Econometrics: Lecture 9

Joint Statistical Inference on OLS Estimator

Hang Miao

Rutgers University

January 21, 2021

Overview

1. Joint Statistical Inference with known σ^2 and Large n
   - Wald Test
   - Hypothesis Testing
   - P-Value
2. Joint Statistical Inference with Unknown σ^2 and Large n
3. Joint Statistical Inference with Unknown σ^2 and Small n
4. Joint Statistical Inference with known σ^2 and Small n
5. Appendix
   - Properties of Idempotent Matrix
   - χ^2, t, F-Distribution

Joint Statistical Inference with known σ^2 and Large n

Linear Relationship between the entris of β

Motivation

Previously we introduce the technique of how to test an individual parameter against the null hypothesis in favor of the alternative hypothesis. What if we are interested in testing against a null hypothesis which is a joint relationship of several parameters?

Linear Null and Alternative Hypothesis

Suppose the null and alternative hypothesis is $H_0 : \boldsymbol{A}\beta = \boldsymbol{a}$. The alternative is $H_1 : \boldsymbol{A}\beta \neq \boldsymbol{a}$ Where the value of matrix $\boldsymbol{A}$ and vector $\boldsymbol{a}$ are known and specified by the hypothesis.

Dimension of the matrix $\boldsymbol{A}$ and vector $\boldsymbol{a}$

The number of column of $\boldsymbol{A}$ is always $p+1$. The number of rows of $\boldsymbol{A}$ is determined by how many restrictive equations imposed to the null hypothesis. We assume there is no redundant restrictive equations, so the $rank(\boldsymbol{A})$ is equal to the number of rows of $\boldsymbol{A}$ and number of rows of $\boldsymbol{a}$: $rank(\boldsymbol{A}) = nrows(\boldsymbol{A}) = nrows(\boldsymbol{a}) \equiv d$

Wald Test

Wald Test Statistic

$$W = (\boldsymbol{A}\hat{\beta}_{ols} - \boldsymbol{a})' \left[\boldsymbol{A}\sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{A}'\right]^{-1} (\boldsymbol{A}\hat{\beta}_{ols} - \boldsymbol{a})$$

Distribution of Wald Test Statistic

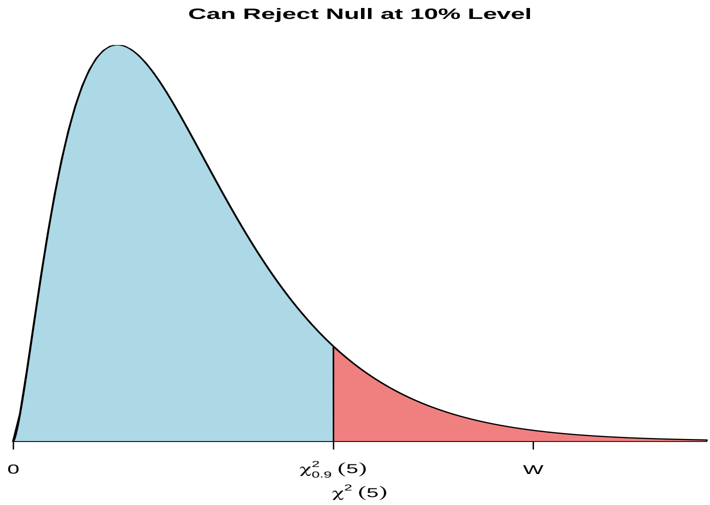
Under assumption 1-4 and Null Hypothesis $H_0 : \boldsymbol{A}\beta = \boldsymbol{a}$, the Wald Test Statistic follows χ^2-distribution with degree of freedom d: $W \sim \chi^2(d)$

Proof of $W \sim \chi^2(d)$

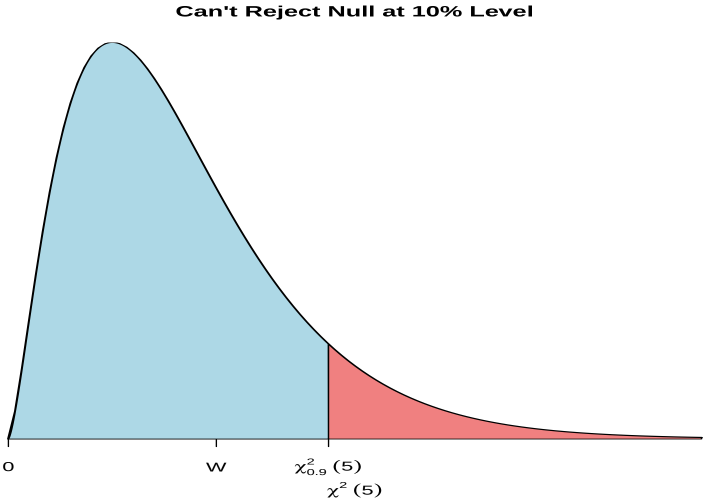
$$\begin{aligned} W &= (\boldsymbol{A}\hat{\beta}_{ols} - \boldsymbol{a})' \left[\boldsymbol{A}\sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{A}'\right]^{-1} (\boldsymbol{A}\hat{\beta}_{ols} - \boldsymbol{a}) \\ &= (\boldsymbol{A}\hat{\beta}_{ols} - \boldsymbol{A}\beta)' \left[\boldsymbol{A}\sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{A}'\right]^{-1} (\boldsymbol{A}\hat{\beta}_{ols} - \boldsymbol{A}\beta) \\ &= \left(\boldsymbol{A}(\hat{\beta}_{ols} - \beta)\right)' \left[\boldsymbol{A}\sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{A}'\right]^{-1} \left(\boldsymbol{A}(\hat{\beta}_{ols} - \beta)\right) \end{aligned}$$

