
실시간 경기진단 시스템 구축 해외 출장보고서

2025. 12. 22

KDI 경제전망실

1 출장 배경 및 목적

- 데이터 기반 실시간 경기진단 시스템 구축을 위해 우수 해외기관의 전문가를 방문하여 자문을 구하고자 함.
 - Federal Bank of New York (이하 New York Fed) 거시·금융 데이터 요인모형, 나우캐스팅·예측 성능, 머신러닝 기반 차원축소 및 동적 요인모형 설계에의 이론적·실증적 합의 자문을 구하고자 함.
 - Columbia University의 Serena Ng 교수는 요인 모형 등 시계열과 Factor Model의 대가이자, New York Fed 내 나우캐스팅 등을 수행하는 Applied Macro and Metrics Center(AMEC) advisor로 실시간 경기진단 모형 구축에 대한 자문을 구하고자 함.
 - 대부분의 전망기관에서 New York Fed 구축한 나우캐스팅 모형을 벤치마크로 삼고 있음.
 - City University of New York의 윤수인 교수는 대규모·비정형 자료 처리 및 복합 데이터 구조 분석 경험이 풍부함. 경기진단 시스템 운영 측면에서 비정형 자료(unstructured data) 활용 방안, 공간 시계열 데이터 통합분석 전략, 고차원 데이터 분석(high-dimensional data analysis)의 활용 방안에 대하여 자문을 구하고자 함.

2 출장 개요

- 출장지 및 기간
 - 출장지: 미국 뉴욕
 - 출장기간: '25년 12월 8일(월) ~ 12월 12일(금), 3박 5일

- 출장자 (총 3명)
 - 경제전망실 정규철 경제전망실장
 - 재정사회정책연구부&경제전망실 이승희 연구위원
 - 재정사회정책연구부 임영주 전문연구원

- 주요활동
 - New York Fed 거시경제 및 통화연구팀(Macroeconomic and Monetary Studies) 면담
 - Columbia Univ. Serena Ng 교수(Edwin W. Rickert Professor of Economics) 면담
 - CUNY 윤수인 교수(Zicklin School of Business) 면담

3 출장 일정

일자	주요일정(안)
12.8(월)	[이동] 인천(10:10)→뉴욕(현지 시각 10:00 도착)
12.9(화)	[10:00-12:00] Columbia University의 Serena Ng 교수 면담 [17:00-18:30] Morgan Stanley(Rong Fan 박사 등)와 간담회
12.10(수)	[10:30-12:30] CUNY 윤수인 교수 면담 [14:00-15:00] New York Fed 거시경제 및 통화연구팀 면담 [18:30-20:00] JP Morgan Chase(MJ Shin 박사 등)와 간담회
12.11(목)	[이동] 12:00 뉴욕 출발 ~ 익일 17:45 인천공항 도착

4 활동 내용

1. Columbia Univ. Serena Ng 교수 면담

□ 일 시: 2025. 12. 9. (화), 10:00 - 12:00

□ 장 소: Serena Ng 교수 연구실

□ 주요 내용

○ 요인(factor)에 대한 논의

- 전통적인 기존 요인모형이 가정하는 pervasiveness 조건(factor loading 이 $O(N)$ 인 점)이 실제 거시·금융 데이터에서는 지나치게 강할 수 있음.
- 실제 데이터에서는 특정 요인이 일부 산업이나 섹터에만 영향을 미치는 경우가 빈번하며, 이러한 요인은 전체 횡단면에 강하게 퍼지지 않더라도 경제적으로 중요한 정보를 포함
- 따라서 기존 요인모형은 의미 있는 요인을 이론적으로 배제하거나 잡음으로 오인할 위험이 있음.
- 그러나 이러한 강한 혹은 약한 요인에 대한 구분은 이론적 토대에 마련한 것이고 실제 연구에서 목표하고자 하는 모형 예측력 제고와는 다른 문제이며 모형을 풍부하게 하는 것도 중요하지만 오차의 동학이 더 중요한 문제로 판단됨.

