시계열 분석
(출생아 수 예측 모형)

201552001 유승우

목 차

Ⅰ 분석 개요

한국의 출생율이 바닥을 치고 있다. 정부는 지난 19년간 점진적으로 예산을 늘려가며 약 230조를 투입하고, 정책의 방향성까지 바꿔가며 저출생 문제를 극복하고자 했지만, 여전히 2020년 출생율이 OECD 회원국 중 한국이 꼴찌이며, 평균 0.84명으로 유일하게 0명대를 기록하고 있다.

우리나라의 출생율 하락세는 계속해서 진행 중이므로, 이를 극복하기 위해 출생율 예측으로 앞으로의 정책의 방향성과 예산집행의 기초자료로 활용하기 위해 분석을 실시한다.

[그림 1.1] 최근 19년 출생율 변화[1]

[그림 1.2] 최근 저출생·고령화 예산 규모[2]

1) http://news.tf.co.kr/read/ptoday/1849187.htm
2) https://m.news.zum.com/articles/54972035

Ⅱ. 시계열 분석

1. 데이터 설명

○ 데이터 설명

1) 시군구/성/월별 출생

자료갱신일: 2020-08-26 / 수록기간: 월, 년 1997.01 ~ 2019.12 / 자료문의처 : 02-2012-9114, 042-481-

일괄설정 + 항목[1/3] 시군구별[1/306] 시점[276/299]

새창보기 주석정보

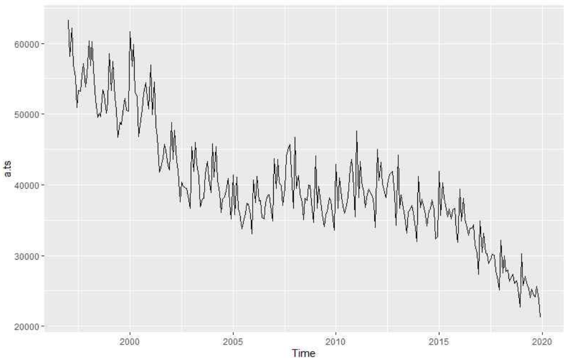
시점	전국 계 (명)
2019. 12	21,228
2019. 11	23,727
2019. 10	25,613
2019. 09	24,090
2019. 08	24,371
2019. 07	25,222
2019. 06	23,992
2019. 05	25,299
2019. 04	26,104
2019. 03	27,049
2019. 02	25,710
2019. 01	30,271
2018. 12	22,767
2018. 11	25,301
2018. 10	26,474
2018. 09	26,066
2018. 08	27,381
2018. 07	27,033
2018. 06	26,357
2018. 05	27,949
2018. 04	27,734
2018. 03	29,987

[그림 2.1.1] 월별 출생아 수3)

KOSIS 국가 통계 포털에서 1997. 01 ~ 2019. 12 기간의 월별 출생아 수를 사용하였다.

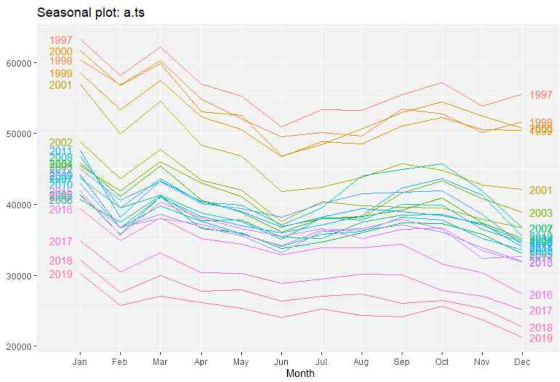
3) https://kosis.kr/statisticsList/statisticsListIndex.do?vwcd=MT_ZTITLE&menuId=M_01_01#content-group

○ 데이터 분포

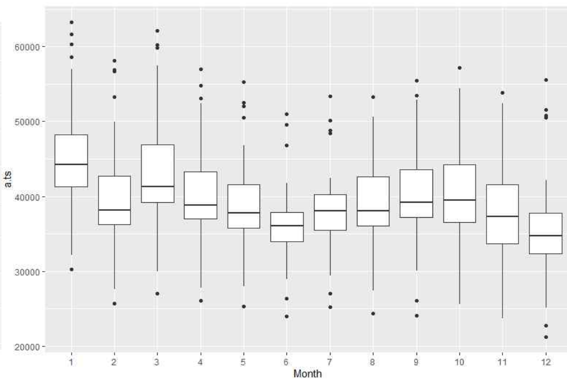


[그림 2.1.2] 출생아 수 분포

출생아 수 분포에서 출생아 수가 감소하는 추세와 계절성이 보이고, 변동 폭의 변화도 크지 않게 보인다.



[그림 2.1.3] 년도별 월별 출생아 수



[그림 2.1.4] 월별 출생아 수

월별 출생아 수는 주로 1~3월의 출생아 수가 많았고, 6~7월, 12월이 적게 나타났다

○ Train & Test 자료 분리

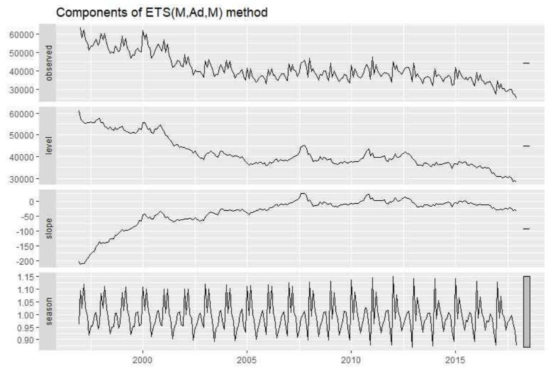
Train	1997. 01 ~ 2017. 12 기간의 출생아 수
Test	2018. 01 ~ 2019. 12 기간의 출생아 수

[표 2.1.1] Train & Test 자료 분리

시계열 예측 모형의 예측 오차를 비교하기 위해서 2년치 데이터를 Test data로 분리시켜 분석을 진행한다.

