

Forecasting Korea's GDP growth rate based on the dynamic factor model

Kyoungseo Lee^a, Yaeji Lim^{1,a}

^aDepartment of Applied Statistics, Chung-Ang University

Abstract

GDP represents the total market value of goods and services produced by all economic entities, including households, businesses, and governments in a country, during a specific time period. It is a representative economic indicator that helps identify the size of a country's economy and influences government policies, so various studies are being conducted on it. This paper presents a GDP growth rate forecasting model based on a dynamic factor model using key macroeconomic indicators of G20 countries. The extracted factors are combined with various regression analysis methodologies to compare results. Additionally, traditional time series forecasting methods such as the ARIMA model and forecasting using common components are also evaluated. Considering the significant volatility of indicators following the COVID-19 pandemic, the forecast period is divided into pre-COVID and post-COVID periods. The findings reveal that the dynamic factor model, incorporating ridge regression and lasso regression, demonstrates the best performance both before and after COVID.

Keywords: dynamic factor model, OLS, Ridge, Lasso, ARIMA, GDP growth rate

1. Introduction

거시경제학은 GDP, 인플레이션, 물가, 실업률 등 경제주체들의 상호작용에 의해서 발생하는 범국가적 경제 현상에 대해 연구하는 학문이다. 거시경제 변수들은 한 나라의 거시경제 현상을 파악하고 정책에 대한 의사 결정을 하는데 유용하게 활용되므로 경제학 뿐만 아니라 통계 분야에서도 활발하게 연구되고 있는 자료이다. 그 중 GDP는 한 나라의 가계, 기업, 정부 등 모든 주체가 일정 기간 동안 창출한 재화와 서비스의 가치를 나타내는 지표이며, GDP를 통하여 국가의 전반적인 경제 수준을 파악할 수 있다. 따라서 다양한 연구가 GDP 분석을 위해 이루어져왔다. 대표적으로 Giannone 등 (2008)은 다량의 월간 데이터를 사용하여 현재 분기 GDP 성장률에 대한 nowcasting 방법론을 연구하였다. Nowcasting은 GDP 통계보다 더 일찍 발표되는 여러 실물 및 금융 지표들을 효율적으로 결합하여 실시간 또는 그에 가까운 속보성을 확보하여 GDP를 예측하는 것이다. 또한, Angelini 등 (2011)은 유로 지역에서 분기별 GDP 성장률의 조기 추정치를 구하기 위하여 적시에 월간 데이터를 활용하는 모델들을 평가했다.

대표적인 GDP 분석 방법론으로 다변량 시계열 방법론인 dynamic factor model (DFM)이 있다. DFM은 다수의 데이터를 적은 수의 공통 요인으로 설명해주는 방법으로, Sargent와 Sims (1977)와 Geweke (1977)이 처음 제안하였다. Factor model (FM)은 시계열 데이터를 공통 요인(common component)과 특이 요인(idiosyncratic

This research was supported by the National Research Foundation of Korea (NRF) funded by the Korea government (2022R1F1A1074134).

¹Corresponding author: Department of Applied Statistics, Chung-Ang University, 84 Heukseok-ro, Dongjak-gu, Seoul 06974, Korea. E-mail: yaeji.lim@gmail.com

component)의 선형 결합으로 분해하는 방법론이며, 기존의 FM이 시차에 따른 영향을 고려하지 않았다면, DFM은 해당 관계를 고려한 모형이므로 시계열 자료분석에 보다 적절한 방법론이다. DFM은 거시경제학 자료 분석에 활발히 이용되어 왔는데, 특히 각국의 GDP 성장률을 예측하는데 많이 활용되고 있다. 뉴욕 연준은 DFM에 베이저안 추정법과 Kalman-filtering 기법을 적용하여 실질 GDP 성장률의 nowcasting 결과를 실시간으로 발표하고 있다. 국내에서도 DFM과 통계 방법론을 결합하여 GDP 성장률을 예측하는 연구가 다양하게 진행되어 왔다. 대표적으로, Lee 등 (2022)은 DFM과 순차 데이터에 특화된 딥러닝 알고리즘인 long short-term memory (LSTM)을 상호보완적으로 적용한 GDP nowcasting 시스템을 제안하였고, Kim과 Kim (2016)은 분기 GDP 성장률을 예측하는데 DFM과 AR모형을 비교하였다.