○ 데이터 선별에 대한 논의

- Boivin and Ng(2006)에서는 잡음이 많은 변수가 요인 추정에 장애요인으로 작용할 수 있음을 강조하며 실제 분석에서 적절한 변수를 선별하는 것의 중요성을 강조
- 그러나 최근 연구들이나 약한 요인(e.g., Bai and Ng, 2023)에 대한 연구들은 약한 신호를 가지는 변수라도 예측에 도움이 될 수 있는 가능성을 제시
- 따라서 예측력 제고를 위해서는 명백히 예측에 유용한 정보가 거의 없는 변수들은 사전에 제거하되 다른 변수들에 대해서는 가장 실무적으로 많이 쓰이는 변수 조합들의 horse race를 통해 적절한 변수를 선택하는 과정이 반드시 필요

○ 빅데이터 관련 논의

- Boivin and Ng(2006)의 연구는 시계열 예측에 있어 많은 데이터가 예측에 항상 좋지는 않다는 중요한 통찰을 제공한 연구
- 그러나 예측에 필요한 서로 다른 정보를 가지고 있는 데이터들은 물론 많을수록 예측에 도움이 될 수 있고 반대로 중복된 정보가 많다면 이러한 데이터들을 포함하는 것은 예측에 크게 도움이 되지 않을 가능성이 높음.
- 따라서 데이터는 마찬가지로 데이터 조합들의 horse race를 통해 예측에 필요한 데이터와 그 범위를 설정하는 것이 바람직하다고 판단됨.

○ 혼합주기 문제 관련

- 나우캐스팅에서 혼합주기 문제(e.g., 분기별 타겟 변수를 예측하기 위해 월별 변수를 설명변수로 이용하는 것)를 해결하는 것은 필수적임.
- 이에 대해서는 상태공간모형을 이용하거나 MIDAS 등의 방법을 활용할 수 있으나 Ng and Scanlan(2024)의 방법을 활용하는 것도 나우캐스팅 과제 수행에 큰 도움이 될 것으로 판단됨.
- 구체적으로 이 방법은 저빈도 지수를 타겟 변수로 설정하고 이와 관련 있는 고빈도 데이터에서의 정보를 저빈도 지수의 결측치로 처리하고, 고빈도 데이터에서 추정된 다중 요인으로 보완
- 이는 현재 수행 중인 나우캐스팅 과제에서도 즉시 적용할 수 있는 방안으로 보이며 이를 바탕으로 더 다양한 고빈도 자료를 바탕으로 나우캐스팅을 수행할 수 있을 것으로 기대됨.

2. CUNY 윤수인 교수(Zicklin School of Business) 면담

□ 일 시: 2025. 12. 10 (수), 10:30 - 12:30

□ 주요 내용

○ 일반적인 모형 설정 관련

- 지역별 경제활동 나우캐스팅의 시공간 모형화
- 전국 단위 예측에 활용되는 Dynamic Factor Model(DFM)을 기본 골격으로 두고, 지역 간 공간적 상관관계(spatial correlation)를 결합하는 접근이 비교적 자연스러운 설정으로 판단됨.

- 인접 지역 및 공간 가중행렬 정의

- 지리적 인접성 외에도 교통·산업·물류 연관성을 고려할 수 있으나 지역 간 의존성은 spatial correlation 구조로 처리하고, 교통·산업·물류 요인은 공변량으로 포함하는 방식이 해석력과 안정성 측면에서 적절함

○ 구체적 모형 설정 관련

- 비선형성 및 구조적 변화 반영

- regime switching, time-varying parameter(TVP), 비선형 state-space 모형 등이 활용 가능
- 지역별 이질성이나 구조 변화는 계층적(hierarchical) 구조를 통해 부분적으로 반영

- 공간·시간 의존성 결합 절차

- 시간 차원의 동학 구조(AR, state-space, factor dynamics)를 우선 설정
- 이후 spatial lag, spatial error, spatial random effects 형태로 공간 상관구조를 결합하는 단계적 접근이 일반적

- 계산 부담 문제

- 대규모 시공간 모형은 계산 비용이 큼
- 시·도 단위 자료가 제한적인 경우, full spatio-temporal model보다는 DFM에 spatial random effects를 결합하고 spillover 효과는 오차항에서 spatial correlation으로 처리하며 오차항 u_t 에 시간적 상관을 추가하는 계층적 구조가 현실적인 대안으로 판단됨.

