2} + \alpha_2 x_{t-1} + Z_{t-2}^\top \beta + \varepsilon_t, \quad (9)$$

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 x_t + Z_{t-2}^\top \beta + \varepsilon_t. \quad (10)$$

LASSO 혹은 ElasticNet을 활용하여 축소추정을 진행할 경우, Z_{t-2} 중 일부 변수의 계수만 0이 아닌 값으로 남는다.

본 연구는 기본 축소추정 기법에 더해 적응형(Adaptive) Ridge, LASSO, ElasticNet(Zou, 2006)을 고려한다. 적응형 방식은 각 변수에 서로 다른 가중치를 적용하여 변수마다 계수의 축소 정도를 다르게 조절하는 방법으로, 효과적으로 각 변수의 중요도를 반영할 수 있다. 또한, 본 연구에서는 x_{t-1} 혹은 x_t 가 예측력 향상에 기여하는 기간과 그렇지 않은 기간이 있을 수 있다는 점을 고려하여, x_{t-1} 혹은 x_t 에 제약을 적용한 경우와 적용하지 않은 경우를 모두 포함해 예측을 수행한다. 축소추정 기법에 대한 자세한 설명은 부록에 제시한다.

혼합형 기법 FA 기법과 변수선택 기법(LASSO, ElasticNet)이 결합된 경우, 밀집 구조와 희소 구조가 모두 반영된 혼합 구조가 형성된다. 이는 FA 기법에 따른 밀집 구조와 축소추정을 통한 희소 구조를 동시에 활용하는 방식이다. 본 연구는 (11) 및 (12)와 같은 혼합형 방식을 고려한다.

$$y_t = \alpha_0 + \alpha_1 y_{t-2} + \alpha_2 x_{t-1} + Z_{t-2}^\top \beta + Z_{F,t-2}^\top \gamma + \varepsilon_t, \quad (11)$$

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 x_t + Z_{t-2}^\top \beta + Z_{F,t-2}^\top \gamma + \varepsilon_t. \quad (12)$$

$Z_{F,t-2}$ 는 요인을 통해 전체 변수를 요약한 밀집 구조를 반영하고, Z_{t-2} 에서는 일부 변수의 계수만 0이 아닌 값으로 남아 희소 구조를 반영한다. 혼합형 기법에서도 축소추정 기법과 마찬가지로 적응형 방식 및 x_{t-1} 혹은 x_t 에 제약을 적용할지 여부를 추가로 고려하여 예측을 진행한다.

선형방법 요약 표 2는 본 연구에서 활용하는 선형방법을 요약한다. 선형방법은 요인 활용 여부, 축소추정 기법 적용 여부, 그리고 1년 시차항 포함 여부를 기준으로 구분된다.

2.2.3. 비선형방법

비선형 머신러닝 방법을 사용할 때는 설명변수를 이용하는 방식이 다른데, 이는 계산 비용과 적합 과정의 제약을 고려하여 Medeiros *et al.* (2021)과

구조	요인	축소추정	1년 시차항
AR/ARX + FA	O	X	O / X
AR/ARX + Shrinkage Method	X	O	O / X
AR ARX + FA + LASSO/ElasticNet	X	O	O / X

표 2: 선형방법 요약. 이 표는 본 연구에서 활용하는 선형방법을 요약한다. 선형방법은 요인 활용 여부, 축소추정 기법 적용 여부, 그리고 1년 시차항 포함 여부를 기준으로 구분된다.

Table 2: SUMMARY OF LINEAR METHODS. This table summarizes the linear methods utilized in this study. The linear approaches are categorized based on the use of factor analysis, the application of shrinkage estimation techniques, and the inclusion of a one-year lag term.

한희준 (2023, 2024)의 방식을 따르기 때문이다. 이에 따라 비선형방법은 1차 나우캐스트 단계에서만 적용한다.

앞서 설명한 고차원 설명변수 벡터 Z_{t-2} 는 서비스 및 상품 CPI를 제외한 93개 설명변수로 익익월 중순에 최신화가 완료된다는 점을 반영하여 예측 시점에서 실제로 이용 가능한 최신 정보가 $t-2$ 시점이라는 사실을 고려해 구성한 것이다. 이 $t-2$ 시점의 93개 설명변수에 로그 차분한 디플레이터와 로그 차분한 전월의 서비스 혹은 상품 CPI x_{t-2} 를 추가한 $t-2$ 시점의 총 95개 변수들에서 주성분분석을 통해 4개의 요인을 추출한다. Z_{t-2}^a 를 $t-2$ 시점의 95개 변수와 해당 변수들로부터 추출한 4개의 요인($P_{1,t-2}, P_{2,t-2}, P_{3,t-2}, P_{4,t-2}$)을 포함한 것으로 정의하면, (13)과 같이 표기할 수 있다.

$$Z_{t-2}^a = (Z_{t-2}, y_{t-2}, x_{t-2}, P_{1,t-2}, P_{2,t-2}, P_{3,t-2}, P_{4,t-2}), \quad (13)$$

시계열의 의존성을 반영하기 위해 Medeiros *et al.* (2021)는 1차에서 4차까지의 시차변수들을 설명변수로 취급하였는데, 이를 따르면 각 머신러닝 방법에 사용하는 설명변수의 수는 총 $396 = (95 + 4) \times 4$ 개가 된다. 참고로 Medeiros *et al.* (2021)은 FRED-MD의 123개 변수와 이 변수들로부터 추출한 4개의 주성분요소(Principal Component Factor)들의 과거값들(1차에서 4차까지의 시차변수)을 포함하여 총 508개의 설명변수를 사용하였다.

이를 이용하여 1차 나우캐스트를 실행하는 기법은 (14)와 같이 표현할 수 있다.

$$y_t = f(Z_{t-2}^a, Z_{t-3}^a, Z_{t-4}^a, Z_{t-5}^a), \quad (14)$$

이때 $f(\cdot)$ 는 다양한 머신러닝 기법을 나타낸다. 본 연구에서 활용한 머신러닝 기법으로는 Tree 기반 방법인 랜덤 포레스트(Random forest), XGBoost, Boruta Algorithm과 신경망 기반 방법인 인공신경망(Neural network), LSTM(Long short-term memory) 등을 고려한다. 또한, 총 396개의 설명변수를 사용하는 경우의 선형 기법인 축소추정 기법(Ridge, LASSO, ElasticNet)과 CSR(Complete subset regression) 등도 함께 비교한다.²

각 머신러닝 기법에 대한 자세한 설명은 부록에 수록한다.

3. 예측 절차

본 연구는 구간이동 방식을 활용하여 산업활동동향 디플레이터를 예측한다. 구간이동 방식에서 설정한 추정 기간(Window size)은 120개월(10년)이며, 이에 따라 예측 대상 기간은 2014년 8월부터 2025년 6월까지 총 131개월이다. 예측은 1차 나우캐스트와 2차 나우캐스트로 구분하여 수행한다. 또한 Medeiros *et al.* (2021)과 같이 직접 예측 방식을 적용한다. Marcellino, Stock and Watson (2006) 등 기존 연구에 따르면 직접 예측 방식은 모형 오지정에 강건하다는 장점이 있다.

본 연구에서는 종속변수를 로그 차분하고 각 방법에 반영하여 예측을 수행한다. 다만 예측의 최종 대상은 로그 차분값이 아니라 서비스 및 소매판매 디플레이터의 수준값이므로, 각 방법으로부터 얻은 로그 차분 예측치는 (15)에 따라 수준값으로 변환하여 최종 예측값을 산출한다.

$$\hat{Y}_t = Y_{t-h} \cdot \exp\left(\sum_{j=0}^{h-1} \hat{y}_{t-j}\right), \quad (15)$$

여기서 1차 나우캐스트의 경우 $h = 2$ 이고, 2차 나우캐스트의 경우 $h = 1$ 이며, 최종 예측값은 \hat{Y}_t 이다. 실제값과 각 방법의 예측값의 차이는 RMSE(Root Mean Squared Error)를 이용해 평가한다. 또한, 각 방법이 벤치마크 모형 대비 얼마나 우수한 예측력을 가지는지 평가하기 위해, 각 방법의 RMSE를 벤치마크 모형의 RMSE로 나눈 비율을 사용하여 상대적 예측력을 비교한다. 이 비율이 1보다 작으면 해당 방법이 벤치마크보다 우수한 예측력을 보였음을 의미한다.

