Lecture Notes:
Topic in Econometrics

Based on lectures by **Marko Mlikota** in Spring semester, 2025

Draft updated on October 6, 2025

This is the lecture note taken in the course *Topic in Econometrics* taught by **Marko Mlikota** at Graduate Institute of International and Developoment Studies, Geneva as part of the International Economics program (Semester III, 2025). The content is partly based on the course notes provided by the professor and supplemented by many other references I read myself. The main reason is that the original notes are found a bit ambiguous and I want to further clarify.

Currently, these are just drafts of the lecture notes. There can be typos and mistakes anywhere. So, if you find anything that needs to be corrected or improved, please inform at jingle.fu@graduateinstitute.ch.

Contents

Lecture 1.

Introduction to Bayesian Econometrics

1.1 Introduction

Definition 1.1.1 (Proportional Function)**.** We say $f(x)$ is proportional to $g(x)$ if there exists a constant c, such that $f(x) = c \cdot g(x)$ for all x in the domain of interest. We denote this relationship as $f(x) \propto g(x)$.

If y is a R.V. with pdf $f(y) \propto \exp(-\lambda y)$ for $y \geq 0$ and 0 otherwise, then we know $f(y) = c \cdot \exp(-\lambda y)$. To find c, we use the fact that the total probability must equal 1:

$$\int_0^\infty f(y)dy = 1 \Rightarrow \int_0^\infty c \cdot \exp(-\lambda y)dy = 1 \tag{1.1}$$

Calculating the integral, we have:

$$c \cdot \left[-\frac{1}{\lambda} \exp(-\lambda y) \right]_0^\infty = 1 \tag{1.2}$$

Evaluating the limits, we get:

$$c = \lambda \tag{1.3}$$

Now looking at the normal distribution: $y \sim \mathcal{N}\left(\mu, \sigma^2\right)$ we have

$$p(y|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2} \right) \tag{1.4}$$

$$\propto \exp\left(-\frac{1}{2\sigma^2}(y-\mu)^2 \right) \tag{1.5}$$

For a simple linear regression model $y_i = \theta + u_i$, where $\mathbb{E}[u_i|\theta = 0]$, we know $\mathbb{E}[y_i|\theta] = \theta$.

- Least-squares estimator:

$$\hat{\theta}_{LS} = \arg\min_\theta \sum_{i=1}^n (y_i - \mathbb{E}[y_i|\theta])^2 \tag{1.6}$$

$$= \arg\min_\theta (y_i - \theta)^2 \tag{1.7}$$

$$= \frac{1}{n} \sum_{i=1}^n y_i \tag{1.8}$$

- Maximum likelihood estimator(Assuming $y_i|\theta \sim \mathcal{N}\left(\theta, \sigma^2\right)$):

$$\hat{\theta}_{ML} = \arg\max_\theta \prod_{i=1}^n p(y_i|\theta) \tag{1.9}$$

$$\propto \arg\max_\theta \prod_{i=1}^n \exp\left(-\frac{1}{2\sigma^2}(y_i - \theta)^2 \right) \tag{1.10}$$

$$\propto \arg\min_\theta \sum_{i=1}^n (y_i - \theta)^2 \tag{1.11}$$

$$= \frac{1}{n}\sum_{i=1}^{n} y_i = \hat{\theta}_{LS} \tag{1.12}$$

It's easy to see that:

$$\mathbb{E}[\hat{\theta}|\theta] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} y_i|\theta\right] = \theta \tag{1.13}$$

$$\mathbb{V}[\hat{\theta}|\theta] = \mathbb{V}\left[\frac{1}{n}\sum_{i=1}^{n} y_i|\theta\right] = \frac{\sigma^2}{n} \tag{1.14}$$

Even without assuming that $\hat{\theta}|\theta \sim \mathcal{N}\left(\theta, \frac{\sigma^2}{n}\right)$, we know by the Central Limit Theorem that:

$$\sqrt{n}\left(\hat{\theta} - \theta\right) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2\right) \Rightarrow \hat{\theta}|\theta \sim \mathcal{N}\left(\theta, \frac{\sigma^2}{n}\right) \tag{1.15}$$

Definition 1.1.2 (Posterior)**.** The posterior distribution of a parameter θ given data y is defined as:

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)} \propto p(y|\theta)p(\theta) \tag{1.16}$$

where $p(y|\theta)$ is the likelihood, $p(\theta)$ is the prior distribution of θ, and $p(y)$ is the marginal likelihood. This shows how, given a prior belief about θ and observed data y, we can update our belief to form the posterior distribution.

