Queen's Economics Department Working Paper No. 1319

Hypothesis Testing for Arbitrary Bounds

Jeffrey Penney
Queen's University

Department of Economics
Queen's University
94 University Avenue
Kingston, Ontario, Canada
K7L 3N6

10-2013

Hypothesis Testing for Arbitrary Bounds

Jeffrey Penney[1]
Department of Economics, Queen's University
penneyj@econ.queensu.ca

version 4.0.0 - October 3, 2013

I derive a rigorous method to help determine whether a true parameter takes a value between two arbitrarily chosen points for a given level of confidence via a multiple testing procedure which strongly controls the familywise error rate. For any test size, the distance between the upper and lower bounds can be made smaller than that created by a confidence interval. The procedure is more powerful than other multiple testing methods that test the same hypothesis. This test can be used to provide an affirmative answer about the existence of a negligible effect.

Keywords: familywise error, multiple testing, null effect, partial identification, precise zero

JEL Classification: C12, C18

1. Introduction

In statistical analysis, it is often of interest to determine whether the value of an unknown parameter lies within a certain range. There are two primary reasons for this. The first is to determine if a true parameter lies between the upper and lower bounds of what is considered usual for the issue at hand. The second is to investigate the possibility of a "null result"; to illustrate, consider a scenario where a biostatistician would like to determine if a new drug treatment has a negligible effect on the incidence of a specific side effect usually associated with the class of drugs that the new drug is a member of.

The current practice is usually based on confidence intervals, normally ones derived from inverting a t-statistic. Typically, an investigator would arbitrarily pick an upper and a lower bound for the parameter of interest, and then examine whether the entire range of the estimated confidence interval falls within these chosen bounds. The problem with this approach is that the confidence interval is symmetric and selected for the investigator, which may result in inference about the range of values of the true parameter that is too conservative. The issue of what constitutes a null result is especially nebulous (Krantz, 1999), especially since the failure to reject a null hypothesis could simply be due to a lack of precision.

This paper proposes a formal method that assuages the aforementioned problems. After the investigator selects the upper and lower bounds for the test, the procedure then deter-

[1]Acknowledgements: I would like to thank Nicola Lacetera, Lealand Morin, and Benoit Perron for their comments and suggestions. All remaining errors are my own.

mines whether sufficient evidence exists to conclude that the true parameter lies within the prescribed range. The method is constructed using a combination of hypothesis tests and is thus a multiple testing procedure. Section 2 introduces the notation, formulates the problem, and discusses existing methods. The proposed testing procedure is outlined in Section 3. In Section 4, I apply it to an illustrative empirical example. This test has a number of desirable technical properties, which are proven in Section 5.

2. Problem Formulation and Existing Methods

Label the lower bound τ_l, and the upper bound τ_u; $\tau_l < \tau_u$. The estimated coefficient of interest is $\widehat{\beta}$. The two relevant hypotheses are:

$$H_{o,l} : \beta \leq \tau_l \qquad H_{a,l} : \beta > \tau_l \tag{1}$$

$$H_{o,u} : \beta \geq \tau_u \qquad H_{a,u} : \beta < \tau_u \tag{2}$$

The p-value for test (1) shall be designated $p_{o,l}$ and the p-value for test (2) is $p_{o,u}$. For simplicity of exposition, I combine both of these hypotheses to form a *combined test*:

$$H_{o,c} : H_{o,l} \veebar H_{o,u} \qquad H_{a,c} : \tau_l < \beta < \tau_u \tag{3}$$

Care needs to be taken in interpretation of the outcome of the combined test; in particular, if both nulls fail to be rejected, there is a potential contradiction[2]. In the case where both nulls fail to be rejected, users should conclude that insufficient evidence exists to claim $\tau_l < \beta < \tau_u$ at the level of confidence α.

In a multiple testing scenario, the probability of rejecting at least one true null hypothesis is called the familywise error rate (FWE), here denoted as α. A test is said to have *strong* control of the FWE at level α if it is equal to or below level α regardless of which null hypotheses are true in $\mathcal{H} = \{H_{o,1}, \ldots, H_{o,m}\}$.

One recently developed method that can be applied to an arbitrary bounds problem is a partitioning procedure as proposed by Finner and Stassburger (2002). There are several key differences between their procedure and the one proposed in this paper. Their procedure, when applied to this bounds problem[3], reverses the direction of the signs of the null hypotheses; this necessitates a different interpretation of the outcomes of the joint hypothesis test[4]. The method also requires the generation of non-standard test statistics and critical values. Other recently developed multiple testing methods are not valid to be applied to the problem examined here, such as Romano and Wolf (2005) and Bittman et al. (2008), because of the dependence of the hypotheses and the fact that the nulls operate in different directions.

