

# 서울시 아파트 실거래 가격 지수 예측

고영희

## 1. 분해법(Decomposition Method)

### 1) 데이터 시각화

데이터 시각화를 통해 데이터 분해법에 사용될 결합방식을 지정한다. 데이터의 증가폭이 일정하면 가법모형(additive model)을 사용하고, 데이터의 증가폭이 커지면 승법모형(multiplicative model)을 사용한다.

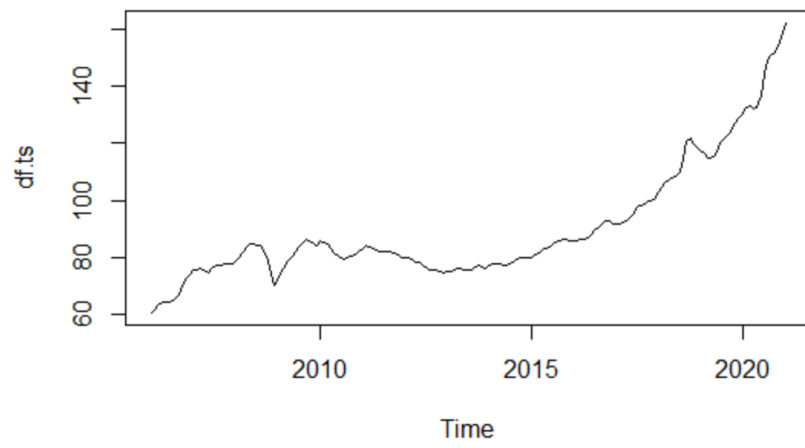
```
> head(df)

> df.ts<-ts(data=df$value,frequency = 12, start=c(2006,01))

> plot(df.ts)
```

Output]

|   | year | month | value |
|---|------|-------|-------|
| 1 | 2006 | 1     | 60.6  |
| 2 | 2006 | 2     | 61.7  |
| 3 | 2006 | 3     | 63.2  |
| 4 | 2006 | 4     | 64.1  |
| 5 | 2006 | 5     | 64.1  |
| 6 | 2006 | 6     | 64.4  |



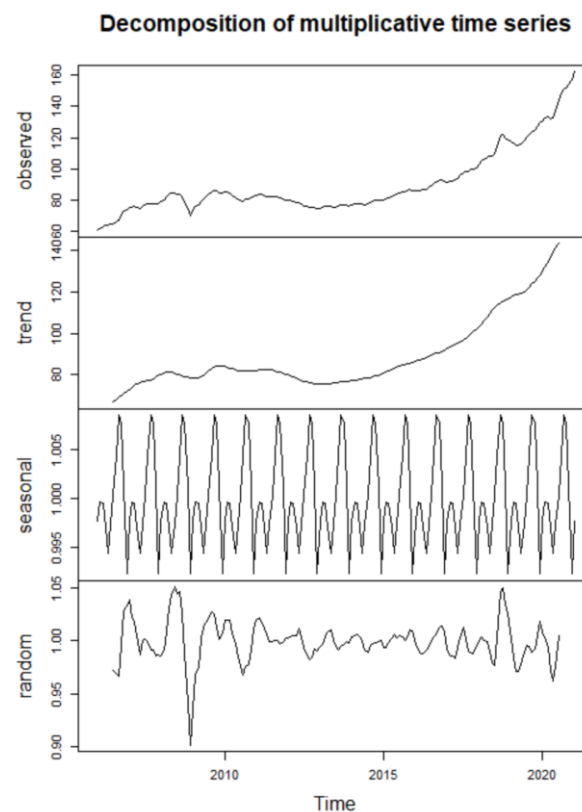
- ✓ 아파트 실거래 가격지수의 증가폭은 점차 커지기 때문에 승법 모델을 사용한다.