다음 장에서는 위와 같은 예측 절차를 기반으로 산출된 선형방법, 비선형방법의 예측 결과를 벤치마크 모형의 예측 결과와 비교하여 제시하며, 벤치마크 모형의 결과는 부록의 표 9에 제시한다.

² 해당 머신러닝 기법들은 경제학 분야에서 폭넓게 활용되고 있다. 일례로 노시현·한희준(2025)은 해당 방법론을 적용하여 상장기업의 부도 예측을 분석한 바 있다.

4. 예측실험 결과

본 장에서는 2장에서 제시한 자료와 방법을 적용하여 3장의 예측 절차를 거쳐 수행한 산업활동동향 디플레이터의 예측실험 결과를 제시한다. 각 방법의 예측 성능은 RMSE를 기준으로 평가하고, 벤치마크 모형의 RMSE와 비교하여 상대적 성능을 판단한다. 이를 통해 본 연구에서 제시한 선형 및 비선형방법이 벤치마크 모형 대비 어느 정도 예측력을 개선하는지 구체적으로 평가한다.

4.1. 선형방법

선형방법의 결과는 1년 시차항 적용 여부에 따라 구분한 뒤, FA 기법 / 축소추정 기법 / 혼합형 기법 각각에서 1차와 2차 나우캐스트에서 가장 우수한 예측력을 보인 방법의 결과를 제시한다. 이를 통해 1년 시차항 적용 및 예측 방법에 따른 예측력 차이를 확인할 수 있다. 표 3 및 표 4는 각각 서비스와 소매 판매 예측에서, 각 방법의 RMSE를 벤치마크 모형의 RMSE로 나누어 상대적 예측력 지표를 제시한 표이다.³ 1차 나우캐스트와 2차 나우캐스트 각각에서 가장 뛰어난 예측 성능을 보인 예측방법은 밑줄로 표시한다.

서비스 디플레이터의 경우에는 AR + Ridge 방식이 가장 우수한 성능을 보였다. 실제로 1차 나우캐스트에서는 벤치마크 모형 대비 약 9.4% 낮은 RMSE, 2차 나우캐스트에서는 약 4.3% 낮은 RMSE를 기록하며 두 시점 모두에서 일관된 우위를 보였다. 이는 서비스 디플레이터 예측에서 밀집 구조가 예측력 제고에 유리하다는 점을 시사한다.

반면, 소매판매 디플레이터에서는 ARX + FA + LASSO/ElasticNet이 가장 낮은 RMSE를 기록하였다. 1차 나우캐스트에서는 벤치마크 모형 대비 약 7.0%, 2차 나우캐스트에서는 약 7.5%의 RMSE 개선을 보이며 두 시점에서 모두 우수한 성능을 나타냈다. 이는 소매판매 디플레이터가 고차원 설명변수의 전체 흐름과 변수 선택이 동시에 필요한 특성을 지니고 있음을 보여준다. 또한, 2차 나우캐스트에서는 서비스와 소매판매 디플레이터 모두에서 1년 시차항을 반영한 방식이 가장 좋은 예측력을 기록하였다. 이는 2차 나우캐스트에서 물가의 연간흐름을 고려한 추가 정보가 예측력을 높이는 데 기여했음을 시사한다.

³본문에 제시하지 않은 전체 예측 결과는 부록의 표 9 - 표 12에 제시한다. 각 방법의 세부적인 RMSE 및 MAE(Mean Absolute Error) 결과는 저자에게 요청 시 제공받을 수 있다.

	1차 나우캐스트	2차 나우캐스트
벤치마크 모형 (ARX 모형)	1.000	1.000
FA 기법		
ARX + FA	1.011	1.042
축소추정 기법		
AR + Ridge (적응형)	<u>0.906</u>	0.976
혼합형 기법		
ARX + FA + ElasticNet (P)	0.954	0.976
AR + FA + ElasticNet (적응형, P)	0.936	1.007
1년 시차항 적용		
FA 기법		
ARX + FA	1.005	1.012
축소추정 기법		
AR + Ridge (적응형)	0.916	<u>0.957</u>
혼합형 기법		
AR + FA + ElasticNet	0.912	0.988
AR + FA + LASSO (적응형, P)	0.938	0.961

표 3: 선형방법 결과 (서비스 디플레이터). 이 표는 선형방법의 서비스 디플레이터 예측 결과를 제시한다. 1년 시차항 적용 여부에 따라 구분한 뒤, FA 기법 / 축소추정 기법 / 혼합형 기법 각각에서 1차와 2차 나우캐스트에서 가장 우수한 예측력을 보인 방법의 결과를 제시하며, 각 방법의 RMSE를 벤치마크 모형의 RMSE로 나누어 상대적 예측력 지표를 제시한다. ‘P’는 서비스 CPI에 축소 제약을 적용한 경우를 의미하며, 1차 나우캐스트와 2차 나우캐스트 각각에서 가장 뛰어난 예측 성능을 보인 예측방법은 밑줄로 표시한다.

Table 3: LINEAR METHOD RESULTS (SERVICE DEFLATOR). This table presents the nowcasting results for the service deflator using linear methods. The models are categorized by whether a one-year lag term is applied. For each category—Factor Analysis (FA), Shrinkage Estimation, and Hybrid methods—we report the results from the model showing the highest predictive power in the first and second nowcasts. Relative performance indicators are provided by dividing the RMSE of each method by the benchmark model’s RMSE. ‘P’ denotes the application of shrinkage constraints to the service CPI. The method with the best forecasting performance in each nowcast is underlined.

	1차 나우캐스트	2차 나우캐스트
벤치마크 모형 (ARX 모형)	1.000	1.000
FA 기법		
ARX + FA	0.981	0.998
축소추정 기법		
ARX + LASSO	0.950	0.992
혼합형 기법		
ARX + FA + LASSO (P)	0.964	1.021
ARX + FA + LASSO (적응형)	<u>0.930</u>	1.097
1년 시차항 적용		
FA 기법		
ARX + FA	0.967	0.947
축소추정 기법		
AR + ElasticNet	0.964	1.054
ARX + ElasticNet	0.987	0.935
혼합형 기법		
ARX + FA + ElasticNet (P)	1.075	<u>0.925</u>
ARX + FA + ElasticNet (P)	0.997	1.033

표 4: 선형방법 결과 (소매판매 디플레이터). 이 표는 선형방법의 소매판매 디플레이터 예측 결과를 제시한다. 1년 시차항 적용 여부에 따라 구분한 뒤, FA 기법 / 축소추정 기법 / 혼합형 기법 각각에서 1차와 2차 나우캐스트에서 가장 우수한 예측력을 보인 방법의 결과를 제시하며, 각 방법의 RMSE를 벤치마크 모형의 RMSE로 나누어 상대적 예측력 지표를 제시한다. 'P'는 상품 CPI에 축소 제약을 적용한 경우이며, 1차 나우캐스트와 2차 나우캐스트 각각에서 가장 뛰어난 예측 성능을 보인 예측방법은 밑줄로 표시한다.

Table 4: LINEAR METHOD RESULTS (RETAIL SALES DEFLATOR). This table presents the forecasting results for the retail sales deflator using linear methods. The results are categorized by the inclusion of a one-year lag term. For each category—Factor Analysis (FA), Shrinkage Estimation, and Hybrid methods—the results of the model with the highest predictive power in the first and second nowcasts are reported. Relative forecasting performance is presented as the RMSE of each method divided by the benchmark model's RMSE. 'P' indicates the application of shrinkage constraints to the commodities CPI. The forecasting methods that exhibit the best performance in the first and second nowcasts, respectively, are underlined.