[2]See Finner and Strassburger (2002) for a guide as to how to interpret these cases.

[3]This is done in Example 4.2 on page 1207 in Finner and Strassburger (2002).

[4]For example, consider the case of large standard errors wherein the two nulls will not be rejected for many different values of τ_l and τ_u.

3. Testing Procedure

The steps to test the combined hypothesis (3) are as follows:

1. Choose the lower bound τ_l and the upper bound τ_u.
2. Conduct a test of the null hypothesis of the lower bound $H_{o,l} : \beta \leq \tau_l; H_{a,l} : \beta > \tau_l$ and record its p-value $p_{o,l}$.
3. Conduct a test of the null hypothesis of the upper bound $H_{o,u} : \beta \geq \tau_u; H_{a,u} : \beta < \tau_u$ and record its p-value $p_{o,u}$.
4. Define the p-value of the combined test $p_{o,c} = sup\{p_{o,l}, p_{o,u}\}$. If $p_{o,c} \leq \alpha$, reject the combined null hypothesis and conclude sufficient evidence exists that $\tau_l < \beta < \tau_u$.

It is important to note that the bounds should not be interpreted as a confidence interval. This testing procedure says nothing about the chance that the population parameter is between the upper and lower limit; rather, rejecting the combined null hypothesis leads us to conclude that it is unlikely that the population parameter is outside the bounds. If the combined null hypothesis is true, there is at most an α chance that the test wrongly produces an affirmative result.

4. Empirical Example

Example 1. A biostatistician working for a pharmaceutical company is performing an analysis of a drug currently in clinical trials. According to the country's regulations, a drug must list as a side effect any condition that occurs over 3 percentage points more or less often in patients taking the drug compared to placebo. Hence, the investigator sets $\tau_l = -3$ and $\tau_u = 3$. He estimates a regression that includes a dummy variable for those in the experimental group. The estimated coefficient on the dummy is $\widehat{\beta} = -0.09$, with $SE(\widehat{\beta}) = 1.72$. The 95% confidence interval on the parameter estimate is $[-3.47, 3.29]$, exceeding both limits. He employs the arbitrary bounds test. Since $|\widehat{\beta} - \tau_l| = |-0.09 - -3| < |-0.09 - 3| = |\widehat{\beta} - \tau_u|$, he examines the lower bound since it will yield the larger p-value. Performing the one-tail test of $H_{o,l} : \beta \leq \tau_l$ gives a p-value of 0.045. The investigator concludes that sufficient evidence exists to claim $-3 < \widehat{\beta} < 3$ at level $\alpha = 5\%$: the drug is not different from the placebo for the incidence of the side effect.

5. Mathematical Properties

I first discuss an important technical issue. Denote the set of null hypotheses $\mathcal{H} = \{H_{o,l}, H_{o,u}\}$. Since the case that all null component hypotheses are true is not possible, $\mathcal{H}$ is said to be closed under arbitrary intersections; furthermore, a global null hypothesis[5] does not exist (Finner and Strassburger, 2002). Because of this, the technical property of coherence is

[5] A global null hypothesis is defined as the case where all null hypotheses in a set of null hypotheses $\mathcal{H}$ are true.

satisfied. A test is *coherent* if whenever a joint null hypothesis fails to be rejected, all of its components also fail to be rejected; for example, a test that fails to reject the joint null hypothesis $H_{12} : \beta_1 = \beta_2 = 0$ will also fail to reject both $H_1 : \beta_1 = 0$ and $H_2 : \beta_2 = 0$ separately. Since there are no such implications in the combined test, this property is satisfied vacuously. Coherence is a necessary feature of any multiple testing procedure, as any incoherent test can be replaced with a coherent test that has at least as much power and will reject in every situation where the noncoherent test will reject (Sonnemann and Finner, 1988).

This bounds test has the following technical properties, which I prove in turn. I first present a technical lemma for later use.

Lemma 1. *The inequality $Pr(p_{o,l} \leq \alpha | p_{o,u} < p_{o,l}) \leq Pr(p_{o,u} \leq \alpha | p_{o,u} < p_{o,l})$ holds.*

Proof. Suppose instead $Pr(p_{o,l} \leq \alpha | p_{o,u} < p_{o,l}) > Pr(p_{o,u} \leq \alpha | p_{o,u} < p_{o,l})$. Using Bayes's theorem and cancelling out the denominator, the equality can be contradicted. □

Theorem 1. *The proposed test has the following properties.*

(i) The familywise error rate is strongly controlled for at level α.

(ii) $FWE = \alpha$ is possible.

(iii) $FWE \overset{a.s.}{\to} \alpha$.