본 논문에서는 다양한 DFM 기반의 예측모형을 고려하여 한국 GDP 성장률을 예측하고자 한다. 먼저, DFM을 활용하여 다수의 거시경제 변수들에 내재되어 있는 요인을 구하고, 이 요인들에 다양한 회귀분석 방법론을 적용하여 분기별 한국 GDP 성장률을 예측하고자 한다. 기존 연구에 활용되었던 최소제곱법과 더불어, ridge와 lasso 방법론을 추가로 고려하여, 요인들의 다중공선성을 해결하고자 하였고 예측력에 있어서 다양한 방법론들을 비교하고자 한다. Li와 Chen (2014) 또한 lasso 회귀분석 방법론을 사용하여 경제지표를 예측하였지만, 그들은 lasso 회귀계수와 DFM에서 얻은 factor를 모두 예측변수로 사용한 회귀분석식을 고려하였다. 이와 달리, 본 연구에서는 factor들을 예측변수로 활용한 회귀모형에서 패널티 항을 이용하여 lasso 등을 적용하였다.

논문은 다음과 같이 구성된다. 2장에서는 DFM에 대한 개념과 기존 DFM 추정 방법론에 대해 소개한다. 3장에서는 DFM기반의 시계열 예측방법론들을 소개한다. 4장에서는 예측에 사용되는 한국 거시경제 자료들에 대해 설명하고, 3장에서 소개한 예측 방법을 적용한 결과를 비교한다. 5장에서는 결론에 대하여 논한다.

2. Factor model

본 장에서는 먼저 static FM과 비교하여 DFM을 소개한다.

2.1. Dynamic factor model

시계열 자료 $\mathbf{x}_t = (x_{1t}, \dots, x_{nt})'$, $t \in \mathbb{Z}$ 에 대해 다음과 같은 factor model을 고려하자.

$$\mathbf{x}_t = \mathbf{B}\mathbf{f}_t + \boldsymbol{\xi}_t = \boldsymbol{\chi}_t + \boldsymbol{\xi}_t. \quad (2.1)$$

이 때, $\mathbf{f}_t = (f_{1t}, \dots, f_{rt})'$ 는 factor로 정상성을 만족하는 r -차원 벡터이며, $\mathbf{B} \in \mathbb{R}^{n \times r}$ 는 loading 행렬이다. \mathbf{B} 와 \mathbf{f}_t 를 곱한 $\boldsymbol{\chi}_t = (\chi_{1t}, \dots, \chi_{nt})'$ 를 공통 요인이라고 하며, $\boldsymbol{\xi}_t = (\xi_{1t}, \dots, \xi_{nt})'$ 는 특이 요인이라고 한다.

Static FM에서는 \mathbf{f}_t 는 \mathbf{x}_t 에 시차에 상관없이 동일한 영향을 미친다. 공통 요인과 특이 요인에 대해서는 아래와 같이 상관성이 없다고 가정하며, 특이 요인 간 상관성도 없다고 가정한다.

$$\begin{aligned} \text{Cov}(\chi_{it}, \xi_{js}) &= 0, \quad t, s \in \mathbb{Z}, \quad i, j = 1, \dots, n, \\ \text{Cov}(\xi_{it}, \xi_{js}) &= 0, \quad t, s \in \mathbb{Z}, \quad i, j = 1, \dots, n, \quad i \neq j. \end{aligned}$$

DFM은 위의 static FM을 확장한 것으로, 데이터를 공통 요인과 특이 요인으로 분해해서 내재되어 있는 공통 요인을 파악한다는 점은 동일하나, 공통 요인의 시차에 따른 영향을 반영한다는 차이가 있다. 모형은 아래와 같다.

