The outline for Unit 3

UNIT 1. Introduction: The regression model. ✓

UNIT 2. Estimation principles. ✓

UNIT 3: Hypothesis testing principles.

3.1 Wald test.
3.2 Lagrange Multiplier.
3.3 Likelihood Ratio Test.
3.4 Comparison between LR, Wald and LM tests.

Addendum A. Hypothesis testing using bootstrap.

UNIT 4: Heteroscedasticity in the regression model.

UNIT 5: Endogeneity of regressors.

Economic restrictions

In many cases, economic theory suggests restrictions on the parameters of a model. For example, a demand function is supposed to be homogeneous of degree zero in prices and income. If we have a Cobb-Douglas model,

$$\ln q = \beta_0 + \beta_1 \ln p_1 + \beta_2 \ln p_2 + \beta_3 \ln m + \varepsilon,$$

then we need that

$$k^0 \ln q = \beta_0 + \beta_1 \ln kp_1 + \beta_2 \ln kp_2 + \beta_3 \ln km + \varepsilon,$$

so, the only way to guarantee this for arbitrary k is to set

$$\beta_1 + \beta_2 + \beta_3 = 0,$$

which is a *parameter restriction.*

Linear restrictions - 1

The general formulation of linear equality restrictions is the model

$$\begin{aligned} y &= X\beta + \varepsilon \\ R\beta &= r \end{aligned}$$

where R is a $Q \times K$ matrix, $Q < K$ and r is a $Q \times 1$ vector of constants.

- We assume R is of rank Q, so that there are no redundant restrictions.
- We also assume that $\exists \beta$ that satisfies the restrictions: they aren't infeasible.

Linear restrictions - 2

Let's consider how to estimate β subject to the restrictions $R\beta = r$. The most obvious approach is to set up the Lagrangean

$$\min_{\beta} s(\beta) = \frac{1}{n}(y - X\beta)'(y - X\beta) + 2\lambda'(R\beta - r).$$

The Lagrange multipliers are scaled by 2, which makes things less messy. The fonc are

$$\begin{aligned} D_{\beta}s(\hat{\beta}, \hat{\lambda}) &= -2X'y + 2X'X\hat{\beta}_R + 2R'\hat{\lambda} \equiv 0 \\ D_{\lambda}s(\hat{\beta}, \hat{\lambda}) &= R\hat{\beta}_R - r \equiv 0, \end{aligned}$$

which can be written as

$$\begin{bmatrix} X'X & R' \\ R & 0 \end{bmatrix} \begin{bmatrix} \hat{\beta}_R \\ \hat{\lambda} \end{bmatrix} = \begin{bmatrix} X'y \\ r \end{bmatrix}.$$

Linear restrictions - 3

We get

$$\begin{bmatrix} \hat{\beta}_R \\ \hat{\lambda} \end{bmatrix} = \begin{bmatrix} X'X & R' \\ R & 0 \end{bmatrix}^{-1} \begin{bmatrix} X'y \\ r \end{bmatrix}.$$

So,

$$\begin{aligned} \begin{bmatrix} \hat{\beta}_R \\ \hat{\lambda} \end{bmatrix} &= \begin{bmatrix} (X'X)^{-1} - (X'X)^{-1}R'P^{-1}R\,(X'X)^{-1} & (X'X)^{-1}R'P^{-1} \\ P^{-1}R\,(X'X)^{-1} & -P^{-1} \end{bmatrix} \begin{bmatrix} X'y \\ r \end{bmatrix} \\ &= \begin{bmatrix} \hat{\beta} - (X'X)^{-1}R'P^{-1}\left(R\hat{\beta} - r\right) \\ P^{-1}\left(R\hat{\beta} - r\right) \end{bmatrix} \\ &= \begin{bmatrix} \left(I_K - (X'X)^{-1}R'P^{-1}R\right) \\ P^{-1}R \end{bmatrix} \hat{\beta} + \begin{bmatrix} (X'X)^{-1}R'P^{-1}r \\ -P^{-1}r \end{bmatrix} \end{aligned}$$

where $P = R(X'X)^{-1}R'$.

Linear restrictions - 4

If the number of restrictions is small, we can impose them by substitution and write

$$
\begin{aligned}
y &= X_1\beta_1 + X_2\beta_2 + \varepsilon \\
\begin{bmatrix} R_1 & R_2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} &= r
\end{aligned}
$$

where R_1 is $Q \times Q$ nonsingular. Supposing the Q restrictions are linearly independent, one can always make R_1 nonsingular by reorganizing the columns of X. Then $\beta_1 = R_1^{-1}r - R_1^{-1}R_2\beta_2$. Substitute this into the model

$$
\begin{aligned}
y &= X_1R_1^{-1}r - X_1R_1^{-1}R_2\beta_2 + X_2\beta_2 + \varepsilon \\
y - X_1R_1^{-1}r &= \left[X_2 - X_1R_1^{-1}R_2\right]\beta_2 + \varepsilon
\end{aligned}
$$

or with the appropriate definitions, $y_R = X_R\beta_2 + \varepsilon$.