	분류	1차 나우캐스트	2차 나우캐스트
서비스 디플레이터	AR/ARX	AR	AR
	희소 구조 (Sparse)	X	X
	밀집 구조 (Dense)	O (Ridge)	O (Ridge)
	1년 시차항 적용	X	O
소매판매 디플레이터	AR/ARX	ARX	ARX
	희소 구조 (Sparse)	O (LASSO)	O (ElasticNet)
	밀집 구조 (Dense)	O (Factor)	O (Factor)
	1년 시차항 적용	X	X

표 5: 선형방법 결과 요약. 이 표는 선형방법의 예측 결과를 각 디플레이터별로 정리하여, 각 예측 시점에서 어떤 구조가 가장 우수한 성능을 보였는지를 보여준다.
 Table 5: SUMMARY OF LINEAR METHOD RESULTS. This table summarizes the forecasting results of the linear methods for each deflator, highlighting which model structure achieved the best performance at each forecasting point.

선형방법 결과 요약 표 5는 위와 같은 선형방법의 예측 결과를 각 디플레이터별로 정리하여, 각 예측 시점에서 어떤 구조가 가장 우수한 성능을 보였는지를 보여준다. 표 5에서는 AR/ARX, 희소 구조, 밀집 구조, 그리고 1년 시차항 포함 여부에 대해, 각 시점에서 최적의 성능을 기록한 방식을 O 기호로 표시한다.

4.2. 비선형방법

표 6 및 표 7은 각각 서비스와 소매판매 디플레이터 예측에서, 각 방법의 RMSE를 벤치마크 모형의 RMSE로 나누어 상대적 예측력 지표를 제시한 표이다.

표 6의 서비스 디플레이터 예측의 결과는 벤치마크 모형의 RMSE에 비교한 각 모형의 예측오차 비율이 모든 경우 1보다 크다는 것을 나타낸다. 이는 벤치마크 모형보다 예측력이 우수한 모형이 존재하지 않는다는 것을 의미한다. 이러한 결과가 나타난 가장 큰 이유는 벤치마크 모형과 표 6의 방법들이 설명변수로 사용하는 정보에 차이가 있기 때문으로 추측된다. 벤치마크 모형은 로그 차분한 전월의 서비스 CPI x_{t-1} 를 사용하고 있으나 표 6의 모형들은 이를 사용하지 않고 $t-2$ 시점까지의 변수들만을 사용한다. 비록 표 6의 비선형방법들은 시차변수들까지 포함하여 많은 (총 396개의) 변수를 설명변수로 사용하고는 있지만 $t-1$ 시점의 로그 차분한 서비스 CPI 변수를 포함하지 않고 있는데, 표 6의 결과는 전월의 서비스 CPI 변수가 반영하는 정보가 디플레이터 나우캐스팅에 매우 중요하다는 것을 의미한다.

	1차 나우캐스트	
	로그 차분 변환	차분 변환
선형 및 회귀 기반 방법		
Ridge	2.509	1.275
LASSO	2.497	1.291
LASSO (적응형)	2.509	1.291
ElasticNet	2.490	1.271
ElasticNet (적응형)	2.498	1.280
CSR (Complete Subset Regression)	2.535	1.346
Tree 기반 방법		
랜덤 포레스트 (Random Forest)	2.524	1.275
XGBoost	2.476	<u>1.218</u>
Boruta Algorithm (변수 선택)	1.231	1.225
신경망 기반 방법		
신경망	2.530	1.326
LSTM (Long Short-term Memory)	2.791	1.473

표 6: 비선형방법 결과 (서비스 디플레이터). 이 표는 비선형방법의 서비스 디플레이터 예측 결과를 제시한다. 선형 및 회귀 기반 방법 / Tree 기반 방법 / 신경망 기반 방법으로 구분하여 예측 결과를 제시하며, 각 방법의 RMSE를 벤치마크 모형의 RMSE로 나누어 상대적 예측력 지표를 제시한다. 비선형 방법 중 가장 뛰어난 예측성능을 보인 예측방법은 밑줄로 표시한다.

Table 6: NONLINEAR METHOD RESULTS (SERVICE DEFLATOR). This table presents the nowcasting results for the service deflator using nonlinear methods. The results are categorized into linear and regression-based methods, tree-based methods, and neural network models. Relative forecasting performance is presented as the RMSE of each method divided by the benchmark model's RMSE. The nonlinear method exhibiting the best forecasting performance is underlined.

표 7은 소매판매 디플레이터 예측의 결과를 나타낸다. 표 6의 결과처럼 벤치마크 모형의 RMSE에 비교한 각 모형의 예측오차 비율이 모든 경우 1보다 큰 것으로 나타났다. 이는 앞서 설명한 것처럼 로그 차분한 전월의 상품 CPI x_{t-1} 가 반영하는 정보가 소매판매 디플레이터 나우캐스팅에 매우 중요하기 때문에 나타난 결과임을 추측할 수 있다.

표 6와 표 7은 두 디플레이터 예측에서 공통적으로, 비선형 방법이 로그 차분 변환보다 차분 변환을 적용했을 때 더 우수한 예측력을 보인다는 점을

	1차 나우캐스트	
	로그 차분 변환	차분 변환
선형 및 회귀 기반 방법		
Ridge	1.556	1.553
LASSO	1.521	1.521
LASSO (적응형)	1.570	1.564
ElasticNet	1.518	1.516
ElasticNet (적응형)	1.563	1.560
CSR (Complete Subset Regression)	1.611	1.599
Tree 기반 방법		
랜덤 포레스트 (Random Forest)	1.555	1.565
XGBoost	1.536	1.528
Boruta Algorithm (변수 선택)	1.468	<u>1.459</u>
신경망 기반 방법		
신경망	1.731	1.735
LSTM (Long Short-term Memory)	1.868	1.849

표 7: 비선형방법 결과 (소매판매 디플레이터). 이 표는 비선형방법의 소매판매 디플레이터 예측 결과를 제시한다. 선형 및 회귀 기반 방법 / Tree 기반 방법 / 신경망 기반 방법으로 구분하여 예측 결과를 제시하며, 각 방법의 RMSE를 벤치마크 모형의 RMSE로 나누어 상대적 예측력 지표를 제시한다. 비선형 방법 중 가장 뛰어난 예측성능을 보인 예측방법은 밑줄로 표시한다.

Table 7: NONLINEAR METHOD RESULTS (RETAIL SALES DEFLATOR). This table presents the nowcasting results for the retail sales deflator using nonlinear methods. The results are categorized into linear and regression-based methods, tree-based methods, and neural network models. Relative forecasting performance is presented as the RMSE of each method divided by the benchmark model's RMSE. The nonlinear method exhibiting the best forecasting performance is underlined.

보여준다. 또한 Tree 기반 방법이 두 디플레이터 예측 모두에서 비선형 모형 중 가장 높은 예측 성능을 나타낸다. 이러한 결과는 향후 비선형 방법을 확장하거나 개선하여 추가 연구를 수행할 때 유용한 시사점을 제공한다.

5. 결론

본 연구는 서비스 디플레이터와 소매판매 디플레이터를 대상으로 다양한 선형 및 비선형 예측 방법의 성과를 비교·분석하였다. 분석에는 AR/ARX 모형을 비롯하여 요인분석, 축소추정 기법(Ridge, LASSO, ElasticNet), 1년 시차항 포함 여부, 자료 변환, 그리고 Random Forest, XGBoost, LSTM 등의 비선형 머신러닝 방법을 포함하였다. 예측은 구간이동 추정과 직접 예측 방식을 적용하여 수행하였으며, 정보 가용성에 따라 1차 및 2차 나우캐스트 환경을 구분하여 분석하였다.

