Confirmation of parametric hypotheses

Pascal Lavergne, Simon Fraser University

September 2007

Preliminary and incomplete – Comments welcome

Abstract

Econometricians commonly want to confirm some precise hypothesis, such as a consequence of economic theory, an economic hypothesis, or a econometric modeling assumption. In practice, researchers routinely use significance tests that can only fail to reject the hypothesis of interest. Generalizing previous work by Romano (2005), I propose a new theoretical framework for confirmation of multivariate restrictions in parametric models and I construct tests that are asymptotically unbiased, minimax, and uniformly most powerful invariant.

Keywords: Hypothesis testing, Parametric methods.

Address correspondence to: Pascal Lavergne, Dept. of Economics, Simon Fraser University, 8888 University Drive, Burnaby BC, V5A 1S6 CANADA. Email: pascal_lavergne@sfu.ca

If by the truth of Newtonian mechanics we mean that it is approximately true in some appropriate well defined sense we could obtain strong evidence that it is true; but if we mean by its truth that it is exactly true then it has already been refuted. I.J. Good (1981)

1 Introduction

Most econometric tests can be viewed as significance tests that aim at rejecting sharp restrictions on an econometric model. Such tests are routinely used to assess the relevance of explanatory variables in econometric models. As noted by Andrews (1989), "most practitioners realize that just because a test fails to reject a hypothesis one cannot claim to accept it. Nevertheless, it is common for this to be ignored, since the practitioner is often in a position where he would like the outcome of the test to provide useful inferences ..." Moreover, a rather common objective in econometric modeling is to confirm that a precise hypothesis holds. For instance, practitioners often want to test whether a particular parametric model is correctly specified. As Nakamura, Nakamura and Duleep (1990) point out, " in the case of the test of significance ... the researcher is typically hoping that the null hypothesis of insignificance will be rejected. ... On the other hand, in the case of a specification error test, a researcher often hopes the null will be *accepted* ..." Though researchers routinely use significance tests for specification checks, see Godfrey (1988) for a review, such tests cannot confirm that the model is correctly specified and thus fail to provide an adequate answer. Specification testing is by no means an atypical situation, and there are many instances where we would like to confirm an econometric hypothesis. Such an hypothesis can be a consequence of economic theory, for instance homogeneity in prices and income in demand analysis, as implied by consumer rationality; an economic hypothesis, such as purchasing power parity; or a key assumption to estimate a structural model, such as overidentifying restrictions.

The traditional significance testing framework makes it impossible to accept the null hypothesis because a failure to reject it may be the result of low statistical power. Andrews (1989) proposes to use the inverse power function as an aid to interpretation of nonsignificant outcomes and provides theoretical justifications. More broadly, a recent literature in applied statistics advocates the use of post-experiment power calculations, though some criticize this approach, see Hoenig and Heisey (2001) and the references therein. Indeed, when our aim is not to reject a null hypothesis but to confirm it, we would rather control the power of the test than its level. This statistical issue yields a methodological one, since we are never in a position to confirm the hypothesis of interest. From a scientific perspective, we can only have weak evidence to support our model or theory, in the sense that "the data cannot reject" it.

Significance testing has been criticized by some authors, on these and other grounds, see e.g. Leamer (1983). Many throw the baby with the water and recommend a Bayesian approach instead. The goal of this paper is to develop a completely classical approach to confirmation of parametric hypotheses in completely parametric models. The considered alternative hypothesis is that the norm of some function of the parameters is close to zero. The hypothesis depends on the sample size, and specifically concentrates around zero as the sample size increases, taking into account that as information increases we can be more precise. I then propose a confirmation test for this hypothesis and show that this test had several desirable properties. Specifically, it is asymptotically unbiased, minimax, and most powerful against zero. Moreover, it is asymptotically uniformly most powerful among tests invariant to orthogonal transformations (rotations and rotoinversions) of the restrictions. Our test is based on the usual Wald statistic for testing restrictions on parameters. Roughly speaking, the confirmation test is like a significance test in reverse. However, the null and alternative hypotheses are different and the critical value is not determined under the assumption that the restriction holds. Finally a key difference with a significance test is that, because the hypothesis of interest varies with the sample size, the power of a confirmation test never attains one. In nature, the test is "tough" with the alternative hypothesis to be confirmed that corresponds to our model or theory, as it does not confirm a true alternative hypothesis with some probability that is bounded away from zero. The paper is organized as follows. In Section 2, I define the hypotheses of interest and relate these hypotheses to previous work. In Section 3, I derive the confirmation test and its properties. Section 4 is devoted to the practice of the test.

