

ARIMA모형 차수결정

다음과 같은 방법을 사용하면 ARIMA 모형의 차수를 결정하기 위한 정보를 얻을 수 있다.

- Augmented Dickey-Fuller 검정
- 자기상관계수 함수
- 편자기상관계수 함수

이러한 방법을 실제 시계열 자료에 적용하여 ARIMA 모형의 차수를 알아보자.

In [1]:

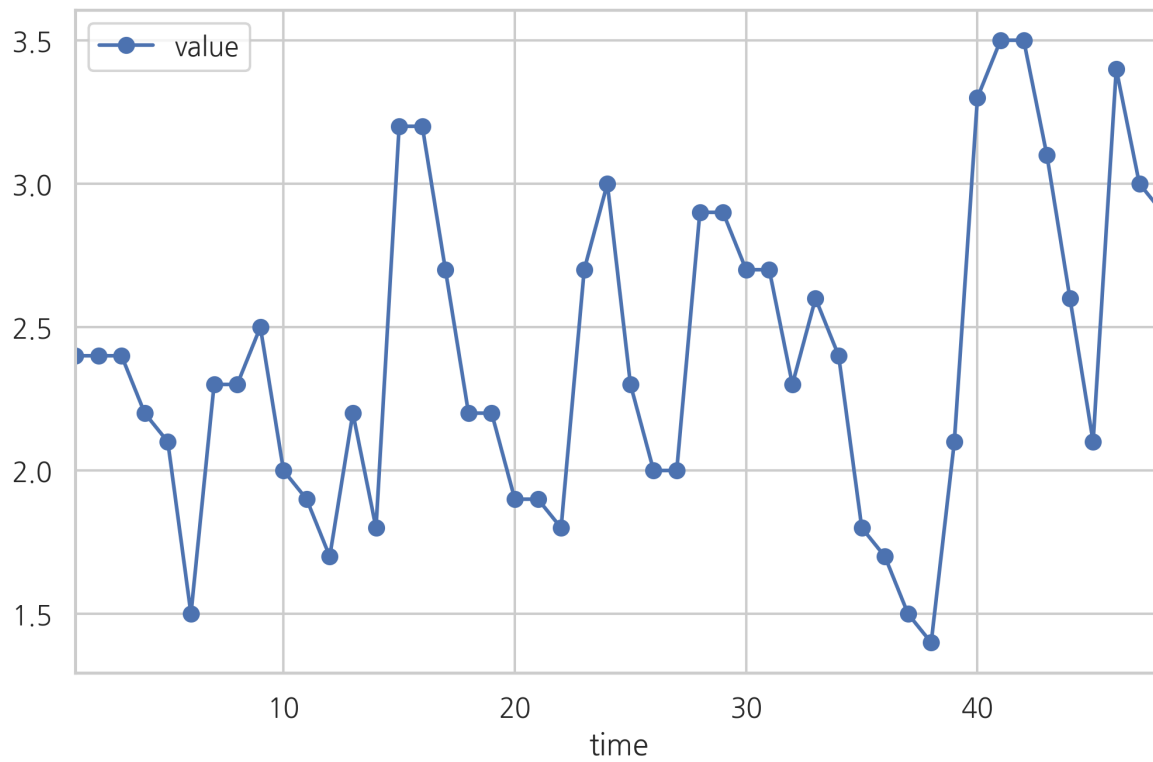
```
def yearfraction2datetime(yearfraction, startyear=0):  
    import datetime  
    import dateutil  
    year = int(yearfraction) + startyear  
    month = int(round(12 * (yearfraction - year)))  
    delta = dateutil.relativedelta.relativedelta(months=month)  
    date = datetime.datetime(year, 1, 1) + delta  
    return date
```

황체형성 호르몬 수치 시계열

다음 자료는 황체형성 호르몬 수치 시계열에 대한 시계열 자료이다.