예측 결과는 다음과 같은 특징을 보인다. 첫째, 서비스 디플레이터의 경우 AR + Ridge 방식이 전반적으로 가장 우수한 예측 성과를 나타낸다. 둘째, 소매 판매 디플레이터에서는 ARX 구조에 요인분석과 축소추정을 결합한 방식이 상대적으로 우수하며, 1차 나우캐스트에서는 LASSO, 2차 나우캐스트에서는 ElasticNet이 각각 가장 낮은 RMSE를 기록한다. 셋째, 1년 시차항은 공통적으로 2차 나우캐스트 환경에서 예측력을 개선하는 역할을 하며, 두 디플레이터 모두에서 해당 시점에 시차항을 포함한 모형이 가장 낮은 RMSE를 보인다. 넷째, 비선형 머신러닝 방법은 모든 설정에서 선형 방법 대비 낮은 예측 성과를 보인다. 다만 이는 비선형 방법의 본질적 한계라기보다, 높은 계산 비용과 모형 적합 과정의 제약으로 인해 서비스·상품 CPI를 ARX 모형과 같이 분리하여 반영하지 못하고, 설명변수를 동일한 시점에서 일괄적으로 처리한 설계상의 제약에 기인한 결과로 해석될 수 있다. 이러한 점은 나우캐스팅 환경에서 변수의 시점 정보가 예측 성능에 중요한 역할을 함을 시사한다.

이상의 결과는 산업활동동향 디플레이터의 나우캐스팅에서 예측 방법의 성과가 디플레이터의 특성과 예측 시점에 따라 상이하게 나타날 수 있음을 보여준다. 특히 서비스 디플레이터와 소매판매 디플레이터가 서로 다른 예측 구조에서 최적 성과를 보였다는 점은, 산업활동동향 디플레이터가 단일한 구조로 설명되기보다는 지표별로 상이한 정보 구조와 변동 특성을 내포하고 있을 가능성을 시사한다. 또한 2차 나우캐스트 환경에서 시차항의 역할이 두 디플레이터 모두에서 중요하게 나타난 점은, 예측 시점에 따라 동태적 정보의 활용 방식이 예측 성과에 영향을 미칠 수 있음을 보여준다.

비선형 머신러닝 방법의 상대적으로 낮은 예측 성과는, 본 연구에서 각 예측 시점별로 이용 가능한 최신 정보를 충분히 반영하지 못한 모형 설계의 제약과 밀접하게 관련되어 있다. 이러한 결과는 산업활동동향 디플레이터 나우캐스팅에서 시점별로 이용 가능한 정보를 충분히 반영하는 것이 중요하다는 것을 시사한다. 향후 연구에서는 설명변수의 시점 처리를 보다 정교하게 반

영하거나, 각 기법의 구조를 나우캐스팅 환경에 맞게 세분화함으로써 비선형 방법의 예측 성능을 추가적으로 검토할 수 있을 것이다.

향후 서비스 디플레이터와 소매판매 디플레이터 간 예측 구조의 차이가 어떠한 경제적 요인에서 비롯되는지, 그리고 이러한 차이가 물가 결정 및 기대 형성과정과 어떻게 연결되는지에 대해서는 추가적인 분석이 필요하다. 거시 경제 모형을 기반으로 산업별 디플레이터 변동의 기저 요인을 보다 체계적으로 규명한다면, 산업활동동향 디플레이터 예측에 대한 이해를 한층 심화시킬 수 있을 것으로 기대된다.

부록

A. 자료

표 8은 본 연구에서 활용한 각 변수의 변환 방식을 제시한다. 시계열 y_t 에 대한 Tcode는 다음과 같은 데이터 변환을 의미한다: (1) y_t ; (2) Δy_t ; (5) $\Delta \log y_t$. 변수들은 통계청 국가통계포털 및 한국은행경제통계시스템을 통해 수집하였다.

변수	Tcode
Service Consumer Price Indices	5
Commodity Consumer Price Indices	5
Capital goods Production index (seasonally adjusted)	5
Capital goods Shipment index (seasonally adjusted)	5
Capital goods Inventory index (seasonally adjusted)	5
Intermediate goods Production index (seasonally adjusted)	5
Intermediate goods Shipment index (seasonally adjusted)	5
Intermediate goods Inventory index (seasonally adjusted)	5
Consumers' goods Production index (seasonally adjusted)	5
Consumers' goods Shipment index (seasonally adjusted)	5
Consumers' goods Inventory index (seasonally adjusted)	5
Whole country Production index (seasonally adjusted)	5
Whole country Shipment index (seasonally adjusted)	5
Whole country Inventory index (seasonally adjusted)	5

Continued on next page

변수 설명	Tcode
Monthly survey of Mining and Manufacturing Production index (seasonally adjusted)	5
Monthly survey of Mining and Manufacturing Shipment index (seasonally adjusted)	5
Monthly survey of Mining and Manufacturing Inventory index (seasonally adjusted)	5
Monthly survey of Mining and Quarrying Production index (seasonally adjusted)	5
Monthly survey of Mining and Quarrying Shipment index (seasonally adjusted)	5
Monthly survey of Manufacturing Production index (seasonally adjusted)	5
Monthly survey of Manufacturing Shipment index (seasonally adjusted)	5
Monthly survey of Manufacturing Inventory index (seasonally adjusted)	5
Monthly survey of Electricity, gas and steam Production index (seasonally adjusted)	5
Monthly survey of Electricity, gas and steam Shipment index (seasonally adjusted)	5
Unemployment rate	2
Total economically active population (thousand persons)	5
Total employed persons (thousand persons)	5
Unemployed persons (duration < 3 months)	5
Unemployed persons (3 months ≤ duration < 6 months)	5
Unemployed persons (6 months ≤ duration < 12 months)	5
Unemployed persons (6 months and over)	5
Unemployed persons (12 months and over)	2
House price index (national)	5
House price index (Seoul capital area)	5
House price index (provincial area)	5

Continued on next page

변수 설명	Tcode
National future tendency business condition BSI	1
National future tendency sales growth BSI	1
National future tendency profitability BSI	1
National future tendency financial situation BSI	1
National future tendency human resources BSI	1
Index of all industry production: Industrial production (seasonally adjusted)	5
Index of all industry production: Construction (seasonally adjusted)	5
Index of all industry production: Service industry (seasonally adjusted)	5
Index of all industry production: Public administration (seasonally adjusted)	5
Monetary base (average, seasonally adjusted)	5
Currency in circulation	5
Central bank liabilities to depository corporations	5
M1	5
M2	5
CD rate (91 days)	2
CP rate (91 days)	2
Treasury bond yield (1-year)	2
Treasury bond yield (3-year)	2
Treasury bond yield (5-year)	2
Treasury bond yield (10-year)	2
Corporate bond yield (AA-)	2
Bank of Korea base rate	2
Corporate bond yield (BBB-)	2
Weighted average lending rate to households (housing loans)	5
Treasury bond spread (10-year minus 1-year)	5
KRW/USD exchange rate	5
KRW/CNY exchange rate	5
KRW/JPY (per 100 yen) exchange rate	5

Continued on next page

변수 설명	Tcode
KRW/EUR exchange rate	5
KRW/GBP exchange rate	5
KRW/CAD exchange rate	5
KRW/CHF exchange rate	5
CPI: Finished consumer goods	5
CPI: Food and non-alcoholic beverages	5
CPI: Clothing and footwear	5
CPI: Furnishings and household maintenance	5
CPI: Health	5
CPI: Transport	5
CPI: Education	5
CPI: Miscellaneous goods and services	5
Domestic supply price index: Raw materials	5
Domestic supply price index: Intermediate goods and services	5
Domestic supply price index: Final goods and services	5
Domestic supply price index: Raw and intermediate materials	5
Import price index: Agricultural, forestry and marine products	5
Import price index: Bituminous coal	5
Import price index: Crude oil	5
Import price index: Liquefied natural gas	5
Import price index: Manufacturing products	5
Export price index: Agricultural, forestry and marine products	5
Export price index: Manufacturing products	5
KOSPI index	5
KOSDAQ index	5
VKOSPI index	1
Expected inflation	5
Composite leading indicator	5
Composite lagging indicator	5
Oil price	5
US CPI (CPIAUCSL)	5

Continued on next page

변수 설명	Tcode
US 10-year treasury bond yield (IRLTLT01USM156N)	2

표 8: 변수 및 변수 변환 코드 (Tcode). 이 표는 본 연구에서 활용한 각 변수의 변환 방식을 제시한다. 시계열 y_t 에 대한 Tcode는 다음과 같은 데이터 변환을 의미한다: (1) y_t ; (2) Δy_t ; (5) $\Delta \log y_t$.