2 The testing problem

2.1 Background

The aim of a significance test is to test a sharp null hypothesis about a parameter θ. For a scalar parameter, the hypotheses of interest would typically be of the kind

$$H_0 : \ \theta = 0 \qquad \text{against} \qquad H_1 : \ \theta \neq 0 \ .$$

If we aim at proving that θ is indeed different from zero, the significance test will give us a proper answer. If however we would like to show that θ is zero, a significance test will either reject or fail to reject this hypothesis, but it can never be used to confirm it. How can we cope with this problem? One possible answer comes from other domains of applied statistics, namely from biometrics. Equivalence tests somewhat reverse the role of the considered hypotheses compared to a classical significance test, and thus allow to control the probability of a false confirmation. From an historical perspective, equivalence tests were triggered by the introduction of special regulations by the Food and Drug Administration in the U.S., as well as drug regulation authorities of many industrial countries, which require a positive proof that a generic drug had an equivalent bio-effect to the primary manufacturer's formulation, see e.g. FDA (2001). Wellek's monograph (2003) reviews most of the relevant literature. An equivalence test considers the null and alternative hypotheses

$$H : \ |\theta| \geq \delta \qquad \text{against} \qquad K : \ |\theta| < \delta \ .$$

Typically, θ is the expected difference in the bio-effect of two drugs to be compared and the threshold C is set by the drug regulation authority. An equivalence test then aims to confirm the equivalence hypothesis H_1 and controls for the probability of a false confirmation through the level of the test.

A different framework has been considered by Romano (2005), see also Borovkov (1998) and Janssen (2000). Since the above equivalence testing problem becomes degenerate when information accumulates, in the sense that the values C and 0 can be almost perfectly distinguished as the sample size increases, Romano proposes to let the null and alternative hypothesis change with the sample size, and specifically considers as the alternative hypothesis to be confirmed an interval that concentrates around 0 as the sample size increases, that is,

$$H_n: \ |\theta| \geq \delta/\sqrt{n} \qquad \text{against} \qquad K: \ |\theta| < \delta/\sqrt{n} \ .$$

Such a framework is suitable for econometric applications. It reflects that for most economic purposes "genuinely interesting hypotheses are neighborhoods, not points. No parameter is exactly equal to zero; many may be so close that we can act as if they were zero," as argued by Leamer (1988), see also McCloskey (2001). This also formalizes the fact, as sample size and information increase, the neighborhood we are interested in becomes narrower, and the hypothesis of interest more precise. In what follows, I generalize Romano's framework to propose confirmation tests of general parametric hypotheses.

2.2 Testing framework

I now introduce the basic setup considered in the remaining of this paper. We observe a sample X of independent observations from a probability distribution P_θ that belongs to a parametric family $\{P_\theta : \theta \in \Theta\}$. Under the usual assumptions, see below, the maximum likelihood estimator $\widehat{\theta}$ exists, is $\sqrt{n}$-consistent for the unknown true value θ^* and is efficient. Its asymptotic $\sqrt{n}$-variance is $I^{-1}(\theta^*)$, the inverse of the information matrix at θ^*. We are interested in $r(\theta)$, where $r(\cdot)$ is a continuous function of $\mathbb{R}^p$ to $\mathbb{R}^r$ and the matrix $R(\theta) = \nabla_\theta r(\theta)$ is of full rank in a neighborhood of θ^*, where ∇_θ denotes differentiation with respect to θ. The hypotheses of interest are defined as

$$H_n: \ \|r(\theta^*)\| \geq \delta/\sqrt{n} \qquad \text{against} \qquad K_n: \ \|r(\theta^*)\| < \delta/\sqrt{n} \ ,$$

where we define

$$\|r(\theta)\|^2 = r'(\theta)\left(R'(\theta)I^{-1}(\theta)R(\theta)\right)^{-1} r(\theta) \ .$$

That is, the norm of the vector $r(\theta)$ is defined through the asymptotic $\sqrt{n}$-variance of its efficient estimator. This norm has several advantages. First, it scales the parameter so that deviations from zero are made comparable across the different components. Second, it renders the norm unitless, which is useful when considering functions of parameters with possibly different units. Third, it has the interpretation of a signal-to-noise ratio. Fourth, as we will see below, it yields test statistics that are routinely computed by statistical and econometric softwares.

