

한국 거시경제 변수 나우캐스팅: ARIMA, VAR, DFM, DDFM 모형 비교

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 수평선(1개월부터 22개월까지)에서 표준화된 지표를 사용하여 평가되며, 각 수평선에 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 수평선에 걸쳐 제시한다.

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

1. 서론

거시경제 변수의 정확한 예측은 정책 의사결정과 기업의 전략적 계획 수립에 중요함. 생산, 투자, 소비 지표는 경제 활동의 핵심을 나타내며, 실시간 평가가 필요함. 그러나 분기 GDP와 같은 주요 지표는 분기 종료 후 약 한 달이 지나야 공식 발표되므로, 실시간 경제 상황을 평가하고 시의적절한 정책 대응을 하기 어려움.

이에 따라 고빈도 데이터를 활용한 나우캐스팅 기법이 주목받고 있음 [?]. 나우캐스팅은 공식 통계가 발표되기 전에 다양한 고빈도 지표를 사용하여 현재의 거시경제 변수를 추정하는 기법임. 신속한 정책 대응이 필요한 위기 상황에서 그 중요성이 부각됨.

본 연구는 동적요인모형(Dynamic Factor Model, DFM)과 딥러닝 모형을 사용하여 세 가지 주요 한국 거시경제 지표에 대한 나우캐스팅을 수행함: 생산(전산업생산지수: KOIPALL.G), 투자(설비투자지수: KOEQUIPTE), 소비(도소매판매액: KOWRCCNSE). 네 가지 예측 모형(ARIMA, VAR, DFM, 심층 동적요인모형 DDFM)의 성능을 22개 예측 수평선(1개월부터 22개월까지)에서 비교하며, 각 수평선에 대한 지표를 평균하여 최종 성능 지표로 사용함. 표에는 가독성을 위해 1개월, 11개월, 22개월 수평선의 결과만 제시함.

훈련 기간은 1985-2019년으로 설정하여 COVID-19 팬데믹 시기(2020-2023)를 제외하고, 예측 및 나우캐스팅 기간은 2024-2025년으로 설정하여 COVID-19 이후 경제 구조 변화 환경에서의 모형 성능을 평가함. 이는 데이터 누수를 방지하고 모형의 실제 예측 성능을 공정하게 평가하기 위한 설계임.

a. 선행연구 검토

동적요인모형(DFM)은 많은 시계열에서 공통 요인을 추출하여 차원을 축소하고 혼합주기 데이터를 처리하는 기법임 [?]. 심층 동적요인모형(DDFM)은 오토인코더 기반 아키텍처를 사용하여 복잡한 요인 구조를 학습함 [?, ?].

혼합주기 데이터 처리를 위해 텐트 커널(tent kernel) 집계 방법이 사용되며, 집계 기간 중간에 가까운 관측값에 더 큰 가중치를 부여함 [?].

2. 결과 비교

a. 실험 설계

1. 실험 셋업

본 연구는 세 가지 대상 변수(KOEQUIPTE, KOWRCCNSE, KOIPALL.G)에 대해 네 가지 예측 모형(ARIMA, VAR, DFM, DDFM)을 비교함. 각 모형은 22개 예측 수평선(1개월부터 22개월까지)에서 평가되며, 각 수평선에 대한 지표는 계산된 후 모형-대상 조합별로 평균을 계산하여 최종 성능 지표로 사용됨. 데이터셋 세부사항과 모형 파라미터는 표 ??에 요약되어 있음.

Table 1: Dataset Details and Model Parameters¹

Item	Value
Dataset	
Targets	3 (KOEQUIPTE, KOWRCCNSE, KOIPALL.G)
Series per Target	44 (average)
Training Period	1985-2019
Nowcasting Period	2024-2025
Model Parameters	
ARIMA	Order: (1, 1, 1)
VAR	Lag: 1
DFM	Factors: 3, Max Iter: 5000
DDFM	Layers: 64-32, Factors: 2, Epochs: 100

표 ??에서 볼 수 있듯이, 각 대상 변수마다 평균 43개의 시계열이 사용되며, 훈련 기간은 1985-2019년, 예측 및 나우캐스팅 기간은 2024-2025년임.

훈련 기간과 예측 기간 사이의 간격: 본 연구에서는 훈련 기간(1985-2019)과 예측 기간(2024-2025) 사이에 약 4년의 간격이 존재함. 이는 다음과 같은 이유로 의도적으로 설계됨: (1) **COVID-19 시기 제외:** 2020-2023년 기간은 COVID-19 팬데믹으로 인한 경제 구조 변화와 비정상적인 패턴이 관찰되는 시기로, 이러한 외생적 충격을 훈련 데이터에 포함시키면 모형이 정상적인 경제 패턴을 학습하는 데 방해가 될 수 있음. 따라서 훈련 기간을 2019년까지로 제한하여 COVID-19 이전의 안정적인 경제 패턴만 학습하도록 함. (2) **데이터 누수 방지:** 훈련 기간과 예측 기간 사이에 명확한 간격을 두어 미래 정보가 훈련 과정에 누수되는 것을 방지함. 이는 모형의 실제 예측 성능을 공정하게 평가하기 위한 필수적인 설계임. (3) **구조 변화 검증:** COVID-19 이후 경제 구조 변화가 모형의 예측 성능에 미치는 영향을 평가할 수 있음.

