

대한민국 거시경제 Nowcasting: 고빈도 데이터 통합과 동적요인모형 분석을 중심으로

ABSTRACT

본 연구는 세 가지 주요 한국 거시경제 변수에 대한 나우캐스팅을 위해 네 가지 예측 모형(ARIMA, VAR, 동적요인모형, 심층 동적요인모형)의 성능을 비교한다. 대상 변수는 생산(전산업생산지수: KOIPALL.G), 투자(설비투자지수: KOEQUIPTE), 소비(도소매판매액: KOWRCCNSE)이다. 모형들은 22개 예측 시점(1개월부터 22개월까지)에서 표준화된 지표를 사용하여 평가되며, 각 시점에 대한 지표를 평균하여 최종 성능 지표로 사용한다. 이를 통해 서로 다른 시계열 규모 간 공정한 비교가 가능하다. 실험적 평가를 통해 모형 성능을 대상 변수와 예측 시점에 걸쳐 제시한다.

키워드: 나우캐스팅, 동적요인모형, 고빈도 데이터, 거시경제 예측, 딥러닝

1 서론

1.1 연구 배경

거시경제 변수의 정확한 예측은 정책 의사결정과 기업의 전략적 계획 수립에 중요함. 특히 **nowcasting**는 공식 통계가 발표되기 전 현재 시점의 거시경제 변수를 실시간으로 추정하는 것으로, 정책 의사결정자와 시장 참여자에게 시의성 있는 정보를 제공함 [1].

Nowcasting는 forecasting과 구분되는 중요한 특징을 가짐:

- **Forecasting:** 과거 데이터를 기반으로 미래 값을 예측하는 것으로, 예측 시점이 참조 기간 이후임.
- **Nowcasting:** 현재 또는 최근 과거 기간의 값을 추정하는 것으로, 공식 통계 발표 전에 실시간 데이터 흐름을 활용하여 추정함.

Nowcasting의 핵심은 **vintage 정보세트**의 개념과 밀접히 연관됨. 각 거시경제 지표는 발표 시차(publication lag)가 다르며, 예를 들어 산업생산지수는 참조월의 다음 달 말에 발표되지만, 한국은행의 BSI/ESI/CSI 등 심리지수는 참조월 말에 발표되어 약 3-5주 선행함. 따라서 목표 월에 대해 예측을 수행하는 시점(예: 4주 전, 1주 전)에 따라 사용 가능한 정보세트가 달라지며, 이를 **vintage**라고 함. DFM과 DDFM은 Kalman filter를 통해 이러한 비동기적 데이터 발표와 결측치를 자연스럽게 처리할 수 있어 nowcasting에 특히 적합함 [1, 2].

본 연구는 세 가지 주요 실험을 통해 거시경제 변수 예측 및 nowcasting 성능을 평가함:

- **예측 실험:** 과거 데이터를 기반으로 미래 값을 예측하는 실험으로, ARIMA, VAR, DFM, DDFM 모형을 비교함. 대상 변수는 생산(KOIPALL.G), 투자(KOEQUIPTE), 소비(KOWRCCNSE)이며, 다단계 예측 성능을 평가함.

- **Nowcasting 실험:** 공식 통계 발표 전 현재 시점 거시경제 변수를 추정하는 실험으로, DFM, DDFM, MAMBA 모형을 활용함. 각 목표 월에 대해 4주 전, 1주 전 시점에서 예측을 수행하며, 시리즈별 발표 시차를 기준으로 미발표 데이터를 마스킹하여 실제 운영 환경을 시뮬레이션함.
- **고빈도 데이터 실험:** 고빈도 지표(주별 전력거래량, BSI)의 단일변수 예측에서의 활용 가능성을 탐색하는 실험으로, MIDAS-AR, ARX, XGBoost 모형을 비교함.

각 실험의 구체적인 설정과 결과는 실험 설계 및 결과 섹션에서 상세히 제시됨.

1.2 선행연구 검토

동적요인모형(DFM)

- 많은 시계열에서 공통 요인을 추출해 소수의 동태적 요인으로 설명하는 대표적 차원축소 기법으로, 관측식과 상태식을 갖는 state-space 형태를 취함 [3].
- 대규모 이질적 거시 지표 간의 공분산 구조를 소수 요인으로 집약해 수십 수백 개 변수의 동시 예측이 가능하며, Kalman filter를 통해 누락 · 비동기 데이터(혼합주기, jagged edges)를 자연스럽게 처리할 수 있다는 점에서 나우캐스팅에 핵심적으로 활용됨 [1, 2].
- 뉴욕 연준 Nowcast 플랫폼 등 실무 시스템 역시 DFM 기반으로 실시간 발표 흐름에 맞춰 주별 업데이트를 수행해 예측치를 개선하는 방식을 채택함 [2].

심층 동적요인모형(DDFM)

- 오토인코더 기반 비선형 인코더를 사용해 요인 구조를 학습함으로써 전통적 DFM의 선형 가정을 완화함 [5].
- 비선형 인코더는 고차원 거시 데이터의 복잡한 상호작용을 더 적은 요인으로 포착하면서도, 요인층 뒤에는 여전히 선형 state-space를 두어 필터링 · 스무딩 안정성을 유지함.
- 최근 연구들은 고빈도 외생 변수와 결합하거나 비선형 활성화 · 정규화 기법을 도입해 예측 성능을 개선하고 있으며, 이러한 딥 요인 접근을 DFM과 병행 비교하여 실증적으로 평가함.

혼합 데이터 샘플링(MIDAS)

- MIDAS(Mixed Data Sampling)는 서로 다른 주기의 데이터를 통합하여 예측하는 모형으로, 고빈도 데이터(주간, 일간)와 저빈도 데이터(월간)를 함께 활용함 [6].
- MIDAS 모형은 고빈도 변수를 저빈도 종속변수에 매핑하기 위해 다양한 가중치 함수(exp-Almon, Beta, Almon 등)를 사용하며, 이를 통해 주별 · 일별 데이터를 월별 · 분기별 예측에 활용할 수 있음.
- Clements & Galvão (2008)는 MIDAS-AR 모형의 추정 절차를 제안하여 자기회귀 성분과 고빈도 변수의 효과를 동시에 추정할 수 있도록 함 [7].

- 고빈도 지표(주별 전력거래량, BSI)의 단일변수 예측에서의 활용 가능성을 탐색하기 위해 MIDAS-AR 모형을 수행함.

1.3 이론적 배경

1.3.1 동적요인모형

- **요인모형:** 많은 시계열에서 공통 요인을 추출하여 소수의 요인으로 설명하는 차원축소 기법 [3].
- **동적요인모형(DFM):** 관측식과 상태식을 갖는 state-space 형태로, 대규모 이질적 거시 지표 간의 공분산 구조를 소수 요인으로 집약함 [3, 1].
- **칼만필터:** 실시간 데이터 흐름을 재귀적으로 처리하여 각 시점의 예측을 업데이트하며, 비동기적 데이터 발표와 결측치를 자연스럽게 처리함 [1, 2].
- **EM 알고리즘:** state-space 모형의 파라미터 추정을 위한 최대우도 추정 방법 [2].
- **VAE (Variational Autoencoder):** 오토인코더 기반 비선형 인코더를 사용하여 요인 구조를 학습하는 딥러닝 접근법 [5].

1.3.2 SSM

- **상태 공간 모델(State Space Model):** 관측 변수와 은닉 상태 변수를 분리하여 모델링하는 프레임워크로, 시계열 데이터의 동적 특성을 포착함.
- **MAMBA:** 선택적 상태 공간 모델(Selective State Space Model)을 기반으로 선형 시간 복잡도로 장기 의존성을 효과적으로 포착하는 최신 딥러닝 아키텍처 [?]. 비선형 선택 메커니즘을 통해 입력에 따라 상태 전이를 동적으로 조정하여 시계열 데이터의 복잡한 패턴을 학습함.

1.3.3 혼합주기

- **혼합주기 데이터 처리:** 서로 다른 주기(주간, 월간, 분기)의 데이터를 통합하여 예측하는 방법으로, tent kernel, MIDAS 가중치 함수 등을 활용함 [8, 6].
- **Mariano & Murasawa (2003):** 주간/월간/분기 데이터를 state-space 모형에서 통합 처리하는 방법론을 제안함 [8].
- **MIDAS:** 고빈도 데이터를 저빈도 종속변수에 매핑하기 위한 가중치 함수를 사용하는 혼합 데이터 샘플링 방법 [6, 7].

2 실험 설계

2.1 데이터

예측 실험 데이터

- 대상 변수:** 생산(전산업생산지수: KOIPALL.G), 투자(설비투자지수: KOEQUIPTE), 소비(도소매판매액: KOWR-CCNSE) 3개 변수
- 생산 부문 모형:** 총 41개 변수로 구성됨. 고용, 산업생산, 서베이(기업경기, 소비자 동향) 등 주요 월간 지수와 주간 데이터를 포함함. 기업경기동향 조사는 해당월 중 발표되어 속보성이 높으며, 주가지수 등 금융변수, 뉴스 심리지수, 미국 경제정책불확실성 지수를 포함함. 상세 변수 구성은 부록의 표 6를 참조함.
- 투자 부문 모형:** 총 41개 변수로 구성됨. 고용, 설비투자, 건설 등 주요 지표와 속보성 높은 서베이(기업경기, 소비자 동향) 등 주요 월간 지수와 주간 데이터 9개를 포함함. 고빈도 데이터는 주가지수 등 금융변수, 뉴스심리지수, 미국 경제정책불확실성 지수 및 투자관련 섹터 주가지수를 포함함. 상세 변수 구성은 부록의 표 7를 참조함.
- 모형:** ARIMA, VAR, DFM, DDFM 4개 모형을 비교함

Nowcasting 실험 데이터

- 대상 변수:** 생산(전산업생산지수: KOIPALL.G), 투자(설비투자지수: KOEQUIPTE) 2개 변수
- 데이터 구성:** 생산 및 투자 부문 모형의 변수 구성을 활용하며, 주간 및 월간 데이터를 혼합하여 사용함. 혼합 주기 처리 방법은 동적요인모형 섹션(2.2.1)을 참조함.
- 모형:** DFM, DDFM, MAMBA 모형을 활용함
- 평가 시점:** 각 목표 월에 대해 4주 전, 1주 전 시점에서 예측을 수행하며, 시리즈별 발표 시차(publication lag)를 기준으로 미발표 데이터를 마스킹함

고빈도 데이터 실험 데이터

- 종속변수:** 월별 전산업생산지수(계절조정)의 전월대비 성장률 및 전년동월비
- 설명변수:** 주별 전력거래량(로그-STL 계절조정 후 주간 성장률), 월별 BSI(수준 및 전년동월비)
- 표본 분할:** Train(2002–2020년), Validation(2021–2022년), Test(2023–2024년)
- Vintage:** h0(전월 말), h1–h4(당월 1–4주)
- 모형:** MIDAS-AR(1), AR(1) 벤치마크, 선형 ARX, XGBoost 기반 비선형 모형을 비교함

데이터 전처리

- 변환:** 시계열별 변환 유형('lin', 'log', 'chg' 등) 적용
- 결측치 처리:** forward-fill → backward-fill → naive forecaster 순차 적용

- 표준화:

- ARIMA/VAR: 원본 스케일 유지
- DFM/DDFM: StandardScaler 적용 (평균 0, 표준편차 1)

추가 데이터셋(실험 미활용)

- 산업생산지수 nowcasting을 위한 고빈도 공공데이터 조사를 수행하였으며, 실험에 직접 활용하지 않은 추가 데이터 소스들을 정리함. 상세 내용은 부록 A를 참조함.
- 주요 후보: 한국전력거래소 전력수급현황 API, 한국은행 뉴스심리지수, NLIC 주별 해상운임지수, 한국은행 BSI/ESI/CSI/CBSI, 항만 물동량 통계 등

2.2 예측 모형

2.2.1 동적요인모형

- 동적요인모형(DFM)은 많은 시계열에서 공통 요인을 추출해 소수의 동태적 요인으로 설명하는 대표적 차원축소 기법으로, 관측식과 상태식을 갖는 state-space 형태를 취함 [3]. 대규모 이질적 거시 지표 간의 공분산 구조를 소수 요인으로 집약해 수십 수백 개 변수의 동시 예측이 가능하며, Kalman filter를 통해 누락·비동기 데이터 (흔합주기, jagged edges)를 자연스럽게 처리할 수 있다는 점에서 나우캐스팅에 핵심적으로 활용됨 [1, 2].
- DFM의 기본 구조는 다음과 같음:

$$y_t = \lambda'_i f_t + e_t \quad (1)$$

$$f_t = A_1 f_{t-1} + A_2 f_{t-2} + A_3 f_{t-3} + A_4 f_{t-4} + u_t \quad (2)$$

여기서 y_t 는 관측 데이터, f_t 는 은닉 요인(latent factors) 벡터임.

