

공식발표 통계지표의 적시성 확보를 위한 대안 데이터 파이프라인 구축제안⁺

(Proposal an Alternative Data Pipeline to Secure the
Timeliness for Official Statistical Indicators)

조 용 복¹⁾, 김 도 원^{2)*}
(Yongbok Cho and Dowan Kim)

요 약 본 연구는 공식발표 통계지표의 적시성 확보를 위해 기존 Nowcasting 방법론을 살펴보고 실시간 경기 현황 분석이 가능한 Real-time nowcasting 모형을 운용하기 위한 대안 데이터와 그 수집 체계를 점검한다. 공공영역과 민간영역에서 경기지표를 예측할 수 있는 고빈도 실시간 데이터를 탐색하고, 나아가 데이터의 수집, 가공, 모형화를 위한 클라우드 기반의 구축과정을 제안한다. 더불어 Real-time nowcasting 모형 추정 및 데이터 관리에 있어 고려해야 할 요소를 확인함으로써 적시성 및 안정성을 갖춘 공식 통계지표의 예측 프로세스를 제시한다.

핵심주제어: 공식 통계, 실시간 경제지표, Nowcasting, 대안 데이터, 파이프라인

Abstract This study provides a comprehensive analysis of recent studies conducted on the topic of nowcasting in order to enhance the accuracy and promptness of official statistical data. Furthermore, we propose an alternative approach involving the utilization of real-time data and its corresponding collection methods to effectively operate a real-time nowcasting model capable of accurately capturing the current economic condition. We explore high-frequency real-time data that can predict economic indicators in both the public and private sectors and propose a pipeline for data collection processing and modeling that is based on cloud platforms. Furthermore we validate the essential elements required for the implementation of real-time nowcasting, as well as their data management protocols to ensure the reliability and consistency needed for accurate forecasting of official statistical indicators.

Keywords: Official statistics, Real time Economic and Monetary Union indicator, Nowcasting, Alternative data, Pipeline

* Corresponding Author: abtop@hansung.ac.kr

+ 이 논문은 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. RS-2023-00214172).

Manuscript received August 28, 2023 / revised September 24, 2023 / accepted October 01, 2023

1) 동아대학교 경영정보학과, 제1저자

2) 한성대학교 사회과학부, 교신저자

1. 서 론

각국의 공식 통계생산기관(통계청, 중앙은행 등)은 여러 가지 경제지표를 필요한 주기별로 집계하여 공표하고 있다. 그중 국내총생산(GDP: Gross Domestic Product), 산업생산지수(IP:

Industrial Production Index) 등은 해당 국가 경제의 활동성 및 경기 국면을 판단할 수 있는 중요한 공식 통계다.

이와 같은 경제활동 지표는 통화정책(기준금리의 결정 등) 및 재정정책(조세 결정, 국채발행량, 산업별 공공지출 금액) 결정에 가장 중요한 기준이 되며, 가계와 기업 등 주요 경제 주체들의 의사 결정(설비투자 규모, 인력 채용 등)에 핵심 지표로 활용되고 있다(Banbura, M., 2013). 따라서 해당 지표에 대한 정확하고 적시성 있는 공식 통계치의 생산은 국가 경제 및 개별경제 주체들에게 올바른 정보를 정확한 시간에 제공해 사회 시스템 안정을 유지하는 데 매우 중요한 역할을 담당하고 있다.

앞서 언급한 공식 통계지표의 생산을 위해서는 경제 주체들의 활동에 관한 광범위하고 정확한 데이터가 요구되며, 따라서 많은 경제 주체들의 활동 결과물에 대한 정교한 집계 작업이 요구된다. 이에 정확한 공식 통계치의 공표(확정치 발표)까지 상당한 시간이 소요되고 있다. 예를 들어 분기별로 생산되는 GDP의 경우 확정치의 생산까지는 1년 이상의 시간이 소요되는 문제를 가지고 있다(Statistics Korea, 2021).

공식 통계자료의 발표 시차에 의해 발생하는 경제지표 정보의 불확실성은 경제시스템 내 참가자들에게 많은 혼란을 일으킬 수 있다. 따라서 민간에서는 각 경제전망 기관 연구자와 금융기관 연구 인력으로 이루어진 풀(Pool)을 활용하여 해당 지표에 대한 전망치를 조사(Survey)하고 그 대표 값을 예상치로 사용하는 등 다양한 방법으로 공식지표 발표 시차 발생에 의해 생겨나는 불확실성을 해소하기 위해 노력하고 있다. 나아가, 각국의 중앙은행은 적시성(Timeless)을 갖춘 예측치를 제공하기 위해 '실시간 경제전망' (Nowcasting) 모형에 관한 연구를 2000년대 중반부터 시작해 현재까지 활발히 진행 중이다(Giannone, D. et al., 2008; Khan, M. A. A., 2022).

기존 Nowcasting 모형에 관한 연구는 분기별로 발표되는 GDP의 추정을 위해 월 단위로 발표되는 경제지표 (산업생산지수, 경기 선행 및 동행지수) 등을 사용한 전통 계량 경제학 모형

을 중심으로 발전하였다. 하지만 이와 같은 모형을 추정하기 위해서는 입력변수인 월별 경제지표의 정확한 데이터(확정치)가 필요하다. 하지만, 대부분의 전통적 분기별 모형에서 입력변수로 사용되는 월별 지표의 확정치 발표까지 2개월 이상의 시간이 소요되는 등 분기별 모형과 동일한 공표 시차 발생의 문제를 가지게 된다(Statistics Korea, 2021).

기존 Nowcasting 방법론들은 제약조건이 많은 계량 경제학 모형[가. 단일방정식 모형: 1. 교량 방정식(BE: Bridge Equation), 2. 혼합주기 모형(MIDAS: Mixed Data Sampling) 및 나. 다변수 동적 모형: 1. 동적 요인모형(DFM: Dynamic Factor Model), 2. 벡터 자기 회귀모형(VAR: Vector Auto Regression)]을 기반으로 한 선형 모형을 활용한 연구가 주를 이루었다(Yi, et al., 2022).

최근 기계학습(Machine Learning)모형을 적용하여 경제 변수 간 비선형 상관관계를 반영하기 위한 연구가 제안되고 있으나, 입력변수로서 기존 경제지표만을 활용하는 한계를 벗어나기는 힘든 상황이다. 또한, 데이터의 시계열 관측치 개수보다 머신러닝 모형에서 추정해야 할 모수가 많을 경우 (신경망 모형 등) 모형 적용과정에 한계를 나타내고 있다(이연희 외, 2022).

따라서, 본 연구에서는 앞서 제시한 공식 통계 중 특히 GDP(분기별 공표)와 산업생산지수(월별 공표)에 대해 Nowcasting 모형을 효과적으로 운용하기 위한 실시간 대안(Real-time Alternative) 데이터를 살펴보고, 이러한 데이터를 한곳에 집중하는 실시간 데이터 레이크(Real-time Data Lake)제안하였다. 그리고 이를 지속적 업데이트하여 지속 가능한 공식 통계의 실시간 추정체계(Real-time Nowcasting System)를 확보하기 위한 데이터 파이프라인(Data Pipeline)을 제시하고자 한다.

2. 이론적 배경

공식발표 GDP의 경우 집계 주기가 통상 분기별로 이루어지며 정확한 확정치의 생산까지는 1

년 이상의 시간이 소요되어, 이를 보완하기 위해 해당 분기가 종료된 시점을 기준으로 28일 이내에 속보치를, 또한 70일 이내에 잠정치를 발표하고 있다(Statistics Korea, 2021).

GDP를 이용한 성장률 지표는 한 국가의 경기 상황을 판단할 수 있는 가장 중요한 지표로 정부의 통화정책 및 재정정책의 기준이 되며, 국가 경제시스템 내 경제 주체들의 중요 의사 결정에 결정적 벤치마크 역할을 하고 있다. 실제로 금융시장에서는 GDP의 속보치, 잠정치 및 확정치가 발표되는 시점에 기존의 경제 주체들의 기대치와 시간이 지남에 따라 공표되는 공식 통계치, 즉 속보치와 잠정치 간의 차이를 보정하기 위해 발표 시점과 그 직후, 해당 통계지표와 관련된 많은 금융 상품들의 변동성이 커지는 경우가 자주 발생한다.³⁾

이와 같은 문제점을 해결하기 위해 OECD 및 미국, 영국 등 선진국 통계기관에서는 잠정치와 확정치의 차이에 대한 분석 (Revision Analysis)을 실시하고, 여러 가지 통계적 기법을 이용하여 잠정치와 확정치의 오차를 줄이고자 노력하고 있다(Kim, 2006). 또한, 여러 번의 글로벌 금융위기를 겪으며 거시경제 안정성 확보를 위한 경기조절 정책이 세계 경제 안정을 확보할 수 있는 중요한 수단으로 부상되며 시의성(Timeliness) 있는 공식 경기지표의 집계 및 추정은 더욱 주목받게 되었다. 이에 따라 각국의 중앙은행과 학계에서는 적시성 있는 공식 통계의 예측치를 제공하기 위한 '실시간 경제전망'(Nowcasting) 모형에 관한 연구가 활발히 진행되었다.

2.1 전통적 Nowcasting 방법론

Nowcasting에 관한 연구들은 각 국가의 GDP 및 그 성장률을 시의성 있게 추정하기 위해 GDP와 동일한 발표주기를 가지는 분기 데이터 및 그보다 발표주기가 짧은 월별 데이터를 이용

하여 전통적 계량 경제학 모형을 적용하는 방식으로 발전되었다. 이러한 모형들은 단일방정식 모형으로 지칭되는 교량 방정식(BE) 모형 및 혼합주기모형(MIDAS)과 다변수 간 상호작용 및 설명변수들의 동태적 움직임을 고려한 동적 요인모형(DFM) 및 벡터 자기 회귀모형(VAR)으로 구분할 수 있다.

