Workspace for 'SM_Lecture2_3'

Page 1 (row 2, column 1)

$$\varepsilon_t = \alpha \varepsilon_{t-1} + \eta_t$$
$$\varepsilon_{t-1} = \alpha \varepsilon_{t-2} + \eta_{t-1}$$
$$\vdots$$
$$\varepsilon_{t-n} = \alpha \varepsilon_{t-n-1} + \eta_{t-n}$$

$$\therefore \quad \varepsilon_t = \sum_{j=0}^{\infty} \alpha^j \eta_{t-j}$$

$$\Rightarrow \quad E(\varepsilon_t) = \sum_{j=0}^{\infty} \alpha^j \cdot E(\eta_{t-j}) \quad , \text{ where } \eta_t \sim N(0, \sigma^2)$$
$$= 0$$

$$Var(\varepsilon_t) = \sum_{j=0}^{\infty} Var(\alpha^j \eta_{t-j})$$
$$= \sum_{j=0}^{\infty} \alpha^{2j} \cdot Var(\eta_{t-j})$$
$$= \frac{\sigma^2}{1-\alpha^2}$$

$\Rightarrow$ Stationary Conditions $\quad \varepsilon_t \sim N\left(0, \frac{\sigma^2}{1-\alpha^2}\right)$

$j=1$ case

$$cov(\varepsilon_t, \varepsilon_{t-j}) = E(\varepsilon_t \cdot \varepsilon_{t-1}) - E(\varepsilon_t)E(\varepsilon_{t-1})$$
$$= E(\varepsilon_t \cdot \varepsilon_{t-1})$$
$$= E\left[(\eta_t + \alpha\eta_{t-1} + \alpha^2\eta_{t-2} + \cdots)(\eta_{t-1} + \alpha\eta_{t-2} + \alpha^2\eta_{t-3} + \cdots)\right]$$
$$= E\left\{\left[\eta_t + \alpha(\underline{\eta_{t-1} + \alpha\eta_{t-2} + \cdots})\right]\left[\underline{\eta_{t-1} + \alpha\eta_{t-2} + \alpha^2\eta_{t-3} + \cdots}\right]\right\}$$
$$= E\left\{\eta_t\left[\eta_{t-1} + \alpha\eta_{t-2} + \alpha^2\eta_{t-3} + \cdots\right] + \alpha(\eta_{t-1} + \alpha\eta_{t-2} + \cdots)^2\right\} \quad , \text{ where } cov(\eta_t, \eta_{t-j}) = E(\eta_t \cdot \eta_{t-j}) = 0$$
$$= \alpha \cdot E\left[\varepsilon_{t-1}^2\right]$$
$$= \alpha \cdot Var(\varepsilon_{t-1})$$
$$= \alpha\left(\frac{\sigma^2}{1-\alpha^2}\right)$$

$$cov(\varepsilon_t, \varepsilon_{t+h}) = E(\varepsilon_t \cdot \varepsilon_{t-h}) - E(\varepsilon_t) \cdot E(\varepsilon_{t-h}) \quad , \text{ since } E(\varepsilon_i) = 0$$
$$= E(\varepsilon_t \cdot \varepsilon_{t-h})$$
$$= E\left\{\left[\sum_{j=0}^{\infty} \alpha^j \eta_{t-j}\right]\left[\sum_{j=0}^{\infty} \alpha^j \eta_{t-h-j}\right]\right\}$$
$$= E\left\{\left[\eta_t + \alpha\eta_{t-1} + \alpha^2\eta_{t-2} + \cdots + \alpha^n\eta_{t-n} + \cdots\right]\left[\eta_{t-h} + \alpha\eta_{t-h-1} + \alpha^2\eta_{t-h-2} + \cdots + \alpha^n\eta_{t-h-n} + \cdots\right]\right\}$$
$$= E\left\{\left[\eta_t + \alpha\eta_{t-1} + \cdots + \alpha^h(\eta_{t-h} + \alpha\eta_{t-h-1} + \alpha^2\eta_{t-h-2} + \cdots)\right]\left[\eta_{t-h} + \alpha\eta_{t-h-1} + \alpha^2\eta_{t-h-2} + \cdots\right]\right\}$$
$$= E\left\{\left[\eta_t + \alpha\eta_{t-1} + \cdots + \alpha^{h-1}\eta_{t-h+1}\right]\left[\eta_{t-h} + \alpha\eta_{t-h-1} + \alpha^2\eta_{t-h-2} + \cdots\right] + \alpha^h(\eta_{t-h} + \alpha\eta_{t-h-1} + \alpha^2\eta_{t-h-2} + \cdots)^2\right\}$$
$$= \alpha^h \cdot E\left[\varepsilon_{t-h}^2\right]$$
$$= \alpha^h \cdot \left(\frac{\sigma^2}{1-\alpha^2}\right)$$