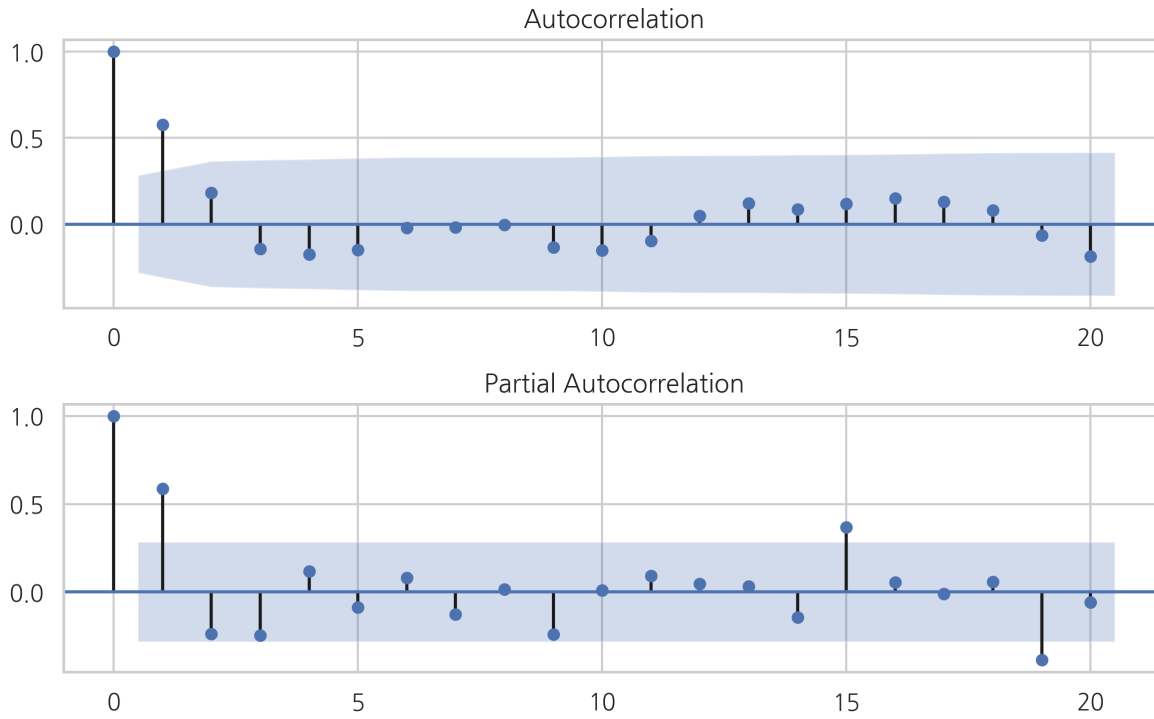
In [2]:

```
data = sm.datasets.get_rdataset("lh")  
df = data.data  
df.plot(x="time", y="value", style=["o-"])  
plt.show()
```



In [3]:

```
ax1 = plt.subplot(211)
sm.graphics.tsa.plot_acf(df.value, lags=20, ax=ax1)
ax2 = plt.subplot(212)
sm.graphics.tsa.plot_pacf(df.value, lags=20, ax=ax2)
plt.tight_layout()
plt.show()
```



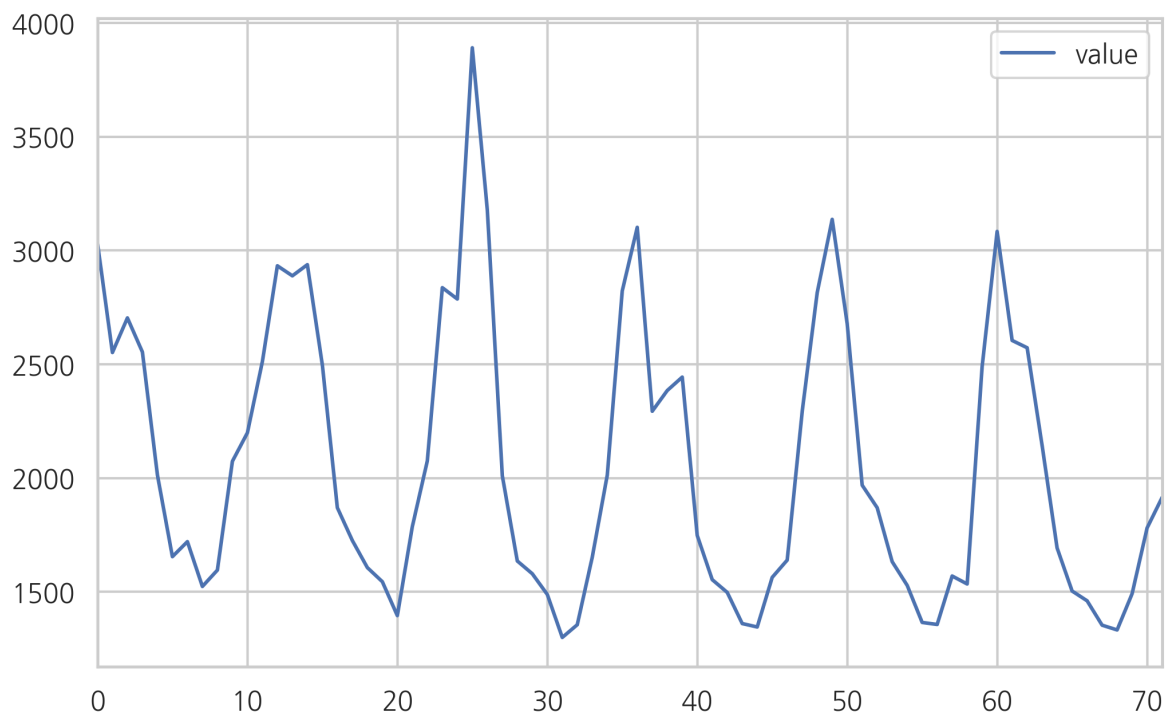
ACF가 지수함수적으로 감소하는 주기적 파형을 보이고 PACF가 2차항까지 유의한 값을 가지므로 AR(2) 모형으로 볼 수 있다.