$$
\begin{aligned}
\because\ (\hat{\boldsymbol{\beta}}_{ols} - \boldsymbol{\beta})\,|\,\boldsymbol{X} &\sim N\Big(\boldsymbol{0},\ \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\Big) \quad \textit{(by Central Limit Theorem)} \\
\therefore\ \boldsymbol{A}(\hat{\boldsymbol{\beta}}_{ols} - \boldsymbol{\beta}) &\sim N\Big(\boldsymbol{0},\ \boldsymbol{A}\sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{A}'\Big) \\
&\qquad \textit{(by properties of Multivariate Normal Distribution)} \\
\boldsymbol{A}\sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{A}' &= \boldsymbol{Q}'\boldsymbol{\Lambda}\boldsymbol{Q} \quad \textit{(Eigen Decomposition)} \\
W &= \Big(\boldsymbol{A}(\hat{\boldsymbol{\beta}}_{ols} - \boldsymbol{\beta})\Big)'\left[\boldsymbol{Q}'\boldsymbol{\Lambda}\boldsymbol{Q}\right]^{-1}\Big(\boldsymbol{A}(\hat{\boldsymbol{\beta}}_{ols} - \boldsymbol{\beta})\Big) \\
&= \Big(\boldsymbol{A}(\hat{\boldsymbol{\beta}}_{ols} - \boldsymbol{\beta})\Big)'\boldsymbol{Q}'\boldsymbol{\Lambda}^{-\frac{1}{2}}\boldsymbol{\Lambda}^{-\frac{1}{2}}\boldsymbol{Q}\Big(\boldsymbol{A}(\hat{\boldsymbol{\beta}}_{ols} - \boldsymbol{\beta})\Big) \\
&= \Big(\boldsymbol{\Lambda}^{-\frac{1}{2}}\boldsymbol{Q}\boldsymbol{A}(\hat{\boldsymbol{\beta}}_{ols} - \boldsymbol{\beta})\Big)'\Big(\boldsymbol{\Lambda}^{-\frac{1}{2}}\boldsymbol{Q}\boldsymbol{A}(\hat{\boldsymbol{\beta}}_{ols} - \boldsymbol{\beta})\Big) \\
\boldsymbol{\Lambda}^{-\frac{1}{2}}\boldsymbol{Q}\boldsymbol{A}(\hat{\boldsymbol{\beta}}_{ols} - \boldsymbol{\beta}) &\sim N\Big(\boldsymbol{0},\ \boldsymbol{\Lambda}^{-\frac{1}{2}}\boldsymbol{Q}\,\boldsymbol{Q}'\boldsymbol{\Lambda}\boldsymbol{Q}\,\boldsymbol{Q}'\boldsymbol{\Lambda}^{-\frac{1}{2}}\Big) = N\Big(\boldsymbol{0},\ \boldsymbol{I}_d\Big) \\
\therefore\ W &= \sum_{i=1}^{d} Z_i^2 \quad \text{where } Z_i \sim N(0,1) \\
W &\sim \chi^2(d)
\end{aligned}
$$

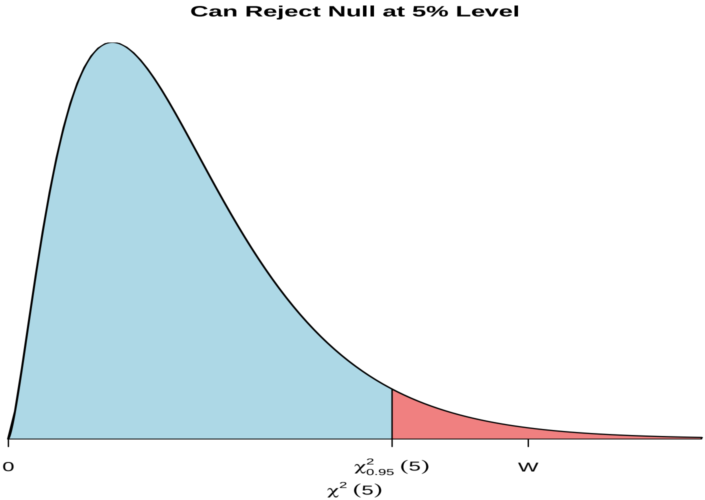
Reject the $H_0 : \boldsymbol{A}\beta = \boldsymbol{a}$ at 10% Significant Level



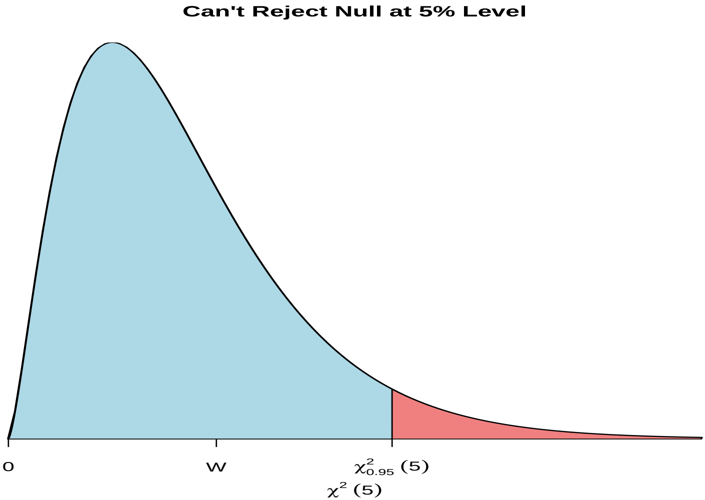
Can't Reject the $H_0 : \boldsymbol{A}\beta = \boldsymbol{a}$ at 10% Significant Level