- DFM은 state-space 형태로 표현되며, measurement equation과 transition equation으로 구성됨. EM 알고리즘으로 파라미터 추정, 칼만 필터와 스무더로 요인 추정 [2]. 칼만 필터는 실시간 데이터 흐름을 재귀적으로 처리하여 각 시점의 예측을 업데이트하며, 데이터의 품질과 시의성을 기반으로 가중치를 부여함. 이는 nowcasting에 특히 유용한 특성으로, 비동기적 데이터 발표와 결측치를 자연스럽게 처리할 수 있음 [1].
- DFM 모형에서 요인 식별을 위한 factor loading 제약 가정이 nowcasting 성과를 저해하는 요소로 추정됨에 따라, 요인식별 가정 없이 DFM 모형을 추정하고 nowcasting 성과를 측정함. 요인 개수는 Ahn & Horenstein (2013)의 Eigenvalue ratio 테스트 등을 참고하여 설정함 [?].
- 주/월 흔합 주기 처리: 주간 데이터와 월간 데이터를 함께 활용하는 흔합 주기 모형을 사용함. DFM의 기본 frequency를 주간('w')으로 설정하고, dfm-python의 mixed_freq=True 옵션을 통해 tent kernel이 자동으로 적용됨. 흔합주기 변환은 Mariano & Murasawa (2003) 방법을 따름 [8].

- 월간 지수의 전월대비 상승률은 주간지수의 전월비 상승률 4개의 평균값으로 표현됨:

$$y_t^m = \frac{1}{4}y_t + \frac{1}{4}y_{t-1} + \frac{1}{4}y_{t-2} + \frac{1}{4}y_{t-3} = \frac{1}{4}(I + L + L^2 + L^3)y_t \quad (3)$$

$$\Delta y_t^m = y_t^m - y_{t-4}^m = \frac{1}{4}(I + L + L^2 + L^3)(y_t - L^4y_t) \quad (4)$$

$$= \frac{1}{4}\Delta^4 y_t + \frac{1}{4}\Delta^4 y_{t-1} + \frac{1}{4}\Delta^4 y_{t-2} + \frac{1}{4}\Delta^4 y_{t-3} \quad (5)$$

여기서 y_t^m 는 월간 수준값, Δy_t^m 는 월간 전월대비 상승률, y_t 는 주간 수준값, $\Delta^4 y_t$ 는 주간 전월비 상승률, L 은 시차 연산자임.

- 혼합주기 모형에서 분기 성장률은 Mariano & Murasawa (2003)에 따라 시차(lag) 4개 월간 성장률의 가중합으로 표현됨 [8]:

$$y_t^q = \frac{1}{3}y_t + \frac{1}{3}y_{t-1} + \frac{1}{3}y_{t-2} = \frac{1}{3}(I + L + L^2)y_t \quad (6)$$

$$\Delta y_t^q = y_t^q - y_{t-3}^q = \frac{1}{3}(I + L + L^2)(y_t - L^3y_t) \quad (7)$$

$$= \frac{1}{3}(I + L + L^2)(\Delta y_t + \Delta y_{t-1} + \Delta y_{t-2}) \quad (8)$$

$$= \frac{1}{3}\Delta y_t + \frac{2}{3}\Delta y_{t-1} + \frac{3}{3}\Delta y_{t-2} + \frac{2}{3}\Delta y_{t-3} + \frac{1}{3}\Delta y_{t-4} \quad (9)$$

여기서 y_t^q 는 분기 수준값, Δy_t^q 는 분기 성장률, y_t 는 월간 수준값, Δy_t 는 월간 성장률, L 은 시차 연산자임.

- 분기지수의 전분기 대비 성장률은 주간지수 상승률 20개 시차의 가중 평균값으로 표현됨:

$$\Delta y_t^q = \frac{1}{3}(I + L + L^2)(\Delta y_t^m + \Delta y_{t-1}^m + \Delta y_{t-2}^m) \quad (10)$$

$$= \frac{1}{3}\Delta y_t^m + \frac{2}{3}\Delta y_{t-1}^m + \frac{3}{3}\Delta y_{t-2}^m + \frac{2}{3}\Delta y_{t-3}^m + \frac{1}{3}\Delta y_{t-4}^m \quad (11)$$

$$= \frac{1}{12}(\Delta^4 y_t + \Delta^4 y_{t-1} + \Delta^4 y_{t-2} + \Delta^4 y_{t-3}) + \frac{1}{12}(\Delta^4 y_{t-4} + \dots) \quad (12)$$

$$+ \frac{3}{12}(\Delta^4 y_{t-8} + \dots) + \frac{2}{12}(\Delta^4 y_{t-12} + \dots) + \frac{1}{12}(\Delta^4 y_{t-16} + \dots + \Delta^4 y_{t-19}) \quad (13)$$

따라서 공동요인 5개, 잔차항의 자기회귀, 5개 주간 지표, 30개 월간 지표, 6개 분기 데이터로 구성된 DFM 모형을 가정하면, $20 \times 5 + (5 + 4 \times 30 + 20 \times 6) = 345$ 개의 상태변수가 필요하여 모형 추정이 어려워짐. 주간과 월간 데이터만을 활용하여 모형 복잡성을 관리함.

2.2.2 DDFM

- 심층 동적요인모형(DDFM)은 오토인코더 기반 비선형 인코더를 사용해 요인 구조를 학습함으로써 전통적 DFM의 선형 가정을 완화한다 [5]. 비선형 인코더는 고차원 거시 데이터의 복잡한 상호작용을 더 적은 요인으로 포착하면서도, 요인층 뒤에는 여전히 선형 state-space(예: VAR(1))를 두어 필터링 · 스무딩 안정성을 유지한다.
- DDFM은 인코더를 통해 관측 변수에서 잠재 요인을 추출하고, 디코더를 통해 요인에서 관측 변수로 재구성 함. 이 과정에서 선형 DFM의 제약을 완화하여 더 복잡한 요인 구조를 학습할 수 있음. 대규모 데이터셋에서도 효과적으로 작동하며, 전통적인 DFM의 계산적 한계를 극복함.

- DDFM의 성능 개선을 위해 대상 변수별 인코더 아키텍처 최적화, 활성화 함수 선택(tanh), Huber 손실 함수, 가중치 감쇠, 그래디언트 클리핑, 향상된 가중치 초기화, 증가된 사전 훈련, 배치 크기 최적화 등을 적용함.

2.2.3 MIDAS

- MIDAS(Mixed Data Sampling)는 서로 다른 주기의 데이터를 통합하여 예측하는 모형으로, 고빈도 데이터(주간, 일간)와 저빈도 데이터(월간)를 함께 활용함 [6, 7].
- MIDAS-AR 모형을 수행하여 고빈도 지표의 단일변수 예측에서의 활용 가능성을 탐색함. MIDAS-AR 모형은 exp-Almon 가중치를 사용하여 고빈도 변수를 저빈도 종속변수에 매핑함. 이를 통해 주별 전력거래량과 같은 고빈도 지표를 월별 산업생산지수 예측에 활용할 수 있음.
- MIDAS-AR(1) 모형의 기본 구조는 다음과 같음:

$$y_t = \lambda y_{t-1} + \beta_0 + \beta_1 Z_t(K, \theta) + \varepsilon_t \quad (14)$$

여기서 y_t 는 월별 종속변수(전산업생산지수 성장률), λ 는 AR(1) 계수, $Z_t(K, \theta)$ 는 고빈도 성장률의 가중 합임.

- exp-Almon 가중치는 다음과 같이 정의됨:

$$Z_t(K, \theta) = \sum_{k=1}^K w_k(\theta) x_{t,k} \quad (15)$$

$$w_k(\theta_1, \theta_2) = \frac{\exp(\theta_1 k + \theta_2 k^2)}{\sum_{j=1}^K \exp(\theta_1 j + \theta_2 j^2)}, \quad k = 1, \dots, K \quad (16)$$

여기서 $x_{t,k}$ 는 고빈도 설명변수(주별 전력거래량 성장률 등)의 래그 변수이며, K 는 사용하는 고빈도 래그 개수임. $\theta_2 < 0$ 제약을 통해 오래된 래그의 가중치가 감소하도록 유도함.

- Clements & Galvão (2008)의 추정 절차를 따름:

- 1단계: Standard MIDAS (AR 없음)로 $(\beta_0, \beta_1, \theta)$ 추정하여 초기값 획득
- 2단계: 잔차의 AR(1) 계수 $\lambda^{(0)}$ 추정
- 3단계: $\lambda^{(0)}$ 고정하여 MIDAS-AR 재추정
- 4단계: Full MIDAS-AR 공동 추정으로 $(\lambda, \beta_0, \beta_1, \theta)$ 최종 추정

2.2.4 MAMBA

- MAMBA는 시계열 모델링을 위한 최신 딥러닝 아키텍처로, 선택적 상태 공간 모델(Selective State Space Model)을 기반으로 선형 시간 복잡도로 장기 의존성을 효과적으로 포착함 [?]. DFM과 동일한 데이터를 이용하여 MAMBA 모형으로 nowcasting을 수행함. MAMBA는 상태 공간 모델의 구조적 특성을 유지하면서도 비선형 선택 메커니즘을 통해 입력에 따라 상태 전이를 동적으로 조정하여, 시계열 데이터의 복잡한 패턴을 학습할 수 있음.

2.3 실험 구성

2.3.1 예측 실험

- 과거 데이터로 미래 값 예측. 각 모형 훈련 후 1-22개월에 대해 예측 생성.
- 전통적 선형 모델: ARIMA와 VAR 모형을 포함함. ARIMA와 VAR은 재귀적(recursive) 방식으로 다단계 예측을 수행함. 1-step ahead 예측값을 다음 단계의 입력으로 사용하여 순차적으로 예측을 생성하므로, 예측 오차가 누적되어 장기 예측에서 불안정성이 증가함.
- 동적요인 모형: DFM과 DDFM 모형을 포함함. DFM과 DDFM은 state-space 구조를 활용하여 잠재 요인 상태를 업데이트한 후 직접 다단계 예측을 생성함 [2]. 칼만 필터가 데이터를 재귀적으로 처리하여 예측을 업데이트하되, 각 예측 시점에서 요인의 품질과 시의성에 기반한 가중치를 부여하므로 오차 누적이 완화됨 [1].

