Testing Parametric Models

Robert A. Miller

Structural Econometrics

November 2021

Introduction

Some intuition

- Let H_0 denote a *null hypothesis*.
- Let H_A denote the *alternative hypothesis*, the complement of H_0.
- Evidence we collect against the null might lead us to *reject* H_0.
- Lacking evidence to the contrary, we might continue to believe H_0 is true, or *fail to reject* H_0.
- We can commit two types of mistakes:
  1. A *Type 1 error* occurs (a false positive) when we falsely reject H_0. The *size* of the test, or the probability of landing in the *critical region*, gives the probability of committing a Type 1 error.
  2. A *Type 2 error* occurs (a false negative) when we do not reject a false H_0. The *power* of the test is the probability of rejecting H_0 when H_A is true and that probability is well defined.
- Given a sample, we might minimize the expected loss from making either of these two mistakes, and balance it against the benefits of collecting more information.

Introduction

A formalism

- The notation used in estimation readily adapts to hypothesis testing:
  - Ω is a population, or the set of all possible histories/orderings.
  - $\omega^* \in \Omega$ is one ordering of the population, or a specific history.
  - $\mathcal{F}_N$ is a $\sigma-$algebra induced on Ω with a sample of $N = 1, 2, \ldots$
- Given this framework we might:
  - construct a test statistic, $T_N(\omega)$, an $\mathcal{F}_N-$measurable random variable.
  - fix a size, α, the probability of a Type 1 error.
  - choose a critical region, c_α, for rejecting H_0 that minimizes the probability of a Type 2 error.
  - reject H_0 iff $t_N \equiv T_N(\omega^*) \in c_\alpha$.
- The size of the test is set by convention: the notion is that only under exceptional circumstances should the null hypothesis be rejected.
- For this reason we tend to include (exclude) significant (insignificant) variables from our final regressions.

- When H_i fully specifies a probability distribution, we say H_i is *simple*.
- In this case H_i induces a probability space $(\Omega, \mathcal{F}_N, P_i)$.
- More generally suppose Θ_i denotes class of probability distributions for $i = \{0, A\}$ containing elements $\theta_i \in \Theta_i$, and define:
  - $H_0 : \theta \in \Theta_0$
  - $H_A : \theta \in \Theta_A$
- When Θ_i contains more than one element we say H_i is *composite*.
- For example if $X_n(\omega)$ is *iid* normal with mean μ and variance σ^2:
  - $\Theta_0 = \{(\mu, \sigma) : \mu = 2, \sigma = 2\}$ and $\Theta_A = \{(\mu, \sigma) : \mu = 2, \sigma = 5\}$ pits two simple hypotheses against each other.
  - $\Theta_0 = \{(\mu, \sigma) : \mu = 2, \sigma = 2\}$ and $\Theta_A = \{(\mu, \sigma) : \mu = 0, \sigma > 0\}$ pits a simple hypothesis H_0 against a composite hypothesis H_A.

Testing Simple Hypotheses

A framework

- Suppose H_i is simple for $i = \{0, A\}$.
- Let P_i denote the probability distribution characterizing H_i.
- We can now define:

$$\text{the size of the test by } \; P_0 \left\{ t_N \in c_\alpha \right\} = \alpha \; . \tag{1}$$

$$\text{the power of the test by } P_A \left\{ t_N \in c_\alpha \right\} . \tag{2}$$

- Choose c_α to maximize (2) subject to (1).
- That is for a given tolerance of rejecting the null hypothesis (such as incorrectly diagnosing a healthy patient as sick), we maximize the probability of rejecting a false null hypothesis (correctly diagnosing a sick patient).

- To characterize the *best* test for two simple hypotheses, suppose:
  - $X(\omega) : \Omega \rightarrow \mathbb{R}$ is the random variable generating the sample.
  - $f_i(\cdot)$ denotes the *pdf* derived from P_i of observing x given H_i.
  - $\Lambda(x) \equiv f_A(x) / f_0(x)$ is the likelihood ratio for x of H_A relative to H_0.
  - x_N is the sample outcome and $\Lambda_N \equiv f_A(x_N) / f_0(x_N)$.

Lemma (Neyman-Pearson)

The best (most powerful) test of H_0 against H_A of size α sets $c_\alpha \equiv [\Lambda_\alpha, \infty]$ for $\Lambda_\alpha \in \mathbb{R}^+$ solving:

$$P_0\{\Lambda(x) \geq \Lambda_\alpha\} = \alpha$$

- The lemma gives a straightforward procedure for conducting the test:
  - Form $F_\Lambda(\cdot)$, the cumulative distribution function of Λ under H_0.
  - Solve $F_\Lambda(\Lambda_\alpha) = \alpha$ to obtain Λ_α.
  - Reject H_0 iff $t_N \equiv \Lambda_N \geq \Lambda_\alpha$.