(iv) $\forall \alpha$, one can bound the estimated parameter $\widehat{\beta}$ in a narrower area compared to using a confidence interval based on a t-statistic.

(v) $Pr(H_{o,c}$ is rejected$| \tau_l < \beta < \tau_u) \to 1$ as $n \to \infty$.

(vi) This is an asymptotically uniformly most powerful (AUMP) test.

Proof. (i) Without loss of generality, consider the case where (2) is true at the point $\beta = \tau_u$. We have two possibilities for test statistics: $p_{o,c} = sup\{p_{o,l}, p_{o,u}\} = p_{o,u}$ and $p_{o,c} = p_{o,l}$. Denote the expression $(\beta = \tau_u) \equiv \theta$. Then, the FWE is

$$
\begin{aligned}
&= Pr(p_{o,u} \leq \alpha | p_{o,u} \geq p_{o,l}, \theta) Pr(p_{o,u} \geq p_{o,l} | \theta) + Pr(p_{o,l} \leq \alpha | p_{o,u} < p_{o,l}, \theta) Pr(p_{o,u} < p_{o,l} | \theta) \\
&\leq Pr(p_{o,u} \leq \alpha | p_{o,u} \geq p_{o,l}, \theta) Pr(p_{o,u} \geq p_{o,l} | \theta) + Pr(p_{o,u} \leq \alpha | p_{o,u} < p_{o,l}, \theta) Pr(p_{o,u} < p_{o,l} | \theta) \\
&= Pr(p_{o,u} \leq \alpha | \theta) = \alpha.
\end{aligned}
$$

The inequality follows by Lemma 1, and the second equality follows from the law of total probability. An analogous proof follows for the case of (1) being true at the point $\beta = \tau_l$. Since this proof shows that the FWE is controlled at level α regardless of which null hypothesis is true, the test strongly controls for the FWE. □

Proof. (ii) Consider the case where the null $H_{o,u}$ is true, $\beta = \tau_u$, and $\frac{\widehat{\beta} - \tau_u}{se(\widehat{\beta})} = t^{-1}(df, 1 - \alpha)$ where $t^{-1}(df, \cdot)$ is the quantile function of Student's t distribution with df degrees of freedom. By continuity, $\exists \tau_l : \frac{\widehat{\beta} - \tau_l}{se(\widehat{\beta})} = t^{-1}(df, \alpha) = |t^{-1}(df, 1 - \alpha)|$, thus $p_{o,l} = p_{o,u}$. Thus, $p_{o,l}$ is irrelevant. Then, $Pr(p_{o,u} \leq \alpha | \beta = \tau_u) = \alpha$, giving us exactly FWE $= \alpha$. □

Proof. (iii) Without loss of generality, consider (2) to be true at the point $\beta = \tau_u$. Then, $Pr(sup\{p_{o,l}, p_{o,u}\} = p_{o,u}) \overset{a.s.}{=} 1$, since $\widehat{\beta} \overset{p}{\to} \beta = \tau_u$ implies that $|\widehat{\beta} - \tau_u| < |\widehat{\beta} - \tau_l| \Rightarrow t_{o,u} < t_{o,l}$

as $n \to \infty$. The FWE is thus

$$
\begin{aligned}
&= Pr(p_{o,u} \leq \alpha | p_{o,u} \geq p_{o,l}, \theta) Pr(p_{o,u} \geq p_{o,l} | \theta) + Pr(p_{o,l} \leq \alpha | p_{o,u} < p_{o,l}, \theta) Pr(p_{o,u} < p_{o,l} | \theta) \\
&\stackrel{a.s.}{=} Pr(p_{o,u} \leq \alpha | \theta) = \alpha
\end{aligned}
$$

where the almost surely equality follows because $Pr(p_{o,u} \geq p_{o,l}|\theta) \stackrel{a.s.}{=} 1$ and $Pr(p_{o,u} \leq \alpha | p_{o,u} \geq p_{o,l}, \theta) \stackrel{a.s.}{=} Pr(p_{o,u} \leq \alpha|\theta)$. An analogous proof can be constructed for the case of $\beta = \tau_l$.