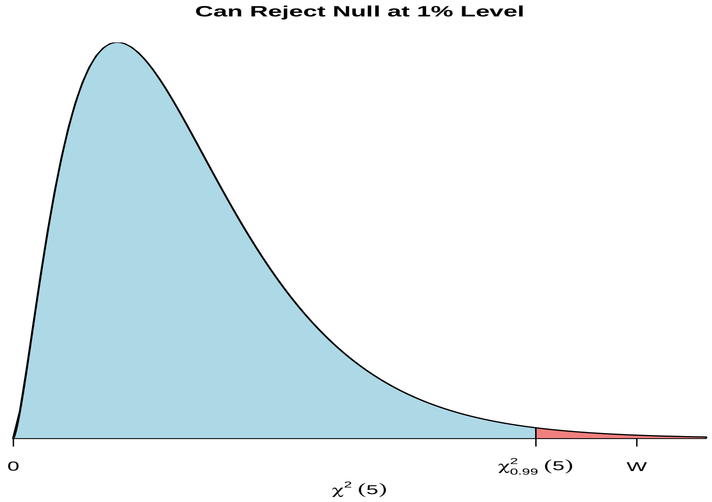
Reject the $H_0 : \boldsymbol{A}\beta = \boldsymbol{a}$ at 5% Significant Level



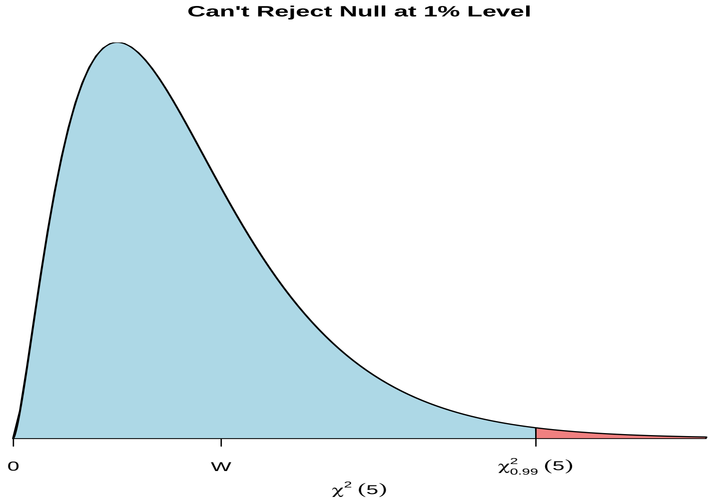
Can't Reject the $H_0 : \boldsymbol{A}\beta = \boldsymbol{a}$ at 5% Significant Level



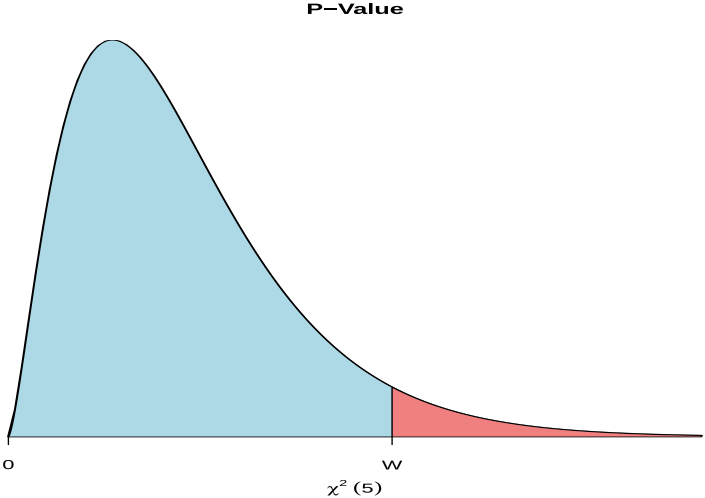
Reject the $H_0 : \boldsymbol{A}\beta = \boldsymbol{a}$ at 1% Significant Level



Can't Reject the $H_0 : \boldsymbol{A}\beta = \boldsymbol{a}$ at 1% Significant Level



P-Value for $H_0 : \boldsymbol{A\beta} = \boldsymbol{a}$



Joint Statistical Inference with Unknown σ^2 and Large n

Wald Test

Estimate σ^2

If σ^2 is unkown, suppose assumption 1-4 hold, then we could estimate it by the previous unbiased and consistent estimator:

$$\hat{\sigma}^2 = \frac{1}{n-p-1}\sum_{i=1}^{n}\hat{u}_i^2 = \frac{\hat{\boldsymbol{u}}'\hat{\boldsymbol{u}}}{n-p-1} = \frac{\text{SSR}}{n-p-1}$$

Wald Test Statistic

$$W = (\boldsymbol{A}\hat{\beta}_{ols} - \boldsymbol{a})'\left[\boldsymbol{A}\,\hat{\sigma}^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{A}'\right]^{-1}(\boldsymbol{A}\hat{\beta}_{ols} - \boldsymbol{a})$$

Distribution of Wald Test Statistic

Since $\hat{\sigma}^2 \xrightarrow{p} \sigma^2$, under the null hypothesis $H_0 : \boldsymbol{A}\beta = \boldsymbol{a}$, the Wald Test Statistic follows χ^2-distribution with degree of freedom d: $W \sim \chi^2(d)$. The hypothesis testing and P-value is same as previous case.