2. ETS 모델 분석

○ ETS 모형 적합



[그림 2.2.1] 분해 요소별 분포

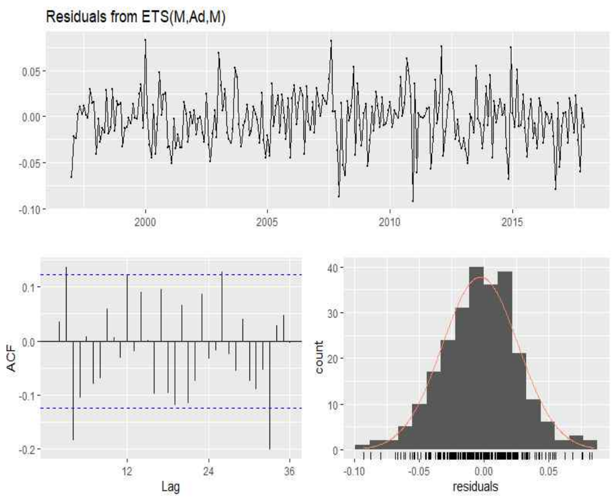
평활화 지수	
alpha	0.7617
beta	0.0042
gamma	0.2383
phi	0.9676

[표 2.2.1] 평활화 지수

분석을 시작하기에 앞서 Train data의 변동폭의 변화가 크지 않으므로 변환은 하지 않고 분석을 진행한다.

ETS모델 적합 결과 ETS(M, Ad, M) 모형이 적합되었다. alpha가 0.7617로 level에서는 큰 변동이 있고, gamma가 0.2383으로 계절에서는 약간의 변동이 있으며, beta가 0.0042로 기울기는 일정하게 나타났다.

○ ETS 모형 진단



[그림 2.2.2] 잔차 독립성 검정 결과

잔차의 독립성검정 (Ljung-Box test)	
통계량(Q)	47.462
df	7
P-value	4.535e-08

[표 2.2.2] 잔차 독립성 검정 결과

H_0 : 잔차는 독립이다.
H_1 : 잔차는 독립이 아니다.

ETS모델을 사용하기 전 오차의 가정을 만족 여부를 확인하기 위해 잔차의 독립성 검정을 실시한다. 검정 결과 P-value가 4.535e-08으로 귀무가설을 기각하여 잔차는 독립이 아니라는 충분한 근거가 있다.

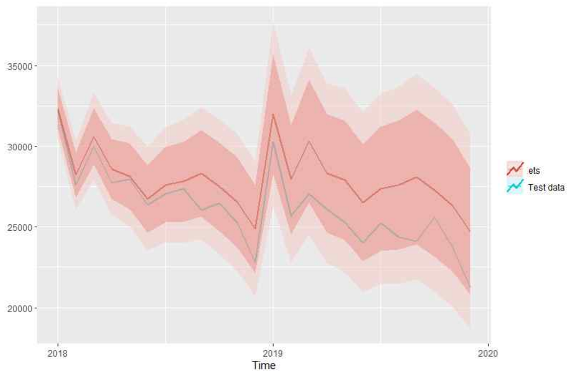
따라서, 분포가 정규성을 띄는 것으로 보이고, 독립성 가정이 위반되었다. 독립이 아니어도 예측값에는 큰 문제가 없지만, 예측 구간이 좁아질 수 있어 신뢰성이 떨어질 수 있다는 것을 인지하고 예측을 진행한다.

○ 예측 결과

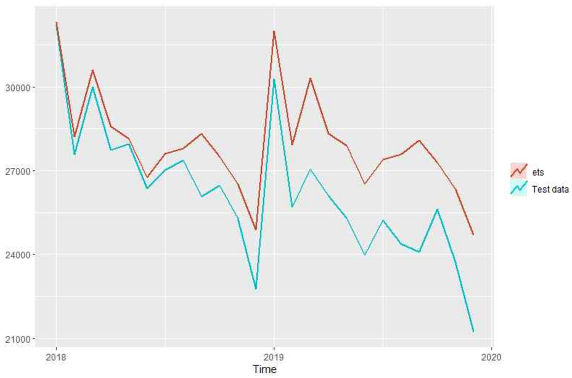
RMSE	MASE	MAPE
2077.466	0.654	7.009

[표 2.2.3] 예측 결과 주요 지표

ETS(M, Ad, M) 모형을 통한 예측과 Test data의 차이는 RMSE가 2077.466으로 나타났고, MASE가 0.653으로 1보다 작게 나타났다.



[그림 2.2.3] ETS모형 예측 1



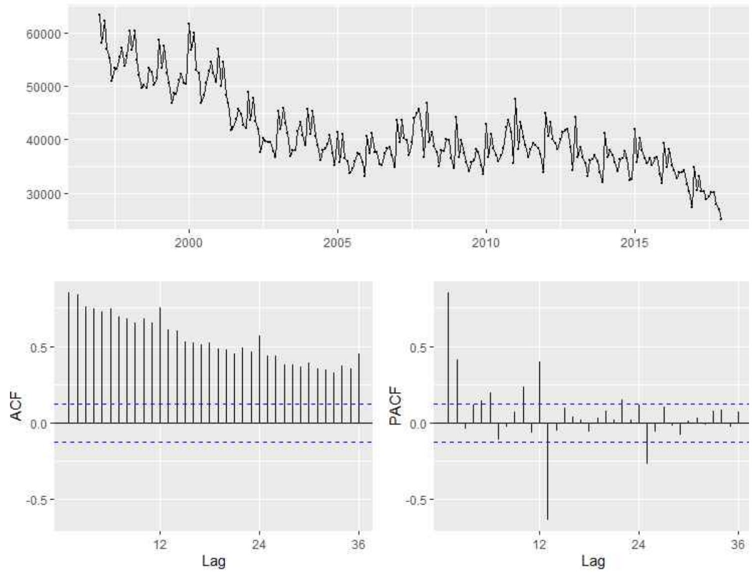
[그림 2.2.4] ETS모형 예측 2

예측이 Test data와 약간의 차이가 있지만, 형태가 비슷하게 나타났고, 신뢰구간 안에 포함되어있는 것을 확인할 수 있다.

