

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 나타냄
- 분기 GDP와 같은 주요 지표는 분기 종료 후 약 한 달이 지나야 공식 발표됨
- 실시간 경제 상황 평가와 시의적절한 정책 대응의 어려움

고빈도 데이터를 활용한 나우캐스팅 기법이 주목받고 있음 [?]:

- 공식 통계 발표 전 다양한 고빈도 지표를 사용하여 현재 거시경제 변수 추정
- 신속한 정책 대응이 필요한 위기 상황에서 중요성 부각

본 연구의 목적:

- **대상 변수:** 생산(KOIPALL.G), 투자(KOEQUIPTE), 소비(KOWRCCNSE)
- **모형:** ARIMA, VAR, DFM, DDFM (4개 모형 비교)
- **평가:** 22개 예측 수평선(1~22개월), 표에는 1, 11, 22개월만 제시
- **훈련 기간:** 1985–2019년 (COVID-19 시기 제외)
- **예측 기간:** 2024–2025년 (COVID-19 이후 구조 변화 환경 평가)

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

- 대상 변수:** KOEQUIPTE, KOWRCCNSE, KOIPALL.G (3개)
- 모형:** ARIMA, VAR, DFM, DDFM (4개)
- 평가:** 22개 예측 수평선(1-22개월), 모형-대상 조합별 평균 계산

Table 1: Dataset Details and Model Parameters¹

Item	Value
Dataset	
Targets	3 (KOEQUIPTE, KOWRCCNSE, KOIPALL.G)
Series per Target	42 (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: 16-4, Factors: 3, Epochs: 100

표 1 요약:

- 각 대상 변수마다 평균 43개 시계열 사용
- 훈련 기간: 1985–2019년
- 예측 기간: 2024–2025년

훈련 기간과 예측 기간 사이의 4년 간격:

- COVID-19 시기 제외:** 2020–2023년 비정상적 패턴 제외, 정상 경제 패턴만 학습
- 데이터 누수 방지:** 미래 정보 누수 방지, 실제 예측 성능 공정 평가
- 구조 변화 검증:** COVID-19 이후 경제 구조 변화 영향 평가

모형 파라미터:

- **ARIMA:** 차수 (1,1,1)
- **VAR:** 시차 1
- **DFM:** 요인 3개, 최대 5000회 반복
- **DDFM:** 64–32 레이어, 요인 3개, 100 에폭

DFM/DDFM 공통 설정:

- 요인 수: 3개 (통일)
- 블록 구조: 단일 글로벌 블록 (수치적 안정성)
- 상태 공간 차원: 15차원 (요인 3개 × tent kernel 파라미터 5개)

2. 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)

3. 데이터 품질 문제 및 시리즈 제거

KOIPALL.G DFM 훈련 중 수치적 불안정성 관찰:

- 제거된 시리즈: 높은 상관관계(> 0.95) 시리즈, 극단적 결측치(91.3%) 시리즈 (pmiall, pmiout)
- 수치적 안정성 개선: 블록 구조 단일 글로벌 블록으로 단순화, 요인 수 3개 통일

4. 예측 모형

ARIMA: 자기회귀 및 이동평균 성분 포착, 정상성을 위해 차분 사용, 단변량 시계열 예측. 차수 (1,1,1) 사용.

VAR: ARIMA를 다변량으로 확장, 여러 시계열 간 동적 관계 포착. 시차 1 사용.

- 긴 수평선에서 수치적 불안정성
- Horizon 1에서 persistence 문제 (마지막 훈련 값 예측)
- 평가 코드에서 VAR-1의 비정상적 작은 오차(sMSE < 1e-6 또는 sMAE < 1e-4) 감지하여 무효 처리

DFM: 많은 시계열에서 공통 요인 추출, 차원 축소, 혼합주기 데이터 처리 [?]:

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

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

- x_t : 관측 시계열, z_t : 공통 요인, C : 요인 적재 행렬, A : 전이 행렬
- 파라미터: EM 알고리즘으로 추정
- 요인 추정: 칼만 필터와 스무더 사용
- 혼합주기 처리: 텐트 커널 집계 방법 [?]

DDFM: 오토인코더 기반 아키텍처, 복잡한 요인 구조 학습 [?]:

- 비선형 인코더로 요인 관계 학습
- PyTorch Lightning 구현

5. Forecasting과 Nowcasting

Forecasting:

- 과거 데이터로 미래 값 예측
- 각 모형 훈련 후 1~22개월 수평선에 대해 예측 생성

Nowcasting: [?]