## 2) 시계열 분해

시계열 데이터를 시간 경과에 따른 장기적인 변동을 의미하는 추세성분(Trend Component,  $T_t$ ), 규칙적인 주기가 반복되는 변동을 의미하는 계절성분(Seasonal Component,  $S_t$ )과 시간에 관계없이 우연한 원인에 의해 나타나는 변동을 의미하는 불규칙 성분(Irregular Component,  $I_t$ )로 분해한다. 이때 각 성분을 곱셈으로 결합하는 승법모형을 사용한다. 즉  $X_t = T_t * S_t * I_t$  모형을 세우며  $\hat{x}_t = \hat{T}_t * \hat{S}_t$  를 통해 예측한다.

```
> dec.df.mul <- decompose(df.ts,type='multiplicative')  
  
> plot(dec.df.mul)
```

Output]



- ✓ 중심화 이동 평균을 사용해 추세성분을 추정하였다.
- ✓ 추세성분과 계절 성분으로 잘 분해되었고, 예측할 수 없는 2008년 금융위기 충격과 2018년 대량의 아파트 투기로 인한 데이터는 불규칙 성분에서 잘 나타나고 있다.

### 3) 계절 조정

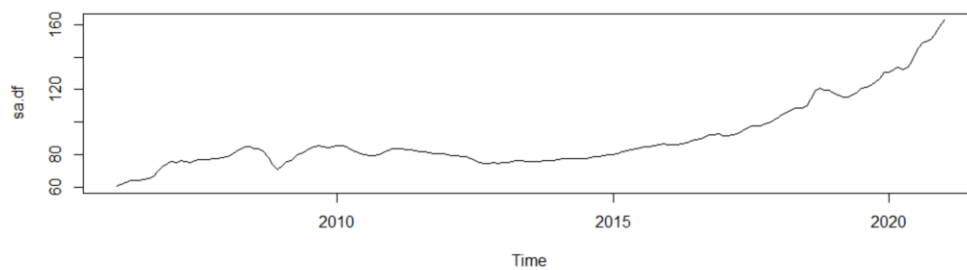
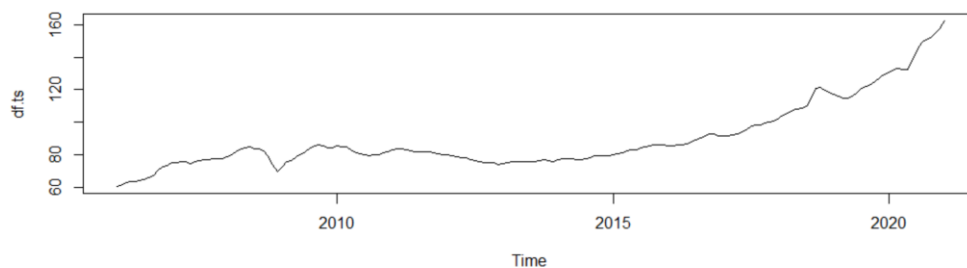
시계열 분해 시, 승법 모델을 사용했기 때문에 다음과 같은 식을 통해 계절조정을 시행한

$$\text{다. } SA_t = \frac{x_t}{\hat{s}_t}$$

```
> sa.df <- dec.df.mul$x / dec.df.mul$seasonal  
  
> head(sa.df)  
  
> par(mfrow=c(2,1))  
> plot(df.ts)  
> plot(sa.df)  
  
> dev.off()
```

Output]

|      | Jan      | Feb      | Mar      | Apr      | May      | Jun      |
|------|----------|----------|----------|----------|----------|----------|
| 2006 | 60.74179 | 61.72150 | 63.22797 | 64.29136 | 64.45985 | 64.57728 |



- ✓ 계절조정 이전의 시계열 그림 (df.ts)와 계절조정 이후의 시계열 그림 (sa.df) 가 크게 달라지지 않았다.
- ✓ 이를 통해 아파트 실거래 가격 지수는 계절의 영향을 많이 받지 않는다는 것을 알 수 있다.

#### 4) 회귀를 통한 예측

2021년 1월까지 데이터가 있기 때문에 2021.02 부터 2022.09 데이터까지 총 20개의 데이터를 예측한다. 이때 2021.01의 인덱스는 181이기 때문에 2021.02 ~ 2022.09까지의 인덱스는 182 ~ 201로 지정한다. 예측해야 하는 값이 2월부터 시작하기 때문에 추세성분 예측값과 각 계절별 예측값을 더해 최종 예측값을 구할 때 계절성분의 인덱스가 하나씩 더한 것으로 지정한다.

```
> tt<-1:length(sa.df)
> reg<-lm(sa.df~tt)
> reg

> dsn.vec<-predict(reg,newdata=data.frame(tt=182:201))
> dsn.vec

> predict.vec<-rep(NA,length(dsn.vec))
> for (i in 1:length(dsn.vec)){
  rem<-i%%12
  predict.vec[i]<-dsn.vec[i]+dec.df.mul$figure[rem+1]
}
> predict.vec
```