3. ARIMA 모델 분석

○ 정상성 여부 확인

ARIMA 모델을 분석에 사용하기 위해서는 시계열 데이터가 정상성을 만족해야 하는 가정이 필요하다. 따라서, 정상성 여부을 먼저 확인하며, Train data의 변동폭의 변화가 크지 않으므로 변환은 하지 않고 분석을 진행한다.



[그림 2.3.1] Train data의 ACF, PACF

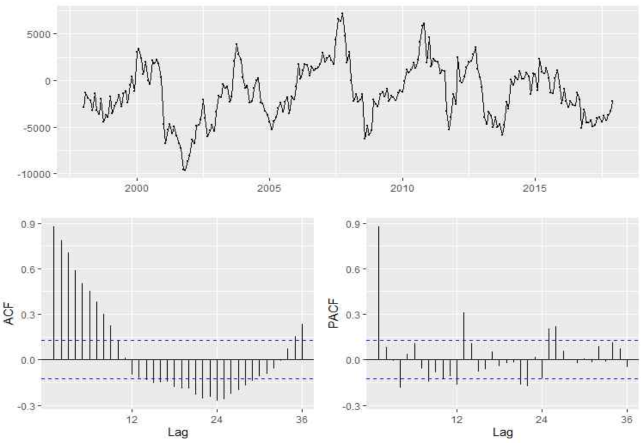
그림[2.3.1]을 보면 Train data의 ACF가 감소하는 추세를 보이고 있으며, 12, 24, 36시점에서 조금 증가한 것을 보아 계절 변동 요인이 있는 것으로 보인다. 따라서, 출생아 수 데이터는 정상성을 나타내지 못하는 데이터로 보인다.

○ 일반 & 계절 차분 차수 결정

단위근 검정	
ndiffs	1
nsdiffs	1

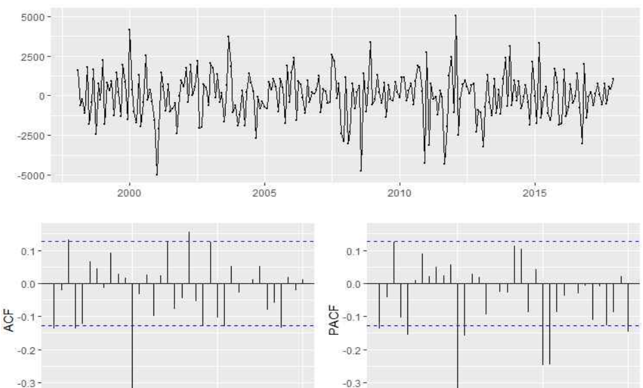
[표 2.3.1] 단위근 검정

Train data가 정상성을 나타내지 못하므로 차분 차수를 결정해야 한다. 차분 차수에 대한 검정으로 단위근 검정을 실시한 결과 ndiffs, nsdiffs가 각각 1씩 나타났고, 위의 그림[2.3.1]의 ACF를 보면 일반차분과 계절차분이 모두 필요한 것으로 보인다.



[그림 2.3.2] 계절 차분 실시 후 ACF, PACF

먼저, 계절 차분을 실시한 결과 여전히 ACF가 천천히 감소하는 추세이지만, 계절 변동 요인은 많이 감소한 것으로 보인다. 여전히 정상성을 만족하지 않는 것으로 보고, 1차 차분을 실시한다.



[그림 2.3.3] 계절 & 일반 차분 실시 후 ACF, PACF

1차 차분까지 실시한 결과 정상성을 만족하는 것으로 보이고, 일반 & 계절 차분의 차수 d와 D는 1이 적합하다고 판단된다.

○ ARIMA 모형 적합

AR과 MA의 차수를 결정하기 위한 첫 번째 방법으로 ACF, PACF를 근거로 선택하는 것과 두 번째 방법인 auto.arima를 사용하여 최소의 AICc를 가지는 모형을 선택하는 방법이 있다.

첫 번째 방법으로 그림[2.3.3]을 비계절형 관점에서 보면 ACF, PACF 모두 1시점이 같은 상황에 2시점이 들어가있는 형태를 보인다. 이것을 절단으로 본다면 2시점에서 더 많이 들어가 있는 ACF를 절단으로 보고 MA1 혹은 MA2 모형으로 볼 수 있고, 둘 다 감소로 본다면 ARMA모형으로 볼 수도 있을 것이다.

계절형 관점에서 보면 ACF가 절단, PACF가 감소하는 형태를 띄고 있으므로, 계절형의 차수는 SMA1이 적합하다고 판단된다.

두 번째 방법인 auto.arima를 통하여 나타난 최소의 AICc 모형은 ARIMA(0,1,4)(0,1,1)[12] 이고, 두 번째는 ARIMA(2,1,2)(0,1,1)[12] 이다.

두 모형 모두 예측한 후 비교하여 최종 모델을 선택해 보겠다.