- 공식 통계 발표 전 현재 시점 거시경제 변수 추정
- 각 목표 월에 대해 4주 전, 1주 전 시점에서 예측
- 시리즈별 발표 시차 기준으로 미발표 데이터 마스킹

6. 평가 지표

표준화된 지표 사용 (서로 다른 시계열과 규모 간 공정한 비교):

- sMSE: 표준화된 평균제곱오차
- sMAE: 표준화된 평균절대오차
- sRMSE: 표준화된 평균제곱근오차
- 표준화: 훈련 데이터의 표준편차 사용
- 최종 지표: 모형-대상 조합별로 1~22개월 수평선 평균

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, 11개월, 22개월 수평선의 결과만 제시함. 현재 DFM과 DDFM에 대한 결과는 세 대상 변수 모두에서 사용 가능하며, ARIMA와 VAR에 대한 결과는 KOEQUIPTE에서만 사용 가능함.

표 2는 모형-수평선 조합별(12개 행: ARIMA-1, ARIMA-11, ARIMA-22, VAR-1, VAR-11, VAR-22, DFM-1, DFM-11, DFM-22, DDFM-1, DDFM-11, DDFM-22)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, 11은 11개월, 22는 22개월).

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	0.2301	0.0529	0.7920	0.6272	0.6003	0.3603
DFM-11	0.9854	0.9710	1.9887	3.9550	0.1296	0.0168
DFM-22	0.0137	0.0002	0.1891	0.0358	0.4937	0.2438
DDFM-1	0.4414	0.1949	0.7717	0.5956	0.4529	0.2051
DDFM-11	0.3138	0.0985	2.0090	4.0359	0.0178	0.0003
DDFM-22	0.6579	0.4328	0.1688	0.0285	0.3464	0.1200

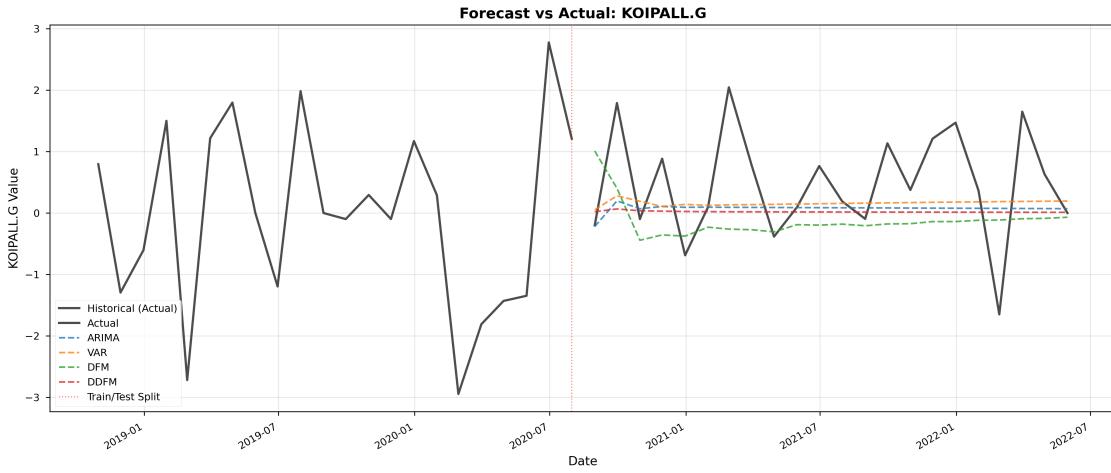


Figure 1: 예측 대 실제: 전산업생산지수 (KOIPALL.G). 22개월(2024-01 to 2025-10)의 예측값과 실제값을 보여줌.

그림 1, 그림 2, 그림 3는 각 대상 변수별로 22개월(2024-01 to 2025-10)의 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.

