# A UNIVERSAL UPPER BOUND ON POWER FUNCTIONS

By Allan H. Würtz[1]

School of Economics
University of New South Wales
Sydney 2052, Australia.

June 5, 1997.

The finite sample power function is shown to be bounded by a power upper bound. The same power upper bound is valid for any test of a hypothesis. In normal regression models the power upper bound is simple to calculate including cases where the hypotheses are composite and where the regression function is nonlinear.

Keywords: Composite hypothesis, finite sample power, Lagrange multiplier test, likelihood ratio test, nonlinear regression models, nonmonotonic power, uniformly most powerful test.

---

I am grateful to Tue Gørgens, Joel Horowitz, Hidehiko Ichimura, Forrest Nelson, George Neumann and Tim Robertson for helpful discussion and suggestions. I am especially indebted to Gene Savin.

## 1. INTRODUCTION

In their paper on testing in econometrics, Keuzenkamp and Magnus (1995) note that the predominant methodology to testing in econometrics is that of Neyman and Pearson. In this methodology the emphasis is on size and power. In practice, however, power considerations do not play a predominant role in empirical work. In a study of all empirical work in the American Economic Review 1980-1989, McCloskey and Ziliak (1996, table 1) find two striking numbers on the use of statistical significance; 70.3% use statistical significance as decisive in an empirical argument but only 4.4% consider the power of the test. One barrier to a widespread use of power considerations is that the power function is not, in general, readily available. One result in this paper is an upper bound on the power function that is readily available in non-linear regression models.

The power function plays an important role in the design of controlled experiments and in the analysis of natural experiments. In the design of experiments, the power function is used to determine the sample size: The objective is to achieve the desired power against alternatives of interest. In econometric studies the design is usually not under the control of the investigator. Here the question is whether the data generated by a natural experiment is informative. The power function can be used to determine if sample size is large enough to discriminate between the null and empirically relevant alternatives.

This paper presents a nontrivial upper bound on the power function of a test. The power upper bound does not depend on the test. Instead, the power upper bound depends on the model, the hypothesis and the significance level. This implies universality of the power upper bound in that it bounds the power of any test of the hypothesis. As a consequence, for a given alternative of interest if the value of the power upper bound is low, then no test exists which has high power.

The power upper bound is shown to have a simple form in a normal regression model. In particular, the power upper bound is easy to apply with nonlinear regression functions and nonlinear composite hypotheses. This implies that the power upper bound can be found for tests where it may be intractable to derive the exact power function.

The application of the power upper bound is illustrated with several examples. First, the power upper bound is applied to testing a nonlinear composite hypothesis in a linear model. Compared to the power function of a LR test of that hypothesis, the power upper bound is close for relevant alternatives. Second, the power upper bound is applied to a linear hypothesis in a nonlinear model. It is often implicitly assumed that the power function is a monotonically increasing function of the distance between the null and the alternative. This does not hold for many tests of nonlinear regression functions: The power function is nonmonotonic. This occurs even though the test is consistent. A nonmonotonic power function is illustrated by a LM test of a slope coefficient in an exponential model. For this case, the power upper bound is also nonmonotonic and, hence, nonmonotonic power functions may be detected using the power upper bound.

The power upper bound supplements existing techniques on approximating power functions. The use of asymptotic local power functions can be limiting in several ways; they can be complicated to derive and evaluate. Nelson and Savin (1990) have shown that generalizing the local behavior to global behavior can be potentially very misleading. An advantage of the power upper bound is that it is valid globally for any test. This makes it possible to investigate the power as a function of a whole range of alternatives of interest, not only local alternatives. The power upper bound also bounds the power envelope. Finding power envelopes and point optimal tests have been extensively investigated by Dufour (1989), Dufour and King (1991), Elliott, Rothenberg and Stock (1996) and King (1987). To find the empirical power of tests Monte Carlo methods are commonly used. If the power upper bound is low, then a Monte Carlo investigation is not of interest. If the power upper bound is high, then a Monte Carlo study may usefully be performed for a test statistic. The difference between the power of the test and the power upper bound indicates whether it is worthwhile to search for a better test.