This formulation naturally generalizes Romano's (2005) framework, who considers a unique restriction on parameters that can assume negative as well as positive values, so that his results does not apply in our framework. In the particular case of linear regression models, our testing framework appears naturally from a discrimination approach, see Lavergne (1998). A test of H_n against K_n will control for the probability of a false confirmation through the level of the test. The threshold δ can be given an economic interpretation beyond its statistical role, and can be chosen depending on the particular application.

In general, the hypotheses H_n and K_n involve all the parameters of the model, even if the restrictions of interest $r(\theta^*)$ are only on a subset of parameters. Also these hypotheses are invariant to orthogonal transformations (rotations and rotoinversions) of the restrictions, in the sense that only the norm of $r(\theta^*)$ is of interest. Specifically, if A is an orthogonal matrix, then

$$\|Ar(\theta)\|^2 = r'(\theta)A'\left(AR'(\theta)I^{-1}(\theta)R(\theta)A'\right)^{-1} Ar(\theta) = \|r(\theta)\|^2 \ .$$

Hence, invariance should be understood not with respect to the parameters, but with respect to the restrictions. It is thus natural to search for a test that is also invariant in that sense.

3 Theory and methods

Given that $R(\theta)$ is of full rank in a neighborhood of θ^*, we can always reparameterize the problem, at least locally, so that the first r components of the parameter corresponds to $r(\theta)$. Hence we consider the hypotheses

$$H_n : \ \|\theta_1^*\| \geq \delta/\sqrt{n} \qquad \text{against} \qquad K_n : \ \|\theta_1^*\| < \delta/\sqrt{n} \ ,$$

where $\theta^* = (\theta_1^*, \theta_2^*)$. From the properties of maximum likelihood, $\sqrt{n}\widehat{\theta}_1$ is asymptotically distributed as $N\left(\sqrt{n}\theta_1^*, I^{11}\left(\theta^*\right)\right)$, where $I^{11}(\theta^*)$ is the block of $I^{-1}(\theta^*)$ corresponding to θ_1^*. Provided the information matrix is known, the hypotheses reduce to restrictions on the mean $\gamma_1^* = \sqrt{n}\theta_1^*$ of the distribution of $\sqrt{n}\widehat{\theta}_1$, i.e., the alternative hypothesis is

$$\gamma_1^{*\prime} \left(I^{11}(\theta^*)\right)^{-1} \gamma_1^* < \delta^2 \ .$$

Our testing problem is then asymptotically equivalent to testing a restriction on the norm of the mean vector of a normal distribution in a sample of size one. Such asymptotic equivalence is often used in statistics to simplify the problem under study, see Lehman and Romano (2005). In what follows, we follow a similar approach and rely on results by Borovkov (1998) to derive some optimal tests of H_n against K_n. Specifically, we will search for tests that are minimax and uniformly most powerful invariant.

3.1 Assumptions and definitions

We begin by introducing the basic assumption that will be maintained throughout our analysis. This assumption is standard and ensures that the maximum likelihood estimator has the usual properties, see e.g. Borovkov (1998).

Assumption A *(a) (Identification)* $P_{\theta_1} \neq P_{\theta_2}$ *for* $\theta_1 \neq \theta_2$ *(b) The distributions* P_θ *have a density function* $f_\theta(\cdot)$ *with respect to a common dominating measure* ν*. (c) The set* Θ *is compact. (d) The densities* $\sqrt{f_\theta}(\cdot)$ *are continuously differentiable in* θ *almost everywhere,*

$$I(\theta) \equiv \mathbb{E}_\theta \left[\nabla_\theta l(x,\theta) \nabla'_\theta l(x,\theta)\right]$$

exists, is continuous and positive definite. (d) The function $l(\cdot,\theta) = \log f_\theta(\cdot)$ is twice continuously differentiable in θ almost everywhere. There exists a function $l(x)$ such that $\|\nabla^2_{\theta,\theta'} l(x,\theta)\| < l(x)$ and $\mathbb{E}_\theta l(X) < \infty$ uniformly in θ.