Table 8: VARIABLE TRANSFORMATIONS AND TCODES. This table presents the transformation methods applied to each variable in this study. For a given time series y_t , the Tcode indicates the following data transformations: (1) y_t ; (2) Δy_t ; and (5) $\Delta \log y_t$.

B. 벤치마크 모형

	서비스 디플레이터		소매판매 디플레이터	
	1차 나우캐스트	2차 나우캐스트	1차 나우캐스트	2차 나우캐스트
RMSE	0.00428	0.00251	0.00705	0.00265
MAE	0.00346	0.00193	0.00570	0.00204

표 9: 벤치마크 모형 (ARX) 예측 결과. 이 표는 벤치마크 모형의 1차 및 2차 나우캐스트에 대한 RMSE와 MAE 결과를 제시한다.

Table 9: FORECASTING RESULTS OF BENCHMARK MODEL (ARX). This table presents the Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) for the first and second nowcasts of the benchmark model.

C. 선형방법

C.1. 축소추정 기법 해설

LASSO, Elastic Net 등의 축소추정 기법(Shrinkage Method)은 OLS 추정과 마찬가지로 잔차의 제곱의 합을 최소화하여 추정을 진행하지만, 다음과 같이 제약 함수(Penalty Function)를 추가로 고려하여 설명변수의 계수를 축소시킨다.

$$\hat{\beta}_{\text{shrinkage}} = \arg \min_{\beta} \left[\sum_{i=1}^n (y_i - X_i^T \beta) + p(\beta, \lambda) \right],$$

이 식에서 X 는 모든 설명변수를 포함하는 벡터이며, 해당 설명변수들은 모두 표준화(Standardize)되어 반영된다. $p(\beta, \lambda)$ 는 제약 함수이며, 이는 설명력이

약한 변수의 계수를 작게 하거나 설명변수의 차원을 축소하는 역할을 한다. 각 축소추정 기법마다 제약함수는 다르게 정의된다.

λ 는 축소추정 기법의 튜닝 파라미터(Tuning Parameter)이다. $\lambda = 0$ 인 경우, 제약함수 없이 추정이 이뤄지기 때문에, 이는 OLS와 같다. 본 연구에서는 교차검증(Cross-validation)을 통해 튜닝 파라미터를 결정한다.

LASSO는 다음과 같이 L^1 제약 함수를 가진다.

$$p(\beta, \lambda) = \lambda \sum_{j=1}^m |\beta_j|,$$

적응형 LASSO는 두 단계로 추정을 진행한다. 첫번째 단계에서는 OLS, LASSO, Ridge 등을 통해 초기 추정치(Initial Estimate)를 구한다. 두번째 단계에서는 첫번째 단계에서 구한 계수를 활용해 가중치 $w_j = |\tilde{\beta}_{\text{initial},j}|^{-1}$ 를 설정하고, L^1 제약에 가중치 w_j 를 추가해 다음과 같은 제약함수를 이용해 추정을 진행한다.

본 연구에서는 Ridge 추정치를 이용하여 가중치를 $w_j = |\tilde{\beta}_{\text{ridge},j}|^{-1}$ 로 정하였다.

ElasticNet은 다음과 같은 L^1 과 L^2 의 가중평균 제약을 둔다.

$$p(\beta, \lambda) = \lambda \sum_{j=1}^m \{\alpha \beta_j^2 + (1 - \alpha) |\beta_j|\},$$

본 연구에서는 Medeiros *et al.* (2021)과 같이 $\alpha = 0.5$ 로 설정하였다. 또한 적응형 ElasticNet 및 적응형 Ridge도 고려했는데, 이는 적응형 LASSO와 마찬가지로 각각의 제약에 가중치 w_j 가 추가된 제약함수를 사용한다. 본 연구에서는 적응형 ElasticNet과 Ridge도 가중치를 $w_j = |\tilde{\beta}_{\text{ridge},j}|^{-1}$ 로 설정하여 추정을 진행하였다.

C.2. 전체 예측 결과

표 10 - 표 11은 선형 예측 방법과 각 선형방법별 예측 결과(벤치마크 모형 대비 RMSE 비율)를 제시한다. 각 표에서 'P'는 서비스 CPI 및 상품 CPI에 축소 제약을 적용한 경우를, 'Ada'는 적응형 축소추정 방식을 의미한다.

방법	1차 나우캐스트			2차 나우캐스트		
	로그차분	차분	수준	로그차분	차분	수준
AR + FA	1.088	1.083	1.116	1.185	1.180	1.205
ARX + FA	1.011	1.005	1.164	1.042	1.037	1.239
AR + LASSO	0.985	0.978	1.051	1.080	1.062	1.168
ARX + LASSO	0.973	0.964	1.076	0.995	1.000	1.209
AR + Ridge	0.985	0.983	1.047	1.065	1.068	1.168
ARX + Ridge	0.956	0.968	1.092	0.999	0.995	1.189
AR + Elastic	0.977	0.985	1.060	0.991	1.023	1.109
ARX + Elastic	0.948	0.951	1.138	0.979	0.989	1.157
AR + FA + LASSO	0.957	0.963	1.056	1.002	1.012	1.097
ARX + FA + LASSO	0.936	0.956	1.112	0.982	0.983	1.150
AR + FA + LASSO P	0.969	0.960	1.068	0.997	1.017	1.116
ARX + FA + LASSO P	0.954	0.962	1.109	0.979	1.003	1.169
AR + FA + Elastic	0.947	0.956	1.063	1.017	1.008	1.090
ARX + FA + Elastic	0.938	0.948	1.120	0.987	0.990	1.152
AR + FA + Elastic P	0.946	0.961	1.064	1.026	1.000	1.119
ARX + FA + Elastic P	0.954	0.959	1.118	0.976	0.995	1.156
AR + Ada LASSO	0.935	0.926	1.018	0.987	1.015	1.161
ARX + Ada LASSO	0.934	0.931	1.097	1.010	1.001	1.225
AR + Ada Ridge	0.906	0.949	1.060	0.969	1.041	1.166
ARX + Ada Ridge	0.914	0.920	1.128	0.976	0.976	1.193
AR + Ada Elastic	0.928	0.918	1.005	1.009	1.009	1.162
ARX + Ada Elastic	0.939	0.924	1.093	0.991	0.981	1.221

방법	1차 나우캐스트			2차 나우캐스트		
	로그차분	차분	수준	로그차분	차분	수준
AR + FA + Ada LASSO	0.948	0.928	1.133	1.019	1.031	1.174
ARX + FA + Ada LASSO	0.945	0.925	1.127	1.024	1.001	1.217
AR + FA + Ada LASSO P	0.937	0.923	1.030	1.007	0.997	1.176
ARX + FA + Ada LASSO P	0.939	0.934	1.075	1.009	0.976	1.225
AR + FA + Ada Elastic	0.947	0.929	1.134	1.015	1.035	1.173
ARX + FA + Ada Elastic	0.939	0.920	1.106	1.020	1.006	1.217
AR + FA + Ada Elastic P	0.936	0.921	1.028	1.007	1.000	1.177
ARX + FA + Ada Elastic P	0.937	0.940	1.071	1.008	0.965	1.220
1년 시차항 적용						
AR + FA	1.038	1.032	1.169	1.076	1.073	1.249
ARX + FA	1.005	1.001	1.339	1.012	1.010	1.337
AR + LASSO	0.958	0.960	1.117	1.044	1.046	1.196
ARX + LASSO	0.968	0.960	1.252	0.986	0.981	1.283
AR + Ridge	0.963	0.974	1.108	1.044	1.041	1.199
ARX + Ridge	0.983	0.989	1.253	0.976	0.991	1.266
AR + Elastic	0.926	0.937	1.118	0.976	0.979	1.128
ARX + Elastic	0.923	0.932	1.278	0.972	0.994	1.241
AR + FA + LASSO	0.937	0.927	1.165	0.981	0.986	1.225
ARX + FA + LASSO	0.927	0.922	1.278	0.977	0.976	1.248
AR + FA + LASSO P	0.935	0.941	1.115	0.973	0.989	1.198