Joint Statistical Inference with Unknown σ^2 and Small n

F Test

When the sample size n is small, Central Limit Theory can not be applied, so the additional assumption have to be made so that we could have a statistic with known dsitribution. The normality assumption below is the same one we made getting the $\hat{\beta}_k$ follows t-distribution.

Assumptions For Getting F-distribution

5. uncertainty u follows Normal Distribution: $\boldsymbol{u}|\boldsymbol{X} \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_n)$
   i.e $u_i|\boldsymbol{X} \sim N(0, \sigma^2)$

Since σ^2 is unknown, we still use the unbiased and consistent estimator $\hat{\sigma}^2$ together with assumption 1-4, we construct the F test statistic similar to the Wald test statistic:

F test statistic (Wald Principle)

$$\text{F} = (\boldsymbol{A}\hat{\beta}_{ols} - \boldsymbol{a})' \left[\boldsymbol{A}\,\hat{\sigma}^2 (\boldsymbol{X}'\boldsymbol{X})^{-1} \boldsymbol{A}'\right]^{-1} (\boldsymbol{A}\hat{\beta}_{ols} - \boldsymbol{a})/d$$
$$\text{F} \sim F(d, n-p-1)$$

Another Expression for F Statistic

Another equivalent expression for the F test statistic is as follow:

F test statistic (Likelihood-Ratio Principle)

$$\begin{aligned} \mathrm{F} &= \frac{(\mathrm{SSR}_R - \mathrm{SSR}_U)/d}{\mathrm{SSR}_U/(n-p-1)} \\ \mathrm{F} &\sim F(d, n-p-1) \end{aligned}$$

Definition (SSR_R Restricted Sum of Squared Residual)

Where SSR_R is Restricted Sum of Squared Residual. It is the minimum of SSR where the estimated $\tilde{\boldsymbol{\beta}}$ is under the restriction that $\boldsymbol{A}\tilde{\boldsymbol{\beta}} = \boldsymbol{a}$

Definition (SSR_U Unrestricted Sum of Squared Residual)

Where SSR_U is Unrestricted Sum of Squared Residual. It is the minimum of SSR where the estimated $\tilde{\boldsymbol{\beta}}$ is under no restriction. Therefore $\tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}_{ols}$, and SSR_U is just the usual SSR obtained by the $\hat{\boldsymbol{\beta}}_{ols}$

Proof F-statistic follows F-distribution

$$
\begin{aligned}
\text{F} &= (\boldsymbol{A}\hat{\beta}_{ols} - \boldsymbol{a})' \left[\boldsymbol{A}\,\hat{\sigma}^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{A}'\right]^{-1} (\boldsymbol{A}\hat{\beta}_{ols} - \boldsymbol{a})/d \\
&= \frac{(\boldsymbol{A}\hat{\beta}_{ols} - \boldsymbol{a})' \left[\boldsymbol{A}\,\sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{A}'\right]^{-1} (\boldsymbol{A}\hat{\beta}_{ols} - \boldsymbol{a})}{d} \left(\frac{(n-p-1)\hat{\sigma}^2}{(n-p-1)\sigma^2}\right)^{-1}
\end{aligned}
$$

We want to show:

- $(\boldsymbol{A}\hat{\beta}_{ols} - \boldsymbol{a})' \left[\boldsymbol{A}\,\sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{A}'\right]^{-1} (\boldsymbol{A}\hat{\beta}_{ols} - \boldsymbol{a}) \sim \chi^2(d)$
- $\frac{(n-p-1)\hat{\sigma}^2}{\sigma^2} \sim \chi^2(n-p-1)$
- $(\boldsymbol{A}\hat{\beta}_{ols} - \boldsymbol{a})' \left[\boldsymbol{A}\,\sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{A}'\right]^{-1} (\boldsymbol{A}\hat{\beta}_{ols} - \boldsymbol{a}) \sim \chi^2(d)$ and $\frac{\hat{\sigma}^2}{\sigma^2}$ are independent

$(\boldsymbol{A}\hat{\boldsymbol{\beta}}_{ols} - \boldsymbol{a})' \left[\boldsymbol{A}\,\sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{A}'\right]^{-1}(\boldsymbol{A}\hat{\boldsymbol{\beta}}_{ols} - \boldsymbol{a}) \sim \chi^2(d)$

