

Hypothesis Testing

Robert A. Miller

Econometrics 2

November 2019

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.

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 σ -algebra induced on Ω with a sample of $N = 1, 2, \dots$
- Given this framework we might like to:
 - 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 $t_N(\omega) \in c_\alpha$.
- 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.

Introduction

Simple versus composite hypotheses

- 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 $x(\omega)$ 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:

$$\text{the size of the test by } P_0 \{t_N(\omega) \in c_\alpha\} = \alpha . \quad (1)$$

$$\text{the power of the test by } P_A \{t_N(\omega) \in c_\alpha\} . \quad (2)$$

- 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).

Testing Simple Hypotheses

Neyman-Pearson Lemma

- To characterize the *best* test for two simple hypotheses, suppose:
 - $x(\omega) : \Omega \rightarrow \mathbb{R}$ is the source of the sample evidence.
 - $f_i(\cdot)$ 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 $t_N \equiv \Lambda_N \geq \Lambda_\alpha$.

Testing Simple Hypotheses

Proof of Neyman-Pearson Lemma (one of two)

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

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

- But:

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

Testing Simple Hypotheses

Proof of Neyman-Pearson Lemma (two of two)

- 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:

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

Testing Composite Hypotheses

Uniformly most powerful tests

- 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:
 - $P_{\theta_0} \{t_N(\omega) \in c_\alpha\} = \alpha$
 - $P_\theta \{t_N(\omega) \in c_\alpha\} \rightarrow 1$ as $N \rightarrow \infty$ for all $\theta \in \Theta_A - \Theta_0$

Asymptotic Properties of Test Statistics

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

- For example, suppose $\sqrt{N}(\theta_N - \mu) \sim N(0, \Sigma)$, 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:

$$H_0 : \mu = \theta_0$$

$$H_A : \mu \neq \theta_0$$

- Define the test statistic by:

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

- $t_N \sim \chi_{df=s}^2$ under H_0 but under H_A the distribution is not defined.

Asymptotic Properties of Test Statistics

A consistent test in this multivariate normal example

- Putting some probability mass in the right tail of $\chi_{df=s}^2$ for any size α produces a consistent test.
- For example choose positive constants \underline{c}_α and \bar{c}_α satisfying:

$$P_0 \{t_N(\omega) \in [0, \underline{c}_\alpha]\} = P_0 \{t_N(\omega) \in [\bar{c}_\alpha, \infty]\} = \frac{\alpha}{2}$$
$$\Rightarrow P \{t_N(\omega) \in \underline{c}_\alpha | \mu \neq \theta_0\} > P \{t_N(\omega) \in [\bar{c}_\alpha, \infty] | \mu \neq \theta_0\} \rightarrow 1$$

- 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 $\delta \in \mathbb{R}^s$.

Likelihood Ratio Test

The likelihood ratio test

- 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:
 - $H_A : \theta \in \Theta \subset \mathbb{R}_{interior}^s$
 - $H_0 : \theta \in \Theta_0 \subset \mathbb{R}_{interior}^{s-q}$ for some $q \in \{1, 2, \dots, s\}$
 - $L_N(\theta)$ 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)

$$t_N \xrightarrow{d} \chi_q^2.$$

Testing in the GMM Framework

A framework of testing GMM estimators

- $(\Omega, \mathcal{F}, P_i)$ is $\{x_n\}$ stationary and ergodic for $i \in \{0, A\}$, and:
 - $\theta_{10} \in \mathbb{R}^{p_1}$
 - $\theta_{20} \in \mathbb{R}^{p_2}$.
- Suppose:
 - $h_1(x_n, \theta_{10})$ is $q_1 \times 1$, where $q_1 \geq p_1$
 - $h_2(x_n, \theta_{10}, \theta_{20})$ is $q_2 \times 1$, where $q_2 \geq p_2$.
- Under H_A :
 - $E[h_1(x_n, \theta_{10})] = 0$
 - An identification condition for θ_{10} is satisfied.
- Under $H_0 \subseteq H_A$:
 - $E[h_1(x_n, \theta_{10})] = E[h_2(x_n, \theta_{10}, \theta_{20})] = 0$
 - Given θ_{10} an identification criterion for θ_{20} is satisfied.

Testing in the GMM Framework

Notation

- For all $\theta_1 \in \mathbb{R}^{p_1}$ and $\theta_2 \in \mathbb{R}^{p_2}$ we suppose there exists:

$$h_{1N}(x_n, \theta_1) = h_1(x_n, \theta_1) + o_p(\sqrt{N})$$

$$h_{2N}(x_n, \theta_1, \theta_2) = h_2(x_n, \theta_1, \theta_2) + o_p(\sqrt{N})$$

$$h_N(\theta) = \frac{1}{N} \sum_{n=1}^N h_N(x_n, \theta) \equiv \frac{1}{N} \sum_{n=1}^N \begin{pmatrix} h_{1N}(x_n, \theta_1) \\ h_{2N}(x_n, \theta_1, \theta_2) \end{pmatrix}$$

$$h_{1N}(\theta) = \frac{1}{N} \sum_{n=1}^N h_{1N}(x_n, \theta_1)$$

$$S_N \xrightarrow{P} S_0 = \sum_{j=-\infty}^{\infty} E[h(x_n, \theta_0) h(x_{n-j}, \theta_0)']$$

$$S_N'' \xrightarrow{P} S_0'' = \sum_{j=-\infty}^{\infty} E[h_1(x_n, \theta_0) h_1(x_{n-j}, \theta_0)']$$

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:

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

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

$$J_1^{(N)} = \min_{\theta_1} \{ N h_{1N}(\theta_1)' (S_N'')^{-1} h_N(\theta_1) \}$$

- Form the test statistic:

$$t_N = J^{(N)} - J_1^{(N)}$$

Testing in the GMM Framework

The asymptotic distribution of the test statistic

Lemma

$t_N \xrightarrow{d} \chi_{q_2-p_2}^2$ under H_0 .