$$\mathbf{x}_t = \mathbf{B}(L)\mathbf{u}_t + \boldsymbol{\xi}_t, \quad (2.2)$$

$$x_{it} = \chi_{it} + \xi_{it}, \quad \text{where } \chi_{it} = \sum_{j=1}^q b_{ij}(L)u_{jt}.$$

이 때, L 은 lag operator이고, $\mathbf{u}_t = (u_{1t}, \dots, u_{jt})'$ 는 dynamic factor로 orthonormal한 white noise이다. DFM에서의 공통 요인은 dynamic factor의 시차변수들로 설명될 수 있으며, static FM과 마찬가지로 공통 요인과 특이 요인의 상관성에 대한 가정이 적용된다. Forni 등 (2000)은 더 나아가 특이 요인 사이에 약간의 dependence가 있으며, 자기상관관계도 존재할 수 있다는 점을 허용한 일반화된 DFM을 제시했다.

2.2. Factor estimation

본 연구에서는 Doz 등 (2011)이 제안한 Kalman filtering을 이용한 factor추정 방법을 활용하였다. 해당 논문에서는 크게 두 단계의 절차를 거쳐 factor를 추정하는데, 첫 번째 단계에서는 주성분 분석을 통해 얻은 principal componet들에 대해 단순 회귀분석을 진행하여 모수들을 추정하며 두 번째 단계에서는 Kalman filter and smoother를 통하여 factor를 재추정한다. Doz 등 (2011)는 논문에서 이렇게 얻은 추정 factor가 consistent함을 증명하였고, EURO 지역의 GDP 성장률에 대하여 적용하여 그 우수성을 보여주었다. 본 연구에서는 Doz 등 (2011)이 제안한 두 단계의 factor 추정방법을 사용하였으며, 방법론에 대한 자세한 내용은 해당 논문에서 확인할 수 있다.

3. Forecasting method

본 절에서는 DFM 기반의 예측 모델을 소개한다.

- Common component based forecasting

$$\mathbf{x}_t = \chi_t + \xi_t. \tag{3.1}$$

Eichler 등 (2011)는 위의 DFM모델에서 common component를 구한 후 $T + 1$ 시점의 GDP 성장률을 $\hat{\chi}_T$ 로 추정하였다.

- Ordinary least squares (OLS) based forecasting

Stock과 Watson (2002)은 다음의 회귀분석 모델을 사용하여 예측모델을 제안하였다.

$$\mathbf{y} = \mathbf{F}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{3.2}$$

여기서 $\mathbf{y} = (y_1, \dots, y_T)'$ 는 GDP값이며, $\mathbf{F} = (\mathbf{f}_1, \dots, \mathbf{f}_T)'$ 로 DFM에서 얻은 factor이며, 오차벡터 $\boldsymbol{\epsilon} = (\epsilon_1, \epsilon_2, \dots, \epsilon_T)'$ 은 $\mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$ 의 분포를 따른다고 가정한다. 이 때 $\boldsymbol{\beta}$ 의 최소제곱추정량은

$$\hat{\boldsymbol{\beta}} = (\mathbf{F}'\mathbf{F})^{-1} \mathbf{F}'\mathbf{y}$$

로 주어진다.

- Ridge regression based forecasting

다중공선성이 존재할 경우 최소제곱추정량 $\hat{\boldsymbol{\beta}}$ 은 적절하지 않은 추정량이 될 수 있다. 이에 대한 대안으로 ridge regression이 제안되었다 (Hoerl과 Kennard, 1970).

$$\hat{\boldsymbol{\beta}}_j = \arg \min_{\boldsymbol{\beta}} \left\{ \sum_{i=1}^T \left(y_i - \sum_j \beta_j f_{ij} \right)^2 + \lambda \sum_j \beta_j^2 \right\},$$

여기서 λ 는 penalty term으로 불필요한 변수는 축소시켜 주는 효과가 있다. 본 논문에서는 Stock과 Watson (2002)의 OLS 예측방법론을 확장하여 ridge panely를 고려한 회귀분석 모형을 (3.2)에 적용하였다.