$$
\begin{aligned}
\boldsymbol{A}\hat{\boldsymbol{\beta}}_{ols} - \boldsymbol{a} &= \boldsymbol{A}\hat{\boldsymbol{\beta}}_{ols} - \boldsymbol{A}\boldsymbol{\beta} = \boldsymbol{A}\big(\hat{\boldsymbol{\beta}}_{ols} - \boldsymbol{\beta}\big) \\
&= \boldsymbol{A}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{u} \\
\because\ \boldsymbol{u} &\sim N(0, \sigma^2\boldsymbol{I}_n) \quad \textit{from new assumption 5} \\
\therefore\ \boldsymbol{A}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{u} &\sim N\Big(0,\ \boldsymbol{A}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\sigma^2\boldsymbol{I}_n\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{A}'\Big) \\
\therefore\ \boldsymbol{A}\hat{\boldsymbol{\beta}}_{ols} - \boldsymbol{a} &\sim N\Big(0,\ \boldsymbol{A}\sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{A}'\Big) \\
\boldsymbol{A}\sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{A}' &= \boldsymbol{Q}'\boldsymbol{\Lambda}\boldsymbol{Q} \quad \textit{(Eigen Decomposition)} \\
W &= \Big(\boldsymbol{A}\big(\hat{\boldsymbol{\beta}}_{ols} - \boldsymbol{\beta}\big)\Big)' \left[\boldsymbol{Q}'\boldsymbol{\Lambda}\boldsymbol{Q}\right]^{-1} \Big(\boldsymbol{A}\big(\hat{\boldsymbol{\beta}}_{ols} - \boldsymbol{\beta}\big)\Big) \\
&= \Big(\boldsymbol{A}\big(\hat{\boldsymbol{\beta}}_{ols} - \boldsymbol{\beta}\big)\Big)' \boldsymbol{Q}'\boldsymbol{\Lambda}^{-\frac{1}{2}}\boldsymbol{\Lambda}^{-\frac{1}{2}}\boldsymbol{Q} \Big(\boldsymbol{A}\big(\hat{\boldsymbol{\beta}}_{ols} - \boldsymbol{\beta}\big)\Big) \\
&= \Big(\boldsymbol{\Lambda}^{-\frac{1}{2}}\boldsymbol{Q}\boldsymbol{A}\big(\hat{\boldsymbol{\beta}}_{ols} - \boldsymbol{\beta}\big)\Big)' \Big(\boldsymbol{\Lambda}^{-\frac{1}{2}}\boldsymbol{Q}\boldsymbol{A}\big(\hat{\boldsymbol{\beta}}_{ols} - \boldsymbol{\beta}\big)\Big)
\end{aligned}
$$

$(\boldsymbol{A}\hat{\boldsymbol{\beta}}_{ols} - \boldsymbol{a})' \left[\boldsymbol{A}\,\sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{A}'\right]^{-1} (\boldsymbol{A}\hat{\boldsymbol{\beta}}_{ols} - \boldsymbol{a}) \sim \chi^2(d)$

$$
\begin{aligned}
\boldsymbol{\Lambda}^{-\frac{1}{2}}\boldsymbol{Q}\boldsymbol{A}(\hat{\boldsymbol{\beta}}_{ols} - \boldsymbol{\beta}) &\sim N\Big(\mathbf{0},\ \boldsymbol{\Lambda}^{-\frac{1}{2}}\boldsymbol{Q}\,\boldsymbol{Q}'\boldsymbol{\Lambda}\boldsymbol{Q}\,\boldsymbol{Q}'\boldsymbol{\Lambda}^{-\frac{1}{2}}\Big) = N\Big(\mathbf{0},\ \boldsymbol{I}_d\Big) \\
\therefore W &= \sum_{i=1}^{d} Z_i^2 \quad \text{where } Z_i \sim N(0,1) \\
W &\sim \chi^2(d)
\end{aligned}
$$

$\frac{(n-p-1)\hat{\sigma}^2}{\sigma^2} \sim \chi^2(n-p-1)$

$$\begin{aligned}
\frac{(n-p-1)\hat{\sigma}^2}{\sigma^2} &= \frac{\hat{\boldsymbol{u}}'\hat{\boldsymbol{u}}}{\sigma^2} \\
&= \frac{\boldsymbol{u}'\boldsymbol{M}\boldsymbol{u}}{\sigma^2} \\
&= \frac{\boldsymbol{u}'\boldsymbol{Q}'\boldsymbol{\Lambda}\boldsymbol{Q}\boldsymbol{u}}{\sigma^2} \quad \textit{(by eigen-decomposition)} \\
&= \frac{\boldsymbol{u}'\boldsymbol{Q}'\begin{bmatrix}\boldsymbol{I}_{(n-p-1)} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0}\end{bmatrix}\boldsymbol{Q}\boldsymbol{u}}{\sigma^2} \quad \textit{(eigenvalues are either 1 or 0)} \\
&= \frac{\sum_{i=1}^{n-p-1} u_i^2}{\sigma^2} = \sum_{i=1}^{n-p-1}\left(\frac{u_i}{\sigma}\right)^2 \sim \chi(n-p-1)
\end{aligned}$$

$$\left(\because u_i \sim N(0,\sigma^2) \implies \frac{u_i}{\sigma} \sim N(0,1)\right)$$

$(\boldsymbol{A}\hat{\beta}_{ols} - \boldsymbol{a})' \left[\boldsymbol{A}\,\sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{A}'\right]^{-1} (\boldsymbol{A}\hat{\beta}_{ols} - \boldsymbol{a})$ and $\frac{\hat{\sigma}^2}{\sigma^2}$ are independent