방법	1차 나우캐스트			2차 나우캐스트		
	로그차분	차분	수준	로그차분	차분	수준
ARX + FA + LASSO P	0.928	0.935	1.248	0.974	0.981	1.270
AR + FA + Elastic	0.912	0.928	1.164	0.988	0.998	1.245
ARX + FA + Elastic	0.948	0.925	1.261	0.971	0.971	1.276
AR + FA + Elastic P	0.952	0.941	1.118	0.997	1.006	1.182
ARX + FA + Elastic P	0.933	0.934	1.260	0.990	0.971	1.267
AR + Ada LASSO	0.921	0.920	1.130	0.959	1.018	1.198
ARX + Ada LASSO	0.932	0.945	1.213	0.981	0.982	1.282
AR + Ada Ridge	0.916	0.923	1.142	0.957	1.021	1.191
ARX + Ada Ridge	0.948	0.934	1.232	0.993	0.974	1.305
AR + Ada Elastic	0.914	0.927	1.122	0.974	1.030	1.214
ARX + Ada Elastic	0.949	0.923	1.254	1.017	0.998	1.269
AR + FA + Ada LASSO	0.933	0.912	1.165	1.002	1.043	1.225
ARX + FA + Ada LASSO	0.931	0.909	1.278	1.006	0.988	1.248
AR + FA + Ada LASSO P	0.938	0.938	1.115	0.961	1.039	1.198
ARX + FA + Ada LASSO P	0.932	0.919	1.248	0.977	1.020	1.270
AR + FA + Ada Elastic	0.920	0.952	1.164	0.996	1.039	1.245
ARX + FA + Ada Elastic	0.927	0.909	1.261	1.002	0.981	1.276
AR + FA + Ada Elastic P	0.943	0.943	1.118	0.982	1.021	1.182

방법	1차 나우캐스트			2차 나우캐스트		
	로그차분	차분	수준	로그차분	차분	수준
ARX + FA + Ada Elastic P	0.918	0.919	1.260	0.991	0.996	1.267

표 10: 벤치마크 모형과 RMSE 비교 (서비스 디플레이터). 이 표는 본 연구에서 활용한 모든 선형방법의 결과를 제시하며, 선형방법의 RMSE를 벤치마크 모형의 RMSE로 나눈 결과를 보여준다.

Table 10: COMPARISON OF RMSE WITH BENCHMARK MODEL (SERVICE DEFLATOR). This table presents the results for all linear methods utilized in this study, showing the ratio of the RMSE of each linear method to the RMSE of the benchmark model.

방법	1차 나우캐스트			2차 나우캐스트		
	로그차분	차분	수준	로그차분	차분	수준
AR + FA	1.319	1.314	1.421	2.175	2.166	2.274
ARX + FA	0.981	0.980	1.294	0.998	0.998	1.873
AR + LASSO	1.014	1.003	1.410	1.347	1.279	2.253
ARX + LASSO	0.950	0.964	1.218	0.992	0.992	1.802
AR + Ridge	1.014	1.001	1.400	1.347	1.272	2.259
ARX + Ridge	0.956	0.964	1.231	1.039	0.986	1.784
AR + Elastic	1.021	0.999	1.395	1.050	1.034	2.033
ARX + Elastic	0.993	0.969	1.302	1.004	1.005	1.840
AR + FA + LASSO	0.999	0.981	1.372	1.087	1.031	1.953
ARX + FA + LASSO	0.970	0.973	1.290	1.038	1.019	1.716
AR + FA + LASSO P	0.993	0.990	1.394	1.054	1.013	1.947
ARX + FA + LASSO P	0.964	0.967	1.314	1.021	0.995	1.786
AR + FA + Elastic	0.987	0.989	1.386	1.090	1.034	1.971
ARX + FA + Elastic	0.985	0.980	1.290	1.014	0.996	1.733
AR + FA + Elastic P	0.993	1.000	1.391	1.030	1.054	1.958

방법	1차 나우캐스트			2차 나우캐스트		
	로그차분	차분	수준	로그차분	차분	수준
ARX + FA + Elastic P	0.980	0.960	1.300	1.024	1.010	1.754
AR + Ada LASSO	0.999	1.018	1.507	1.102	1.212	2.510
ARX + Ada LASSO	0.966	0.989	1.292	1.135	1.104	1.850
AR + Ada Ridge	0.982	1.109	1.428	1.029	1.525	2.352
ARX + Ada Ridge	0.977	0.971	1.275	1.112	1.080	1.913
AR + Ada Elastic	1.002	1.019	1.480	1.133	1.239	2.529
ARX + Ada Elastic	0.965	0.980	1.317	1.051	1.111	1.884
AR + FA + Ada LASSO	0.996	1.002	1.482	1.097	1.157	2.456
ARX + FA + Ada LASSO	0.930	0.971	1.316	1.097	1.071	1.884
AR + FA + Ada LASSO P	1.002	0.976	1.454	1.141	1.174	2.415
ARX + FA + Ada LASSO P	0.970	0.987	1.315	1.114	1.121	1.895
AR + FA + Ada Elastic	1.006	1.007	1.458	1.093	1.223	2.401
ARX + FA + Ada Elastic	0.957	0.976	1.314	1.093	1.080	1.885
AR + FA + Ada Elastic P	1.006	0.978	1.451	1.158	1.210	2.356
ARX + FA + Ada Elastic P	0.966	0.968	1.312	1.109	1.102	1.894
1년 시차항 적용						
AR + FA	1.290	1.283	1.406	2.117	2.112	2.267
ARX + FA	0.967	0.965	1.125	0.947	0.945	1.271
AR + LASSO	1.079	1.071	1.445	1.441	1.367	2.411

방법	1차 나우캐스트			2차 나우캐스트		
	로그차분	차분	수준	로그차분	차분	수준
ARX + LASSO	0.982	0.963	1.250	1.009	1.046	1.270
AR + Ridge	1.051	1.067	1.458	1.445	1.376	2.418
ARX + Ridge	0.994	1.001	1.224	1.053	1.022	1.284
AR + Elastic	0.964	1.016	1.540	1.054	1.068	2.021
ARX + Elastic	0.987	1.002	1.280	0.935	0.954	1.260
AR + FA + LASSO	1.064	1.062	1.567	1.111	1.077	2.132
ARX + FA + LASSO	1.054	1.014	1.248	0.950	1.001	1.276
AR + FA + LASSO P	1.030	1.019	1.566	1.035	1.066	2.402
ARX + FA + LASSO P	1.060	1.028	1.232	0.959	0.932	1.223
AR + FA + Elastic	1.033	1.012	1.549	1.105	1.071	2.115
ARX + FA + Elastic	1.084	1.016	1.281	0.981	0.968	1.292
AR + FA + Elastic P	1.026	1.007	1.551	1.077	1.026	2.072
ARX + FA + Elastic P	1.075	1.008	1.262	0.925	0.931	1.208
AR + Ada LASSO	1.021	1.145	1.470	1.045	1.634	2.407
ARX + Ada LASSO	1.007	1.017	1.231	1.055	1.020	1.349
AR + Ada Ridge	1.029	1.135	1.487	1.044	1.638	2.434
ARX + Ada Ridge	1.035	1.002	1.209	1.025	1.084	1.343
AR + Ada Elastic	1.103	1.172	1.570	1.183	1.298	2.437
ARX + Ada Elastic	1.060	1.050	1.314	1.073	1.106	1.307
AR + FA + Ada LASSO	1.075	1.042	1.567	1.002	1.206	2.132

방법	1차 나우캐스트			2차 나우캐스트		
	로그차분	차분	수준	로그차분	차분	수준
ARX + FA + Ada LASSO	1.060	1.050	1.248	1.047	1.052	1.276
AR + FA + Ada LASSO P	1.030	1.089	1.566	1.002	1.221	2.402
ARX + FA + Ada LASSO P	1.018	1.030	1.232	1.021	1.009	1.223
AR + FA + Ada Elastic	1.040	1.040	1.549	1.007	1.261	2.115
ARX + FA + Ada Elastic	1.053	1.029	1.281	1.076	1.085	1.292
AR + FA + Ada Elastic P	1.075	1.022	1.551	1.003	1.266	2.072
ARX + FA + Ada Elastic P	0.997	0.997	1.262	1.033	1.052	1.208

표 11: 벤치마크 모형과 RMSE 비교 (소매판매 디플레이터). 이 표는 본 연구에서 활용한 모든 선형방법의 결과를 제시하며, 선형방법의 RMSE를 벤치마크 모형의 RMSE로 나눈 결과를 보여준다.