Corollary

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

Corollary

Let $h_2(x_n, \theta_1, \theta_2) = h_2(\theta_1, \theta_2)$. (In this case the restrictions do not depend on the data.) Then $t_N \xrightarrow{d} \chi_{q_2-p_2}^2$ under H_0 . In particular $t_N \xrightarrow{d} \chi_{q_2}^2$ if $p_2 = 0$.

Restrictions on the Parameter Space

Motivation

- Many tests undertaken in published empirical work are essentially restrictions on the parameter space.
- These include t -tests, F -tests, and their nonlinear GMM analogues (given by the corollary above).
- Given (Ω, \mathcal{F}, P) with the $\{x_n\}_{n=1}^N$ stationary and ergodic, we now define H_A as the orthogonality conditions:

$$E[h(x_n, \theta_0)] = 0$$

and model H_0 as the parametric restrictions:

$$g(\theta_0) = 0$$

- In terms of the previous notation and the second corollary of the previous slide, we might express $\theta_0 \equiv (\theta_{01}, \theta_{02})$ and assume $h_2(\theta_{01}, \theta_{02}) = 0$ under H_0 with $E[h_1(x_n, \theta_{10})] = 0$ under H_0 implying H_0 imposes $q_2 - p_2$ restrictions on θ_{01} .

Restrictions on the Parameter Space

Motivation

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

$$h_N(\theta) \equiv \frac{1}{N} \sum_{n=1}^N h_N(x_n, \theta) \equiv \frac{1}{N} \sum_{n=1}^N \left[h(x_n, \theta) + o_p(\sqrt{N}) \right]$$

- S_N , a consistent estimate of the asymptotic covariance for $\sqrt{N}h_N(\theta_0)$

$$Q_N(\theta) \equiv \frac{\partial h_N(\theta)'}{\partial \theta} S_N^{-1} \frac{\partial h_N(\theta)}{\partial \theta}$$

- $Q_N(\theta)$, a consistent estimate of the asymptotic covariance for θ
- Now define the three estimators:

$$\theta_{un} = \operatorname{argmin} \{ h_N(\theta)' S_N^{-1} h_N(\theta) \}$$

$$\theta_r = \operatorname{argmin} \{ h_N(\theta)' S_N^{-1} h_N(\theta) : g(\theta) = 0 \}$$

$$\theta_r^* = \operatorname{argmin} \left\{ (\theta_0 - \theta) Q_N(\theta_r)^{-1} (\theta_0 - \theta) : g(\theta) = 0 \right\}$$

Restrictions on the Parameter Space

Four test statistics

1 Wald:

$$W_N = N g(\theta_{un})' \left(\frac{\partial g(\theta_{un})}{\partial \theta} Q_N(\theta_{un})^{-1} \frac{\partial g(\theta_{un})'}{\partial \theta} \right)^{-1} g(\theta_{un})$$

2 J -statistic:

$$\begin{aligned} t_N &= N \{ [h_N(\theta_{un})' S_N^{-1} h_N(\theta_{un})] - [h_N(\theta_r)' S_N^{-1} h_N(\theta_r)] \} \\ &= J_N(\theta_{un}) - J_N(\theta_r) \end{aligned}$$

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

$$L_N = N \left[h_N(\theta_r)' S_N^{-1} \frac{\partial h_N(\theta_r)}{\partial \theta} \right] Q_N^{-1} \left[\frac{\partial h_N(\theta_r)'}{\partial \theta} S_N^{-1} h_N(\theta_r) \right]$$

4 Minimum chi-squared (MC):

$$c_N = N (\theta_{un} - \theta_r^*)' Q_N(\theta_{un})^{-1} (\theta_{un} - \theta_r^*)$$

Restrictions on the Parameter Space

Exact equivalence in the linear model

- From our previous lemmas on testing in the GMM framework:

$$t_N \xrightarrow{d} \chi_{q_2 - p_2}^2$$

where $q_2 - p_2$ is the number of restrictions implied by $g(\theta_0) = h(\theta_{01}, \theta_{02}) = 0$.

- In the linear model the following equalities hold:

Lemma

If h is linear in θ then $t_N = L_N = c_N$.

Lemma

If g and h are linear in θ then $w_N = t_N$.

Restrictions on the Parameter Space

Asymptotic equivalence in nonlinear models

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

Lemma

$$\begin{aligned}t_N &= w_N + o_p(1) \\ &= L_N + o_p(1) \\ &= c_N + o_p(1)\end{aligned}$$

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 :

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

by a WLLN and subject to some regularity conditions

$$\theta_{qml} \xrightarrow{P} \operatorname{argmax}_{\theta \in \Theta} \mathbb{E} \{ \ln [dF(x_n, \theta)] \} \equiv \check{\theta} \in \Theta$$

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

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

An Information Matrix Test

Intuition for IM test

- Recall that we showed that

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

where

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

- Under H_0 notice $\check{\theta} = \theta_0$, 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 .