A Note on a Family of Criteria for Evaluating Test Statistics

Gordon Anderson[*] Teng Wah Leo[†]

University of Toronto St. Francis Xavier University

December 30, 2013

_____

[*]Department of Economics, University of Toronto. Email Address: `anderson@chass.utoronto.ca`

[†]Department of Economics, St. Francis Xavier University. Email Address: `tleo@stfx.ca`

Abstract

In noting that the usual criteria for choosing an optimal test, Uniform Power and Local Power are at opposite ends of a spectrum of dominance criteria, a complete "Power Dominance" family of criteria for classifying and choosing optimal tests on the basis of their power characteristics is identified, wherein successive orders of dominance attach increasing weight to power close to the null hypothesis. Indices of the extent to which a preferred test has superior power characteristics over other members in its class, and an index of the proximity of a test to the envelope function of alternative tests are also provided. The ideas are exemplified using various optimal test statistics for Normal and Laplace population distributions.

1 Introduction

Choosing between alternative test instruments is usually based upon comparison of the power properties of competing tests, and frequently it is a matter of visual comparison of the respective power curves and the envelope of competing power curves (see for example Conover et al. (1978), Juhl and Xiao (2003), Omelka (2005) and Ramsey (1971)). For example, in considering a test[1] $H_0 : \mu \leq \mu_0$ versus $H_1 : \mu > \mu_0$ of size α, the choice of test instrument T_k is usually based upon power considerations where, given β is the probability of a Type 2 Error, i.e. $\beta = \Pr(\text{Fail to reject } H_0 | H_1)$, the power function of the test is $1 - \beta(\alpha, \mu_1, T_k)$, where $\mu_1 \in H_1$ (see for example Lehmann and Romano (2005)). For convenience in what follows, since we wish to consider the power function as a cumulative density function, note that for consistent tests, the power function $1 - \beta(\alpha, \mu_1, T_k) \in [\alpha, 1]$, so that letting $F(\mu_1 | \alpha, T_k) = \frac{1 - \alpha - \beta(\alpha, \mu_1, T_k)}{1 - \alpha}$, $F(\mu_1 | \alpha, T_k) \in [0, 1]$, and $f(\mu_1 | \alpha, T_k) = \frac{dF(\mu_1 | \alpha, T_k)}{d\mu_1}$ corresponds to a probability distribution of μ_1, for $\mu_1 > \mu_0$. Generally the test of choice, if it exists, is the Uniformly Most Powerful (UMP) test, which for test statistics T_k, $k = 1, 2, \ldots$, defines T_1 as the UMP test if:

$$F(\mu_1 | \alpha, T_1) \geq F(\mu_1 | \alpha, T_k) \ \forall \ \mu_1 \in H_1 \ \text{ and all } \ k > 1$$

(with strict inequality holding somewhere). When no such test exists, resort is usually made to the Locally Most Powerful (LMP) test criterion which defines T_1 as the LMP test if:

$$\frac{dF(\mu_1 | \alpha, T_1)}{d\mu_1} \geq \frac{dF(\mu_1 | \alpha, T_k)}{d\mu_1} \ \forall \ \mu_1 \in (\mu_0, \mu_0 + \delta] \ \text{ for } \delta \text{ small}$$

These comparison criteria for statistical tests are very similar in nature and spirit to the extremes of the family of stochastic dominance criteria used by economists in empirical welfare and finance literatures. In essence, they are an expression, on the part of the statistical investigator, of a preference for power, whereas Economists have used them as an expression of preference for wellbeing or portfolio returns. What is interesting here is that in the context of economic comparisons, the UMP and LMP criterion are at opposite ends of the comparison spectrum in that literature, and as will be seen, in between these ends are a plethora of comparisons which suit the particular needs of

---

[1]A similar analysis is possible for two sided tests, but is omitted here for reasons of brevity.

the economic analysis, and so too it could be for statisticians when considering power properties of tests.