Table 11: COMPARISON OF RMSE WITH BENCHMARK MODEL (RETAIL SALES DEFLATOR). This table presents the results for all linear methods utilized in this study, showing the ratio of the RMSE of each linear method to the RMSE of the benchmark model.

D. 비선형방법

D.1. 비선형 머신러닝 기법 해설

Complete Subset Regression(CSR) CSR은 Elliott, Gargano and Timmermann (2013)에 의해 소개된 앙상블(Ensemble) 방법으로 총 m 개의 설명변수에서 k 개의 설명변수로 이루어진 모든 가능한 조합을 이용하여 선형 추정 및 예측을 하고 이 예측치들의 평균을 최종 예측치로 삼는 것이다. 다만 설명변수의 수가 너무 많은 경우 추정해야 할 수가 급격히 증가하기 때문에($m = 24$ 그리고 $k = 4$ 인 경우 12,650번의 추정이 필요함), Medeiros *et al.* (2021)은 사전에 각 설명변수를 추정하여 t -통계값을 구하고 이의 절댓값 크기가 큰 25개의

변수만을 선별하고 $k = 4$ 를 설정하여 CSR을 실시하였다. 본 연구에서도 이와 같은 방법을 따랐다.

Random Forest 모형 랜덤 포레스트는 Breiman (2001)에 의해 소개되었으며, 결정 트리(Decision Tree)를 기반으로 하는 Bootstrap aggregation(Bagging)을 분산이 작아지도록 개선한 방법이다. 랜덤 포레스트는 각 부트스트랩 샘플에서 분할을 수행할 때마다 사용할 설명 변수를 무작위로 선택한다. 각 노드에서 분할을 수행할 때, 전체 m 개의 예측 변수 중에서 m^* 개만 무작위로 선택하여 후보로 사용한다. 선택된 m^* 개의 변수 중 하나만을 이용하여 해당 노드에서 최적의 분할을 수행한다. 새로운 노드를 만들 때마다 다시 m^* 개의 새로운 설명 변수를 무작위로 선택한다(고정된 설명 변수가 아님). 따라서 부트스트랩 샘플에서 만들어지는 결정 트리들을 서로 상관없게 만들고(Decorrelate) 결과적으로 배깅에 비해 더욱 분산을 감소시키게 된다. m^* 값은 일반적으로 회귀 문제에서는 $m^* = \frac{m}{3}$ 로 선택하고, 본 연구에서도 이를 따랐다.

XGBoost 모형 부스팅은 순차적으로(sequentially) 학습을 진행한다. 각 의사결정나무는 이전에 학습된 의사결정나무의 정보를 활용하여 성장한다. 부스팅은 부트스트랩 샘플을 사용하지 않고, 원본 데이터가 반복적으로 수정되면서 새로운 의사결정나무를 학습한다. 부스팅은 여러 개의 약한 학습기를 결합하면서 점진적으로 예측 성능을 향상시키는 방식으로 동작한다. 부스팅 알고리즘 중 Chen and Guestrin (2016)이 소개한 XGBoost(Extreme Gradient Boosting)는 그래디언트 부스팅을 개선한 알고리즘으로, 정규화(regularization) 기능을 추가하여 기존 그래디언트 부스팅의 약점인 과대적합을 방지한다. 특히 XGBoost는 병렬 CPU 환경에서 병렬 학습이 가능하기 때문에 기존 순차적으로 학습하는 그래디언트 부스팅보다 빠르게 학습할 수 있다. XGBoost는 머신러닝 경진대회인 캐글 경연대회(Kaggle contest)에서 상위를 차지한 많은 데이터 과학자가 이를 이용하면서 널리 알려져 있다.

인공신경망(Neural Network) 모형 그림 2은 은닉층(Hidden layer)이 두 개이고 각각이 네 개의 노드(Node)를 가진 인공신경망 모형을 나타낸다. 본 연구에서 이용한 인공신경망 모형은 세 개의 은닉층과 각 은닉층이 32, 16, 8개의 노드를 지닌 것으로 설정하였으며, 활성화 함수(Activation function)로는 일반적인 Rectifier 함수를 사용하였고 학습 횟수(Epoch)는 100으로 정하였다.

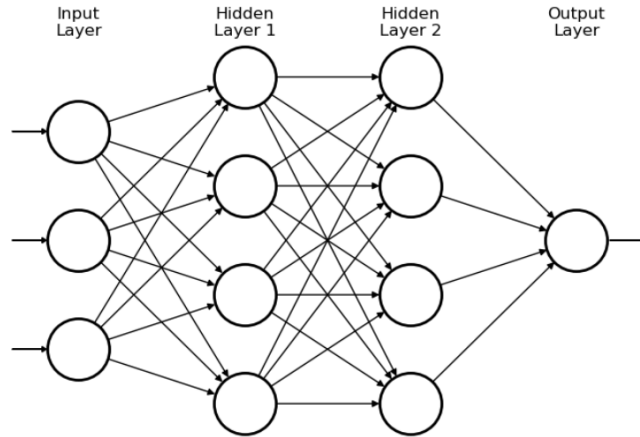


그림 2: 인공신경망 (한희준, 2025). 이 그림은 인공신경망의 구조를 보여준다.
Figure 2: NEURAL NETWORK (HAN, 2025). This figure illustrates the structure of an artificial neural network.

LSTM 모형 기존 RNN (Recurrent Neural Network) 모형은 BPTT (Back-Propagation Through Time) 방식을 사용하는데, 역전파(Back Propagation)시 Vanishing / Exploding Gradient 문제 때문에 시계열의 장기 의존성(Long-Term Dependence)을 제대로 반영하지 못하는 한계가 있다. LSTM 모형은 이 한계를 극복하기 위해 장기 기억을 담당하는 그림 3과 같이 Cell State를 도입하여 정보를 장기간에 걸쳐 저장 및 접근할 수 있게 한다.

LSTM 모형은 분석 시 학습 횟수(Epoch), 한 번에 처리할 데이터의 크기(Batch Size) 등의 하이퍼 파라미터를 조정할 수 있는데, 본 연구에서는 학습 횟수(Epoch)를 100으로, 그리고 한 번에 처리할 데이터의 크기(Batch Size)를 25로 설정하였다.

D.2. 전체 예측 결과

표 12 - 표 13는 비선형 예측 방법과 각 모형별 예측 결과(벤치마크 모형 대비 RMSE 비율)를 제시한다.

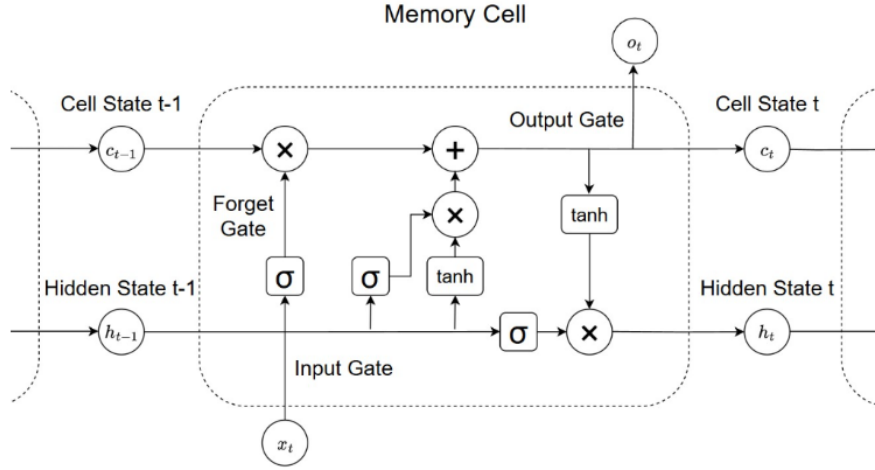


그림 3: LSTM (한희준, 2025). 이 그림은 LSTM 구조를 보여준다.
 Figure 3: LSTM (HAN, 2025). This figure illustrates the structure of a Long Short-Term Memory (LSTM).