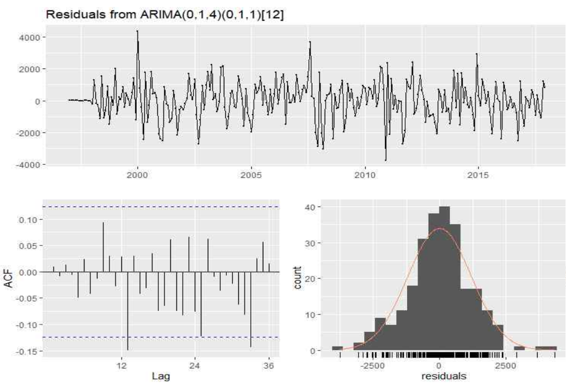
ARIMA모형	AICc
ARIMA(0,1,4)(0,1,1)[12]	4096.654
ARIMA(2,1,2)(0,1,1)[12]	4098.021

[표 2.3.2] 최소의 AICc 모델

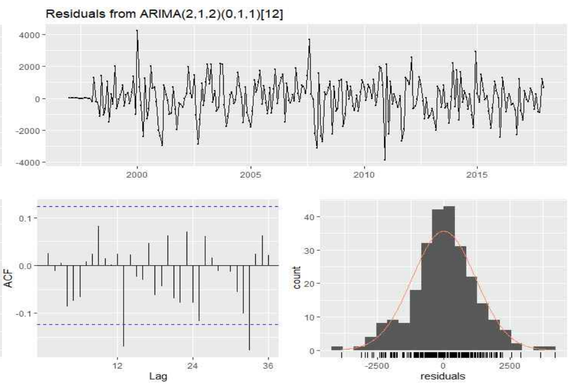
○ ARIMA 모형 진단

잔차의 독립성검정 (Ljung-Box test)			
모형	통계량	df	P-value
ARIMA(0,1,4)(0,1,1)[12]	21.664	19	0.3013
ARIMA(2,1,2)(0,1,1)[12]	24.588	19	0.1745

[표 2.3.3] 모형별 잔차의 독립성검정



[그림 2.3.4] ARIMA(0,1,4)(0,1,1)[12]의 독립성 검정



[그림 2.3.5] ARIMA(2,1,2)(0,1,1)[12]의 독립성검정

H_0 : 잔차는 독립이다.
H_1 : 잔차는 독립이 아니다.

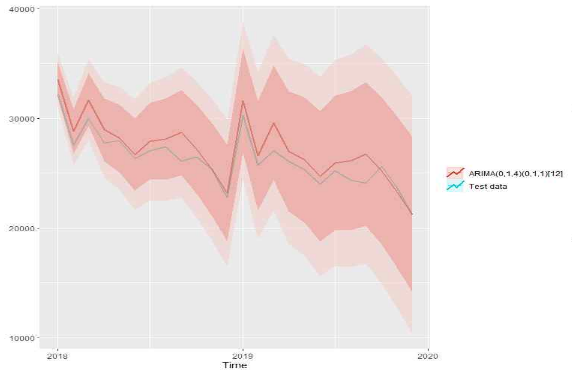
두 모형의 P-value가 각각 0.3013, 0.1745이고, 유의수준 5%에서 귀무가설을 기각하지 못하므로, 잔차는 독립이라는 충분한 근거가 있다. 각각의 ACF 에서도 2시점에서 튀어나와 있지만, 그렇게 큰 차이가 나지는 않다고 판단되므로 두 모형의 잔차는 백색잡음이라고 할 수 있다.

○ 예측 결과

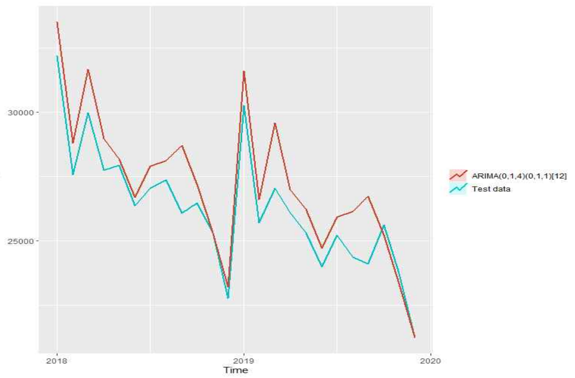
모형	RMSE	MASE	MAPE
ARIMA(0,1,4)(0,1,1)[12]	1274.239	0.383	3.876
ARIMA(2,1,2)(0,1,1)[12]	1173.846	0.352	3.583

[표 2.3.4] 예측 결과 주요 지표

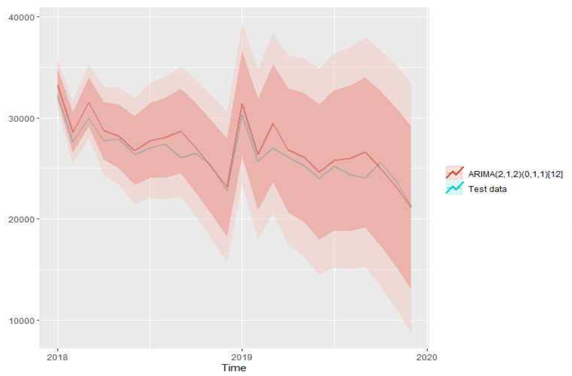
Test data와 ARIMA(2,1,2)(0,1,1)[12] 모형의 예측 오차가 ARIMA(0,1,4)(0,1,1)[12] 모형과의 예측 오차보다 더 작다.



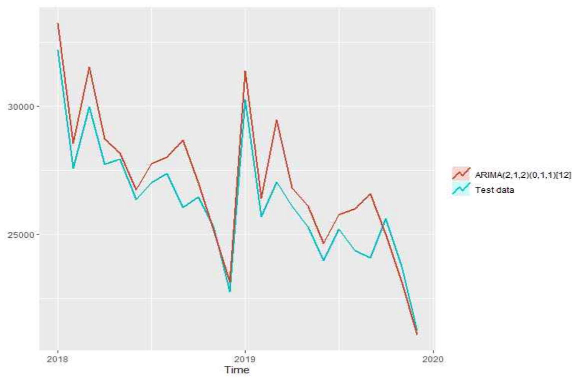
[그림 2.3.6] ARIMA(0,1,4)(0,1,1)[12] 예측 1



[그림 2.3.7] ARIMA(0,1,4)(0,1,1)[12] 예측 2



[그림 2.3.8] ARIMA(2,1,2)(0,1,1)[12] 예측 1



[그림 2.3.9] ARIMA(2,1,2)(0,1,1)[12] 예측 2

두 모형의 예측이 Test data와 약간의 차이가 있지만, 신뢰구간 안에 포함되어있는 것을 확인할 수 있다. 모형의 예측 그래프들을 눈으로 봤을 때 차이점을 찾기 힘들고, 매우 비슷하게 나타난 것으로 보인다.