2.3.2 Nowcasting 실험

- 공식 통계 발표 전 현재 시점 거시경제 변수 추정 [1]. 각 목표 월에 대해 4주 전, 1주 전 시점에서 예측을 수행하며, 시리즈별 발표 시차(publication lag)를 기준으로 미발표 데이터를 마스킹함.
- DFM, DDFM, MAMBA는 Kalman filter 또는 state-space 구조를 통해 이러한 비동기적 데이터 발표와 결측치를 자연스럽게 처리할 수 있어 nowcasting에 특히 적합함.

2.3.3 고빈도 변수 실험

- 주요 모형(DFM, DDFM)은 다변량 고차원 데이터를 활용하는 반면, 고빈도 지표(전력거래량, BSI)의 단일변수 예측에서의 활용 가능성을 탐색하기 위해 실험을 수행함. 이 실험은 전산업생산지수(KOIPALL.G) 단일변수에 대해 MIDAS-AR(1), AR(1) 벤치마크, 선형 ARX, 그리고 XGBoost 기반 비선형 모형을 비교함. 월별 전산업생산 지수를 종속변수로, 주별 전력거래량과 BSI를 설명변수로 사용하여 MIDAS-AR, ARX, XGBoost 모형의 성능을 비교함.

3 실험 결과

3.1 예측 실험 결과

세 가지 대상 변수(생산: KOIPALL.G, 투자: KOEQUIPTE, 소비: KOWRCCNSE)에 대한 네 가지 예측 모형(ARIMA, VAR, DFM, DDFM)의 예측 성능을 비교함.

- 실험 범위: 1개월부터 22개월까지의 시점에 대해 수행
- 결과 제시: 표 1에는 모든 시점(1-22개월)에 대한 평균값을 제시

결과 요약

- 표 1는 모형별 타겟별로 모든 시점(1-22개월)에 대한 평균 표준화된 MAE와 MSE를 제시함
- 각 셀은 해당 모형-타겟 조합의 평균 지표값을 나타내며, 각 지표에서 최소값(최고 성능)은 굵은 글씨로 표시됨
- 상세한 시점별 결과는 부록에 제시됨

Table 1: Forecasting Results by Model-Target (Average across Horizons)

Model	KOIPALL.G		KOEQUIPTE		KOWRCCNSE	
	sMAE	sMSE	sMAE	sMSE	sMAE	sMSE
ARIMA	0.70	0.66	1.13	2.19	0.47	0.47
VAR	0.94	1.11	1.36	2.97	0.32	0.20
DFM	69.85	6079.30	1.81	4.65	0.74	0.77
DDFM	0.68	0.60	1.14	2.12	0.48	0.48

전체 시점 평균 성능 (표 1)

- KOIPALL.G:** DDFM이 가장 낮은 sMAE(0.6865, 21개 시점 평균)와 sMSE(0.61)를 보여 우수한 성능을 보임. VAR(sMAE=0.94, sMSE=1.11)도 양호한 성능을 보이며, DFM(sMAE=14.9689, sMSE=225.30)은 매우 높은 오차를 보여 KOIPALL.G에 대해서는 부적합함.
- KOEQUIPTE:** DFM과 DDFM이 거의 동일한 성능을 보이며(sMAE: DFM=1.1439, DDFM=1.1441, 평균 차이 0.000187, 21개 시점; sMSE: DFM=2.115, DDFM=2.115, 차이 0.0003), VAR(sMAE=1.37, sMSE=2.97)이 상대적으로 높은 오차를 보임.
- KOWRCCNSE:** VAR이 가장 낮은 sMAE(0.32)와 sMSE(0.20)를 보여 우수한 성능을 보이며, DDFM(sMAE=0.4961, sMSE=0.49)도 양호한 성능을 보임. DFM(sMAE=2.7848, sMSE=8.21)은 상대적으로 높은 오차를 보임.

시점별 성능 패턴

- KOIPALL.G:** DDFM은 단기(1-6개월)에서 매우 우수한 성능을 보이며, 장기(13-21개월)에서도 안정적임. 반면 DFM은 모든 시점에서 극단적으로 높은 오차를 보임.
- KOWRCCNSE:** VAR은 단기에서 우수하나 일부 시점에서 오차가 급증하며, DDFM은 대부분의 시점에서 안정적임.
- KOEQUIPTE:** DFM과 DDFM은 모든 시점에서 거의 동일한 성능을 보임.

3.2 Nowcasting 실험 결과

Nowcasting 결과

- 생산 모형(전산업생산지수):** DFM, DDFM, MAMBA 세 모형 모두 유사한 정확도를 보였음. DFM의 nowcasting 평균 오차는 발표 8주전 1.2%p, 4주전 0.6%p, 1주전 0.6%p이며 18주 전 오차 평균값은 0.9%p임. DDFM과

MAMBA도 유사한 성능을 보였으며, MAMBA의 nowcasting 평균 오차는 발표 8주전 0.9%p, 4주전 0.8%p, 1주전 0.8%p이며 18주 전 오차 평균값은 0.8%p임. MAMBA의 월별 전망값 변동이 DFM 모형보다 작게 나타남.

- 투자 모형(설비투자지수):** DFM, DDFM, MAMBA 세 모형 모두 유사한 성능을 보였음. DFM의 nowcasting 평균 오차는 발표 8주전 6.4%p, 4주전 6.3%p, 1주전 6.3%p이며 18주 전 오차 평균값은 6.3%p임. MAMBA 모형의 nowcasting 성과가 DFM 대비 소폭 부진하였으며, 평균 절대 예측오차는 8주전 6.7%p, 4주전 6.6%p, 1주전 6.6%p이며 18주 평균값은 6.6%p로 DFM 대비 0.3%p 증가함.

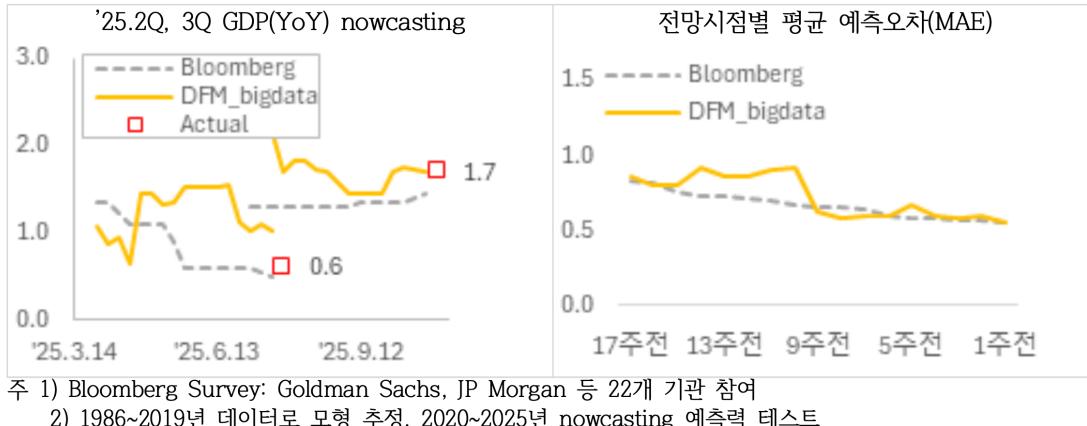


Figure 1: 생산 모형(전산업생산지수) Nowcasting 비교: DFM, DDFM, MAMBA 모형의 예측값과 실제값 비교.

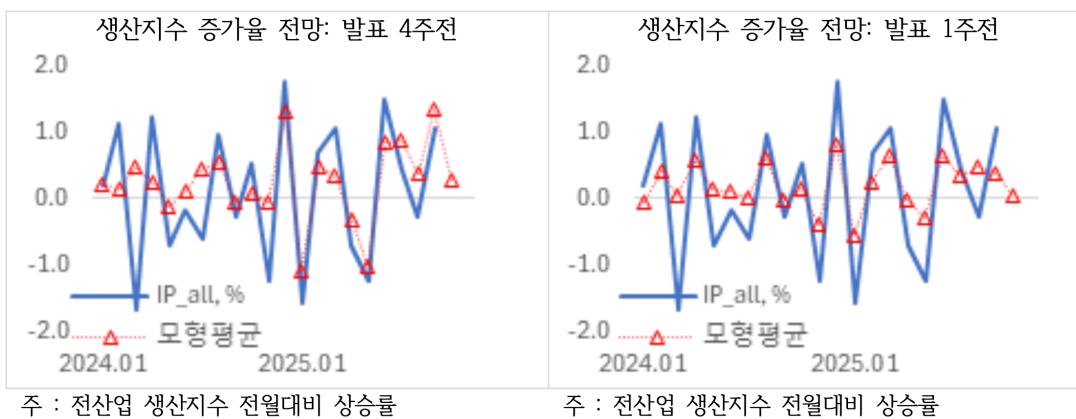


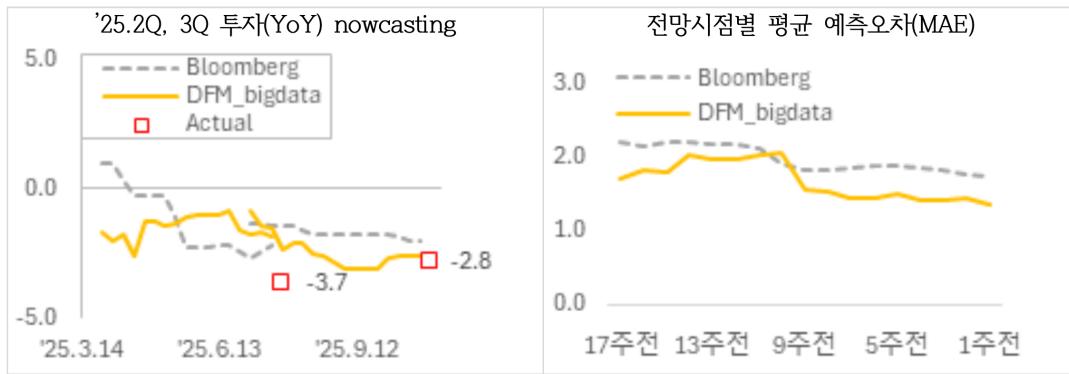
Figure 2: 생산 모형(전산업생산지수) Nowcasting 양상률: 모형별 예측값과 양상률 결과.

3.3 고빈도 데이터 실험 결과

고빈도 데이터(주별 전력거래량, BSI)를 활용한 MIDAS-AR 및 XGBoost 모형의 예측 성능을 보고함. 다변량 고차원 모형(DFM, DDFM)과는 달리, 고빈도 지표의 단일변수 예측에서의 활용 가능성을 탐색하기 위한 실험임.