Example 1. Taking a simple example:

$$y_i|\theta \sim \mathcal{N}\left(\theta, 1\right) \Rightarrow p(y_i|\theta) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(y_i - \theta)^2\right) \tag{1.17}$$

Suppose $\theta \sim \mathcal{N}\left(\theta, \frac{1}{\lambda}\right)$.

$$p(\theta|y) \propto p(y|\theta)p(\theta) \tag{1.18}$$

$$= (2\pi)^{-\frac{n}{2}} \exp\left(-\frac{1}{2}(y_i - \theta)^2\right) \cdot \frac{1}{\sqrt{2\pi\frac{1}{\lambda}}} \exp\left(-\frac{1}{2\frac{1}{\lambda}}(\theta - \theta_0)^2\right) \tag{1.19}$$

$$\propto \exp\left(-\frac{1}{2}\sum_{i=1}^{n}(y_i - \theta)^2 - \frac{\lambda}{2}(\theta - \theta_0)^2\right) \tag{1.20}$$

$$\propto \exp\left(-\frac{1}{2}\left[(n+\lambda)\theta^2 - 2\left(\sum_{i=1}^{n} y_i + \lambda\theta_0\right)\theta\right]\right) \tag{1.21}$$

$$\theta|y \sim \mathcal{N}\left(\frac{1}{n+\lambda}\left(\sum_{i=1}^{n} y_i + \lambda\theta_0\right), \frac{1}{n+\lambda}\right) \tag{1.22}$$

We guess that $\theta|y \sim \mathcal{N}\left(\bar{\theta}, \bar{V}\right)$, then we can write:

$$p\left(\theta|y\right) \propto \exp\left(-\frac{1}{2}\bar{V}^{-1}(\theta - \bar{\theta})^2\right) \tag{1.23}$$

$$\propto \exp\left(-\frac{1}{2}\left[\bar{V}^{-1}\theta^2 - 2\bar{V}^{-1}\bar{\theta}\theta\right]\right) \tag{1.24}$$

then, we know that:

$$\bar{V}^{-1} = n + \lambda \tag{1.25}$$

and

$$\begin{align}
\bar{\theta} &= \frac{1}{n+\lambda}\left(\sum_{i=1}^{n} y_i + \lambda\theta_0\right) \tag{1.26}\\
&= \frac{1}{n+\lambda}\cdot\left[n\cdot\sum_{i=1}^{n} y_i + \lambda\,\theta_0\right] \tag{1.27}\\
&\to \begin{cases}\theta_0, & \text{if } \lambda \to \infty;\\ \hat{\theta}, & \text{if } \lambda \to 0 \text{ and/or } n \to \infty.\end{cases} \tag{1.28}
\end{align}$$

In general, we can push θ_0 to 0 by re-centering y_i. Then we have:

$$\hat{\theta} = \frac{n}{n+\lambda}\underbrace{\frac{1}{n}\sum_{i=1}^{n} y_i}_{\hat{\theta}_{ML}} \tag{1.29}$$

then,

$$\begin{align}
\mathbb{E}[\hat{\theta}|\theta] &= \mathbb{E}\left[\frac{1}{n+\lambda}\sum_{i=1}^{n} y_i|\theta\right] \tag{1.30}\\
&= \frac{1}{n+\lambda}\sum_{i=1}^{n}\mathbb{E}[y_i|\theta] = \frac{1}{n+\lambda}\sum_{i=1}^{n}\theta = \frac{n}{n+\lambda}\theta \tag{1.31}
\end{align}$$

for any $\lambda > 0$, this $\hat{\theta}$ is biased.