Linear restrictions - 5

This model satisfies the classical assumptions, *supposing the restriction is true*. One can estimate by OLS. The variance of $\hat{\beta}_2$ is $V(\hat{\beta}_2) = (X_R' X_R)^{-1} \sigma_0^2$ and the estimator is $\hat{V}(\hat{\beta}_2) = (X_R' X_R)^{-1} \hat{\sigma}^2$ where one estimates σ_0^2 in the normal way, using the restricted model, *i.e.*,

$$\widehat{\sigma_0^2} = \frac{\left(y_R - X_R \widehat{\beta_2}\right)' \left(y_R - X_R \widehat{\beta_2}\right)}{n - (K - Q)}$$

To recover $\hat{\beta}_1$, use the restriction. To find the variance of $\hat{\beta}_1$, use the fact that it is a linear function of $\hat{\beta}_2$, so

$$\begin{aligned} V(\hat{\beta}_1) &= R_1^{-1} R_2 V(\hat{\beta}_2) R_2' \left(R_1^{-1}\right)' \\ &= R_1^{-1} R_2 \left(X_2' X_2\right)^{-1} R_2' \left(R_1^{-1}\right)' \sigma_0^2 \end{aligned}$$

Properties of the restricted estimator - 1

We have that

$$\begin{aligned}
\hat{\beta}_R &= \hat{\beta} - (X'X)^{-1}R'P^{-1}\left(R\hat{\beta} - r\right) \\
&= \hat{\beta} + (X'X)^{-1}R'P^{-1}r - (X'X)^{-1}R'P^{-1}R(X'X)^{-1}X'y \\
&= \beta + (X'X)^{-1}X'\varepsilon + (X'X)^{-1}R'P^{-1}\left[r - R\beta\right] \\
&\quad -(X'X)^{-1}R'P^{-1}R(X'X)^{-1}X'\varepsilon \\
\hat{\beta}_R - \beta &= (X'X)^{-1}X'\varepsilon \\
&+ (X'X)^{-1}R'P^{-1}\left[r - R\beta\right] \\
&- (X'X)^{-1}R'P^{-1}R(X'X)^{-1}X'\varepsilon
\end{aligned}$$

Mean squared error is

$$MSE(\hat{\beta}_R) = \mathrm{E}(\hat{\beta}_R - \beta)(\hat{\beta}_R - \beta)'$$

Properties of the restricted estimator - 2

Noting that the crosses between the second term and the other terms expect to zero, and that the cross of the first and third has a cancellation with the square of the third, we obtain

$$\begin{aligned} MSE(\hat{\beta}_R) &= (X'X)^{-1}\sigma^2 \\ &+ (X'X)^{-1}R'P^{-1}\left[r - R\beta\right]\left[r - R\beta\right]' P^{-1}R(X'X)^{-1} \\ &- (X'X)^{-1}R'P^{-1}R(X'X)^{-1}\sigma^2 \end{aligned}$$

So, the first term is the OLS covariance and

- If the restriction is true, the second term is 0, so we are better off. *True restrictions improve efficiency of estimation.*

- If the restriction is false, we may be better or worse off, in terms of MSE, depending on the magnitudes of $r - R\beta$ and σ^2.

Testing: t-test

Suppose one has the model

$$y = X\beta + \varepsilon$$

and one wishes to test the *single restriction* $H_0 : R\beta = r$ vs. $H_A : R\beta \neq r$.

Under H_0, with normality of the errors,

$$R\hat{\beta} - r \sim \mathcal{N}\left(0, R(X'X)^{-1}R'\sigma_0^2\right)$$

so

$$\frac{R\hat{\beta} - r}{\sqrt{R(X'X)^{-1}R'\sigma_0^2}} = \frac{R\hat{\beta} - r}{\sigma_0\sqrt{R(X'X)^{-1}R'}} \sim \mathcal{N}\left(0, 1\right).$$

The problem is that σ_0^2 is unknown. One could use the consistent estimator $\widehat{\sigma_0^2}$ in place of σ_0^2, but the test would only be valid asymptotically in this case.

Testing: t-test - 2

Proposition 1:

$$\frac{N(0,1)}{\sqrt{\frac{\chi^2(q)}{q}}} \sim t(q) \tag{1}$$

as long as the $N(0,1)$ and the $\chi^2(q)$ are independent.

Proposition 2: If $x \sim N(\mu, I_n)$ is a vector of n independent r.v.'s., then

$$x'x \sim \chi^2(n, \lambda) \tag{2}$$

where $\lambda = \sum_i \mu_i^2 = \mu'\mu$ is the *noncentrality parameter.*

When a χ^2 r.v. has the noncentrality parameter equal to zero, it is referred to as a central χ^2 r.v., and it's distribution is written as $\chi^2(n)$, suppressing the noncentrality parameter.

Testing: t-test - 3

Proposition 3: If the n dimensional random vector $x \sim N(0, V)$, then $x'V^{-1}x \sim \chi^2(n)$.