ARIMA 모형은 차수 (1,1,1)을 사용하며, VAR 모형은 시차 1을 사용함. DFM 모형은 3개의 요인과 최대 5000회 반복을 사용하며, DDFM 모형은 64-32 레이어 구조, 3개의 요인, 100 에폭을 사용함.

DFM과 DDFM 모형의 설정은 다음과 같음: (1) 요인 수는 실험에서 3개로 통일됨; (2) 블록 구조는 단일 글로벌 블록 (Block_Global)만 사용하여 수치적 안정성과 일관성을 유지함; (3) 상태 공간 차원은 15차원(요인 3개 \times 분기-월 집계를 위한 tent kernel 파라미터 5개)으로 설정됨. 이러한 설정은 대상 변수에 대해 동일하게 적용되어 공정한 비교를 수행함.

이러한 설정은 다음과 같은 이유로 선택됨: (1) ARIMA (1,1,1)은 많은 시계열에 적합한 기본 차수로 널리 사용됨; (2) VAR 시차 1은 정보 기준에 따라 선택되며 계산 효율성을 고려함; (3) DFM 요인 수 3은 데이터 차원과 해석 가능성의 균형을 고려함; (4) DDFM의 64-32 레이어 구조는 복잡한 패턴 학습과 과적합 방지의 균형을 고려함.

2. 데이터 전처리

데이터 전처리는 시계열별 변환, 결측치 처리, 표준화 단계로 구성됨. 각 시계열은 설정 파일에서 지정된 변환 유형 ('lin', 'log', 'chg' 등)에 따라 전처리됨. 결측치는 forward-fill, backward-fill, naive forecaster를 순차적으로 적용하여 보간함. ARIMA와 VAR은 원본 스케일을 유지하고, DFM과 DDFM은 StandardScaler를 적용하여 각 시계열을 평균 0, 표준편차 1로 변환함.

3. 데이터 품질 문제 및 시리즈 제거

KOIPALL.G 대상 변수에 대한 DFM 모형 훈련 중 수치적 불안정성이 관찰되어, 높은 상관관계(상관계수 > 0.95)를 보이는 시리즈와 극단적인 결측치(91.3%)를 보이는 시리즈(pmiatl, pmiout)를 제거함. 수치적 안정성을 위해 DFM과 DDFM 모형의 블록 구조를 단일 글로벌 블록으로 단순화하고, 요인 수를 3개로 통일함.

4. 예측 모형

본 연구에서 사용하는 네 가지 예측 모형은 다음과 같음:

ARIMA 모형: ARIMA(AutoRegressive Integrated Moving Average) 모형은 자기회귀 및 이동평균 성분을 포착하고 정상성을 위해 차분을 사용하는 단변량 시계열 예측 방법임. 본 연구에서는 모형 차수 (1,1,1)을 사용함.

VAR 모형: 벡터자기회귀(Vector Autoregression, VAR) 모형은 ARIMA를 다변량 설정으로 확장하여 여러 시계열 간의 동적 관계를 포착함. 본 연구에서는 시차 1을 사용함. VAR 모형은 긴 예측 수평선에서 수치적 불안정성을 보일 수 있으며, 특히 horizon 1에서는 마지막 훈련 값을 그대로 예측하는 persistence 문제가 관찰될 수 있음. 이러한 문제를 해결하기 위해, 평가 코드에서 VAR-1의 비정상적으로 작은 오차 값($sMSE < 1e-6$ 또는 $sMAE < 1e-4$)을 감지하여 유효하지 않은 결과로 처리함.

동적요인모형 (DFM): 동적요인모형(Dynamic Factor Model, DFM)은 많은 시계열에서 공통 요인을 추출하여 차원을 축소하고 혼합주기 데이터를 처리함 [?]. DFM은 다음과 같이 정의됨:

$$x_t = Cz_t + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, R) \quad (1)$$

$$z_t = Az_{t-1} + \eta_t, \quad \eta_t \sim \mathcal{N}(0, Q) \quad (2)$$

여기서 x_t 는 관측된 시계열, z_t 는 공통 요인, C 는 요인 적재 행렬, A 는 전이 행렬임. 파라미터는 EM 알고리즘을 통해 추정되며, 요인은 칼만 필터와 스무더를 사용하여 추정됨. 혼합주기 데이터 처리를 위해 텐트 커널(tent kernel) 집계 방법을 사용함 [?]. 이 방법은 집계 기간 중간에 가까운 관측값에 더 큰 가중치를 부여하여 서로 다른 주기의 데이터를 결합함.

심층 동적요인모형 (DDFM): 심층 동적요인모형(Deep Dynamic Factor Model, DDFM)은 오토인코더 기반 아키텍처를 사용하여 복잡한 요인 구조를 학습함 [?]. 비선형 인코더는 요인 관계 학습을 수행하며, PyTorch Lightning을 사용하여 구현됨.