Table 1: Available variables of G20 countries

	GDP	Cons.	Inv.	CPI	PPI	I.P.	Int.	Une.
Korea	O	O	O	O	O	O	O	O
Japan	O	O	O	O	O	O	O	O
India	O	-	O	O	-	O	O	-
Indonesia	O	-	O	O	-	O	O	-
United States	O	O	O	O	-	O	O	O
United Kingdom	O	O	O	O	O	O	O	O
France	O	O	O	O	O	O	O	O
Germany	O	O	O	O	O	O	O	O
Canada	O	O	O	O	-	O	O	O
Italy	O	O	O	O	O	O	O	O
Australia	O	O	O	O	-	O	O	O
Brazil	O	-	O	O	-	O	O	-
Mexico	O	O	O	O	O	O	-	O
Russia	O	-	O	O	O	O	O	-
Turkiye	O	O	O	O	O	O	-	O
South Africa	O	O	O	O	O	O	O	-
EU	O	-	O	O	O	O	-	O

- Lasso regression based forecasting

Lasso regression 또한 ridge regression과 함께 penalized regression 방법 중 하나이며, 회귀 계수가 축소되는 ridge regression과 달리 일부 회귀계수를 0으로 만들어줌으로써 변수선택 기능이 있다 (Tibshirani, 1996). 아래는 lasso regression의 모형이다.

$$\hat{\beta}_j = \arg \min_{\beta} \left\{ \sum_{i=1}^T \left(y_i - \sum_j \beta_j f_{ij} \right)^2 + \lambda \sum_j |\beta_j| \right\}.$$

위 식에서 알 수 있듯이 ridge regression의 penalty term에는 β_j^2 이, lasso regression의 penalty term에는 $|\beta_j|$ 이 사용되며, 이로 인해 lasso regression에서는 변수선택이 이루어진다. 본 논문에서는 회귀모형 (3.2)에 lasso regression을 적용하여 예측모형을 생성하였다.

4. Forecasting Korea's GDP growth rate

4.1. Data

본 논문에서는 2005년부터 2022년까지의 G20 국가들의 분기별 거시경제 변수들에 DFM을 적용하여 한국의 GDP를 예측하였다. 거시경제 지표로 총소득 증감률(GDP), 민간최종소비지출 증감률(Cons.), 총고정자본형성 증감률(Inv.), 소비자물가지수(CPI), 생산자물가지수(PPI), 산업생산지수(I.P.), 장기금리(Int.), 실업률(Une.)을 선정하였다. 데이터는 OECD data에서 수집하였고, 4개 이상 변수에 대한 데이터를 얻을 수 있는 국가들만 분석에 포함했다. 이에 따라 중국, 사우디아라비아, 아르헨티나의 데이터는 제외하였으며, 각 나라별 사용항 변수는 Table 1과 같다.

변수는 총 117개이고, time point는 18년 간 매 분기 데이터를 수집하여 총 72개이다. 모형에는 예측하고자 하는 분기의 직전 분기까지의 데이터를 사용하였다. 예를 들어, 2022년 4분기 한국의 GDP 성장률을 예측하고자 하면, 2005년 1분기부터 2022년 3분기까지의 데이터를 사용하였다. 2020년 COVID의 영향으로



Figure 1: Quarterly GDP growth rate of Korea.

지표의 변동성이 클 것을 감안하여 예측 시기를 팬데믹 이전과 이후로 나눠서 모형을 적용해보았다 (Figure 1). COVID 이전은 2018년부터 2019년까지 총 8분기의 GDP 성장률을, COVID 이후는 2021년부터 2022년까지 총 8분기의 GDP 성장률을 예측하였다. 또한, 예측 시점 직전까지의 데이터를 training data로 사용하였는데, 예를들어 2018년 1분기를 예측하는 경우에는 2005년 1분기부터 2017년 4분기까지의 자료를 사용하여 모형을 만들었으며, 2018년 2분기를 예측할 때는 2005년 1분기부터 2018년 1분기까지의 데이터를 사용하였다.

