

시계열 분석 모형 및 머신 러닝 분석을 이용한 수출 증가율 장기예측 성능 비교

남성휘
부경대학교 경제학과 박사 수료

Comparison of long-term forecasting performance of export growth rate using time series analysis models and machine learning analysis

Seong-Hwi Nam^a

^aCompletion of Doctoral Course, Department of Economics, Pukyong National University, South Korea

Received 10 December 2021, Revised 28 December 2021, Accepted 29 December 2021

Abstract

In this paper, various time series analysis models and machine learning models are presented for long-term prediction of export growth rate, and the prediction performance is compared and reviewed by RMSE and MAE. Export growth rate is one of the major economic indicators to evaluate the economic status. And It is also used to predict economic forecast. The export growth rate may have a negative (-) value as well as a positive (+) value. Therefore, Instead of using the ReLU function, which is often used for time series prediction of deep learning models, the PReLU function, which can have a negative (-) value as an output value, was used as the activation function of deep learning models. The time series prediction performance of each model for three types of data was compared and reviewed. The forecast data of long-term prediction of export growth rate was deduced by three forecast methods such as a fixed forecast method, a recursive forecast method and a rolling forecast method. As a result of the forecast, the traditional time series analysis model, ARDL, showed excellent performance, but as the time period of learning data increases, the performance of machine learning models including LSTM was relatively improved.

Keywords: Time Series, Forecasting, Deep Learning, ARDL, VAR, LSTM, Machine Learning, Long Term Prediction, AR, ANN, DNN, SVR, KNN, Rolling, Recursive, Fixed, Trade Growth Rate

JEL Classifications:

^a First Author, namseonghwe@naver.com

© 2021 The Korea Trade Research Institute. All rights reserved.

I. 서론

1.1 연구배경

대외무역지표는 우리나라의 경제 현황을 평가하기 위한 주요 경제 지표 중 하나이다. 최근까지 수출주도성장이 우리나라의 경제적 성장에 미친 영향과 특히, 글로벌 중간재 서플라이체인(Supply Chain)의 전략적 중요성을 고려해볼 때, 수출 관련 지표는 우리나라의 경제 성장을 예측하는 데 중요한 변수가 될 수 있을 것으로 생각된다. 일부에서는, 수출에 대한 낙수효과가 국내 경제에 미치는 영향이 크지 않다는 의견도 있으나, 대표적인 경제 지표인, 국내 총생산은 소비(+)와 투자(+), 수출(+), 수입(-)으로 구성되며, 정부의 거시경제 정책은 소비, 투자, 수출을 변화시키는 것에 목적을 두고 있다. 이러한 이유로 국내 경기 예측을 위하여, 다양한 방법으로 수출과 관련한 예측이 시도되고 있다.

(강한균·김성권, 2006)은 VAR 모형을 이용하여, 한국의 대 중국·미국 FDI가 부품수출에 미치는 영향을 분석하였으며, (이서형, 2014)은 구조적 VAR 모형을 이용하여, 환율전가 효과를 분석하였다. (조중형, 2014)은 VAR 모형을 통해, 소형 승용차의 분기별 대미수출 수요예측을 분석하였다. (최종일·박순찬, 2016)은 Bayesian VAR 모형을 이용하여, 정보통신산업의 수출을 전망하는 연구를 수행하였다. (채가영·김학민, 2020)은 VAR 모형을 이용하여, 한국의 대 중국 전자상거래 수출 분야에 대한 연구를 진행하였다. (이은경, 2016)은 ARIMA 모델을 활용하여 한국의 고등교육 팽창과 경제성장 관계를 분석하였다. (이성윤·안기명, 2020)은 ARIMA, VAR, VEC 모형을 이용하여 부산항 물동량을 예측하는 연구를 수행하였다.

상기의 연구들은 전통적인 시계열 모형을 이용하여, 연구에 활용된 변수간의 관계를 분석하여, 최종적으로는 보다 정확한 경제 전망을 하고자 하는 데에 그 목적이 있다고 할 수 있다. 하지만 일부에서는 시계열분석 모형이 선형 모형을 가정하기 때문에, 비현실적이고 예

측 효율성이 떨어질 수 있다는 의견(배성완·유정석, 2018)도 있다. 부동산 시장 예측에 대한 연구로 (배성완·유정석, 2018)의 연구 등과 함께, 최근에는 경제 예측 분야에서 다양한 시계열 분석 방법이 시도되고 있으며, 특히, 머신러닝 분석 도구를 통한 연구가 이어지고 있다. 최근의 머신러닝을 활용한 거시경제 관련 연구를 살펴보면, (이재득, 2021)은 머신러닝(의사결정나무, 랜덤 포레스트, 그랜디언트 부스팅)을 활용하여 투자와 수출, 환율, 고용의 관계를 분석하였다. (김단우, 2021)은 딥러닝 및 머신러닝 알고리즘을 기반으로, 환율 예측 모형 설계 및 평가에 대한 연구를 수행하였다.

상기의 분석 내용을 종합해 보면, 국내 수출을 포함한 거시경제의 동향 평가 및 전망에 대하여, 시계열 분석 방법을 활용한 분석이 일부 이루어졌으나, 국내 총 수출 지표에 대한 전망 분석이 많지 않다는 점과, 최근 주로 활용되고 있는 머신러닝 분석 방법을 우리나라의 수출 및 거시경제 예측에 활용한 사례가 많지 않다는 것을 알 수 있다. 참고로, 국외 연구 중, (Volkan Ülke 외, 2018)은 인플레이션 예측에 대하여 CPI, Core-CPI, PCE, Core-PCE에 대하여, 시계열 분석 모형 및 머신러닝 분석 모형을 활용하여, 예측 결과에 대한 RMSE를 비교하여, 지표별 예측 모형의 성능을 비교 평가하기도 하였다. 이에 따라 본 연구의 구성은 다음과 같다. 먼저 시계열 분석 및 머신러닝에 대한 분석 모형 및 분석 자료를 소개한 후, 우리나라의 총 수출 관련 데이터 중, 특히 수출 증가율 예측에 대한 실증 분석 과정을 소개하고자 한다.

1.2 연구목적

본 연구를 통하여, 수출 증가 지표를 대상으로, 시계열 분석 방법인 전통적 시계열 분석 도구 및 머신러닝의 성능을 비교하고자 한다. 수출 증가 지표로는 우리나라의 수출 원계열의 전년 동기 대비 수출 증가율, 계절조정계열의 전기 대비 수출 증가율, 계절조정계열의 로그차분 변수를 활용하였다. 전통적 시계열 분석 도구로는 자기회귀모형(Autoregressive Model),

벡터자기회귀모형(Vector Autoregressive Model), 자기회귀시차분포모형(Autoregressive Distributed Lag Model)이 사용되었으며, 머신러닝 시계열 분석 도구로는 KNN 회귀(K-Nearest Neighbors Regression), 서포트 벡터 머신 회귀(Support Vector Machine Regression), 다층퍼셉트론 회귀(Multi Layer Perceptron Regression), 심층신경망(Deep Neural Networks)이 사용되었다. 이러한 전통적 시계열 분석 도구 및 머신러닝 분석 도구를 활용하여, 수출 증가율을 예측함으로써, 예측 성능을 비교 검토하고, 머신러닝의 수출 증가율 예측에 대한 활용 가능성을 확인하고자 한다.

1.3 기존 연구와의 차이점

본 연구와 관련된 연구 중에서, 국외 연구에서는 (Volkan Ülke 외, 2018)이 인플레이션 예측을 위해 머신러닝 및 전통적 시계열 분석 도구를 활용한 RMSE와 R2 값을 비교 검토하는 연구를 수행하였고, 국내 연구에서는 부동산 시장 예측에 대하여 (배성완·유정석, 2018)이 머신러닝 분석과 함께 전통적인 시계열 분석 방법을 활용한 예측치와 실측치 간 RMSE 및 MAE를 비교 검토하는 연구를 시행하였다.

(Volkan Ülke 외, 2018)의 연구에서는 (Karlik and Olgac, 2011)의 활성화 함수별 성과분석 연구를 참고하여 다양한 활성화 함수를 ANN(Artificial Neural Network) 분석에 이용하였으나, 인플레이션 지표 관련 시계열 예측에 있어서 ANN 분석이 뚜렷한 성과를 보여주지는 못하였다. {(Karlik and Olgac, 2011)의 연구에서는 Sigmoid 계열 함수, tanh(Hyperbolic Tangent) 함수, RBF(Radial Basis Function) 및 CSF(Conic Section Function)가 소개되고 있고, 이러한 함수들은 도출되는 함수값이 일정한 범위를 가지는 것이 특징이다.} 또한, 해당 연구에서 월별 예측 기간을 최대 1년(단기, 중기)으로 설정한 것과 달리 본 연구에서는 5년 기간 동안의 장기 예측 성과 비교를 실시하였다.