Output]

```
Call:
lm(formula = sa.df ~ tt)

Coefficients:
(Intercept)          tt
      61.1443       0.3174
```

|          |          |          |          |          |          |          |          |          |          |
|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|
| 1        | 2        | 3        | 4        | 5        | 6        | 7        | 8        | 9        | 10       |
| 118.9122 | 119.2296 | 119.5471 | 119.8645 | 120.1819 | 120.4993 | 120.8167 | 121.1341 | 121.4515 | 121.7689 |
| 11       | 12       | 13       | 14       | 15       | 16       | 17       | 18       | 19       | 20       |
| 122.0863 | 122.4037 | 122.7211 | 123.0385 | 123.3559 | 123.6733 | 123.9907 | 124.3081 | 124.6255 | 124.9430 |

[1] 119.9119 120.2292 120.5441 120.8589 121.1791 121.4996 121.8210 122.1426 122.4592  
 [10] 122.7701 123.0786 123.4014 123.7208 124.0381 124.3529 124.6677 124.9880 125.3085  
 [19] 125.6299 125.9515

- ✓ 계절성분을 제거했을 때 회귀식을 통한 예측값은 2번째 output 과 같고 최종 예측값을 구하기 위해 각 계절성분을 더해 3번째 output과 같은 2021.02 ~ 2022.09 까지 최종 아파트 실거래 가격 지수를 예측한다.

## 2. 지수 평활법을 통한 예측

### 1) 단순 지수 평활법

지수 평활법은 단순 지수 평활법, 선형 지수 평활법, 계절지수 평활법이 있다. 먼저 단순 지수 평활법은  $ES_{t+1} = w * x_{t+1} + (1 - w) * F_t(1)$ ,  $F_n(l) = ES_t$  로 예측한다. 이는 새롭게 주어지는 데이터에 W만큼의 가중치를 주고 (t+1)시점에는 (1-W)만큼의 가중치를 준다는 의미이다. 즉 n 시점과 n+1 시점의 값이 같을 것으로 예측한다. R프로그램에서 HoltWinters 함수를 사용하고, beta와 gamma의 값을 False로 주어 선형추세와 계절적 변동이 존재하지 않는 모형을 적합하고 alpha 평활상수를 지정하지 않아 SSE가 가장 작은 값으로 최적화 시킨다. alpha값은 약 0.99로 적합되었다.

```
> hw1 <- HoltWinters(df.ts, beta=F, gamma=F)

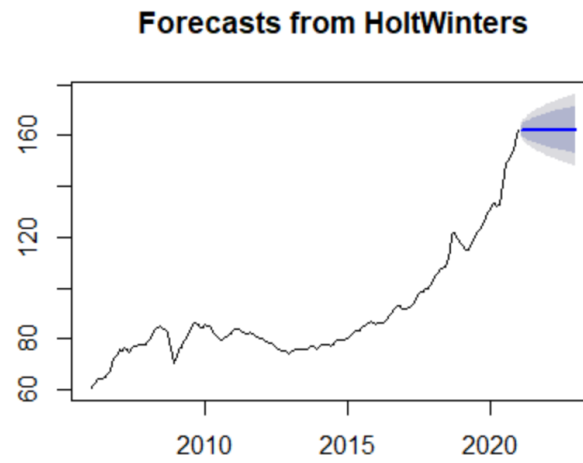
> forecast(hw1)

> plot(forecast(hw1))
```