호흡기질환 사망자수

다음 시계열 자료는 호흡기질환 사망자수에 대한 연구 자료이다.

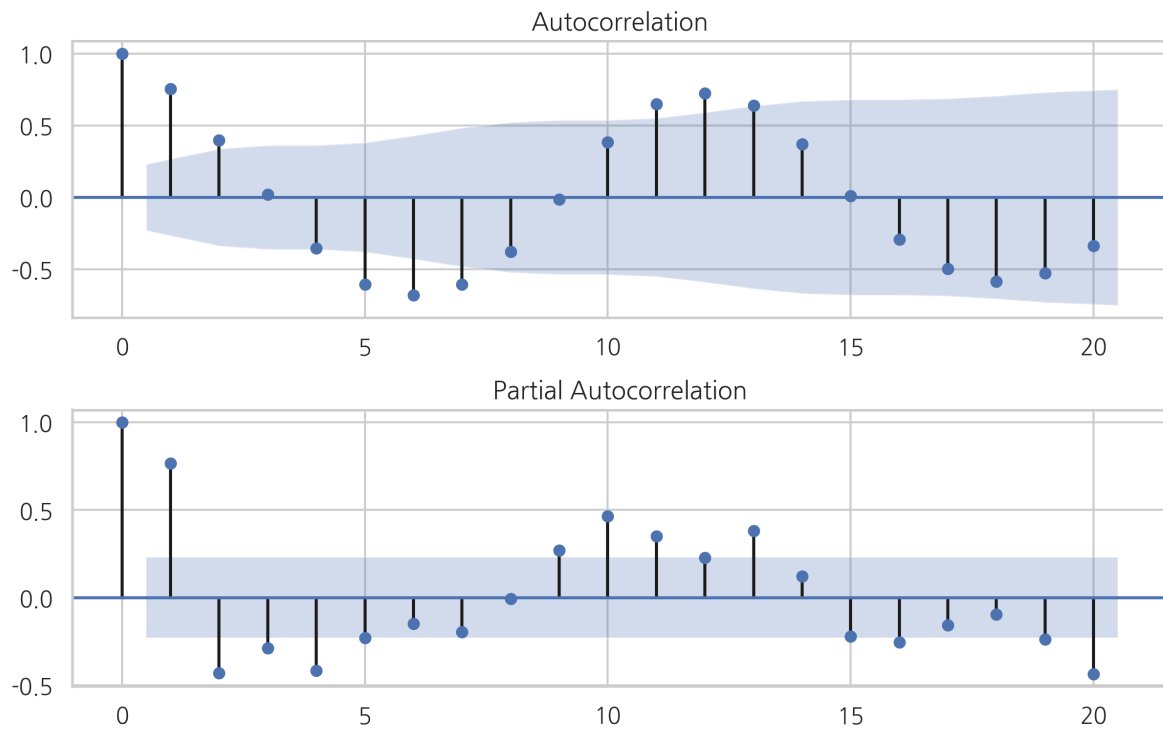
In [4]:

```
data = sm.datasets.get_rdataset("deaths", "MASS")  
df = data.data  
df.plot(y="value")  
plt.show()
```



In [5]:

```
ax1 = plt.subplot(211)
sm.graphics.tsa.plot_acf(df.value, lags=20, ax=ax1)
ax2 = plt.subplot(212)
sm.graphics.tsa.plot_pacf(df.value, lags=20, ax=ax2)
plt.tight_layout(pad=0.4, w_pad=0.5, h_pad=1.0)
plt.show()
```