교량 방정식 모형은 GDP를 설명하는 월별 거시경제 변수를 개별적으로 예측(Forecast)하고, 이 예측값들을 이용해 해당 분기의 GDP를 예측하는 모형이다. 또한, 혼합주기모형은 분기별 GDP 성장률 예측을 위해 속보성이 높은 월별 거시경제 지표(지출 부문 또는 생산부문별)를 분기화 시키지 않고 직접 단일방정식 또는 연립 방정식의 설명변수로 모형화하는 방법이다. 따라서 혼합주기모형은 교량 방정식 모형 이용 시 수반되는 월별 결측치 추정과정의 예측 오차를 줄여 경제전망에 대한 예측력이 향상되는 것으로 알려져 있다(Clements and Galvao, 2008; Lee et al., 2017; Lee, 2021).

다변수 모형 중 동적 요인모형은 다수의 월별 거시경제 데이터를 사용하고 통계적 기법인 주 성분 분석(PCA: Principal Component Analysis) 등을 사용하여 자료들 간의 상호의존 관계를 소수의 요인으로 요약한 후 해당 요인을 이용하여 GDP 성장률을 예측하는 모형이다(Stock and Watson, 2002). 이 모형은 정보의 손실을 최소화하며 모형에 사용된 데이터의 주기 불일치(Frequency Mismatch) 및 결측값의 문제를 해결하고 입력변수의 임의성을 피할 수 있다는 장점이 있어 주요국 중앙은행의 경기지표 예측 시 많이 활용되는 모형이다(Yi et al., 2022). 또한, 여러 변수 간 상관관계와 시차를 모형화할 수 있는 대표적인 계량경제 모형인 벡터 자기 회귀 모형을 기반으로 한 모형은 거시경제 변수뿐만 아니라 그보다 관측주기가 짧은 금융변수를 사용하여 경제 변수 간 시차에 따른 상관관계를 반영하여 경기상황을 판단하는 방법으로 발전되어 왔다(Rovertson and Tallman, 1999; Sung, 2001).

3) 특히 경기상황에 대한 기대와 통화정책에 민감한 금리(채권 등) 시장은 해당 통계치들이 발표되는 시점에 해당 거시경제 변수에 영향을 많이 받는 자산가격이 민감하게 반응하는 경향을 보인다.

Table 1 Nowcasting Model of Major Countries

Country	Model
US	DFM with PCA (Weekly)
UK	MIDAS,DFM
Euro(ECB)	DFM, BVAR(Bayesian VAR)
Germany	DFM
Japan	BE,MIDAS, factor-MIDAS
Canada	BE, MIDAS, DFM, BVAR
Korea	DFM

Reference: Yi, et al. (2022) and Lewis, D. J. (2022)

각국의 중앙은행은 해당 모형들을 이용하여 당분기 경제전망을 고도화하기 위한 노력을 계속 기울이고 있으며, 각 중앙은행에서 사용하고 있는 주요 Nowcasting모형은 Table 1에서 주어진 바와 같다.

2.2 전통적 Nowcasting 데이터

앞서 제시된 전통적 Nowcasting 모형들은 주로 분기별로 공표되는 GDP 성장률의 공식발표에 대한 시차 개선을 위해 월별로 집계되는 거시경제 변수 및 실시간 업데이트되는 금융시장 변수를 주로 사용하여 해당 지표의 분기 내 상황을 좀 더 정확하게 반영하고자 하였다.

Table 2와 같이 대표적인 주요 거시경제 변수로써 소비자 및 생산자 물가를 이용하여 물가 상승률을 반영하였으며, 수출입 물량지수 및 경상수지 데이터를 이용해 대외거래량을 경제 성장률에 반영하고자 하였다. 나아가 국내 및 세계경기 반영을 위해 국내 산업별 또는 OECD 국가별 산업생산 지수와 기업 심리에 대한 실적 및 전망 자료를 입력데이터로 사용하였다.

또한, 금융시장의 대표적인 실시간 지표인 국고채 및 회사채 금리, 나아가 이들의 장단기 스

Table 2 Input Variables for Nowcasting Model

Category	Variables	US	Japan	Korea
National Account	GDP		O	O
	Consumption	O	O	O
	Construction	O	O	O
Labor	Unemployment Rate		O	O
	Employment to Population Ratio			O
	Unemployment Initial Claims	O		
International Trade	Export	O	O	O
	Import			
	Export Price	O	O	O
	Import Price			
Consumption Price	Retail Sales	O	O	O
	Consumer Price	O	O	O
	Producer Price			O
Production Manufacturing	Manufacturing Price Index	O	O	O
	Inventory Index	O		O
Surveys	BSI by Industry	O	O	O
	Economic Sentiment Index	O	O	O
	Consumer Sentiment Index	O	O	O

프레드 자료와 주가 및 주택지수를 이용하여 경기 판단 변수(GDP)의 분기별 추정에서 발생하는 정보의 공백을 줄이고자 노력하였다.

2.3 전통적 Nowcasting 모형의 문제점

하지만 분기 또는 월별 거시경제 변수를 주로 사용하는 전통적 경기 진단 Nowcasting 모형들은 입력변수의 상이한 발표주기와 해당 월별 변수의 발표 시점의 불일치에서 발생할 수밖에 없는 몇 가지 문제점을 내포하고 있다.

첫째, 분기별 GDP 성장률 예측에 입력변수로서 발표주기가 짧은 월별 경제 변수를 이용할 경우 발생하는 주기 불일치 문제이다. 경제 이론에 입각한 계량경제 모형들은 입력변수들의 관측주기가 같은 것으로 가정하며, 그보다 짧은 공표 주기의 데이터를 모형에 함께 포함 시킬 경우 모형 추정에 문제를 일으키게 된다(Yi, et al, 2022). 주기 불일치의 문제를 해결하기 위해 월별 경제 변수의 평균, 중앙값 등 대푯값을 이용하여 분기별 데이터에 주기를 맞추는 방법이 있으나, 이는 월별 경제 변수가 가지고 있는 정보의 손실을 초래하는 단점이 존재한다. 반면, 월별로 존재하지 않는 분기별 데이터를 다양한 방법으로 보간(Interpolation)하면 합리적 대응치를 사용할 수 있지만 관측되지 않는 값을 추정 과정에서 발생하는 오류는 확인할 방법이 없다(Hwang, 2017).

둘째, 월별로 관측되는 입력변수의 공표 시점의 차이로 인해 분석 시계열의 마지막 시점이 변수별로 달라지는 자료의 불균형 문제이다. 일반적인 계량경제 모형들은 모형을 사용하고자 하는 시점에 어느 한 변수라도 결측값이 발생할 경우, 같은 시점에 관측된 다른 변수들도 입력값으로 사용하지 못하게 되는 시계열 완료 시점 불일치 데이터의 문제(Ragged-edge Problem)가 발생한다(Wallis, K. F., 1986; Ferrara, L. et al., 2010). 더욱이 월별로 관측되는 거시경제변수의 값마저도 확정치가 아닌 잠정치인 경우가 대부분이다. 이는 잠정치와 확정치가 순차적으로 변경될 때 발생하는 오차가 Nowcasting 모형의 오차를 확대시킬 수 있어 각 전망 시점별로 입

수 가능한 데이터셋, 즉 빈티지 데이터셋(Vintage Dataset)을 사용하여 모형을 추정하고 평가해야만 한다(Croushore, and Stark, 2003; Yi, et al., 2022).

셋째, 모형에 사용되는 입력변수가 많을 경우 차원의 저주(Curse of Dimensionality)의 문제가 발생하기 때문에 예측하고자 하는 경기변수에 영향을 미치는 최적의 변수를 선택해야 하는데, 이 과정에서 많은 가정과 연구자의 주관적 판단에 의존하게 되는 문제가 발생한다. 변수의 선택 방법별 입력변수의 선택 편의와 거시경제적 관점에서 연구자가 생각하는 변수 간 관계의 선택적 편향이 작용할 수밖에 없는 것이다.

넷째, 변수 간 상호의존관계를 모형에 반영하는데 있어 선형적인 관계만 모형화 가능하다. 이는 경기상황에 따라 변화하는 거시경제 변수의 비대칭적 상관관계의 반영이 불가능하여 호황 시와 경제위기 시 과소 또는 과대 추정의 오차가 발생하게 된다(Smith and Vahey, 2016; Cho and Lee, 2022).

2.4 대안 데이터를 이용한 Nowcasting

앞서 제시한 문제점을 보완하기 위하여 최근 한국은행과 금융감독원에서는 기존의 동적 요인 모형(DFM)과 다양한 머신러닝 기법을 빅데이터에 적용해 GDP 성장률 등 공식 통계 데이터에 대해 좀 더 정확성을 확보하고 시의성 있는 정보를 제공하고자 다양한 시도를 하고 있다(Choi, 2019; Yi, et al., 2022). 학계에서도 최근 데이터 집적기술의 발달로 증가하고 있는 다양한 대안 데이터를 활용한 연구들이 증가하고 있다. Park(2020)은 한국거래소의 산업별 전력사용량 자동 검침(AMR: Automated Meter Reading) 데이터를 이용하여 산업별 생산지수를 예측하기 위한 연구를 진행하였다. 또한, Lim and Choi (2021)은 분기 실질 GDP Nowcasting을 위해 경기심리지표(기업경기실사지수, 소비자 동향지수)를 이용한 2단계 추정방법을 사용했으며, 이러한 경기심리지표를 추정하기 위해 뉴스 기사에 대한 텍스트 마이닝을 실시하고 딥러닝 기법을 적용하였다.

또한, 각 중앙은행 역시 공식 거시경제 데이터의 집계 시간에서 발생하는 오차를 줄이기 위해 실시간 데이터와 머신러닝 기법을 활용한 Nowcasting 모형의 도입을 적극적으로 검토하고 있다. 미 연방은행 (FRB: Federal Reserve Bank of New York)은 전통적 경제통계인 실업급여청구 데이터와 대안 데이터로서 전력사용량 정보를 이용하여 주(Weekly) 단위 경기 전망 모형을 구축하였다(Lewis et. al, 2021). 일본 중앙은행은 GDP 추청을 위해 주간 소매 판매와 일별 인터넷 검색량 데이터를 활용하였고(Nakazawa, 2022), 월별 산업 생산(Industrial Production)을 추정하기 위해 시간대별로 관측되는 스마트폰 모빌리티 데이터와 전력지수를 사용하여 Nowcasting 모형을 개선하였다(Furukawa et. al., 2022).