To derive tests with desirable properties, we need to introduce some definitions. But first we recast our hypotheses in a more general framework. Suppose we want to test $H_n = \{\theta^* \in \Theta_{1n}\}$ against $K_n = \{\theta^* \in \Theta_{2n}\}$, where

$$\Theta_{in} = \Gamma_i n^{-1/2} \qquad i = 1, 2,$$

and Γ_i is independent of n. This means that $\theta^* \in \Theta_{in}$ iff $\theta^* = \gamma n^{-1/2}$ for some $\gamma \in \Gamma_i$. Specifically, our simplified testing framework sets $\Gamma_1 = \{\theta : \|\theta\| \geq \delta\}$ and $\Gamma_2 = \{\theta : \|\theta\| \leq \delta\}$. But in what follows we will also use other definitions that makes the two hypotheses disjoints, for instance $\Gamma_2 = \{\theta : \|\theta\| \leq \delta'\}$, with $\delta' < \delta$. Finally we introduce the further assumption that the sets Γ_i are bounded. This allows us to abstract from technical complications that are not essential to our problem. From a practical perspective, it makes sense to a priori exclude the possibility that θ_1 is infinitely away from the alternative hypothesis.

Definition 1 *A test π belongs to $\mathcal{K}^\alpha$, the class of tests of asymptotic level α if*

$$\lim_{n\to\infty} \sup_{\theta\in\Theta_{1n}} \mathbb{E}_\theta \pi(X) \leq \alpha \ .$$

Suppose that a priori distributions Π_i are given on Γ_i, which induce on Θ_{in} some distributions denoted by Q_{in}, $i = 1, 2$. We can then define a Bayes test for testing $H_{Q_{1n}}$ against $H_{Q_{2n}}$, where $H_{Q_{in}}$ is the hypothesis where θ^* is chosen randomly with a priori distribution Q_{in}. The class Bayes tests of level α is denoted by $\mathcal{K}^\alpha_{Q_{1n}}$.

Definition 2 *A test $\pi^* \in \mathcal{K}^\alpha_{Q_{1n}}$ is asymptotically Bayes in $\mathcal{K}^\alpha_{Q_{1n}}$ if for any other test $\pi \in \mathcal{K}^\alpha_{Q_{1n}}$*

$$\liminf_{n\to\infty} \left(\mathbb{E}_{Q_{2n}} \pi^*(X) - \mathbb{E}_{Q_{2n}} \pi(X)\right) \geq 0 \ .$$

Definition 3 *A test* $\pi^* \in \mathcal{K}^\alpha$ *is asymptotically minimax in* $\mathcal{K}^\alpha$ *if for any other test* $\pi \in \mathcal{K}^\alpha$

$$\liminf_{n\to\infty} \left(\inf_{\theta\in\Theta_{2n}} \mathbb{E}_\theta \pi^*(X) - \inf_{\theta\in\Theta_{2n}} \mathbb{E}_\theta \pi(X) \right) \geq 0 \ .$$

In what follows, invariance refers to invariance with respect to the restrictions, as explained above.

Definition 4 *Assume the* Γ_i *is invariant to orthogonal transformations,* $i = 1, 2$. *A test* $\pi^* \in \mathcal{K}^\alpha$ *is asymptotically uniformly most powerful invariant if for any invariant* $\pi \in \mathcal{K}^\alpha$

$$\liminf_{n\to\infty} \left(\mathbb{E}_\theta \pi^*(X) - \mathbb{E}_\theta \pi(X) \right) \geq 0 \qquad \forall \theta \in \Theta_{2n} \ .$$

Definition 5 *Two tests* π_1 *and* π_2 *are asymptotically equivalent if*

$$\lim_{n\to\infty} \sup_{\theta\in\Theta_{1n}\cup\Theta_{2n}} \left| \mathbb{E}_\theta \pi_1(X) - \mathbb{E}_\theta \pi_2(X) \right| = 0 \ .$$

It follows readily that if π_1 has some of the properties in Definitions 1-4, the asymptotic equivalent test π_2 has the same property, see Borovkov (1998, Section 54.2).

3.2 Reduction of the Problem

To formalize our heuristics, we define two testing problems.