The paper is organized as follows. The power upper bound is derived in a general form in Section 2 and explicitly for a normal regression model in section 3. The remainder of the paper is focused on the use of the power upper bound in a normal regression model. In Section 4 and 5 the power upper bound is applied to a test of a nonlinear composite hypothesis and to show the nonmonotonicity of a power function. Section 6 concludes the paper.

## 2. GENERAL POWER UPPER BOUND

In this section the general upper bound on the power of a test is derived. The power upper bound is obtained by bounding each element of a decomposition of the power function.

Consider a model of a n-dimensional random vector Y described as a density or frequency function $f(y|\theta,X)$. For the sake of exposition, attention is here restricted to the case where $f(y|\theta,X)$ is a density. The function $f(y|\theta,X)$ corresponds to a likelihood function where n is the sample size and y is the n-vector of observations. Assume that X is a $n\times m$ fixed regressor matrix. The function $f(y|\theta,X)$ depends on a k-dimensional parameter vector $\theta$ belonging to a compact parameter space $\Omega \subset \Re^k$. It is assumed that $\theta$ is the only unknown in the specification.

The hypothesis concerns $\theta$. Let the null hypothesis, $H^0$, and the alternative hypothesis, $H^A$, be represented as disjoint subsets of the parameter space $\Omega$. A test is assumed to divide the sample space into a rejection and non-rejection region. The power function of the test, $\pi(\theta)$, equals the probability of observing a sample point in the rejection region of the test when $\theta$ is the true parameter value (see Neyman and Pearson (1933b)). The power function evaluated at $\theta^0 \in H^0$ gives a rejection probability under the null hypothesis. If the null hypothesis is composite, then many values of $\theta^0$ satisfy $H^0$. The largest of the rejection probabilities under the null hypothesis is the size of the test (see Neyman and Pearson (1933a) or Lehmann (1959)).

The power function can be decomposed into a sum of two functions. For this purpose, define a new function.

DEFINITION 1: The *power gain function* $\Delta\pi(\theta^0,\theta)$ of a test with rejection region R is a mapping from $\Omega\times\Omega \rightarrow [-1,1]$ defined as

$$\Delta\pi(\theta^0,\theta) \equiv \pi(\theta) - \pi(\theta^0) = \int_R \left( f(y|\theta,X) - f(y|\theta^0,X) \right)dy.$$

The power gain function equals the change in rejection probability of the test if $\theta$ is the true value rather than $\theta^0$. The power function can then be decomposed as

$$\pi(\theta) = \pi(\theta^0) + \Delta\pi(\theta^0,\theta), \quad \theta^0 \in H^0, \qquad \textbf{(1)}$$

In the case where $H^0$ is composite, different values of $\theta^0 \in H^0$ can be chosen and, thus, different decompositions are possible. For any decomposition, that is $\theta^0 \in H^0$, $\pi(\theta^0)$ is bounded by the size of the test. Hence, the main obstacle in bounding the power function using (1) will be to find a function which bounds the power gain function, $\Delta\pi(\theta^0,\theta)$.

The power gain function of a test is bounded by a function which does not depend on the test. This function will be called the displacement, $D(\theta^0,\theta)$. It measures the distance between the two densities $f(y|\theta^0,X)$ and $f(y|\theta,X)$ using the $L^1$ norm.

DEFINITION 2: The *displacement function* $D(\theta^0,\theta)$ is a mapping from $\Omega\times\Omega \rightarrow [0,1]$ defined as

(i) $$D(\theta^0,\theta) = \tfrac{1}{2} \| f(y|\theta^0,X) - f(y|\theta,X) \|_1 = \tfrac{1}{2} \int_{\Re^n} |f(y|\theta^0,X) - f(y|\theta,X)| dy.$$

where $\|\bullet\|_1$ is the $L^1$ norm with respect to the Lebesgue measure.

The displacement can be interpreted as the probability mass which distinguishes the two densities. This is seen more clearly from the two following expressions, which are equivalent to the definition of the displacement function.