Output]

| Point    | Forecast | Lo 80    | Hi 80    | Lo 95    | Hi 95    |
|----------|----------|----------|----------|----------|----------|
| Feb 2021 | 162.1997 | 160.3328 | 164.0665 | 159.3445 | 165.0548 |
| Mar 2021 | 162.1997 | 159.5596 | 164.8397 | 158.1620 | 166.2373 |
| Apr 2021 | 162.1997 | 158.9663 | 165.4330 | 157.2547 | 167.1447 |
| May 2021 | 162.1997 | 158.4661 | 165.9332 | 156.4897 | 167.9096 |
| Jun 2021 | 162.1997 | 158.0254 | 166.3739 | 155.8158 | 168.5836 |
| Jul 2021 | 162.1997 | 157.6270 | 166.7723 | 155.2065 | 169.1929 |
| Aug 2021 | 162.1997 | 157.2607 | 167.1386 | 154.6461 | 169.7532 |
| Sep 2021 | 162.1997 | 156.9197 | 167.4796 | 154.1246 | 170.2747 |
| Oct 2021 | 162.1997 | 156.5994 | 167.7999 | 153.6348 | 170.7645 |
| Nov 2021 | 162.1997 | 156.2965 | 168.1028 | 153.1715 | 171.2278 |
| Dec 2021 | 162.1997 | 156.0083 | 168.3910 | 152.7309 | 171.6685 |
| Jan 2022 | 162.1997 | 155.7330 | 168.6663 | 152.3098 | 172.0895 |





- ✓ 데이터가 점차 증가하는 경향성을 무시하고 미래 20개의 값이 상수처럼 취급되므로 이의 예측력은 낮을 것으로 예상할 수 있다.

## 2) 선형 지수 평활법

국소적으로 선형 추세 모델을 가정해  $F_n(l) = L_n + T_n * l$  값을 예측한다.  $L_n$ 은  $n$ 시점에서 시계열의 평균 수준을 의미하고,  $T_n$ 은  $n$  시점에서 선형추세의 기울기로 예측한다. R프로그래밍에서 HoltWinters 함수를 사용해 적합하되 gamma값을 False로 두어 선형추세는 가지나 계절적 변동이 존재하지 않는 것으로 적합한다. 이때  $n$ 시점에서 시계열의 평균 수준을 의미하는 평활상수 alpha는 1로 적합되었고, 기울기  $T_n$ 에 관한 평활상수도 1로 적합되었다.

```
> hw2<-HoltWinters(df.ts,gamma=FALSE)

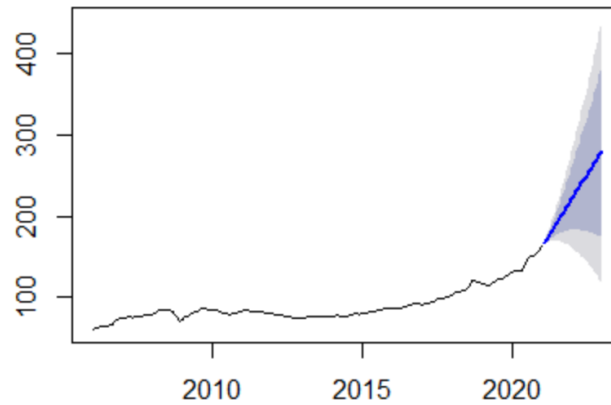
> forecast(hw2)

> plot(forecast(hw2))
```

Output]

| Point    | Forecast | Lo 80    | Hi 80    | Lo 95    | Hi 95    |
|----------|----------|----------|----------|----------|----------|
| Feb 2021 | 167.1    | 165.5948 | 168.6052 | 164.7980 | 169.4020 |
| Mar 2021 | 172.0    | 168.6343 | 175.3657 | 166.8527 | 177.1473 |
| Apr 2021 | 176.9    | 171.2682 | 182.5318 | 168.2868 | 185.5132 |
| May 2021 | 181.8    | 173.5558 | 190.0442 | 169.1916 | 194.4084 |
| Jun 2021 | 186.7    | 175.5373 | 197.8627 | 169.6282 | 203.7718 |
| Jul 2021 | 191.6    | 177.2416 | 205.9584 | 169.6407 | 213.5593 |
| Aug 2021 | 196.5    | 178.6905 | 214.3095 | 169.2628 | 223.7372 |
| Sep 2021 | 201.4    | 179.9018 | 222.8982 | 168.5214 | 234.2786 |
| Oct 2021 | 206.3    | 180.8897 | 231.7103 | 167.4384 | 245.1616 |
| Nov 2021 | 211.2    | 181.6664 | 240.7336 | 166.0322 | 256.3678 |
| Dec 2021 | 216.1    | 182.2420 | 249.9580 | 164.3186 | 267.8814 |
| Jan 2022 | 221.0    | 182.6255 | 259.3745 | 162.3112 | 279.6888 |