- Letting c'_α denote any other critical region of size α aside from c_α:

$$\begin{aligned} \alpha &= \int_{c_\alpha} f_0(x)\,dx = \int_{c'_\alpha} f_0(x)\,dx \\ &\Longrightarrow \int_{c_\alpha - c'_\alpha} f_0(x)\,dx = \int_{c'_\alpha - c_\alpha} f_0(x)\,dx \\ &\Longrightarrow \int_{c_\alpha - c'_\alpha} \Lambda_\alpha f_0(x)\,dx = \int_{c'_\alpha - c_\alpha} \Lambda_\alpha f_0(x)\,dx \end{aligned}$$

- But:

$$\begin{aligned} \Lambda(x) &\in c_\alpha - c'_\alpha \Longrightarrow \Lambda(x) \in c_\alpha \Longrightarrow f_A(x) \geq \Lambda_\alpha f_0(x) \\ \Lambda(x) &\in c'_\alpha - c_\alpha \Longrightarrow \Lambda(x) \in c_\alpha^{complement} \Longrightarrow f_A(x) < \Lambda_\alpha f_0(x) \end{aligned}$$

- Integrating over the regions $x \in c_\alpha - c'_\alpha$ and then $x \in c'_\alpha - c_\alpha$:

$$\begin{aligned}\int_{c_\alpha - c'_\alpha} f_A(x)\,dx &\geq \int_{c_\alpha - c'_\alpha} \Lambda_\alpha f_0(x)\,dx \\ &= \int_{c'_\alpha - c_\alpha} \Lambda_\alpha f_0(x)\,dx \geq \int_{c'_\alpha - c_\alpha} f_A(x)\,dx\end{aligned}$$

  with strict inequality if the regions are different.
- Adding the region $x \in c_\alpha \cap c'_\alpha$ we obtain:

$$\begin{aligned}\int_{c_\alpha} f_A(x)\,dx &= \int_{c_\alpha \cap c'_\alpha} f_A(x)\,dx + \int_{c_\alpha - c'_\alpha} f_A(x)\,dx \\ &\geq \int_{c_\alpha \cap c'_\alpha} f_A(x)\,dx + \int_{c'_\alpha - c_\alpha} f_A(x)\,dx = \int_{c'_\alpha} f_A(x)\,dx\end{aligned}$$

- When H_A is a composite hypothesis, the critical region typically depends on which value of $\theta \in \Theta_A$ holds.
- In the example of the normal distribution described above, the solution to maximizing (2) subject to (1) depends on whether $\sigma = 1$ or $\sigma = 3$ under H_A.
- A *uniformly most powerful test* (UMP) exists when c_α does not vary with $\theta \in \Theta_A$.
- Roughly speaking, a UMP for composite hypotheses exists, when all three of the following conditions hold:
  1. only one parameter to be estimated.
  2. the likelihood ratio is monotone in that parameter.
  3. the test is one sided.
- When a UMP exists, following the Neyman-Pearson approach:
  - form the likelihood ratio for any of the alternatives in H_A.
  - compute its cumulative distribution function.
  - match the probability of c_α to the size designated by α.

Asymptotic Properties of Test Statistics

Consistent tests

- We encountered computation problems in deriving the *exact* (finite sample) distribution of estimators for nonlinear models.
- Aside from a few special cases, deriving the exact distribution of the likelihood ratio also demands a numerical approach.
- This has led applied researchers to focus on the large sample or asymptotic properties of test statistics.
- A desirable asymptotic property is that for any given size, the probability of a Type 2 error should vanish as the sample increases.
- Reverting to the notation let $H_0 : \theta \in \Theta_0$ and $H_A : \theta \in \Theta_A$: we say $T_N(\omega)$ is *consistent* when there exists some $\theta_0 \in \Theta_0$ such that as $N \to \infty$ for all $\theta \in \Theta_A$:
    - $P_0 \{t_N \in c_\alpha\} = \alpha$ (holding constant the probabilty of a false positive)
    - $P_A \{t_N \in c_\alpha\} \to 1$ (the probability of a false negative goes to zero)