$$p(X_1, X_2, \ldots, X_n) = p(X_n \mid X_1, \ldots, X_{n-1})\, p(X_{n-1} \mid X_1, \ldots, X_{n-2}) \cdots p(X_2 \mid X_1)\, p(X_1)$$
$$= \frac{p(X_1, \ldots, X_{n-1}, X_n)}{\cancel{p(X_1, \ldots, X_{n-1})}} \cdot \frac{\cancel{p(X_1, \ldots, X_{n-2}, X_{n-1})}}{\cancel{p(X_1, \ldots, X_{n-2})}} \cdots \frac{\cancel{p(X_1, X_2)}}{\cancel{p(X_1)}} \cdot \cancel{p(X_1)}$$
$$= p(X_1, \ldots, X_{n-1}, X_n)$$

Workspace for 'SM_Lecture2_3'
Page 2 (row 2, column 2)

Maximum Likelihood Estimation of AR(1) Error Model

Maximum Likelihood Estimation of SAR

2. Statistical Modelling (3)

Statistical Modelling & Machine Learning

Jaejik Kim

Department of Statistics
Sungkyunkwan University

STA3036

Independence Assumption이 위배되었을 때 : error, 즉 ε_i's 들이 correlated 되어 있을 때

수업에서 다룰 것들 : Time-Series , Spatial - Dependencies

Data with Time Dependency

- Data: $(y_t, \boldsymbol{x}_t),\ t = 1, \ldots, T$.
  - $(y_t, \boldsymbol{x}_t)$ are measured for the same object at discrete time points (e.g., hourly, weekly, monthly, yearly data).

- Model: $Y_t = \overbrace{f(\boldsymbol{X}_t; \boldsymbol{\theta})}^{\text{평균을 나타내기 위해선 어떤 모형을 써도 상관없다.}} + \epsilon_t,\ t = 1, \ldots, T)$.
  - $\epsilon_t,\ t = 1, \ldots, T$ have constant variance. - 이렇게 가정한다
  - ϵ_t's are correlated (time dependency) $\Rightarrow$ Y_t's are correlated.
  - ϵ_t's have a stationary process (i.e., covariance between ϵ_t's depends only on time difference).
  - ARMA (p,q) time series modelling for ϵ_t.

$$\epsilon_t = \overbrace{\sum_{j=1}^{p} \alpha_j \epsilon_{t-j}}^{AR} + \overbrace{\sum_{j=1}^{q} \phi_j \eta_{t-j}}^{MA} + \eta_t,$$

where $\eta_t \overset{iid}{\sim} N(0, \sigma^2)$.

$Y_t = f(X_t; \theta) + \varepsilon_t$ 에서 $f(X_t; \theta)$ 는 fixed component 이고, ε_t 만 random component 이기 때문에 ε_t 가 Correlated 된다면 Y_t 역시 Correlated 된다.

Regression Model with AR(1) Error

≈ Markov Process : 현 시점의 값은 바로 전 시점의 값에만 의존한다. (Auto-Regressive with lag-1)

- Regression Model with AR(1) Error:

$$
\begin{aligned}
Y_t &= f(\boldsymbol{X}_t; \boldsymbol{\theta}) + \epsilon_t, \\
\epsilon_t &= \alpha \epsilon_{t-1} + \eta_t, \quad \rightarrow \text{AR(1)}
\end{aligned}
$$

  where α is an autocorrelation parameter satisfying $|\alpha| < 1$ (stationary condition), and $\underbrace{\eta_t \overset{iid}{\sim} N(0, \sigma^2)}_{\text{White Noise}}$.