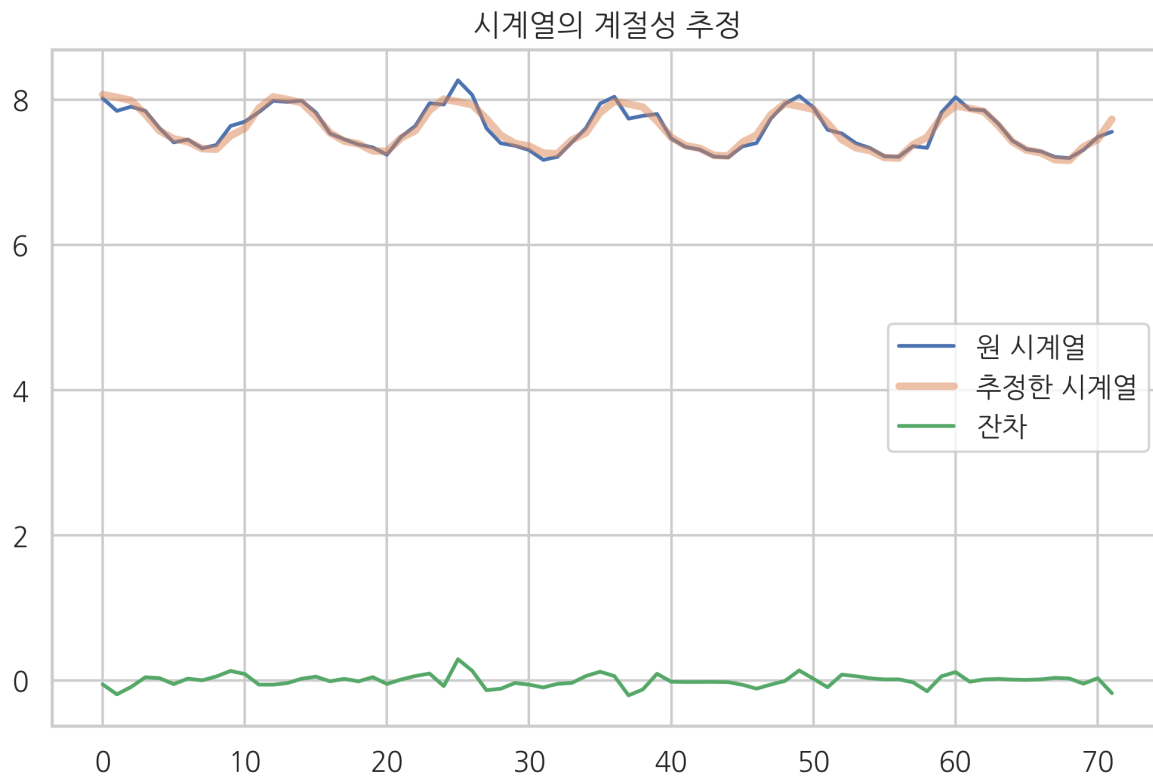


In [6]:

```
df["datetime"] = df.time.map(yearfraction2datetime)
df["month"] = df.datetime.dt.month
df["logvalue"] = np.log(df.value)

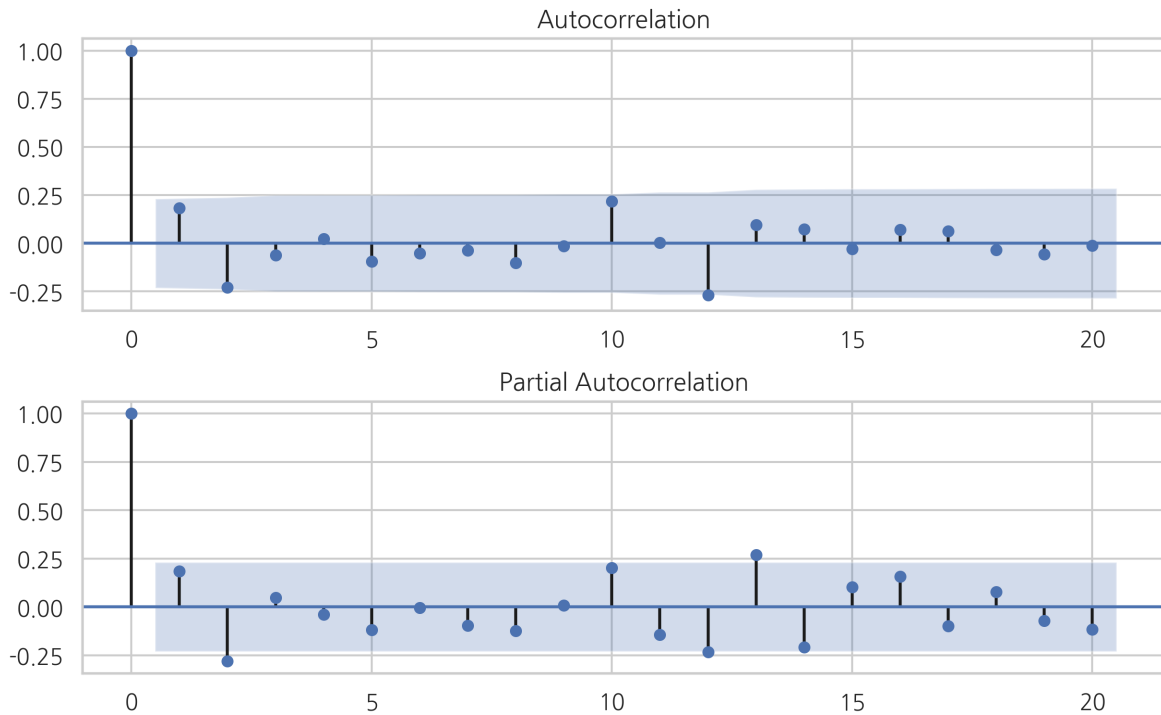
result = sm.OLS.from_formula('logvalue ~ C(month) + time + 0', data=df).fit()
y = result.resid

plt.plot(df.logvalue, label="원 시계열")
plt.plot(result.fittedvalues, lw=3, alpha=0.5, label="추정한 시계열")