Asymptotic Properties of Test Statistics

Testing a normally distributed estimator of the parameters against a null vector

- For example, suppose $\sqrt{N}\left(\theta_N - \mu\right) \sim N\left(0, \Sigma\right)$, where:
  - $\mu \in \mathbb{R}^s$
  - Σ is a known $s \times s$ positive definite matrix.
- Consider the simple null hypothesis H_0 and composite alternative H_A respectively defined by:

$$\begin{aligned} H_0 &: \mu = \theta_0 \\ H_A &: \mu \neq \theta_0 \end{aligned}$$

- Define the test statistic by:

$$t_N = N\left(\theta_N - \theta_0\right)' \Sigma^{-1}\left(\theta_N - \theta_0\right)$$

- $t_N \sim \chi^2_{df=s}$ under H_0 but under H_A the distribution is degenerate at infinity.

- This last point follows from:

$$\begin{aligned} t_N &= N\left(\theta_N - \theta_0\right)' \Sigma^{-1} \left(\theta_N - \theta_0\right) \\ &= N\left(\theta_N - \mu\right)' \Sigma^{-1} \left(\theta_N - \mu\right) + N\left(\mu - \theta_0\right)' \Sigma^{-1} \left(\mu - \theta_0\right) \end{aligned}$$

  and that $N\left(\mu - \theta_0\right)' \Sigma^{-1} \left(\mu - \theta_0\right) \to \infty$ for all $\mu \neq \theta_0$.
- Putting some (any) probability mass in the right tail of $\chi^2_{df=s}$ for any size α produces a consistent test.
- For example choose positive constants $\underline{c}_\alpha$ and $\overline{c}_\alpha$ satisfying:

$$\begin{aligned} & P_0\left\{t_N \in [0, \underline{c}_\alpha]\right\} = P_0\left\{t_N \in [\overline{c}_\alpha, \infty]\right\} = \frac{\alpha}{2} \\ \Rightarrow \quad & P\left\{t_N \in c_\alpha \,|\mu \neq \theta_0\right\} > P\left\{t_N \in [\overline{c}_\alpha, \infty] \,|\mu \neq \theta_0\right\} \to 1 \end{aligned}$$

- Following this line of enquiry would lead us to consider asymptotically most powerful unbiased tests, in which we consider local alternative hypotheses such as $H_A : \mu = \theta_0 + N^{-1/2}\delta$ for fixed $\delta \in \mathbb{R}^s$.

- When testing two simple hypotheses we showed that the likelihood ratio produces the *best* test, maximizing power subject to a given size.
- For consistent tests, pitting H_0 against H_A, is no different from testing H_0 against the composite $H_0 \cup H_A$.
- Accordingly suppose:
    - $H_A : \theta \in \Theta \subset \mathbb{R}^s_{interior}$
    - $H_0 : \theta \in \Theta_0 \subset \mathbb{R}^{s-q}_{interior}$ for some $q \in \{1, 2, \ldots, s\}$
    - $L_N(\theta)$ denotes the likelihood.
    - $\theta_i^{(N)}$ denotes the MLE for $i = \{0, A\}$ under H_i.
- Define the *likelihood ratio statistic* (LR) as:

$$t_N \equiv 2\left\{\ln L_N\left(\theta_A^{(N)}\right) - \ln L_N\left(\theta_0^{(N)}\right)\right\}$$

Theorem (Wilk's theorem)

$t_N \xrightarrow{d} \chi^2_q$.

Testing in the GMM Framework

A framework for testing GMM estimators

- $(\Omega, \mathcal{F}, P_i)$ is $\{X_n\}$ stationary and ergodic for $i \in \{0, A\}$, and:
  - $\theta^{(1)} \in \mathbb{R}^{p_1}$.
  - $\theta^{(2)} \in \mathbb{R}^{p_2}$.
  - $\theta' \equiv \left(\theta^{(1)\prime}, \theta^{(2)\prime}\right)$
- Suppose:
  - $h_1\left(x_n, \theta^{(1)}\right)$ is $q_1 \times 1$, where $q_1 \geq p_1$
  - $h_2\left(x_n, \theta^{(1)}, \theta^{(2)}\right)$ is $q_2 \times 1$, where $q_2 \geq p_2$.
- Under H_A:
  - $E\left[h_1\left(X_n, \theta_0^{(1)}\right)\right] = 0$
  - An identification condition for $\theta_0^{(2)}$ is satisfied.
- Under $H_0 \subseteq H_A$:
  - $E\left[h_1\left(X_n, \theta_0^{(1)}\right)\right] = E\left[h_2\left(X_n, \theta_0^{(1)}, \theta_0^{(2)}\right)\right] \equiv E\left[h_2\left(X_n, \theta_0\right)\right] = 0$
  - Given $\theta_0^{(1)}$ an identification criterion for $\theta_0^{(1)}$ is satisfied.

