Econometrics I, Testing

Department of Economics
Stanford University

September, 2008

Part I
- Null hypothesis $H_0$. Alternative hypothesis $H_A$.
- Sample $\mathbf{x} = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$.
- Rejection region: $R \in \mathbb{R}^n$ such that $H_0$ is rejected if $\mathbf{x} \in R$.
- Simple hypothesis; composite hypothesis.
  - Simple $H_0$ vs simple $H_A$;
  - Simple $H_0$ vs Composite $H_A$;
  - Composite $H_0$ vs Composite $H_A$. 
Principle of testing (of finding $R$):

- Find a scalar statistic $t(x)$ such that under the null $H_0$:
  \[ t(x) \xrightarrow{p} 0 \]
  while under the alternative $H_A$:
  \[ t(x) \xrightarrow{p} c > 0. \]

- Derive (asymptotic or exact) distribution of $t(x)$ under $H_0$:
  \[
  \text{Under } H_0: \quad P(a_n t(x) \leq x) \xrightarrow{A} F(x)
  \]
  where $a_n \to \infty$ when $n \to \infty$

- Reject $H_0$ if $a_n t(x)$ is larger than $\alpha$th critical value of $F(x)$.

- Under $H_A$: $a_n t(x) \to \infty$. 
Trinity of tests

- Likelihood ratio test.
- Need to estimate model under both $H_0$ and $H_A$.
- Wald test.
- No need to estimate under $H_0$.
- Score function test.
- Only need to estimate under $H_0$. 
• Type I error: error of rejecting $H_0$ when it is true.
• Size: probability of type I error (under $H_0$ by definition).
  \[ P\left(a_n t(x) \geq F^{-1}(1 - \alpha)\right) \approx \alpha \quad \text{when } H_0 \text{ is true.} \]
• Type II error: error of accepting $H_0$ when $H_A$ is true.
• Power: probability of rejecting $H_0$ when $H_A$ is true.
• Therefore, power = 1 - $P$(type II error) = 1 - $\beta$.
• Power of asymptotic test is usually 1:
  \[ P\left(a_n t(x) \geq F^{-1}(1 - \alpha)\right) \rightarrow 1 \quad \text{when } H_A \text{ is true.} \]
• Consistent test. Local asymptotic power.
• Definition 9.2.2: Let \((\alpha_1, \beta_1)\) and \((\alpha_2, \beta_2)\) be the characteristics of two tests. The first test is better (more powerful) than the second test if \(\alpha_1 \leq \alpha_2\) and \(\beta_1 \leq \beta_2\) with a strict inequality holding for at least one of the \(\leq\).

• Definition 9.2.4: \(R\) is the most powerful test of size \(\alpha\) if \(\alpha(R) = \alpha\) and for any test \(R_1\) of size \(\alpha\), \(\beta(R) \leq \beta(R_1)\). (It may not be unique.)

• Definition 9.5.2: \(R\) is the most powerful test of level \(\alpha\) if \(\alpha(R) \leq \alpha\) and for any test \(R_1\) of level \(\alpha\) (that is, such that \(\alpha(R_1) \leq \alpha\), \(\beta(R) \leq \beta(R_1)\).

• “level \(\alpha\)” is needed with discrete sample when \(\alpha(R) \neq \alpha\) for all \(R\). Not needed with randomized test.

• Randomized test: toss a coin before deciding which of \(R_1\) and \(R_2\) to use. Can achieve any desired \(\alpha\).
• Bayes testing: loss matrix,

<table>
<thead>
<tr>
<th>Decision</th>
<th>State of Nature</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_0$</td>
<td>$0$</td>
</tr>
<tr>
<td>$H_1$</td>
<td>$\gamma_1$</td>
</tr>
</tbody>
</table>

• Bayesian expected loss

$$
\phi(R) = \underbrace{\gamma_1 \pi(H_0)}_{\eta_0} P(x \in R|H_0) + \underbrace{\gamma_2 \pi(H_1)}_{\eta_1} P(x \in R^c|H_1) = \eta_0 \alpha(R) + \eta_1 \beta(R)
$$

$$
= \eta_0 \int_{x \in R} f(x|H_0) \, dx + \eta_1 \int_{x \in R^c} f(x|H_1) \, dx.
$$
Optimal $R$ that minimizes $\phi (R)$:

$$R_0 = \{ x : \eta_0 f (x | H_0) \leq \eta_1 f (x | H_1) \} = \{ x : \frac{L (x | H_1)}{L (x | H_0)} > \frac{\eta_0}{\eta_1} \}$$

$$= \{ x : \log L (x | H_1) - \log L (x | H_0) > \log \left( \frac{\eta_0}{\eta_1} \right) \}.$$

Likelihood ratio test minimizes Bayes risk.

is also most powerful.