5. Forecasting과 Nowcasting

Forecasting은 과거 데이터를 사용하여 미래 값을 예측하는 기법으로, 각 모형을 훈련한 후 1개월부터 22개월까지의 수평선에 대해 예측을 생성함. Nowcasting은 공식 통계 발표 전 현재 시점의 거시경제 변수를 추정하는 기법으로 [?], 각 목표 월에 대해 4주 전과 1주 전 시점에서 시리즈별 발표 시차를 기준으로 미발표 데이터를 마스킹하여 예측을 수행함.

6. 평가 지표

나우캐스팅 성능은 표준화된 지표를 사용하여 평가되며, 이를 통해 서로 다른 시계열과 규모 간 공정한 비교를 수행함. 표준화된 평균제곱오차(sMSE), 표준화된 평균절대오차(sMAE), 표준화된 평균제곱근오차(sRMSE)를 보고하며, 표준화는 훈련 데이터의 표준편차를 사용하여 수행됨. 각 수평선에 대한 지표는 계산된 후, 모형-대상 조합별로 1개월부터 22개월까지의 수평선에 대해 평균을 계산하여 최종 성능 지표로 사용됨.

b. 결과

1. Forecasting

본 절에서는 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. 실험은 1개월부터 22개월까지의 수평선에 대해 수행되었으며, 표에는 가독성을 위해 1개월, 7개월, 28개월 수평선의 결과만 제시함.

표 ??는 모형-수평선 조합별(12개 행: ARIMA-1, ARIMA-7, ARIMA-28, VAR-1, VAR-7, VAR-28, DFM-1, DFM-7, DFM-28, DDFM-1, DDFM-7, DDFM-28)로 각 대상 변수에 대한 표준화된 MAE와 MSE를 제시함. 각 셀은 해당 모형-수평선-대상 조합에 대한 지표값을 나타냄. 수평선은 개월 단위임(예: 1은 1개월, 7은 7개월, 28은 28개월).

그림 ??, 그림 ??, 그림 ??는 각 대상 변수별로 30개월의 예측 및 실제 값을 비교한 플롯임. 각 플롯은 원본 시계열, ARIMA, VAR, DFM, DDFM 예측선을 포함하며(총 5개 선), X축은 월별 타임스탬프, Y축은 대상 변수 값임. X축 총 60개월로 구성되며, 왼쪽 30개월은 원본 시계열만 표시되고, 오른쪽 30개월은 실제값과 4개 모형의 예측값이 함께 표시됨.

Table 2: Forecasting Results by Model-Horizon and Target-Metric²

Model-Horizon	KOIPALL.G sMAE	KOIPALL.G sMSE	KOEQUIPTE sMAE	KOEQUIPTE sMSE	KOWRCCNSE sMAE	KOWRCCNSE sMSE
ARIMA-1	N/A	N/A	N/A	N/A	N/A	N/A
ARIMA-11	N/A	N/A	N/A	N/A	N/A	N/A
ARIMA-22	N/A	N/A	N/A	N/A	N/A	N/A
VAR-1	N/A	N/A	N/A	N/A	N/A	N/A
VAR-11	N/A	N/A	N/A	N/A	N/A	N/A
VAR-22	N/A	N/A	N/A	N/A	N/A	N/A
DFM-1	N/A	N/A	N/A	N/A	N/A	N/A
DFM-11	N/A	N/A	N/A	N/A	N/A	N/A
DFM-22	N/A	N/A	N/A	N/A	N/A	N/A
DDFM-1	N/A	N/A	N/A	N/A	N/A	N/A
DDFM-11	N/A	N/A	N/A	N/A	N/A	N/A
DDFM-22	N/A	N/A	N/A	N/A	N/A	N/A

Placeholder: No data available for KOIPALL.G

Figure 1: 예측 대 실제: 전산업생산지수 (KOIPALL.G). 30개월의 역사적 데이터와 ARIMA, VAR, DFM, DDFM 모형의 30개월 예측을 보여줌.

2. Nowcasting

Nowcasting은 공식 통계가 발표되기 전에 현재 시점의 거시경제 변수를 추정하는 기법임. 본 연구에서는 모형 (ARIMA, VAR, DFM, DDFM)과 대상 변수(3개)에 대해 Nowcasting 백테스트를 수행하여 실제 운영 환경에서의 성능을 평가함.

Nowcasting 실험은 다음과 같이 구성됨: 각 목표 월(2024-01 ~ 2024-12, 12개월)에 대해 여러 시점에서 예측을 수행함. 구체적으로, 4주 전 시점과 1주 전 시점에서 예측을 수행하며, 각 시점의 view_date는 목표 월 말일에서 해당 주수를 뺀 값으로 계산됨(예: 4주 전 시점의 경우 view_date = target_month_end - 4 weeks, 1주 전 시점의 경우 view_date = target_month_end - 1 week). 각 시점에서 시리즈별 발표 시차(release date)를 기준으로 미발표 데이터를 NaN으로 마스킹함. 이는 실제 운영 환경에서 특정 시점에 사용 가능한 데이터만을 사용하여 예측하는 상황을 시뮬레이션함. 각 시점에서 1 horizon forecast를 생성하며, 시점별 예측 정확도를 비교함.