- For all $\theta^{(1)} \in \mathbb{R}^{p_1}$ and $\theta^{(2)} \in \mathbb{R}^{p_2}$ we suppose there exists:

$$
\begin{aligned}
h_{1N}\left(x_n, \theta^{(1)}\right) &= h_1\left(x_n, \theta^{(1)}\right) + o_p\left(\sqrt{N}\right) \\
h_{2N}\left(x_n, \theta\right) &= h_2\left(x_n, \theta\right) + o_p\left(\sqrt{N}\right) \\
h_N\left(\theta\right) &= \frac{1}{N}\sum_{n=1}^{N} h_N\left(x_n, \theta\right) \equiv \frac{1}{N}\sum_{n=1}^{N}\left(\begin{array}{c} h_{1N}\left(x_n, \theta^{(1)}\right) \\ h_{2N}\left(x_n, \theta\right) \end{array}\right) \\
h_{1N}\left(\theta^{(1)}\right) &\equiv \frac{1}{N}\sum_{n=1}^{N} h_{1N}\left(x_n, \theta^{(1)}\right) \\
S_N \xrightarrow{p} S_0 &= \sum_{j=-\infty}^{\infty} E\left[h\left(x_n, \theta_0\right) h\left(x_{n-j}, \theta_0\right)'\right] \\
S_N^{(1)} \xrightarrow{p} S_0^{(1)} &= \sum_{j=-\infty}^{\infty} E\left[h_1\left(x_n, \theta_0^{(1)}\right) h_1\left(x_{n-j}, \theta_0^{(1)}\right)'\right]
\end{aligned}
$$

Testing in the GMM Framework

Two estimators and a test statistic

- Under H_0 there are $q_1 + q_2 - p_1 - p_2$ overidentifying restrictions in the minimization problem:

$$J_N = \min_{\theta} \left\{ N h_N(\theta)' S_N^{-1} h_N(\theta) \right\}$$

- Under H_A there are $q_1 - p_1$ (that is $q_2 - p_2$ fewer) overidentifying restrictions in the minimization problem:

$$J_N^{(1)} = \min_{\theta^{(1)}} \left\{ N h_{1N}\left(\theta^{(1)}\right)' \left[S_N^{(1)}\right]^{-1} h_N\left(\theta^{(1)}\right) \right\}$$

- Form the test statistic:

$$t_N = J_N - J_N^{(1)}$$

Lemma

$t_N \xrightarrow{d} \chi^2_{(q_2 - p_2) - (q_1 - p_1)}$ *under* H_0.

Corollary

Let $p_1 = q_1 = 0$. *Then* $t_N \xrightarrow{d} \chi^2_{q_2 - p_2}$ *under* H_0.

Corollary

Let $h_2(x_n, \theta) = h_2(\theta)$, *a restriction on the parameter space. Then* $t_N \xrightarrow{d} \chi^2_{q_2 - p_2}$ *under* H_0. *In particular* $t_N \xrightarrow{d} \chi^2_{q_2}$ *if* $p_2 = 0$.

- Many popular tests are restrictions on the parameter space.
- These include t-tests, F-tests, and their nonlinear GMM analogues.
- Given $(\Omega, \mathcal{F}, P)$ with the $\{X_n\}_{n=1}^{N}$ stationary and ergodic (say), define the orthogonality conditions:

$$\begin{aligned} H_A &: E\left[f\left(X_n, \mu_0\right)\right] = 0 \\ H_0 &: E\left[f\left(X_n, \mu_0\right)\right] = g\left(\mu_0\right) = 0 \end{aligned}$$

- In terms of the notation expressed in the second corollary above:
  - $\theta^{(1)} = \mu$ and $\theta^{(2)} \equiv 0$
  - $h_2\left(\theta^{(1)}, 0\right) = g\left(\mu\right)$
  - $E\left[f\left(X_n, \mu_0\right)\right] = E\left[h_1\left(X_n, \theta_0^{(1)}\right)\right] = 0$ under H_A
  - H_0 imposes $q_2 - p_2$ restrictions on μ of the form $E\left[h_2\left(\theta_0^{(1)}, 0\right)\right] \equiv g\left(\mu_0\right) = 0$

- The test statistics for the null hypothesis fall into four major classes (the first three sometimes called the *trinity*):
  - *Wald statistics* are based on deviations of the unconstrained estimates from values consistent with the null.
  - *Lagrange multiplier* or *score statistics* are based on deviations of the constrained estimates from values solving the unconstrained problem.
  - *Distance metric statistics* are based on differences in the between the unconstrained and constrained GMM criterion function.
  - *Minimum chi-square statistics* are based on differences between the constrained and unconstrained parameter estimates.