DFM을 적용하려면 모든 데이터가 정상성을 만족해야 한다. Augmented Dickey-Fuller test와 KPSS test 결과 정상성을 만족하지 않는 데이터는 차분을 하여 정상성을 가지도록 변형하였다. 그 후 국가별 각 변수에 대해 표준화 작업을 진행한 후 모형을 적용하였다.

4.2. Result

본 연구에서는 3장에서 소개한 DFM 기반 예측방법론들을 사용하여 GDP를 예측하였으며, 비교 방법론으로 시계열방법론인 ARIMA 또한 함께 고려하였다. 각 방법론에 대한 구체적인 적용과정은 아래와 같다.

- **[Method1: CC]** Common component based forecasting

1. $T + 1$ 시점의 GDP성장률을 예측하기 위해 직전 분기까지의 데이터에 DFM을 적용한다.
2. Factor를 구한 뒤 loading matrix를 곱해서 common component를 구한다.
3. Common component 행렬에서 한국 GDP 성장률에 해당하는 행을 추출한다.
4. 윗 단계에서 추출한 데이터의 마지막 값을 다음 분기의 GDP 성장률의 예측값으로 사용한다.

- **[Method2: OLS]** Ordinary least squares based forecasting

1. $T + 1$ 시점의 GDP성장률을 예측하기 위해 직전 분기까지의 데이터에 DFM을 적용하여, $(\hat{\mathbf{f}}_1, \dots, \hat{\mathbf{f}}_T)$ 를 얻는다.
2. 아래의 회귀분석 모형을 고려한다.

$$\mathbf{y} = \hat{\mathbf{F}}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

여기서 $\mathbf{y} = (y_2, \dots, y_T)'$ 는 한국의 GDP 성장률값이며, $\hat{\mathbf{F}} = (\hat{\mathbf{f}}_1, \dots, \hat{\mathbf{f}}_{T-1})'$ 는 DFM에서 얻은 factor이다. 즉, time lag 1인 회귀모형을 고려하였으며, 이에 대해 OLS를 활용하여 회귀계수 $\hat{\boldsymbol{\beta}}_{OLS}$ 를 얻는다.

3. 구해진 회귀계수에 T 시점의 factor를 넣어 예측값을 구한다.

$$\hat{y}_{T+1} = \hat{\mathbf{f}}_T' \hat{\boldsymbol{\beta}}_{OLS}.$$

- **[Method3: RR]** Ridge regression based forecasting

OLS와 동일한 과정을 따르며 ridge regression을 적용하여 회귀계수 $\hat{\boldsymbol{\beta}}_{ridge}$ 를 얻는다. 이 때 λ 값은 cross validation 기법으로 정하였다.

- **[Method4: LR]** Lasso regression based forecasting

OLS와 동일한 과정을 따르며 lasso regression을 적용하여 회귀계수 $\hat{\boldsymbol{\beta}}_{lasso}$ 를 얻는다. 이 때, λ 값은 cross validation 기법으로 정하였다.

- **[Method5: AR]** ARIMA forecasting

ARIMA는 시계열 데이터 예측에 가장 많이 사용되는 모형 중 하나이며 자기상관을 나타내는 AR (autoregressive)항과 이동평균을 나타내는 MA (moving average)항이 결합되어 있다. 데이터의 정상성이 만족되지 않는 경우 차분을 해야 하며, 후향 연산자 B 를 사용하여 다음과 같이 나타낼 수 있다.

$$\Phi(\mathbf{B})(1 - B)^d y_t = \delta + \Theta(\mathbf{B})\varepsilon_t.$$

위 식에서 d 는 차분횟수이며, $\Phi(\mathbf{B}) = (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)$, $\Theta = (1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q)$ 이다. 여기서 p 는 AR차수, q 는 MA차수이다. 본 논문에서는 자기상관함수 또는 편자기상관함수 그래프를 통하여 결정한 다음과 같은 AR(1) 모형을 사용하였다.

$$y_t = \delta + \phi_1 y_{t-1}.$$

즉, ARIMA모형에서는 한국 GDP 성장률 자료만 사용하였으며, 다른 나라 자료와 다른 변수는 사용하지 않는다.