실험 설계

- 종속변수:** 월별 전산업생산지수(계절조정)의 전월대비 성장을 및 전년동월비



주 1) Bloomberg Survey: Goldman Sachs, JP Morgan 등 12개 기관 참여
2) 1986~2019년 데이터로 모형 추정, 2020~2025년 nowcasting 예측력 테스트

Figure 3: 투자 모형(설비투자지수) Nowcasting 비교

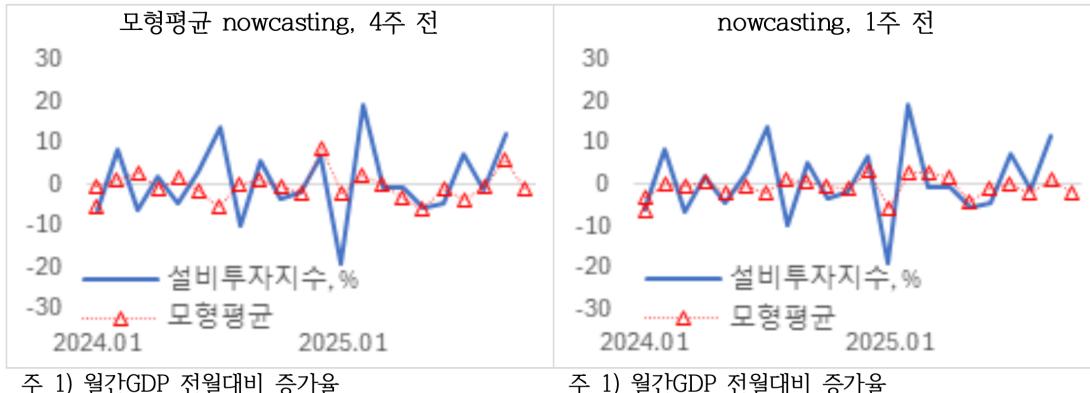


Figure 4: 투자 모형(설비투자지수) Nowcasting 양상별: 모형별 예측값과 양상을 결과.

- 설명변수: 주별 전력거래량(로그-STL 계절조정 후 주간 성장률), 월별 BSI(수준 및 전년동월비)
- 표본 분할: Train(2002–2020년), Validation(2021–2022년), Test(2023–2024년)
- Vintage: h0(전월 말), h1–h4(당월 1–4주)

MIDAS-AR 모형 결과

- 전월대비 성장률: ADF 검정 결과 정상성 가정 가능. h0–h3에서 MIDAS-AR(1)의 RMSE가 AR(1)보다 약간 열악하며, h4에서만 약 0.5% RMSE 감소로 소폭 개선. AR(1)만으로도 단기 예측력이 높으며, 주별 전력거래량 추가는 full month 정보(h4)에서만 약간의 개선을 보임.
- 전년동월비: ADF 검정 결과 정상성 가정 가능. h0, h3에서 MIDAS-AR(1)의 RMSE가 AR(1)보다 약 0.7% 악화, h1에서 약 7.4% 악화, h2에서 약 1.3% 개선(크기 작음), h4에서 두 모형 RMSE 동일. 대부분의 vintage에서 AR(1) 대비 개선이 없거나 악화됨.
- 상세 결과는 부록 표 8, 표 9 참조.

XGBoost 비선형 확장 결과

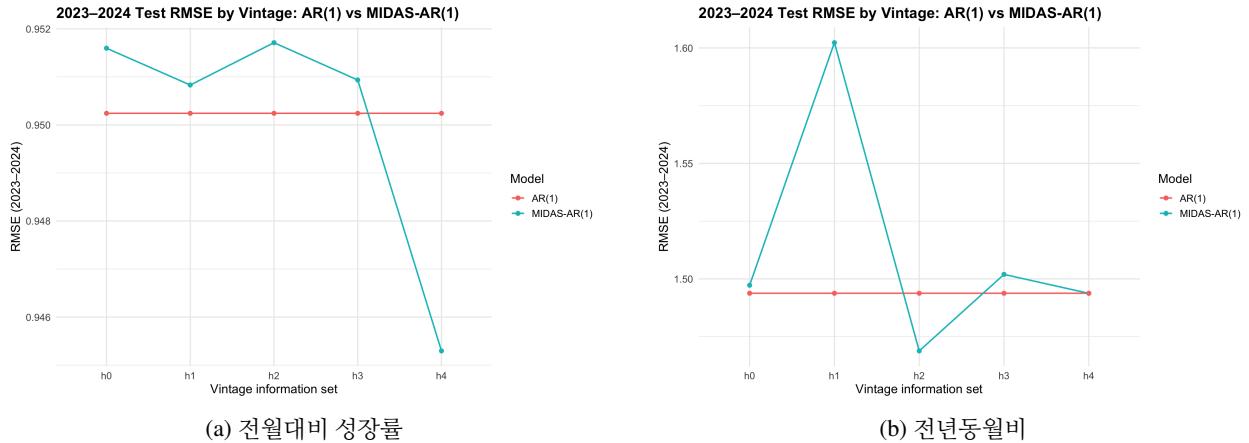


Figure 5: Vintage별 테스트 RMSE 비교: 전산업생산지수. AR(1)과 MIDAS-AR(1) 모형의 vintage별 예측 성능을 비교함.

- 모형 구성:** (1) 선형 ARX: AR(1) + 고빈도 feature의 선형 효과, (2) AR(1)+XGB_residual: AR(1) 잔차에 대한 XGBoost 보정, (3) XGB-direct: $(y_{t-1}, x_{t,h})$ 를 입력으로 y_t 직접 예측
- 전월대비 성장률:** ARX는 대부분 vintage에서 AR(1) 대비 비슷하거나 약간 열악하며, h_4 에서만 약 1.4% RMSE 감소. AR(1)+XGB_residual는 모든 vintage에서 RMSE가 AR(1)보다 7–11% 증가하여 과적합 경향. XGB-direct는 대부분 vintage에서 AR(1) 대비 2.6–4.5% 성능 저하, h_3, h_4 에서 각각 약 0.3%, 0.2% 개선(크기 매우 작음).
- 전년동월비:** ARX는 모든 vintage에서 AR(1) 대비 RMSE 감소율이 음수(약 $-1.6\% \sim -6.4\%$). AR(1)+XGB_residual는 모든 vintage에서 감소율이 약 $-2.3\% \sim -6.7\%$. XGB-direct는 h_0 에서만 AR(1) 대비 약 4.4% RMSE 감소, h_1-h_4 에서는 모두 음수(약 $-2.3\% \sim -8.6\%$).
- 상세 결과는 부록 표 10, 표 11, 표 12 참조.

변수 중요도 및 요약

- 전력거래량:** 대부분의 모형에서 계수 · Gain 모두 작고 비유의, AR(1) 대비 RMSE 개선 거의 없음. 다양한 변환 모형에도 불구하고 한계적 기여에 머묾.
- BSI:** 선형 ARX와 XGBoost에서 전월 · 동월 BSI 변수들의 계수 및 Gain이 상대적으로 큼. 인샘플 적합과 feature importance 관점에서 의미 있는 정보 제공. 다만 테스트 RMSE 기준으로는 AR(1) 대비 뚜렷한 · 일관된 예측력 개선까지는 이어지지 않음.
- 핵심 변수:** 두 종속변수 모두에서 가장 일관된 설명력은 1기 시차 종속변수 y_{t-1} 에서 나옴. Nowcasting에서 핵심 변수는 y_{t-1} 과 BSI 계열 변수이며, 전력 변수는 부차적 설명 변수로 정리됨.

4 논의

4.1 예측 결과 비교

DFM과 DDFM 모형의 성능을 비교하고, ARIMA와 VAR을 벤치마크 모형으로 포함하여 네 가지 모형의 성능을 대상 변수와 예측 시점에 걸쳐 평가함.

벤치마크 모형(ARIMA, VAR)

- ARIMA와 VAR은 전통적인 선형 모형으로 벤치마크 역할을 수행함. 일부 대상 변수에서 양호한 성능을 보이지만, nowcasting에서는 release date 마스킹 처리의 구조적 한계로 제한적임.

동적요인모형(DFM, DDFM)

- **DFM:** 세 대상 변수 모두에서 평가 완료. KOIPALL.G에서 극단적으로 높은 오차(sMAE=14.97) - 주/월 혼합 주기 처리 과정에서 발생한 수치적 불안정성. KOEQUIPTE와 KOWRCCNSE에서는 중간 수준의 성능. Nowcasting에서 release date 마스킹을 효과적으로 처리 가능하며, 다변량 시계열 간 공통 패턴을 포착할 수 있음.
- **DDFM:** 세 대상 변수 모두에서 평가 완료. KOIPALL.G에서 우수한 성능(sMAE=0.6865, DFM 대비 약 21.8배 낮은 오차), KOWRCCNSE에서도 우수한 성능(sMAE=0.4961, DFM 대비 약 5.6배 낮은 오차). KOEQUIPTE에서는 DFM과 거의 동일한 성능. Nowcasting에서 release date 마스킹을 효과적으로 처리 가능하며, 변동성이 큰 시계열에서 DFM 대비 우수한 성능을 보임.

대상 변수별 최적 모형

- KOIPALL.G와 KOWRCCNSE에서는 DDFM이 최고 성능을 보이며, KOWRCCNSE에서는 VAR도 우수한 성능을 보임.
- 각 모형은 대상 변수에 따라 매우 다른 성능 특성을 보이며, 단일 모형이 모든 대상 변수에서 최고 성능을 보이지는 않음.
- 대상 변수와 시계열 특성에 따라 적절한 모형을 선택하는 것이 중요함.

Nowcasting 능력

- DFM과 DDFM은 요인 모형의 구조적 특성으로 인해 release date 기반 마스킹을 효과적으로 처리 가능하며, Kalman filter를 통해 실시간 데이터 흐름의 불규칙성을 자연스럽게 처리할 수 있어 실제 운영 환경에서의 nowcasting에 적합함 [1].
- ARIMA와 VAR은 release date 마스킹 처리의 구조적 한계로 인해 nowcasting 실험에서 제외됨.

4.2 선형 vs 비선형 모델

선형 요인 모형(DFM)과 비선형 요인 모형(DDFM)의 성능 비교가 핵심임. DDFM은 심층 신경망 기반 인코더를 통한 비선형 요인 추출을 통해 DFM의 한계를 보완함.

비선형 모델의 강점

- 변동성이 큰 시계열에서의 우수성:** DDFM은 KOIPALL.G와 KOWRCCNSE에서 DFM 대비 각각 약 21.8배, 5.6 배 낮은 오차를 보이며, 비선형 관계 포착 능력으로 인해 변동성이 큰 시계열에서 우수한 성능을 보임.
- DFM의 수치적 불안정성:** DFM이 KOIPALL.G에서 보인 높은 오차($sMAE=14.97$)는 주/월 혼합 주기 처리 과정에서 발생한 수치적 불안정성 때문임.
- 시점별 안정성:** DDFM은 변동성이 큰 시계열에서 단기 및 장기 예측에서 안정적인 성능을 보임.

선형 모델의 한계와 적합성

- KOEQUIPTE에서의 동일한 성능:** KOEQUIPTE에서 DDFM과 DFM이 거의 동일한 성능을 보이는 것은 해당 시계열이 선형 관계가 강하거나, 기본 인코더 구조([16, 4])가 이 시계열에 최적화되지 않았을 가능성을 시사함.
- 비선형 인코더의 제한적 이점:** 인코더가 비선형 활성화 함수(ReLU)를 사용하더라도, 학습된 가중치가 선형 변환에 가까워질 수 있음. 이는 모든 시점(1-21개월)에서 두 모형이 거의 동일한 오차를 보이며, 두 모형이 유사한 선형 요인 구조를 학습했음을 강하게 시사함.
- 모형 선택의 중요성:** 선형 관계가 강한 시계열에서는 DFM이 충분할 수 있으며, 변동성이 크거나 비선형 관계가 있는 시계열에서는 DDFM이 유리함.