Proof: Factor V^{-1} as PP' (this is the Cholesky factorization). Then consider $y = P'x$. We have

$$y \sim N(0, P'VP)$$

but

$$\begin{aligned} VPP' &= I_n \\ P'VPP' &= P' \end{aligned}$$

so $PVP' = I_n$ and thus $y \sim N(0, I_n)$. Thus $y'y \sim \chi^2(n)$ but

$$y'y = x'PP'x = xV^{-1}x.$$

Testing: t-test - 4

A more general proposition which implies this result is

Proposition 4: If the n dimensional random vector $x \sim N(0, V)$, then

$$x'Bx \sim \chi^2(\rho(B)) \tag{3}$$

if and only if BV is idempotent.

An immediate consequence is

Proposition 5: If the random vector (of dimension n) $x \sim N(0, I)$, and B is idempotent with rank r, then

$$x'Bx \sim \chi^2(r). \tag{4}$$

Testing: t-test - 5

Application:

$$\frac{\hat{\varepsilon}'\hat{\varepsilon}}{\sigma_0^2} = \frac{\varepsilon' M_X \varepsilon}{\sigma_0^2} = \left(\frac{\varepsilon}{\sigma_0}\right)' M_X \left(\frac{\varepsilon}{\sigma_0}\right) \sim \chi^2(n-K)$$

Proposition 6: If the random vector (of dimension n) $x \sim \mathcal{N}(0, I)$, then Ax and $x'Bx$ are independent if $AB = 0$.

Now consider (remember that we have only one restriction in this case)

$$\frac{\frac{R\hat{\beta}-r}{\sigma_0\sqrt{R(X'X)^{-1}R'}}}{\sqrt{\frac{\hat{\varepsilon}'\hat{\varepsilon}}{(n-K)\sigma_0^2}}} = \frac{R\hat{\beta}-r}{\widehat{\sigma_0}\sqrt{R(X'X)^{-1}R'}}$$

This will have $t(n-K)$ distribution if $\hat{\beta}$ and $\hat{\varepsilon}'\hat{\varepsilon}$ are independent.

Testing: t-test - 6

But $\hat{\beta} = \beta + (X'X)^{-1}X'\varepsilon$ and $(X'X)^{-1}X'M_X = 0$, so

$$\frac{R\hat{\beta} - r}{\widehat{\sigma_0}\sqrt{R(X'X)^{-1}R'}} = \frac{R\hat{\beta} - r}{\hat{\sigma}_{R\hat{\beta}}} \sim t(n-K)$$

In particular, for the commonly encountered *test of significance* of an individual coefficient, for which $H_0 : \beta_i = 0$ vs. $H_0 : \beta_i \neq 0$, the test statistic is

$$\hat{\beta}_i / \hat{\sigma}_{\hat{\beta}i} \sim t(n-K)$$

Note: the $t-$ test is strictly valid only if the errors are normally distributed. If one has nonnormal errors, one could use the above asymptotic result to justify taking critical values from the $\mathcal{N}(0,1)$ distribution, since $t(n-K) \xrightarrow{d} \mathcal{N}(0,1)$ as $n \to \infty$. In practice, a conservative procedure is to take critical values from the t distribution if nonnormality is suspected. This will reject H_0 less often since the t distribution is fatter-tailed than is the normal.

Testing: F-test

The F test allows testing multiple restrictions jointly.

Proposition 7: If $x \sim \chi^2(r)$ and $y \sim \chi^2(s)$, then

$$\frac{x/r}{y/s} \sim F(r,s) \tag{5}$$

provided that x and y are independent.

Proposition 8: If the random vector (of dimension n) $x \sim N(0,I)$, then $x'Ax$ and $x'Bx$ are independent if $AB = 0$.

Using these results, and previous results on the χ^2 distribution, it is simple to show that the following statistic has the F distribution:

$$F = \frac{\left(R\hat{\beta} - r\right)' \left(R\left(X'X\right)^{-1}R'\right)^{-1}\left(R\hat{\beta} - r\right)}{q\hat{\sigma}^2} \sim F(q, n-K).$$

Testing: F-test - 2

A numerically equivalent expression is

$$\frac{\left(ESS_R - ESS_U\right)/q}{ESS_U/(n-K)} \sim F(q, n-K).$$

Note: The F test is strictly valid only if the errors are truly normally distributed. The following tests will be appropriate when one cannot assume normally distributed errors.

Testing: Wald-test

The Wald principle is based on the idea that if a restriction is true, the unrestricted model should "approximately" satisfy the restriction. Given that the least squares estimator is asymptotically normally distributed:

$$\sqrt{n}\left(\hat{\beta}-\beta_0\right) \xrightarrow{d} N\left(0, \sigma_0^2 Q_X^{-1}\right)$$

then under $H_0 : R\beta_0 = r$, we have

$$\sqrt{n}\left(R\hat{\beta}-r\right) \xrightarrow{d} N\left(0, \sigma_0^2 R Q_X^{-1} R'\right)$$

so by Proposition 5 we have

$$n\left(R\hat{\beta}-r\right)'\left(\sigma_0^2 R Q_X^{-1} R'\right)^{-1}\left(R\hat{\beta}-r\right) \xrightarrow{d} \chi^2(q)$$

Testing: Wald-test - 2

Note that Q_X^{-1} or σ_0^2 are not observable.