Since $\boldsymbol{A}, \boldsymbol{a}, \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$, σ^2 are constant, and $\hat{\sigma}^2 = \frac{\hat{\boldsymbol{u}}'\hat{\boldsymbol{u}}}{n-p-1}$ it is equivalent to show that $\hat{\beta}_{ols}$ and $\hat{\boldsymbol{u}}$ are independent

$$\begin{aligned}
\hat{\beta}_{ols} &= \beta + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{u} \sim N\left(\beta, \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\right) \\
\hat{\boldsymbol{u}} &= \boldsymbol{M}\boldsymbol{u} \sim N\left(\boldsymbol{0}, \sigma^2\boldsymbol{M}\right) \\
\mathrm{Cov}(\hat{\beta}_{ols},\, \hat{\boldsymbol{u}} \,|\, \boldsymbol{X}) &= \mathrm{E}\left[\,(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{u}\,(\boldsymbol{M}\boldsymbol{u})' \,|\, \boldsymbol{X}\,\right] \\
&= \mathrm{E}\left[\,(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{u}\,\boldsymbol{u}'\boldsymbol{M} \,|\, \boldsymbol{X}\,\right] \\
&= \mathrm{E}\left[\,(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\sigma^2\boldsymbol{I}_n\boldsymbol{M} \,|\, \boldsymbol{X}\,\right] \\
&= \sigma^2\mathrm{E}\left[\,(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{M} \,|\, \boldsymbol{X}\,\right] \\
&= \sigma^2\mathrm{E}\left[\,\boldsymbol{0} \,|\, \boldsymbol{X}\,\right] \\
&= \boldsymbol{0} \\
\Longrightarrow &\quad \text{independence of } \hat{\beta}_{ols},\ \hat{\boldsymbol{u}}
\end{aligned}$$

Proof that the Two F Statistics is Equivalent

$$
\begin{aligned}
\mathrm{F} &= (\boldsymbol{A}\hat{\beta}_{ols} - \boldsymbol{a})' \left[\boldsymbol{A}\,\hat{\sigma}^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{A}'\right]^{-1} (\boldsymbol{A}\hat{\beta}_{ols} - \boldsymbol{a})/d \\
&= \frac{(\boldsymbol{A}\hat{\beta}_{ols} - \boldsymbol{a})' \left[\boldsymbol{A}\,(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{A}'\right]^{-1} (\boldsymbol{A}\hat{\beta}_{ols} - \boldsymbol{a})/d}{\hat{\sigma}^2} \\
&= \frac{(\boldsymbol{A}\hat{\beta}_{ols} - \boldsymbol{a})' \left[\boldsymbol{A}\,(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{A}'\right]^{-1} (\boldsymbol{A}\hat{\beta}_{ols} - \boldsymbol{a})/d}{\mathrm{SSR}/(n-p-1)} \\
\mathrm{F} &= \frac{(\mathrm{SSR}_R - \mathrm{SSR}_U)/d}{\mathrm{SSR}_U/(n-p-1)} = \frac{(\mathrm{SSR}_R - \mathrm{SSR})/d}{\mathrm{SSR}/(n-p-1)}
\end{aligned}
$$

It turns out the only thing we need to prove is:

$$
(\boldsymbol{A}\hat{\beta}_{ols} - \boldsymbol{a})' \left[\boldsymbol{A}\,(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{A}'\right]^{-1} (\boldsymbol{A}\hat{\beta}_{ols} - \boldsymbol{a}) = (\mathrm{SSR}_R - \mathrm{SSR})
$$

Proof of $(\boldsymbol{A}\hat{\beta}_{ols} - \boldsymbol{a})' \left[\boldsymbol{A}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{A}'\right]^{-1}(\boldsymbol{A}\hat{\beta}_{ols} - \boldsymbol{a}) = (\mathrm{SSR}_R - \mathrm{SSR})$

Let $\hat{\beta}_{Rols}$ be the minimizer of the objective function $\mathrm{SSR}(\tilde{\beta})$ constraint on $\boldsymbol{A}\tilde{\beta} = \boldsymbol{a}$ that is:

$$\hat{\beta}_{Rols} = \mathrm{argmin}_{\boldsymbol{A}\tilde{\beta}=\boldsymbol{a}}\ \mathrm{SSR}(\tilde{\beta})$$

Using the Lagrange multiplier to solve this constraint optimization problem:

$$\begin{aligned} \mathcal{L} &= \mathrm{SSR}(\tilde{\beta}) + \lambda(\boldsymbol{A}\tilde{\beta} - \boldsymbol{a}) \\ &= (\boldsymbol{Y} - \boldsymbol{X}\tilde{\beta})'(\boldsymbol{Y} - \boldsymbol{X}\tilde{\beta}) + \lambda(\boldsymbol{A}\tilde{\beta} - \boldsymbol{a}) \end{aligned}$$

The minimizer $\hat{\beta}_{Rols}$ must satisfy the First Order Condition:

$$2\boldsymbol{X}'(\boldsymbol{Y} - \boldsymbol{X}\hat{\beta}_{Rols}) = \lambda\boldsymbol{A}' \quad (1)$$

$$\boldsymbol{A}\hat{\beta}_{Rols} = \boldsymbol{a} \quad (2)$$

from (1) we have

$$
\begin{aligned}
2\boldsymbol{X}'(\boldsymbol{Y} - \boldsymbol{X}\hat{\beta}_{ols} + \boldsymbol{X}\hat{\beta}_{ols} - \boldsymbol{X}\hat{\beta}_{Rols}) &= \lambda \boldsymbol{A}' \\
\Longrightarrow \; 2\boldsymbol{X}'\boldsymbol{X}(\hat{\beta}_{ols} - \hat{\beta}_{Rols}) &= \lambda \boldsymbol{A}' \\
\Longrightarrow \; (\boldsymbol{X}'\boldsymbol{X})^{-1}(\boldsymbol{X}'\boldsymbol{X}\hat{\beta}_{ols} - \lambda \boldsymbol{A}') &= \hat{\beta}_{Rols} \qquad (3)
\end{aligned}
$$