plt.plot(y, label="잔차")
plt.title("시계열의 계절성 추정")
plt.legend()
plt.show()
```



In [7]:

```
ax1 = plt.subplot(211)
sm.graphics.tsa.plot_acf(y, lags=20, ax=ax1)
ax2 = plt.subplot(212)
sm.graphics.tsa.plot_pacf(y, lags=20, ax=ax2)
plt.tight_layout(pad=0.4, w_pad=0.5, h_pad=1.0)
plt.show()
```

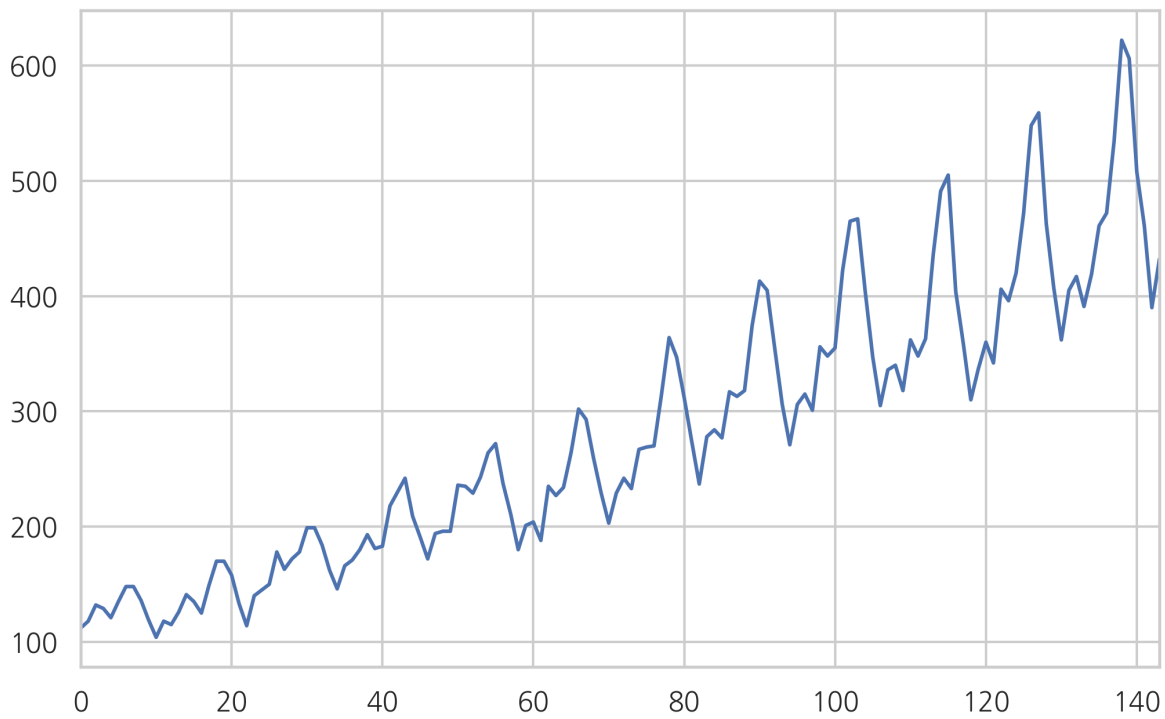


항공운송량

다음 시계열 자료는 항공운송량을 나타낸 것이다.