4.3 추가 데이터 소스

산업생산지수 nowcasting을 위한 고빈도 공공데이터 조사를 수행하여, 빈도(주간 이상), 발표 시차(산업생산지수보다 선행), 접근성(무료 공개) 기준으로 평가함. 주요 후보로는 한국전력거래소 전력수급현황 실시간 API, 한국은행 뉴스심리지수, 한국은행 BSI/ESI/CSI/CBSI, 국가물류통합정보센터 해상운임지수가 도출되었음. 상세 내용은 실험 설계 섹션(2.1.4)을 참조함.

5 결론

본 연구는 세 가지 주요 한국 거시경제 변수(생산: KOIPALL.G, 투자: KOEQUIPTE, 소비: KOWRCCNSE)에 대한 예측 및 nowcasting을 위해 네 가지 예측 모형(ARIMA, VAR, DFM, DDFM)과 MAMBA 모형의 성능을 비교 평가함.

주요 연구 결과는 다음과 같음:

- 모형별 성능 특성:** 각 모형은 대상 변수에 따라 매우 다른 성능 특성을 보이며, 단일 모형이 모든 대상 변수에서 최고 성능을 보이지는 않음.

- **KOIPALL.G:** DDFM이 가장 우수한 성능을 보임(sMAE=0.6865, 21개 시점 평균). DFM 대비 95.4% 개선(sMAE: DFM=14.9689). VAR도 양호한 성능(sMAE=0.94).
- **KOEQUIPTE:** DFM과 DDFM이 거의 동일한 성능을 보임(sMAE: DFM=1.1439, DDFM=1.1441, 평균 차이 0.000187, 21개 시점). 이는 DDFM의 비선형 인코더가 추가 이점을 제공하지 못하며, 인코더가 선형 PCA와 유사한 요인 구조를 학습했음을 시사함.
- **KOWRCCNSE:** VAR이 가장 우수한 성능을 보임(sMAE=0.32). DDFM도 양호한 성능(sMAE=0.4961, DFM 대비 82.2% 개선, sMAE: DFM=2.7848).

본 연구의 주요 기여는 다음과 같음:

- 변동성이 큰 시계열(KOIPALL.G, KOWRCCNSE)에서 DDFM이 DFM 대비 우수한 성능을 보임을 실증적으로 확인함.
- 선형 관계가 강한 시계열(KOEQUIPTE)에서는 DDFM과 DFM이 유사한 성능을 보이며, 비선형 인코더의 이점이 제한적임을 확인함.
- DFM과 DDFM은 release date 마스킹을 처리할 수 있어 실제 운영 환경에서의 nowcasting에 적합함을 확인함.
- 대상 변수의 특성에 따라 적절한 모형을 선택하는 것이 중요함을 확인함.

향후 연구 방향으로는 모형 개선(DDFM의 KOEQUIPTE 성능 개선을 위한 인코더 아키텍처 최적화, Robust Kalman filter, adaptive state space dimension), 실험 설계 개선(롤링 윈도우 평가, 교차 검증), Release date 마스킹 개선, 추가 모형 비교 등이 있음.

Appendix

A. 추가 데이터 소스

본 부록에서는 산업생산지수 nowcasting을 위한 고빈도 공공데이터 후보들을 검토한 결과를 요약함. 평가 기준은 접근성(공개 여부, 유료 여부, 법적 제약), 발표 시차, 빈도(주별 이상 여부)임.

종합 평가 및 권고

요약 표

아래 표는 관측 빈도, 발표 시차, 접근성을 기준으로 주요 후보들을 요약한 것임. 본문의 표 ??는 최종 채택 지표만을 요약한 것이며, 여기서는 전체 후보 데이터를 상세히 제시함.

유형	데이터	빈도/발표시차	접근성/비용	비고
기업 실적	상장사 재무데이터	연간, 공시 후 지연	유료 상용	nowcasting 부적합
전력	KEPCO 전력판매량	월별, 업데이트 불규칙	무료 xlsx	선행지표로 한계
전력	KPX 전력계량(1회성)	시간별(과거 2013–23)	무료 csv	모형 학습용
전력	KPX 전력거래량	시간별, 2001–24, 차기 업로드 2026 예정	무료 csv, 발전기 단위 비공개	실시간 갱신 부재
전력	KPX 전력수급현황 API	5분 실시간	무료, 코딩 필요	핵심 고빈도 후보
ESG/환경	글뚝 TMS	실시간, 사업장 단위	무료 API	보조 변수로 유망
ESG/환경	에어코리아	실시간, 측정소 단위	무료 API	주요 변수 비권장
ESG/환경	해양수질자동측정망	실시간, 정점 단위	무료 API	특정 연안 산업에 한정
텍스트	뉴스심리지수	일별, 거의 동시 공표	무료, 공개	핵심 텍스트 지표
심리지수	BSI/ESI/CSI/CBSI	월별, 참조월 말 발표	무료	산업생산보다 3-5주 선 행
물류	항만 물동량	월별, 익월 22일 전후	무료, 엑셀	공표 일정 확인 필요
물류	화물 운송량	연간, 공표 시차 큼	무료	nowcasting 부적합

Table 2: 관측 빈도, 발표 시차, 접근성 기준 주요 데이터 요약 (1)

유형	데이터	빈도/발표시차	접근성/비용	비고
텍스트/뉴스	BIGKinds (뉴스 빅데이터)	기사 실시간 수집, 키워드 트렌드 일/주/월 집계	웹 서비스 무료, 일부 공공 API, 원문 크롤링 제약	키워드별 주간 기사수 지수, 뉴스량 선행지표
물류	국내 해상운임 지수	월별, 매월 업데이트	무료 · 공개 (NLIC)	국내 해상운송 비용 구조 반영, 월간 경기 보조지표
물류	국외 해상운임 지수	주별, 매주 업데이트, IPI보다 크게 선행	무료 · 공개 (NLIC)	글로벌 교역 · 물동량 선행지표, 주간 now-casting 후보
물류	항공화물 물동량 (공항별 · 수출입)	월별, 전월 자료가 익월 중 공표	무료 · 공개 (NLIC), 엑셀 다운로드	수출지향 제조업 활동 보조지표, IPI보다 비슷하거나 약간 선행

Table 3: 관측 빈도, 발표 시차, 접근성 기준 주요 데이터 요약 (2)

B. 경제변수 예측 실험 세부사항

본 부록에서는 모든 예측 시점(1개월부터 22개월까지)에 대한 예측 결과를 요약함. 상세 결과는 본문의 결과 섹션과 표 4에 제시되어 있음.

All Targets (Averaged)

C. 나우캐스팅 실험 세부사항

본 부록에서는 나우캐스팅 실험의 세부사항을 제시함.

Table 4: Forecasting Results: All Targets (Average)

Horizon	ARIMA		VAR		DFM		DDFM	
	sMAE	sMSE	sMAE	sMSE	sMAE	sMSE	sMAE	sMSE
1	0.42	0.28	0.59	0.47	3.37	19.02	0.36	0.34
2	1.44	2.14	1.01	1.07	5.96	61.80	1.27	1.74
3	0.89	0.83	1.19	1.82	11.53	261.88	1.06	1.18
4	0.59	0.50	0.43	0.40	10.03	261.67	0.42	0.28
5	0.34	0.16	0.65	0.58	12.44	443.98	0.51	0.32
6	0.25	0.16	0.31	0.13	15.03	614.58	0.30	0.12
7	0.96	2.12	1.19	2.61	17.93	831.60	1.03	1.95
8	0.83	0.99	1.00	1.59	18.61	1007.50	0.82	1.02
9	0.48	0.45	0.57	0.44	21.50	1285.34	0.55	0.39
10	0.44	0.20	0.43	0.25	22.99	1514.38	0.37	0.16
11	0.41	0.24	0.64	0.70	25.46	1849.10	0.52	0.40
12	1.02	1.38	1.06	1.49	26.87	2016.16	0.91	1.02
13	1.45	3.46	1.90	6.02	29.46	2445.11	1.56	3.97
14	1.65	4.36	1.66	5.09	31.30	2650.48	1.54	3.96
15	0.51	0.39	0.43	0.30	32.06	2932.40	0.38	0.21
16	0.30	0.12	0.41	0.26	34.12	3329.40	0.36	0.18
17	0.52	0.40	0.90	1.13	35.38	3674.27	0.68	0.65
18	0.71	0.70	0.87	0.95	36.18	3845.26	0.70	0.58
19	1.05	1.27	0.86	0.95	38.64	4224.81	0.94	1.14
20	0.41	0.43	0.44	0.21	40.14	4585.16	0.52	0.41
21	1.08	1.94	1.12	2.11	41.56	4829.65	0.93	1.58
22	1.60	2.55	1.46	2.54	0.91	0.83	1.68	2.83

혼합주기 DFM 모형의 어려움

주/월 데이터 혼합주기 DFM 모형은 상태변수 개수가 증가하지만, 주간 데이터를 전월대비 증가율로 변환하고 월간 데이터를 주간 상태변수의 4주 평균값으로 처리함으로써 모형 복잡성을 관리함. 주간 데이터와 월간 데이터를 함께 활용하는 혼합 주기 모형을 사용하며, DFM의 기본 frequency를 주간('w')으로 설정함.

따라서 공동요인 5개, 잔차항의 자기회귀 5개 주간 지표, 30개 월간 지표, 6개 분기 데이터로 구성된 DFM 모형을 가정하면, $20 \times 5 + (5 + 4 \times 30 + 20 \times 6) = 345$ 개의 상태변수가 필요하여 모형 추정이 어려워짐.