2 Power Dominance & the Optimal Test

In economics literature, stochastic dominance is used as a criteria for choosing between two potential distributions of a random variable x (usually income, consumption or a portfolio return) in order to find the distribution which maximizes $\mathbf{E}(U(x))$ based upon the properties of the function $U(.)^2$, where $U(.)$ represents a felicity function either of agents in a society under the income size distribution of x (the incomes of its members) or that derived from the size distribution of portfolio returns x. Working with $U(x)$, where for simplicity x is continuously defined over the range $[a, b]$, and two alternative states defined by pdf's $f(x)$ and $g(x)$ describing the distribution of x across agents in those states, the family of Stochastic Dominance techniques address the issue: "which state is preferred if the objective is the largest $\mathbf{E}(U(x))$?". Formally, when $U(x)$ is such that $(-1)^{i+1}\frac{\partial U(x)}{\partial x^i} > 0$, for $i = \{1, \ldots, J\}$, a sufficient condition for:

$$\mathbf{E}_f(U(X)) - \mathbf{E}_g(U(X)) = \int_a^b U(x)(dF - dG) \geq 0$$

is given by[3]:

$$F_J(x) \leq G_J(x) \ \forall \ x \in [a, b]$$
$$\text{and } F_J(x) < G_J(x) \text{ for some } x \in [a, b]$$
$$\text{where } F_i(x) = \int_a^x F_{i-1}(z)dz$$
$$\text{and } F_0(x) = f(x)$$

An important notion regarding dominance relations is that dominance at order h implies dominance at all orders $h' > h$, and a useful lemma drawn from Davidson and Duclos (2000) is that if F first order dominates G over some region $(-\infty, a)$, then F will dominate G over the whole range of x at some higher order. The Wellbeing implication of various

---

[2]Sometimes poorness indices $P(x)$ are studied (Atkinson 1987) in which case $U(x) = -P(x)$.

[3]The intuition for how the result is arrived at follows in the discussion below.

dominance orders are[4]:

$U_1 \sim$ (requiring $\frac{dU(x)}{dx} > 0$) yields the *first order dominance* rule (referred to as Utilitarian societal preferences), and is an expression of preference for more of x, without regard to spread.

$U_2 \sim$ (requiring $\frac{dU(x)}{dx} > 0$, $\frac{d^2U(x)}{dx^2} < 0$) yields the *second order dominance* rule (referred to as Daltonian societal preferences), and is an expression of preference for more x, with weak preference for reduced spread.

$U_3 \sim$ (requiring $\frac{dU(x)}{dx} > 0$, $\frac{d^2U(x)}{dx^2} < 0$, and $\frac{d^3U(x)}{dx^3} > 0$) yields a *third order dominance* rule, and is an expression of preference for more x, with weak preference for reduced spread, especially at the low end of the distribution.

$\vdots$

$U_\infty \sim$ or *infinite order dominance*, is referred to as Rawlsian societal preference since it attaches infinite weight to the poorest individual, and can be examined in the context of the relative incomes of the poorest individuals in 2 equally populated societies[5].

In the Wellbeing and Finance fields, the stochastically dominating distribution is the preferred distribution, reflecting as it does the desire for greater average $U(x)$. In the present "Power" context, the stochastically dominated (*power dominant*) distribution will be the preferred distribution, reflecting as it does faster accumulating "correct" rejection probabilities over alternative tests as μ_1 increases. In other words, the question is "which power function is preferred if the objective is smallest $\mathbf{E}(U(x))$ or the fastest accumulation of mass as μ_1 moves away from H_0?". In this context, the UMP test criterion can be seen to be closely associated with the first order stochastic dominance criterion, since it is first order stochastically dominated by (or *power dominate*) the power functions of all other tests (that is $F(\mu_1|\alpha, T_1)$ is first order stochastically dominated by all other $F(\mu_1|\alpha, T_k)$, for $k > 1$). Intuitively, we seek the test which accumulates power most rapidly with increasing μ_1. In a similar fashion, the LMP test criterion can be seen to be very similar to the infinite order stochastic dominance criterion, since the test whose power function

---

[4]The comparison procedures have been empirically implemented in several ways, see for example Anderson (1996, 2004), Barrett and Donald (2003), Davidson and Duclos (2000), Linton et al. (2005), Knight and Satchell (2008), and McFadden (1989).