$$\begin{align}
\mathbb{V}[\hat{\theta}|\theta] &= \mathbb{V}\left[\frac{1}{n+\lambda}\sum_{i=1}^{n} y_i|\theta\right] \tag{1.32}\\
&= \frac{1}{(n+\lambda)^2}\sum_{i=1}^{n}\mathbb{V}[y_i|\theta] \tag{1.33}\\
&= \frac{n}{(n+\lambda)^2} \tag{1.34}\\
&< \frac{1}{n} = \mathbb{V}[\hat{\theta}_{ML}|\theta] \text{ for any } \lambda > 0. \tag{1.35}
\end{align}$$

1.2 Hypothesis Testing

We want to test $H_0 : \theta = 0$ vs $H_1 : \theta \neq 0$. $\varphi \in \{0, 1\}$ is a test function, where $\phi = 1$ means accept, then the size of the test is defined as:

$$\alpha = \mathbb{P}(\varphi = 1|\theta = 0) = \mathbf{1}\{\theta < \theta_0\} \tag{1.36}$$

We have:

$$\mathbb{P}(\theta|y)\begin{cases} p\left(\theta \in \Theta_0|y\right);\\ p\left(\theta \notin \Theta_0|y\right) = 1 - p\left(\theta \in \Theta_0|y\right).\end{cases} \tag{1.37}$$

Then, the posterior odds ratio is defined as:

$$\frac{p\left(\theta \in \Theta_0|y\right)}{p\left(\theta \in \Theta_1|y\right)} = \frac{p\left(\theta \in \Theta_0|y\right)}{1 - p\left(\theta \in \Theta_0|y\right)} \tag{1.38}$$

The Bayes factor is defined as:

$$BF = \frac{\text{Post. Odds}}{\text{Prior Odds}} \tag{1.39}$$

Example 2. Suppose $\theta \in \{0, 1\}$, and $y|\theta \in \{0, 1, 2, 3, 4\}$,

	0	1	2	3	4
$p(y\|\theta = 0)$	75%	14%	4%	3.7%	3.3%
$p(y\|\theta = 1)$	70%	25.1%	4%	0.5%	0.4%

Table 1.1: Example

Suppose $y = 2$, then the hypothesis test results are:

$$\mathcal{H}_0 : \theta = 0 \rightarrow p(y \geq 2|\theta = 0) = 11\% \tag{1.40}$$

$$\mathcal{H}_1 : \theta = 1 \rightarrow p(y \geq 2|\theta = 1) = 4.9\% \tag{1.41}$$

The Bayes factors are:

$$BF = \frac{p\left(\theta = 1|y = 2\right)}{p\left(\theta = 0|y = 2\right)} \tag{1.42}$$

$$= \frac{p(y = 2|\theta = 1)p(\theta = 1)}{p(y = 2|\theta = 0)p(\theta = 0)} \tag{1.43}$$

Consider $c(y)$ such that $\mathbb{P}\left[\theta \in c(y)|\theta\right] = 1 - \alpha$, e.g. 95%, then the decision rule is:

$$\{\theta_0 \in \Theta : \varphi(\theta_0, \alpha) = 1\} \tag{1.44}$$

Under Bayesian approach, we say: $\mathbb{P}\left[\theta \in c(y)|y\right] = 1 - \alpha$, e.g. 95%, then we have the Highest Posterior Density(HPD) region:

$$c(y) = \{\theta : p(\theta|y) \geq k_\alpha\} \tag{1.45}$$

where k_α is such that $\mathbb{P}\left[\theta \in c(y)|y\right] = 1 - \alpha$.