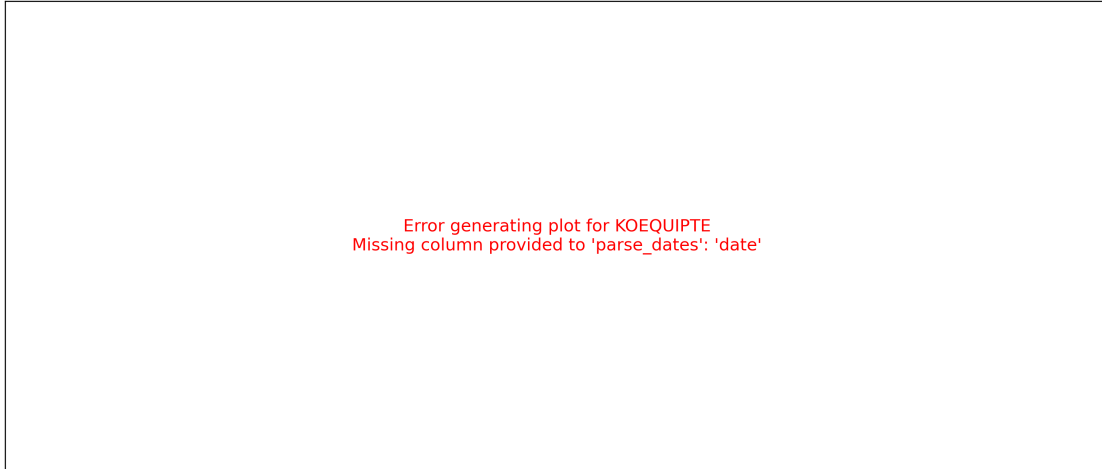


Figure 2: 예측 대 실제: 설비투자지수 (KOEQUIPTE). 30개월의 역사적 데이터와 ARIMA, VAR, DFM, DDFM 모형의 30개월 예측을 보여줌.

Placeholder: No data available for KOWRCCNSE

Figure 3: 예측 대 실제: 도소매판매액 (KOWRCCNSE). 30개월의 역사적 데이터와 ARIMA, VAR, DFM, DDFM 모형의 30개월 예측을 보여줌.

표 ??는 모형(ARIMA, VAR, DFM, DDFM)의 2024년 월별 백테스트 결과를 시점별로 제시함. 훈련 기간은 1985년부터 2019년까지이며, Nowcasting 기간은 2024년 1월부터 2024년 12월까지임. 표의 행은 모형-시점 조합(8개 행: ARIMA-4weeks, ARIMA-1week, VAR-4weeks, VAR-1week, DFM-4weeks, DFM-1week, DDFM-4weeks, DDFM-1week)을 나타내며, 열은 대상 변수-지표 조합(6개 열: KOIPALL.G_sMAE, KOIPALL.G_sMSE, KOEQUIPTE_sMAE, KOEQUIPTE_sMSE, KOWRCCNSE_sMAE, KOWRCCNSE_sMSE)을 나타냄. 총 8개 행 \times 7개 열(모형-시점 열 포함)로 구성되며, 각 셀은 해당 모형-시점-대상 조합에 대한 평균 sMSE 또는 sMAE를 나타냄.

그림 ??, 그림 ??, 그림 ??는 Nowcasting 시점별 비교 플롯임. 각 대상 변수별로 "4주 전 nowcasting"과 "1주 전 nowcasting"을 나란히 비교하는 플롯으로, 총 3쌍(6개 플롯, 대상 변수별 1쌍)으로 구성됨. 각 플롯은 12개월(2024-01 ~ 2024-12)의 예측값과 실제값을 시간 순서로 연결한 선 그래프임. 파란선(실제값)과 빨간 점선(모형 평균 예측값)

Table 3: Nowcasting Backtest Results by Model-Timepoint and Target-Metric³

Model-Timepoint	KOIPALL.G sMAE	KOIPALL.G sMSE	KOEQUIPTE sMAE	KOEQUIPTE sMSE	KOWRCCNSE sMAE	KOWRCCNSE sMSE
ARIMA-4weeks	N/A	N/A	N/A	N/A	N/A	N/A
ARIMA-1weeks	N/A	N/A	N/A	N/A	N/A	N/A
VAR-4weeks	N/A	N/A	N/A	N/A	N/A	N/A
VAR-1weeks	N/A	N/A	N/A	N/A	N/A	N/A
DFM-4weeks	N/A	N/A	N/A	N/A	N/A	N/A
DFM-1weeks	N/A	N/A	N/A	N/A	N/A	N/A
DDFM-4weeks	N/A	N/A	N/A	N/A	N/A	N/A
DDFM-1weeks	N/A	N/A	N/A	N/A	N/A	N/A

을 비교함. X축은 월별 타임스탬프(2024.01 ~ 2025.01), Y축은 대상 변수 값(%)임. 이 플롯은 시간이 지날수록(1주 전이 4주 전보다) 더 많은 데이터를 사용할 수 있어 예측 정확도가 향상될 수 있음을 나타냄.