(배성완·유정석, 2018)의 연구에서는 일부 기간에서 우수한 시계열 예측 성과를 보였던, LSTM 분석 예측에서, ReLU 함수를 활성화 함

수로 사용하였다. 이는 예측되는 종속변수의 값이 정(+)의 값을 가질 때는 적절한 방법이라 할 수 있으나, 부(-)의 값을 가질 때는 그렇지 않을 수 있다. 이는 ReLU 함수를 활용할 때, 입력값이 부(-)의 값을 가질 경우 함수의 결과값이 "0"을 도출하기 때문이다. 또한, 해당 연구에서 월별 예측 기간을 최대 1년(중기)으로 설정한 것과 달리 본 연구에서는 5년 기간 동안의 예측 성과 비교를 실시하였다.

본 연구에서는 수출 증가율에 대한 예측을 실행하는 것이 연구의 목적이며, 거시 경제 지표에서 수출 증가율이 경기를 예측하는데 주요한 지표로서 활용되는 것을 감안해 볼 때, 그 의미를 찾아볼 수 있다. 이러한 수출 증가율 지표를 장기적으로 예측하기 위하여 본 연구에서는 기존 국내 연구에서 연속 시계열 분석에 있어서 좋은 성과를 보여주었던 LSTM과 함께 DNN의 활성화 함수로 PReLU(Parametric Rectified Linear Unit)을 사용하였으며, PReLU 함수는 함수의 출력값으로 특정한 범위 없이 부(-)의 값을 가질 수 있다. 그리고 일부 프로그램 패키지를 활용한 MLP의 경우, PReLU 함수를 지원하지 않기 때문에, 데이터 전처리를 수행한 후 tanh(Hyperbolic Tangent) 함수를 사용하여 예측치를 도출하였다.

또한, 기존 연구에서 예측 기간별 차이를 두고 구간고정법(Fixed Forecasting Method)을 활용한 연구를 수행한 것과 달리 본 연구에서는 예측 기간에 대하여 구간확장법(Recursive Forecasting Method), 구간이동법(Rolling Forecasting Method)의 3가지 예측방법을 사용하였다. 그리고 (Volkan Ülke 외, 2018)의 연구에서 인플레이션 관련 지표들 중에서 특정 지표마다 우수한 성과를 보이는 분석 도구가 달랐다는 점에 착안하여, 우리나라의 월별 총수출 데이터를 이용하여 원계열의 전년동기 데이터, 계절 조정 계열의 전기 데이터, 계절 조정 계열의 로그 차분 데이터를 도출하고 이를 통한 시계열 분석 도구별 예측 오차 분석을 수행하였다. 본 연구에서 정(+) 혹은 부(-)의 값을 가지는 증가율에 대하여, 예측 구간 설정 방법 및 데이터 종류에 따라 머신러닝 및 전통적 시

Table 1. 기초통계량

	단위근 검정	평균	중위수	최대값	최소값	표준편차
전년동기대비 증가율 (원계열)	우리나라 총 수출액	10.02148	9.62018	53.85097	-34.51996	15.27940
	환율	2.16644	0.14226	99.90131	-31.17172	13.94242
	수출물가지수	-0.52354	-0.06510	15.18754	-21.57466	7.66666
	전 세계 총 수출액	6.97900	7.49491	33.87417	-35.80912	11.53179
전기대비 증가율 (계절조정계열)	우리나라 총 수출액	0.91266	0.75164	25.57337	-30.04392	6.21851
	환율	0.15119	-0.00150	45.76040	-10.61115	3.06823
	수출물가지수	-0.06592	0.01588	3.52720	-7.69594	1.15202
	전 세계 총 수출액	0.54746	0.67474	11.40880	-12.52336	2.32452
로그 차분 (계절조정계열)	우리나라 총 수출액	0.00713	0.00749	0.22772	-0.35730	0.06334
	환율	0.00110	-0.00002	0.37679	-0.11217	0.02787
	수출물가지수	-0.00073	0.00016	0.03466	-0.08008	0.01163
	전 세계 총 수출액	0.00519	0.00672	0.10804	-0.13380	0.02330

계열 분석 도구를 활용한 예측 성과 비교 연구를 수행하고자 한다.

II. 자료 분석

2.1 자료

데이터는 1982년 5월부터 2019년 12월까지의 자료를 사용하였으며, 회귀 모형을 설정함에 있어, 국내 수출함수와 관련한 선행 연구 내용을 참고하였다. 먼저, 수출함수는 (이상호, 1998)의 연구에서 로그 차분 변수인 수출량(종속변수)과 수출단위가격, 상대국의 수출단위가격, 해외물가수준, 환율, 오차수정항(설명변수) 등의 선형식으로 추정되었으며, (김안호, 2001)의 연구에서는 로그 변수인 수출량(종속변수)과 수출가격, 세계수입가격, 세계수입물량(설명변수)의 선형식으로 추정되었다. 이 외에도, (홍성욱 외, 2017)의 보고서에는 아래와 같은 수출함수 기본 행태식을 도출하였다.

$$X_A = f(I_B, P_{AB}, E_A, S_{AB}, \dots) \quad (1)$$

X_A : A국의 B국 수출

I_B : B국의 수입수요요인(실질GDP, 총수입 등)

P_{AB} : A국과 B국 사이의 가격요인

E_A : 환율요인

S_{AB} : A국과 B국 사이의 특이효과(정책요인, 유가요인)

이러한 내용을 참고하여, 우리나라의 수출 총액 증가율을 종속변수로 설정하였으며, 자료는 KOSIS 국가통계포털 우리나라 수출 총액 월별 데이터를 사용하였다. 설명 변수로는 수출물가지수(달러기준), 환율, 전세계수출총액 증가율¹⁾로 설정하였다. 수출물가지수(달러기준)는 한국은행 경제통계시스템 자료를 이용하였으며, 환율은 기획재정부 e나라지표 월별 환율 데이터를 사용하였다. 총 수출액은 IMF의 총 수출액 데이터를 사용하였다.

1) 전 세계 수출 및 수입 총액에 대한 측정 오차를 고려하여, 전 세계 수입 총액 대신 전 세계 수출 총액을 사용하였음.

Table 2. 단위근 검정

월별 총수출, Korea	단위근 검정	Constant	No Constant	Constant+Trend
전년동기대비 증가율(원계열)	ADF	-7.391***	-6.101***	-7.699***
	PP	-7.279***	-5.716***	-7.672***
전기대비 증가율 (계절조정계열)	ADF	-32.302***	-31.270***	-32.476***
	PP	-33.884***	-31.205***	-34.563***
로그 차분 (계절조정계열)	ADF	-31.890***	-31.309***	-31.977***
	PP	-33.454***	-31.794***	-33.828***
월별 환율	단위근 검정	Constant	No Constant	Constant+Trend
전년동기대비 증가율(원계열)	ADF	-3.720***	-3.686***	-3.712**
	PP	-4.689***	-4.641***	-4.685***
전기대비 증가율 (계절조정계열)	ADF	-13.826***	-13.817***	-13.812***
	PP	-13.324***	-13.322***	-13.308***
로그 차분 (계절조정계열)	ADF	-13.714***	-13.714***	-13.700***
	PP	-13.194***	-13.199***	-13.179***
월별 수출물가지수	단위근 검정	Constant	No Constant	Constant+Trend
전년동기대비 증가율(원계열)	ADF	-2.233	-2.226**	-2.333
	PP	-3.920***	-3.911***	-3.999***
전기대비 증가율 (계절조정계열)	ADF	-10.508***	-10.499***	-10.558***
	PP	-10.602***	-10.592***	-10.652***
로그 차분 (계절조정계열)	ADF	-10.472***	-10.459***	-10.524***
	PP	-10.559***	-10.544***	-10.610***
월별 총 수출, World	단위근 검정	Constant	No Constant	Constant+Trend
전년동기대비 증가율(원계열)	ADF	-4.956***	-4.133***	-5.001***
	PP	-5.084***	-4.149***	-5.125***
전기대비 증가율 (계절조정계열)	ADF	-19.672***	-18.749***	-19.701***
	PP	-20.316***	-19.661***	-20.333***
로그 차분 (계절조정계열)	ADF	-19.383***	-18.577***	-19.411***
	PP	-20.065***	-19.488***	-20.081***

*** 0.01 ** 0.05 *0.1

2.2 분석방법

2.2.1 단위근 검정

시계열 자료의 정상성을 확인하기 위하여, 우리나라 총 수출액, 환율, 수출물가지수 및 세계 총 수출액 데이터의 단위근 검정을 실시하였으며, 검정방법으로 ADF 검정(Augmented Dickey Fuller) 및 PP(Phillips-Perron) 검정을 수행하였다. 단위근 검정 결과는 아래와 같으

며, 전년동기대비 증가율, 전기대비 증가율, 로그 차분 데이터 모두 정상성을 가지는 것으로 확인되었다.