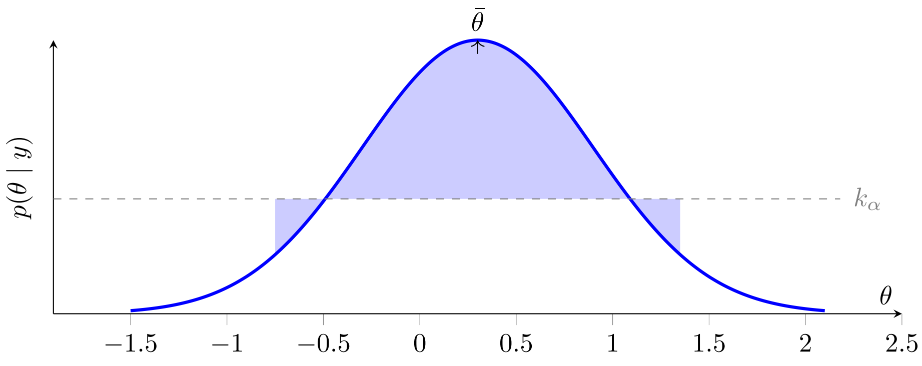

Figure 1.1: HPD Region Example

Now we consider a simple linear regression model:

$$y_i = x_i'\beta + u_i, \quad u_i \sim \mathcal{N}(0, \sigma^2) \tag{1.46}$$

then $y_i|x_i, \beta \sim \mathcal{N}(x_i'\beta, \sigma^2)$.

Denote $\theta = (\beta', \sigma^2)'$ as the parameter of interest, then the likelihood function is:

$$\begin{aligned}
p(y|x,\theta) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(y_i - x_i'\beta)^2\right) && (1.47)\\
&= \prod \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - x_i'\beta)^2\right) && (1.48)\\
&= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - x_i'\beta)^2\right) && (1.49)\\
&= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta)\right) && (1.50)
\end{aligned}$$

and the Maximum Likelihood Estimator will be:

$$\begin{aligned}
\hat{\theta}_{ML} &= \arg\max_{\theta} p(y|x,\theta) && (1.51)\\
&= \arg\min_{\theta} (y - X\beta)'(y - X\beta) && (1.52)
\end{aligned}$$

which we would solve:

$$\begin{aligned}
\hat{\beta} &= (X'X)^{-1}X'y && (1.53)\\
\hat{\sigma}^2 &= \frac{1}{n}\sum_{i=1}^{n}(y_i - x_i'\hat{\beta})^2 && (1.54)
\end{aligned}$$

Example 3. Suppose $\beta \sim \mathcal{N}(\beta_0, \sigma^2 V_0)$, then

$$p(\beta) = (2\pi\sigma^2)^{-\frac{k}{2}}|V_0|^{-\frac{1}{2}} \exp\left(-\frac{1}{2\sigma^2}(\beta - \beta_0)'V_0^{-1}(\beta - \beta_0)\right) \tag{1.55}$$

then the posterior distribution is:

$$\begin{aligned}
p(\beta|y) &\propto p(y|\beta)p(\beta) && (1.56)\\
&= (2\pi\sigma^2)^{-\frac{n+k}{2}}|V_0|^{-\frac{1}{2}} \exp\left(-\frac{1}{2\sigma^2}(\beta - \beta_0)'V_0^{-1}(\beta - \beta_0)\right) \cdot \exp\left(-\frac{1}{2\sigma^2}(Y - X\beta)'(Y - X\beta)\right) && (1.57)\\
&\propto \exp\left(\frac{1}{2\sigma^2}\left[-\beta'X'Y - Y'X\beta + \beta X'X\beta + \beta_0'V_0^{-1}\beta_0 - \beta'V_0^{-1}\beta_0 - \beta_0'V_0^{-1}\beta\right]\right) && (1.58)\\
&\propto \exp\left(-\frac{1}{2\sigma^2}\left[\beta'(X'X + V_0^{-1})\beta - 2(X'Y + V_0^{-1}\beta_0)'\beta\right]\right) && (1.59)
\end{aligned}$$

This let us guess that $\beta|Y \sim \mathcal{N}(\bar{\beta}, \sigma^2\bar{V})$, with:

$$\begin{aligned}
\bar{V} &= \left[X'X + V_0^{-1}\right]^{-1} && (1.60)\\
\bar{\beta} &= \bar{V}\left(X'Y + V_0^{-1}\beta_0\right) = \left(X'X + V_0^{-1}\right)^{-1}\left(X'X\hat{\beta}_{ML} + V_0^{-1}\beta_0\right). && (1.61)
\end{aligned}$$