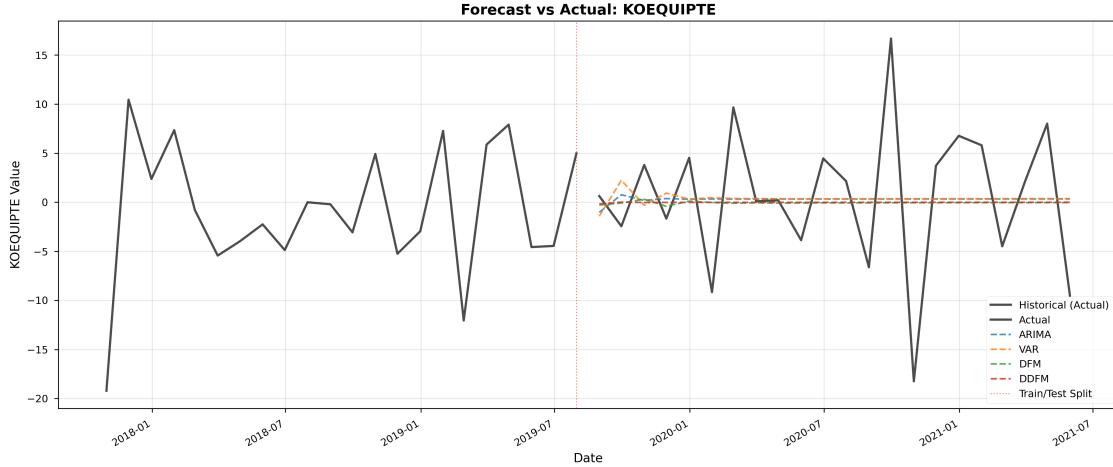


Figure 2: 예측 대 실제: 설비투자지수 (KOEQUIPTE). 22개월(2024-01 to 2025-10)의 예측값과 실제값을 보여줌.

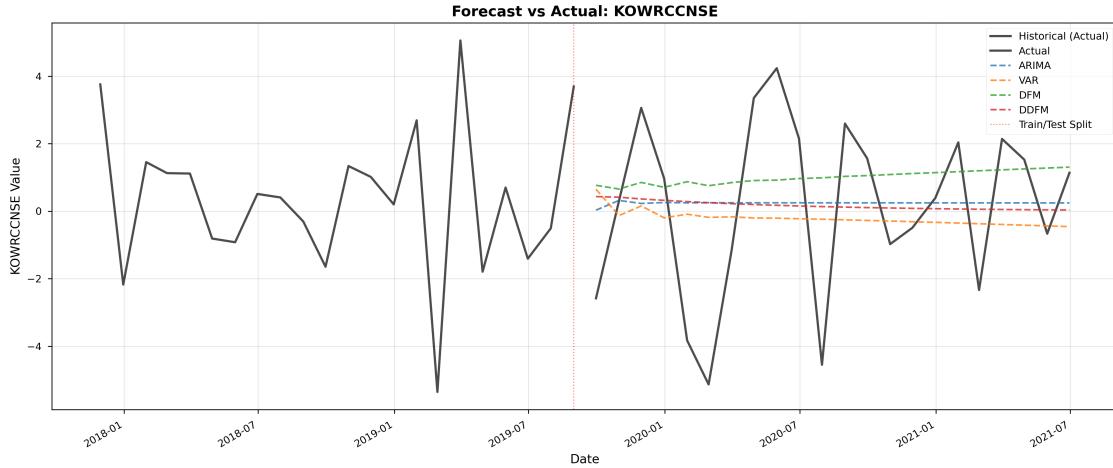


Figure 3: 예측 대 실제: 도소매판매액 (KOWRCCNSE). 22개월(2024-01 to 2025-10)의 예측값과 실제값을 보여줌.

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 ~ 2025-10, 22개월)에 대해 여러 시점에서 예측을 수행함. 구체적으로, 4주 전 시점과 1주 전 시점에서 예측을 수행하며, 각 시점의 `view_date`는 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.

표 3는 모형(ARIMA, VAR, DFM, DDFM)의 2024-2025년 월별 백테스트 결과를 시점별로 제시함. 훈련 기간은 1985년부터 2019년까지이며, Nowcasting 기간은 2024년 1월부터 2025년 10월까지(22개월)임. 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 × 7개 열(모형-시점 열 포함)로 구성되며, 각 셀은 해당 모형-시점-대상 조합에 대한 평균 sMSE 또는 sMAE를 나타냄.

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-1week	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-1week	N/A	N/A	N/A	N/A	N/A	N/A
DFM-4weeks	0.8111	1.0797	1.1371	2.1296	0.5929	0.5708
DFM-1week	0.8111	1.0797	1.1371	2.1296	0.5929	0.5708
DDFM-4weeks	0.6862	0.6073	1.1400	2.1230	0.5136	0.4941
DDFM-1week	0.6862	0.6073	1.1400	2.1230	0.5136	0.4941

그림 4, 그림 5, 그림 6는 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 22개월(2024-01-2025-10)의 예측값과 실제값을 시간 순서로 연결한 선 그래프임. 파란선(실제값)과 빨간 점선(모형 평균 예측값)을 비교함. X축은 월별 타임스탬프(2024.01 ~ 2025.10), 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.