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \delta_1 \Delta y_{t-1} + \dots + \delta_{p-1} \Delta y_{t-p+1} + \varepsilon_t, \text{ null}(\gamma = 0) \quad (2)$$

- ADF 검정

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \varepsilon_t, \text{ null}(\gamma = 0) \quad (3)$$

- PP 검정²⁾

Table 3. 정보기준 통계량(전년 동기 대비 증가율, AR)

Lag	AIC	BIC
AR(1)	2894.625	2906.539
AR(1/2)	2860.073	2875.958
AR(1/3)	2861.071	2880.927

Table 4. 정보기준 통계량(전년 동기 대비 증가율, VAR)

Lag	AIC	FPE	LR
0	28.6471	3.2e+07	-
1	22.6033	77015	2715.5
2	21.6454	29553.2	457.28
3	21.4278	23774.6	128.62
4	21.3934*	22971.8*	47.288
5	21.4029	23194.9	27.767
6	21.4161	23507.9	26.135
7	21.4217	23645.4	29.53*
8	21.4443	24195.2	21.952

Significance Level : 0.05

- Constant : 상수항 포함
- No Constant : 상수항 및 추세 미포함
- Constant + Trend : 상수항 및 추세 포함

2.2.2 자기회귀모형(Autoregressive Model)

(Volkan Ülke 외, 2018)의 연구에서 자기회귀모형을 사용한 거시경제지표 예측 평가를 진행하였다. 본 연구에서도 자기회귀모형을 사용하였으며, 편자기상관함수(PACF) 및 자기상관함수(ACF) 그래프 분석 및 AIC, BIC 통계량을 감안하여, 자기회귀 차수를 “2”로 설정하였다.

$$Y_t = C + \sum_{i=1}^p \beta_i Y_{t-i} + \varepsilon_t \quad (4)$$

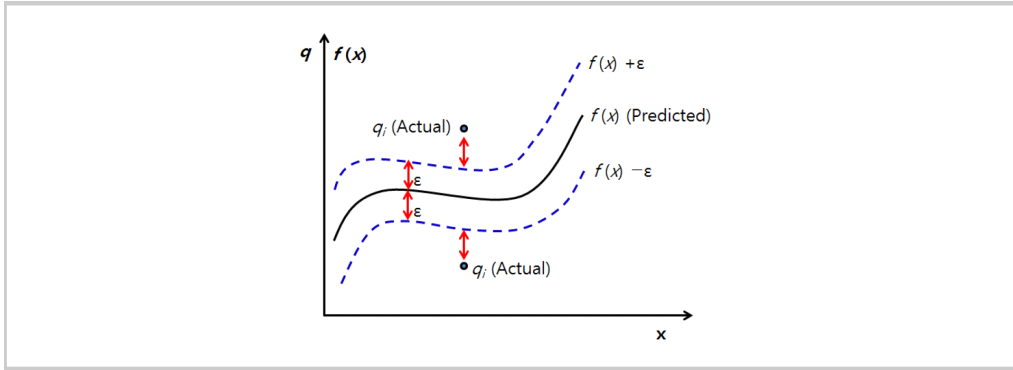
- AR(2,0) 모형

- 2) DF 검정에서는 오차항의 정규분포를 가정하나, PP 검정에서는 오차항의 자기상관 혹은 이분산이 있다고 가정하여, DF 검정에 대한 비모수적 수정을 제시하고 있음.

2.2.3 벡터자기회귀모형(Vector Autoregressive Model)

벡터자기회귀모형에서의 축약형 모형은 아래와 같으며, 전년 동기 대비 증가율 데이터를 이용하여, AIC 값을 확인 후, 벡터자기회귀모형의 시차길이를 결정하였다. 시차길이는 AIC와 FPE에서 낮은 통계량을 보인 “4”차수로 설정하였다. 이 후, 장단기메모리(Long Short Term Memory, LSTM)에서 사용될 시계열 데이터의 연속 차수 등을 고려하여, 데이터 종류에 상관없이 차수는 “4”로 통일하였다. 벡터자기회귀모형 설정을 위해 (신관호 외, 2007, 수출에 영향을 미치는 주요 결정 요인 분석 보고서(산업자원부 제출 보고서))를 일부 참고하였으며, 해당 보고서에서 수출의 주요 결정 요인으로, 수출가격, 환율 등이 언급되었으며, 이를 참고하여, 모형설정을 위한 변수로 우리나라 총 수출액, 세계 경기 변동을 반영하기 위한 세계 총 수출액, 환율, 우리나라 수출물가지수(달러기준)를

Fig. 1. SVR 개념도³⁾



사용하였다. 아래는 모형설정의 축약형 모형이다.

$$\begin{bmatrix} EX_t \\ e_t \\ EP_t \\ WEX_t \end{bmatrix} = B(L) \begin{bmatrix} EX_{t-1} \\ e_{t-1} \\ EP_{t-1} \\ WEX_{t-1} \end{bmatrix} + \begin{bmatrix} u_{EX,t} \\ u_{e,t} \\ u_{EP,t} \\ u_{WEX,t} \end{bmatrix} \quad (5)$$

EX : 우리나라 총 수출액
e : 환율 (KRW/USD)
EP : 수출물가지수
WEX : 전 세계 총 수출액

2.2.4 자기회귀시차분포모형 (Autoregressive Distributed Lag Model)

자기회귀시차분포모형은 (Volkan Ülke 외, 2018)의 연구에서 소비자 가격 지수 및 개인 소비 지출 예측에서 좋은 성과를 보여 주었다. 본 연구에서 벡터자기회귀모형과의 성과 비교 및 과적합을 방지하기 위해 자기회귀시차분포모형 역시 시차길이를 “4”로 제한하였다.

$$\begin{aligned} EX_t = & C + \sum_{k=1}^l \beta_{1k} EX_{t-k} \\ & + \sum_{k=0}^m \beta_{2k} e_{t-k} + \sum_{k=0}^n \beta_{3k} EP_{t-k} \\ & + \sum_{k=0}^o \beta_{4k} WEX_{t-k} + \varepsilon_t \end{aligned} \quad (6)$$

2.2.5 KNN 회귀(K-Nearest Neighbors Regression)

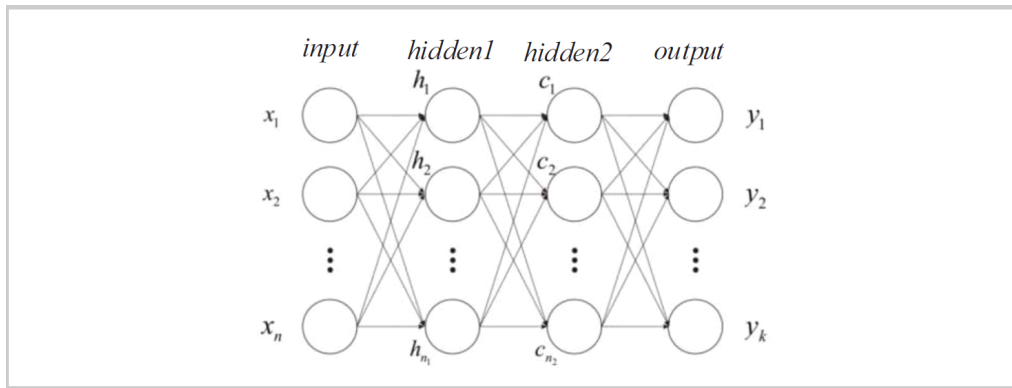
KNN 회귀는 특징 공간 내의 K개의 가까운 훈련 데이터를 입력 값으로 가지며, K개의 최근접 이웃이 가지는 평균값을 출력값으로 가지는 비모수 방식의 알고리즘이다. “가까운 이웃”을 찾는 과정은 데이터 간 거리를 계산하는 방식으로 이루어지며, KNN 회귀에서 사용되는 유클리디안 거리 계산 모형의 예시는 다음과 같다. 본 연구에서 K값은 교차검증을 통하여 n-neighbors(1~12)에서, 예측치와 실측치 간, 최소 RMSE를 가지는 “8”로 설정하였다.

$$d(X_q, X_i) = \sqrt{\sum_{r=1}^p (x_{qr} - x_{ir})^2} \quad (7)$$

2.2.6 서포트 벡터 머신 회귀(Support Vector Machine Regression)

기계 학습 분야의 하나인 서포트 벡터 머신은 초평면 혹은 초평면의 집합으로 구성되어 있으며, 특정 데이터의 분류 혹은 회귀 분석의 목적으로 사용되어진다. 분류의 목적으로 사용되어졌을 때, 초평면에서 가장 가까운 클래스의 데이터 점들 간의 거리를 가장 크게 하는 초평면을 선택하여 분류 작업을 진행하게 되는

3) 다수 인자 모델에 대한 SVR 기반 선형과 비선형회귀 분석의 비교(김준봉 외, 2021)

Fig. 2. MLP 구조도⁴⁾

데, 선형 구분이 되지 않을 경우 더 높은 차원으로 대응시켜 분리를 하는 방법이 사용되기도 한다. 반면에, 회귀 분석은 -무감도 손실함수를 계산 과정에 도입하여, 임의의 실수값을 예측하도록 고안한 것이다. SVR의 개념도는 아래와 같다.