  ϵ_t 가 발산하지 않고, covariance 가 lag 에만 의존한다는 것을 만족하기 위해서는 $|\alpha| < 1$ 이 꼭 필요하다 (unit-root condition)

- AR(1) error: From the AR(1) model and recursive calculations, we obtain

$$\epsilon_t = \sum_{j=0}^{\infty} \alpha^j \eta_{t-j}.$$

☆ Likelihood Function을 만들기 위해서는 Y_t 들의 Variance-Covariance matrix를 알아야 한다.
그러기 위해선 Y_t 들의 joint 분포를 명확히 아는게 중요하다. 이 분포를 알기 위해선 ϵ_t 들의 Covariance Matrix가 어떻게 되는지 알아야 한다.

Properties of AR(1) Error

- Since $\epsilon_t = \sum_{j=0}^{\infty} \alpha^j \eta_{t-j}$ and $E(\eta_t) = 0$ for all t,

$$E(\epsilon_t) = 0.$$

- Since η_t's are independent and $Var(\eta_t) = \sigma^2$ for all t,

$$Var(\epsilon_t) = \frac{\sigma^2}{1-\alpha^2}.$$

- Covariance of ϵ_t and ϵ_{t-j}:

$$Cov(\epsilon_t, \epsilon_{t-j}) = \alpha^j \left(\frac{\sigma^2}{1-\alpha^2} \right), \; j \neq 0.$$

Maximum Likelihood Estimation of AR(1) Error Model

Method 1: Likelihood function from multivariate normal density.

- $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_T)^\top \sim MVN(\mathbf{0}, \boldsymbol{\Sigma})$, where

$$\boldsymbol{\Sigma} = \frac{\sigma^2}{1-\alpha^2}\begin{pmatrix} 1 & \alpha & \cdots & \alpha^{T-1} \\ \alpha & 1 & \cdots & \alpha^{T-2} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha^{T-1} & \alpha^{T-2} & \cdots & 1 \end{pmatrix}.$$

- $\boldsymbol{y} = (y_1, \ldots, y_T)^\top \sim MVN(\underbrace{\boldsymbol{f}, \boldsymbol{\Sigma}}_{\text{assumed means and variances}})$.
- Log-likelihood function:

$$l(\boldsymbol{\theta}; \boldsymbol{y}, \alpha, \sigma^2) = -\frac{1}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}(\boldsymbol{y}-\boldsymbol{f})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\boldsymbol{f}).$$

Maximum Likelihood Estimation of AR(1) Error Model

Method 2: Likelihood function from conditional density.

- Relationship between joint density and conditional densities:

$$p(x_1, x_2, \ldots, x_n) = p(x_n|x_1, \ldots, x_{n-1})p(x_{n-1}|x_1, \ldots, x_{n-2}) \cdots p(x_2|x_1)p(x_1).$$

걘 얘가 원래는 맨 데, AR(1) structure의 Markovian property에 의해서 얘로 바꿔쓸 수 있음

- AR(1) structure:
  - Current status depends only on the previous status (Markovian property).
  - i.e., X_t depends only on X_{t-1}
    $\Rightarrow X_t$ is independent of $X_{t-2}, X_{t-3}, \ldots, X_1$.

$\sim p(X_{n-1} | X_{n-2})$

$$\begin{aligned} p(x_1, x_2, \ldots, x_n) &= p(x_n|x_{n-1})p(x_{n-1}|x_1, \ldots, x_{n-2}) \cdots p(x_2|x_1)p(x_1) \\ &= \left[\prod_{t=2}^{n} p(x_t|x_{t-1})\right] p(x_1). \end{aligned}$$

얘까지 random variable로 두면 계산이 너무 복잡해짐.
=> X_1도 given 이라고 생각하자.

Maximum Likelihood Estimation of AR(1) Error Model

- AR(1) error: $\epsilon_t = \alpha\epsilon_{t-1} + \eta_t$, $\eta_t \overset{iid}{\sim} N(0, \sigma^2)$.
- $\epsilon_t | \epsilon_{t-1} \sim N(\alpha\epsilon_{t-1}, \sigma^2)$.
- Since $E(\epsilon_t) = 0$ and $Var(\epsilon_t) = \frac{\sigma^2}{1-\alpha^2}$, $\epsilon_1 \sim N\left(0, \frac{\sigma^2}{1-\alpha^2}\right)$.
- $Y_t | Y_{t-1} \sim N(f(\boldsymbol{X}_t; \boldsymbol{\theta}) + \alpha\epsilon_{t-1}, \sigma^2)$.
- $Y_1 \sim N\left(f(\boldsymbol{X}_1; \boldsymbol{\theta}), \frac{\sigma^2}{1-\alpha^2}\right)$.