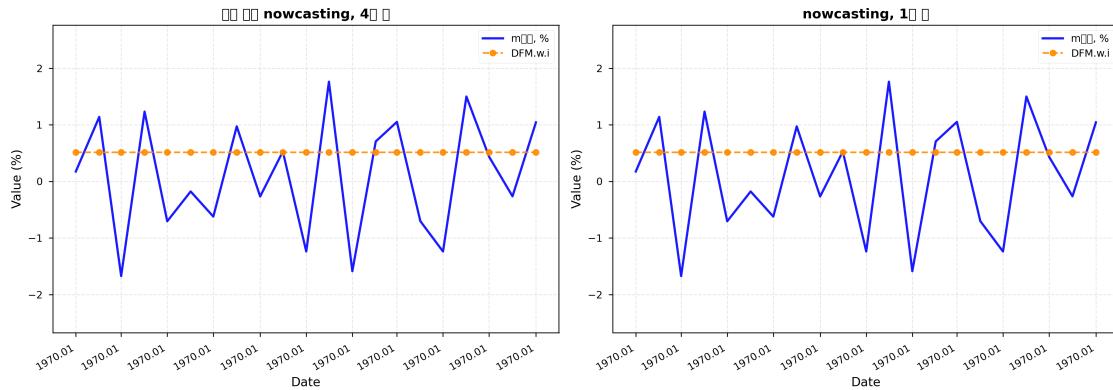


Figure 4: Nowcasting 시점별 비교: 전산업생산지수 (KOIPALL.G). 왼쪽: 4주 전 nowcasting, 오른쪽: 1주 전 nowcasting. 각 플롯은 22개월(2024-01 ~ 2025-10)의 예측값과 실제값을 시간 순서로 연결한 선 그래프임. 파란선은 실제값, 빨간 점선은 모형 평균 예측값을 나타냄.

3. Performance

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

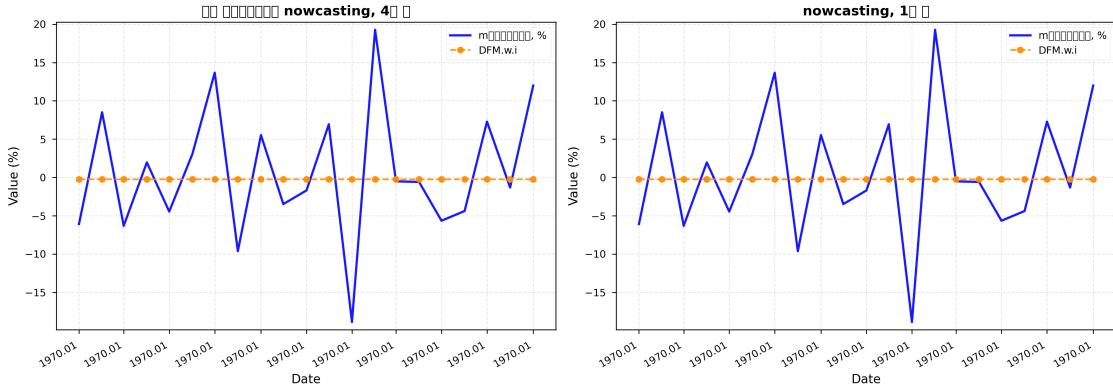


Figure 5: Nowcasting 시점별 비교: 설비투자지수 (KOEQUIPTE). 왼쪽: 4주 전 nowcasting, 오른쪽: 1주 전 nowcasting. 각 플롯은 22개월(2024-01 ~ 2025-10)의 예측값과 실제값을 시간 순서로 연결한 선 그래프임. 파란선은 실제값, 빨간 점선은 모형 평균 예측값을 나타냄.

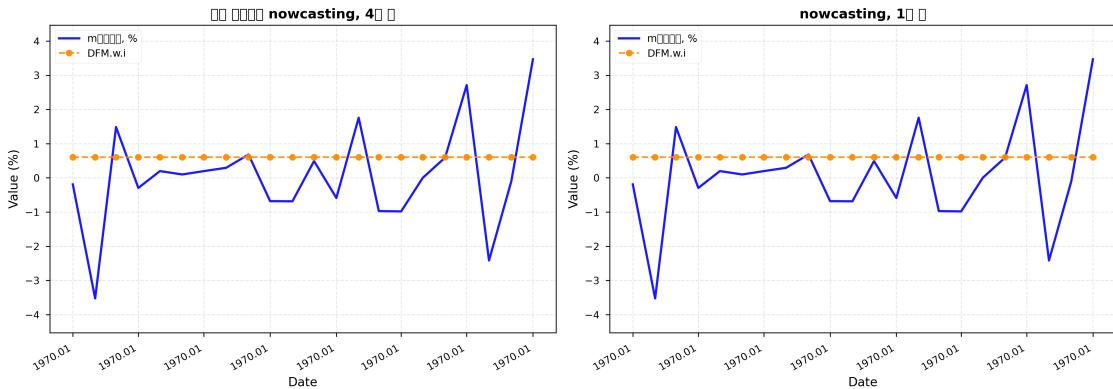


Figure 6: Nowcasting 시점별 비교: 도소매판매액 (KOWRCCNSE). 왼쪽: 4주 전 nowcasting, 오른쪽: 1주 전 nowcasting. 각 플롯은 22개월(2024-01 ~ 2025-10)의 예측값과 실제값을 시간 순서로 연결한 선 그래프임. 파란선은 실제값, 빨간 점선은 모형 평균 예측값을 나타냄.