□

Proof. (iv) A $(1-\alpha)\%$ confidence interval is defined as $\widehat{\beta} - se(\widehat{\beta}) \cdot t^{-1}(df, \alpha/2)$, $\widehat{\beta} + se(\widehat{\beta}) \cdot t^{-1}(df, \alpha/2)$. Set $\tau_l = \widehat{\beta} - se(\widehat{\beta}) \cdot t^{-1}(df, \alpha/2)$, and $\tau_h = \widehat{\beta} + se(\widehat{\beta}) \cdot t^{-1}(df, \alpha/2)$. See that

$$
\tau_l = \widehat{\beta} - se(\widehat{\beta}) \cdot t^{-1}(df, \alpha/2) \Rightarrow t^{-1}(df, \alpha/2) = \frac{\widehat{\beta} - \tau_l}{se(\widehat{\beta})} > t^{-1}(df, \alpha)
$$

where the right-hand side expression of the second equal sign is the t-statistic for hypothesis (1), and $t^{-1}(df, \alpha)$ is its critical value. Similarly, we see that $\tau_u = \widehat{\beta} + se(\widehat{\beta}) t^{-1}(df, \alpha/2) \Rightarrow \frac{\widehat{\beta} - \tau_u}{se(\widehat{\beta})} = -t^{-1}(df, \alpha/2) = t^{-1}(df, 1-\alpha/2) < t^{-1}(df, 1-\alpha)$ where the left-hand side expression of the second equal sign is the t-statistic for hypothesis (2). Thus, we can pick a larger value for τ_l and a smaller value for τ_u, giving us a range of values narrower than that created by a confidence interval for a given level of α. □

Remark. (iv) In fact, we can set $p_{o,l} = \alpha$, $p_{o,u} = \alpha$, and use these values in the t-statistics for the hypothesis tests (1) and (2) to derive figures for τ_l and τ_u that would form an interval that would be equivalent to a $(1-2\alpha)$ confidence interval. It is easy to see that this is the maximum possible length of the interval.

Proof. (v) Assume $\tau_l < \beta < \tau_u$ and, without loss of generality, $|\beta - \tau_u| < |\beta - \tau_l|$. Since $\widehat{\beta} \xrightarrow{p} \beta$, $Pr(|\widehat{\beta} - \tau_u| < |\widehat{\beta} - \tau_l|) \stackrel{a.s.}{=} 1$ as $n \to \infty$, thus $Pr(sup\{p_{o,l}, p_{o,u}\} = p_{o,u}) = 1$. Therefore, we need only consider $H_{o,u}$ to reject the combined null $H_{o,c}$. We thus require $z^{-1}(\alpha) \geq \frac{\widehat{\beta} - \tau_u}{se(\widehat{\beta})}$, where $z^{-1}(\alpha)$ is the α-quantile of the standard normal distribution. Clearly, $\lim_{n\to\infty} \frac{\widehat{\beta} - \tau_u}{se(\widehat{\beta})} = -\infty \leq z^{-1}(\alpha)$ since $\lim_{n\to\infty} se(\widehat{\beta}) = 0$ and $\beta < \tau_u$. Hence, the null hypothesis $H_{o,u}$ will be rejected with probability 1 asymptotically, which means $H_{o,c}$ will be as well. The proof of $|\beta - \tau_u| > |\beta - \tau_l|$ is similar, and the case of $|\beta - \tau_u| = |\beta - \tau_l|$ is trivial. □

Proof. (vi) Since $H_{o,l} \cap H_{o,u} = \emptyset$, that is, both nulls that make up the combined test cannot be true simultaneously, a global null hypothesis does not exist. Therefore, the combined null hypothesis must necessarily be tested through its individual components $H_{o,l}$ and $H_{o,u}$. Asymptotically, the Likelihood Ratio test and the t-test are equivalent; therefore, $H_{o,l}$ and $H_{o,u}$ are both AUMP at level α. Since each component of the combined test is AUMP, the combined test is AUMP as the rejection probability is maximized through the sum of its parts at the level FWE $\xrightarrow{a.s.} \alpha$. □

References

[1] Bittman, Richard M., Joseph P. Romano, Carlos Vallarino, and Michael Wolf. (2008): "Optimal testing of multiple hypotheses with common effect direction," Institute for Empirical Research in Economics, Working Paper No. 307.

[2] Choi, Sungsub, W.J. Hall, and Anton Schick. (1996): "Asymptotically Uniformly Most Powerful Tests in Parametric and Semiparametric Models," *Annals of Statistics*, vol. 24: 841-861.

[3] Finner, Helmut, and Klaus Strassburger. (2002): "The Partitioning Principle: A Powerful Tool in Multiple Decision Theory," *Annals of Statistics*, vol. 30: 1194-1213.

[4] Krantz, David H. (1999) "The Null Hypothesis Testing Controversy in Psychology," *Journal of the American Statistical Association*, vol. 44: 1372-1381.

[5] Romano, Joseph P., and Michael Wolf. (2005): "Stepwise Multiple Testing as Formalized Data Snooping," *Econometrica*, vol. 73: 1237-1282.

[6] Sonnemann, E., and Finner, H. (1988): Vollständigkeitssätze für multiple Testprobleme. In Multiple Hypothesenprüfung, Ed. P. Bauer, G. Hommel, and E. Sonnemann, pp. 121-135. Berlin: Springer.