The test statistic we use substitutes the consistent estimators. Use $(X'X/n)^{-1}$ as the consistent estimator of Q_X^{-1}. With this, there is a cancellation of $n's$, and the statistic to use is

$$\left(R\hat{\beta} - r\right)' \left(\widehat{\sigma_0^2} R(X'X)^{-1}R'\right)^{-1} \left(R\hat{\beta} - r\right) \xrightarrow{d} \chi^2(q)$$

- The Wald test is a simple way to test restrictions without having to estimate the restricted model.

- Note that this formula is similar to one of the formulae provided for the F test.

Testing: Score-test

In some cases, an unrestricted model may be nonlinear in the parameters, but the model is linear in the parameters under the null hypothesis. For example, the model

$$y = (X\beta)^{\gamma} + \varepsilon$$

is nonlinear in β and γ, but is linear in β under $H_0 : \gamma = 1$.

Estimation of nonlinear models is a bit more complicated, so one might prefer to have a test based upon the restricted, linear model. The *score test* is useful in this situation.

- Score-type tests are based upon the general principle that the gradient vector of the unrestricted model, evaluated at the restricted estimate, should be asymptotically normally distributed with mean zero, if the restrictions are true. The original development was for ML estimation, but the principle is valid for a wide variety of estimation methods.

Testing: Score-test - 2

We have seen that

$$\begin{aligned}\hat{\lambda} &= \left(R(X'X)^{-1}R'\right)^{-1}\left(R\hat{\beta}-r\right) \\ &= P^{-1}\left(R\hat{\beta}-r\right)\end{aligned}$$

Given that

$$\sqrt{n}\left(R\hat{\beta}-r\right) \xrightarrow{d} N\left(0, \sigma_0^2 R Q_X^{-1} R'\right)$$

under the null hypothesis,

$$\sqrt{n}\hat{\lambda} \xrightarrow{d} N\left(0, \sigma_0^2 P^{-1} R Q_X^{-1} R' P^{-1}\right)$$

or

$$\sqrt{n}\hat{\lambda} \xrightarrow{d} N\left(0, \sigma_0^2 \lim n\,(nP)^{-1} R Q_X^{-1} R' P^{-1}\right)$$

since the n's cancel and inserting the limit of a matrix of constants changes nothing.

Testing: Score-test - 3

However,

$$\begin{aligned} \lim nP &= \lim nR(X'X)^{-1}R' \\ &= \lim R\left(\frac{X'X}{n}\right)^{-1}R' \\ &= RQ_X^{-1}R' \end{aligned}$$

So there is a cancellation and we get

$$\sqrt{n}\hat{\lambda} \xrightarrow{d} N\left(0, \sigma_0^2 \lim nP^{-1}\right)$$

In this case,

$$\hat{\lambda}'\left(\frac{R(X'X)^{-1}R'}{\sigma_0^2}\right)\hat{\lambda} \xrightarrow{d} \chi^2(q)$$

since the powers of n cancel. To get a usable test statistic substitute a consistent estimator of σ_0^2.

Testing: Score-test - 4

- This makes it clear why the test is sometimes referred to as a Lagrange multiplier test. It may seem that one needs the actual Lagrange multipliers to calculate this. If we impose the restrictions by substitution, these are not available. Note that the test can be written as

$$\frac{\left(R'\hat{\lambda}\right)'(X'X)^{-1}R'\hat{\lambda}}{\sigma_0^2} \overset{d}{\to} \chi^2(q)$$

However, we can use the fonc for the restricted estimator:

$$-X'y + X'X\hat{\beta}_R + R'\hat{\lambda} = 0$$

to get that $R'\hat{\lambda} = X'(y - X\hat{\beta}_R) = X'\hat{\varepsilon}_R$. Substituting this into the above, we get $\frac{\hat{\varepsilon}'_R X(X'X)^{-1}X'\hat{\varepsilon}_R}{\sigma_0^2} \overset{d}{\to} \chi^2(q)$ but this is simply $\hat{\varepsilon}'_R \frac{P_X}{\sigma_0^2}\hat{\varepsilon}_R \overset{d}{\to} \chi^2(q)$.

Testing: Score-test - 5

To see why the test is also known as a score test, note that the fonc for restricted least squares

$$-X'y + X'X\hat{\beta}_R + R'\hat{\lambda} = 0$$

give us

$$R'\hat{\lambda} = X'y - X'X\hat{\beta}_R$$

and the rhs is simply the gradient (score) of the unrestricted model, evaluated at the restricted estimator. The scores evaluated at the unrestricted estimate are identically zero. The logic behind the score test is that the scores evaluated at the restricted estimate should be approximately zero, if the restriction is true. The test is also known as a *Rao test*, since P. Rao first proposed it in 1948.