[5]Knight and Satchell (2008) have empirically implemented this order of dominance comparison.

is infinite order stochastically dominated by (or power dominate) the power functions of all other tests is the LMP test (i.e. $\frac{dF(\mu_1|\alpha, T_1)}{d\mu_1} \geq \frac{dF(\mu_1|\alpha, T_k)}{\mu_1}$ for all $\mu_1 \in (\mu_0, \mu_0 + \delta]$ for δ small)[6]. Second order power dominance comparisons can best be understood by considering the pdf's of two power functions, that for expositional convenience are symmetric with equal means, but different spreads. The first order power dominance comparison (UMP comparison) would yield no conclusion because the cdf's would cross at the mean (see for example comparisons in Anderson (2001)). The second order power dominance comparison however would yield a decision favoring the distribution with the largest variance, since the integral of its cdf would lay everywhere above that of the distribution with the smaller variance. In a similar fashion, the third order power dominance comparison can be best understood by considering two distributions with equal mean and variance, that are left (positively) skewed, so that the third order power dominance relationship will favour the one that is most left skewed. Essentially successive orders of "power dominance" can be seen to attach increasing weight to mass close to the null hypothesis. Intuitively, power function crossovers closer to the null require successively higher orders of integration to satisfy a power dominance criterion, so that power dominance at order "i" represents a stronger form of power dominance than that at order "$i + k$", for $k > 0$.

Thus we are led to consider a family of *Power Dominance* criteria of the form:

$$\int_{\mu_0}^{\mu_1} \left(F_j(z|\alpha, T_1) - F_j(z|\alpha, T_k) \right) dz \quad \geq \quad 0 \; , \forall \; \mu_1 > \mu_0 \text{ and } k \neq 1$$

$$\text{where } F_{j+1}(\mu_1|\alpha, T)dz = \int_{\mu_0}^{\mu_1} F_j(z|\alpha, T)dz \; , \quad \text{and} \quad F_0 = f$$

as a j'th order power dominance criterion.

Furthermore, note that

$$PB_{j+1} \quad = \quad \int_{\mu_0}^{\infty} \left(F_j(\mu_1|\alpha, T_1) - F_j(\mu_1|\alpha, T_k) \right) d\mu_1$$

provides a measure of the *power benefit* of test 1 over test k at the j'th order comparison, in other words it is a measure of how much better T_1 is than T_k at order j. In this context,

---

[6]It follows from the Davidson and Duclos (2000) lemma that if Test 1 first order power dominates locally, as the (LMP) does, it will dominate at some higher order over the whole range of μ_1.

where the metric of the unstandardized measure is related to the units of $(\mu_0)^j$, PB_{j+1} could be divided by a value of $(\mu_0)^j$, thereby making it a unit free measure.

Finally the Envelope of alternative power functions is frequently used for comparison purposes (see Omelka (2005), and Juhl and Xiao (2003) for example), and proximity to it gauged by visual inspection. The proximity could be quantified by a modification of PB to the following form:

$$PBE_j \;\; = \;\; - \int_{\mu_0}^{\infty} \left[F(\mu_1 | \alpha, T_j) - ENV_{k=1}^K \left(F(\mu_1 | \alpha, T_k) \right) \right] d\mu_1 \;\; , j \in \{1, \ldots, K\}$$

where $ENV_{k=1}^K = \sup_{k=1}^K F(z | \alpha, T_k)$, and PBE_j is a measure of the aggregate distance of the power curve of the j'th test statistic from the envelope. Note the envelope will always first order power dominate any of the individual tests from which it is constructed. PBE_j is thus a measure of the first order power dominance of the envelope over the j'th test, and its proximity to 0 is a measure of how superior in power terms the j'th test statistic is.

3 Examples

3.1 Example 1

The following illustrative example is for a sample mean ($\sim N(\mu, \sigma^2/n)$) and sample median ($\sim N(\mu, \pi\sigma^2/2n)$) comparison using the above criteria for test size α.