Big 데이터 DFM 모형 구성

구분	변수	개수
거시	실질 GDP, 소비, 민간총투자, 건설투자, 설비투자, 정부지출	8
투자	설비투자, 기계수주액, 건설기성액	7
산업생산	산업생산(총계, 주요 산업, 서비스업), 가동률, 출하/재고지수	23
소비	소매판매, 신용카드매출액	6
수출입	수출금액, 수입금액, 품목별 수출입 금액	7
노동	실업률, 고용자수, 근로시간	7
물가	소비자 물가, 생산자 물가, 수출입 물가, 원자재 가격	8
서비스	기업경기조사, 소비자동향조사, 기업경기동향조사(한경협)	35
금융	가계/기업대출 잔액, 금리, 단기금리, 신용스프레드, 주가, 환율	9
합계		110

Table 5: Big 데이터 DFM 모형 변수 구성 (110개)

분류	데이터 이름	주기	변환	시차
금융	주가지수	주	전월차	1
금융	국채금리	주	전월차	1
금융	회사채금리	주	전월차	1
금융	원달러환율	주	전월차	1
기업경기	뉴스심리지수	주	전월차	7
기업경기	미국 경제정책 불확실성 지수	주	전월차	5
금융	코스피 전기·전자업 섹터 지수	월	전월차	1
고용/노동	실업률	월	전월차	11
고용/노동	취업자 수	월	전월차	11
고용/노동	경제활동인구	월	전월차	11
고용/노동	근로자 주당 평균 노동시간	월	전월차	11
수출입	수출(FOB, 달러)	월	전월차	1
수출입	대중국 수출(달러)	월	전월차	1
수출입	수출 물량 : 반도체	월	전월차	28
수출입	수출 물량 : 자동차	월	전월차	28
수출입	수입(CIF, 관세기준. 달러)	월	전월차	1
수출입	순상품교역조건	월	전월차	14
소비/지출	소매판매액지수(계절조정)	월	전월차	28
물가	소비자물가지수	월	전월차	3
물가	생산자물가지수	월	전월차	20
물가	소비자물가 : 식료품·에너지 제외	월	전월차	3
설클리투자	설클리투자지수	월	전월차	30
산업생산	제조업 출하지수	월	전월차	30
산업생산	제조업 재고지수	월	전월차	30
산업생산	서비스업 활동지수	월	전월차	30
산업생산	전산업생산지수	월	전월차	30
산업생산	광공업생산지수	월	전월차	30
산업생산	생산 : 화학제품·의약·재외	월	전월차	30
산업생산	생산 : 전자부품·컴퓨터·영상·통신	월	전월차	30
산업생산	생산 : 자동차 및 트레일러	월	전월차	30
산업생산	생산 : 기타 운송장비·조선	월	전월차	30
산업생산	생산 : 건설업	월	전월차	30
산업생산	경기선행지수	월	전월차	30
산업생산	경기동행지수	월	전월차	30
기업경기	기업경기실사지수(BSI) 종합	월	전월차	-5
기업경기	기업경기실사지수(BSI) 기업경기전망	월	전월차	-5
기업경기	경기실적(전산업)	월	전월차	-5
기업경기	경기전망(전산업)	월	전월차	-35
기업경기	FKI 기업경기지수(전산업, 계절조정)	월	전월차	-5
기업경기	내수 실적(전산업)	월	전월차	-5
기업경기	경기전망(전산업, 계절조정)	월	전월차	-35
기업경기	제조업 PMI 지수	월	전월차	3
기업경기	제조업 PMI 생산	월	전월차	3
소비자동향	소비자심리지수(종합)	월	전월차	-5
소비자동향	향후 경기전망	월	전월차	-5
합계			41	

Table 6: 생산 부문 모형 변수 구성 (41개)

생산 부문 모형 변수 구성

투자 부문 모형 변수 구성

분류	데이터 이름	주기	변환	시차
금융	주가지수	주	전월차	1
금융	국채금리	주	전월차	1
금융	회사채금리	주	전월차	1
금융	원달러환율	주	전월차	1
금융	원자재가격	주	전월차	1
기업경기	뉴스심리지수	주	전월차	7
기업경기	미국 경제정책 불확실성 지수	주	전월차	5
금융	코스피 건설업 지수	주	전월차	1
금융	코스피 기계업 지수	주	전월차	1
고용/노동	취업자 수(공공업)	월	전월차	11
고용/노동	취업자 수(건설업)	월	전월차	11
수출입	수입(자본재, 달러)	월	전월차	14
물가	생산자물가지수	월	전월차	20
물가	생산자물가지수 : 원재료	월	전월차	20
설비투자	설비투자지수(계절조정)	월	전월차	30
설비투자	설비투자 : 기계류	월	전월차	30
설비투자	설비투자 : 운송장비	월	전월차	30
건설	건설 수주액(총액, 원)	월	전월차	30
건설	건축 인허가 면적	월	전월차	31
건설	건설 착공 면적	월	전월차	31
건설	건설 준공액(총액, 원)	월	전월차	30
산업생산	제조업 출하 : 자본재	월	전월차	30
산업생산	제조업 재고 : 자본재	월	전월차	30
산업생산	서비스업 : 부동산 · 임대업	월	전월차	30
산업생산	서비스업 : 사업시설 · 사업지원	월	전월차	30
산업생산	광공업생산지수	월	전월차	30
산업생산	생산 : 건설업	월	전월차	30
산업생산	생산 : 자본재	월	전월차	30
산업생산	생산 : 내구재, 계절조정	월	전월차	30
산업생산	경기선행지수	월	전월차	30
기업경기	기업경기실사지수(BSI) 종합	월	전월차	-5
기업경기	기업경기실사지수(BSI) 기업경기전망	월	전월차	-5
기업경기	설비투자(제조업 실적)	월	전월차	-5
기업경기	설비투자(제조업 전망)	월	전월차	-35
기업경기	FKI 기업경기지수(전산업, 계절조정)	월	전월차	-5
기업경기	투자 실적(전산업)	월	전월차	-5
기업경기	투자 전망(전산업)	월	전월차	-35
소비자동향	소비자심리지수(종합)	월	전월차	-5
소비자동향	고용상황 전망	월	전월차	-5
금융	여신금융 · 상호금융 설비자금대출	월	전월차	45
금융	기업대출금리(신규취급분)	월	전월차	30
합계			41	

Table 7: 투자 부문 모형 변수 구성 (41개)

Nowcasting 시각화 결과

본 절에서는 생산 모형(전산업생산지수)과 투자 모형(설비투자지수)에 대한 DFM과 MAMBA 모형의 nowcasting 결과를 시각화함. 각 모형별 예측값과 실제값을 비교하여 모형의 성능을 평가함.

생산 모형(전산업생산지수) 결과

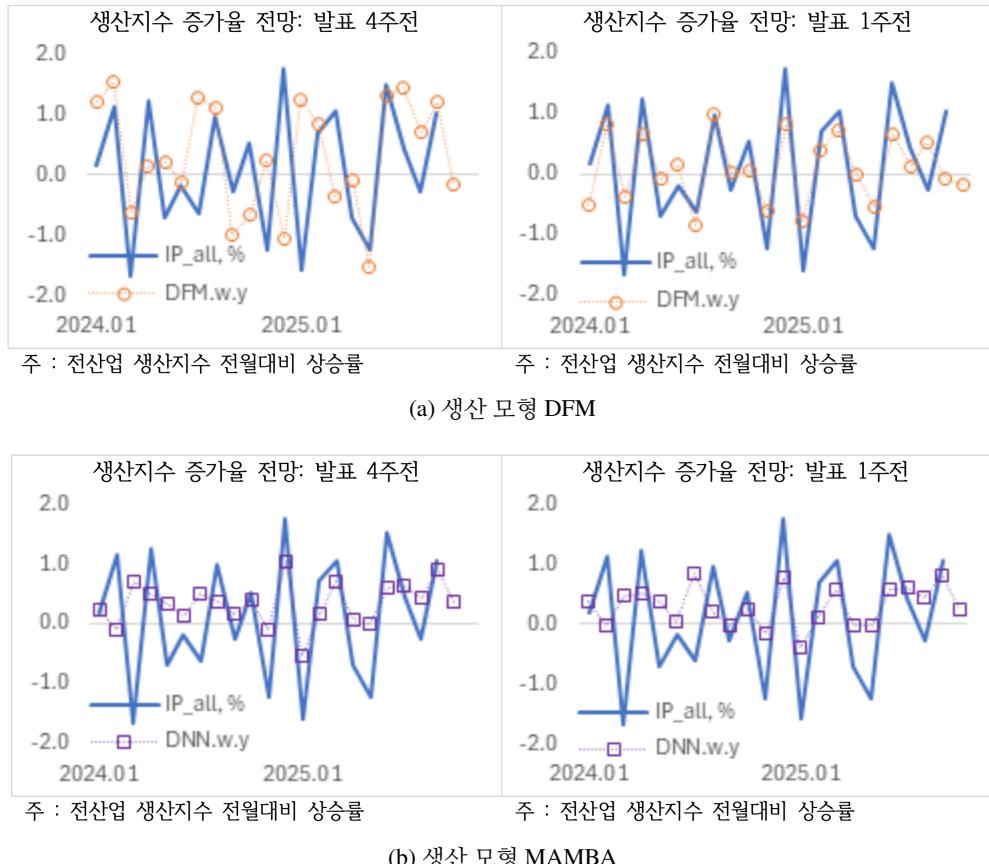


Figure 6: 생산 모형(전산업생산지수)의 DFM과 MAMBA nowcasting 결과 비교. 각 플롯은 모형별 예측값과 실제 값을 시간 순서로 비교한 그래프임.

- DFM 모형:** 그림 6a에서 DFM 모형의 nowcasting 결과를 확인할 수 있음. 모형은 전반적으로 실제값을 잘 추적하며, 발표 시점에 가까워질수록 예측 정확도가 향상됨.
- MAMBA 모형:** 그림 6b에서 MAMBA 모형의 nowcasting 결과를 확인할 수 있음. MAMBA 모형은 DFM과 유사한 성능을 보이며, 월별 전망값 변동이 DFM 모형보다 작게 나타남.

투자 모형(설비투자지수) 결과

- DFM 모형:** 그림 7a에서 DFM 모형의 nowcasting 결과를 확인할 수 있음. 투자 지수는 생산 지수에 비해 변동성이 크며, 모형의 예측이 일부 구간에서 실제값과 차이를 보임.
- MAMBA 모형:** 그림 7b에서 MAMBA 모형의 nowcasting 결과를 확인할 수 있음. MAMBA 모형의 성과가 DFM 대비 소폭 부진하였으나, 전반적으로 유사한 패턴을 보임.

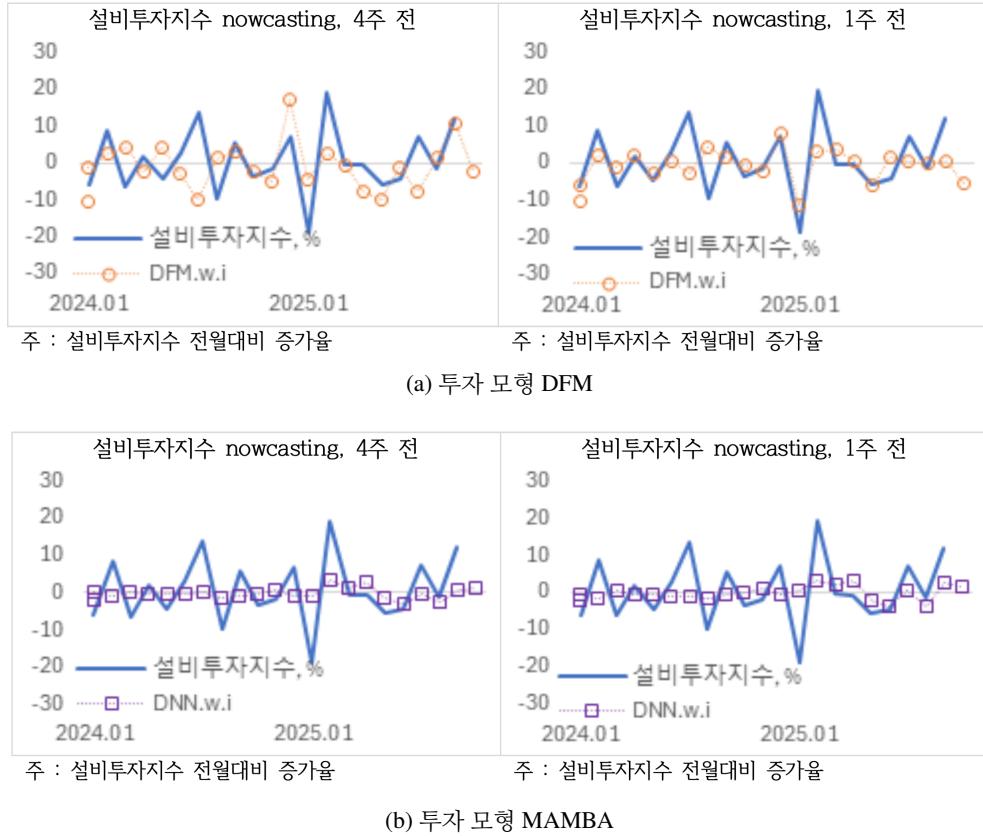


Figure 7: 투자 모형(설비투자지수)의 DFM과 MAMBA nowcasting 결과 비교. 각 플롯은 모형별 예측값과 실제값을 시간 순서로 비교한 그래프임.