더불어 Nowcasting 모형의 입력변수 간 비선형적 상호의존관계를 반영하기 위해 다양한 기계학습 모형을 적용하여 대표적인 전통적 Nowcasting 방법론인 동적 요인모형(DFM)의 성능을 향상하기 위한 연구가 지속되고 있다(Soybilgen and Yazgan, 2021; Lee, 2021; Longo, Riccaboni, and Rungi, 2022; Hopp, 2022).

3. 실시간 데이터 파이프라인

본 장에서는 공식 통계의 적시성 확보를 위한 Nowcasting 모형 운용에 필요한 입력변수로서 실시간 대안 데이터를 살펴본 뒤 이를 확보하고 안정적으로 공급하기 위한 데이터 파이프라인 및 데이터 레이크 구축 방법을 제안한다.

3.1 실시간 대안 데이터

2.2절에서 살펴본 바와 같이 기존의 Nowcasting 모형은 공식발표 통계자료를 이용하거나 실시간으로 집계되는 금융시장 데이터를 사용하여 적시성을 확보하고자 노력하였다. 하지만 최근 ICT (Information and Communications Technologies) 기술의 발전과 실시간 데이터 저장 및 공유기능

의 강화로 경제지표 공식발표 전에 이를 실시간으로 추정하여 볼 수 있는 대안 데이터가 늘어나고 있다.

A. 공공영역의 실시간 데이터

빅데이터 집적기술의 발달과 공공성을 가진 데이터의 접근성을 확보하기 위하여 공공영역에서도 다양한 데이터가 실시간으로 집계되어 공개되고 있다. 이 중 경제지표의 Nowcasting 모형에 입력변수로 사용할 수 있는 데이터를 살펴보고자 한다.

첫째, 한국도로 공사에서는 고속도로 공공데이터 포털의 TCS(Toll Collection System)를 통해 전국의 고속도로 요금소 통과기준 데이터를 제공하고 있다. 고속도로 영업소 출구 통과 차량을 기준으로 6가지 차종별 통행량 정보를 실시간으로 집계하고 함께 집계정보도 일별, 월별로 시차 없이 제공하고 있다.⁴⁾

교통량은 경제지표 및 사회적 지표와 밀접한 연관성이 있는 것으로 밝혀져 있다. 구체적으로 Kim et al.(2010)은 30년간의 고속도로 통행량 및 충전장에 대한 연간 데이터를 사용해 연간 GDP와의 관계를 분석하여 통행량과는 0.987, 연장 통행량과는 0.870의 상관관계가 있는 것을 확인하였다. Chung et al.(2010)은 TCS 자료의 차종별 통행량이 주중과 주말, 그리고 휴일에 1종(소형) 차량과 2~5종(중대형 및 화물차 등) 차량이 다른 통행 패턴이 나오는 것을 확인했다.

교통량과 경제지표 간의 밀접한 관계가 있다는 기존의 연구결과는 있으나 연간 기준의 전통적 계량경제모형을 이용한 장기(저빈도) 모형이라는 한계를 가지고 있다. 하지만 TCS에서 제공하는 실시간 교통량 데이터를 이용하여 Nowcasting 모형의 적시성을 확보한다면 빠른 정책적 대응 및 경제 시스템내 참여자들에게 시의성 있는 경기판단을 위한 정보를 제공할 수 있을 것이다. 더욱이 TCS 데이터는 각 지역별 통행량을 실시간으로 파악할 수 있어 Lee(2021) 및 Lee(2022) 등이 지적하고 있는 연간 집계 지

4) 차종 구분은 1종(소형차), 2종(중형차), 3종(대형차), 4종(대형화물차), 5종(특수화물차), 6종(경형 자동차)이다.

역내총생산(GRDP: Gross Regional Domestic Product)의 문제점을 획기적으로 개선할 수 있을 것이라 기대한다.

추가로 일반국도, 국가지원지방도, 지방도의 교통량을 조사하여 통계정보를 제공하는 국토교통부의 교통량정보제공 시스템(TMS: Traffic Monitoring System)이 있다. 하지만 TMS의 데이터베이스에서는 연간단위의 집계정보만을 제공하고 있어 본 연구의 목적인 Nowcasting 모형의 적용에는 한계점이 존재한다. 하지만 TMS의 정보집계를 위해 실시간 교통량정보 수집장치인 AVC(Automatic Vehicle Classification) 및 VDS(Vehicle Detection System)를 운영하고 있어 실시간으로 데이터를 수집할 수 있는 파이프라인만 구축한다면 추가적인 투자 없이도 전국을 작은 지역으로 나누어 지역 내 활동량을 반영한 경제지표를 추정할 수 있을 것이다.

둘째, 산업통상자원부 산하 한국전력 거대소에서 운영하고 있는 전력통계정보시스템(EPIS: Electric Power Statistics Information System)의 전력사용 및 발전량 데이터를 활용할 수 있다. 해당 데이터베이스에서는 실시간 전력 수급 데이터와 산업별(용도별), 지역별 및 제조업종별 판매전력량을 제공하고 있다. 또한, 지역별 발전량 데이터도 함께 제공하고 있어 다양한 시각의 전력사용량을 이용한 Nowcasting 모형의 정확성을 높이는데 기여할 수 있는 실시간 데이터라 할 수 있다.

전력 및 에너지 소비와 경제 성장 간의 관계는 오래전부터 에너지 경제학 분야에서 중요한 연구 주제였으며 그 인과 관계가 충분히 검증되었다.⁵⁾ Park et al.(2016)은 1970년부터 2013년의 연간 GDP와 산업별 전력 판매량 데이터를 벡터 오차수정모형(VECM: Vector Error Correction Model) 및 Granger 인과 관계 검정을 이용해 산업별로 전력 소비와 GDP 성장 간

의 관계가 성장가설 및 상호의존 가설을 보임을 확인하였다. Park and Kim (2013) 또한 에너지 소비와 산업부문, 수송부문, 가정 및 상업 부문과의 관계를 VECM 모형을 사용하여 검증하였으며, 산업부문의 성장을 에너지 소비증가 자료가 선행해 증가하는 것을 밝혔다. 이외에도 전력 소비와 경제 성장 간 인과 관계를 규명한 다양한 연구결과들이 존재한다(Cho and Kang, 2012). 기존 연구결과들은 주로 장기간의 GDP 또는 기타 경제 성장률 지표와 전력사용량의 저빈도 (분기 이상) 자료를 이용하여 경제지표와 전력사용량 간의 장기적인 균형 관계와 인과 관계에 집중된 분석을 실시하였다. 하지만 최근 데이터 실시간 집적기술의 발달로 지역별, 산업별 전력 소모량을 실시간으로 모니터링 가능하게 됨으로써 기업의 생산활동에 대한 활성도를 빠르게 추정하여 볼 수 있게 되었다.

셋째, 해양수산부가 운영하는 ‘해운 항만물류정보시스템(PORT-MIS)은 국내 주요 선박종류별, 해외지역별, 선종별 실시간 입출항 데이터를 실시간(최소 5분 단위)으로 제공하고 있다. 항만별 정보를 구분하여 제공하고 있기 때문에 부산, 울산, 인천 등 주요 수출입 항만 지역의 지역별 총생산 및 산업생산 지표의 Nowcasting에도 유용하게 활용할 수 있는 정보라 할 수 있다. 항만의 물동량 정보는 내수 시장 규모가 협소하고 무역 의존도가 높은 국내 산업구조의 특성을 반영한 산업의 활동성에 대한 지표라 할 수 있으며, 이는 국내 경기상황을 판단할 수 있는 중요한 정보라 할 수 있다.

이에 항만의 물동량과 지역경제와의 인과 관계를 밝힌 경제학 관련 연구들이 이루어졌다. Mo and Lee(2013)은 지역별 항만 물동량과 산업생산지수와의 그랜저 인과 관계 검정(Granger Causality Test)을 통해 항만 물동량과 지역 산업생산지수 간에는 양방향 인과 관계가 존재한다는 것을 밝혀냈으며, Choi(2007)는 장기적으로 항만의 물동량과 지역 산업 성장의 관계가 유의함을 밝혀내었다. 더욱이 Ann and Lee(2017)은 항만 컨테이너 물동량과 세계 GDP를 대륙별로 분석하여 각 대륙별 컨테이너 물동량증가가 GDP 증가에 영향을 미치며 그 크기가 지역별

5) 에너지 소비와 경제 성장 간의 관계는 경제 성장이 에너지 소비를 유도한다는 보존가설(Conservation Hypothesis), 에너지 소비가 경제 성장을 유도한다는 성장가설 (Growth Hypothesis) 등 다양한 가설이 존재하나 본 연구의 목적은 Nowcasting 모형에 사용되는 입력변수로서 전력사용량의 가능성을 확인하는 것으로 성장가설 및 상호의존가설 주장한 연구 유사한 접근이라 할 수 있다.

로 다르다는 것을 확인하였고, Lee and Lee (2022)은 2011~2021년간 세계 GDP와 컨테이너 물동량 간의 상관관계가 90% 이상임을 보였다. 이처럼 항만 물동량과 거시경제지표와의 인과 관계는 충분히 검증되어 항만 물동량의 실시간 정보를 활용하여 Nowcasting 모형의 정확성을 높일 수 있는 근거를 제시하고 있다.

넷째, 고용노동부의 실업급여 청구 건수를 이용하여 취업 시장 및 노동시장의 기대를 반영할 수 있다. 고용노동부는 고용행정통계를 통해 산업별 취업 현황 및 실업급여 관련 데이터를 공개하고 있다. 통계청에서 집계 발표하고 있는 경기종합지수 중 경기선행지수의 경우 여러 가지 정보(재고 순환지표, 소비자자기대지수, 기계류 내수출하주수, 수출입물가비율, 구인 구직비율, 건설수주액, KOSPI 지수, 장단기 금리 차 등)를 기초자료로 산출되지만, 그중에서 특히 고용 관련 지수는 경기선행지수를 구성하는데 가장 중요한 항목으로 여겨진다.