The power function for a sample mean test is given by:

$$F(z | \alpha, \overline{X}) \;\; = \;\; 1 - Z \left(\frac{\mu_0 - \mu_1}{\frac{\sigma}{\sqrt{n}}} + Z_{1-\alpha} \right)$$

The power function for sample median test is given by:

$$F(z | \alpha, X_{median}) \;\; = \;\; 1 - Z \left(\frac{\mu_0 - \mu_1}{\sqrt{\frac{\pi}{2n}} \sigma} + Z_{1-\alpha} \right)$$

Note that consistent with the mean test being the uniformly most powerful test,

$$F(z | \alpha, \overline{X}) \;\; \geq \;\; F(z | \alpha, X_{median}) \;\; , \forall \; \mu_1 > \mu_0 \;\; \text{and} \;\; \sigma > 0$$

since,

$$\frac{\mu_0 - \mu_1}{\sqrt{\frac{\pi}{2n}} \sigma} \;\; > \;\; \frac{\mu_0 - \mu_1}{\frac{\sigma}{\sqrt{n}}} \;\; \forall \; \mu_1 > \mu_0 \;\; \text{and} \;\; \sigma > 0$$

5

Hence the test using the mean is the *first order dominated* test, which is equivalent to a test from the UMP class. The index of relative power would be:

$$PB_1 = \frac{1}{\mu_0} \int_{\mu_0}^{\infty} \left[Z \left(\frac{\mu_0 - \mu_1}{\sqrt{\frac{\pi}{2n}} \sigma} + Z_{1-\alpha} \right) - Z \left(\frac{\mu_0 - \mu_1}{\frac{\sigma}{\sqrt{n}}} + Z_{1-\alpha} \right) \right] d\mu_1$$

which is positive for all finite $\sigma > 0$, and diminishing in σ.

3.2 Example 2

A second, perhaps more illuminating example is taken from Omelka (2005), and relates to the power comparisons of the Locally Most Powerful Test (the sign test), the Wald Test (the sample median), and the Maximum Likelihood Test for the location parameter when data follow a Laplace (Double Exponential) distribution. Omelka (2005) is followed in considering a random sample of size $n = 10$ drawn from a population described by the pdf:

$$f(x) = \frac{1}{2} e^{-|x-\theta|}$$

The test being considered is $H_0 : \theta \leq 0$ against $H_1 : \theta > 0$, and the nominal size is 0.0546875. The 3 tests under consideration for this case are defined as:

1. Locally Most Powerful (Sign Test Statistic): $\sum_{i=1}^{n} \mathbb{I}(X_i > 0) \geq K$ (c.v. $K = 8$ (Lehmann (1975)))

2. Wald (Median Test Statistic): $X_{median} \geq C_\alpha$ ($C_\alpha = 0.625$)

3. Maximum Likelihood: $X_{median} \geq 0$ and $2 \sum_{i=1}^{n} (|X_i| - |X_i - X_{median}|) \geq C_\alpha$ ($C_\alpha = 2.86$)

Monte Carlo simulations (50,000 replications in each of 1,000 samples taken at 0.005 increments of θ from 0 to 4.995) facilitate generation of the power curves of the three test statistics which are presented in figure 1.

Figure 2 illustrates the various orders of integration[7] of the difference in the respective power curves (Power(Sign Test) - Power(Wald Test)), and shows the power dominance of the sign test at the *fifth order of dominance*, since it is the fifth curve that is everywhere non-negative. The index of fifth order power dominance of the LMP statistic is 0.135.