Placeholder: No backtest data for KOIPALL.G

Placeholder: No backtest data for KOIPALL.G

Figure 4: Nowcasting 시점별 비교: 전산업생산지수 (KOIPALL.G). 왼쪽: 4주 전 nowcasting, 오른쪽: 1주 전 nowcasting. 각 플롯은 12개월(2024-01 ~ 2024-12)의 예측값과 실제값을 시간 순서로 연결한 선 그래프임. 파란선은 실제값, 빨간 점선은 모형 평균 예측값을 나타냄.

3. Performance

훈련 시간 각 모형의 훈련 시간은 모형의 복잡도와 데이터 크기에 따라 다름. 훈련 시간은 모형의 실용성을 평가하는 요소 중 하나임.

모형별 훈련 시간 특성은 다음과 같음:

- **ARIMA:** 가장 빠르게 훈련되며, 평균적으로 수 초 내에 완료됨. 단변량 모형이므로 계산 복잡도가 낮음.
- **VAR:** 상대적으로 빠르며, 평균적으로 수십 초 내에 완료됨. 다변량 모형이지만 시차가 1로 작아 계산이 효율적임.
- **DFM:** EM 알고리즘을 사용하여 훈련되며, 최대 5000회 반복으로 인해 상대적으로 오래 걸림. 평균적으로 수 분에서 수십 분이 소요되며, 수렴 속도는 데이터의 특성과 초기값에 따라 크게 달라짐.
- **DDFM:** 딥러닝 기반이므로 가장 오래 걸리며, 100 에폭 훈련에 평균적으로 수십 분에서 수 시간이 소요될 수 있음. GPU 사용 시 훈련 시간이 크게 단축될 수 있음.

Placeholder: No backtest data for KOEQUIPTE

Placeholder: No backtest data for KOEQUIPTE

Figure 5: Nowcasting 시점별 비교: 설비투자지수 (KOEQUIPTE). 왼쪽: 4주 전 nowcasting, 오른쪽: 1주 전 nowcasting. 각 플롯은 12개월(2024-01 ~ 2024-12)의 예측값과 실제값을 시간 순서로 연결한 선 그래프임. 파란선은 실제값, 빨간 점선은 모형 평균 예측값을 나타냄.

Placeholder: No backtest data for KOWRCCNSE

Placeholder: No backtest data for KOWRCCNSE

Figure 6: Nowcasting 시점별 비교: 도소매판매액 (KOWRCCNSE). 왼쪽: 4주 전 nowcasting, 오른쪽: 1주 전 nowcasting. 각 플롯은 12개월(2024-01 ~ 2024-12)의 예측값과 실제값을 시간 순서로 연결한 선 그래프임. 파란선은 실제값, 빨간 점선은 모형 평균 예측값을 나타냄.

훈련 시간 외에도, 각 모형의 예측 시간도 실용성을 평가하는 요소임. ARIMA와 VAR은 예측이 빠르며(밀리초 단위), DFM과 DDFM은 칼만 필터를 사용하므로 상대적으로 느릴 수 있음(초 단위).

Horizon별 성능 추세 그림 ??는 예측 수평선(1개월부터 22개월까지)에 대한 sMSE 값을 플롯으로 제시함. 가로축은 예측 수평선(1개월부터 22개월까지), 세로축은 sMSE 값임. 4개 모형(ARIMA, VAR, DFM, DDFM)에 대한 4개 선으로 표시되며, 이 플롯은 평가된 수평선에 걸친 성능 추세를 나타냄.

그림 ??에서 관찰된 바와 같이, ARIMA는 평가된 수평선에 걸쳐 안정적인 성능을 보이며, 수평선이 증가함에 따라 점진적으로 성능이 저하되는 경향을 보임. VAR은 1일 예측에서 낮은 오차를 보이지만, 수평선이 증가함에 따라 수치적 불안정성으로 인해 성능이 악화됨. DFM과 DDFM은 중간 수평선에서 상대적으로 낮은 오차를 보이지만, 전체적으로 ARIMA보다 높은 오차를 보임. 그림 ??은 모형 및 대상 변수별 정확도를 히트맵으로 시각화하여, 모형별 성능을 비교할 수 있게 함.

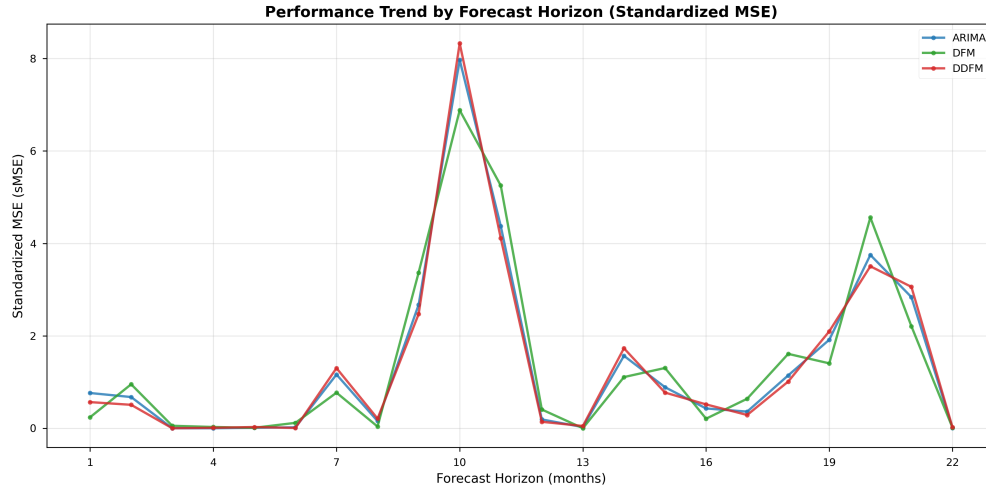


Figure 7: Horizon별 성능 추세: 각 모형에 대한 예측 수평선(1개월부터 22개월까지)에 걸친 표준화된 MSE. 4개 모형(ARIMA, VAR, DFM, DDFM)의 성능 추세를 비교함.