### Forecasts from HoltWinters



- ✓ 선형 지수 평활을 시행해 예측값을 구했을 때는 앞서 단순 지수 평활보다 더 좋은 예측을 갖는다고 할 수 있다.
- ✓ 왜냐하면 선형 지수 평활은 2021년 1월 이전의 데이터가 증가하는 추세를 잘 반영하고 있기 때문이다.
- ✓ 선형지수 평활법은 시간이 지남에 따라 신뢰구간의 길이가 점점 커지는 것을 알 수 있다.

### 3) 계절 지수 평활법

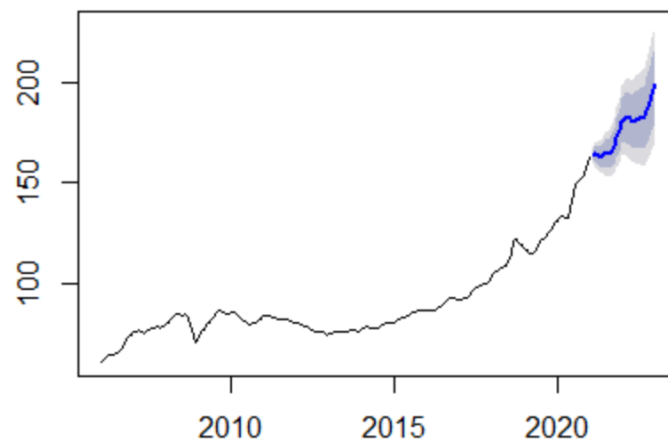
선형 지수 평활법에 계절성분을 추가한 것으로 계절성분이 가법적 계절변동인지, 승법적 계절변동인지에 따라 모형이 변화한다. 서울시 아파트 실거래 가격지수 데이터에서는 시각화를 통해 시간이 지남에 따라 증가폭이 점차 커지는 것을 보고 승법적 계절변동을 사용한다.  $F_n(l) = L_n + T_n * l + S_{n+l-L}$  식을 통해 예측값을 구한다. 이 때  $L_n$ 은 n시점에서 데이터의 평균수준을 의미하고  $T_n$ 은 기울기 값,  $S_{n+l-L}$ 은 계절 성분값을 의미한다. 평활상수 alpha는 0.94로 적합되었고, beta 평활상수는 0.04, gamma는 1로 적합되었다.

```
> hw3<-HoltWinters(df.ts,seasonal='multiplicative')  
  
> forecast(hw3)  
  
> plot(forecast(hw3))
```

Output]

| Point    | Forecast | Lo 80    | Hi 80    | Lo 95    | Hi 95    |
|----------|----------|----------|----------|----------|----------|
| Feb 2021 | 163.8161 | 161.1028 | 166.5294 | 159.6665 | 167.9657 |
| Mar 2021 | 164.2271 | 160.4342 | 168.0201 | 158.4264 | 170.0279 |
| Apr 2021 | 163.6623 | 159.0019 | 168.3227 | 156.5349 | 170.7898 |
| May 2021 | 162.4715 | 157.0613 | 167.8817 | 154.1973 | 170.7457 |
| Jun 2021 | 163.8969 | 157.7236 | 170.0703 | 154.4557 | 173.3382 |
| Jul 2021 | 164.1591 | 157.3010 | 171.0172 | 153.6706 | 174.6476 |
| Aug 2021 | 164.3299 | 156.8197 | 171.8401 | 152.8440 | 175.8158 |
| Sep 2021 | 165.2255 | 157.0531 | 173.3980 | 152.7268 | 177.7243 |
| Oct 2021 | 168.0180 | 159.1064 | 176.9297 | 154.3888 | 181.6472 |
| Nov 2021 | 172.4876 | 162.7538 | 182.2214 | 157.6011 | 187.3741 |
| Dec 2021 | 175.8962 | 165.3951 | 186.3973 | 159.8362 | 191.9562 |
| Jan 2022 | 180.5375 | 169.3633 | 191.7117 | 163.4480 | 197.6270 |
| Feb 2022 | 182.1635 | 170.2600 | 194.0669 | 163.9587 | 200.3683 |
| Mar 2022 | 182.4505 | 169.9763 | 194.9246 | 163.3729 | 201.5281 |
| Apr 2022 | 181.6566 | 168.6882 | 194.6250 | 161.8231 | 201.4900 |
| May 2022 | 180.1727 | 166.7652 | 193.5801 | 159.6678 | 200.6775 |
| Jun 2022 | 181.5927 | 167.5396 | 195.6458 | 160.1003 | 203.0851 |

