Forecasting

- suppose we observe $z_1, \dots, z_n$.

(목표) predict the future value z_{n+m}, where $m = 1, 2, 3, \dots$

(가정) (1) $\{Z_t\}$ is staionary and (2) the model parameters are known.

- (prob 3.26.) $z_1, \dots, z_n$ be a sample of size n from a causal AR(1) process, $Z_t = \phi Z_{t-1} + \epsilon_t$. Let $\hat{\phi}$ be the Yule-Walker estimator of ϕ.

(1) $\hat{\phi} - \phi = O_p(n^{-1/2})$.

(2) $z_{n+1}^n - \hat{z}_{n+1}^n = O_p(n^{-1/2})$ where

- z_{n+1}^n: the one-step-ahead forecast of z_{n+1} (1) given the data $z_1, \dots, z_n$ (2) based on the known parameter, ϕ.
- z_{n+1}^n: the one-step-ahead forecast of z_{n+1} (1) given the data $z_1, \dots, z_n$ (2) based on the estimated parameter $\hat{\phi}$.

- 결국 true-parameter를 알고 있다고 가정해도 무방하다. (모르면 예측해서 끼워넣으면 된다. 그렇게 끼워넣어도 $O_p(n^{-1/2})$ 만큼만 차이날테니까)

- z_{n+m}을 $\hat{z}_{n+m}$으로 예측한다고 하자. 어쨋든 우리는 $z_1, \dots, z_n$ 과 known-paramer를 기반으로 z_{n+m}을 예측할 것이므로 (어차피 알 수 있는 정보가 이게 다잖음?) $\hat{z}_{n+m}$을 아래와 같이 표현할 수 있다.

$$\hat{z}_{n+m} = g(z_1, z_2, \dots, z_n; \phi_1, \dots, \phi_p; \theta_1, \dots, \theta_q)$$

편의상 아래와 같이 쓰자.

$$\hat{z}_{n+m} = g(\boldsymbol{z}, \boldsymbol{\phi}, \boldsymbol{\theta})$$

교재에 따라서 간단하게 아래와 같이 쓰기도 한다.

$$\hat{z}_{n+m} = g(\boldsymbol{z})$$

혹은 n개의 자료를 관측했다는 점을 강조하기 위하여 아래와 같이 쓰기도 한다.

$$\hat{z}_{n+m}^n = g(\boldsymbol{z})$$

- 어떻게 예측해야 잘했다고 할까? 즉 $\hat{z}_{n+1}$을 어떻게 구하면 잘했다고 할까? $g(\boldsymbol{z})$가 아래를 최소화 하도록 하면 되지 않을까?

$$E\big(x_{n+m} - g(\boldsymbol{z})\big)^2$$

참고로 위의식을 *mean square error*라고 한다.

- 이러한 $g(\boldsymbol{z})$는 아래와 같이 구할 수 있음이 알려져있다.

$$\hat{z}_{n+m} = g(\boldsymbol{z}) = E(z_{n+m}|\boldsymbol{z})$$

결국 이는 z_{n+m}을 $E(z_{n+m}|\boldsymbol{z})$와 같이 예측한다는 의미이다. 이때 $E(z_{n+m}|\boldsymbol{z})$를 *minimum mean square error predictor* 라고 한다.

- $g(\boldsymbol{z})$의 형태를 아래와 같이 제한하자.

$$g(\boldsymbol{z}) = \alpha_0 + \sum_{k=1}^{n} \alpha_k z_k$$

즉 $g(\boldsymbol{z})$를 관측치들의 선형결합정도로만 표현하자. ARMA(p,q)모델 같은 경우는 이정도 예측자(*predictor*)로도 충분하다. 이러한 예측자를 선형예측자(*linear predictor*)라고 한다. 많은 선형예측자중에서 아래식을 가장 작게 만드는 예측자를 최적선형예측자(*best linear predictor*)라고 부르고 줄여서 BLP라고 한다.

$$E\big(x_{n+m} - g(\boldsymbol{z})\big)^2$$

(정리) 자료를 $z_1, \dots, z_n$을 관찰한 상황을 가정하자. m시점뒤의 값 z_{n+m}에 대한 최적선형예측자를 구하기 위해서는 아래식을 연립하여 풀면 된다.

$$E\big((z_{n+m} - \hat{z}_{n+m})z_1\big) = 0$$
$$E\big((z_{n+m} - \hat{z}_{n+m})z_2\big) = 0$$
$$\dots$$
$$E\big((z_{n+m} - \hat{z}_{n+m})z_n\big) = 0$$

note: 여기에서 m시점뒤의 값 z_{n+m}의 최적선형예측자는 선형예측자중에서 z_{n+m}의 값을 가장 잘 맞추는 예측자라는 뜻이다. (이때 잘 맞추는다는 것은 mean square error를 최소화 한다는 의미임)

note: m시점뒤의 값 z_{n+m}의 최적선형예측자는 아래와 같은 모양을 가져야만 한다. (왜냐하면 이것도 결국 선형이기 때문)

$$\hat{z}_{n+m} = \alpha_0 + \sum_{k=1}^{n} \alpha_k z_k$$

note: 위에서 $\hat{z}_{n+m}$은 다양하게 표현할 수 있다. 최적선형예측자라는 의미를 강조하기 위해서 아래와 같이 표현하기도 하고

$$\hat{z}_{n+m} = z_{n+m}^{BLP}$$

이것도 결국 $z_1, \dots, z_n$의 함수라는 점을 강조하기 위하여 아래와 같이 표현하기도 한다.

$$\hat{z}_{n+m} = g(z_1, \dots, g_n) = g(\boldsymbol{z})$$

또한 $z_1, \dots, z_n$의 선형결합이라는 점을 강조하기 위하여 아래와 같이 표현하기도 한다.

$$\hat{z}_{n+m} = \alpha_0 + \sum_{k=1}^{n} \alpha_k z_k$$

- 위의 연립방정식을 *prediction equation* 이라고 부른다. 그리고 이 연립방정식으로 통하여 $\alpha_0, \alpha_1, \dots, \alpha_n$을 구할 수 있다.

- 참고로

$$\hat{z}_{n+m} = \alpha_0 + \sum_{k=1}^{n} \alpha_k z_k$$

은 아래와 같이 표현할 수 있다.

$$(\hat{z}_{n+m} - \mu) = \sum_{k=1}^{n} \alpha_k (z_k - \mu)$$

note: 이게 성립하는 이유는

$$\hat{z}_{n+m} = \alpha_0 + \sum_{k=1}^{n} \alpha_k z_k$$

의 양변에 평균을 취함으로써 쉽게 알 수 있다.

note: 위에서 μ라고 당당하게 쓸 수 있는 이유는 true-model을 알고 있기 때문이다.

- 1시점뒤를 예측하여보자. 1시점뒤의 최적선형예측자 $\hat{z}_{n+1}$은 아래와 같이 쓸 수 있다.

$$\hat{z}_{n+1} = \alpha_n z_n + \dots \alpha_1 z_1$$

$\alpha_1 = \phi_{n1}, \dots, \alpha_n = \phi_{nn}$이라고 두면

$$\hat{z}_{n+1} = \phi_{n1} z_n + \dots \phi_{nn} z_1$$

와 같이 쓸 수 있다.