We can calculate the probability $p(y)$,

$$\begin{aligned}
p(y) &= \frac{p(y|\beta)p(\beta)}{p(\beta|y)} \\
&= \frac{(2\pi\sigma^2)^{-\frac{n+k}{2}}|V_0|^{-\frac{1}{2}}\exp\left(-\frac{1}{2\sigma^2}(\beta-\beta_0)'V_0^{-1}(\beta-\beta_0)\right)\cdot\exp\left(-\frac{1}{2\sigma^2}\left(Y-X\beta\right)'\left(Y-X\beta\right)\right)}{(2\pi\sigma^2)^{-\frac{k}{2}}|\bar{V}|^{-\frac{1}{2}}\exp\left(-\frac{1}{2\sigma^2}(\beta-\bar{\beta})'\bar{V}^{-1}(\beta-\bar{\beta})\right)} \\
&= (2\pi\sigma^2)^{-\frac{n}{2}}\left(\frac{|V_0|}{|\bar{V}|}\right)^{-\frac{1}{2}}\exp\left(-\frac{1}{2\sigma^2}\left(Y'Y+\beta_0'V_0^{-1}\beta_0-\bar{\beta}'\bar{V}^{-1}\bar{\beta}\right)\right)
\end{aligned}$$

Or, we can integrate out β:

$$p(y) = \int p(y|\beta)p(\beta)d\beta = \mathbb{E}_\beta[p(y|\beta)]$$

1.3 Ridge Regression

Under the previous normal prior assumption, we can simplify the prior to $\beta_j \sim \mathcal{N}(0, \sigma^2\lambda^{-1}I)$, then we have $V_0 = \lambda^{-1}I$.

$$\begin{aligned}
\beta|y &\sim \mathcal{N}(\bar{\beta}, \sigma^2\bar{V}) \\
\bar{V}^{-1} &= \left(X'X+\lambda I\right)^{-1} \\
\bar{\beta} &= \left(X'X+\lambda I\right)X'Y
\end{aligned}$$

The Ridge regression estimator is:

$$\overline{\beta} = \arg\min_\beta (Y-X\beta)'(Y-X\beta)+\lambda\beta'\beta \tag{1.62}$$

With prior λ and $\sigma^2 = 1$, the MDD expression from above can be simplified to:

$$p(y) = \frac{(2\pi)^{-\frac{n}{2}}\exp\left(-\frac{1}{2}Y'Y\right)\left|\lambda^{-1}I_k\right|^{-\frac{1}{2}}}{\left|X'X+\lambda I\right|^{\frac{1}{2}}\exp\left(-\frac{1}{2}\overline{\beta}'\overline{V}^{-1}\overline{\beta}\right)} \tag{1.63}$$

Taking logs, we get:

$$\begin{aligned}
\log p(y) &= c - \frac{1}{2}Y'Y + \frac{1}{2}\overline{\beta}'\overline{V}^{-1}\overline{\beta} - \frac{1}{2}\log\left|X'X+\lambda I_k\right| \\
&= c - \frac{1}{2}\left[Y'Y - Y'X\overline{V}X'Y\right] - \frac{1}{2}\log\lambda^{-k}\left|X'X+\lambda I\right| \\
&= c - \frac{1}{2}\left[Y'Y - Y'X\left(X'X+\lambda I_k\right)^{-1}X'Y\right] - \frac{1}{2}\log\left|\lambda^{-1}X'X+I\right|
\end{aligned}$$

where $c = -\frac{n}{2}\log(2\pi)$ is a constant that doesn't depend on Y or λ.