(ii) $$D(\theta^0,\theta) = \int_U (f(y|\theta,X) - f(y|\theta^0,X))dy,$$

where $U = \{ y \mid f(y|\theta,X) > f(y|\theta^0,X) \}$ and

(iii) $$D(\theta^0,\theta) = \int_V (f(y|\theta^0,X) - f(y|\theta,X))dy,$$

where $V = \{ y \mid f(y|\theta^0,X) > f(y|\theta,X) \}$. In (ii) the difference between $f(y|\theta,X)$ and $f(y|\theta^0,X)$ is integrated over the sample points where $f(y|\theta,X)$ dominates $f(y|\theta^0,X)$, and opposite for (iii).

The displacement is associated with other measures of the difference between two distributions. The displacement times 2, the $L^1$ norm, is the total variation of the probability measures corresponding to $f(y|\theta^0,X)$ and $f(y|\theta,X)$, see for example Rudin (1987). Hence, the displacement is a subprobability measure. The $L^1$ norm difference between two distributions is considered briefly by Pitman (1979), who, however, focused on the $L^2$ norm difference. The $L^2$ norm is not smaller than the square of the $L^1$ norm. Other useful inequalities on the $L^1$ norm difference between two distributions are provided by Devroye and Györfi (1985). The displacement can be applied to situations where the Kullback-Leibler information number traditionally has been used, see Plott, Wit and Yang (1996). The Kullback-Leibler information number is the expected difference of the logarithm of the two density functions, see for example Zacks (1971). Note, the displacement is related to identification; $D(\theta^0,\theta)$ equals zero if and only if $\theta_0$ and $\theta$ are unidentified.

Before proving that the change in rejection probability, $\Delta\pi(\theta^0,\theta)$, is bounded by the displacement, $D(\theta^0,\theta)$, consider an illustrative example. Suppose one observation, y, is drawn from one of the two distributions $N(\mu,4)$, $\mu = 0$ or $\mu = 4$. Suppose the null hypothesis $\mu = 0$ is tested against $\mu = 4$ by the following test: Accept the null hypothesis if $y < 3$ and reject if $y \geq 3$. The example is illustrated in Figure 1. The vertical line at $y = 3$ divides the sample space into the acceptance and the rejection region of the test. In the figure, the size of the test is area I, namely, the probability of rejecting the null hypothesis when it is true. The power gain, $\Delta\pi(0,4)$, is marked as area II. Hence, area II together with area I equals the power of the test, $\pi(4)$. Using (ii) the displacement equals area II together with area III. The displacement exceeds the change in rejection probability by area III. Hence, the displacement and the size bounds the power of the test; that is, areas I, II and III bound areas I and II.

The result from the example holds in general. The displacement measures the absolute difference between the two densities over the whole sample space whereas the power gain measures the difference between the two densities over a subset of the sample space, namely, the rejection region of the test. This is formally stated in the following lemma.

LEMMA (Displacement and power gain): *Assume a test of the hypothesis* $H^0$ *about* $f(y|\theta,X)$. *Then, for any* $\theta^0$ *and* $\theta$, $\Delta\pi(\theta^0,\theta) \leq D(\theta^0,\theta)$.

PROOF: Let R be the rejection region of the test. The power gain is

$$\int_R (f(y|\theta,X) - f(y|\theta^0,X))dy \;\leq\; \int_R |f(y|\theta,X) - f(y|\theta^0,X)|dy$$

and also

$$\begin{aligned}\int_R (f(y|\theta,X) - f(y|\theta^0,X))dy \;&=\; -\int_{\Re^n\setminus R} (f(y|\theta,X) - f(y|\theta^0,X))dy \\ &\leq\; \int_{\Re^n\setminus R} |f(y|\theta,X) - f(y|\theta^0,X)|dy\end{aligned}$$

Add the inequalities and divide by 2 to get

$$\int_R (f(y|\theta,X) - f(y|\theta^0,X))dy \;\leq\; \frac{1}{2}\int_{\Re^n} |f(y|\theta,X) - f(y|\theta^0,X)|dy$$

Thus, $\Delta\pi(\theta^0,\theta) \leq D(\theta^0,\theta)$. Q.E.D.

There exists no test with a power gain larger than the displacement since the displacement does not depend on a test. There exists, however, a test which power gain equals the displacement. This follows directly from the formulation (ii) of the displacement since the right hand side equals the power gain of a test with rejection region U.