따라서, 최종 ARIMA 모형은 오차가 더 작은 ARIMA(2,1,2)(0,1,1)[12] 모형이고, 모형식은 $(1+0.614B+0.707B^2)(1-B^{12})(1-B)$Yt = $(1+0.378B+0.642B^2)(1-0.65B^{12})\epsilon^t$ 이다.

ARIMA(2,1,2)(0,1,1)[12]	ar1	ar2	ma1	ma2	sma1
계 수	-0.614	-0.707	0.378	0.642	-0.650

[표 2.3.5] 최종 ARIMA 모형의 모수

4. ARMA 오차 회귀 모델 분석

○ 회귀 모형 적합

시계열 자료를 회귀 모델의 반응 변수로 놓고 추세와 계절성을 설명변수로 설정하여 분석을 실시한다. 추세변수는 선형 또는 다항 회귀모형으로, 계절변수는 더미변수 또는 sine과 cosine 함수의 선형결합으로 계절요소를 표현하는 Fourier 변수로 설명한다. 오차항은 비정상요소가 제거된 상태이기 때문에 정상성을 가정하고 분석을 진행할 수 있지만, 시계열 자료의 경우 대부분 정상성을 만족하지 못하므로 ARMA모형으로 설명한다.

시계열 자료의 회귀식
$Y_t = T_t + S_t + \epsilon_t$

[표 2.4.1] 시계열 자료의 회귀식

Train 데이터를 tslm()을 사용하여 출생아 수를 반응 변수로 두고, 설명변수인 추세변수와 계절변수는 각각 Train 데이터의 관측된 시점과 월별 더미변수로 두며, 오차항을 백색잡음이라 가정하고 회귀 모형에 적합하였다.

	Estimate	P - value
model		2.2e-16
Intercept	2010000	< 2e-16
TIME	-982.3	< 2e-16
MONTHJan	8982	5.37e-13
MONTHFeb	3622	0.002316
MONTHMar	7312	2.23e-09
MONTHApr	3759	0.001577
MONTHMay	2622	0.026725
MONTHJun	-3.339	0.997696
MONTHJul	1661	0.159025
MONTHAug	2422	0.040528
MONTHSep	4031	0.000714
MONTHOct	4103	0.000576
MONTHNov	1796	0.127931

[표 2.4.2] 회귀 모형 적합 결과

모형 적합 결과 전체적인 모델과 절편, 추세변수가 유의하게 나타났고, 12월 대비 월별 더미변수들 대부분이 유의하게 나타났지만, 6, 7, 11월은 유의하지 않게 나타났다.

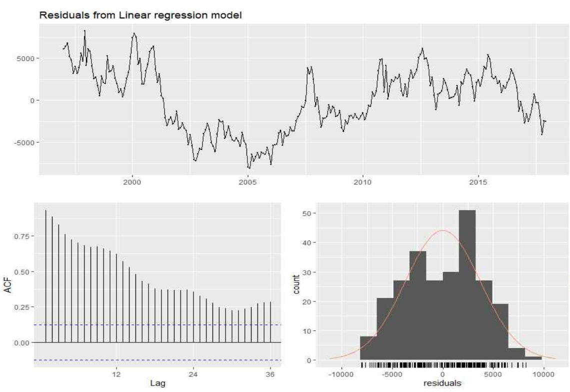
백색잡음 가정 회귀 모형식
$Y_t = 2010000 - 982.3\,Time + 8982\,MONTHJan \cdots + 1796\,MONTHNov + \epsilon_t$

[표 2.4.3] 백색잡음 가정 회귀 모형식

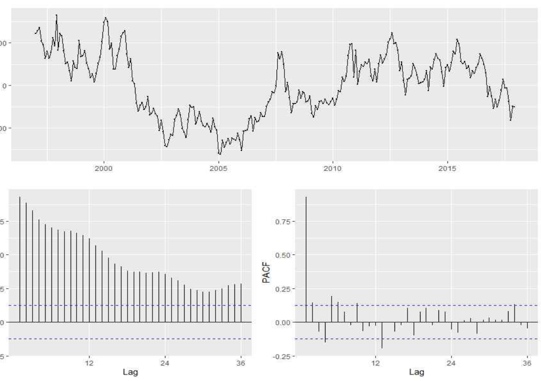
○ 회귀 모형 진단

잔차의 독립성검정 (Breusch-Godfrey test)	
통계량	225.9
df	24
P-value	< 2.2e-16

[표 2.4.4] 회귀모형 잔차의 독립성 검정



[그림 2.4.1] 회귀모형 잔차의 독립성 검정



[그림 2.4.2] 회귀모형 잔차의 ACF, PACF

H_0 : 잔차는 독립이다.
H_1 : 잔차는 독립이 아니다.

적합 후 잔차의 독립성 검정 결과 유의수준 5%에서 잔차는 독립이 아니라는 충분한 근거가 있다. 따라서, 이 모형의 잔차는 백색잡음이라 할 수 없으며, 잔차에 ARMA모형을 사용하여 정상성 만족을 충족시키려 한다. 잔차의 ACF, PACF를 보았을 때 AR(1)과 SAR(1)이 식별이 되었다.

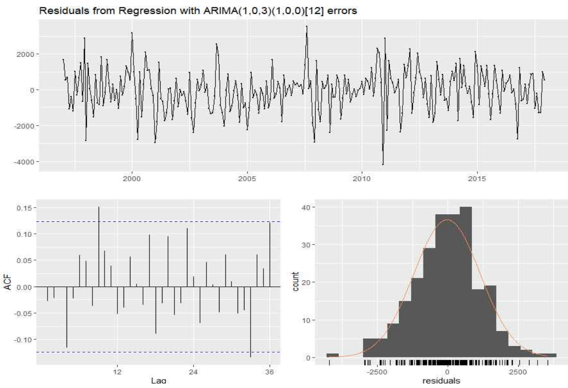
○ ARMA 오차 회귀모형 적합 I

Auto.arima()를 사용하여 반응변수에는 Train 데이터를 설명변수에는 Train의 관측 시점과 월별 더미 변수를 넣고 회귀모형에 적합한 후 남아있는 잔차에 대해 ARIMA모형을 적합한 결과 ARMA(1, 0, 3)(1, 0, 0)[12]이 적합되었다.