The penalty term:

$$\begin{aligned}
\log\left|\lambda^{-1}X'X+I\right| &= -\frac{1}{2}\log\left|n\left(\frac{1}{n}\sum_i x_i x_i' + \frac{\lambda}{n}I\right)\right| \\
&= -\frac{k}{2}\log n - \frac{1}{2}\log\left|Q_n + \frac{\lambda}{n}I\right|
\end{aligned}$$

Definition 1.3.1. The Bayesian Information Criterion (BIC) is defined as:

$$BIC = \log p(y|\hat{\beta}_{ML}) - \frac{k}{2}\log n \tag{1.64}$$

where $\hat{\beta}_{ML}$ is the maximum likelihood estimate of β.

For $\beta \sim \mathcal{N}\left(\beta_0, \sigma^2 V_0\right)$,

- Ridge: $\beta \sim \mathcal{N}(0, \sigma^2\lambda^{-1}I)$, $\beta_j \sim \mathcal{N}(0, \sigma^2\lambda^{-1})$
- Lasso: $\beta \sim \mathit{Laplace}(\cdot)$, $p(\beta_j) \propto \exp\left(-\lambda\left|\beta_j\right|\right)$, where the mode solves

$$\min_{\beta}\left\{(Y - X\beta)'(Y - X\beta) + \tilde{\lambda}\sum_{j=1}^{k}|\beta_j|\right\}$$

.

1.3.1 Model Selection

Suppose we have regression models M_j, with parameters θ_j

Suppose we have two regression models:

$$M_1 : y_i = \beta_1^1 x_{1i} + \beta_2^1 x_{2i} + u_i, \quad M_2 : y_i = \beta_1^2 x_{1i} + v_i \tag{1.65}$$

with $p\left(\beta_1^1, \beta_2^1|M_1\right) = \mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}\right)$, and for model 2, we have $p\left(\beta_1^2|M_2\right) = \mathcal{N}\left(0, 1\right)$ $p\left(\beta_1^2|M_2\right) = \delta_0$

Then $p\left(\beta_1^2, \beta_2^2|M_2\right) = \mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}1 & 0\\0 & 0\end{bmatrix}\right)$, so, we can write:

$$p(\beta) = \mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}1 & 0\\0 & \lambda\end{bmatrix}\right)$$

with $\lambda \in \{0, 1\}$.

$$p(\theta|y) = \sum_j \pi_{j,n} p(\theta|y, M_j)$$

For $y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + u_i$, with $p(\beta|M_j) = \mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}1 & 0\\0 & \lambda\end{bmatrix}\right)$, we have:

$$\lambda = \begin{cases} 0, & \pi_{2,0} \\ 1, & \pi_{1,0} \end{cases}$$

, then

$$\begin{aligned} p(\beta, \lambda) &= p(\beta|\lambda)p(\lambda) \\ &= \pi_{1,0}\cdot\mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}\right) + \pi_{2,0}\cdot\mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}1 & 0\\0 & 0\end{bmatrix}\right) \end{aligned}$$

so,

$$p(\beta, \lambda|y) = p(\beta|\lambda, y)p(\lambda|y)$$

$$= \pi_{1,n} p(\beta|y, \lambda = 1) + \pi_{2,n} p(\beta|y, \lambda = 0)$$

To get $p(\theta, \lambda|y)$, we can consider $p(\lambda|y)$, whose mode is given by: $\arg\max_\lambda p(y|\lambda)$, and we can also consider the posterior $p(\theta|y)$, which equals $\int p(\theta|y, \lambda)p(y|\lambda)d\lambda$.

Lecture 2.

Appendix

2.1 Yule-Walker

2.2 Kronecker Products & Vector Operator

Recommended Resources

Books

[1] Helmut Lütkepohl. *New Introduction to Multiple Time Series Analysis*. New York: Springer, 2005

[2] James H. Stock and Mark W. Watson. *Introduction to Econometrics*. 4th ed. New York: Pearson, 2003

[3] Jeffrey M. Wooldridge. *Introductory Econometrics: A Modern Approach*. 7th ed. Cengage Learning, 2020

[4] Bruce E. Hansen. *Econometrics*. Princeton, New Jersey: Princeton University Press, 2022