Restrictions on the Parameter Space

Notation

- Let μ denote an $s \times 1$ parameter vector to be estimated and define:

$$f_N(\mu) \equiv \frac{1}{N}\sum_{n=1}^{N} f_N(x_n, \mu) \equiv \frac{1}{N}\sum_{n=1}^{N}\left[f(x_n, \mu) + o_p\left(\sqrt{N}\right)\right]$$

  - S_N, a consistent estimate of the asymptotic covariance for $\sqrt{N} f_N(\mu_0)$

$$Q_N(\mu) \equiv \frac{\partial f_N(\mu)'}{\partial \mu} S_N^{-1} \frac{\partial f_N(\mu)}{\partial \mu}$$

  - $Q_N(\mu)$, a consistent estimate of the asymptotic covariance for μ

- Now define the three estimators:

$$\begin{aligned}
\mu_{un} &= \operatorname{argmin}\left\{f_N(\mu)' S_N^{-1} f_N(\mu)\right\} \\
\mu_r &= \operatorname{argmin}\left\{f_N(\mu)' S_N^{-1} f_N(\mu) : g(\mu) = 0\right\} \\
\mu_r^* &= \operatorname{argmin}\left\{(\mu_{un} - \mu) Q_N(\mu_r)^{-1} (\mu_{un} - \mu) : g(\mu) = 0\right\}
\end{aligned}$$

1. Wald:

$$W_N = Ng\left(\mu_{un}\right)'\left(\frac{\partial g\left(\mu_{un}\right)}{\partial \mu} Q_N\left(\mu_{un}\right)^{-1}\frac{\partial g\left(\mu_{un}\right)}{\partial \mu}'\right)^{-1} g\left(\mu_{un}\right)$$

2. J-statistic:

$$\begin{aligned} t_N &= N\left\{\left[f_N\left(\mu_{un}\right)' S_N^{-1} f_N\left(\mu_{un}\right)\right] - \left[f_N\left(\mu_r\right)' S_N^{-1} f_N\left(\mu_r\right)\right]\right\} \\ &= J_N\left(\mu_{un}\right) - J_N\left(\mu_r\right) \end{aligned}$$

3. Lagrange multiplier (LM) (or 'gradient test' or 'efficient score'):

$$L_N = N\left[f_N\left(\mu_r\right)' S_N^{-1}\frac{\partial f_N\left(\mu_r\right)}{\partial \mu}\right] Q_N^{-1}\left[\frac{\partial f_N\left(\mu_r\right)'}{\partial \mu} S_N^{-1} f_N\left(\mu_r\right)\right]$$

4. Minimum chi-squared (MC):

$$c_N = N\left(\mu_{un} - \mu_r^*\right)' Q_N\left(\mu_{un}\right)^{-1}\left(\mu_{un} - \mu_r^*\right)$$

Restrictions on the Parameter Space

Exact (asymptotic) equivalence in the linear (nonlinear) model

- In the linear model the following equalities hold:

Lemma

If f is linear in μ then:

$$t_N = L_N = c_N$$

If f and g are both linear in μ then:

$$w_N = t_N = L_N = c_N$$

- More generally in nonlinear models, the statistics converge in probability to the same random variable, and hence have the same asymptotic distribution:

Lemma

$$w_N = t_N + o_p(1) = L_N + o_p(1) = c_N + o_p(1)$$

An Information Matrix Test

Motivation

- In the standard J test H_A does not restrict the data: if there are more equations than parameters, the test simply checks how well these overidentifying restrictions are satisfied.
- This begs the questions of how to test H_0 when all the orthogonality conditions are assumed under H_0.
- One important type of conditional moment test is the *information matrix* (IM) test.
- The basic idea is that if the model is:
  - correct the information matrix asymptotically equals minus the Hessian.
  - incorrect, then equality will not generally hold, because proving the information matrix equality exploits the fact that the joint density of the data is the likelihood function.
- The concept of *Quasi-Maximum Likelihood* (QML) estimation is useful for deriving the test.