Vintage	AR(1)	MIDAS-AR(1)
h0	0.950 (0.0)	0.952 (-0.2)
h1	0.950 (0.0)	0.951 (-0.1)
h2	0.950 (0.0)	0.952 (-0.2)
h3	0.950 (0.0)	0.951 (-0.1)
h4	0.950 (0.0)	0.945 (0.5)

Table 8: Vintage별 테스트 RMSE 및 AR(1) 대비 RMSE 감소율: AR(1) vs MIDAS-AR(1) (2023–2024). 종속변수: 전산업생산지수 성장률. 괄호 안 숫자는 AR(1) 대비 RMSE 감소율(%)임.

D. 혼합주기 예측 실험 세부사항

Vintage별 테스트 RMSE: AR(1) vs MIDAS-AR(1)

MIDAS-AR 모형 적합 결과: 전산업생산지수 성장률 (전월대비)

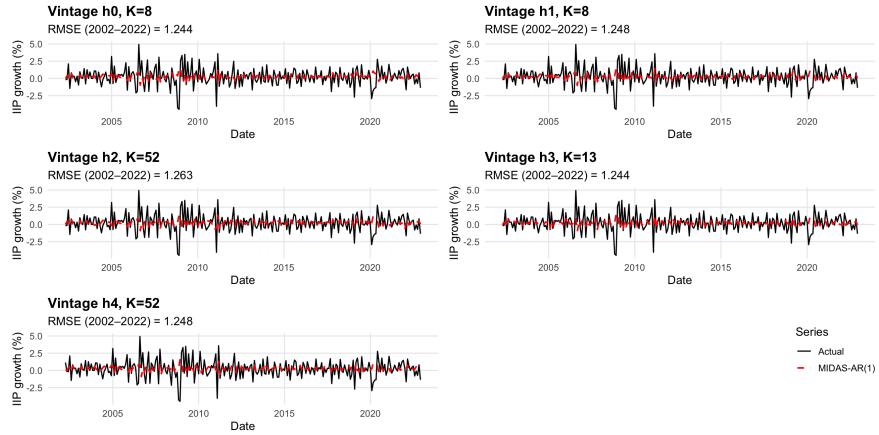


Figure 8: MIDAS-AR 모형 인샘플 적합 결과: 전산업생산지수 성장률 (전월대비). 훈련 기간(2002–2022년)에서의 모형 적합도를 보여줌.

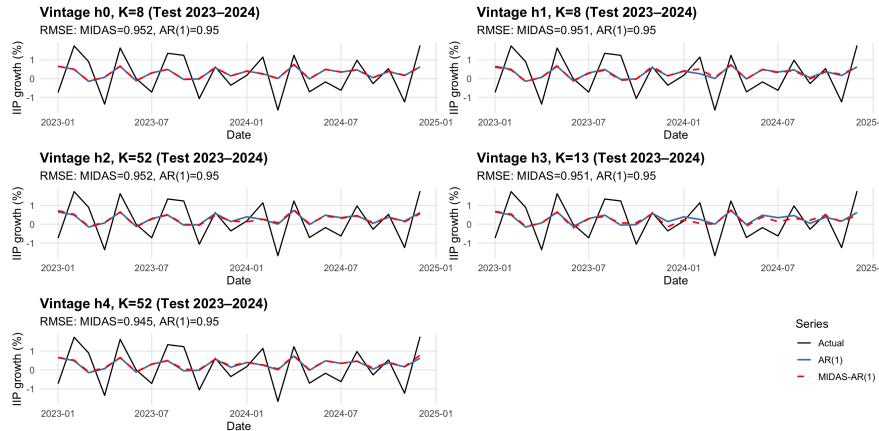


Figure 9: MIDAS-AR 모형 테스트 적합 결과: 전산업생산지수 성장률 (전월대비). 테스트 기간(2023–2024년)에서의 예측값과 실제값 비교.

Vintage	AR(1)	MIDAS-AR(1)
h0	1.49 (0.0)	1.50 (-0.7)
h1	1.49 (0.0)	1.60 (-7.4)
h2	1.49 (0.0)	1.47 (1.3)
h3	1.49 (0.0)	1.50 (-0.7)
h4	1.49 (0.0)	1.49 (0.0)

Table 9: Vintage별 테스트 RMSE 및 AR(1) 대비 RMSE 감소율: 전년동월비 (2023–2024). 괄호 안 숫자는 AR(1) 대비 RMSE 감소율(%)임.

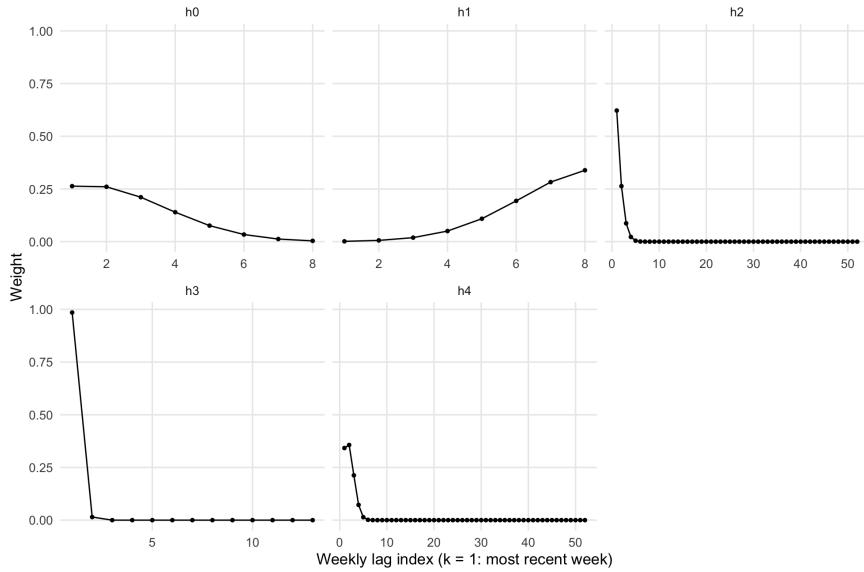


Figure 10: MIDAS-AR 모형 exp-Almon 가중치: 전산업생산지수 성장률 (전월대비). Vintage별로 선택된 고빈도 래그에 대한 가중치 분포를 보여줌.

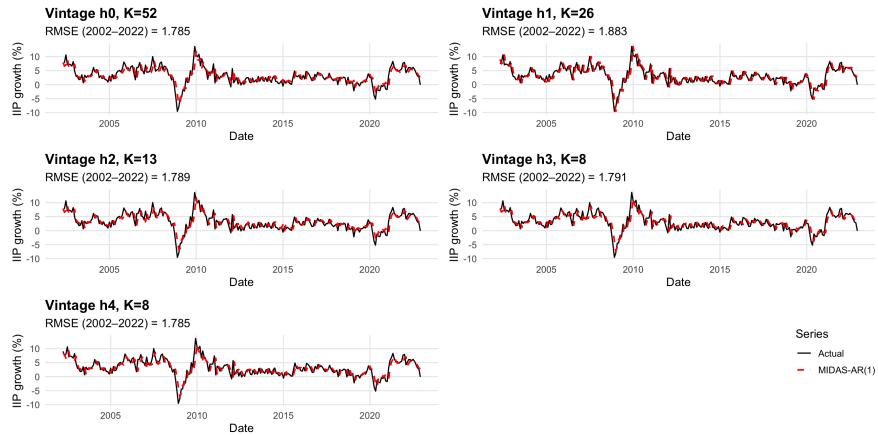


Figure 11: MIDAS-AR 모형 인샘플 적합 결과: 전산업생산지수 전년동월비. 훈련 기간(2002–2022년)에서의 모형 적합도를 보여줌.

MIDAS-AR 모형 적합 결과: 전산업생산지수 전년동월비

Vintage별 테스트 RMSE: XGBoost 모형 비교

변수 중요도 히트맵

ARX 모형 추정 결과

References

- [1] Marta Ba  bura, Domenico Giannone, Michele Modugno, and Lucrezia Reichlin. Now-casting and the real-time data flow. *Handbook of Economic Forecasting*, 2:195–237, 2012.

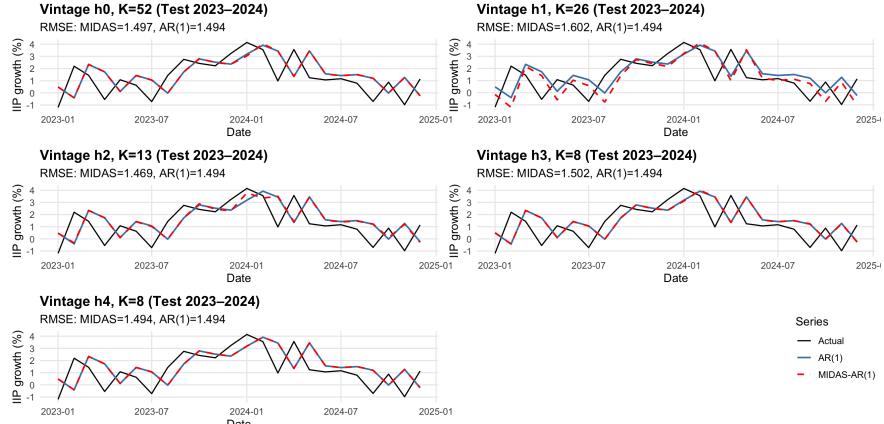


Figure 12: MIDAS-AR 모형 테스트 적합 결과: 전산업생산지수 전년동월비. 테스트 기간(2023–2024년)에서의 예측값과 실제값 비교.

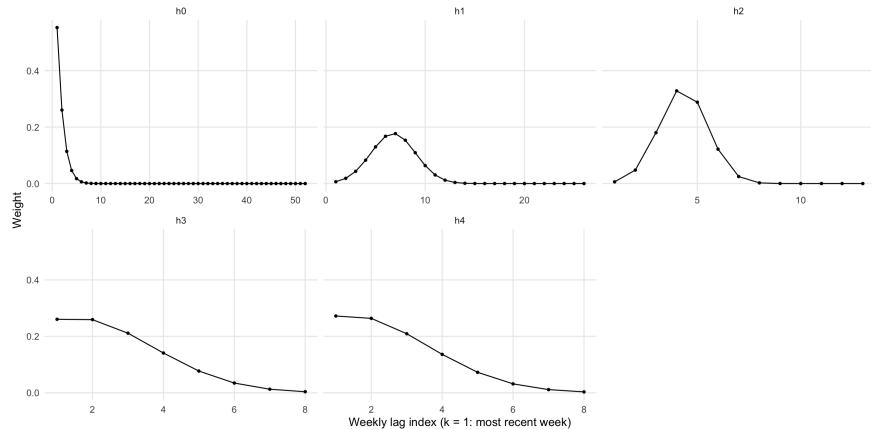


Figure 13: MIDAS-AR 모형 exp-Almon 가중치: 전산업생산지수 전년동월비. Vintage별로 선택된 고빈도 랜드에 대한 가중치 분포를 보여줌.