$X_t | X_{t-1}$만 random variable로 둘 경우 등분산을 갖게 된다.

$\Rightarrow L(\beta) \approx \prod_{i=1}^{n} p(Y_t | Y_{t-1})$

$= \prod_{t=2}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(y_t - X_t^T\beta - \alpha\varepsilon_{t-1})^2}{2\sigma^2}\right]$

$\Rightarrow \min_{\beta,\alpha} \sum (y_t - X_t^T\beta - \alpha\varepsilon_{t-1})^2$ (Typical Form of LSE)

- Log-likelihood function:

$$\cancel{l(\boldsymbol{\theta}; \boldsymbol{y}, \alpha, \sigma^2)} = \sum_{t=2}^{T} \log p(Y_t | Y_{t-1}) + \log p(Y_1).$$

$\ell(\theta, \alpha, \sigma^2, Y)$

$\log p(Y_1 | Y_0)$은 없으니까

Maximum Likelihood Estimation of AR(1) Error Model

Estimation Algorithm:

1. Set the initial parameter vectors $\hat{\boldsymbol{\theta}}$.
2. Compute residuals $r_t = y_t - f(\boldsymbol{x}_t; \hat{\boldsymbol{\theta}}),\ t = 1, \ldots, T$.
3. Estimate the AR(1) model parameters α and σ^2 using the residuals $r_1, \ldots, r_T$.
4. Construct $\boldsymbol{\Sigma}$ using $\hat{\alpha}$ and $\hat{\sigma}^2$ obtained from Step 3.
5. Find $\hat{\boldsymbol{\theta}}$ minimizing $(\boldsymbol{y} - \boldsymbol{f})^\top \underline{\boldsymbol{\Sigma}^{-1}} (\boldsymbol{y} - \boldsymbol{f})$.

   AR(1)와 같은 모델을 사용하지 않을 경우 Σ를 추정하기엔 너무나 많은 parameter들이 있다.
6. Repeat Steps 2–5 until $\boldsymbol{\theta}$ is converged.

3주차 1차시 끝

Data with Spatial Correlations

2공간 사이에 거리에 따라 Correlation이 생긴다. (거리 ↑ Correlation ↓)

Time-Series 와 Spatial Correlation 끼리의 공통점 ; 특정한 시간점 또는 공간점에서의 데이터는 한개씩 밖에 존재하지 않는다.

- Data are observed at spatial points in 2 or 3 dimensional space (e.g., house price in a city, house income in a city, the number of infectious persons in an area, etc.)
- There exist correlations between spatial points. EX) 상권, 주택가
- Basically, as distance between two spatial points increases, the correlation decreases.
- There are various approaches for spatial prediction problems (spatial autoregressive model, spatial error model, kriging, etc.)

  ≈ smoothing

(Data)
S개의 공간점

- Data: $(\underbrace{y_s, \boldsymbol{x}_s}_{\text{Spatial}}),\ s = 1, \ldots, S.$
- Spatial autoregressive model: or Spatial lag Model

$$\underbrace{\boldsymbol{y} = \rho \boldsymbol{W} \boldsymbol{y}}_{} + \overbrace{\boldsymbol{X}\beta}^{\text{OLS part}} + \epsilon.$$

왜 autoregressive 인가? 자기를 예측하는데 자신이 쓰인다.

$y_i = \rho W_i y + \varepsilon$, $W = \begin{bmatrix} W_1 \\ W_2 \\ \vdots \\ W_S \end{bmatrix}$

$\Rightarrow W_i y = W_{ij} \cdot y_j$

$\Rightarrow y_i = \rho W_{ij} \cdot y_j + \varepsilon$

Spatial Weight Matrix : Spatial Dependencies를 - 공간점 사이의 잡는 역할 ☆

- $\boldsymbol{y}$: $S \times 1$ output variable vector.
- ρ: Spatial autocorrelation parameter. scalar AR에서 α같은 존재랄까...
- $\boldsymbol{W}$: $S \times S$ weight matrix that accounts for the spatial dependencies among spatial units. W_{ij} : i번째 공간점을 j번째 공간점으로 예측할 때의 j번째 공간점의 weight (공간적 영향력)
  ☆ W는 Symmetric Matrix가 아니다.
- $\boldsymbol{X}$: $S \times p$ input matrix.
- β: Coefficient vector of $\boldsymbol{X}$.
- $\epsilon \sim MVN(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$.