The Lemma applied with the decomposition (1) of the power function provides an upper bound on the power function, namely, $\pi(\theta) \leq \pi(\theta^0) + D(\theta^0,\theta)$ for any $\theta^0 \in H^0$. For a given significance level, there exists a universal power upper bound which is valid for any test of the hypothesis. This result is stated in the following Theorem.

THEOREM (Universal power upper bound, PUB): *Let $\alpha$ be the significance level. Then the power function of any test of $H^0$ with size $\leq \alpha$ is bounded by* $PUB(\theta;\theta^0) \equiv \alpha + D(\theta^0,\theta)$*, $\theta^0 \in H^0$. If $H^0$ is composite, then the lowest of the power upper bounds is* $PUB^*(\theta) \equiv \alpha + Inf_{\theta^0 \in H^0}(D(\theta^0,\theta))$.

PROOF: The decomposition (1) of the power function and the Lemma gives

$$\pi(\theta) = \pi(\theta^0) + \Delta\pi(\theta^0,\theta) \leq \pi(\theta^0) + D(\theta^0,\theta).$$

Since the size $\leq \alpha$, this implies that for all $\theta^0 \in H^0$, $\pi(\theta^0) \leq \alpha$. Hence, $\pi(\theta) \leq \alpha + D(\theta^0,\theta) \equiv PUB(\theta;\theta^0)$.

Taking the infimum (INF) on both sides gives

$$\pi(\theta) = INF_{\theta^0 \in H^0}\pi(\theta) \leq INF_{\theta^0 \in H^0}(\alpha + D(\theta^0,\theta)) =$$

$$\alpha + INF_{\theta^0 \in H^0}D(\theta^0,\theta) \equiv PUB^*(\theta)$$

Q.E.D.

The power upper bound does not depend on a specific test. This property implies that the power upper bound is valid for any test and, thus, is universal. For example, the same power upper bound applies to the three classical tests, LR, LM and Wald tests with size 0.05. When testing hypotheses on misspecification the alternative is typically not specified explicitly. In order to apply the power upper bound to such cases, it is necessary to specify a model under the alternative. Hence, it can be useful to specify a class of models and then find the infimum of the power upper bound to express the power against this class of models. It is important to emphasize that to use the form of the power upper bound shown above, a density has to be specified. This, however, can be generalized to measures. In addition, if the assumption on the parameter vector θ to be finite is relaxed, then the power upper bound can be extended to testing in semi- and non-parametric contexts.

The power upper bound is related to uniformly most powerful tests when such tests exist. The displacement equals the power gain of some test. Hence, if a uniformly most powerful test exists, its power gain must equal the displacement for some alternative. Therefore the power function of a uniformly most powerful test will touch the power upper bound for at least one parameter value under the alternative.

## 3. POWER OF TESTS IN A NORMAL REGRESSION MODEL

In order to apply the power upper bound, the displacement has to be evaluated. In this section the displacement is expressed in a simple form for a normal regression model. This implies that the power function of any test in the normal regression model is easy to bound.

Consider the normal regression model

$$Y \sim N(\, G(\beta,X)\, ,\, \Sigma\, ), \qquad \textbf{(2)}$$

where Y is a n-vector, G is any regression function of dimension n, β is a vector of parameters of the regression function and Σ is the n×n covariance matrix. In the notation of the previous section, β and Σ are the parameters in θ and the density function $f(y|\theta,X)$ is the multivariate normal density with mean G(β,X) and covariance Σ.

Let the hypothesis of interest be on the parameters β of the regression function. Hence, the null and alternative hypotheses can be restated as $\beta \in H^0_\beta$ and $\beta \in H^A_\beta$, respectively. Since the hypothesis of interest is on β, Σ is unconstrained and the same under the null and alternative hypothesis.