2.2.7 다층퍼셉트론 회귀(Multi Layer Perceptron Regression)

다층 퍼셉트론은 프랑크 로젠블랫(Frank Rosenblatt, 1957)에 의해 고안된 인공신경망의 한 종류인 퍼셉트론을 여러층으로 구성된 인공신경망이다. 피드 포워드 신경망(Feed Forward Neural Network)의 기본적인 형태이며, 입력층, 은닉층, 출력층으로 구성되며, 각 층에 활성화 함수(Activation Function)를 두고 입력값을 처리한다. 아래의 내용은 각 퍼셉트론의 입력과 활성화함수를 요약 정리한 것이다. 본 연구에서는 Hyper Parameter 설정에 대하여, 최적 RMSE를 가질 수 있도록 교차검증을 실시하였으며, Hidden Layer Node(100-100-100), Max Iteration(200), Optimizer(Adam), Learning Rate(0.001)으로 설정하였다.

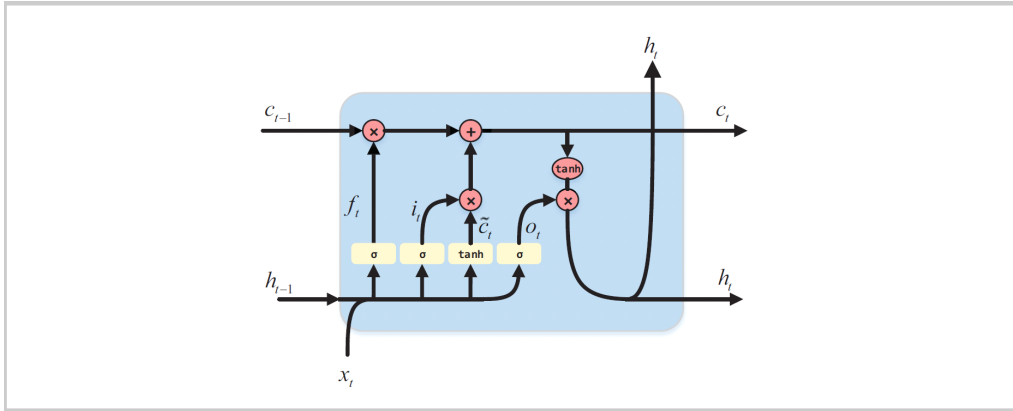
2.2.8 심층신경망 (Deep Neural Networks)

심층신경망은 입력층과 출력층 사이에 다수의 은닉층을 구성한 인공 신경망이며, 일반적으로 피드 포워드 신경망(Feed Forward Neural Network) 형태로 구성되며, 순환적으로 적용되는 순환 신경망(Recurrent Neural Network) 등으로 발전되었다. 장단기메모리(Long Short Term Memory, LSTM)는 순환 신경망 기법의 하나로써, 시계열 예측 분야에 활용되기도 한다. 본 연구에서는 Hyper Parameter 설정에 대하여, 최적 RMSE를 가질 수 있도록 교차검증을 실시하였으며, Dropout(0.2), Hidden Layer Node(50-50-50-50-50), Epochs(200), Optimizer(Adam), Learning Rate(0.001)으로 설정하였다.

2.3 자료분석

시계열 예측 과정은 우리나라 총 수출액의 전년동기 대비 증가율, 전기 증가율, 로그 차분 값의 1982년 05월부터 2014년 12월까지의 데이터를 가지고, 상기에 언급된 시계열 분석 모형을 활용하여 예측을 진행하였으며, 예측에 대한 정확도 평가는 예측치와 실제 2015년부터 2019년까지의 60개의 월별 데이터에 대한 RMSE와 MAE를 계산하여, 모형별 비교를 통하여 실시하였다.

4) Prediction of Solar Photovoltaic Power Generation Based on MLP and LSTM neural networks(Di Huang 외, 2020)

Fig. 3. LSTM 구조도⁵⁾

$$\text{RMSE}(\theta) = \sqrt{\{\sum_{i=1}^n (y_i - x_i)^2\}/n} \quad (8)$$

$$\text{MAE}(\theta) = |\sum_{i=1}^n (y_i - x_i)|/n \quad (9)$$

2.4 분석결과

시계열 예측 결과는 구간고정법(Fixed Forecasting Method), 구간확장법(Recursive Forecasting Method), 구간이동법(Rolling Forecasting Method)의 3가지 예측방법을 통해 도출하였다.

구간고정법에서는 1982년 05월부터 2014년 12월까지(LSTM⁶⁾) 혹은 1982년 09월부터 2014년 12월까지의 월별 데이터를 표본으로 설정하여, 2015년 1월부터 2019년 12월까지의 우리나라의 월별 총 수출 예측치를 추정하였다. 구간확장법에서는 t기에서 t+h기에 대한 예측치를 추정할 때, t기는 구간고정법에서 설정한 최초 기간은 동일하게 하되, 추정표본에 대하여, 1기씩 t기를 확장시켜 예측치를 추정하는 방법을

선택하였다. 구간이동법에서는 t-R+1기에서 R개의 추정표본수를 유지하여, 구간확장법과 유사하게 t기를 1기씩 확장시켜 예측치를 추정하도록 하였다.

또한 추정하는 예측치를 구간고정법에는 2015년 1월부터 2019년 12월까지의 우리나라의 월별 총 수출 예측치를 추정하도록 설정한 반면, 구간확장법 및 구간이동법에서는 t+1기에 대한 예측과 t+6기에 대한 예측치를 추정하도록 하였다.

종속변수로는 우리나라의 수출 총액 월별 데이터를 사용하였으며, 설명 변수로는 수출물가지수(달러기준), 환율, 전세계수출총액 데이터를 사용하였다. 이 때, 시계열 예측에 사용되는 데이터는 변수별로 전년동기 대비 증가율, 전기 대비 증가율, 로그 차분값을 사용하였다. 이 때, 전년 동기 대비 증가율은 데이터의 원계열을 사용하였으며, 전기 대비 증가율 및 로그 차분값에서는 계절조정계열을 사용하였으며, 계절조정계열은 X-12-ARIMA 를 통해 도출하였다.

표5-표10에서는 시계열 예측 분석 도구별 예측치와 실측치 사이의 RMSE 및 MAE 수치를 정리한 것이며, 사용된 데이터에 따라, 전년동기 대비 증가율 데이터를 사용하였을 때의 RMSE 및 MAE 수치와 전기 대비 증가율 데이터를 이용하였을 때의 RMSE 및 MAE 수치, 로그 차분값 데이터를 사용할 경우의 RMSE 및 MAE 수치를 나타낸 것이다. Fixed는 구간고정

5) Prediction of Solar Photovoltaic Power Generation Based on MLP and LSTM neural networks(Di Huang 외, 2020)

6) LSTM에서 각 설명변수의 5기의 시계열 연속 데이터와 설명변수의 최근 기수와 동일한 월의 종속변수 데이터에 대한 예측 작업을 진행하였으며, 이에 따라, LSTM을 적용함에 있어 설명변수에 대해서는 1982년 05월 데이터부터 훈련 표본으로 채웠었음.

Table 5. RMSE (데이터:전년동기대비증가율)

	Root Mean Squared Error									
	VAR	ARDL	AR(2)	DNN	LSTM	MLP	KNN	SVR Linear	SVR RBF	SVR Polynomial
Fixed	17.0528	7.1235***	8.3281**	8.4747*	8.9270	9.8627	11.0093	8.8074	8.4776	17.9592
Recursive+1	7.5954**	6.9950***	16.2473	8.7642	8.2457*	8.8236	9.7532	8.4890	8.3158	15.6412
Recursive+6	7.6364**	7.0401***	16.4587	8.7878	8.5912	9.3325	10.6043	8.6218	8.4039*	16.3821
Rolling+1	7.4539**	6.8902***	16.3155	8.4847	8.8155	8.6134	9.6166	8.1377	8.1209*	15.6412
Rolling+6	7.4840**	6.9222***	16.5537	8.6660	8.9760	10.8587	10.4913	8.2688	8.2028*	16.3821

*** The most precise ** The secondly precise * The thirdly precise

Table 6. MAE (데이터:전년동기대비증가율)

	Mean Absolute Error									
	VAR	ARDL	AR(2)	DNN	LSTM	MLP	KNN	SVR Linear	SVR RBF	SVR Polynomial
Fixed	14.2498	5.7998***	6.8245**	7.0989	6.9976*	7.9529	8.7745	7.4501	7.0211	16.0711
Recursive+1	6.0568**	5.6825***	14.2232	7.3449	6.1683*	7.1532	7.8404	7.1498	6.8773	14.0631
Recursive+6	6.0897**	5.7089***	14.4209	7.3442	6.1669*	7.6645	8.6388	7.2704	6.9701	14.7290
Rolling+1	5.9367**	5.6244***	14.2888	6.9793	6.9733	6.9052	7.6035	6.7931	6.6212*	14.0631
Rolling+6	5.9583**	5.6442***	14.5139	7.2561	6.8890	8.3326	8.4052	6.9091	6.7031*	14.7290

*** The most precise ** The secondly precise * The thirdly precise

법을 나타내며, Recursive+1은 구간확장법에서 t+1기에 대한 예측을 수행하였을 경우이며, Recursive+6은 구간확장법에서 t+6기에 대한 예측을 의미한다. 반면에 Rolling+1은 구간이동법에서 t+1기에 대한 예측을 수행하였음을 의미하며, Rolling+6은 구간이동법에서 t+6기에 대한 예측을 수행하였음을 나타낸다.