Plug (3) into (2) we have

$$
\begin{aligned}
\boldsymbol{A}(\boldsymbol{X}'\boldsymbol{X})^{-1}(\boldsymbol{X}'\boldsymbol{X}\hat{\beta}_{ols} - \lambda \boldsymbol{A}') &= \boldsymbol{a} \\
\Longrightarrow \; \lambda &= \left(\boldsymbol{A}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{A}'\right)^{-1}(\boldsymbol{A}\hat{\beta}_{ols} - \boldsymbol{a}) \qquad (4)
\end{aligned}
$$

Plug (4) into (3) we have

$$
\begin{aligned}
\hat{\beta}_{Rols} &= \hat{\beta}_{ols} - \left(\boldsymbol{A}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{A}'\right)^{-1}(\boldsymbol{A}\hat{\beta}_{ols} - \boldsymbol{a})(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{A}' \\
\hat{\beta}_{ols} - \hat{\beta}_{Rols} &= \left(\boldsymbol{A}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{A}'\right)^{-1}(\boldsymbol{A}\hat{\beta}_{ols} - \boldsymbol{a})(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{A}' \\
&= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{A}'\left(\boldsymbol{A}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{A}'\right)^{-1}(\boldsymbol{A}\hat{\beta}_{ols} - \boldsymbol{a}) \qquad (5)
\end{aligned}
$$

$$(\boldsymbol{A}\hat{\beta}_{ols} - \boldsymbol{a})' \left[\boldsymbol{A}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{A}'\right]^{-1} (\boldsymbol{A}\hat{\beta}_{ols} - \boldsymbol{a}) = (\mathrm{SSR}_R - \mathrm{SSR})$$

$$\begin{aligned}
\mathrm{SSR}_R &= (\boldsymbol{Y} - \boldsymbol{X}\hat{\beta}_{Rols})'(\boldsymbol{Y} - \boldsymbol{X}\hat{\beta}_{Rols}) \\
&= (\boldsymbol{Y} - \boldsymbol{X}\hat{\beta}_{ols} + \boldsymbol{X}\hat{\beta}_{ols} - \boldsymbol{X}\hat{\beta}_{Rols})'(\boldsymbol{Y} - \boldsymbol{X}\hat{\beta}_{ols} + \boldsymbol{X}\hat{\beta}_{ols} - \boldsymbol{X}\hat{\beta}_{Rols}) \\
&= \left(\boldsymbol{Y} - \boldsymbol{X}\hat{\beta}_{ols} + \boldsymbol{X}(\hat{\beta}_{ols} - \hat{\beta}_{Rols})\right)' \left(\boldsymbol{Y} - \boldsymbol{X}\hat{\beta}_{ols} + \boldsymbol{X}(\hat{\beta}_{ols} - \hat{\beta}_{Rols})\right) \\
&= \left(\boldsymbol{Y} - \boldsymbol{X}\hat{\beta}_{ols}\right)' \left(\boldsymbol{Y} - \boldsymbol{X}\hat{\beta}_{ols}\right) + (\hat{\beta}_{ols} - \hat{\beta}_{Rols})'\boldsymbol{X}'\boldsymbol{X}(\hat{\beta}_{ols} - \hat{\beta}_{Rols}) \\
&= \mathrm{SSR} + (\hat{\beta}_{ols} - \hat{\beta}_{Rols})'\boldsymbol{X}'\boldsymbol{X}(\hat{\beta}_{ols} - \hat{\beta}_{Rols})
\end{aligned}$$

$$
\begin{aligned}
(\mathrm{SSR}_R - \mathrm{SSR}) &= (\hat{\beta}_{ols} - \hat{\beta}_{Rols})' \boldsymbol{X}'\boldsymbol{X} (\hat{\beta}_{ols} - \hat{\beta}_{Rols}) \\
&= (\boldsymbol{A}\hat{\beta}_{ols} - \boldsymbol{a})' \left(\boldsymbol{A} \left(\boldsymbol{X}'\boldsymbol{X} \right)^{-1} \boldsymbol{A}' \right)^{-1} \boldsymbol{A} (\boldsymbol{X}'\boldsymbol{X})^{-1} \boldsymbol{X}'\boldsymbol{X} \\
&\qquad (\boldsymbol{X}'\boldsymbol{X})^{-1} \boldsymbol{A}' \left(\boldsymbol{A} \left(\boldsymbol{X}'\boldsymbol{X} \right)^{-1} \boldsymbol{A}' \right)^{-1} (\boldsymbol{A}\hat{\beta}_{ols} - \boldsymbol{a}) \\
&= (\boldsymbol{A}\hat{\beta}_{ols} - \boldsymbol{a})' \left(\boldsymbol{A} \left(\boldsymbol{X}'\boldsymbol{X} \right)^{-1} \boldsymbol{A}' \right)^{-1} \boldsymbol{A} (\boldsymbol{X}'\boldsymbol{X})^{-1} \boldsymbol{A}' \\
&\qquad \left(\boldsymbol{A} \left(\boldsymbol{X}'\boldsymbol{X} \right)^{-1} \boldsymbol{A}' \right)^{-1} (\boldsymbol{A}\hat{\beta}_{ols} - \boldsymbol{a}) \\
&= (\boldsymbol{A}\hat{\beta}_{ols} - \boldsymbol{a})' \left(\boldsymbol{A} \left(\boldsymbol{X}'\boldsymbol{X} \right)^{-1} \boldsymbol{A}' \right)^{-1} (\boldsymbol{A}\hat{\beta}_{ols} - \boldsymbol{a})
\end{aligned}
$$

Joint Statistical Inference with known σ^2 and Small n

Wald Test

Because the sample size n is small, Central Limit Theory can not be applied, the same additional assumption has to be made.