ARMA 오차 회귀식 I
$Y_t = 2232410.8 - 1092.923\,Time + 8483.474\,MONTHJan \cdots + 1659.472\,MONTHNov + \eta_t$
$(1-0.871B)(1-0.354B^{12})\eta_t = (1-0.043B+0.179B^2+0.200B^3)\epsilon_t$
$\epsilon_t \sim N(0,\,1494613)$

[표 2.4.5] ARMA 오차 회귀식 I

○ ARMA 오차 회귀 모형 I 진단



[그림 2.4.3] ARMA 오차 회귀식 I 잔차 독립성 검정

잔차의 독립성검정 (Ljung-Box test)	
통계량	27.818
df	6
P-value	0.0001017

[표 2.4.6] ARMA 오차 회귀식 I 잔차 독립성검정

H_0 : 잔차는 독립이다.
H_1 : 잔차는 독립이 아니다.

잔차의 독립성 검정을 한 결과 유의수준 5%에서 귀무가설을 기각하여 독립이 아니라고 나타나지만, 정규분포를 따르고 있고, 분산도 동일하게 나타난 것으로 보인다. 따라서, 신뢰구간에 대한 신빙성은 떨어지지만, 예측값에는 큰 문제가 없다고 볼 수 있다.

○ ARMA 오차 회귀 모형 적합Ⅱ

두 번째로 계절 변수에 Fourier 변수를 사용한다. Dummy 변수는 11개의 월별 변수를 사용하여 계절 변수를 설명하였지만, Fourier은 k 주기를 어떻게 설정하냐에 따라 더 작은 수의 변수로 설명할 수도 있는 장점이 있다. 최적 주기 k값을 찾기 위해 1부터 [m/2 = 6] 까지의 주기별 모형 비교를 통해 최소의 AICc 모델을 탐색한다.

k	1	2	3	4	5	6
AICc	4546.389	4509.893	4498.526	4448.533	4386.155	4327.793

[표 2.4.7] k 주기별 AICc

탐색 결과 k = 6에서의 회귀 모델이 최소의 AICc를 가지며, 11개의 sin, cos 변수를 사용한다. 앞서 Dummy로 만든 오차 회귀 모델과의 변수 개수가 같고, 모델 역시 ARMA(1, 0, 3)(1, 0, 0)[12] error로 적합되었다.

ARMA 오차 회귀식Ⅱ
$Y_t = 2235493.1 - 1092.924\,Time + 1239.668\sin(\frac{2\pi t}{12})$ $+ 1108.131\cos(\frac{2\pi t}{12}) \cdots \cdots -1015.484\cos(\frac{2\pi t}{2}) + \eta_t$
$(1-0.871B)(1-0.354B^{12})\eta_t = (1-0.043B+0.179B^2+0.200B^3)\epsilon_t$
$\epsilon_t \sim N(0, 1494613)$

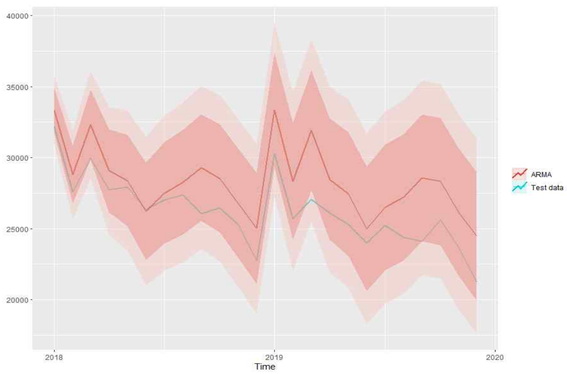
[표 2.4.8] ARMA 오차 회귀식Ⅱ

○ 예측 결과

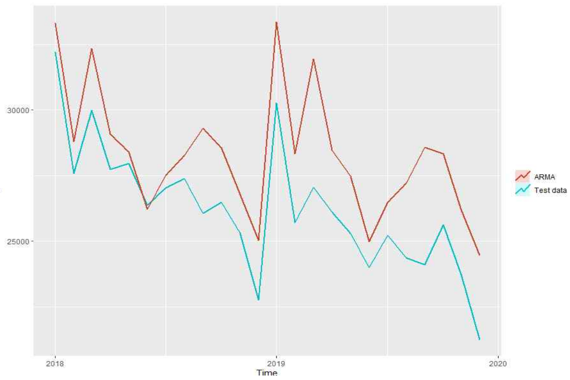
모형	RMSE	MASE	MAPE
Regression with ARMA(1, 0, 3)(1, 0, 0)[12] error	2400.222	0.777	8.131

[표 2.4.9] ARMA 오차 회귀모형 예측 결과 주요 지표

앞의 두 개의 ARMA 오차 회귀 모형 모두 ARMA(1, 0, 3)(1, 0, 0)[12] error로 적합되었고, RMSE가 2400.222, MASE가 0.777로 1보다 작게 나타났다.



[그림 2.4.5] ARMA 오차 회귀 예측 그래프1



[그림 2.4.6] ARMA 오차 회귀 예측 그래프2

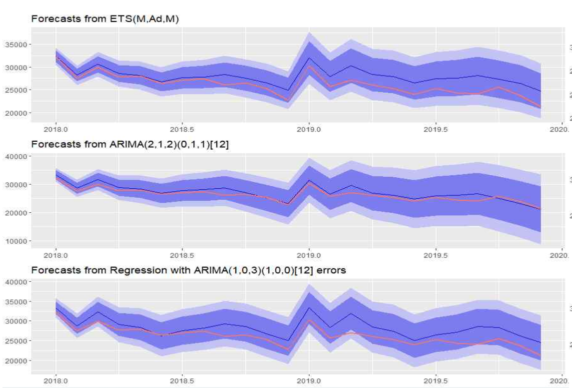
예측이 신뢰구간 안에 포함되어있는 것을 확인할 수 있지만, test 데이터와는 조금 거리가 있어 보인다.