방법	1차 나우캐스트		
	로그차분	차분	수준
Ridge	2.509	1.275	11.817
LASSO	2.497	1.291	1.456
Ada LASSO	2.509	1.291	1.411
ElasticNet	2.490	1.271	1.558
Ada ElasticNet	2.498	1.280	1.506
CSR (Complete Subset Regression)	2.535	1.346	1.314
랜덤포레스트 (Random Forest)	2.524	1.275	4.801
XGBoost	2.476	1.218	3.036
Boruta Algorithm	1.231	1.225	3.738
신경망 (Neural Network)	2.530	1.326	11.094
LSTM (Long Short-term Memory)	2.791	1.473	8.635

표 12: 벤치마크 모형과 RMSE 비교 (서비스 디플레이터). 이 표는 비선형방법의 RMSE를 벤치마크 모형의 RMSE로 나눈 결과를 제시한다.
 Table 12: COMPARISON OF RMSE WITH BENCHMARK MODEL (SERVICE DEFLATOR). This table presents the ratio of the RMSE of nonlinear methods to the RMSE of the benchmark model.

방법	1차 나우캐스트		
	로그차분	차분	수준
Ridge	1.556	1.553	6.560
LASSO	1.521	1.521	1.979
Ada LASSO	1.570	1.564	1.845
ElasticNet	1.518	1.516	2.063
Ada ElasticNet	1.563	1.560	1.815
CSR (Complete Subset Regression)	1.611	1.599	1.657
랜덤포레스트 (Random Forest)	1.555	1.565	3.700
XGBoost	1.536	1.528	2.880
Boruta Algorithm	1.468	1.459	3.078
신경망 (Neural Network)	1.731	1.735	6.303
LSTM (Long Short-term Memory)	1.868	1.849	4.402

표 13: 벤치마크 모형과 RMSE 비교 (소매판매 디플레이터). 이 표는 비선형방법의 RMSE를 벤치마크 모형의 RMSE로 나눈 결과를 제시한다.

Table 13: COMPARISON OF RMSE WITH BENCHMARK MODEL (RETAIL SALES DEFLATOR). This table presents the ratio of the RMSE of nonlinear methods to the RMSE of the benchmark model.

참고문헌

김동석 (2003). *산업연관표 시계열화를 통한 한국의 산업구조변화 분석*, 한국개발연구원 (KDI).

(Translated in English) Kim, D. S. (2003). *Analysis of Changes in Korea's Industrial Structure through Time-Series Input-Output Tables*, Korea Development Institute (KDI).

이승윤 · 이은경 · 한진현 (2017). “기계학습과 혼합주기 모형을 활용한 단기전망시스템,” *조사통계월보* 7, 한국은행.

(Translated in English) Lee, S. Y., Lee, E. K., and Han, J. H. (2017). “Short-term Forecasting System Using Machine Learning and Mixed-Frequency Models,” *BOK Monthly Bulletin* 7, Bank of Korea.

이현창 · 최동규 · 김용건 · 허정 (2022). “디지털 신기술을 이용한 실시간 당분기경제전망 (GDP nowcasting) 시스템 개발,” *BOK이슈노트* 26, 한국은행.

(Translated in English) Lee, H. C., Choi, D. G., Kim, Y. G., and Hur, J. (2022).

“Development of a Real-time Quarterly Economic Forecasting (GDP Nowcasting) System Using Digital Technologies,” *BOK Issue Note* 26, Bank of Korea.

한희준 (2023). “랜덤 포레스트(Random Forest)의 시계열 적용에 관한 연구: 한국 물가상승률 예측 사례 분석,” *경제학연구* 71, 37–73.

(Translated in English) Han, H. (2023). “A Study on the Time Series Application of Random Forest: An Analysis of Korean Inflation Forecasting,” *The Korean Economic Review* 71, 37–73.

한희준 (2024). “머신러닝 기법을 이용한 한국 인플레이션 예측: 시계열 특성과 머신러닝 예측의 관계에 관한 사례 연구,” *국제경제연구* 30, 21–56.

(Translated in English) Han, H. (2024). “Forecasting Korean Inflation Using Machine Learning Techniques: A Case Study on the Relationship between Time Series Characteristics and Machine Learning Forecasts,” *International Economic Journal* 30, 21–56.

한희준 (2025). *머신러닝과 경제·금융 시계열 예측: R과 Python을 활용한 실습 중심 입문서*, 자유아카데미.

(Translated in English) Han, H. (2025). *Machine Learning and Economic/Financial Time Series Forecasting: A Practice-Oriented Introductory Text Using R and Python*, Free Academy.

노시현·한희준 (2025). “머신러닝과 오버샘플링을 이용한 상장기업 부도예측 연구,” *Journal of Economic Theory and Econometrics* 36, 59–82.

(Translated in English) Noh, S. H. and Han, H. (2025). “Bankruptcy Prediction for Listed Companies in Korea Using Machine Learning and Oversampling Methods,” *Journal of Economic Theory and Econometrics* 36, 59–82.

Bernanke, B. S. and J. Boivin (2003). “Monetary policy in a data-rich environment,” *Journal of Monetary Economics* 50, 525–546.

Breiman, L. (2001). “Random forests,” *Machine Learning* 45, 5–32.

Chen, T. and C. Guestrin (2016). “XGBoost: A scalable tree boosting system,” *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794.

- Elliott, G., A. Gargano, and T. Timmermann (2013). “Complete subset regressions,” *Journal of Econometrics* 177, 357–373.
- Goulet Coulombe, P., M. Leroux, D. Stevanovic, and S. Surprenant (2022). “How Is machine learning useful for macroeconomic forecasting?” *Journal of Applied Econometrics* 37, 920–964.
- Hoerl, A. E. and R. W. Kennard (1970). “Ridge regression: Biased estimation for nonorthogonal problems,” *Technometrics* 12, 55–67.
- Marcellino, M., J. H. Stock, and M. W. Watson (2006). “A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series,” *Journal of Econometrics* 135, 499–526.
- Masini, R. P., M. C. Medeiros, and E. F. Mendes (2023). “Machine learning advances for time series forecasting,” *Journal of Economic Surveys* 37, 76–111.
- McCracken, M. W. and S. Ng (2016). “FRED-MD: A monthly database for macroeconomic research,” *Journal of Business & Economic Statistics* 34, 574–589.
- Medeiros, M. C., G. F. Vasconcelos, Á. Veiga, and E. Zilberman (2021). “Forecasting inflation in a data-rich environment: The benefits of machine learning methods,” *Journal of Business & Economic Statistics* 39, 98–119.
- Stock, J. H. and M. W. Watson (1998). “Diffusion indexes,” NBER Working Paper No. 6702.
- Stock, J. H. and M. W. Watson (2002a). “Macroeconomic forecasting using diffusion indexes,” *Journal of Business & Economic Statistics* 20, 147–162.
- Stock, J. H. and M. W. Watson (2002b). “Forecasting using principal components from a large number of predictors,” *Journal of the American Statistical Association* 97, 1167–1179.
- Stock, J. H. and M. W. Watson (2006). “Forecasting with many predictors,” in *Handbook of Economic Forecasting*, Vol. 1, 515–554.

Tibshirani, R. (1996). "Regression shrinkage and selection via the Lasso," *Journal of the Royal Statistical Society Series B* 58, 267–288.

Zou, H. (2006). "The adaptive Lasso and Its oracle properties," *Journal of the American Statistical Association* 101, 1418–1429.

Zou, H. and T. Hastie (2005). "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society Series B* 67, 301–320.