To apply the power upper bound, the n-dimensional integral in the displacement needs to be evaluated. With a normal density this n-dimensional integral simplifies to a one-dimensional integral, namely, the standard normal cdf, Φ (see proof of Corollary in

Appendix). The displacement is

$$D(\beta^0,\beta;\Sigma) \; = \; 2\Phi\left(\frac{d(\beta^0,\beta;\Sigma)}{2}\right) \; - \; 1$$

where the function $d(\beta^0,\beta;\Sigma)$ is the following standardized distance between the two points $G(\beta,X)$ and $G(\beta^0,X)$ in $\Re^n$:

$$d(\beta^0,\beta;\Sigma) \; = \; \|P^{-1}(G(\beta^0,X) - G(\beta,X))\|$$

where $\|\bullet\|$ is the Euclidean norm and P is a nonsingular matrix satisfying that $PP' = \Sigma$. Hence, the displacement and the power upper bound are straightforward to compute.

If $H^0_\beta$ is composite, then the choice of $\beta^0$ matters for the value of the displacement. Any value of $\beta^0 \in H^0_\beta$ is valid. The value of the displacement can be minimized over $\beta^0 \in H^0_\beta$. Since $\Phi$ is a strictly increasing function, minimizing the displacement is equivalent to minimizing the distance $d(\beta^0,\beta;\Sigma)$. For a given value of $\beta$, $[P^{-1}\, G(\beta,X)]$ is a fixed n-vector. Hence, the minimization problem amounts to minimizing the Euclidean distance between this fixed n-vector and the n-vector $[P^{-1}\, G(\beta^0,X)]$ with respect to $\beta^0 \in H^0_\beta$. This is precisely a nonlinear least squares problem; Estimate $\beta^0$ in a regression of $[P^{-1}\, G(\beta,X)]$ as the dependent variable on the regression function $[P^{-1}\, G(\beta^0,X)]$. Suppose an estimate exists and denote it $\beta^{0*}$. Then the sum of squared residuals, SSR, is

$$SSR = [P^{-1}G(\beta,X) - P^{-1}G(\beta^{0*},X)]'[P^{-1}G(\beta,X) - P^{-1}G(\beta^{0*},X)].$$

The square root of the SSR is equal to the minimized $d(\beta^0,\beta;\Sigma)$. Note, if $H^0_\beta$ is not closed, then $\beta^{0*}$ may not exist. The infimum of $d(\beta^0,\beta;\Sigma)$ with respect to $\beta^0 \in H^0_\beta$, however, still exists. Also note that the nonlinear least squares estimate coincides with the maximum likelihood estimate.

The results on the normal regression model are collected in the following Corollary.

COROLLARY (Power upper bound in a normal regression model): *Let $\alpha$ be the significance level. Then the power function of any test of $H^0_\beta$ in model (2) with size $\leq \alpha$ is bounded by*

$$PUB(\beta,\Sigma;\beta^0) \;=\; \alpha \;+\; 2\Phi\left(\frac{d(\beta^0,\beta;\Sigma)}{2}\right) \;-\; 1$$

*where* $d(\beta^0,\beta;\Sigma) = \{[P^{-1}G(\beta,X) - P^{-1}G(\beta^0,X)]'[P^{-1}G(\beta,X) - P^{-1}G(\beta^0,X)]\}^{0.5}$.

*If* $H^0_\beta$ *is composite, then the lowest of the power upper bounds is*

$$PUB^*(\beta,\Sigma) \;=\; \alpha \;+\; 2\Phi\left(\frac{d^*}{2}\right) \;-\; 1$$

*where* $d^* = INF_{\beta^0 \in H^0_\beta}\, d(\beta^0,\beta;\Sigma)$ *is the square root of the sum of squared residuals from the nonlinear least squares regression of* $P^{-1}G(\beta^0,X)$ *on* $P^{-1}G(\beta,X)$*.*

PROOF: see the Appendix.

The power upper bound is easy to calculate for any test in the normal regression model. In words, evaluate the regression function at β and at a $\beta^0 \in H^0_\beta$, obtaining two n-vectors. Then multiply the difference of the two n-vectors by $P^{-1}$ from the decomposition of the covariance matrix Σ to get a new n-vector. The Euclidean length of this n-vector equals the distance $d(\beta^0,\beta;\Sigma)$.

When $H^0_\beta$ is composite, the Corollary shows how to select the lowest of the power upper bounds. For example, consider the case where the elements of a normal random variable Y are independent with the same variance $\sigma^2$. In this case the nonlinear least squares regression used in estimating the constrained normal regression model can, with one change, be used to find the optimal $\beta^{0*}$: Replace the observations on the dependent variable Y with the 'artificial' observations G(β,X). The artificial observations G(β,X) are, in fact, predicted values of Y if β is the true parameter.