Spatial Weight Matrix

W는 추정해야하는 대상이 아니라 지정해줘야하는 대상이다.

- Spatial weight matrix $\boldsymbol{W} = (w_{ij};\ i, j = 1, \ldots, S)$:
  - w_{ij}: Spatial influence of unit j on unit i.
  - $w_{ii} = 0$ (i.e., all diagonal elements of $\boldsymbol{W}$ are 0).

  자기자신에게 영향을 주어서도 안되고 줄 수도 없으니까

- Construction of $\boldsymbol{W}$:

  2공간의 거리로 구하는 method

  1. Weights based on distance:
     - k-Nearest neighbor weights.
     - Radial distance weights.
     - Power distance weights.
     - Exponential distance weights.
     - Double power distance weights.

  구, 동, ...

  2공간의 경계를 통해 구하는 method

  2. Weights based on boundaries:
     - Spatial contiguity weights.
     - Shared-boundary weights.

  거리, 경계 모두 사용하는 method

  3. Combined distance-boundary weights.

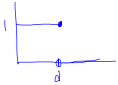
Construction of $\boldsymbol{W}$

Weights based on distance (1):

- k-Nearest neighbor weights: W_{ij} : i를 fix시키고 모든 j의 weight를 구한다
  - i번째 unit에 가장 가까운 N개의 units
  - d_{ij}: Distance between unit i and unit j. $j = 1, \ldots, S,\ i \neq j$
  - $N_k(i)$: A set containing the k closet units to unit i based on $d_{ij},\ j = 1, \ldots, S,\ i \neq j$.
  - j번째 Unit이 i번째 Unit에 대해서 가장 가까운 k개 중 하나일 때
  - If $j \in N_k(i)$, then $w_{ij} = 1$. Otherwise, $w_{ij} = 0$. $\Rightarrow W_{ij} \in \{0, 1\}$
  - For the symmetric matrix of $\boldsymbol{W}$, if $j \in N_k(i)$ or $i \in N_k(j)$, then $w_{ij} = 1$. Otherwise, $w_{ij} = 0$.

=> 단점 : 공간점이 균일하지 않다면 예측력이 떨어진다. → 이러한 단점을 보완

- Radial distance weights:
  - Tuning Parameter
  - d: Threshold distance.
  - If d_{ij} is larger than d, units i and j have no spatial influence.
  - No diminishing effect of spatial influence up to d.
  - If $d_{ij} \leq d$, then $w_{ij} = 1$. Otherwise, $w_{ij} = 0$.

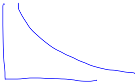


Construction of W

☆ 앞선 방법들의 문제는 균일한 weight를 준다는데 있다. → 이러한 단점을 보완한 method들

Weights based on distance (2):

- Power distance weights:
  - It considers diminishing effect of spatial influence.
  - $w_{ij} = d_{ij}^{-\alpha}$. ? d_{ij}는 어떻게 정하는데
  - $\alpha > 0$. Typical choice of α is 1 or 2.

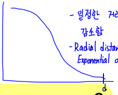


- Exponential distance weights:
  - Diminishing effect of spatial influence.
  - $w_{ij} = \exp(-\alpha d_{ij})$.

    Tuning Parameter

    직접 정해줘야 함
  - $\alpha > 0$.

- Double-power distance weights:
  - Bell-shaped function & threshold distance d.
  - If $d_{ij} \leq d$, then $w_{ij} = \left[1 - (d_{ij}/d)^k\right]^k$. Otherwise $w_{ij} = 0$.
  - Typical choice of k is 2, 3, or 4.