In [8]:

```
data = sm.datasets.get_rdataset("AirPassengers")  
df = data.data  
df["value"].plot()  
plt.show()
```



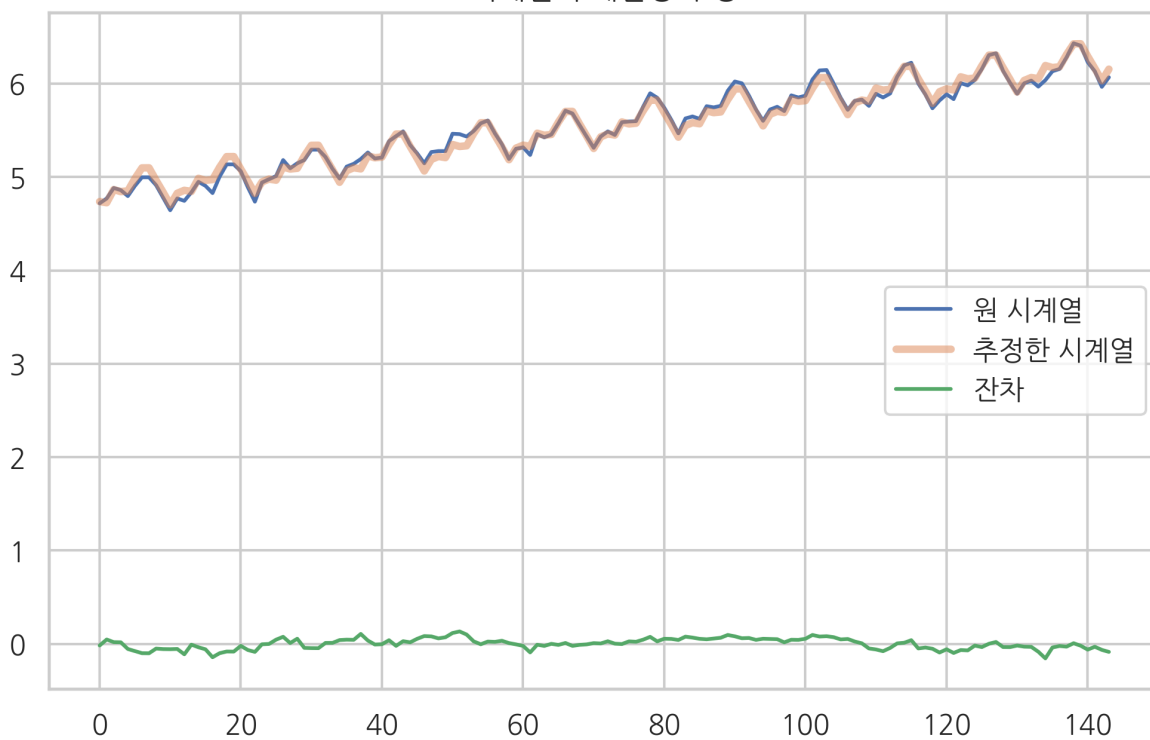
In [9]:

```
df["datetime"] = df.time.map(yearfraction2datetime)
df["month"] = df.datetime.dt.month
df["logvalue"] = np.log(df.value)

result = sm.OLS.from_formula('logvalue ~ C(month) + time + 0', data=df).fit()
y = result.resid

plt.plot(df.logvalue, label="원 시계열")
plt.plot(result.fittedvalues, lw=3, alpha=0.5, label="추정한 시계열")
plt.plot(y, label="잔차")
plt.title("시계열의 계절성 추정")
plt.legend()
plt.show()
```

시계열의 계절성 추정



In [10]:

```
ax1 = plt.subplot(211)
sm.graphics.tsa.plot_acf(y, lags=20, ax=ax1)
ax2 = plt.subplot(212)
sm.graphics.tsa.plot_pacf(y, lags=20, ax=ax2)
plt.tight_layout(pad=0.4, w_pad=0.5, h_pad=1.0)
plt.show()
```