---

[7]Integration was approximated using the trapezoidal rule, which for small increments of θ (0.005), provides a more than adequate approximation.
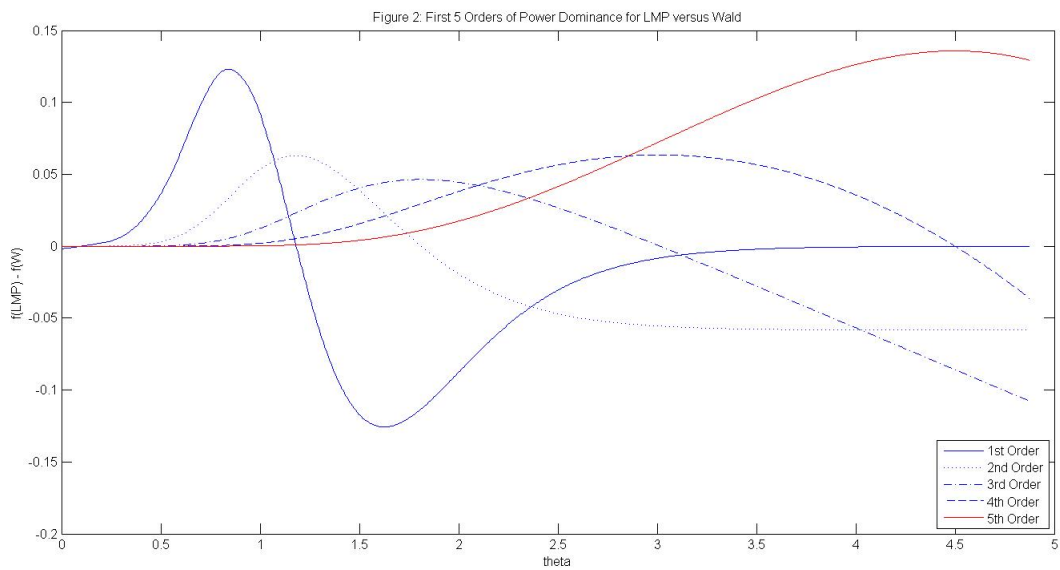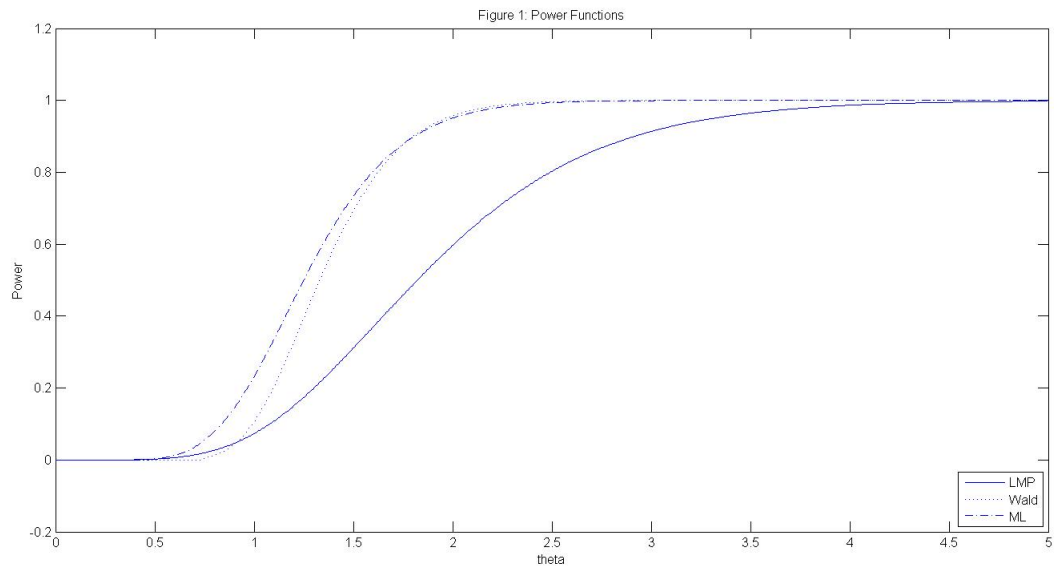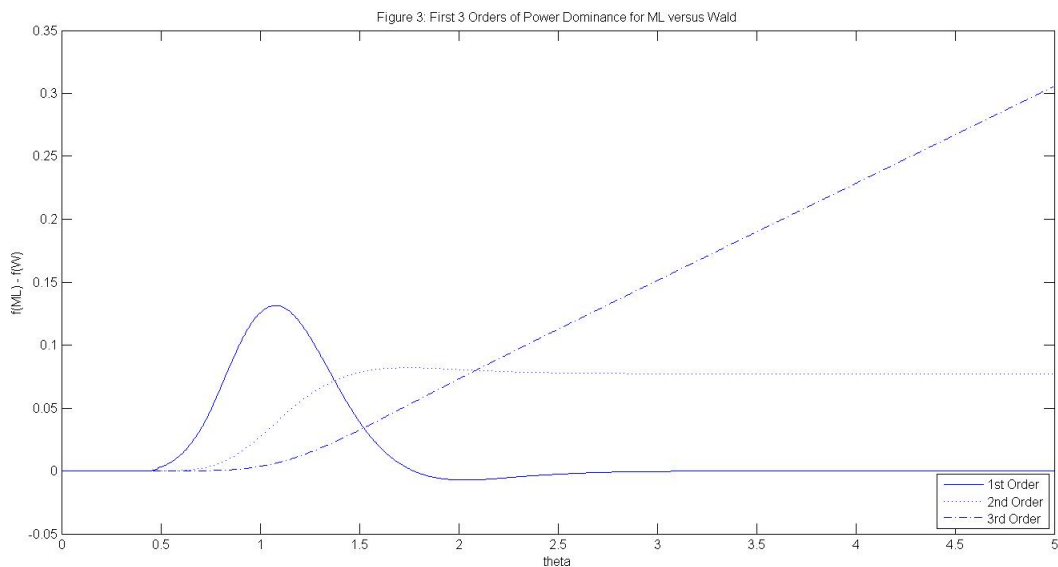
Figure 1: Power Functions