Problem A: We have a sample x_n of size n from P_θ and we want to test $H_n = \{\theta \in \Theta_{1n}\}$ against $K_n = \{\theta \in \Theta_{2n}\}$, where $\Theta_{in} = \Gamma_i n^{-1/2}$, $i = 1, 2$, and Γ_i is independent of n. We denote by $\widehat{\theta}$ the maximum likelihood estimator of θ, which is asymptotically $N\left(\theta^*, n^{-1} I^{-1}(\theta^*)\right)$.

Problem B: We have a sample x of size 1 with distribution $N\left(\gamma, I^{-1}\right)$ and we want to test $H = \{\gamma \in \Gamma_i\}$ against $K = \{\gamma \in \Gamma_2\}$. We denote by $\pi_{\Pi_1\Pi_2}$ the Bayes test of level α for Problem B with a priori distributions Π_i, $i = 1, 2$.

Lemma 3.1 *Suppose that Assumption A is valid in a neighborhood of* θ^* *and let* π *a test of level* α *for Problem B.*

1. *If $\pi(x)$ is unbiased of level α in Problem B, $\pi(\sqrt{n}\widehat{\theta})$ is asymptotically unbiased in $\mathcal{K}_\alpha$ in Problem A.*

2. *If $\pi(x)$ is minimax of level α against K, and thus Bayes for a priori distributions Π_1 and Π_2, with*

$$\begin{aligned}\mathbb{E}_{\Pi_1}\pi(x) &= \sup_{\gamma\in\Gamma_1}\mathbb{E}_\gamma\pi(x)\ , \\ \mathbb{E}_{\Pi_2}\pi(x) &= \inf_{\gamma\in\Gamma_2}\mathbb{E}_\gamma\pi(x)\ ,\end{aligned}$$

   $\pi(\sqrt{n}\widehat{\theta})$ is asymptotically minimax in $\mathcal{K}_\alpha$ against K_n.

3. *If $\pi(x)$ is most powerful of level α against some point in K, $\pi(\sqrt{n}\widehat{\theta})$ is most powerful in $\mathcal{K}_\alpha$ against the corresponding point in K_n.*

4. *Assume the Γ_i is invariant to orthogonal transformations, $i = 1, 2$. If $\pi(x)$ is uniformly most powerful invariant of level α against K, $\pi(\sqrt{n}\widehat{\theta})$ is asymptotically uniformly most powerful invariant in $\mathcal{K}_\alpha$ against K_n.*

Proof.

1. This is a consequence of the definition of asymptotic equivalence.

2. See Theorem 54.1 of Borovkov (1998).

3. See Theorem 54.2 of Borovkov (1998).

4. This is a consequence of Theorem 54.3 of Borovkvov (1998).

5. Since $\pi(x)$ is uniformly most powerful invariant of level α against K, it is Bayes for invariant a priori distributions Π_1 and Π_2 and independent of the invariant a priori Π_2. Let Q_{in}, $i = 1, 2$, be the a priori distributions induced by Π_i. For any other test $\pi' \in \mathcal{K}_\alpha$,

$$\limsup_{n\to\infty}\mathbb{E}_{Q_{1n}}\pi'(x_n) = \lim_{n\to\infty}\sup_{\theta\in\Theta_{1n}}\pi'(x_n) \le \alpha\ ,$$

so that π' is an asymptotically α-level test for the hypotheses induced by Q_{in}, $i = 1, 2$. Now, as π is the Bayes α-level test for these hypotheses, for any $\theta \in \Theta_{2n}$, there is an invariant Q_{2n} such that

$$\lim_{n\to\infty} \mathbb{E}_\theta \pi(x_n) = \lim_{n\to\infty} \mathbb{E}_{Q_{2n}} \pi(x_n) \geq \lim_{n\to\infty} \mathbb{E}_{Q_{2n}} \pi'(x_n) = \lim_{n\to\infty} \mathbb{E}_\theta \pi'(x_n) ,$$

and the result follows. ■

From the above result, it is sufficient to characterize the properties of tests for Problem B. To simplify the exposition, we first treat the following problem. Suppose $X \in \mathbb{R}^p$ is multivariate normal $N(\mu, \Sigma)$ with unknown mean μ and known nonsingular covariance matrix Σ. Consider testing

$$H \ : \mu'\Sigma^{-1}\mu \geq \delta^2 \qquad \text{against} \qquad K \ : \mu'\Sigma^{-1}\mu < \delta^2 ,$$

from one observation x from X. We define the test $\pi(x)$ as the one that rejects H when

$$x'\Sigma^{-1}x < C_\alpha ,$$

where C_α is such that $\Pr[Z < C_\alpha] = \alpha$, with Z distributed as a non-central chi-square with r degrees of freedom and non-centrality parameter δ^2.