이에 고용 및 일자리 현황과 경기지표와의 관

계에 관한 연구는 경제학의 세부 분야 중 경기 예측뿐만 아니라 노동경제학 분야에서도 활발하게 이루어졌다. 오쿤의 법칙(OKUN's Law)으로 잘 알려진 GDP 성장률과 고용증가율의 정(+) 상관관계(Wen, Y., and Chen, M. (2012))로부터 시작된 논의는 경기 성장과 고용 및 실업률 간의 인과 관계에 연구로 이어졌고, 두 가지 요소 간의 영향력 방향성에 대해서는 아직 다양한 의견들이 이어지고 있다(Couch, K. A., and Fairlie, R., 2010; Shin, 2014; Aragon F.M. et. al., 2018; Kim, 2021). 하지만 Table 3에서 보는 바와 같이 각국의 중앙은행은 경기상황을 판단 또는 예측하는 여러 가지 지표에 항상 고용 관련 데이터를 포함하고 있는 것을 볼 수 있다. 더욱이 미국의 '주간 신규 실업수당 청구건수: U.S. Initial Unemployment Claims'의 경우 통화정책의 중요한 지표로 활용되고 있어 주간 지표가 발표되는 순간에 실적치와 기대치의 차이를 조정하기 위한 금융시장의 변동성 증가 현상이 빈번히 관찰된다. 국내에서도 실업급여 지급액과

Table 3 Employment Indicators and Components in the Economic Leading Index

Country	Non-financial Index	Financial Index
OECD	-Working Hours per week in Manufacturing	
	-New Order for Durable Goods	-NYSE Index
	-Housing Starts Index	-Spread for Interest rates
	-ISM Index	
	-Consumer Confidence Index	
US	-Average Working Hours per week in Manufacturing	-S&P 500 Index
	-Unemployment Insurance Initial Claims	-Leading Credit Index
	-ISM Index	(Swaps, CDS Premium, Spread for Corporate Bonds)
	-Building Permits	- Spread for Interest Rates
	-Consumer Confidence Index	
Japan	-New employment	
	-Actual Machinery Orders	-TOPIX Index
	-Index of Tertiary Industry Activity	-Nikkei Index(42 type)
	-Producer Price Index	-Gross Rate for M2
	-Consumer Confidence Index	
Korea	-Employment to Population Ratio	
	-Construction	-KOSPI Index
	-Price Ratio Export per Import	-Spread for Interest Rates
	-Manufacturing Shipment Index	

지급자 수를 집계하여 월간 단위로 사업장 규모 별로 발표하고 있으나 해당 사업장의 지역 및 산업에 대한 정보와 결합한다면 기업이 바라보는 향후 경기상황을 좀 더 객관적으로 측정할 수 있는 지표로 활용 가능할 것이며 현재의 실업률을 발표하는 월간 단위의 고용 지표 산출체계를 개선할 수 있을 것이다.

다섯째, 실시간 항공 데이터를 이용하여 앞서 제시한 해양운송뿐만 아니라 항공을 이용한 산업체의 수송 및 여객 수송을 통한 경기 전망이 가능할 것이다. 일찍이 항공업을 중심으로 경기와 항공 여객 수요에 대한 연구가 많이 이루어졌다(Kim and Shin, 2019; Lim. et al, 2020.). 하지만, 기존 연구들은 항공 및 물류 산업의 관점에서 산업의 활성화에 영향을 미치는 요소로서 거시경제 변수 및 인플레이션과 유가 등을 사용한 연구들이었다.

최근 경기 전망을 위해 항공 감시 시스템(ADSB: Automatic Dependent Surveillance Broadcast) 데이터를 활용한 연구가 해외 학자들을 중심으로 시도되고 있다(Miller. et al., 2020). 국내에서는 항공 여객 및 화물 관련 통계가 국토 교통부에서 운영하고 있는 항공 정보 포털을 통해 일별 단위로 제공되고 있다. 하지만 이를 활용한 경기 전망에 대한 국내 연구는 부족한 상황이라 본 연구에서 제안하는 실시간 Nowcasting 데이터 파이프라인을 통해 항공 물류 실시간 데이터의 경제지표 예측 효과를 확인해 보고자 한다. 이는 실시간 수집되고 있는 항공 물류 데이터 활용에 새로운 영역을 개발하는 것이며 항공, 여객산업과 경기지표와의 관계에 관한 연구 분야에 새로운 시각을 제공할 수 있을 것이다.

앞서 제시된 실시간 자료 외에도 공공영역에서 생산되는 실시간 자료는 계속 늘어나고 있고 이를 활용한 실시간 데이터 파이프라인을 구축한다면 Nowcasting 모델의 성능향상을 기대해 볼 수 있을 것이다.

B. 민간영역의 실시간 데이터

앞서 살펴본 공공영역의 공개 데이터 외에도

민간영역에서 생성되고 있는 경제 주체들의 경제활동에 대한 실시간 데이터는 급격히 증가하고 있다. 이 중 경제지표의 Nowcasting 모형에 입력변수로 사용할 수 있는 대표적인 데이터를 살펴보고자 한다.

우선 현재 소비자들의 경제활동 지표를 확인하기 위해 실시간 카드사용 데이터를 활용할 수 있다. 2020년 5월 기준 여신금융협회의 조사에 따르면 국내 GDP의 약 50% 정도가 민간소비를 통해 발생하며 이 중 민간 최종소비지출의 72% 정도가 신용카드의 신용판매를 통해 이루어지고 있다고 한다. 체크카드 등 신용판매 방식이 아닌 지급 수단 이용까지 합치면 전체 민간소비 지급 결제에서 전자방식의 지급 수단의 사용이 90%에 달하고 있다(Seo, 2021). 현금보다는 카드에 의한 결제방식의 비중이 더 많은 것을 고려한다면 카드사용 및 전자결제를 통한 실적 데이터는 소매 경제활동을 측정할 수 있는 최적의 자료이다.

호주 통계청(Australian Bureau of Statistics), 유럽중앙은행(European Central Bank), 뉴질랜드 통계청(New Zealand Statistics) 등 각국의 중앙은행에서도 가계의 소비 동향 및 기업판매 지표 산출을 위하여 신용카드거래 실적 및 직불카드 거래 내역을 활용하고 있다. Arango. et al.(2011), Seo(2017) 및 Shin and Seo(2022) 등도 공식 통계인 경기지표 산출시 신용카드 및 직불카드의 실시간 정보가 유용할 수 있음을 확인하였다.

이처럼 신용카드 사용실적 데이터의 가치를 인정받기 시작하며 신용카드사들은 경쟁적으로 빅데이터 조직을 신설하고 자사 내 고객의 소비 행태와 가맹점별 매출 정보를 가공한 후, 이를 활용한 데이터 공급 및 컨설팅 사업에 적극적으로 진출하고 있다. 더욱이, 개별 카드회사의 데이터를 공공영역에서도 구입하여 익명 처리 등 개인 정보 보호에 대한 이슈를 처리한 후 활용할 수 있도록 공개하고 있어 그 활용도는 날로 높아지고 있다. 또한, 신용카드 정보와 더불어 POS(Point of Sale) 시스템에서 생성되는 지역별, 업종별 매출 통계 데이터가 금융데이터 거래소 등을 통해 수집할 수 있다. 나아가 국세청

등의 정보를 추가로 통합한다면 신용카드 사용에 누락 될 수 있는 현금영수증 및 현금거래로 인한 부가세 정보들이 추가될 수 있을 것이며, 이는 실제 경제 주체들의 경제활동을 가장 정확, 신속하게 반영할 수 있는 실시간 데이터가 될 것이라 기대한다.

앞서 소개한 민간데이터는 일정한 주기별로 집계되어 데이터베이스 형태(정형 데이터)로 구축할 수 있다. 또한, 웹 크롤링(Web Crawling) 및 스크래핑(Scraping) 등 별도의 데이터 수집(Data Collection) 기술을 활용한다면 텍스트(Text) 등 다양한 실시간 비정형 데이터를 수집하여 Nowcasting 모형의 고도화에 활용 가능할 것이다.

먼저, 검색엔진이 보유한 검색어 정보를 활용하여 실시간으로 경제 주체들의 관심사에 대한 정보를 수집하고 경기상황 및 불확실성 관련 정보를 추출하여 공식경제 지표의 공표 시차 발생 문제를 해소하는데 기여할 것이라 기대한다. 검색 정보의 경우 2000년대 초반부터 사람들의 행동 양식 및 관심사를 잘 반영하는 가치 있는 데이터로 알려져 있다. 이러한 검색 데이터는 2016년 초박빙의 미국 대선 결과에 대해 대부분의 여론 조사회사와는 다른 예측을 하였고, 이러한 예측은 적중한 것으로도 나타났다.

McLauren and Shanbogue(2011)은 영국의 실업률 예측에 있어 검색 서비스 회사의 “Job Seeker’s Allowance: JSA”의 빈도를 지수화하여 사용하였고, Choi and Varian(2012) 역시 미국의 실업청구자 수 (Initial Claims for Unemployment) 및 자동차에 대한 수요를 해당 회사의 검색 데이터를 사용하여 예측하였다. 국내에서도 Lee and Hwang(2014)과 Kwon et al.(2015) 등은 경기지수 및 실업률 예측을 위해 실업에 관련된 키워드(청년, 청년실업, 급여, 실업급여, 경제, 문제, 구직, 정부, 일자리, 기업 등)에 대한 국내 검색 서비스 회사의 검색 트렌드를 이용할 경우 기존 시계열 모형의 설명력이 증가하는 것을 보여 주었다.

두 번째로, 취업 사이트에서 이루어지고 있는 고용과 일자리의 실시간 추적을 통해 고용 주체, 즉 생산 주체들이 생각하는 경기상황에 대

한 기대치를 가늠해 볼 수 있을 것이다. 특히 대표적인 취업 포털 사이트인 잡코리아 및 사람인 등에서는 전체 산업별, 지역별, 채용형태별(정규직 및 계약직), 직급별(직무 성숙도의 정도) 채용정보를 수집하여 가공한 후 별도의 파생변수로 요약정보를 만들어 2차 변수로 제공하고 있다.

이러한 실시간 취업 정보 수집의 경우 앞서 제시한 공공부문의 고용 지표(실업청구 수당)가 가진 정보와는 달리 각 생산 주체가 기대하고 있는 경기상황 및 현재 매출 상황 등의 정보가 빠르게 반영된다는 장점을 가지고 있다. 실시간 구인 정보는 경기 선행지표에 포함된 고용통계를 가장 앞단에서 수집할 수 있는 속보성을 갖춘 데이터라 할 수 있다. 또한, 해당 데이터는 산업별, 지역별로 집계가 가능하여 지역별 산업생산 지표 및 GDP 성장률 추정을 위한 Nowcasting 모형의 성능향상에 기여할 수 있다.