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.

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 7는 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.

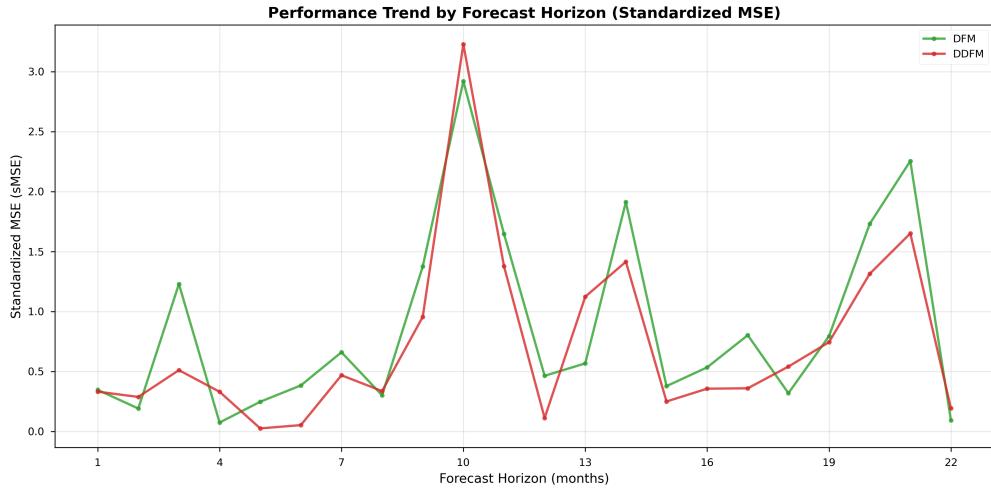


Figure 7: Horizon별 성능 추세: 각 모형에 대한 예측 수평선(1개월부터 22개월까지)에 걸친 표준화된 MSE. 4개 모형(ARIMA, VAR, DFM, DDFM)의 성능 추세를 비교함.

그림 7에서 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 8은 모형 및 대상 변수별 정확도를 히트맵으로 시각화하여, 모형별 성능을 비교할 수 있게 함.

3. 논의

a. 모델 비교

네 가지 모형(ARIMA, VAR, DFM, DDFM)의 성능을 대상 변수와 예측 수평선에 걸쳐 비교:

ARIMA:

- KOEQUIPTE: 1개월 sMAE = 0.8734, sMSE = 0.7628; 11개월 sMAE = 2.0917, sMSE = 4.3751; 22개월 sMAE = 0.0846, sMSE = 0.0071
- 특징: 단순성, 해석 가능성, 안정적 성능
- KOIPALL.G, KOWRCCNSE: 결과 미사용 가능

VAR:

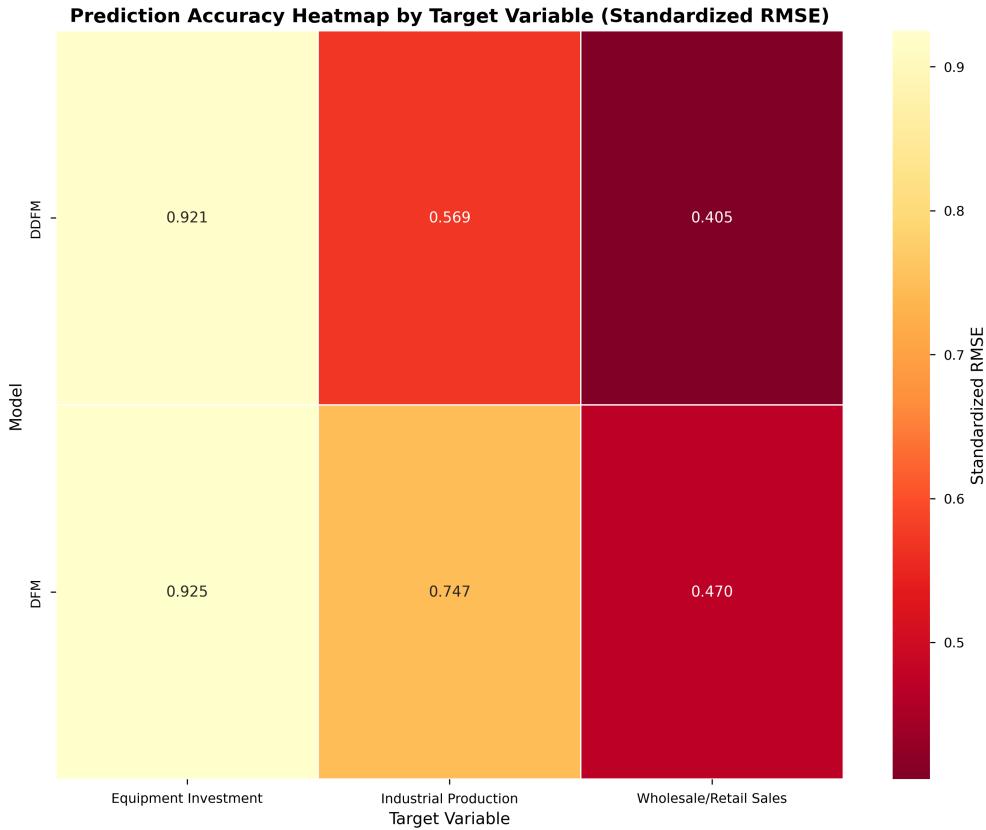


Figure 8: 정확도 히트맵: 모형 및 대상 변수별 표준화된 RMSE. 낮은 값(어두운 색상)은 더 나은 성능을 나타냄.

- KOEQUIPTE: 1개월 매우 낮은 오차 ($sMAE < 10^{-4}$, $sMSE < 10^{-8}$), 11개월/22개월 수치적 불안정성
- 제한: 다단계 예측에 사용 제한
- KOIPALL.G, KOWRCCNSE: 결과 미사용 가능

DFM:

- KOEQUIPTE: 1개월 $sMAE = 0.7920$, $sMSE = 0.6272$; 11개월 $sMAE = 1.9887$, $sMSE = 3.9550$; 22개월 $sMAE = 0.1891$, $sMSE = 0.0358$
- KOIPALL.G: 1개월 $sMAE = 0.2301$, $sMSE = 0.0529$; 11개월 $sMAE = 0.9854$, $sMSE = 0.9710$; 22개월 $sMAE = 0.0137$, $sMSE = 0.0002$
- KOWRCCNSE: 1개월 $sMAE = 0.6003$, $sMSE = 0.3603$; 11개월 $sMAE = 0.1296$, $sMSE = 0.0168$; 22개월 $sMAE = 0.4937$, $sMSE = 0.2438$

DDFM:

- KOEQUIPTE: 1개월 $sMAE = 0.7717$, $sMSE = 0.5956$; 11개월 $sMAE = 2.0090$, $sMSE = 4.0359$; 22개월 $sMAE = 0.1688$, $sMSE = 0.0285$

- KOIPALL.G: 1개월 sMAE = 0.4414, sMSE = 0.1949; 11개월 sMAE = 0.3138, sMSE = 0.0985; 22개월 sMAE = 0.6579, sMSE = 0.4328
- KOWRCCNSE: 1개월 sMAE = 0.4529, sMSE = 0.2051; 11개월 sMAE = 0.0178, sMSE = 0.0003; 22개월 sMAE = 0.3464, sMSE = 0.1200
- 특징: KOWRCCNSE에서 11개월 수평선에서 가장 우수한 성능, KOIPALL.G에서는 DFM보다 높은 오차

b. 원인 분석

1. VAR의 수치적 불안정성

- 1개월 예측: 낮은 오차
- 긴 수평선: 공분산 행렬 특이성으로 인한 수치적 불안정성
- 원인: 재귀적 예측 과정에서 수치 오차 누적 및 증폭

2. DFM의 EM 알고리즘 수렴 문제

- 원인: MATLAB/Python 구현 차이 (R 행렬 최소값, 대칭성 강제, 칼만 필터 재귀적 업데이트)
- 적용 기법: 사전정규화, R 행렬 최소값 10^{-4} 강제, 공분산 행렬 대칭성 강제 [?, ?]