Figure 2: First 5 Orders of Power Dominance for LMP versus Wald

Figure 3 illustrates various orders of integration of the difference in the respective power curves (Power(Likelihood Ratio Test) - Power(Wald Test)), and shows the power dominance of the Likelihood Ratio Test at the *second order*. The index of second order power dominance for this comparison is 0.077.

As for the PBE index of proximity to the envelope, the values for the LMP, Wald and ML statistics are correspondingly 0.6398, 0.0821 and 0.0049, confirming the price of the LMP test in overall power for values of θ further away from the null, and the advantages

7

Figure 3: First 3 Orders of Power Dominance for ML versus Wald

of the ML statistic in overall power across the range of values of θ as inferred by visual inspection in Omelka (2005).

References

Anderson, G. J. (1996). Nonparametric Tests for Stochastic Dominance in Income Distributions. *Econometrica*, 64, 1183–1193.

Anderson, G. J. (2001). The Power and Size of Nonparametric Tests for Common Distributional Characteristics. *Econometric Reviews*, 20, 1–30.

Anderson, G. J. (2004). Toward an Empirical Analysis of Polarization. *Journal of Econometrics*, 122, 1–26.

Atkinson, A. (1987). On the Measurement of Poverty. *Econometrica*, 55, 749–764.

Barrett, G. F. and Donald, S. G. (2003). Consistent Tests for Stochastic Dominance. *Econometrica*, 71, 71–104.

Conover, W. J., Wehmanen, O., and Ramsey, F. L. (1978). A Note on the Small-Sample Power Functions for Nonparametric Tests of Location in the Double Exponential Family. *Journal of the American Statistical Association*, 73, 188–190.

Davidson, R. and Duclos, J.-Y. (2000). Statistical Inference for Stochastic Dominance and for the Measurement of Poverty and Inequality. *Econometrica*, 68, 1435–1464.

Juhl, T. and Xiao, Z. (2003). Power Functions and Envelopes for Unit Root Tests. *Econometric Theory*, 2, 240–253.

Knight, J. and Satchell, S. (2008). Testing for Infinite Order Stochastic Dominance with Applications to Finance, Risk and Income Inequality. *Journal of Economics and Finance*, 32, 35–46.

Lehmann, E. L. (1975). *Nonparametrics: Statistical Methods Based on Ranks*. San Francisco: Holden-Day.

Lehmann, E. L. and Romano, J. P. (2005). *Testing Statistical Hypotheses*. 3 edn. Springer.

Linton, O., Maasoumi, E., and Whang, Y.-J. (2005). Consistent Testing for Stochastic Dominance under General Sampling Schemes. *Review of Economic Studies*, 72, 735–765.

McFadden, D. (1989). Testing for Stochastic Dominance. In T. B. Fomby and T. K. Seo (Eds.), *Studies in the Economics of Uncertainty*. New York: Springer-Verlag.

Omelka, M. (2005). The Behaviour Of Locally Most Powerful Tests. *Kybernetica*, 41, 699–712.

Ramsey, F. L. (1971). Small Sample Power Functions for Non-Parametric Tests of Location in the Double Exponential Family. *Journal of the American Statistical Association*, 66, 149–151.