표5-표6에서 전년동기대비증가율 데이터를 통해 추정된 예측치와 실측치에 대한 RMSE와 MAE 수치가 가장 낮은 분석 도구는 ARDL 인 것으로 확인되었다. 다음으로 구간고정법에서는 AR(2)와 DNN(with RMSE)이 높은 정확도를 보였고, 구간확장법 및 구간이동법에서는 VAR 이 좋은 성과를 나타내었다. 머신 러닝 분석 도구 중, LSTM(with MAE)은 구간고정법 및 구간확장법(1기 예측)에서 좋은 성과를 보였고, SVR RBF는 구간확장법(6기 예측, with

RMSE) 혹은 구간이동법에서 좋은 성과를 나타내었다.

표7-표8에서 전기대비증가율 데이터를 통해 추정된 예측치와 실측치에 대한 RMSE와 MAE 수치가 가장 낮은 분석 도구는 ARDL 인 것으로 확인되었다. 다음으로 구간고정법에서는 AR(2) 이 높은 정확도를 보였고, 구간확장법 및 구간이동법에서는 VAR이 좋은 성과를 나타내었다. 머신 러닝 분석 도구 중, LSTM(with RMSE)은 구간고정법 및 구간확장법(1기 예측)에서 좋은 성과를 보였고, SVR Linear(with RMSE)는 구간확장법(6기 예측) 및 구간이동법에서 좋은 성과를 보였다. LSTM(with MAE)은 구간고정법, 구간확장법 및 구간이동법(6기 예측)에서 좋은 성과를 보였고, SVR Linear(with MAE)는 구간이동법(6기 예측)에서 좋은 성과를 나타내었다.

Table 7. RMSE (데이터:전기대비증가율)

	Root Mean Squared Error									
	VAR	ARDL	AR(2)	DNN	LSTM	MLP	KN N	SVR Linear	SVR RBF	SVR Polyno mial
Fixed	5.6828	4.2919***	5.0366**	5.4059	5.2563*	5.5367	5.8122	5.3113	5.3260	5.6225
Recursive+1	4.4348**	4.2828***	5.6814	5.3890	5.0447*	5.3760	5.8994	5.3169	5.3283	5.6346
Recursive+6	4.4385**	4.2830***	5.6822	5.3602	5.5092	5.4504	5.9030	5.3133*	5.3237	5.6384
Rolling+1	4.3558**	4.2168***	5.6767	5.3629	5.3554	5.4767	5.5361	5.3163*	5.3189	5.6346
Rolling+6	4.3567**	4.2137***	5.6773	5.3739	5.5369	6.0175	5.5488	5.3067*	5.3161	5.6384

*** The most precise ** The secondly precise * The thirdly precise

Table 8. MAE (데이터:전기대비증가율)

	Mean Absolute Error									
	VAR	ARDL	AR(2)	DNN	LSTM	MLP	KNN	SVR Linear	SVR RBF	SVR Polyno mial
Fixed	4.4068	3.4240***	3.9275**	4.1646	3.9983*	4.2463	4.5049	4.0975	4.1059	4.2934
Recursive+1	3.4922**	3.4179***	4.4282	4.1738	3.7239*	4.2120	4.4860	4.0960	4.1070	4.3197
Recursive+6	3.4873**	3.4165***	4.4286	4.1307	3.9649*	4.2192	4.4705	4.0947	4.1120	4.3264
Rolling+1	3.4268**	3.3740***	4.4214	4.1213	4.0027*	4.2082	4.3194	4.0961	4.0964	4.3197
Rolling+6	3.4171**	3.3694***	4.4216	4.1479	4.2004	4.5710	4.3133	4.0886*	4.1006	4.3264

*** The most precise ** The secondly precise * The thirdly precise

Table 9. RMSE (데이터:로그차분값)

	Root Mean Squared Error									
	VAR	ARDL	AR(2)	DNN	LST M	MLP	KNN	SVR Linear	SVR RBF	SVR Polynomi al
Fixed	0.0576	0.0440***	0.0509**	0.0562	0.0570	0.0546	0.0586	0.0539*	0.0541	0.0577
Recursive+1	0.0452**	0.0440***	0.0575	0.0572	0.0535*	0.0598	0.0603	0.0540	0.0539	0.0579
Recursive+6	0.0452**	0.0440***	0.0575	0.0604	0.0545	0.0525*	0.0602	0.0539	0.0540	0.0579
Rolling+1	0.0446**	0.0436***	0.0575	0.0579	0.0469*	0.0554	0.0557	0.0539	0.0538	0.0579
Rolling+6	0.0446**	0.0435***	0.0575	0.0546	0.0505*	0.0548	0.0559	0.0539	0.0537	0.0579

*** The most precise ** The secondly precise * The thirdly precise

Table 10. MAE (데이터:로그차분값)

	Mean Absolute Error									
	VAR	ARDL	AR(2)	DNN	LSTM	MLP	KNN	SVR Linear	SVR RBF	SVR Polynom ial
Fixed	0.0439	0.0343***	0.0384**	0.0432	0.0437	0.0418	0.0446	0.0409	0.0409*	0.0435
Recursive+1	0.0344**	0.0344***	0.0440	0.0423	0.0397*	0.0451	0.0452	0.0409	0.0402	0.0437
Recursive+6	0.0344**	0.0344***	0.0440	0.0467	0.0419	0.0419	0.0446	0.0409	0.0404*	0.0436
Rolling+1	0.0343**	0.0343***	0.0440	0.0455	0.0353*	0.0425	0.0424	0.0408	0.0402	0.0437
Rolling+6	0.0342**	0.0342***	0.0440	0.0402	0.0374*	0.0425	0.0425	0.0408	0.0404	0.0436

*** The most precise ** The secondly precise * The thirdly precise

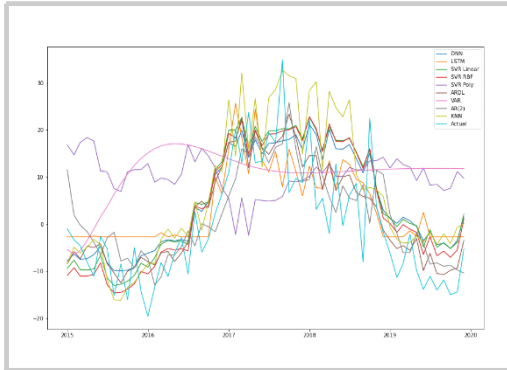
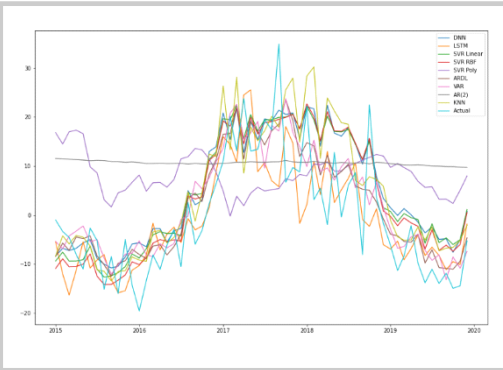
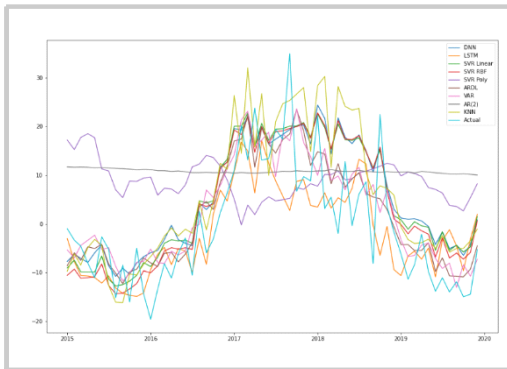
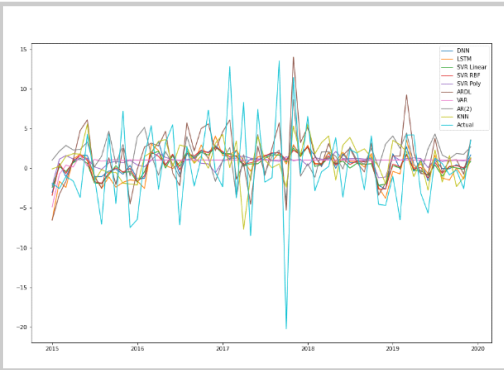
Fig. 4. 월별 수출액 전년 동기 대비 증가율 (구간고정법 예측)**Fig. 5. 월별 수출액 전년 동기 대비 증가율 (구간확장법 1기예측)****Fig. 6. 월별 수출액 전년 동기 대비 증가율 (구간확장법 6기예측)****Fig. 7. 월별 수출액 전년 동기 대비 증가율 (구간이동법 1기예측)**

표9-표10에서 로그차분값 데이터를 통해 추정된 예측치와 실측치에 대한 RMSE와 MAE 수치가 가장 낮은 분석 도구는 ARDL인 것으로 확인되었다. 다음으로 구간고정법에서는 AR(2)이 높은 정확도를 보였고, 구간확장법 및 구간이동법에서는 VAR이 좋은 성과를 나타내었다. 머신 러닝 분석 도구 중, SVR Linear(with RMSE)은 구간고정법에서 좋은 성과를 보였고, LSTM(with RMSE)은 구간확장법(1기 예측) 및 구간이동법에서 좋은 성과를 보였다. MLP(with RMSE)은 구간확장법(6기 예측)에서 좋은 성과를 보였다. SVR RBF(with MAE)는 구간고정법 및 구간확장법(6기 예측)에서 좋은 성과를 보였고, LSTM(with MAE)은 구간확장법(1기 예측) 및 구간이동법에서 좋은 성과를 보였다.