다음으로, 뉴스 기사는 경제 주체들의 경제활동 중 특이사항을 실시간으로 반영하기에 동 시간대 경제지표 예측에 유용하게 사용할 수 있는 비정형 데이터이다. 경제 이벤트 및 경제에 영향을 미칠 수 있는 여러 가지 사건에 대한 뉴스 데이터를 수집 후 그 텍스트에 담겨있는 감성(Sentiment) 또는 논조(Tone)를 파악하기 위하여 텍스트 마이닝 방법론을 적용하면 실시간으로 경제 주체들의 경기에 대한 심리를 파악할 수 있을 것으로 기대된다.

이미 텍스트 데이터, 특히 미디어의 뉴스 데이터에서 나타나는 감성과 여러 금융시장 지표와의 관계는 많은 연구가 이루어져 왔으며(Nyma et al., 2015; Hong et al, 2017; Adyan, et al. 2018; Kim and Lee, 2018; Fraiberger. et al.,2021) 최근 공식발표되는 거시경제 지표들과의 관계에 대한 연구도 활발히 진행되고 있다. Song and Shin(2017)은 경기심리지수를 11개 종합일간지와 9개 경제일간지의 온라인 뉴스 기사를 활용하여 예측하고 뉴스 기사를 활용한 경기심리지수가 기존 서베이 방법의 경기심리지수에 1~2개월 선행함을 확인하였다. Hong and Min(2021)은 국내 제조업 중 주력 3대 부문인 ICT 부문(반도체, 디스플레이)과 장비 부문(자

동차, 조선), 소재 부문 (석유화학, 철강)에 대하여 각 부문을 2개 업종씩 세부 선정하여 뉴스 기사에 나타난 감성을 토픽 모델링 방법으로 추출하였다. 이렇게 추출된 지표를 활용하여 산업별 생산지수 지표와 시차 검증을 수행한 결과 약 한 달 (25일~27일) 정도 빠르게 정보를 파악할 수 있음을 확인하였다. 이와 같은 뉴스 데이터에 대한 감성 분석 모델링은 향후 경제, 주식 관련 금융 커뮤니티 및 SNS상의 투자자 또는 경제 주체들의 감성 분석 및 통화정책 관련 의사 결정문 분석을 위한 기반이 될 수 있다 (Chang 2015; Chung and Park 2022).

마지막으로 모바일 사용량을 이용한 경제 시스템 내 구성원들의 경제 활동성을 반영할 수 있다. 중앙은행 중 일본의 경우 월별 산업생산지수의 Nowcasting을 위해 스마트폰 사용량 데이터를 주간 단위로 집계하여 사용하고 있다 (Furukawa et. al., 2022).

모빌리티 정보를 이용한 또 다른 연구로는 Matsumura. et al., (2021)은 지역을 격자로 나누는 모바일폰 GPS 위치 수집 정보를 이용하여 산업생산시설이 있는 지역 내 활동 데이터와 산업생산 연간 증가율이 0.81의 상관관계를 가지고 있어 거시경제 Nowcasting에 활용가능하다는 것을 확인하였다. Suimon. and Yanai. (2021)은 모바일 집중도를 정의하고 모빌리티 인덱스를 산출하여 GDP 및 산업생산지수 Nowcasting 모형에 적용하였다. 국내연구로 Kim, et al. (2022)는 통계청에서 제공하는 주간별, 지역별, 연령별, 성별별 모빌리티 빅데이터가 소매판매액 지수에 선행하는 모습을 확인하였다.

앞서 소개된 통계청의 주간 데이터를 포함하여, 각 통신사에서 실시간으로 생산하는 모빌리티 기반 유동인구 시스템을 활용한다면 모형의 성능향상을 기대할 수 있다.

C. 전통적 거시 및 금융시장 데이터

Nowcasting 모형의 적시성 확보를 위해 앞서 실시간으로 수집할 수 있는 대안 데이터에 대해 살펴보았다. 이는 기존의 Nowcasting 모형에서 사용하던 전형적 거시경제 변수 및 금융 시장변

수에서 포착하지 못하는 추가정보를 제공하며 적시성을 확보할 수 있다는 장점이 있다. 제안한 실시간 대안 데이터와 함께 기존 Nowcasting 모형에서 입력변수로 많이 사용되던 일 또는 월 단위 금융시장 지표 및 거시경제 변수를 함께 사용한다면 공식 통계지표의 추정 시 필요 정보의 양을 극대화할 수 있을 것이다.

따라서, 일별로 발표되는 주가지수, 시장금리 및 장단기 금리 차를 포함하여 환율 및 유가 정보 등은 대안 데이터를 이용한 모형을 운용하더라도 반드시 같이 포함되어야 할 정보일 것이다. 이는 향후 실시간 대안 데이터를 이용한 추정 모형의 성능에 대한 검증 및 비교분석을 위해서도 필요하며, 더불어 전통적 Nowcasting 모형에 사용된 월간 또는 분기 단위의 거시경제 지표(물가지수, 산업생산지수, 경제 심리지수, 소비자 동향지수, 경기 종합지수-선행, 동행, 후행 등)에 대한 수집이 역시 필요할 것이다.

3.2 실시간 대안 데이터의 확보

앞서 제시한 실시간 Nowcasting 모형의 운용에 필요한 공공 및 민간영역의 데이터는 실시간 데이터베이스를 구축하여 항상성 및 최신성을 유지해야만 한다. 실시간 데이터 수집을 위해 각 데이터 소스에서 발생하는 데이터는 Table 4에 제시된 바와 같이 각종 API(Application Programming Interface)와 웹 크롤링 및 스크래핑의 정기적 수행을 통해 수집할 수 있다.

최근, 빅데이터 수집 기술의 발달과 공공데이터 개방 기조에 따라 공공영역의 데이터는 API를 통해 손쉽게 수집될 수 있다. 이는 공공부문의 각 영역에서 ‘데이터 개방과 국민수요에 맞춘 고품질 데이터를 제공하여 민간의 비즈니스 기회 발굴과 일자리 창출에 기여한다’라는 ‘데이터 기반 정부 혁신’ 프로그램에 기반해 시행되고 있다. 더욱이 ‘공공데이터의 제공 및 이용 활성화에 관한 법률’에 의거 국가기관, 지방자치단체 및 공공 기관에서 생산된 데이터는 현재 공개되지 않은 데이터라 하더라도 소관 공공기관장 및 공공데이터 활용지원센터를 통해 ‘공공데이터 제공 신청서’를 제출하면 ‘개인 정보’ 또는

‘저작권법’에 위배 되지 않는 범위에서 제공을 원칙으로 하고 있다. 따라서 이 연구에서 제시한 대안 데이터의 범위는 향후 계속 확장될 가능성이 있으며 이를 통해 적시성 있는 공식 통계 데이터의 생산 프로세스의 개선을 도모할 수 있을 것이다.

Table 4에 제시된 대안 데이터 중 공공영역의 고속도로 통행량, 전력사용량, 선박 운항 정보는 최소 5분 단위의 집계가 이루어지고 있어 실시간 정보를 수집하기에는 매우 유리한 조건을 갖추고 있다. 하지만, 일 단위 미만의 초 고빈도 데이터를 활용하여 월간 (산업생산) 또는 분기 (GDP 성장률)를 추정하는 Nowcasting 모형을 운용할 경우 단기자료의 특성상 많은 노이즈를 발생하게 될 가능성이 있다. 이에 미국 FRB에서도 주간 단위의 GDP 및 산업생산 추정 모형을 2014년부터 운용하고 있다 (Lewis, D. J. et al., 2022; Choi, 2019).

민간영역 또한 최근 빅데이터 분석을 기반으로 한 비즈니스 모델의 확산에 힘입어 기업의 비즈니스 프로세스상에서 생산되는 다양한 데이터를 데이터 자체에 가치를 부여하고 비즈니스 모델로 발전시켜 새로운 수익구조를 생성하고 있다. 앞서 제시한 카드사용 실적에 대한 데이터의 경우 대부분의 카드사에서 빅데이터 전담 조직을 구성하여 기업 내 생산되는 데이터를 관리하고 가공하여 상품의 형태로 판매하고 있으며, 이를 통해 상당한 매출을 일으켜 기업 내 새로운 비즈니스 모델을 창출하고 있다. 이러한 데이터 판매사업은 2023년 기준 2조 7천억으로 성장한 국내 빅데이터 거래 시장의 성장에 발맞춰 카드사뿐만 아니라 통신 및 기타 금융사들도 해당 분야에 활발히 진출하고 있다. 기업들은 이에 더해 해당 기업이 보유하고 있는 데이터와 타 산업데이터와의 결합을 통해 새로운 가치를 창출하고자 하는 ‘데이터 결합 및 데이터 활용

Table 4 Tool of Real-time Alternative Data

Category	Real-time Data	Tool
Alternative Real-time Data	Highway Traffic	API on the Korean Express co.
	Traffic Volume for General National and Local Roads	API on Traffic Monitoring System
	Electric Power Consumption	API on KEPCO Data Portal
	Ship Operation Information	API on Public Data Portal
	Employment Statistics	API on Public Data Portal
	Air Transportation	API on Air Portal
	Sales Record for Credit Card	API on Financial Data Platform
	POS Record by Industry and Region	API on Financial Data Exchange
	Keyword Statistic on Search Engine	API on Search Engine
	Recruit Announcements	API on Recruitment Site
Traditional Data	News	API on Naver Dev. Center
	Mobility	API on K-Stat and Telecom corp.
	Quarterly GDP	API on ECOS System
	Monthly Production by Industry, Economic Sentiment, Price Index etc.	API on ECOS System
	Daily Stock Index, Interest Rates, Exchange Rates, Oil Price	API on ECOS System

컨설팅 사업'에서 적극적으로 미래 수익원을 발굴하고자 노력하고 있다. 또한, 공공영역에서도 민간영역에서 생성된 데이터를 지역 내 연구자 지원 및 스타트업 생태계 조성을 위해 각 지방자치단체(서울시 및 부산시 등에서 구매 후 개인 정보에 대한 가명처리 후 대상자들에게 제공하여 새로운 가치를 창출하고자 노력하고 있다.