Ⅲ. 결론

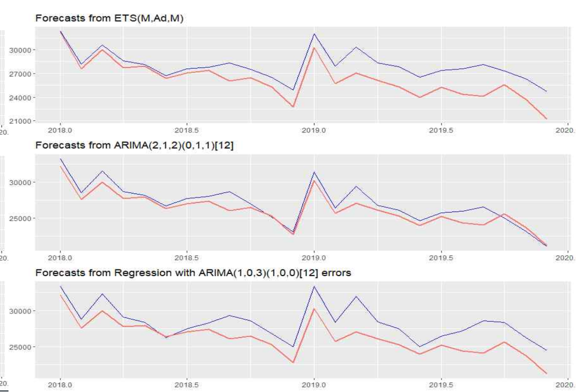
○ 최종 시계열 모형 선택

	ETS	ARIMA	ARMA 오차 회귀
RMSE	2077.466	1173.846	2400.222
MASE	0.654	0.352	0.777
MAPE	7.009	3.583	8.131
백색잡음 가정	백색잡음 O	백색잡음 X	백색잡음 X

[표 3.1] 모형별 예측 비교



[그림 3.1] 모형별 예측 비교 1



[그림 3.2] 모형별 예측 비교 2

모형별 분석 결과 Test data와 오차가 가장 작고 가까운 모형은 ARIMA모델이다.

○ 결론

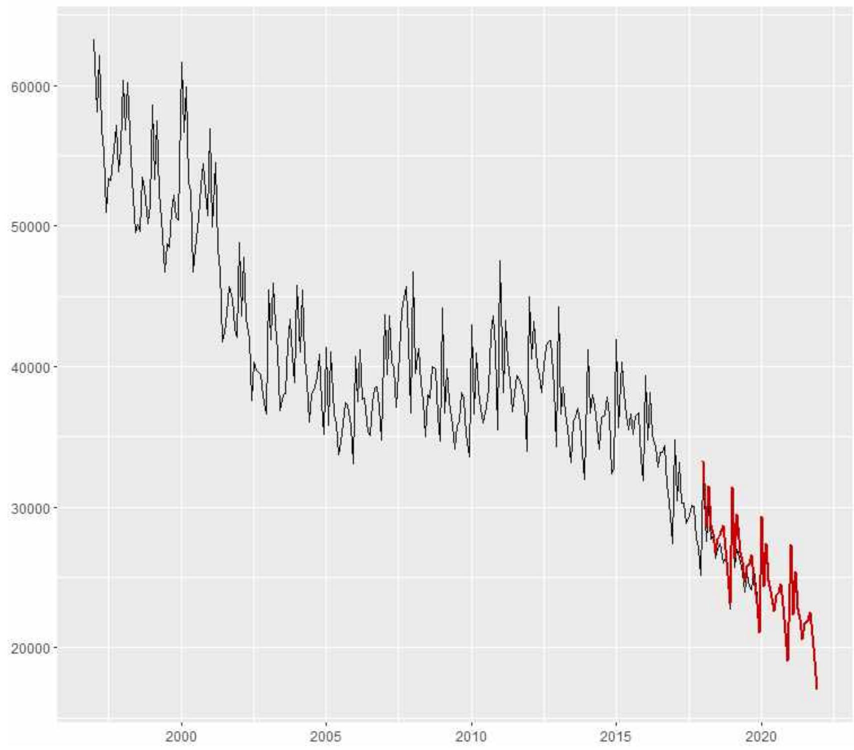


그림 [3.3] 2021년까지의 출생아 수 예측

ARIMA모델의 예측으로는 출생아 수가 앞으로도 계속 줄어들 것으로 예측되며, 저출생 문제가 더욱 심각해질 것으로 예상된다. 예측된 출생아 수를 이용하여 예산을 측정해 지원을 늘리거나 다른 문제 요인들을 파악하여 저출생 문제에 대한 대책이 강구되어야 할 필요가 있다고 보인다.

본 보고서는 1997년 1월부터 2019년 12월까지의 데이터로 예측한 보고서이다. 따라서, 현재 이후의 예측은 20년 1월 이후의 출생아 수를 포함한 새로운 모형으로 예측하는 것이 적합하며, 출생아 수 외의 여러 외부 요인들을 포함하여 모델링을 한다면 더욱 예측력이 높아질 것이다.