$\phi (R)$ is a linear combination of size and type II error.

No $R_1$ s.t. $\alpha (R_1) = \alpha (R_0)$ and $\beta (R_1) < \beta (R_0)$.

Otherwise $\phi (R_1) < \phi (R_0)$.

Frequentist: pick $\eta_0/\eta_1$ for the desired size.
• Example. $H_0 : \mu = \mu_0$, $H_A : \mu = \mu_1$. Assume $\mu_1 > \mu_0$.
• $\{X_t, t = 1, \ldots \}$ i.i.d. $N(\mu, \sigma^2)$. $\sigma^2$ known.
• Log likelihood ratio:
  
  $$-\frac{1}{2\sigma^2} \sum (x_t - \mu_1)^2 + \frac{1}{2\sigma^2} \sum (x_t - \mu_0)^2$$

  $$= \frac{n(\mu_1 - \mu_0)}{\sigma^2} \bar{x} + c$$

• Best rejection region: $\bar{x} > d$ such that

  $$P(\bar{x} > d|H_0) = \alpha.$$ 

• Note that $d$ does not depend on the value of $\mu_1$, as long as

  $$\mu_1 > \mu_0.$$
• Simple Null vs Composite Alternative.

• \( H_0 : \theta = \theta_0 \) against \( H_1 : \theta \in \Theta_1 \).

• Power function for a test \( R \)

\[
Q(\theta) = P(x \in R|\theta).
\]

• \( Q(\theta_0) = \alpha. \) \( Q(\theta_1) = 1 - \beta \) for \( \theta_1 \in \Theta_1 \).

• Definition 9.4.2. \( R_1 \) is uniformly better than \( R_2 \) if \( Q_1(\theta_0) = Q_2(\theta_0), \) \( Q_1(\theta) \geq Q_2(\theta) \forall \theta \in \Theta_1, \) and \( Q_1(\theta_1) > Q_2(\theta_1) \) for at least one \( \theta_1 \in \Theta_1 \).

• Definition 9.4.3. A test \( R \) is uniformly most powerful (UMP) if it is uniformly better than any other test with the same size (the same \( Q(\theta_0) \)).
• UMP tests often do not exist.
• Likelihood ratio test usually is UMP if UMP exists.
• Likelihood ratio test often used even without UMP.
• LR test: reject if
  \[\log L(\theta_0) - \sup_{\theta \in \theta_0 \cup \Theta_1} \log L(\theta) < c.\]
• The second part is just the maximum likelihood estimator:
  \[\log L(\hat{\theta}) = \sup_{\theta \in \theta_0 \cup \Theta_1} \log L(\theta).\]
• Example. $H_0 : \mu = \mu_0$, $H_A : \mu > \mu_0$.

• $\{X_t \sim N(\mu, \sigma^2)\}$ with $\sigma^2$ known.

• Log likelihood ratio statistics:

$$LR = -\frac{1}{2\sigma^2} \sum (x_t - \mu_0)^2 - \sup_{\mu \geq \mu_0} -\frac{1}{2\sigma^2} \sum (x_t - \mu)^2$$

$$= -\frac{1}{2\sigma^2} n (\bar{x} - \mu_0)^2 + \inf_{\mu \geq \mu_0} \frac{1}{2\sigma^2} n (\bar{x} - \mu)^2.$$

• If $\bar{x} \leq \mu_0$, then $\inf_{\mu \geq \mu_0} \frac{1}{2\sigma^2} n (\bar{x} - \mu)^2 = \frac{1}{2\sigma^2} n (\bar{x} - \mu_0)^2$. $LR = 0$.

• Do not reject.

• If $\bar{x} > \mu_0$, $\inf = 0$, $LR = -\frac{1}{2\sigma^2} n (\bar{x} - \mu_0)^2$, reject if $\bar{x} > d$.

• This test is UMP.
• Example: $H_0 : \mu = \mu_0$, $H_A : \mu \neq \mu_0$.

• $\{X_t \sim N(\mu, \sigma^2)\}$ with $\sigma^2$ known.