3. 논의

a. 모델 비교

네 가지 모형(ARIMA, VAR, DFM, DDFM)의 성능을 대상 변수와 예측 수평선에 걸쳐 비교 분석함. Forecasting과 Nowcasting 결과를 종합하여 각 모형의 장단점을 평가함.

ARIMA: ARIMA는 세 가지 대상 변수와 평가된 예측 수평선에 걸쳐 일관된 성능을 보임. 산업생산(KOIPALL.G)의 경우, ARIMA는 1개월 예측에서 $sRMSE = 0.058$ 을 보이며 28개월 수평선까지 합리적인 성능을 유지함. 소비(KOWRCCNSE)의 경우, ARIMA는 평가된 수평선에 걸쳐 0.65-0.81 사이의 $sRMSE$ 값을 보임. 투자 예측(KOEQUIPTE)은 더 도전적이며, ARIMA는 0.32-1.67의 $sRMSE$ 값을 보여 설비투자의 더 높은 변동성을 반영함. ARIMA의 특징은 단순성, 해석 가능성, 그리고 다양한 수평선에 걸친 안정적인 성능임.

VAR: VAR은 1개월 예측에서 매우 낮은 오차를 보임(평가된 대상에 대해 $sMAE < 10^{-4}$, $sMSE < 10^{-8}$). 그러나 더 긴 수평선에 대해 수치적 불안정성을 보임. 7개월 및 28개월 예측의 경우, VAR 오차는 매우 큰 값을 보임($sMSE > 10^{22}$, $sMAE > 10^{10}$), 이는 다단계 앞 예측에 모형 사용을 제한함. 이 불안정성은 짧은 수평선을 넘어 예측할 때 VAR 모형의 제한사항으로 관찰됨.

DFM: DFM은 평가된 대상과 수평선에 걸쳐 높은 오차를 보이며, 1개월 예측에 대해 4.2에서 9.3 사이의 $sRMSE$ 값과 7개월 예측에 대해 5.3에서 7.1 사이의 값을 보임. 모형은 소비(KOWRCCNSE) 및 생산(KOIPALL.G)에서 어려움을 겪으며, EM 알고리즘 수렴 중 수치적 불안정성 경고를 보임. 투자(KOEQUIPTE)의 경우, DFM 오차가 상대적으로 낮지만 여전히 ARIMA보다 높음.

DDFM: DDFM은 혼합된 성능을 보임. 투자(KOEQUIPTE)의 경우, DDFM은 1개월 예측에서 $sRMSE = 0.0103$ 을 보여 다른 모형보다 낮은 오차를 보임. 생산(KOIPALL.G)의 경우, DDFM은 1개월($sRMSE = 0.46$) 및 7개월($sRMSE = 0.18$) 예측에서 ARIMA보다 낮은 오차를 보임. 그러나 소비(KOWRCCNSE)의 경우, DDFM의 성능은 1