[부록] R 코드

```
library(tidyverse)
library(forecast)
library(urca)

# 데이터 불러오기
birth <- read.csv("C:/Data/출생아 수.csv")
birth <- birth[-1,]
birth$전국 <- as.integer(birth$전국)
str(birth)
summary(birth)

# 시계열데이터 변환 & 분포확인
birth.ts <- ts(birth$전국,start=c(1997,1), freq=12)
birth.ts

autoplot(birth.ts)

ggseasonplot(birth.ts, year.labels=TRUE,year.labels.left = TRUE)

ggsubseriesplot(birth.ts)
data.frame(birth.ts=as.numeric(birth.ts), mon=as.factor(cycle(birth.ts))) %>%
  ggplot() +
  geom_boxplot(aes(x=mon,y=birth.ts)) +
  labs(x="Month")

# train & test 분할
train <- window(birth.ts , end=c(2017,12))
test <- window(birth.ts, start=c(2018,1))

## ETS 모델
# ETS 모델 적합 & 가정 검정
fit1 <- ets(train)
summary(fit1)

autoplot(fit1)
checkresiduals(fit1)
```

```
# 예측
fc1 <- forecast(fit1, h = length(test))
accuracy(fc1, test)


autoplot(train) +
  autolayer(test, series="Test data", size=1) +
  autolayer(fc1, series="ets", size=1,PI=FALSE) +
  labs(y= NULL, color=NULL)


autoplot(test, series="Test data", size=1) +
  autolayer(fc1, series="ets", size=1, alpha=0.5) +
  labs(y= NULL, color=NULL)


autoplot(test, series="Test data", size=1) +
  autolayer(fc1, series="ets", size=1, PI = FALSE) +
  labs(y= NULL, color=NULL)



## ARIMA모델
# 정상성만족 확인

train %>% ur.kpss() %>% summary()
ggtsdisplay(train)

# 비계절,계절 차분 확인
ndiffs(train)
nsdiffs(train)

train_d <- diff(train,lag=12)
ggtsdisplay(train_d)
ggtsdisplay(diff(train_d))
```

```
# arima모형 적합
fit2 <- auto.arima(train,d=1,stepwise=FALSE,
                   approximation = FALSE,trace=TRUE)
fit3 <- auto.arima(train,d=1)

# 모델 가정 가설 검정
checkresiduals(fit2)
checkresiduals(fit3)

# 예측
fc2 <- forecast(fit2)
fc3 <- forecast(fit3)

# 정확도 비교
accuracy(fc2, test)
accuracy(fc3, test)

# 예측 그래프

autoplot(test, series="Test data", size=1) +
  autolayer(fc2, series="ARIMA(0,1,4)(0,1,1)[12]", size=1, alpha=0.5) +
  labs(y= NULL, color=NULL)

autoplot(test, series="Test data", size=1) +
  autolayer(fc2, series="ARIMA(0,1,4)(0,1,1)[12]", size=1, PI = FALSE) +
  labs(y= NULL, color=NULL)

autoplot(test, series="Test data", size=1) +
  autolayer(fc3, series="ARIMA(2,1,2)(0,1,1)[12]", size=1, alpha=0.5) +
  labs(y= NULL, color=NULL)

autoplot(test, series="Test data", size=1) +
  autolayer(fc3, series="ARIMA(2,1,2)(0,1,1)[12]", size=1, PI = FALSE) +
  labs(y= NULL, color=NULL)

# 최종 모형
coef(fit3)
confint(fit3)
fit_arima <- fit3
fc_arima <- fc3
```

```
##### ARMA 오차 회귀 모델
# 설명변수 생성
TIME <- time(train)
MONTH <- seasonaldummy(train)

# 회귀모델 적합
fit_g1 <- tslm(train ~ TIME + MONTH)
summary(fit_g1)

# 가정만족 여부 확인
checkresiduals(fit_g1)
ggtsdisplay(fit_g1$residuals)


# ARMA 오차 회귀 모델1 적합
new1 <- cbind(TIME,MONTH)
fit_g2 <- auto.arima(train,
                     xreg=new1,
                     stepwise=F,approximation=F)
summary(fit_g2)
checkresiduals(fit_g2)


# ARMA 오차 회귀 모델1 예측
new_g2 <- cbind(Time = time(test), MONTH = seasonaldummy(test))
fc_g2 <- forecast(fit_g2, xreg=new_g2)


# k별 AICc 값
Time <- time(train)
res <- vector("numeric",6)
for (i in seq(res)){
  xreg <- cbind(Time, fourier(train,K=i))
  fit <- auto.arima(train,xreg=xreg)
  res[i] <- fit$aicc
}
res
(k_min <- which.min(res))
```

```
# ARMA 오차 회귀 모델2 적합
Time <- time(train)
Fourier <- fourier(train, K=k_min)
fit_g3 <- auto.arima(train,xreg=cbind(Time,Fourier),
                      stepwise = F, approximation = F)

summary(fit_g3)
checkresiduals(fit_g3)

# ARMA 오차 회귀 모델2 예측
new_g3 <- cbind(Time = time(test),
                Fourier=fourier(test, K=k_min))

fc_g3 <- forecast(fit_g3, xreg=new_g3)

# test와 ARMA 오차 회귀 모델 그래프
autoplot(test, series="Test data", size=1) +
  autolayer(fc_g3, series="ARMA", size=1, alpha=0.5) +
  labs(y= NULL, color=NULL)


autoplot(test, series="Test data", size=1) +
  autolayer(fc_g3, series="ARMA", size=1, PI = FALSE) +
  labs(y= NULL, color=NULL)


### 모델별 test와의 오차 비교

accuracy(fc_ets, test)
accuracy(fc_arima,test)
accuracy(fc_g3,test)

## 모델별 그래프 (신뢰구간 포함)
p1 <- autoplot(fc_ets, include=0) +
  autolayer(test, size=1) +
  theme(legend.position = "none") +
  labs(x=NULL,y=NULL)
```

```
p2 <- autoplot(fc_arima, include=0) +
  autolayer(test, size=1) +
  theme(legend.position = "none") +
  labs(x=NULL,y=NULL)

p3 <- autoplot(fc_g3, include=0) +
  autolayer(test, size=1) +
  theme(legend.position = "none") +
  labs(x=NULL,y=NULL)

Rmisc::multiplot(p1,p2,p3)

## 모델별 그래프 (신뢰구간 제거)

p1 <- autoplot(fc_ets, include=0,PI=0) +
  autolayer(test, size=1) +
  theme(legend.position = "none") +
  labs(x=NULL,y=NULL)

p2 <- autoplot(fc_arima, include=0,PI=0) +
  autolayer(test, size=1) +
  theme(legend.position = "none") +
  labs(x=NULL,y=NULL)

p3 <- autoplot(fc_g3, include=0,PI=0) +
  autolayer(test, size=1) +
  theme(legend.position = "none") +
  labs(x=NULL,y=NULL)

Rmisc::multiplot(p1,p2,p3)


### 2021년 12월까지의 예측
fc4 <- forecast(fit_arima,h = 48)

autoplot(birth.ts) +
  autolayer(fc4,PI=F,col="red",size=0.8) +
  labs(x=NULL,y=NULL, color=NULL)
```