- 공간 가중행렬 W 는 지리적 인접성 또는 물류량 기반 weights를 선택적으로 적용
- 딥러닝 기반 시공간 모형
 - GNN 등 spatio-temporal deep learning 모델이 제안되고 있으나 해석 가능성 문제로 정책 분석 목적의 실증 연구에서는 아직 제한적으로 활용됨.
- 결측치 및 불균형 시계열 처리
 - 단순 보간 외에도 state-space 또는 factor 기반 접근을 통해 결측을 자연스럽게 처리 가능
- 모형 결과 및 연구 수행 관련
 - 시공간 충격 전이 분석
 - 시공간 모형 기반 impulse response analysis를 통해 지역 간 충격 확산을 분석 가능
 - 정책 시나리오 분석
 - 특정 지역 정책 변화가 타 지역 경제활동에 미치는 spillover 효과를 시뮬레이션 가능
 - 예측력 평가
 - 지역별 예측오차(RMSE, MAE) 비교 및 전체(가중) 평균 예측오차를 활용한 종합 평가가 일반적
 - 참고 자료 및 연구 도구
 - Univ. of Missouri의 Christopher K. Wikle 교수의 시공간 모형 lecture notes 추천
 - 전통적 시공간 통계·계량모형은 R, 머신러닝·딥러닝 기반 모형은 Python이 주로 활용됨.

3. New York Fed 거시경제 및 통화연구팀(Macroeconomic and Monetary Studies) 면담

□ 일 시: 2025. 12. 9. (화), 14:00 - 15:00

□ 장 소: New York Fed 내 회의실

□ 면담자: Geert Mesters(Economic Research Advisor), Martin Almuzara(Research Economist), Donggyu Lee(Research Economist)

□ 주요 내용

- NY Fed Nowcasting Model 개발 배경 및 핵심 특징
 - 미국 GDP 성장률의 장기 하락 추세 반영 위해 시간 추세(time trend) 포함.
 - COVID-19 시기 급격한 변동·이상치(outliers) 대응 필요 →
 - 시간가변적 분산(time-varying variance) + 팬데믹용 특수 요인 추가
 - 기존 모델은 팬데믹 충격을 설명하기 어려워 2.0 버전으로 전면 개편
- 모델의 확장성 및 기계학습 도입 논의
 - 전통적 거시변수만 사용할 경우 비선형 모델(Machine Learning)의 이점은 제한적
 - 텍스트·감성지수 등 비정형 데이터 포함 시 비선형 모델의 효과가 커질 가능성
 - 그러나 ML 모델은 장기추세·공통요인·이상치 구조 등을 기본적으로 반영하지 않아 기존 Nowcast 모델의 특징을 다시 구현해야 하는 부담이 큼.
 - 데이터 길이가 짧은 거시 시계열에서는 ML의 과적합·훈련 어려움 존재
- 모델 운영 방식 및 자동화
 - 매주 금요일 자동 업데이트 시스템으로 Nowcast 산출
 - 내부 데이터베이스(FRED·Haver 등)에서 자동 호출 → 업데이트 비용 매우 낮음.
 - 외부에 제공되는 코드는 재현가능성 높음(일부 비공개 데이터 제외).
 - 정책 담당자에게는 수치뿐 아니라 업데이트 기여도(decomposition) 함께 제공

○ 한국 적용 관련 논의

- 한국은 소규모 개방경제 → 교역상대국·환율·원자재 가격 등 글로벌 변수 포함 필요
- 한국의 신용카드 일별 데이터 활용 가능성에 강한 관심
- 다만 모델은 월별·분기별 구조 유지가 안정적 → 일별 데이터는 주별·월별로 집계해 포함하는 방식 권고
- 변수 선택(variable selection)이 성능의 핵심 → 과거 예측력을 기준으로 변수 조합 테스트 필수

○ 향후 개발 방향(미국 모델)

- GDP 구성요소별(nowcast of components: 소비·투자 등) 확장 추진 중
- 안정적 운영 버전 유지하며 새 버전은 별도 실험·검증 중
- 인플레이션 분석용 별도 모델(MCP)도 운영하며 월 단위 자동 업데이트