뉴스 데이터, SNS 및 투자 토론방의 텍스트 데이터는 웹 크롤링 또는 웹 스크래핑 방식을 이용하여 데이터를 수집할 수 있다. 웹 크롤링 방식의 다양한 데이터 수집용 함수들이 오픈소스로 웹에 공개되어 있으며, 국내 최대 포털 사이트인 네이버는 별도의 개발자 센터를 통해 API 형식으로 뉴스 기사, 검색어 트렌드, 쇼핑인사이드 관련 데이터에 대한 수집을 사용자들에게 쉽게 지원하고 있다. 나아가, 전통적 Nowcasting 모형에 입력변수로 사용되는 거시경제지표 및 금융시장지표 데이터는 한국은행 경제통계시스템(ECOS: Economic Statistics System)을 통해 최소 일(Daily) 단위로 API 시스템을 통해 수집 가능하다. 특히, 한국은행은 2022년 5월 기존(2004년 오픈)의 ECOS 시스템을 전면 개편하고 한국은행이 공표 및 집계하고 있는 대부분 데이터를 오픈 API를 통해 제공하고 있다.

3.3 데이터 파이프라인

앞서 제시한 공공 및 민간 영역 데이터의 수집 및 실시간 Nowcasting 모형의 운용을 위해서는 '데이터의 수집 - 1차 전처리 및 가공 - 이종 데이터의 결합 - 모형의 추정 - 실시간 예측'의 순차적 프로세스가 요구된다. Fig. 1에서는 이러한 순차적 프로세스를 운용하기 위한 데이터 파이프라인의 전체 개념도를 제시하고 있다. 우선 3.1절에서 제시한 공공 및 민간영역의 데이터를 API와 웹 크롤링 방식 등으로 수집하는 프로세스를 구축하고, 이를 통해 실시간으로 수집되는 데이터에 대해 검증 및 보관 과정을 통해 데이터레이크를 구축한다. 이후 주기적 모형의 추정 및 성능검증 모니터링을 통하여 Nowcasting 모형의 예측 성능 유지 및 예측치

실시간 배포 체계를 확보한다.

이러한 전체 프로세스를 안정적으로 운용하기 위해서는 구글 및 AWS(Amazon Web Services)와 같은 클라우드 기반의 데이터 파이프라인의 구축이 필요하다. 클라우드 시스템을 이용한 데이터 파이프라인의 구축은 데이터의 수집을 위한 API 기능 수행과 데이터에 대한 전처리 및 다양한 소스로부터 수집된 데이터들의 결합하여 모형 추정을 추정하고 배포하는 과정을 일괄처리(Batch Processing) 형식으로 구축할 수 있어 모형의 안정적인 운용을 가능하게 한다(Lee et al., 2011). 더욱이, 대안 데이터에 대한 수집을 실시간으로 수행하고 일관된 규칙에 따른 데이터의 지속적 업데이트를 위해서는 충분한 저장공간(실시간 데이터 레이크 - Real Time Data Lake)을 확보할 수 있는 클라우드 기반의 시스템 구축은 반드시 필요하다. 이후 Nowcasting 모형의 업데이트와 성능검증도 정기적으로 일괄 배치 프로세스에서 이루어져야 모형의 항상성 및 신뢰성을 확보할 수 있을 것이다.

API 및 크롤링을 통한 데이터의 수집 후 모형의 데이터 활용을 위한 1차 가공이 이루어져야 한다. 데이터 수집 주기 불일치에 따른 문제를 해결하기 위해 수집 API 실행 배치의 주기를 조율해야 하며 관측주기가 다른 다양한 원천 데이터의 결합을 위한 기준이 마련되어야 할 것이다. 더불어 전통적 모형에서 사용되었던 저빈도(월별 또는 분기별) 거시경제지표 데이터와 주기가 다른 대안 데이터를 함께 사용하는 경우가 있고 있는 데이터의 문제(Ragged-edge Problem)를 해결하기 위해 빈티지 데이터셋의 확보나 결측치의 처리 방안에 대한 기준 마련도 필요할 것이다. 나아가 데이터 파이프라인의 구축 시 전국 단위의 GDP 성장률뿐만 아니라 대안 데이터가 가지는 산업별 집계 가능 장점을 적극적으로 활용하여 개별 산업별로 GDP 성장률 및 월별 산업생산 지표를 산출하는 고도화가 가능하다. 이를 위해 데이터 수집 단계에서부터 산업에 대한 정의와 기준을 마련해야 할 것이며 장기적으로 데이터의 수집을 위한 데이터 관리 거버넌스의 구축도 필요하다.

더불어 모형 추정에 있어 다양한 입력변수를

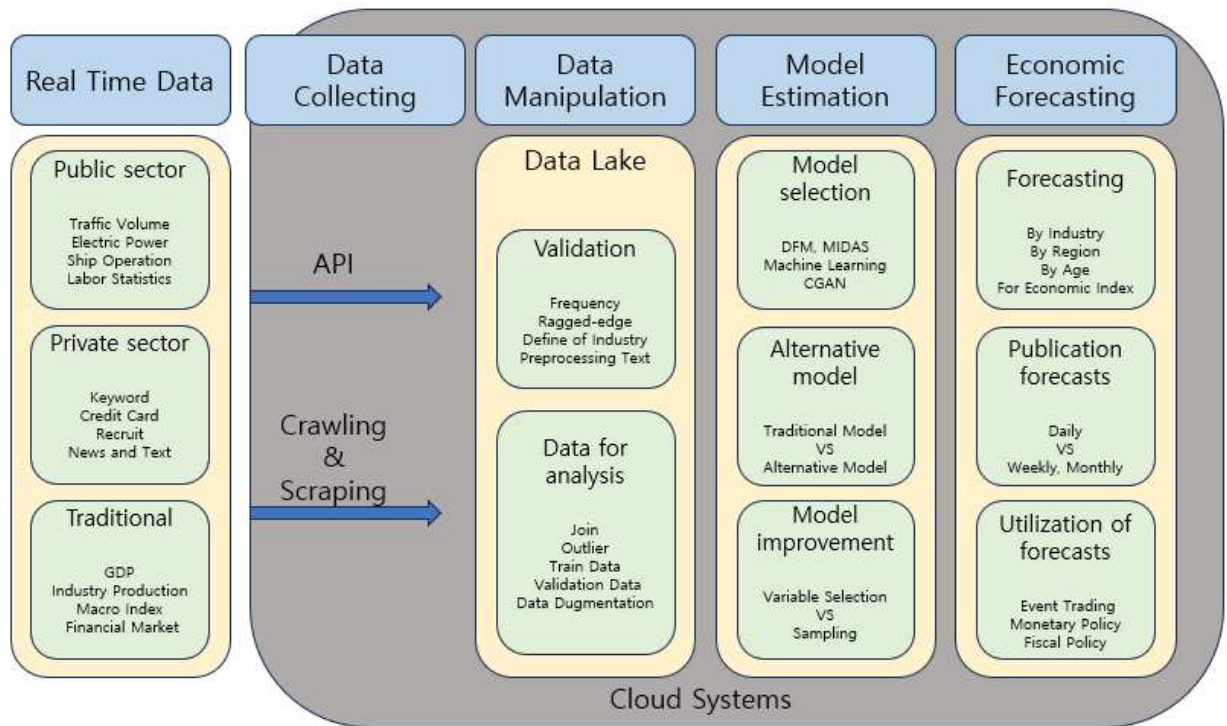


Fig. 1 Concept of Pipeline for Real Time Nowcasting

사용하고 시계열 데이터의 관측치 수가 부족한 경우 발생할 수 있는 차원의 저주 문제에 대한 고려가 필요하다. 이에 대한 문제해결을 위해 기존의 연구 방법들은 변수의 선택이나 주성분 분석 방법 등을 이용하여 접근하였으나(Kim and Swanson, 2018) 최근 연구에서는 지도학습 방법론에서 더욱 뛰어난 성능을 보이는 충분 차원 축약(Sufficient Dimension Reduction: SDR)을 이용한 연구가 늘어나고 있다(Li, 2018; Shin and Seo, 2022). 나아가 실제 모형의 추정 시 변수의 축소 방식 접근이 아닌 Sample 증강기법(Zheng et al., 2021)을 이용한 차원의 저주 방지 및 시계열 예측 일반화 성능 확보 방법이 최근 연구되고 있다. 이를 모형 추정에 적용하여 변수 관점에서 차원의 축소와 Sample 증강의 기법을 비교분석하고 Nowcasting 예측에 활용한다면 좀 더 우수한 예측 성능을 기대할 수 있을 것이다.

추정 모형 또한 제안하는 실시간 대안 데이터를 활용한다면 기존의 계량 경제학에 기반한 전통적 방식을 넘어 머신러닝 알고리즘에 대한 적

용이 가능할 것이다. 각국 중앙은행을 중심으로 머신러닝 기법과 대안 데이터를 이용한 다양한 시도가 이루어지고 있다(Richardson. et al., 2021; Yi, et al., 2022). 더욱이, 경제 변수들 간의 비선형성을 반영하기에 기존 전통적 계량 모형 대비 통계적 머신러닝 모형 및 신경망을 이용한 다양한 딥러닝 모형이 더 적합할 수 있는 것으로 알려졌다(Soybilgen and Yazgan, 2021; Longo, Riccaboni, and Rungi, 2022; Hopp, 2022; Lee, 2021). 나아가 현재 경제상태를 조건부 변수로 제시한 상태에서 입력변수 간의 관계를 고려한 생성모형(CGAN: Conditional Generative Adversarial Network)을 적용한다면 기존에 시도된 앙상블 기법 기반의 모형보다 뛰어난 예측 성능을 기대해 볼 수 있다(Luo., 2022).