[5] Fumio Hayashi. *Econometrics*. Princeton, New Jersey: Princeton University Press, 2000

[6] Jeffrey M. Wooldridge. *Econometric Analysis of Cross Section and Panel Data*. 2nd ed. Cambridge, Massachusetts: The MIT Press, 2010

[7] Joshua Chan et al. *Bayesian Econometric Methods*. 2nd ed. Cambridge, United Kingdom: Cambridge University Press, 2019

[8] Badi H. Baltagi. *Econometric Analysis of Panel Data*. 6th ed. Cham, Switzerland: Springer, 2021

[9] James D. Hamilton. *Time Series Analysis*. Princeton, New Jersey: Princeton University Press, 1994. ISBN: 9780691042893

[10] Takeshi Amemiya. *Advanced Econometrics*. Cambridge, MA: Harvard University Press, 1985

[11] Peng Ding. *A First Course in Causal Inference*. 2023. arXiv: `2305.18793 [stat.ME]`. URL: `https://arxiv.org/abs/2305.18793`

[12] David Walters. *An Introduction to Stochastic Processes and Their Applications*. New York: Dover Publications, 1982

[13] Peter Whittle. *Hypothesis Testing in Time Series Analysis*. Uppsala: Almqvist & Wiksells, 1951

[14] Colorado P. J. Brockwell Fort Collins. *Time Series: Theory and Methods*. New York: Springer, 1991

Others

[15] Roger Bowden. “The Theory of Parametric Identification”. In: *Econometrica* 41.6 (1973), pp. 1069–1074. DOI: `10.2307/1914036`

[16] Robert I. Jennrich. “Asymptotic Properties of Non-linear Least Squares Estimators”. In: *The Annals of Mathematical Statistics* 40.2 (1969), pp. 633–643. DOI: `10.1214/aoms/1177697731`

[17] Michael P. Keane. “A Note on Identification in the Multinomial Probit Model”. In: *Journal of Business & Economic Statistics* 10.2 (1992), pp. 193–200. DOI: `10.1080/07350015.1992.10509906`

[18] Thomas J. Rothenberg. “Identification in Parametric Models”. In: *Econometrica* 39.3 (1971), pp. 577–591. DOI: `10.2307/1913267`

[19] George Tauchen. “Diagnostic Testing and Evaluation of Maximum Likelihood Models”. In: *Journal of Econometrics* 30 (1985), pp. 415–443. DOI: `10.1016/0304-4076(85)90149-6`

[20] Abraham Wald. "Note on the Consistency of the Maximum Likelihood Estimate". In: *The Annals of Mathematical Statistics* 20.4 (1949), pp. 595–601. DOI: 10.1214/aoms/1177729952

[21] Halbert White. "Maximum Likelihood Estimation of Misspecified Models". In: *Econometrica* 50.1 (1982), pp. 1–25. DOI: 10.2307/1912526

[22] Jerzy Neyman. "On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9". In: *Statistical Science* 5.4 (1923), pp. 465–472

[23] Donald B. Rubin. "Comment on "Randomization Analysis of Experimental Data: The Fisher Randomization Test" by D. Basu". In: *Journal of the American Statistical Association* 75.371 (1980), pp. 591–593. DOI: 10.1080/01621459.1980.10477410

[24] Donald B. Rubin. "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies". In: *Journal of Educational Psychology* 66.5 (1974), pp. 688–701. DOI: 10.1037/h0037350

[25] Donald B. Rubin. "Bayesian Inference for Causality: The Importance of Randomization". In: *The Annals of Statistics* 3.1 (1975), pp. 121–131. DOI: 10.1214/aos/1176343238

[26] Donald B. Rubin. "Causal Inference Using Potential Outcomes: Design, Modeling, Decisions". In: *Journal of the American Statistical Association* 100.469 (2005), pp. 322–331. DOI: 10.1198/016214504000001880

[27] Paul W. Holland. "Statistics and Causal Inference(with discussion)". In: *Journal of the American Statistical Association* 81.396 (1986), pp. 945–960. DOI: 10.1080/01621459.1986.10478373