각 방법론에서의 factor 개수는 다음과 같다. **Method1**에서는 1개(또는 3개)의 factor를 추출하고 common component를 구하였다. **Method2**에서도 마찬가지로 1개(또는 3개)의 factor를 추출하고 OLS를 실행하였다. 참고로, 더 많은 수의 factor를 추출한 경우에는 더 낮은 성능을 보였으므로 가장 좋은 성능을 주는 경우에 해당하는 factor 수를 사용하였다. **Method3**와 **Method4**에서는 ridge regression과 lasso regression을 적용하기 위하여 **Method1**과 **Method2**보다 많은 15개(또는 20개)로 결정하였다.

정확도를 측정하기 위한 평가지표로 다음과 같은 RMSE (root mean squared error)를 사용하였다.

$$RMSE = \sqrt{\frac{1}{H} \sum_{T^*=1}^H (\hat{y}_{T^*} - y_{T^*})^2}. \quad (4.1)$$

\hat{y}_T 은 T 시점의 예측값, y_T 은 T 시점의 실제 GDP 성장률을 나타낸다. COVID 이전은 2018년 1분기부터 2019년 4분기까지 8개의 time point를 예측하였고, COVID 이후는 2021년 1분기부터 2022년 4분기까지 8개의 time point를 예측하였으므로 모든 경우 $H = 8$ 이다.

각 방법론의 RMSE결과는 Table 2에 정리되어있다. 전반적으로 ridge regression과 lasso regression이 가장 낮은 RMSE값을 주었다. 이는 DFM을 통해 얻은 factor값들의 상관성이 어느정도 존재하므로 OLS보다는 penalty항을 추가한 regularized regression이 더 적절함을 나타낸다. 실제 2005년 1분기부터 2017년 3분기 데이터에 **Method3**를 적용하여 얻은 15개 DFM의 factor들 간의 상관계수 범위는 $-0.77 \sim 0.74$ 였다. **Method2**

Table 2: RMSE of five forecasting methods

Number of factors	CC		OLS		RR		LR		AR
	1	3	1	3	15	20	15	20	
COVID 이전	0.869	0.968	0.449	0.459	0.442	0.507	0.417	0.510	0.503
COVID 이후	0.799	0.632	0.637	0.819	0.637	0.614	0.662	0.622	0.625

Bold faces indicate best performances.

에서처럼 1개 또는 3개의 적은 수의 factor를 추출하는 경우에는 이러한 다중공선성 문제가 없지만 Table 2의 결과를 비추어볼때, 적은 수의 factor를 사용하는 경우보다는 많은 수의 factor를 추출한 후 변수 선택 또는 변수 축소 방법론을 사용하는 경우가 예측력이 더 우수함을 알 수 있었다.

5. Conclusion

본 논문에서는 G20 국가들의 거시경제 변수를 활용하여 한국의 GDP 성장률을 예측하는 방법을 제안하였다. Common component를 활용한 예측, OLS regression, ridge regression, lasso regression, ARIMA의 5가지 방법론을 적용하여 예측을 하고 성능을 비교하였다. Common component를 활용한 예측과 OLS regression에서는 factor를 1개(또는 3개)를 추출하였고, ridge regression과 lasso regression에서는 그보다 많은 15개(또는 20개)의 factor를 추출하였다. COVID가 발생한 이후 GDP와 관련된 거시경제 지표들이 급락하는 모습을 보였다. 이에 따라 COVID를 기준으로 시기를 나눠서 예측 방법론을 적용하여 보았는데, COVID 이전과 이후 모두 ridge regression과 lasso regression의 RMSE가 가장 낮았다. 한국, 일본, 인도, 인도네시아, 이스라엘 5개의 국가의 데이터를 활용하여 다시 검증을 해보았는데, 마찬가지로 ridge regression과 lasso regression의 성능이 우수하게 나타났다. 선택한 거시경제 변수와 factor 개수에 따라 예측 성능이 달라질 수 있으므로, 연구자가 실험을 통하여 이러한 것들을 결정해야 한다. 즉, 적절한 변수와 factor의 개수를 선택해서 ridge regression과 lasso regression을 적용하면, 기존에 많이 사용하던 시계열 예측 방법론인 ARIMA보다 좋은 성능의 예측 모델을 만들 수 있다.