제안된 실시간 대안 데이터는 지역별 수집이 가능한 데이터이다. 이는 현재 연간단위 또는 1년 지연 시간을 가지고 수집되어, 적시성 있는 지자체의 정책 수립에 어려움을 겪고 있는 지역 내 총생산(GRDP: Gross Regional Domestic Product)의 추정 문제의 해결에도 기여할 수 있

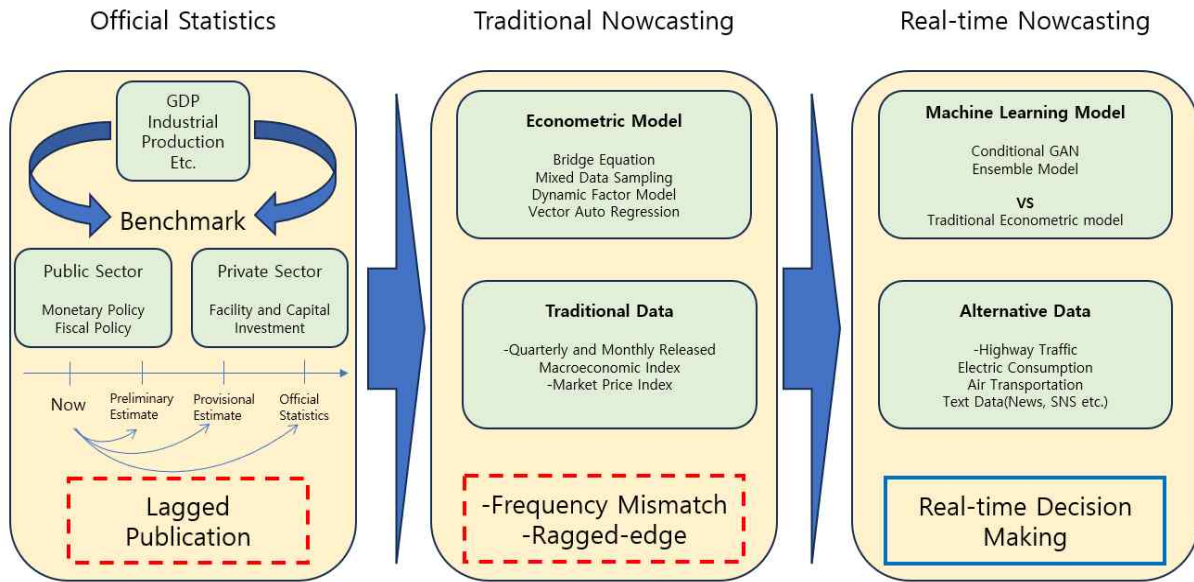


Fig. 2 Advancement of Nowcasting with Real-time Alternative Data

을 것이라 기대한다. 이를 바탕으로 수도권 집중으로 인한 지방소멸과 지역경제 안정화를 위한 정책 수립을 위해 적시성 있는 공식 통계지표를 제공할 수 있는 기틀을 마련할 수 있을 것이다.

마지막으로, 길게는 분기 및 월간 단위의 관측 주기를 가진 데이터와 짧게는 실시간 데이터의 결합과 이를 이용한 모형을 안정적으로 운용하기 위해서는 API 및 웹상에 존재하는 데이터의 수집의 지속성과 항상성 유지가 필요하다. 이를 위해 데이터 수집 과정에서 모형의 추정 및 예측치의 배포에 이르는 전 과정에 걸쳐 데이터 검증 및 시스템 로그 모니터링이 필요하며 이를 위한 관리 체계의 수립되어야 한다. 이는 본 연구에서 제시하는 실시간 데이터를 이용한 Nowcasting 모형 및 예측치의 신뢰성을 담보하여 여러 경제 주체들의 다양한 의사 결정에 효율성을 제공할 수 있을 것이다.

4. 결론

본 연구를 통해 공식발표 통계지표의 공표 시점 지연에 대한 문제를 살펴보고 이를 해결하기 위한 전통적 Nowcasting 방법론 및 대안 데이터를 이용한 관련 연구들을 살펴보았다. 나아

가, 경제지표 Nowcasting 모형의 성능을 향상시킬 수 있는 실시간 대안 데이터를 살펴보고 경제지표와 대안 데이터와의 관계에 대해 검증한 기존 문헌들을 정리하였다. 또한, 실시간 Nowcasting 모형의 효과적 운용을 위한 클라우드 서비스 기반의 데이터 파이프라인 및 데이터 레이크 구축을 제안하고, 구축 시 고려하여야 할 부분을 점검하였다.

이 연구에서는 실시간 대안 데이터가 가지는 공식 경제지표 예측 능력에 관한 후속 연구를 위해 최근 발전하고 있는 데이터 집적기술을 기반으로 Nowcasting 성능을 향상하고 공식발표 통계지표의 적시성을 확보할 방안을 제시하였다. 제안한 실시간 데이터 파이프라인에 기반한 Nowcasting 모형이 안정적으로 운용될 수 있다면 Fig 2.와 같이 기존 전통적 Nowcasting 모형이 가진 몇 가지 약점을 보완하여 통계지표의 집계과정에서 발생하는 시차 지연의 문제를 해결할 수 있을 것이다.

이는 실무적 기여 측면에서 우선, 공공영역에서 신속하고 정확한 공식통계 지표의 추정치를 제공할 수 있어 기준금리의 결정과 같은 통화정책과 조세 수준, 국채발행량 결정 등과 같은 재정정책을 경제 상황에 맞게 정책입안자들이 수립하는데 도움이 줄 수 있을 것이다. 또한, 민간

영역에서도 가계의 소비 의사 결정이나 기업의 설비투자, 인력 채용 계획 수립 등 경제시스템 구성원들의 실시간 의사 결정에 유용한 정보를 제공할 수 있을 것이다. 나아가, 정확하고 신속한 공식통계에 기반한 금융시장 안정과 투자 전략의 수립에도 도움을 줄 수 있을 것으로 기대된다.

학술적 기여도 관점에서 본 연구는 기존 문헌 및 공개되고 있는 공공 및 민간영역 데이터에 기반하여 실시간 대안 데이터를 통한 공식통계 지표 적시성 있는 추정의 가능성을 확인하였다. 하지만, 제시된 대안 데이터와 통계지표 간의 선·후 행성에 관한 실증분석이 포함되지 않고 기존 문헌 연구를 통해 간접적으로 그 연관성을 확인하여 학술적 엄밀성에 그 한계를 지니고 있다.

그럼에도 불구하고 본 연구를 통해 공식통계 지표의 공표시점 지연 문제의 해결을 위해 최근 집적되고 있는 실시간 빅데이터의 활용 가능성을 확인하였으며, 이러한 Nowcasting모형을 운용할 때 유의하여 확인해야 할 사항에 대해 점검하였다. 향후 본 연구에서 제시한 실시간 데이터 파이프라인을 구축하고 전통적 Nowcasting 방법론을 넘어 머신러닝을 이용한 새로운 접근을 통해 공식발표 통계지표의 품질을 개선한다면 공공 및 민간영역에서 이를 이용한 다양한 학술적·실무적 밑거름이 될 수 있을 것이라 기대한다.

References

- Adyan M. R., Choi, H. R. and Lim, S. B. (2018). Korean and English Sentiment Analysis Using the Deep Learning. *Journal of Korea Society of Industrial Information Systems*, 23(3), 59-71.
- Ann, Y. G. and Lee, J. W. (2017). An Empirical Study of Continental Port Traffic Impact on World Economy: Focusing on Container Volume. *Korea Logistics Review*, 27(1), 1-10.
- Aragón, F. M., Rud, J. P. and Toews, G. (2018). Resource Shocks, Employment, and Gender: Evidence from the Collapse of the UK Coal Industry. *Labour Economics*, 52, 54-67.
- Bok, B., Caratelli, D., Giannone, D., Sbordone, A. M. and Tambalotti, A. (2018). Macroeconomic Nowcasting and Forecasting with big data. *Annual Review of Economics*, 10, 615-643.
- Chang, J. H. (2015). Mining Highly Attention Items using a Two-way Decay Mechanism in Data Stream Mining. *Journal of Korea Society of Industrial Information Systems*, 20(2), 1-9.
- Cho, J. H. and Kang, M. O. (2012). Causal Relationship between Electricity Consumption and Economic Growth in Korea. *Environmental and Resource Economics Review*, 21(3), 573-593.
- Cho, Y., and Lee, Y. (2022). Asymmetric Asset Correlation in Credit Portfolios. *Finance Research Letters*, 49, 103037.
- Choi, B. H. (2007). A Study on Causality between Trading Volume of Freight and Industrial Growth in Korea Ports. *JOURNAL OF KOREA PORT ECONOMIC ASSOCIATION*, 23(4), 159-175.
- Choi, H. and Varian, H. (2012). Predicting the Present with Google Trends. *Economic record*, 88, 2-9.
- Choi, K. S. (2019). K-SuperCast: A Big Data based GDP Forecasting Model. *Journal of the Korean Data And Information Science Society*, 30(4), 723-743.
- Chung, J. M. and Park. Y. H. (2022). Social Media Bigdata Analysis Based on Information Security Keyword Using Text Mining, *Journal of Korea Society of Industrial Information Systems*, 27(5), 37-48.
- Chung, S. Y., Baek, S. K. and Kang, J. K. (2010). Analysis of Metropolitan Freeway