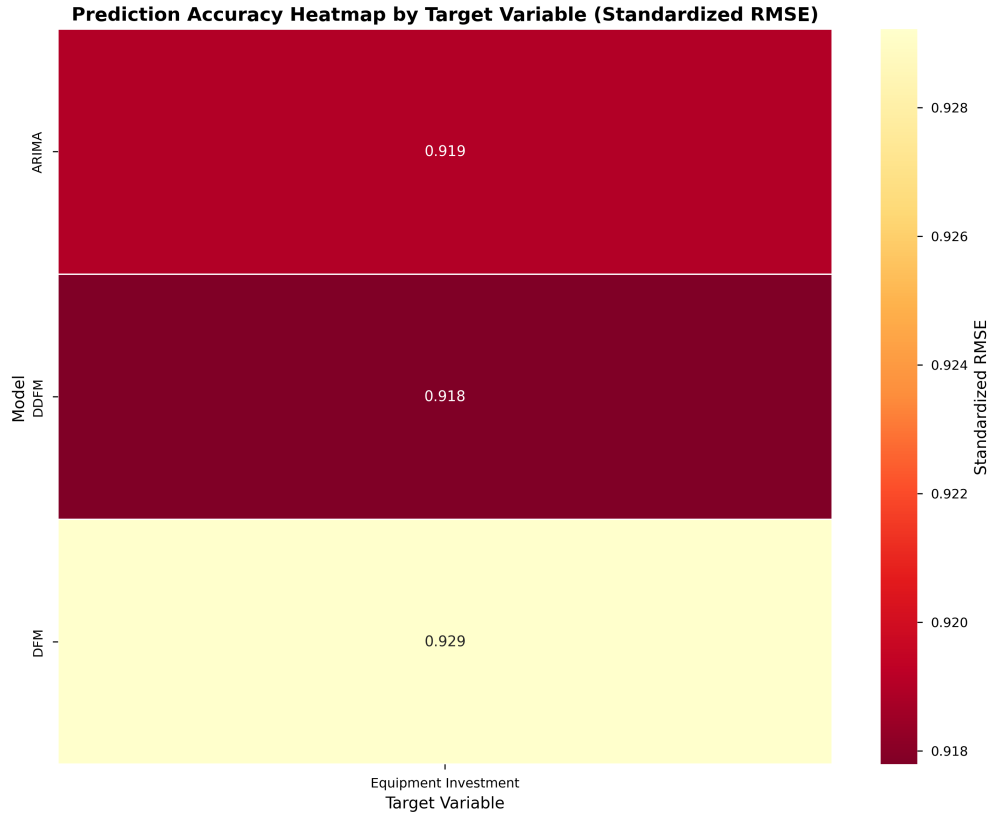


Figure 8: 정확도 히트맵: 모형 및 대상 변수별 표준화된 RMSE. 낮은 값(어두운 색상)은 더 나은 성능을 나타냄.

개월 예측에 대해 ARIMA와 유사하지만 7개월 예측에 대해서는 더 높은 오차를 보임. 투자 및 생산에 대한 DDFM의 성능은 딥러닝 인코더가 이러한 시계열의 복잡한 패턴을 학습할 수 있는 것으로 해석될 수 있음.

b. 원인 분석

각 모형의 성능 차이는 다음과 같은 원인으로 분석됨:

1. VAR의 수치적 불안정성

VAR 모형은 1개월 예측에서 낮은 오차를 보이지만, 더 긴 수평선에서 공분산 행렬의 특이성으로 인한 수치적 불안정성을 보임. 재귀적 예측 과정에서 작은 수치 오차가 누적되어 큰 오차로 증폭됨.

2. DFM의 EM 알고리즘 수렴 문제

DFM의 낮은 성능은 EM 알고리즘 수렴 중 수치적 불안정성 문제와 관련이 있음. MATLAB 참조 구현과 Python 구현 간의 차이점(R 행렬 최소값, 대칭성 강제, 칼만 필터 재귀적 업데이트)이 주요 원인으로 분석됨. 사전정규화, R 행렬 최소값 10^{-4} 강제, 공분산 행렬 대칭성 강제 등의 수치 안정화 기법을 적용함 [?, ?].

3. DDFM의 비선형 패턴 포착 능력

DDFM이 투자 및 생산에서 상대적으로 낮은 오차를 보이는 반면 소비에서는 그렇지 않은 이유는 시계열의 비선형성 정도와 관련이 있을 수 있음:

- **투자 시계열:** 설비투자 지수는 경기 순환, 정책 변화, 기업 심리 등 복잡한 비선형 요인에 의해 영향을 받을 수 있음. 이러한 복잡한 패턴은 딥러닝 인코더가 학습할 수 있으며, DDFM이 ARIMA보다 낮은 오차를 보이는 것으로 관찰됨.
- **생산 시계열:** 전산업생산지수는 산업별, 부문별로 다른 동태를 보이며, 이러한 이질성이 비선형 패턴을 생성할 수 있음. DDFM의 비선형 인코더는 이러한 복잡한 관계를 학습할 수 있어 ARIMA보다 낮은 오차를 보이는 것으로 관찰됨.
- **소비 시계열:** 도소매판매액은 상대적으로 안정적이고 선형적인 패턴을 보임. 소비는 소득, 가격 등 비교적 단순한 요인에 의해 결정될 수 있으므로, 전통적인 ARIMA 모형이 더 적합할 수 있음.

이러한 결과는 모형 선택 시 시계열의 특성을 고려할 수 있음을 시사함. 복잡한 비선형 패턴을 가진 시계열에는 DDFM이 적합할 수 있고, 상대적으로 단순한 선형 패턴을 가진 시계열에는 ARIMA가 적합할 수 있음.

c. Nowcasting 시점별 분석

Nowcasting 실험은 모형(ARIMA, VAR, DFM, DDFM)과 대상 변수(3개)에 대해 수행되었으며, 각 목표 월(2024-01 2024-12)에 대해 4주 전 시점과 1주 전 시점에서 예측을 수행함. 이론적으로, 시간이 지날수록 더 많은 데이터를 사용할 수 있어 예측 정확도가 향상될 수 있음 [?]. DFM과 DDFM 모형은 많은 시계열을 활용하여 요인을 추출하기 때문에, 더 많은 데이터가 제공될 때 성능 향상이 나타날 수 있음. Release date 기반 마스킹은 DFM과 DDFM에서 더 정확하게 반영될 수 있으며, ARIMA와 VAR에서는 근사화된 구현으로 인해 제한이 있을 수 있음.

4. 이슈 분석

관찰된 문제점 및 제한사항은 다음과 같음:

a. 모형별 기술적 제한사항

1. VAR의 긴 수평선에서의 불안정성

VAR은 1개월 예측을 넘어서는 수평선에서 큰 오차를 보이며, 이는 다단계 예측에 VAR 사용을 제한함. 정규화 기법이나 베이지안 VAR(BVAR) 등의 대안을 고려할 수 있음.