References

- Angelini E, Camba-Mendez G, Giannone D, Reichlin L, and Rünstler G (2011). Short-term forecasts of euro area GDP growth, *The Econometrics Journal*, **14**, C25–C44.
- Doz C, Giannone D, and Reichlin L (2011). A two-step estimator for large approximate dynamic factor models based on Kalman filtering, *Journal of Econometrics*, **164**, 188–205.
- Eichler M, Motta G, and Von Sachs R (2011). Fitting dynamic factor models to non-stationary time series, *Journal of Econometrics*, **163**, 51–70.
- Forni M, Hallin M, Lippi M, and Reichlin L (2000). The generalized dynamic-factor model: Identification and estimation, *Review of Economics and Statistics*, **82**, 540–554.
- Geweke J (1977). “The Dynamic Factor Analysis of Economic Time Series,” in *Latent Variables in Socio-Economic Models*, ed. by D.J. Aigner and A.S. Goldberger, Amsterdam: North-Holland.
- Giannone D, Reichlin L, and Small D (2008). Nowcasting: The real-time informational content of macroeconomic data, *Journal of Monetary Economics*, **55**, 665–676.
- Hoerl AE and Kennard RW (1970). Ridge regression: Biased estimation for nonorthogonal problems, *Technometrics*, **12**, 55–67.

- Kim C and Kim H (2016). Study of nowcasting for GDP growth, *Bank of Korea*, **2016**.
- Lee H, Chio D, and Kim Y (2022). Development of the GDP nowcasting system based on the dynamic factor model and deep learning algorithm, *Bank of Korea*, **28(2)**, 1–37.
- Li J and Chen W (2014). Forecasting macroeconomic time series: LASSO-based approaches and their forecast combinations with dynamic factor models, *International Journal of Forecasting*, **30**, 996–1015.
- Sargent TJ and Sims CA (1977). “Business Cycle Modeling Without Pretending to Have Too Much A-Priori Economic Theory,” in *New Methods in Business Cycle Research*, ed. by C. Sims et al., Minneapolis: Federal Reserve Bank of Minneapolis.
- Stock JH and Watson MW (2002). Forecasting using principal components from a large number of predictors, *Journal of the American Statistical Association*, **97**, 1167–1179.
- Tibshirani R (1996). Regression shrinkage and selection via the lasso, *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **58**, 267–288.

Received November 20, 2023; Revised December 7, 2023; Accepted December 11, 2023

동적요인모형에 기반한 한국의 GDP 성장률 예측

이경서^a, 임예지^{1,a}

^a중앙대학교 응용통계학과

요약

GDP는 한 나라의 가계, 기업, 정부 등 모든 경제 주체가 일정 기간 동안 창출한 재화와 서비스의 시장 가치의 합을 나타낸다. GDP를 통하여 국가의 경제 규모를 파악할 수 있으며, 정부의 정책 방향에 영향을 미치는 대표적인 경제 지표이므로 이에 대한 연구가 다양하게 이루어지고 있다. 본 논문에서는 G20 국가들의 주요 거시경제 지표를 활용하여 dynamic factor model 기반의 GDP 성장률 예측 모델을 제시하였다. 추출된 factor를 다양한 회귀분석 방법론과 결합하여 그 결과들을 비교하였으며, 기존의 전통적인 시계열 예측방법인 ARIMA 모델, common component를 이용한 예측 등도 함께 비교하였다. COVID 이후 지표의 변동성이 큰 점을 고려하여 예측 시기를 COVID 전후로 나누었으며, 그 결과 factor에 대해 ridge regression과 lasso regression을 적용하여 예측한 경우 가장 좋은 성능을 나타내었다.

주요용어: 동태요인모형, 릿지 회귀분석, 라쏘 회귀분석, 자동회귀누적이동평균모델, GDP 성장률