- 일정한 거리 이상에서 급격히 weight가 감소함
- Radial distance, Power distance, Exponential distance의 특징을 섞음
- d와 k를 정해줘야 한다
- $d_{ij} > d \Rightarrow W_{ij} = 0$

Construction of W

Weights based on boundaries: The boundaries shared between spatial units play in important role in determining degree of spatial influence. 경계선들이 맞닿아 있는가 아닌가로 Weight가 주어짐

- Spatial Contiguity weights:
  - If units i and j share their boundary, $w_{ij} = 1$. Otherwise $w_{ij} = 0$. EX) 성동구와 종로구는 $W_{ij} = 1$이 부여됨
  - However, even if two units have a shared corner point, this weight returns 1.  경계가 하나의 점에서만 맞닿드려도 가중치가 부여되는 문제가 있다. 이러한 문제를 해결하기 위해 맞닿는 거리에 따라 가중치를 다르게 부여할수도 있다.
  - l_{ij}: Length of shared boundary.
  - If $l_{ij} > 0$, then $w_{ij} = 1$. If $l_{ij} = 0$, $w_{ij} = 0$.
    얘도 tuning parameter인가? 아닌데... 어떻게 일일히 계산하지...? Euclidean Distance

- Shared-boundary weights:
  - Proportional boundary length between unit i and j.
  - l_i: Total boundary length that unit i is shared with all other units (i.e., $\sum_{j=1,\ldots,S,\ j\neq i} l_{ij}$).
  - $w_{ij} = l_{ij}/i_i$.
    i번째 행의 맞닿는 경계선의 길이합

Construction of W

Combined distance-boundary weights:

- Spatial influence represented by both distance and boundary relations.
- Cliff and Ord (1969) proposed the weight by the combination of power distance and boundary-shares as follows:

$$w_{ij} = \frac{l_{ij} d_{ij}^{-\alpha}}{\sum_{k=1,\ldots,S,\ k \neq i} l_{ik} d_{ik}^{-\alpha}}.$$

- power function part
- boundary shares part

※ 보통 구역의 중심을 점으로 두고 구함

where $\alpha > 0$. Typical choice of α is 1.

Normalization of $\boldsymbol{W}$

- Normalization: Normalization of spatial effect for removing scale effects.

- Row normalized weights: $y_i = W_i y$ 의 형태기 때문에, 각 행들끼리 normalize 시킴
  - The sum of each row is 1 (i.e., $\sum_{j=1}^{S} w_{ij} = 1$).

문제점 : 행과 행 사이의 Weight 비교가 불가능하다.

$$w_{ij} \leftarrow \frac{w_{ij}}{\sum_{k=1,\ldots,S,k\neq i} w_{ij}}.$$

보완

- Scalar normalized weights:
  - Row normalization is not appropriate for comparison between rows.
  - Scalar normalization: $\gamma \boldsymbol{W}$, where γ is a positive scalar.
  - $\gamma = 1/\max(w_{ij}) \Rightarrow$ All normalized w_{ij} has a value between 0 and 1 (relative influence intensity). - Weights 간의 상대적 영향력을 알 수 있음

계산 / 해석의 용이성 때문에 λ_{max} 를 사용하기도 한다

  - $\gamma = 1/\lambda_{max}$, where λ_{max} is the largest eigenvalue of $\boldsymbol{W}$.

Maximum Likelihood Estimation of SAR

여기서는 : ρ와 β를 어떻게 추정할 것인가

– 사실 ρ 자체는 우리들의 관심이 아니다. ρ는 단지 W가 주어진 상태에서 W의 크기를 조정해주는 scale-factor 이다.

└ W의 구조를 임의대로 정했기 때문에, 실제로 "얼마만큼" 자기회귀, autoregressive 한지는 알 수 없기 때문.

- SAR Model: Let $\boldsymbol{A} = \boldsymbol{I} - \rho \boldsymbol{W}$

$$
\begin{aligned}
& \boldsymbol{y} = \rho \boldsymbol{W} \boldsymbol{y} + \boldsymbol{X} \underbrace{\boldsymbol{\beta}}_{\text{우리의 관심}} + \boldsymbol{\epsilon} \\
\Rightarrow \quad & (\boldsymbol{I} - \rho \boldsymbol{W}) \boldsymbol{y} = \boldsymbol{X} \boldsymbol{\beta} + \boldsymbol{\epsilon} \quad , \ \text{let } (I - \rho W) y = y^* \\
\Rightarrow \quad & \boldsymbol{\epsilon} = (\boldsymbol{I} - \rho \boldsymbol{W}) \boldsymbol{y} - \boldsymbol{X} \boldsymbol{\beta} \\
\Rightarrow \quad & \boldsymbol{\epsilon} = \boldsymbol{A} \boldsymbol{y} - \boldsymbol{X} \boldsymbol{\beta}. \qquad \varepsilon \sim MVN(0, \sigma^2 I)
\end{aligned}
$$