Lemma 3.2 *The test $\pi(x)$ is of level α. Moreover,*

- *it is unbiased.*
- *it is minimax among α-level tests against $\mu'\Sigma^{-1}\mu = \gamma^2 < \delta^2$ with guaranteed power $\Pr[W < C_\alpha]$, W distributed as a non-central chi-square with r degrees of freedom and non-centrality parameter γ^2.*
- *it is most powerful among α-level tests against $\mu = 0$.*
- *it is uniformly most powerful among α-level tests invariant to orthogonal transformations.*

Proof. We first consider the case $\Sigma = I$ and aim to determine a minimax test of H against

$$K(\gamma) \ : \ \mu'\Sigma^{-1}\mu \leq \gamma^2 \ .$$

for $\gamma^2 < \delta^2$. A minimax test of H against $K(\gamma)$ is Bayes under least favorable a priori distributions. Since the testing problem is invariant under orthogonal transformations, that is any transformation Ax where $A'A = I$, these distributions should also be invariant. Moreover, they should be concentrated on the boundary of the hypotheses. Therefore Q_δ, the uniform distribution on the hypersphere $S(\delta)$ of radius δ, and Q_γ, defined similarly, are the least favorable a priori distributions. The most powerful Bayes test $\pi(x)$ of level α rejects H iff

$$\int_{S(\gamma)} \exp\left[-\frac{1}{2}(x-\mu)'(x-\mu)\right] dQ_\gamma(\mu) > c \int_{S(\delta)} \exp\left[-\frac{1}{2}(x-\mu)'(x-\mu)\right] dQ_\delta(\mu)$$

for some constant c. The left-hand side term writes

$$\exp\left[-\frac{1}{2}(x'x+\gamma^2)\right] \int_{S(\gamma)} \exp\left[x'\mu\right] dQ_\gamma(\mu) \ .$$

Denoting $e_x = x/\|x\|$, the above integral equals

$$\psi\left(\gamma\|x\|\right) = \int_{S(1)} \exp\left[\gamma|x|e_x\mu\right] dQ_1(\mu) = \int_{S(1)} \exp\left[\gamma|x|\mu_1\right] dQ_1(\mu) \ ,$$

by a change of variables, where μ_1 is the first component of μ. The rejection region is

$$\exp\left[-\gamma^2/2\right] \psi\left(\gamma\|x\|\right) > c \exp\left[-\delta^2/2\right] \psi\left(\delta\|x\|\right) \ .$$

Since the function $\psi(\cdot)$ is convex with $\psi(0) = 1$ and $\psi'(0) = 0$, the test writes

$$\|x\|^2 < C \ ,$$

for some constant C. To obtain a Bayes test of level α under Q_δ, we should choose C as in the Theorem.

Let us check that this test is minimax of level α. We have

$$\mathbb{E}_\mu \pi(X) = \mathbb{P}\left[\|X\|^2 < C\right] = \mathbb{P}\left[\chi^2_p(\|\mu\|^2) < C\right] \ .$$

As this probability is decreasing in $\|\mu\|$ for each C,

$$\mathbb{E}_\mu \pi(X) = \mathbb{P}\left[\chi^2_p(\|\mu\|^2) < C\right] \leq \mathbb{P}\left[\chi^2_p(\delta^2) < C\right] \qquad \text{for } \|\mu\|^2 \geq \delta^2$$
$$\mathbb{E}_\mu \pi(X) = \mathbb{P}\left[\chi^2_p(\|\mu\|^2) < C\right] \geq \mathbb{P}\left[\chi^2_p(\gamma^2) < C\right] \qquad \text{for } \|\mu\| \leq \gamma \ ,$$

which yields

$$\sup_{\mu\in H} \mathbb{E}_\mu \pi(X) = \mathbb{E}_\mu \pi(X) \quad \forall \mu \in Q_\delta \quad \text{and} \qquad \inf_{\mu \in K(\gamma)} \mathbb{E}_\mu \pi(X) = \mathbb{E}_\mu \pi(X) \quad \forall \mu \in Q_\gamma \ .$$

Our minimax test is unbiased, most powerful for testing H against $K(\gamma)$ under Q_δ and Q_γ, and since it is independent of γ, it is the most powerful test of H against $K(0)$. Moreover, the test is invariant to any orthogonal transformation, and as any invariant test have constant power on spheres $S(\gamma)$, our minimax test is UMP invariant.