An Information Matrix Test

Quasi-Maximum Likelihood

- Suppose that the observed data $\{x_n\}_{n=1}^{N}$ are *iid*:
  - with distribution function $G(x)$ under H_A
  - $G(x) = F(x, \theta_0)$ for some unknown $\theta_0 \in \Theta$ under H_0.
- Then under H_A:

$$\theta_{qml} \equiv \underset{\theta \in \Theta}{\operatorname{argmax}} \left\{ \frac{1}{N} \sum_{n=1}^{N} \ln \left[\mathrm{d}F(x_n, \theta) \right] \right\}$$

  by a WLLN and subject to some regularity conditions

$$\theta_{qml} \xrightarrow{p} \underset{\theta \in \Theta}{\operatorname{argmax}} \mathrm{E} \left\{ \ln \left[\mathrm{d}F(x_n, \theta) \right] \right\} \equiv \breve{\theta} \in \Theta$$

- Applying a CLT, Lindeberg condition, for some covariance matrix V:

$$\sqrt{N} \left(\theta_{qml} - \breve{\theta} \right) \xrightarrow{d} \mathcal{N}(0, V)$$

- Recall that we showed that

$$\sqrt{N}\left(\theta_{ml}-\theta_0\right)\xrightarrow{d}\mathcal{N}\left(0,\Sigma_0\right)$$

where

$$\Sigma_0=E\left(\frac{\partial \ln \mathrm{d}F\left(x,\theta_0\right)}{\partial\theta}\frac{\partial \ln \mathrm{d}F\left(x,\theta_0\right)}{\partial\theta}'\right).$$

- Under H_0 notice $\breve{\theta}=\theta_0$, that $\theta_{qml}=\theta_{ml}$, and that $V=\Sigma_0$.
- There is, however, no apparent reason why V should be equal to Σ_0 under H_A.

An Information Matrix Test

Testing the specification

- The aim is to test the hypothesis

$$G(x) \in \{F(x,\theta) : \theta \in \Theta\}$$

- We consider the test based on the statistic

$$t_N = \frac{1}{N}\sum_{n=1}^{N} c(x_n, \theta_{qml}),$$

where

$$\int c(x, \breve{\theta})\, dG(x) \equiv t, \qquad \text{and}$$

$$t = 0 \text{ if and only if } G(x) = F(x, \theta_0)$$

for some $\theta_0 \in \Theta$, which is the same as the null hypothesis is true.

- Let c be the outer partials of the score minus the information matrix

$$\begin{aligned} H_0 &: \text{ if the distribution is specified correctly} \\ H_A &: \text{ if not} \end{aligned}$$

Lemma

$$\sqrt{N}\begin{pmatrix} \theta_{qml} - \theta \\ t_N - t \end{pmatrix} \xrightarrow{d} \mathcal{N}\left(0, \Upsilon\Psi\Upsilon'\right)$$

where

$$\Upsilon = \begin{bmatrix} \frac{\partial^2}{\partial\theta\partial\theta'}\mathrm{E}\left(\ln \mathrm{d}F\left(x_n, \breve{\theta}\right)\right) & 0 \\ \frac{\partial}{\partial\theta'}\mathrm{E}\left(c\left(x_n, \breve{\theta}\right)\right) & -I \end{bmatrix}^{-1}$$

and

$$\Psi = \begin{bmatrix} \mathrm{E}\left[\frac{\partial \ln \mathrm{d}F(x_n,\theta)}{\partial\theta}\frac{\partial \ln \mathrm{d}F(x_n,\theta)}{\partial\theta}'\right] & \mathrm{E}\left[\frac{\partial \ln \mathrm{d}F\left(x_n,\breve{\theta}\right)}{\partial\theta}\left(c\left(x_n,\theta\right) - t\right)\right]' \\ \mathrm{E}\left[\left(c\left(x_n,\theta\right) - t\right)\frac{\partial \ln \mathrm{d}F\left(x_n,\breve{\theta}\right)}{\partial\theta}\right] & \mathrm{E}\left[\left(c\left(x_n,\theta\right) - t\right)\left(c\left(x_n,\theta\right) - t\right)'\right] \end{bmatrix}$$

An Information Matrix Test

Quasi ML

The derivatives comes from the log likelihood itself. All we are doing is applying a CLT theorem. How to we calculate this? We need

1. first derivative of likelihood
2. second derivative of the likelihood
3. expectations of these two.

- As in corollary 4, notice that $G(x)$ is not specified parametrically.
- In practice $c(x, \theta)$ should be picked/chosen to detect violates of interest.