Testing: Likelihood ratio-type tests

The Wald test can be calculated using the unrestricted model. The score test can be calculated using only the restricted model. The likelihood ratio test, on the other hand, uses both the restricted and the unrestricted estimators. The test statistic is

$$LR = 2\left(\ln L(\hat{\theta}) - \ln L(\tilde{\theta})\right)$$

where $\hat{\theta}$ is the unrestricted estimate and $\tilde{\theta}$ is the restricted estimate. To show that it is asymptotically χ^2, take a second order Taylor's series expansion of $\ln L(\tilde{\theta})$ about $\hat{\theta}$:

$$\ln L(\tilde{\theta}) \simeq \ln L(\hat{\theta}) + \frac{n}{2}\left(\tilde{\theta} - \hat{\theta}\right)' H(\hat{\theta})\left(\tilde{\theta} - \hat{\theta}\right)$$

(note, the first order term drops out since $D_\theta \ln L(\hat{\theta}) \equiv 0$ by the fonc and we need to multiply the second-order term by n since $H(\theta)$ is defined in terms of $\frac{1}{n}\ln L(\theta)$) so

$$LR \simeq -n\left(\tilde{\theta} - \hat{\theta}\right)' H(\hat{\theta})\left(\tilde{\theta} - \hat{\theta}\right)$$

Testing: Likelihood ratio-type tests - 2

As $n \to \infty, H(\hat{\theta}) \to H_\infty(\theta_0) = -\mathcal{I}(\theta_0)$, by the information matrix equality. So

$$LR \overset{a}{=} n\left(\tilde{\theta} - \hat{\theta}\right)' \mathcal{I}_\infty(\theta_0)\left(\tilde{\theta} - \hat{\theta}\right)$$

We also have that

$$\sqrt{n}\left(\hat{\theta} - \theta_0\right) \overset{a}{=} \mathcal{I}_\infty(\theta_0)^{-1} n^{1/2} g(\theta_0).$$

An analogous result for the restricted estimator is:

$$\sqrt{n}\left(\tilde{\theta} - \theta_0\right) \overset{a}{=} \mathcal{I}_\infty(\theta_0)^{-1}\left(I_n - R'\left(R\mathcal{I}_\infty(\theta_0)^{-1}R'\right)^{-1} R\mathcal{I}_\infty(\theta_0)^{-1}\right) n^{1/2} g(\theta_0).$$

Combining the last two equations

$$\sqrt{n}\left(\tilde{\theta} - \hat{\theta}\right) \overset{a}{=} -n^{1/2}\mathcal{I}_\infty(\theta_0)^{-1}R'\left(R\mathcal{I}_\infty(\theta_0)^{-1}R'\right)^{-1} R\mathcal{I}_\infty(\theta_0)^{-1} g(\theta_0)$$

Testing: Likelihood ratio-type tests - 3

So,

$$LR \overset{a}{=} \left[n^{1/2}g(\theta_0)'\mathcal{I}_\infty(\theta_0)^{-1}R'\right]\left[R\mathcal{I}_\infty(\theta_0)^{-1}R'\right]^{-1}\left[R\mathcal{I}_\infty(\theta_0)^{-1}n^{1/2}g(\theta_0)\right]$$

But since

$$n^{1/2}g(\theta_0) \overset{d}{\rightarrow} N\left(0, \mathcal{I}_\infty(\theta_0)\right)$$

the linear function

$$R\mathcal{I}_\infty(\theta_0)^{-1}n^{1/2}g(\theta_0) \overset{d}{\rightarrow} N(0, R\mathcal{I}_\infty(\theta_0)^{-1}R').$$

We can see that LR is a quadratic form of this rv, with the inverse of its variance in the middle, so

$$LR \overset{d}{\rightarrow} \chi^2(q).$$

The asymptotic equivalence of LR, Wald and score tests

We have seen that the three tests all converge to χ^2 random variables. In fact, they all converge to the *same* χ^2 rv, under the null hypothesis.

We'll show that the Wald and LR tests are asymptotically equivalent. We have seen that the Wald test is asymptotically equivalent to

$$W \stackrel{a}{=} n\left(R\hat{\beta} - r\right)'\left(\sigma_0^2 R Q_X^{-1} R'\right)^{-1}\left(R\hat{\beta} - r\right) \xrightarrow{d} \chi^2(q)$$

Using

$$\hat{\beta} - \beta_0 = (X'X)^{-1}X'\varepsilon$$

and

$$R\hat{\beta} - r = R(\hat{\beta} - \beta_0)$$

The asymptotic equivalence of LR, Wald and score tests - 2

We get

$$\sqrt{n}R(\hat{\beta}-\beta_0) \quad = \quad \sqrt{n}R(X'X)^{-1}X'\varepsilon \; = R\left(\frac{X'X}{n}\right)^{-1} n^{-1/2}X'\varepsilon$$

Substitute this into Wald statistics we get

$$\begin{aligned} W & \stackrel{a}{=} & n^{-1}\varepsilon' X Q_X^{-1} R' \left(\sigma_0^2 R Q_X^{-1} R'\right)^{-1} R Q_X^{-1} X'\varepsilon \\ & \stackrel{a}{=} & \varepsilon' X(X'X)^{-1}R' \left(\sigma_0^2 R(X'X)^{-1}R'\right)^{-1} R(X'X)^{-1}X'\varepsilon \\ & \stackrel{a}{=} & \frac{\varepsilon' A(A'A)^{-1}A'\varepsilon}{\sigma_0^2} \stackrel{a}{=} \frac{\varepsilon' P_R \varepsilon}{\sigma_0^2} \end{aligned}$$

where P_R is the projection matrix formed by the matrix $X(X'X)^{-1}R'$.