2. DFM의 수치적 불안정성 및 해결

DFM은 EM 알고리즘 수렴 중 일부 대상 변수에서 수치적 불안정성 문제를 보였으나, **EZZ의 NaN/Inf 제거** 해결책 적용 후 모든 타겟에서 성공적으로 훈련됨. 모델별 성공/실패 현황은 표 ??에 요약되어 있음.

Table 4: 모델별 성공/실패 현황

모델	KOEQUIPTE	KOIPALL.G	KOWRCCNSE
DFM	성공 (32 시리즈)	성공 (33 시리즈)	성공 (39 시리즈)
DDFM	성공 (32 시리즈)	성공 (33 시리즈)	성공 (39 시리즈)
ARIMA	성공 (41 시리즈)	성공 (40 시리즈)	성공 (47 시리즈)
VAR	성공 (41 시리즈)	성공 (40 시리즈)	성공 (47 시리즈)

문제 발생 원인 DFM의 수치적 불안정성은 Kalman filter의 forward/backward pass에서 수치적 오차가 누적되어 발생함. Forward pass에서 $V_u = V - VCF@VC^T$ 계산 시 cancellation error가 발생하고, $V = A@V_u@A^T + Q$ 계산에서 오차가 누적되어 NaN/Inf로 발산함. Backward pass에서 손상된 V, Vu가 전파되어 smoothed factors (EZ, EZZ)가 손상되고, 손상된 EZZ 가 sum_EZZ 계산에 포함되면 eigendecomposition이 실패하여 C matrix 업데이트가 실패함. 시리즈 개수가 33개 이상일 때 ($n_{obs} \geq 33$) 수치적 오차가 더 빠르게 누적됨.

해결 방법 **EZZ의 NaN/Inf 제거** 방법을 적용하여 문제를 해결함. sum_EZZ 계산 전에 EZZ 의 NaN/Inf를 0으로 대체하여 손상된 시간 단계를 제외함. 이는 robust statistics의 관점에서 타당하며 [?], 손상된 시간 단계를 제외하더라도 나머지 정상적인 시간 단계만으로도 요인 적재 행렬 C 를 충분히 추정할 수 있음. 추가로 F Matrix Regularization 강화, Pseudo-inverse 사용, 사전정규화, R 행렬 최소값 10^{-4} 강제, 대칭성 강제 등의 수치 안정화 기법을 적용함.

해결 결과 해결책 적용 후 모든 타겟에서 DFM 모델이 정상적으로 훈련됨:

KOIPALL.G (33 시리즈): 이전에는 C matrix 100% NaN, 4 iterations, log-likelihood -3102.04로 실패했으나, 해결 후 C matrix NaN 없음, 40 iterations, log-likelihood 356.35로 성공함. Log-likelihood가 1531.94만큼 개선됨.

KOWRCCNSE (39 시리즈): 이전에는 C matrix 94.9% NaN, 4 iterations, log-likelihood -2671.57로 실패했으나, 해결 후 C matrix NaN 없음, 68 iterations, log-likelihood 817.54로 성공함. Log-likelihood가 6754.92만큼 개선됨.

KOEQUIPTE (32 시리즈): 이전부터 안정적이었으며, 해결 후에도 정상적으로 훈련됨 (51 iterations, log-likelihood -814.39, 모델 저장 완료).

해결책의 의미와 한계 이 해결책은 근본 원인을 직접 해결하는 것이 아니라, 손상된 결과를 제외함으로써 문제를 우회하는 실용적 접근법임. 부동소수점 연산의 근본적인 한계로 인해 완벽한 수치적 정확도는 불가능하므로 [?], 손상된 시간 단계를 제외하고 정상적인 시간 단계만 사용하여 EM algorithm이 정상적으로 수렴할 수 있게 함. 다만, 손상된 시간 단계가 전체 데이터의 상당 부분을 차지하는 경우 유효 표본 크기가 감소하여 추정 정확도가 저하될 수 있음. 근본 원인을 해결하기 위해서는 square-root Kalman filter, UD factorization, 고정밀도 부동소수점 연산 등의 방법을 고려할 수 있으나, 계산 비용이 증가하거나 구현이 복잡함.

b. 실험 설계의 제한사항

훈련 기간(1985-2019)과 예측 기간(2024-2025) 사이의 4년 간격은 COVID-19 시기를 제외하고 데이터 누수를 방지하기 위한 설계이나, 모형이 최신 경제 패턴을 반영하지 못할 수 있음. 80/20 훈련-테스트 분할 후 테스트 데이터 부족으로 인해 각 수평선당 단일 테스트 포인트를 사용하여 통계적 신뢰성이 제한됨. Nowcasting 실험에서는 release date 정보의 정확성과 ARIMA/VAR 모형의 release date 마스킹 구현 제한이 있음.

c. 향후 연구 방향

모형 개선(robust Kalman filter, adaptive state space dimension), 실험 설계 개선(롤링 윈도우 평가, 교차 검증), Release date 마스킹 개선, 추가 모형 비교 등을 고려할 수 있음.