For a general Σ, define $y = \Sigma^{-1/2}x$. The distribution of y is $N\left(\varepsilon = \Sigma^{-1/2}\mu, Id\right)$ and our hypotheses are equivalent to

$$H \ : \varepsilon'\varepsilon \geq \delta^2 \qquad \text{against } K \ : \varepsilon'\varepsilon < \delta^2 \ .$$

Apply the above reasoning to y to obtain the desired result. ■

We now treat the more general case. Suppose $X \in \mathbb{R}^{r+q}$ is multivariate normal $N(\mu, \Sigma)$ with unknown mean $\mu = (\mu_1, \mu_2)$ and known nonsingular covariance matrix Σ. Define Σ_{11} as the upper block of Σ corresponding to μ_1. Consider testing

$$H \ : \mu_1' \left(\Sigma_{11}\right)^{-1} \mu_1 \geq \delta^2 \qquad \text{against} \qquad K \ : \mu_1' \left(\Sigma_{11}\right)^{-1} \mu_1 < \delta^2 \ ,$$

from one observation x from X. We define the test $\pi(x)$ as the one that rejects H when

$$x_1' \left(\Sigma_{11}\right)^{-1} x_1 < C_\alpha \ ,$$

where C_α is such that $\Pr\left[Z < C_\alpha\right] = \alpha$ with Z distributed as a non-central chi-square with r degrees of freedom and non-centrality parameter δ^2.

Lemma 3.3 *The test $\pi(x)$ is of level α. Moreover,*

- *it is unbiased.*

- *it is minimax among α-level tests against $\mu_1' (\Sigma_{11})^{-1} \mu_1 = \gamma^2 < \delta^2$ with guaranteed power* $\Pr[W < C_\alpha]$*, W distributed as a non-central chi-square with r degrees of freedom and non-centrality parameter γ^2.*

- *it is most powerful among α-level tests against $\mu = 0$.*

- *it is uniformly most powerful among α-level tests invariant to orthogonal transformations on $\mathbb{R}^r$.*

Proof. Consider $\Sigma = I$. Following the same reasoning as in Lemma 3.1's proof, the minimax test is the Bayes test with a priori uniform distributions on the hyperspheres $S(\delta) = \{\mu : \|\mu_1\| = \delta, \mu_2 = \mu_2^0\}$, where μ_2^0 is an arbitrary fixed point, and $S(\gamma)$. The minimax test of level α is

$$\|x_1\|^2 < C_\alpha$$

and enjoys the same property as the test of Lemma 3.2.

For a general Σ, consider the Cholesky decomposition $\Sigma^{-1} = \Lambda'\Lambda$, where Λ is an invertible, upper triangular matrix, and $y = \Lambda x$. The distribution of y is $N(\varepsilon = \Lambda\mu, I)$, where ε_1 depends only on μ_1. Moreover, denoting by Λ_{11} the corresponding submatrix of Λ, $\Lambda_{11}'\Lambda_{11} = (\Sigma_{11})^{-1}$, since Λ is upper triangular. Hence our hypotheses are equivalent to

$$H \ : \varepsilon_1'\varepsilon_1 \geq \delta^2 \qquad \text{against} \qquad K \ : \varepsilon_1'\varepsilon_1 < \delta^2 \ .$$

Apply the above reasoning to y to obtain the result. ■

3.3 Main result

Theorem 3.4 *Suppose $X_1, \ldots, X_n$ are i.i.d. according to P_θ, $\theta \in \Theta$, that Assumption A holds and $R(\theta)$ is of full rank and bounded away from zero in a neighborhood of θ^*. Then the test that rejects H_n when*

$$nr'(\widehat{\theta}_n) \left(R'(\widehat{\theta}_n) I^{-1}(\widehat{\theta}_n) R(\widehat{\theta}_n) \right)^{-1} r(\widehat{\theta}_n) < C_\alpha \ ,$$

where C_α *is such that* $\Pr\left[\chi^2_r(\delta^2) < C_\alpha\right] = \alpha$*, is in* $\mathcal{K}^\alpha$ *and*