- Traffic Patterns Using TCS Data, *The Korea Institute of Intelligent Transport Systems*, proceeding, 79-83
- Clements, M. P. and Galvão, A. B. (2008). Macroeconomic Forecasting with Mixed-frequency Data: Forecasting Output Growth in the United States. *Journal of Business & Economic Statistics*, 26(4), 546-554.
- Couch, K. A. and Fairlie, R. (2010). Last Hired, First Fired? Black-white Unemployment and the Business Cycle. *Demography*, 47, 227-247.
- Croushore, D. and Stark, T. (2003). A Real-time Data Set for Macroeconomists: Does the Data Vintage Matter?. *Review of Economics and Statistics*, 85(3), 605-617.
- Ferrara, L., Guégan, D. and Rakotomaroahy, P. (2010). GDP Nowcasting with Ragged edge Data: A Semi Parametric Modeling. *Journal of Forecasting*, 29(1-2), 186-199.
- Fraiberger, S. P., Lee, D., Puy, D. and Ranciere, R. (2021). Media Sentiment and International Asset Prices. *Journal of International Economics*, 133, 103526.
- Furukawa, K., Hisano, R., Minoura, Y. and Yagi, T. (2022). A Nowcasting Model of Industrial Production using Alternative Data and Machine Learning Approaches (No. 22-E-16). *Bank of Japan*.
- Hong, S. W. and Min, S. H. (2021). Market Analysis of Key Manufacturing Segments Using News Data. *KIET Monthly Industrial Economics*, 270, 24-34.
- Hong, T. H, Kim, E. M. and Cha, E. J. (2017). The Prediction of Dow Jones and S&P500 Index Using SVM and News Sentiment Analysis. *The Journal of Internet Electronic Commerce Research*, 17(1), 23-36.
- Hopp, D. (2022). Benchmarking Econometric and Machine Learning Methodologies in Nowcasting: *UNCTAD Research Paper* No. 83.
- Hwang, Y. J. (2017). Business Cycles Analysis of Daegu-Gyeongbuk Province Using Mixed-Frequency Data. *Journal of Economics Studies*, 35(1), 115-144.
- Kim, D. H., Won, M. S., Song, T. J. and Hong, J. Y. (2022). Development of Faster Consumption Trend Indicator Using Traffic Mobility Big Data - Focusing on Retail Sales Index -. *Journal of Transport Research*, 29(4), 35-51.
- Kim, D. Y. and Lee, Y. I. (2018). News based Stock Market Sentiment Lexicon Acquisition Using Word2Vec. *The Korea Journal of BigData*, 3(1), 13-20.
- Kim, H. H. and Swanson, N. R. (2018). Methods for Backcasting, Nowcasting and Forecasting using Factor MIDAS: With an Application to Korean GDP. *Journal of Forecasting*, 37(3), 281-302.
- Kim, H. J. (2006). A Case Study of Difference Analysis Between Tentative and Definitive Statistics using Statistical Techniques. *Quarterly National Accounts*, 2(25), 100-118
- Kim, J. (2021). Industrial Growth Elasticity of Employment in Korea: Comparison between Pre and Post-Global Financial Crisis Era. *Ordo Economics Journal*, 24(2), 43-66.
- Relationship between Economic Change and Air Passenger Demand: Focus on Incheon International Airport. *Journal of the Korean Society for Aeronautical Science and Flight operation*, 27(4), 52-64.
- Kim, S. H., Baek, S. K. and Han, D. H. (2010). Relationship between Expressway Traffic Volume and Economic Growth Rate, *Transportation Technology and Policy*, 7(4), 75-79
- Kim, S., Lee, Y. J., Shin, J. and Park, K. Y. (2019). Text Mining for Economic Analysis. Available at SSRN 3405781.
- Kwon, C. M, Hwang, S. W. and Jung J. U.

- (2015). Application of Web Query Information for Forecasting Korean Unemployment Rate. *JOURNAL OF THE KOREA SOCIETY FOR SIMULATION*, 24(2), 31-39.
- Lee, C. D. (2022). Prediction on Busan's Gross Product and Employment of Major Industry with Logistic Regression and Machine Learning Model. *Korea Trade Review*, 47(2), 69-88.
- Lee, G. H. and Hwang, S. P. (2014). Business Cycle Indicator Using Big Data: Compilation of the Naver Search Business Index, *Economic Analysis*, 20(4), 1-38
- Lee, K. H., Choi, H. S. and Chung, Y. D. (2011). Massive Data Processing and Management in Cloud Computing: A Survey, *Journal of KIISE : Databases*, 38(2), 104-125.
- Lee, N. Y. and Lee, K. Y. (2022). A Study on the Advancement of the Container Port Throughput Forecasting Model. *Korea Maritime Institute*. 19(01)
- Lee, S. I. (2021). Development of a GRDP Prediction Model Using Deep Learning: A Case Study of Ulsan, Korea. Available at SSRN 3941637.
- Lee, S. Y., Lee, E. K and Han, J. H. (2017). Short-term Forecasting System using Machine Learning and MIDAS Model. *BOK Monthly Bulletin*, 71(5), 16-41.
- Lewis, D. J., Mertens, K., Stock, J. H. and Trivedi, M. (2021, May). High-frequency Data and A Weekly Economic Index During the Pandemic. *In AEA Papers and Proceedings* (Vol. 111, 326-330). American Economic Association.
- Lewis, D. J., Mertens, K., Stock, J. H. and Trivedi, M. (2022). Measuring Real Activity using A Weekly Economic Index. *Journal of Applied Econometrics*, 37(4), 667-687.
- Lim, J. H., Kim, Y. R. and Choi, Y. C. (2020). A Study on Factors Affecting Korea's Air Freight Trade Field. *Journal of the Korean Society for Aviation and Aeronautics*, 28(4), 117-124.
- Lim, S. S and Choi, H. S. (2021). Real-time Forecasting of Real GDP Using Text Mining. *Journal of Corporation and Innovation*, 44(4), 91-106.
- Longo, L., Riccaboni, M. and Rungi, A. (2022). A Neural Network Ensemble Approach for GDP Forecasting. *Journal of Economic Dynamics and Control*, 134, 104278.
- Luo, C., Li, X., Ye, Y., Feng, S. and Ng, M. K. (2022). Experimental Study on Generative Adversarial Network for Precipitation Nowcasting. *IEEE Transactions on Geoscience and Remote Sensing*, 60, 1-20.
- Matsumura, K., Oh, Y., Sugo, T., & Takahashi, K. (2021). Nowcasting Economic Activity with Mobility data. Available at SSRN 4406104.
- McLaren, N. and Shanbhogue, R. (2011). Using Internet Search Data as Economic Indicators. *Bank of England Quarterly Bulletin*, (2011), Q2.
- Miller, S., Moat, H. S., & Preis, T. (2020). Using Aircraft Location Data to Estimate Current Economic Activity. *Scientific reports*, 10(1), 7576.
- Mo, S. W. and Lee, K. B. (2013). Causality Analysis between Port Trading Volume and Industrial Activity. *Journal of Shipping and Logistics*, 29(2), 221-235.
- Nakazawa, T. (2022). Constructing GDP Nowcasting Models Using Alternative Data (No. 22-E-9). *Bank of Japan*.
- Nyman, R., Kapadia, S. and Tuckett, D. (2021). News and Narratives in Financial Systems: Exploiting Big Data for Systemic Risk Assessment. *Journal of Economic Dynamics and Control*, 127, 104119.
- Park, K. H. and Kim, J. K. (2013). Sectoral Energy Consumption and Economic Growth

- in Korea. *Korea Energy Economic Review*, 12(2), 59-83.
- Park, M. H., Roh, G. K. and Lee, S. E. An Analysis on the Causal Relation Between Electricity Consumption and GDP by industries in KOREA. *Journal of the Korean Institute of Illuminating and Electrical Installation Engineers*, 30(3), 39-45.
- Park, S. K. (2020). Monitoring Real-time Business Conditions based on Electricity Consumption. *KIET Monthly Industrial Economics*, 260, 7-19.
- Richardson, A., van Florenstein Mulder, T. and Vehbi, T. (2021). Nowcasting GDP using Machine-learning Algorithms: A Real-time Assessment. *International Journal of Forecasting*, 37(2), 941-948.
- Robertson, J. C. and Tallman, E. W. (1999). Vector Autoregressions: Forecasting and Reality. *Economic Review-Federal Reserve Bank of Atlanta*, 84(1), 4.
- Seo, J. Y. (2021). The Relationship between Improvement of Private Consumption Sector and Card Usage: Evidence from Korean Card Market. *The Credit Card Review*, 15(3), 1-17.
- Shin, S. H. (2014). Analysis on Structural Change of Growth-employment Relationship in Korea. *Korean Journal of Industrial Relations*, 24(3), 45-61.
- Shin, S. J. and Seo, B. S. (2022). Real-Time Private Consumption Prediction Using Big Data. *BOK Working Paper*, 2022(16)
- Smith, M. S. and Vahey, S. P. (2016). Asymmetric Forecast Densities for US Macroeconomic Variables from a Gaussian Copula Model of Cross-sectional and Serial Dependence. *Journal of Business & Economic Statistics*, 34(3), 416-434.
- Song, M. C. and Shin, K. S. (2017). Construction of Consumer Confidence index based on Sentiment Analysis using News articles. *Journal of Intelligence and Information Systems*, 23(3), 1-27.
- Soybilgen, B. and Yazgan, E. (2021). Nowcasting US GDP using Tree-based Ensemble Models and Dynamic Factors. *Computational Economics*, 57, 387-417.
- Statistics Korea. (2021). Statistics Information Report for Index of All Industry Production: IAI, *Statistics Korea*
- Suimon, Y., and Yanai, M. (2021, July). Using Mobile Phone GPS Data to Nowcast Auto Production Activity and Equity Investment Strategy Based on the Activity Nowcasting. *In 2021 10th International Congress on Advanced Applied Informatics (IIAI-AAI)* (pp. 680-683). IEEE.
- Sung, B. H. (2001). Economic Forecasting using the Bayesian VAR Model. *Economic Analysis*, 7(2), 59-90.
- Wallis, K. F. (1986). Forecasting with an Econometric model: The 'ragged edge' problem. *Journal of Forecasting*, 5(1), 1-13.
- Wegmüller, P. and Glocker, C. (2023). US Weekly Economic Index: Replication and Extension. *Journal of Applied Econometrics*.
- Wen, Y. and Chen, M. (2012). Okun's law: A Meaningful Guide for Monetary Policy?. *Economic Synopses*, 2012(2012-06-08).
- Yi, H. C., Choi, D. K. and Kim, Y. G. (2022). Dynamic Factor Model and Deep Learning Algorithm for GDP Nowcasting. *Economic Analysis*, 28(2), 1-37.
- Zheng, H., Zhou, Z. and Chen, J. (2021). RLSTM: A New Framework of Stock Prediction by using Random Noise for Overfitting Prevention. *Computational Intelligence and Neuroscience*, 2021, 1-14.



조 용 복 (Yongbok Cho)

- 정회원
- 인하대학교 경영학과 경영학사
- 고려대학교 통계학과 경제학사
- 고려대학교 금융공학협동과정
금융공학 석사 및 박사
- (현재) 동아대학교 경영대학

경영정보학과 조교수

- 관심분야: 금융시장, 데이터사이언스, 핀테크, 위험관리, 자산배분 등



김 도 완 (Dowan Kim)

- 정회원
- 고려대학교 금융공학협동과정
금융공학 박사
- (현재) 한성대학교 사회과학부
조교수
- 관심분야: 금융시장, 데이터

사이언스 등