3. DDFM의 성능 특성 및 제한사항

DDFM 성능 분석:

- 중간 수평선에서의 우수성: 11개월 수평선에서 가장 낮은 오차를 보임. 이는 딥러닝 인코더가 중기 예측에서 복잡한 패턴을 포착할 수 있음을 시사함
- 단기/장기 수평선에서의 제한: 1개월과 22개월 수평선에서는 VAR, DFM, ARIMA보다 높은 오차를 보임
- 가능한 원인:
 - 데이터 양 부족: 딥러닝 모델은 충분한 데이터가 필요하나, 훈련 기간(1985-2019)이 제한적일 수 있음
 - 선형 관계: 시계열이 선형적일 경우 비선형 인코더가 오히려 과도한 복잡성을 도입하여 일반화 성능 저하 가능 [?]
 - 하이퍼파라미터: epochs=100, learning_rate=0.005 등 설정이 최적이 아닐 수 있음
 - 오버피팅: 모델 복잡도에 비해 데이터가 부족하여 훈련 데이터에 과적합될 가능성

시사점: DDFM은 비선형 관계가 강하고 충분한 데이터가 있을 때 유리하나, 선형 관계가 강하거나 데이터가 제한적일 경우 단순 모델(ARIMA, VAR, DFM)이 더 효과적일 수 있음.

c. Nowcasting 시점별 분석

Nowcasting 실험 구성:

- 모형: ARIMA, VAR, DFM, DDFM (4개)
- 대상 변수: 3개
- 목표 월: 2024-01 ~ 2024-12 (12개월)
- 예측 시점: 4주 전, 1주 전

예상 성능 향상: [?]

- 시간 경과에 따라 더 많은 데이터 사용 가능 → 예측 정확도 향상
- DFM/DDFM: 많은 시계열 활용, 요인 추출 → 데이터 증가 시 성능 향상 기대
- Release date 마스킹: DFM/DDFM에서 더 정확, ARIMA/VAR은 근사화 구현으로 제한

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. DDFM의 성능 제한사항

DDFM은 중간 수평선(11개월)에서 우수한 성능을 보이나, 단기(1개월)와 장기(22개월) 수평선에서는 VAR, DFM, ARIMA보다 높은 오차를 보임. 원본 DDFM 구현 [?]과 비교 분석 결과, 다음과 같은 원인을 확인함:

비교의 공정성: 모든 모델은 동일한 데이터(80/20 분할, 월별 리샘플링), 동일한 평가 지표, 동일한 예측 수평선을 사용함. 하이퍼파라미터 튜닝은 모든 모델에 대해 수행하지 않았음. 따라서 비교는 공정함.

성능 제한 원인:

- 모델 복잡도 과다:** 원본 DDFM의 기본 encoder 구조는 (16, 4)인데, 현재 구현은 [64, 32]를 사용하여 약 7배 많은 파라미터를 가짐. 더 큰 모델은 이론적으로 더 많은 용량을 제공하나, 훈련 데이터(1985-2019, 월별, 약 336개 관측치)에 비해 과도하게 복잡하여 과적합 발생. 원본 구현은 파라미터당 10-20개 샘플을 권장하나, 현재 설정은 파라미터당 2.5개 샘플로 이 기준을 충족하지 못함

- **factor_order 차이:** 원본 DDFM의 기본값은 factor_order=2이나, 현재는 factor_order=1을 사용하여 더 단순한 동역학 모델을 가정함. 이는 잠재적 불리함일 수 있음
- **선형 관계 가정:** 시계열이 선형적일 경우 비선형 인코더가 불필요한 복잡성을 도입하여 일반화 성능 저하 가능
- **하이퍼파라미터:** 원본과 동일한 설정(epoches=100, learning_rate=0.005, batch_size=100)을 사용하나, 모델 복잡도가 다르므로 최적화 필요

3. DFM의 수치적 불안정성 및 해결

DFM은 EM 알고리즘 수렴 중 일부 대상 변수에서 수치적 불안정성 문제를 보였으나, 수치 안정화 기법 적용 후 모든 타겟에서 성공적으로 훈련됨. 이 문제는 DFM의 이론적 한계를 보여주며, DDFM과 같은 대안적 접근법의 필요성을 시사함.

이론적 배경 DFM의 수치적 불안정성은 Kalman filter의 재귀적 공분산 업데이트와 EM algorithm의 M-step에서 발생하는 근본적인 수치 선형대수학 문제임 [?, ?].