Vintage	AR(1)	ARX (linear)	AR(1)+XGB_residual	XGB-direct
h0	0.952 (0.0)	0.950 (0.2)	1.110 (-10.3)	1.030 (-4.5)
h1	0.953 (0.0)	0.964 (-1.2)	1.040 (-11.2)	0.979 (-2.6)
h2	0.953 (0.0)	0.964 (-1.2)	1.040 (-10.2)	1.000 (-4.4)
h3	0.953 (0.0)	0.964 (-1.2)	1.000 (-7.0)	1.000 (-7.0)
h4	0.953 (0.0)	0.940 (1.4)	1.040 (-7.0)	0.951 (0.2)

Table 10: Vintage별 테스트 RMSE 및 AR(1) 대비 RMSE 감소율 (2023–2024). 종속변수: 전산업생산지수 성장률. 각 셀은 2023–2024년 테스트 구간에서의 RMSE와, 팔호 안의 AR(1) 대비 RMSE 감소율(%)을 함께 보고함. 감소율은 $100 \times (1 - \text{RMSE}_{m,h} / \text{RMSE}_{\text{AR}(1),h})$ 로 정의되며, 양수 값은 동일한 vintage에서 AR(1) 모형보다 예측 오차가 작다는 것을 의미함.

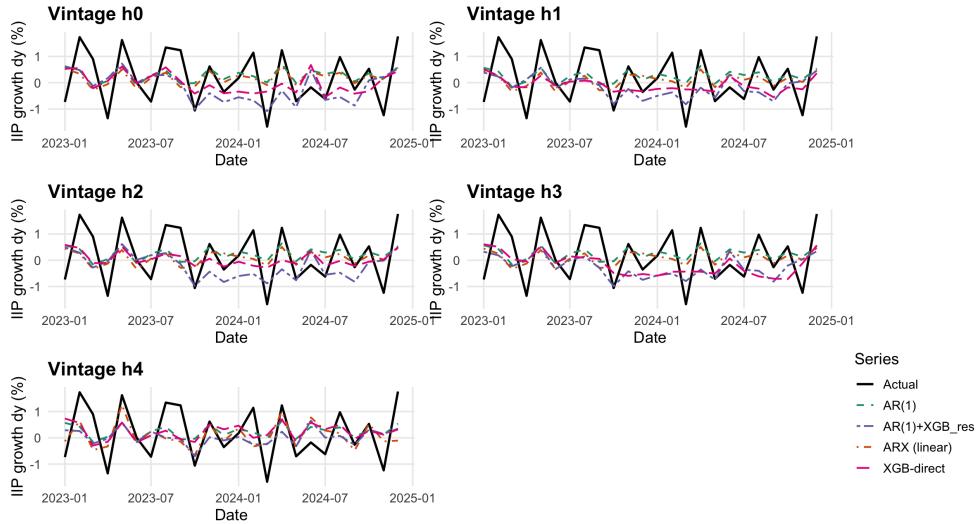


Figure 14: XGBoost 모형 테스트 결과: 전산업생산지수 성장률 (전월대비). AR(1), ARX, AR(1)+XGB_residual, XGB-direct 모형의 vintage별 예측값과 실제값 비교.

Vintage	AR(1)	ARX (linear)	AR(1)+XGB_residual	XGB-direct
h0	1.49 (0.0)	1.51 (-1.6)	1.52 (-2.3)	1.42 (4.4)
h1	1.48 (0.0)	1.58 (-6.4)	1.55 (-4.3)	1.61 (-8.6)
h2	1.48 (0.0)	1.58 (-6.4)	1.58 (-6.7)	1.55 (-4.3)
h3	1.48 (0.0)	1.58 (-6.4)	1.55 (-4.3)	1.58 (-6.7)
h4	1.48 (0.0)	1.53 (-2.9)	1.53 (-2.9)	1.52 (-2.3)

Table 11: Vintage별 테스트 RMSE 및 AR(1) 대비 RMSE 감소율: 전년동월비 (2023–2024). 각 셀은 2023–2024년 테스트 구간에서의 RMSE와, 괄호 안의 AR(1) 대비 RMSE 감소율(%)을 함께 보고함. 감소율은 $100 \times (1 - \text{RMSE}_{m,h} / \text{RMSE}_{\text{AR}(1),h})$ 로 정의되며, 양수 값은 동일한 vintage에서 AR(1) 모형보다 예측 오차가 작다는 것을 의미함.

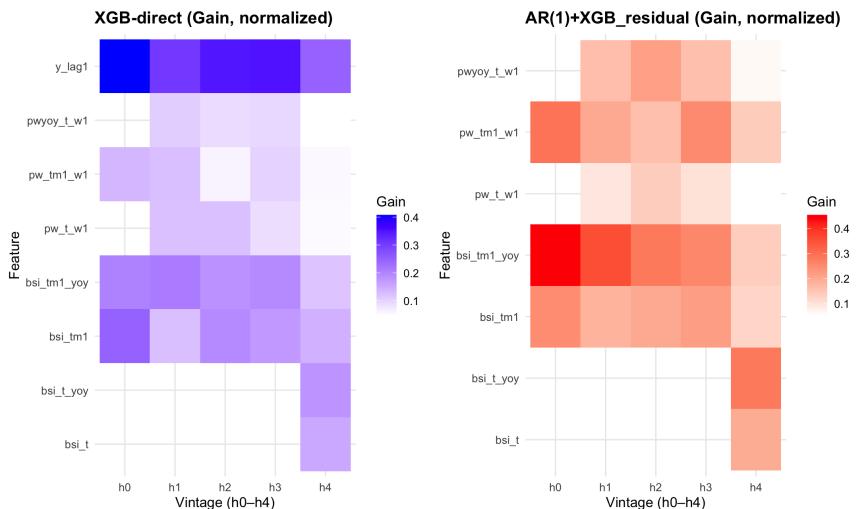


Figure 15: 변수 중요도 히트맵: 전산업생산지수 성장률 (전월대비). 각 변수의 vintage별 중요도를 시각화함.

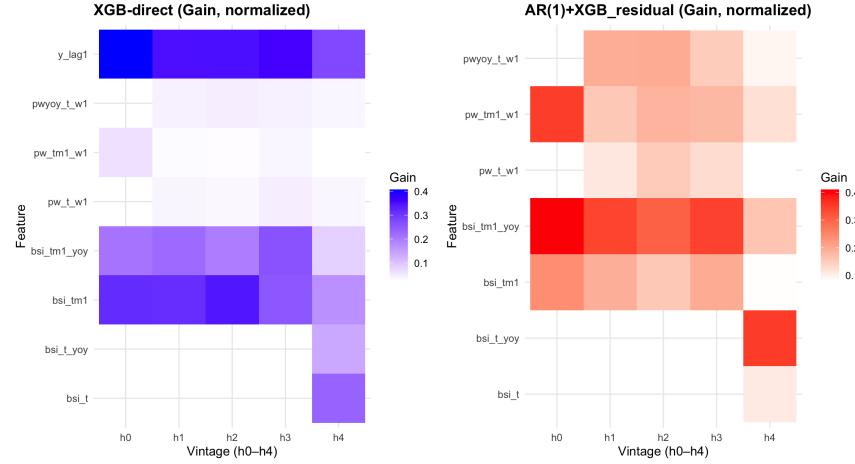


Figure 16: 변수 중요도 히트맵: 전산업생산지수 전년동월비. 각 변수의 vintage별 중요도를 시각화함.

변수	계수 추정치	표준오차	t값
상수항	-1.699	1.906	-0.89
y_{t-1}	-0.403***	0.071	-5.67
$pw_{t-1,w1}$	0.000	0.000	0.21
BSI_{t-1}	-0.076*	0.035	-2.20
BSI_{t-1}^{YoY}	-0.051*	0.022	-2.33
$pw_{t,w1}$	0.000	0.000	-0.21
$pw_{t,w1}^{YoY}$	-0.000	0.005	-0.09
BSI_t	0.096**	0.035	2.76
BSI_t^{YoY}	0.053*	0.022	2.47
R^2		0.239	
조정 R^2		0.205	
관측치 수		188	

Table 12: ARX 모형 추정 결과: 월별 IIP 성장률에 대한 BSI 및 전력거래량의 영향. 종속변수는 전산업생산지수 월별 성장률(y_t)이며, y_{t-1} 은 1기 시차, pw는 월별(또는 주별) 전력거래량 관련 변수, BSI는 기업경기실사지수, “YoY”는 전년동월 대비 변화를 의미함. 두 번째 열은 계수 추정치, 세 번째 열은 표준오차, 네 번째 열은 t값을 나타냄. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

변수	계수 추정치	표준오차	t값
상수항	-11.587***	2.754	-4.21
y_{t-1}	0.398***	0.066	6.07
$pw_{t-1,w1}$	-0.000	0.000	-0.45
BSI_{t-1}	0.113*	0.046	2.48
BSI_{t-1}^{YoY}	-0.076**	0.029	-2.64
$pw_{t,w1}$	0.000	0.000	1.58
$pw_{t,w1}^{YoY}$	-0.008	0.006	-1.28
BSI_t	0.004	0.045	0.08
BSI_t^{YoY}	0.123***	0.028	4.40
R^2		0.788	
조정 R^2		0.779	
관측치 수		188	

Table 13: ARX 모형 추정 결과: 전년동월비에 대한 BSI 및 전력거래량의 영향. 종속변수는 전산업생산지수 월별 전년동월비(y_t)이며, y_{t-1} 은 1기 시차, pw는 전력거래량 관련 변수, BSI는 기업경기실사지수, “YoY”는 전년동월 대비 변화를 의미함. 두 번째 열은 계수 추정치, 세 번째 열은 표준오차, 네 번째 열은 t값을 나타냄. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

- [2] Brandyn Bok, Daniele Caratelli, Domenico Giannone, Argia Sbordone, and Andrea Tambalotti. The frbny staff nowcast. Staff Reports 897, Federal Reserve Bank of New York, December 2019.
- [3] James H. Stock and Mark W. Watson. Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, 97(460):1167–1179, 2002.
- [4] Florian Huber, Gary Koop, Luca Onorante, Michael Pfarrhofer, and Josef Schreiner. Nowcasting in a pandemic using non-parametric mixed frequency vars. *Journal of Econometrics*, 2020. Received 28 August 2020, Received in revised form 17 November 2020, Accepted 30 November 2020, Available online 17 December 2020.
- [5] Paolo Andreini, Cosimo Izzo, and Giovanni Ricco. Deep dynamic factor models. *Working Paper*, 2020. First Version: February 2020, This version: 23 May 2023.
- [6] Eric Ghysels, Pedro Santa-Clara, and Rossen Valkanov. The midas touch: Mixed data sampling regression models. *Review of Economics and Statistics*, 86(4):820–832, November 2004.
- [7] Michael P Clements and Ana Beatriz Galvão. Macroeconomic forecasting with mixed-frequency data: Forecasting output growth in the united states. *Journal of Business & Economic Statistics*, 26(4):546–554, 2008.
- [8] Roberto S. Mariano and Yasutomo Murasawa. A new coincident index of business cycles based on monthly and quarterly series. *Journal of Econometrics*, 117(2):207–231, 2003. Tent kernel aggregation method for mixed-frequency data.