- Note that this matrix is idempotent and has q columns, so the projection matrix has rank q.

The asymptotic equivalence of LR, Wald and score tests - 3

Now consider the likelihood ratio statistic

$$LR \stackrel{a}{=} n^{1/2} g(\theta_0)' \mathcal{I}(\theta_0)^{-1} R' \left(R \mathcal{I}(\theta_0)^{-1} R' \right)^{-1} R \mathcal{I}(\theta_0)^{-1} n^{1/2} g(\theta_0)$$

Under normality, we have seen that the likelihood function is

$$\ln L(\beta, \sigma) = -n \ln \sqrt{2\pi} - n \ln \sigma - \frac{1}{2} \frac{(y - X\beta)' (y - X\beta)}{\sigma^2}.$$

Using this,

$$\begin{aligned} g(\beta_0) &\equiv D_\beta \frac{1}{n} \ln L(\beta, \sigma) \\ &= \frac{X'(y - X\beta_0)}{n\sigma^2} \\ &= \frac{X'\varepsilon}{n\sigma^2} \end{aligned}$$

The asymptotic equivalence of LR, Wald and score tests - 4

Also, by the information matrix equality:

$$\begin{aligned}
\mathcal{I}(\theta_0) &= -H_\infty(\theta_0) \\
&= \lim -D_{\beta'} g(\beta_0) \\
&= \lim -D_{\beta'} \frac{X'(y - X\beta_0)}{n\sigma^2} \\
&= \lim \frac{X'X}{n\sigma^2} \\
&= \frac{Q_X}{\sigma^2}
\end{aligned}$$

so

$$\mathcal{I}(\theta_0)^{-1} = \sigma^2 Q_X^{-1}$$

The asymptotic equivalence of LR, Wald and score tests - 5

Substituting these last expressions into LR test statistics, we get

$$\begin{aligned} LR &\stackrel{a}{=} \varepsilon' X'(X'X)^{-1}R' \left(\sigma_0^2 R(X'X)^{-1}R'\right)^{-1} R(X'X)^{-1}X'\varepsilon \\ &\stackrel{a}{=} \frac{\varepsilon' P_R \varepsilon}{\sigma_0^2} \\ &\stackrel{a}{=} W \end{aligned}$$

This completes the proof that the Wald and LR tests are asymptotically equivalent. Similarly, one can show that, *under the null hypothesis*,

$$qF \stackrel{a}{=} W \stackrel{a}{=} LM \stackrel{a}{=} LR$$

The asymptotic equivalence of LR, Wald and score tests - 6

- The proof for the statistics except for LR does not depend upon normality of the errors, as can be verified by examining the expressions for the statistics.
- The LR statistic *is* based upon distributional assumptions, since one can't write the likelihood function without them.
- However, due to the close relationship between the statistics qF and LR, supposing normality, the qF statistic can be thought of as a *pseudo-LR statistic,* in that it's like a LR statistic in that it uses the value of the objective functions of the restricted and unrestricted models, but it doesn't require distributional assumptions.
- The presentation of the score and Wald tests has been done in the context of the linear model. This is readily generalizable to nonlinear models and/or other estimation methods.

Testing nonlinear restrictions

Testing nonlinear restrictions of a linear model is not much more difficult, at least when the model is linear. Since estimation subject to nonlinear restrictions requires nonlinear estimation methods, which are beyond the score of this course, we'll just consider the Wald test for nonlinear restrictions on a linear model.

Consider the q nonlinear restrictions

$$r(\beta_0) = 0.$$

where $r(\cdot)$ is a q-vector valued function. Write the derivative of the restriction evaluated at β as

$$D_{\beta'} r(\beta)\Big|_{\beta} = R(\beta)$$

Testing nonlinear restrictions - 2

We suppose that the restrictions are not redundant in a neighborhood of β_0, so that

$$\rho(R(\beta)) = q$$

in a neighborhood of β_0. Take a first order Taylor's series expansion of $r(\hat{\beta})$ about β_0:

$$r(\hat{\beta}) = r(\beta_0) + R(\beta^*)(\hat{\beta} - \beta_0)$$

where β^* is a convex combination of $\hat{\beta}$ and β_0. Under the null hypothesis we have

$$r(\hat{\beta}) = R(\beta^*)(\hat{\beta} - \beta_0)$$

Due to consistency of $\hat{\beta}$ we can replace β^* by β_0, asymptotically, so

$$\sqrt{n}r(\hat{\beta}) \stackrel{a}{=} \sqrt{n}R(\beta_0)(\hat{\beta} - \beta_0)$$