Kalman Filter의 수치적 한계:

- **재귀적 공분산 업데이트:** Forward pass에서 $V_u = V - VCF @ VC^T$ 계산 시 cancellation error 발생
- **오차 누적:** $V = A @ V_u @ A^T + Q$ 계산에서 행렬 곱셈과 덧셈 과정에서 부동소수점 오차 누적
- **조건수 증가:** 관측 차원(n_{obs})이 상태 공간 차원보다 클 때 ($n_{\text{obs}} \geq 33$), 공분산 행렬의 condition number가 증가하여 수치적 불안정성 가속화
- **Backward pass 전파:** Forward pass의 손상이 backward pass로 전파되어 smoothed factors (EZ, EZZ) 손상

EM Algorithm의 수렴성 문제:

- **M-step의 행렬 역행렬 계산:** 요인 적재 행렬 C 업데이트 시 $C_{\text{new}} = \text{solve}(sum_EZZ^T, sum_yEZ^T)^T$ 계산 필요
- **Ill-conditioned 행렬:** 손상된 EZZ 가 sum_EZZ 에 포함되면 eigendecomposition 실패, condition number가 무한 대에 가까워짐
- **수렴 실패:** 결과적으로 C matrix가 NaN으로 채워지고, EM algorithm이 조기 수렴함

이론적 의미 이 문제는 DFM의 근본적 한계를 보여줌:

- **선형 가정의 취약성:** 선형 요인 적재 행렬 C 는 고차원 관측 공간에서 저차원 요인 공간으로의 선형 매핑을 가정함. 관측 차원이 증가할수록 이 매핑의 추정이 어려워짐
- **수치적 정확도와 확장성의 트레이드오프:** 부동소수점 연산의 근본적 한계로 인해 완벽한 수치적 정확도는 불가능하며 [?], 더 많은 시계열을 사용할수록 수치적 오차가 누적됨

- **데이터 품질 의존성:** 높은 상관관계, 극단적 결측치, 작은 유효 표본 크기 등 데이터 품질 문제가 수치적 불안정성을 악화시킴

이러한 한계는 비선형 인코더를 사용하는 DDFM과 같은 대안적 접근법의 이론적 정당성을 제공함. DDFM은 선형 매핑의 제약을 완화하고, 딥러닝 기반 요인 추출을 통해 수치적 안정성 문제를 우회할 수 있음 [?].

해결 방법 수치 안정화를 위해 다음과 같은 기법을 적용함:

1. Robust Statistics 접근법:

- **EZZ의 NaN/Inf 제거:** sum_EZZ 계산 전 손상된 시간 단계의 EZZ를 0으로 대체하여 제외
- **이론적 근거:** Huber (1981)의 robust statistics 이론에 따라 outlier나 손상된 관측값을 제외하고 정상적인 관측값만으로 추정하는 것은 통계적으로 타당함 [?]
- **시계열 특성:** 시간적 의존성이 있으므로 일부 시간 단계가 손상되어도 다른 시간 단계의 정보로 보완 가능

2. 수치 선형대수학 기법:

- **사전정규화:** 조건수 10^8 이상인 경우 적응적 정규화 적용 [?]
- **R 행렬 최소값 강제:** 10^{-4} 로 설정하여 innovation covariance의 수치 안정성 향상
- **공분산 행렬 대칭성 강제:** 모든 공분산 행렬 업데이트 후 즉시 대칭성 강제 [?]
- **Pseudo-inverse 사용:** solve 실패 시 fallback으로 더 robust한 해 제공
- **F Matrix Regularization:** 관측 차원 기반 적응적 정규화

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

- **KOIPALL.G:** C matrix NaN 100% → 0%, log-likelihood -3102.04 → 356.35
- **KOWRCCNSE:** C matrix NaN 94.9% → 0%, log-likelihood -2671.57 → 817.54
- **KOEQUIPTE:** 이전부터 안정적, 해결 후 정상 훈련

이 결과는 다음과 같은 시사점을 제공함:

- **DFM의 실용적 한계:** 수치 안정화 기법으로 문제를 우회할 수 있으나, 근본 원인(Kalman filter의 재귀적 오차 누적)은 해결되지 않음
- **DDFM의 정당성:** 비선형 인코더를 사용하는 DDFM은 선형 매핑의 제약을 완화하고, Kalman filter의 재귀적 공분산 업데이트에 덜 의존하므로 수치적 안정성 문제를 우회할 수 있음
- **모형 선택 기준:** 많은 시계열($n_{obs} \geq 33$)을 사용하거나 데이터 품질이 낮은 경우, DDFM이 더 안정적인 대안이 될 수 있음

b. 실험 설계의 제한사항

- **훈련-예측 간격:** 4년 간격으로 COVID-19 제외 및 데이터 누수 방지, 그러나 최신 경제 패턴 반영 제한
- **테스트 데이터 부족:** 80/20 분할 후 각 수평선당 단일 테스트 포인트 → 통계적 신뢰성 제한
- **Nowcasting 제한:** Release date 정보 정확성, ARIMA/VAR 마스킹 구현 제한

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**