### Forecasts from HoltWinters



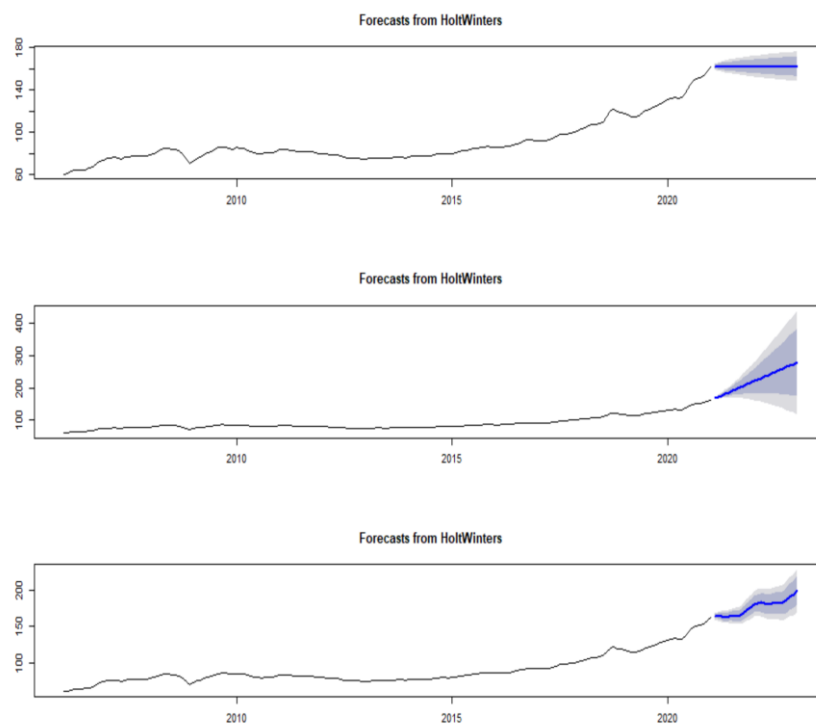
- ✓ 선형지수 평활법에서 계절성분이 추가되었기 때문에 예측선은 직선이 아닌 구불구불한 선의 형태를 볼 수 있다.
- ✓ 이 때의 신뢰구간은 선형지수 평활법보다 좀 더 좁은 구간임을 확인할 수 있다.

#### 4) 평활법 선택

앞서 단순 지수 평활법, 선형지수 평활법, 계절지수 평활법을 통해 서울시 아파트 실거래 가격 지수를 2021년 2월 ~ 2022년 9월 데이터까지 예측해보았다. 세 가지의 예측방법 중, 가장 좋은 최종 방법을 선택하기 위해 3개의 예측선을 비교해보고, SSE가 가장 작은 예측 방법을 선택한다.

```
> par(mfrow=c(3,1))  
  
> plot(forecast(hw1))  
> plot(forecast(hw2))  
> plot(forecast(hw3))  
  
> sse.vec<-rep(NA,3)  
> sse.vec[1]<-hw1$SSE  
> sse.vec[2]<-hw2$SSE  
> sse.vec[3]<-hw3$SSE  
  
> sse.vec
```

Output]



|   |
|---|
| [1]    437.2017    245.6200    754.0099 |
|---|

- ✓ 데이터와 추정 모델 사이의 불일치를 평가하는 척도인 SSE가 가장 작은 모형2(선형  
지수 평활법)을 선택한다.
- ✓ 따라서 2022년 9월의 서울시 아파트 실거래 가격지수는 80% 신뢰구간 [179.5643,  
340.8357] 내 260.2 값을 갖는다고 예측할 수 있다.