- Since $\boldsymbol{\epsilon} \sim MVN(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$, the pdf of $\boldsymbol{\epsilon}$ is

$$
\begin{aligned}
p(\boldsymbol{\epsilon}) &= (2\pi\sigma^2)^{-S/2} \exp\left[-\frac{1}{2\sigma^2} \underline{\boldsymbol{\epsilon}^\top \boldsymbol{\epsilon}} \right] \\
&= (2\pi\sigma^2)^{-S/2} \exp\left[-\frac{1}{2\sigma^2} \underline{(\boldsymbol{A}\boldsymbol{y} - \boldsymbol{X}\beta)^\top (\boldsymbol{A}\boldsymbol{y} - \boldsymbol{X}\beta)} \right].
\end{aligned}
$$

그러나 우리가 필요한건 y의 joint density function

Maximum Likelihood Estimation of SAR

- To construct the likelihood function of $(\rho, \boldsymbol{\beta}, \sigma^2)$, we need the pdf of $\boldsymbol{y}$.
- The pdf of $\boldsymbol{y}$ can be obtained by the transformation of the random vector $\boldsymbol{\epsilon}$ ($\because \boldsymbol{\epsilon} = \boldsymbol{A}\boldsymbol{y} - \boldsymbol{X}\beta$).
- Since $\boldsymbol{y} = \boldsymbol{A}^{-1}\boldsymbol{X}\beta + \boldsymbol{A}^{-1}\boldsymbol{\epsilon}$ is differentiable and monotone within the range of $\boldsymbol{\epsilon}$,

$$
\begin{aligned}
p(\boldsymbol{y}) &= p(\boldsymbol{\epsilon}) \left| \frac{d\boldsymbol{\epsilon}}{d\boldsymbol{y}} \right| \\
&= (2\pi\sigma^2)^{-S/2} \exp\left[-\frac{1}{2\sigma^2} (\boldsymbol{A}\boldsymbol{y} - \boldsymbol{X}\beta)^\top (\boldsymbol{A}\boldsymbol{y} - \boldsymbol{X}\beta) \right] |\boldsymbol{A}|.
\end{aligned}
$$

Maximum Likelihood Estimation of SAR

- Log-likelihood function:

$$
\begin{aligned}
l(\rho, \boldsymbol{\beta}, \sigma^2 | \boldsymbol{y}) &= -\frac{S}{2}\log(\sigma^2) + \log|\boldsymbol{A}| \\
&\quad -\frac{1}{2\sigma^2}(\boldsymbol{Ay} - \boldsymbol{X}\beta)^\top(\boldsymbol{Ay} - \boldsymbol{X}\beta).
\end{aligned}
$$

- MLE of $\boldsymbol{\beta}$ and σ^2: By solving $\frac{\partial l(\rho,\boldsymbol{\beta},\sigma^2|\boldsymbol{y})}{\partial \boldsymbol{\beta}} = 0$ and $\frac{\partial l(\rho,\boldsymbol{\beta},\sigma^2|\boldsymbol{y})}{\partial \sigma^2} = 0$, respectively,

$$
\begin{aligned}
\hat{\beta} &= (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{Ay}, &(1)\\
\hat{\sigma}^2 &= \frac{1}{S}(\boldsymbol{Ay} - \boldsymbol{X}\hat{\beta})^\top(\boldsymbol{Ay} - \boldsymbol{X}\hat{\beta}). &(2)
\end{aligned}
$$

Maximum Likelihood Estimation of SAR

- MLE of ρ: By replacing $(\boldsymbol{\beta}, \sigma^2)$ with $(\hat{\boldsymbol{\beta}}, \hat{\sigma}^2)$,

$$\max_{|\rho|<1} l(\rho|\boldsymbol{y}) = \max_{|\rho|<1} \log|\boldsymbol{A}| - \frac{S}{2}\log(\boldsymbol{A}\boldsymbol{y})^\top(\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top)^\top(\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top)(\boldsymbol{A}\boldsymbol{y}). \quad (3)$$

- ML Estimation procedure:
  1. Find $\hat{\rho}$ by solving the maximization problem (3).
  2. Compute $\boldsymbol{A} = \boldsymbol{I} - \hat{\rho}\boldsymbol{W}$.
  3. Obtain $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma^2}$ using (1) and (2), respectively.