We've already seen the distribution of $\sqrt{n}(\hat{\beta} - \beta_0)$. Using this we get

$$\sqrt{n}r(\hat{\beta}) \xrightarrow{d} N\left(0, R(\beta_0)Q_X^{-1}R(\beta_0)'\sigma_0^2\right).$$

Testing nonlinear restrictions - 3

Considering the quadratic form

$$\frac{nr(\hat{\beta})'\left(R(\beta_0)Q_X^{-1}R(\beta_0)'\right)^{-1}r(\hat{\beta})}{\sigma_0^2} \xrightarrow{d} \chi^2(q)$$

under the null hypothesis. Substituting consistent estimators for $\beta_{0,}Q_X$ and σ_0^2, the resulting statistic is

$$\frac{r(\hat{\beta})'\left(R(\hat{\beta})(X'X)^{-1}R(\hat{\beta})'\right)^{-1}r(\hat{\beta})}{\widehat{\sigma^2}} \xrightarrow{d} \chi^2(q).$$

- This is known in the literature as the *Delta method*, or as *Klein's approximation*.

- Since this is a Wald test, it will tend to over-reject in finite samples. The score and LR tests are also possibilities, but they require estimation methods for nonlinear models, which aren't in the scope of this course.

Empirical comparison of LM, Wald and score tests

Example 1: Lets assume that the DGP satisfy a

$$\ln C = \beta_1 + \beta_2 \ln Q + \beta_3 \ln P_L + \beta_4 \ln P_F + \beta_5 \ln P_K + \epsilon$$

where the variables: are COST (C), OUTPUT (Q), PRICE OF LABOR (P_L), PRICE OF FUEL (P_F) and PRICE OF CAPITAL (P_K).

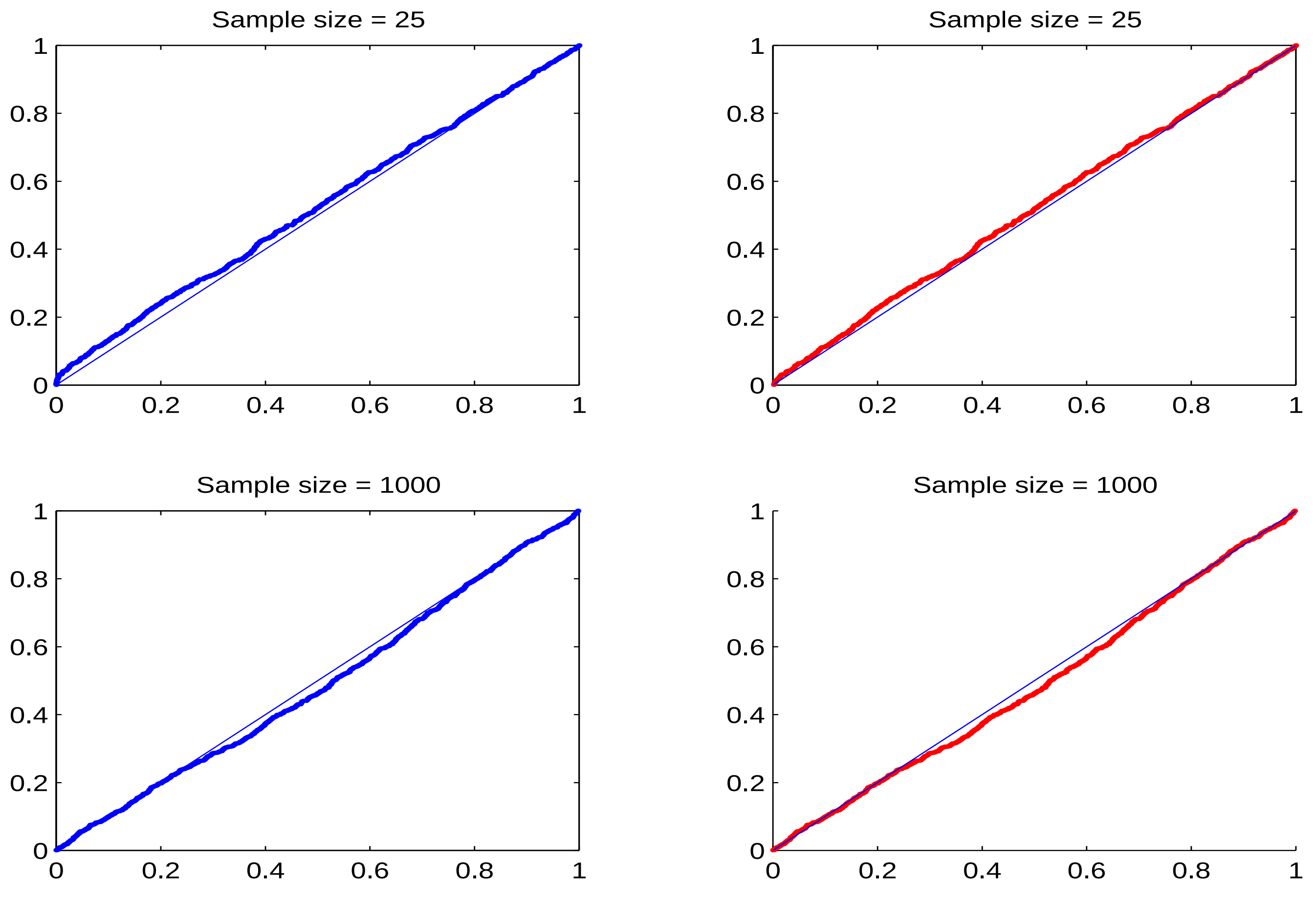
The following restriction are imposed:

- It verify the property of HOD1, i.e., $\sum_{i=3}^{5} \beta_i = 1$.
- It verify the property of CRTS technology, i.e., $\gamma = \frac{1}{\beta_q} = 1$

Compare the Wald and score tests.

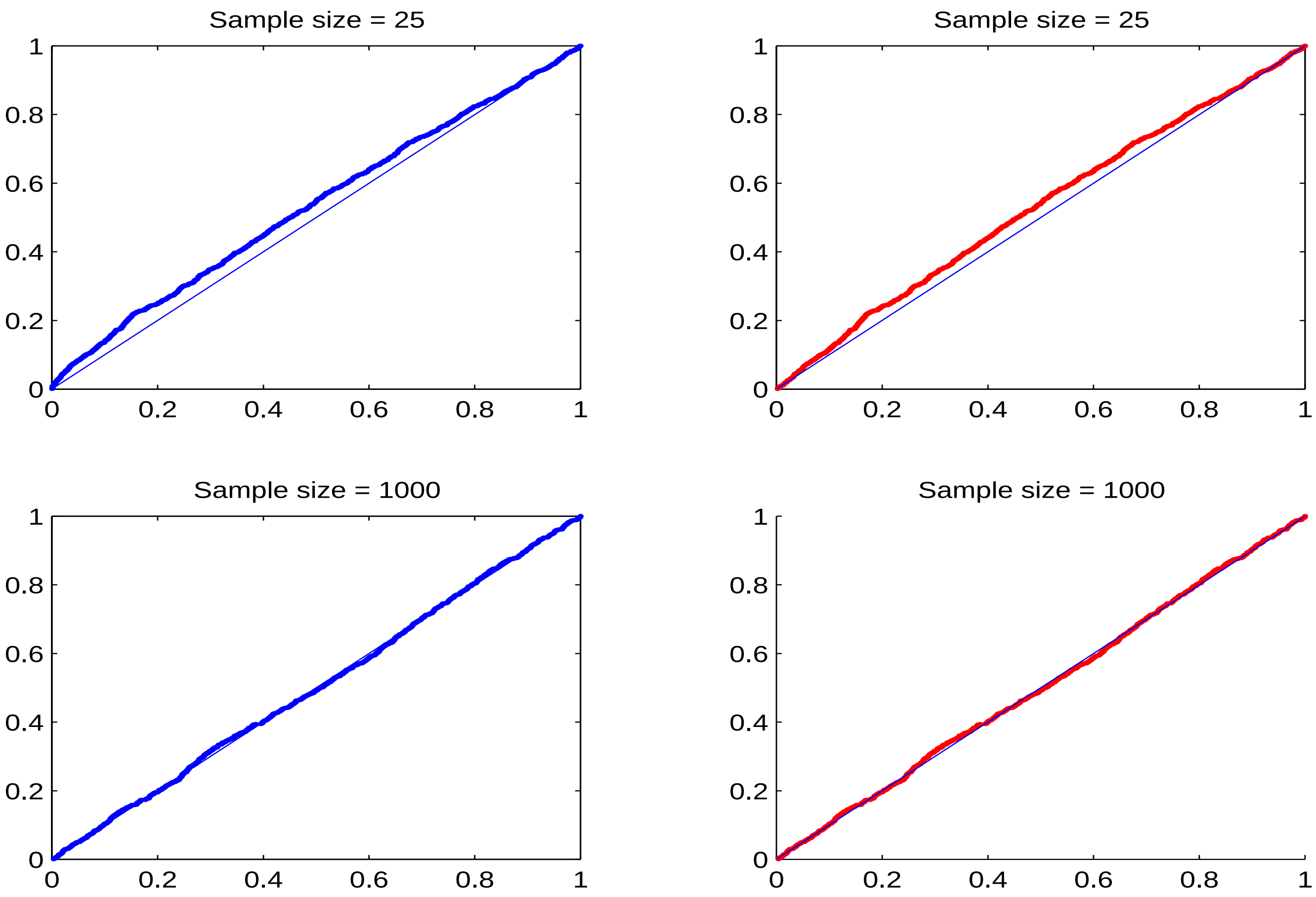
Empirical comparison of LM, Wald and score tests

Example 1 (cont.): HOD1.



Empirical comparison of LM, Wald and score tests

Example 1 (cont.): CRTS.



Empirical comparison of LM, Wald and score tests

Example 1 (cont.): Estimated sizes.