Normality Assumption

5. uncertainty u follows Normal Distribution: $\boldsymbol{u}|\boldsymbol{X} \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_n)$
   i.e $u_i|\boldsymbol{X} \sim N(0, \sigma^2)$

With σ^2 known, the Wald test statistic could be constructed same as the Large n Case. Previously, we have proved that under assumption 1-5 $(\boldsymbol{A}\hat{\beta}_{ols} - \boldsymbol{a})' \left[\boldsymbol{A}\,\sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{A}'\right]^{-1} (\boldsymbol{A}\hat{\beta}_{ols} - \boldsymbol{a}) \sim \chi^2(d)$. So in case of known σ^2 and small n, the test statistic and the associated distribution is summarized as follow :

Wald Test Statistic and Distribution

$$\text{W} = (\boldsymbol{A}\hat{\beta}_{ols} - \boldsymbol{a})' \left[\boldsymbol{A}\,\sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{A}'\right]^{-1} (\boldsymbol{A}\hat{\beta}_{ols} - \boldsymbol{a})$$
$$\text{W} \sim \chi^2(d)$$

Appendix

Properties of $\boldsymbol{M}$

Let $\boldsymbol{M} = \boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$

1. $\boldsymbol{M}$ is symmetric: $\boldsymbol{M}' = \boldsymbol{M}$
2. $\boldsymbol{M}$ is idempotent: $\boldsymbol{M}\boldsymbol{M} = \boldsymbol{M}$
3. $\boldsymbol{M}\boldsymbol{X} = \boldsymbol{0}$
4. $\text{trace}(\boldsymbol{M}) = n - p - 1$
5. The eigenvalues of $\boldsymbol{M}$ are either 0 or 1

The eigenvalues of $\boldsymbol{M}$ are either 0 or 1

Let λ be the eigenvalue and $\boldsymbol{v}$ be the associated eigenvector of $\boldsymbol{M}$

$$\begin{aligned}
\boldsymbol{M}\boldsymbol{v} &= \lambda\boldsymbol{v} \\
\boldsymbol{M}\boldsymbol{v} &= \boldsymbol{M}\boldsymbol{M}\boldsymbol{v} = \boldsymbol{M}(\lambda\boldsymbol{v}) = \lambda\boldsymbol{M}\boldsymbol{v} = \lambda^2\boldsymbol{v} \\
\therefore \lambda^2\boldsymbol{v} &= \lambda\boldsymbol{v} \\
&\Longrightarrow (\lambda^2 - \lambda)\boldsymbol{v} = \boldsymbol{0} \Longrightarrow (\lambda^2 - \lambda) = \lambda(\lambda - 1) = 0 \\
\therefore \quad \lambda &\quad \textit{should be either 0 or 1}
\end{aligned}$$

$\text{trace}(\boldsymbol{M}) = n - p - 1$

We know $\boldsymbol{X}$ is $n \times (p+1)$ matrix, $\boldsymbol{B}$ is $n \times m$ matrix

$$\begin{aligned}
\text{trace}(\boldsymbol{M}) &= \text{trace}(\boldsymbol{I}_n) - \text{trace}(\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}') \\
&= n - \text{trace}(\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}) \\
&= n - \text{trace}(\boldsymbol{I}_{p+1}) \\
&= n - p - 1
\end{aligned}$$

$\text{trace}(\boldsymbol{AB}) = \text{trace}(\boldsymbol{BA})$

Suppose $\boldsymbol{A}$ is $m \times n$ matrix, $\boldsymbol{B}$ is $n \times m$ matrix

$$\text{trace}(\boldsymbol{AB}) = \sum_{i=1}^{m}\sum_{k=1}^{n} a_{ik} b_{ki} = \sum_{k=1}^{n}\sum_{i=1}^{m} b_{ki} a_{ik} = \text{trace}(\boldsymbol{BA})$$

χ^2-Distribution

Definition

- Let $\{Z_n\}$ be a sequence of i.i.d. random variables.
- For each $Z_i \sim N(0,1)$
- Let $W = \sum_{i=1}^{n} Z_i^2$

Then the distribution for the random variable W follows χ^2-Distribution with n degree of freedom, denoted as $\chi^2(n)$

t-Distribution

Definition

- Let $Z \sim N(0,1)$
- Let $W \sim \chi^2(n)$
- Z and W are independent
- Let $T = \frac{Z}{\sqrt{W/n}}$

Then the distribution for the random variable T follows t-Distribution with n degree of freedom, denoted as $t(n)$

F-Distribution

Definition

- Let $W_1 \sim \chi^2(n_1)$
- Let $W_2 \sim \chi^2(n_2)$
- W_1 and W_2 are independent
- Let $F = \frac{W_1/n_1}{W_2/n_2}$

Then the distribution for the random variable F follows F-Distribution with degree n_1 and n_2, denoted as $F(n_1, n_2)$