- *is asymptotically unbiased.*
- *is asymptotically minimax in* $\mathcal{K}^\alpha$ *against* $\|r(\theta^*)\| < \gamma/\sqrt{n}$, $0 < \gamma < \delta$ *with asymptotic guaranteed power* $\Pr\left[\chi^2_r(\gamma^2) < C_\alpha\right]$.
- *is asymptotically most powerful in* $\mathcal{K}^\alpha$ *against* $r\left(\theta^*\right) = 0$.
- *is asymptotically uniformly most powerful in* $\mathcal{K}^\alpha$ *among tests invariant to orthogonal transformations on* $\mathbb{R}^r$.

Proof. We implicitly assume further that the parameter space is included in $\{\theta : \|\theta_1^*\| < M/\sqrt{n}\}$, with M arbitrary large. Under the assumptions of the theorem, we can locally reparameterize the model so that the first r coordinates of θ corresponds to $r\left(\theta\right)$. Then the problem reduces to

$$H_n : \ \|\theta_1^*\| \geq \delta/\sqrt{n} \qquad \text{against} \qquad K_n : \ \|\theta_1^*\| < \delta/\sqrt{n}\ ,$$

where $\theta^* = (\theta_1^*, \theta_2^*) \in \mathbb{R}^r \times \mathbb{R}^{p-r}$. If the asymptotic variance is known, then the test that rejects H_n when

$$nr'(\widehat{\theta}_n)\left(R'(\theta^*)I^{-1}(\theta^*)R(\theta^*)\right)^{-1} r(\widehat{\theta}_n)$$

is less than C_α has the required properties from Lemma 3.1. As $\widehat{\theta}^*$ is consistent for θ^* and the spectrum of $\left(R'(\theta^*)I^{-1}(\theta^*)R(\theta^*)\right)^{-1}$ is bounded away from infinity, the test of the theorem is asymptotically equivalent and shares the same properties. ■

4 Practice

To be completed

4.1 Critical values

4.2 Reporting results

4.3 Examples

REFERENCES

ANDREWS, D.W.K. (1989). Power in econometric applications. *Econometrica* 57, 1059–1090.

BOROVKOV, A.A. (1998). *Mathematical statistics.* Amsterdam: Overseas Publishers Association.

GODFREY, L.G. (1988). *Misspecification Tests in Econometrics*. New York: Cambridge University Press.

GOOD, I.J. (1981). Some logic and history of hypothesis testing. *Philosophical Foundations of Economics*, J.C. Pitt ed., 149–174. Reprinted in *Good Thinking: The Foundations of Probability and Its Applications* (1983) Minneapolis: University of Minessota Press.

HOENIG, J.M. AND HEISEY, D.M. (2001). The abuse of power: the pervasive fallacy of power calculations for data analysis. *American Statistician* 55, 19–24.

JANSSEN, A. (2000). Nonparametric bioequivalence tests for statistical functionals and their efficient power functions. *Statistics and Decisions* 18, 49–78.

LAVERGNE, P. (1998). Selection of regressors in econometrics: parametric and nonparametric methods. *Econometric Reviews* 17, 227–273.

LEAMER, E.E. (1983). Model Choice and Specification Analysis. Z. Griliches and M. Intriligator, eds., Handbook of Econometrics, Volume 1, Amsterdam: North Holland.

LEAMER, E.E. (1988). Things that bother me. *Economic Record* 64, 331–335.

LEHMANN, E.L., AND ROMANO, J.P. (2005). *Testing Statistical Hypotheses.* New York: Springer.

MCCLOSKEY, D.N. (1985). The loss function has been mislaid: the rhetoric of significance tests. *American Economic Review* 75, 201–205.

NAKAMURA, A.O., NAKAMURA, M., AND DULEEP, H.O. (1990). Alternative approaches to model selection. *Journal of Economic Behavior and Organization* 14, 97–125.

ROMANO, J.P. (2005). Optimal testing of equivalence hypotheses. *Annals of Statistics* 33, 1036–1047.

WELLEK, S. (2003). *Testing Statistical Hypotheses of Equivalence*. New York: Chapman and Hall/CRC.