• log likelihood ratio statistics:

$$LR = -\frac{1}{2\sigma^2} \sum (x_t - \mu_0)^2 - \sup_{\mu \neq \mu_0} \frac{1}{2\sigma^2} \sum (x_t - \mu)^2$$

$$= -\frac{1}{2\sigma^2} n (\bar{x} - \mu_0)^2 + \inf_{\mu \neq \mu_0} \frac{1}{2\sigma^2} n (\bar{x} - \mu)^2$$

$$= -\frac{1}{2\sigma^2} n (\bar{x} - \mu_0)^2.$$

• Reject if $|\bar{x} - \mu_0| > d$ where

$$P (|\bar{x} - \mu_0| > d | H_0) = \alpha.$$

• Not a UMP test. Not Neyman-Pearson against, e.g. any $\mu > \mu_0$. But, UMP among tests that assign equal power to $\mu$ equidistant from $\mu_0$. 
• Asymptotic LR test: $H_0 : \theta = \theta_0$, $H_A : \theta \neq \theta_0$.

$$LR = \log L (\theta_0) - \log L \left( \hat{\theta}_{MLE} \right).$$

• One can prove that under $H_0$,

$$2LR \xrightarrow{d} -\chi^2_{\text{dim}(\theta)}.$$

• Intuitively,

$$2LR \approx \sqrt{n} \left( \hat{\theta}_{MLE} - \theta_0 \right)' \frac{1}{n} \frac{\partial^2 \log L(\theta)}{\partial \theta \partial \theta'} \sqrt{n} \left( \hat{\theta}_{MLE} - \theta_0 \right).$$

• Not UMP.

• Not longer $\chi^2$ limit for multivariate inequality test.
• Wald tests: $H_0 : \theta = \theta_0$. $H_A : \theta \neq \theta_0$.

• Typically $\sqrt{n} \left( \hat{\theta} - \theta \right) \xrightarrow{d} N(0, \Sigma)$.

• Intend to reject if $\hat{\theta} - \theta_0$ is sufficient large.

• What metric to use to measure distance $|\hat{\theta} - \theta_0|$?

• Use quadratic norm:

$$a_n \mathbf{t}(\mathbf{x}) = n \left( \hat{\theta} - \theta_0 \right)' \hat{\Sigma}^{-1} \left( \hat{\theta} - \theta_0 \right) \xrightarrow{d} \chi^2_{\text{dim}(\theta)}.$$
• Wald test for linear combinations:

• $H_0 : A\theta = A\theta_0$, $H_A : A\theta \neq A\theta_0$.

• If $\sqrt{n} \left( \hat{\theta} - \theta \right) \xrightarrow{d} N(0, \Sigma)$, then

$$\sqrt{n} A \left( \hat{\theta} - \theta \right) \xrightarrow{d} N(0, A\Sigma A')$$

• Quadratic norm based test statistic

$$n \left( \hat{\theta} - \theta_0 \right)' A' \left( A\hat{\Sigma} A' \right)^{-1} A \left( \hat{\theta} - \theta_0 \right) \xrightarrow{d} \chi^2_{\text{rows of } A}$$

• Not UMP.
• Asymptotic power.
• Under $H_A$, typically $a_n t(x) \xrightarrow{\infty} w.p. \to 1$.
• Under $H_A$, reject w.p. $\to 1$.
• Asymptotic power is 1.
• These are called consistent tests.
• Local alternatives and local power.
• Local alternative, $H_A : \theta = \theta_0 + \frac{c}{\sqrt{n}}$. 
• Example: $H_0 : \mu = \mu_0, \ H_A : \mu > \mu_0$.

• $\{X_t \sim N (\mu, \sigma^2)\}$ with $\sigma^2$ known.

• Reject if $\bar{x} > d$: $d = \frac{\sigma z_\alpha}{\sqrt{n}} + \mu_0$ for size $\alpha$ test.

• Asymptotic power: for fixed $\mu_1 > \mu_0$,

\[
P (\bar{x} > d | \mu_1) = P \left( \sqrt{n} \frac{\bar{x} - \mu_1}{\sigma} \geq z_\alpha + \sqrt{n} \frac{\mu_0 - \mu_1}{\sigma} \right) \to 1.
\]

• Local power, for $\mu_1 = \mu_0 + \frac{c}{\sqrt{n}}$,

\[
P (\bar{x} > d | \mu_1) = P \left( \sqrt{n} \frac{\bar{x} - \mu_1}{\sigma} \geq z_\alpha - \frac{c}{\sigma} \right) 
\to 1 - \Phi \left( z_\alpha - \frac{c}{\sigma} \right) > \alpha.
\]