**Fig. 8. 월별 수출액 전년 동기 대비 증가율
(구간이동법 6기예측)**



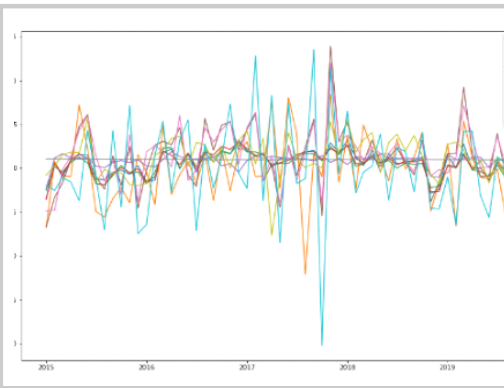
**Fig. 9. 월별 수출액 전기 대비 증가율
(구간고정법)**



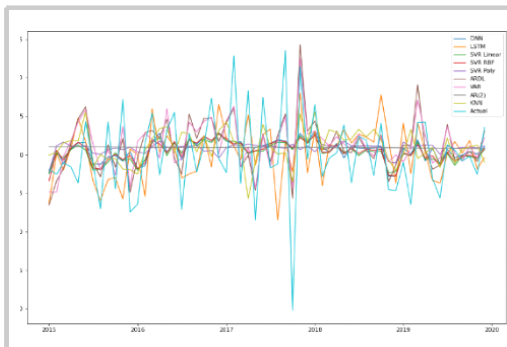
**Fig. 10. 월별 수출액 전기 대비 증가율
(구간확장법 1기예측)**



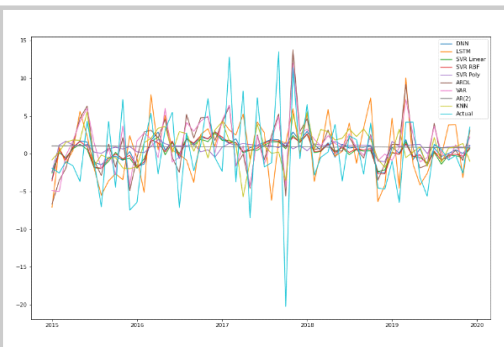
**Fig. 11. 월별 수출액 전기 대비 증가율
(구간확장법 6기예측)**



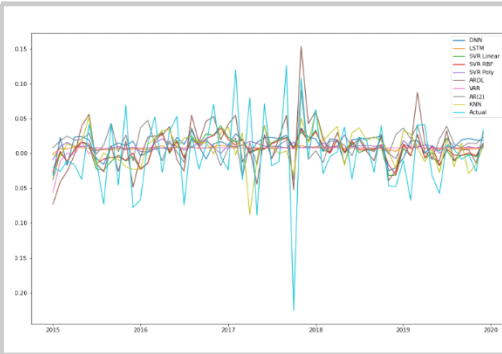
**Fig. 12. 월별 수출액 전기 대비 증가율
(구간이동법 1기예측)**



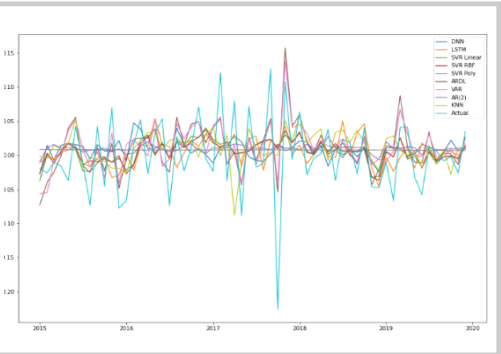
**Fig. 13. 월별 수출액 전기 대비 증가율
(구간이동법 6기예측)**



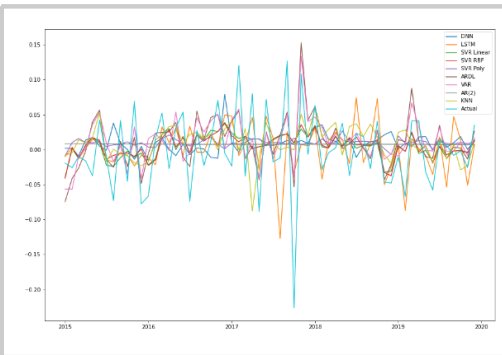
**Fig. 14. 월별 수출액 로그 차분값
(구간고정법 예측)**



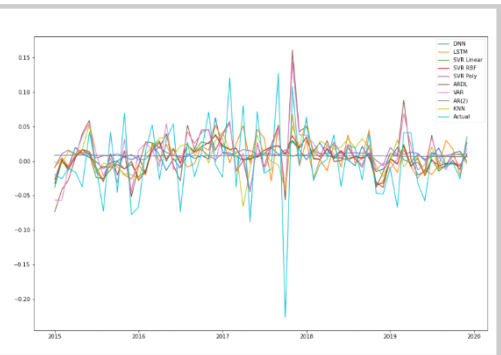
**Fig. 15. 월별 수출액 로그 차분값
(구간확장법 1기예측)**



**Fig. 16. 월별 수출액 로그 차분값
(구간확장법 6기예측)**



**Fig. 17. 월별 수출액 로그 차분값
(구간이동법 1기예측)**



**Fig. 18. 월별 수출액 로그 차분값
(구간이동법 6기예측)**

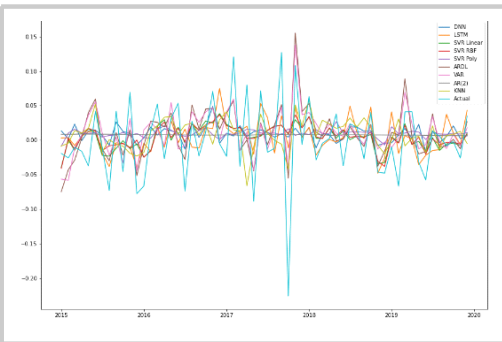


Table 11. ARDL vs 머신러닝 RMSE 비교(데이터:로그차분값)

	Root Mean Squared Error Ratio of ARDL and Machine Learning Models								
	80's			90's			00's		
	LSTM	SVR Linear	SVR RBF	LSTM	SVR Linear	SVR RBF	LSTM	SVR Linear	SVR RBF
Fixed	1.2975	1.2257	1.2317	1.3068	1.2285	1.2401	1.3024	1.2635	1.3290
Recursive+1	1.2165	1.2262	1.2246	1.3004	1.2278	1.2401	1.2746	1.2560	1.2671
Recursive+6	1.2401	1.2264	1.2281	1.3417	1.2306	1.2420	1.2744	1.2505	1.2631
Rolling+1	1.0775	1.2374	1.2339	1.2214	1.2530	1.2537	1.3413	1.2660	1.2750
Rolling+6	1.1607	1.2385	1.2358	1.2097	1.2527	1.2568	1.3271	1.2660	1.2777

Table 12. ARDL vs 머신러닝 MAE 비교(데이터:로그차분값)

	Mean Absolute Error of ARDL and Machine Learning Models								
	80's			90's			00's		
	LSTM	SVR LINEAR	SVR RBF	LSTM	SVR LINEAR	SVR RBF	LSTM	SVR LINEAR	SVR RBF
Fixed	1.2735	1.1932	1.1931	1.2805	1.2048	1.2126	1.2737	1.2531	1.3461
Recursive+1	1.1525	1.1870	1.1686	1.2298	1.1978	1.1912	1.2618	1.2371	1.2560
Recursive+6	1.2196	1.1890	1.1770	1.2758	1.2040	1.2053	1.2852	1.2356	1.2562
Rolling+1	1.0312	1.1917	1.1723	1.1613	1.2183	1.2210	1.3980	1.2407	1.2587
Rolling+6	1.0955	1.1948	1.1819	1.1826	1.2225	1.2373	1.3484	1.2433	1.2643

Table 13. ARDL vs LSTM RMSE 비교(데이터:로그차분값)

	Root Mean Squared Error Ratio of ARDL and LSTMs				
	80's RMSE		ARDL	Ratio(LSTM/ARDL)	
	LSTM PReLU	LSTM ReLU		LSTM PReLU	LSTM ReLU
Fixed	0.0570	0.0547	0.0440	1.2975	1.2450
Recursive+1	0.0535	0.0546	0.0440	1.2165	1.2402
Recursive+6	0.0545	0.0570	0.0440	1.2401	1.2952
Rolling+1	0.0469	0.0499	0.0436	1.0775	1.1443
Rolling+6	0.0505	0.0519	0.0435	1.1607	1.1926

전반적으로, ARDL이 데이터 종류와 구간 설정 방법에 상관없이 좋은 성과를 보인 부분에 주목하여, 머신러닝 분석 도구 중에서 상대적으로 좋은 성과를 나타낸 분석 도구 일부(LSTM, SVR Linear, SVR RBF)를 선정하여 각 분석 도구의 RMSE 및 MAE 값과 ARDL의

RMSE 및 MAE 값을 비교하는 작업을 수행하였다. ARDL의 RMSE 및 MAE 값은 80년대(1982년 05월부터 2014년 12월까지)의 표본 자료를 활용한 예측기간(2015년 01월부터 2019년 12월) 동안의 값을 기준으로 고정하였으며, 머신러닝 분석 도구는 RMSE 및 MAE 도출을 위한

표본 기간을 80년대(1982년 05월부터 2014년 12월까지), 90년대(1991년 01월부터 2014년 12월까지), 00년대(2000년 01월부터 2014년 12월까지)로 나누어 각 표본기간별 도출된 RMSE와 MAE 값과 ARDL의 고정된 기준 값과의 비율 분석을 실시하였다. 표11-표12에서는 각각 기간별 ARDL-머신러닝 분석 도구들 간의 RMSE 및 MAE Ratio를 나타낸다.

또한, LSTM 분석 예측에서 활성화 함수를 ReLU 대신 PReLU를 사용하였을 때, 예측 성능이 얼마나 개선되는지를 비교하기 위해 각 경우의 예측치와 실제 데이터 간의 RMSE를 도출 후, 우수한 성능을 보인 ARDL의 RMSE 결과치와 비교하는 작업을 수행하였다.

III. 결론

우리나라의 총 수출 증가율 장기 예측(5년)에 있어서, 시계열 분석 도구 중, ARDL은 데이터의 종류 및 구간 설정 방법에 상관없이 가장 좋은 성과를 나타내었다. 마찬가지로, AR(2)은 구간고정법에서 장기 시계열 예측에 대하여, 좋은 성과를 보였으며, VAR은 구간확장법 및 구간이동법에서 좋은 성과를 나타내었다. 특히, LSTM은 구간확장법 단기 예측(1기)에서 좋은 성과를 보였다.

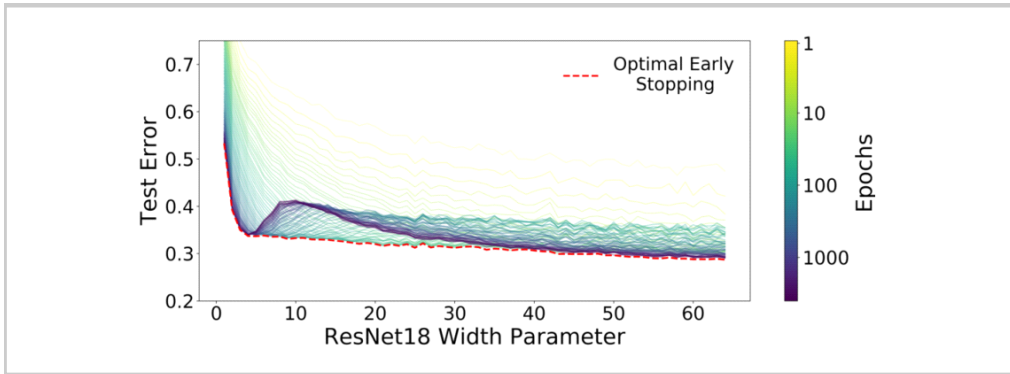
(Volkan Ülke 외, 2018)의 연구에서 인플레이션 관련 지표별, 기간별로 우수한 성과를 보이는 분석 도구가 상이하였으나, ARDL이 상대적으로 좋은 성과를 보이는 경우가 많았다. 해당 연구에서는 ANN(Artificial Neural Network)을 시계열 예측 분석에 사용하였으나, 인플레이션 지표 관련 단기(3-6개월) 및 중기(9개월-12개월) 예측에 있어서 뚜렷한 성과를 보여주지는 못하였다.

(배성완 · 유정석, 2018)의 연구에서 일부 기간에 한하여 우수한 시계열 중기 예측 성과를 보였던, LSTM을 본 연구의 장기 예측에 활용하였다. 본 연구에서 LSTM은 수출 전년 동기 대비 증가율 장기 예측에 있어서는 RMSE 값으로 보았을 때, 구간확장법(1기 예측)에서 좋은 성과를 보였고, MAE 수치로는 구간고정법 및 구

간확장법(1기 예측, 6기 예측)에서 좋은 성과를 나타내었다. 또한 수출 전기 대비 증가율 장기 예측에 있어서는 RMSE 값으로 보았을 때, 구간고정법 및 구간확장법(1기 예측)에서 좋은 성과를 보였고, MAE 수치로는 구간고정법, 구간확장법(1기 예측 및 6기 예측), 구간이동법(1기 예측 및 6기 예측)에서 좋은 성과를 나타내었다. 로그 차분값 장기 예측에 있어서는 RMSE 값으로 보았을 때, 구간확장법(1기 예측) 및 구간이동법(1기 예측 및 6기 예측)에서 좋은 성과를 보였고, MAE 수치로는 구간확장법(1기 예측) 및 구간이동법(1기 예측 및 6기 예측)에서 좋은 성과를 나타내었다. SVR Linear 및 SVR RBF 역시 일부 데이터 및 구간 설정 방법에 따라, 좋은 성과를 보였으며 이 외에도 MLP 및 DNN 역시 일부 좋은 성과를 나타내었다.

분석 도구 별 RMSE 및 MAE 결과값 비교를 통하여 ARDL이 우리나라 수출증가율의 2015년-2019년(60개월) 기간 동안의 장기예측에 있어서 가장 좋은 성과를 보여준다는 것을 확인하였고, 데이터 및 구간 설정에 따라 일부 머신러닝 분석 도구들이 상대적으로 좋은 성과를 내는 것을 확인하였다. 머신러닝 분석 도구 중, 상대적으로 좋은 성과를 내는 LSTM, SVR Linear, SVR RBF를 선택 후 로그차분값 데이터를 이용하여, ARDL과 RMSE 및 MAE 수치를 비교하였다. 이 과정에서, 선별된 머신러닝 분석 도구를 적용함에 있어서, 80년대, 90년대, 00년대로 학습 데이터 기간을 구분하여 예측치의 RMSE 및 MAE를 도출하였고, ARDL은 80년대로 학습 데이터 기간을 고정하여 예측치의 RMSE 및 MAE를 도출하였다. 이렇게 도출된 머신러닝 분석 도구의 RMSE 및 MAE 값은 ARDL의 RMSE 및 MAE 값으로 각각 나누어 RMSE Ratio를 산출하였다. 산출된 데이터를 확인해 보면, LSTM의 일부 RMSE(구간고정법, 구간확장법+1기, 구간확장법+6기) 산출값과 MAE 산출값(구간고정법)을 제외하고는 학습 데이터 기간이 짧아질수록 RMSE 및 MAE 값이 증가하였다.

(Preetum Nakkiran 외, 2019)의 연구에서는 머신러닝 모델의 Deep Double Descent 현상을 설명하고 있으며, 이는 모델의 크기나 데이

Fig. 19. Deep Double Descent⁷⁾

터 크기, 그리고 학습 횟수가 증가함에 따라 과 학습되어 예측 정확도가 떨어질 것인가에 대한 아이디어를 제시하고 있다. 해당 연구에서는 학습 데이터량을 늘려감에 따라, 어느 특정 구간에서 예측 정확도가 떨어졌다가 다시 데이터량을 지속적으로 늘려감에 따라 다시 정확도가 상승되어 감을 보여주고 있다.

이러한 아이디어를 바탕으로 데이터를 늘려감에 따라 LSTM의 일부 구간에서는 정확도가 떨어졌지만, 전반적인 ARDL-머신 러닝 분석 도구의 RMSE 및 MAE Ratio가 하락하는 것을 통하여, 향후 수출 증가율 데이터가 지속적으로 축적되고, 딥러닝 분석 도구에 있어서 학습 횟수(Epoch)를 조정하는 등의 방법을 통하여 머신 러닝 분석 도구의 정확도가 높아질 것으로 기대할 수 있다. 수출 증가율 데이터는 전년 동기 대비 혹은 전기 대비 증가율을 표현한다는 점에서, 시계열 예측의 기본 가정인 시계열 데이터의 정상성을 만족할 가능성이 높다고 할 수 있다. 전기 대비 증가율은 원 데이터의 로그 후 1기 차분한 것과 같은 효과를 가지기 때문이다. 또한 전년 동기 대비 증가율은 계절 효과를 제거해 주는 효과가 있다.

이러한 측면에서 장기 시계열 예측을 위하여 거시 경제 지표 중 하나인 수출 증가율 데이터를 채용하는 것은 그 의미가 있다고 할 수 있

다. 하지만 이러한 증가율 데이터는 부(-)의 값을 일부 가질 수 있으며, 이에 따라, 본 연구에서는 딥러닝에서 연속 시계열 예측에 주로 활용되는 ReLU 함수 대신 PReLU 함수를 사용하였다. 또한 LSTM 분석 예측에 대하여 ReLU 함수 대신 PReLU 함수를 사용하였을 때, RMSE 결과치에 대하여 구간조정법을 제외하고는 구간확장법(1기, 6기 예측) 및 구간이동법(1기, 6기 예측)에서 모두 예측 성능이 개선되는 것을 확인하였다. 하지만 여전히, LSTM의 RMSE를 ARDL의 RMSE와 비교하였을 때, 예측 오차가 작게는 7.75%에서 많게는 29.75% 수준 더 큰 것을 확인하였다.

본 연구 결과만으로도, 특정 시계열 분석 기법이 다른 분석 기법보다 정확하다고 결론짓기는 어려울 것이다. 이는 국내외 선행된 연구((Volkan Ülke 외, 2018), (배성완·유정석, 2018))에서, 데이터 종류에 따라 분석도구들의 성과가 상이하였으며, 또한 본 연구에서도 데이터 종류 및 구간 설정 방법에 따라 일부 분석 도구들의 성과가 상이함을 확인하였고, 일부 머신러닝 시계열 분석 도구와 상대적으로 우수한 성능을 보여줬던 ARDL의 RMSE 및 MAE의 비교 검토 과정을 통해, 학습 데이터 기간이 늘어남에 따라, 머신 러닝 분석 도구를 활용한 시계열 예측 데이터의 정확도가 개선되어 감을 확인하였기 때문이다. 이에 따라, 데이터의 종류 및 구간 설정 방법에 따라 적절한 예측 분석 도구를 설정할 필요가 있다고 할 수 있다.

7) Deep Double Descent : Where Bigger Models And More Data Hurt(Preetum Nakkiran 외, 2019)

하지만 여전히, 부(-)의 값을 가지는 수출 증가율에 있어서 ARDL이 높은 예측 성능을 보였다는 점에 주목할 필요가 있다. 그리고, 이와는 대조적으로 ARDL의 높은 예측 성능 때문에 과적합의 가능성을 생각해 볼 수도 있다. 하지만, 데이터의 종류와 상관없이, 구간고정법의 장기 예측(5년)에서도 ARDL이 높은 성능을 보여주면서, 설정된 ARDL 모형이 일부 기간(단기 등) 예측에 과적합 되지 않았음을 보여주는 것으로 해석해 볼 수 있다. 또한, 과적합의 위험성을 축소하기 위해, ARDL 변수별 시차를 “4” 시차

로 제한하기도 하였다. 이에 더불어, 구간확장법의 단기 예측에서 LSTM이 우수한 성과를 보여주었다는 것에도 주목할 필요가 있다. 이러한 내용을 바탕으로, 향후, 수출 증가율과 같은 거시 경제 지표 중, 부(-)의 데이터를 포함한 시계열 예측에 있어서 시계열 데이터의 지속적인 축적과 함께 다양한 활성화함수의 개발, 적절한 Hyper Parameter 설정(학습횟수 등)을 통하여, 머신 러닝 분석 도구의 정확도가 개선될 것을 기대할 수 있다.

References

- Basurko, Agirre E., Ibarra G. Berastegi and Madariaga I.(2006), “Regression and multilayer perceptron-based models to forecast hourly O3 and NO2 levels in the Bilbao area”, *Environmental Modelling & Software*, 21, 430-446.
- Bae, Seong-Wan and Yu, Jung-Suk(2018), “Predicting the Real Estate Price Index Using Machine Learning Methods and Time Series Analysis Model”, *Housing Studies Review*, 26(1), 107-133.
- Cai, Jiaying and Kim, Hag-Min(2020), “Analysis of South Korea’s Export to China in Cross-Border Electronic Commerce Using VAR Model”, *International Commerce and Information Review*, 22(4), 3-20.
- Cho, Jung-hyeong (2014), “A Study on Demand Forecasting of Export Goods Based on Vector Autoregressive Model : Subject to Each Small Passenger Vehicles Quarterly Exported to USA”, *International Commerce and Information Review*, 16(3), 73-96.
- Cho, Young-Hee and Lee, Gye-Sung(2010), “A Study on Improving Prediction Accuracy by Modeling Multiple Similar Time Series”, *The Institute of Internet, Broadcasting and Communication (IIBC)*, 10(6), 137-143.
- Choe, Jong-Il and Park, Soon-Chan(2016), “A Study on the ICT Industry Export Forecast Using Bayesian VAR Model”, *Journal of International Trade & Commerce*, 12(2), 515-527.
- Hong, Seong-Wook, Shin, Hyun-Soo, Park, Seong-Geun and Kim, Jeong-Hyeon(2017), “A study on the establishment of an export forecast model through the analysis of export determinants by country and economic region, *Korea Institute for Industrial Economics and Trade*, Report No. 2017-837, 1-143.
- Huang, Di, Zhang, Chenyu, Li, Qiang, Han, Huachun, Huang, Dawen, Li, Taotao and Wang, Chen(2020), “Prediction of Solar Photovoltaic Power Generation Based on MLP and LSTM neural networks”, *Preprints of the 4th IEEE Conference on Energy Internet and Energy System Integration*, 2744-2748.
- Joo, Il-Taek and Choi, Seung-Ho(2018), “Stock Prediction Model based on Bidirectional LSTM Recurrent Neural Network”, *Korea Information Electron Communication Technology*, 11(2), 204-208.

- Kang, Han-Gyoun and Kim, Sung-Kwon(2006), “A Comparative Analysis of the Effects of Korean Manufacturing FDI in China and the U.S.A. on the Part Exports in Korea Using VAR”, *Korean Journal of Business Administration*, 19(5), 1645-1662.
- Karlik, Bekir and Vehbi A Olgac(2011), “Performance Analysis of Various Activation Functions in Generalized MLP Architectures of Neural Networks”, *International Journal of Artificial Intelligence And Expert Systems(IJAE)*, 1(4), 111-122.
- Kim, Jun-Bong, Oh, Seung-Chul and Seo, Ki-Sung(2016), “Comparison of MLR and SVR Based Linear and Nonlinear Regressions Compensation for Wind Speed Prediction”, *The Transactions of The Korean Institute of Electrical Engineers*, 65(5), 851-856.
- Kim, Dan-Woo(2021), “Design and Evaluation of Exchange Rate Prediction Model Based on Artificial Intelligence Neural Network and Machine Learning Algorithms”, *Doctorate Thesis*, Gachon University.
- Kim, Ahn-Ho(2001), “A Consistent Estimation of Export and Import in Korea”, *Journal of Industrial Economics and Business*, 14(5), 117-130.
- Lee, Eun-Kyung(2016), “Analysis on the Relationship between Higher Education and Economic Growth: Focused on Korea and Japan”, *The Journal of Economics and Finance of Education*, 25(2),129-150.
- Lee, Sung-Yhun and Ahn, Ki-Myung(2020), “Study on the Forecasting and Relationship of Busan Cargo by ARIMA and VAR-VEC”, *Korean Institute of Navigation and Port Research*, 44(1), 44-52.
- Lee, Sang-Ho(1998), “A Consistent Estimation of Korea’s Export Function”, *Journal of International Trade and Industry Studies*, 3(2), 3-15.
- Martínez, Francisco, María P. Frías, Francisco Charte and Antonio J. Rivera(2019), “Time Series Forecasting with KNN in R: the tsfkn Package”, *The R Journal*, 11(2), 229-242.
- Nakkiran, Preetum, Gal Kaplun, Yamini Bansal, Tristan Yang, Boaz Barak and Ilya Sutskever(2020), “Deep Double Descent: Where Bigger Models and More Data Hurt”, *ICLR 2020 Conference Paper*, 1-24.
- Ülke, Volkan, Afsin Sahin and Abdulhamit Subasi(2018), “A comparison of time series and machine learning models for inflation forecasting: empirical evidence from the USA”, *Neural Comput & Applic*, 30, 1519-1527.
- Xu, Zhikun, Gao, Yabin and Jin, Ying-Ying(2014), “Application of an Optimized SVR Model of Machine Learning”, *International Journal of Multimedia and Ubiquitous Engineering*, 9(6), 67-80.
- Yi, Chae-Deug(2021), “Investment, Export, and Exchange Rate on Prediction of Employment with Decision Tree, Random Forest, and Gradient Boosting Machine Learning Models”, *Korea Trade Review*, 46(2), 281-299.
- Zhu, Shiyou, Lee, Min-Hwan, and Hwang, Kyu-Sun(2010), “The Exchange Rate Pass-through Effects on Korean Import and Consumer Prices Using a Structural VAR Model”, *The Korean Economic and Business Association*, 28(2), 85-108.