## 4. NONLINEAR COMPOSITE HYPOTHESIS

In practice, the usefulness of the power upper bound depends on how easy it is to calculate and how tight the bound is. These questions are addressed in the following comparison between a Monte Carlo estimated power function and the power upper bound when testing a nonlinear composite hypothesis in the classical normal regression model. In this case, the exact finite sample power function is not readily available but the power upper bound is.

In the classical normal regression model, Gregory and Veall (1985) studied a nonlinear composite hypothesis. The classical normal regression model is

$$Y \sim N(\beta_1 + \beta_2 x_2 + \beta_3 x_3, \sigma^2 I). \quad \textbf{(7)}$$

Let the nonlinear composite null hypothesis of interest be $H^0_\beta$: $\beta_2\beta_3 = 1$ against $\beta_2\beta_3 \neq 1$. To test the hypothesis consider the LR test. The LR test statistic is given by

$$LR = n \ln( SSR^0 / SSR^A ) \quad \textbf{(8)}$$

(see Amemiya (1985)), where $SSR^0$ and $SSR^A$ are the sums of squared residuals under $H^0$ and $H^A$, respectively.

Monte Carlo simulations were used to estimate the finite sample power function of the LR test. The sample size is n = 50. The regressor $x_2$ is the perfect standard normal $x_{2i} = \Phi^{-1}(i/(n+1))$, i = 1,..,n, where $\Phi$ is the standard normal cdf. The other regressor $x_3$ is also a perfect standard normal given by $x_{3i} = (-1)^i x_{2i}$ for $i \leq 25$ and $x_{3i} = (-1)^{i+1} x_{2i}$ for $i > 25$. The correlation between $x_2$ and $x_3$ equals -0.12. The alternatives of interest were chosen to be $\beta_2 > 1$, $\beta_3 = 1$ and the remaining parameters were $\beta_1 = 1$ and $\sigma^2 = 1$. The 0.05 size-corrected critical value was found by searching over the following grid of parameter values: $\sigma \in [0.5,2]$, $\beta_1 \in [0,2]$, $\beta_2 \in [2/3,1.5]$ and $\beta_3 \in [2/3,1.5]$ with the restriction imposed by $H^0$. At each parameter point, 2000 Monte Carlo repetitions were used to find the empirical distribution of the LR test statistic. The unrestricted model is estimated by ordinary least squares and the restricted model by a Gauss-Newton algorithm.

The Monte Carlo estimated power function of the LR test is shown in Figure 2 for different values of $\beta_2$ with the remaining parameters held fixed. The power function increases monotonically to one. For alternatives of $\beta_2$ larger than 2 the power of the LR test is larger than 0.75.

The power upper bound is valid on the power of any test of $H^0$ including the LR test. Since the hypothesis is composite, the lowest power upper bound, $PUB^*$, is calculated. The lowest power upper bound is found by minimizing the distance $d(\beta^0,\beta;\Sigma)$ for $\beta^0 \in H^0_\beta$, where $\beta = (\beta_1,\beta_2,\beta_3)$. The minimum distance $d(\beta^0,\beta;\Sigma)$ equals the square root of the sum of squared residuals from the nonlinear least squares regression of $(\beta_1+\beta_2 x_{2i}+\beta_3 x_{3i})$ as the dependent variable on $\beta_1^0+\beta_2^0x_{2i}+\beta_3^0x_{3i}$ under the restriction $\beta_2^0\beta_3^0 = 1$. The regression is similar to the regression used to find the restricted estimates for the LR test statistic. The only difference is that $y_i$ is replaced by $\beta_1+\beta_2x_{2i}+\beta_3x_{3i}$.

The power upper bound is also shown in Figure 2. The power upper bound mimics the behavior of the power function. The tightness of the bound varies with the parameter values. The largest difference between the upper bound and the power function is 0.26 for $\beta_2$ equal to 1.4. For $\beta_2$ between 1 and 3, the average difference is 0.13. This difference should be judged in the context of the small sample size, n = 50.

The difference between the power upper bound and the power function depends on the specific test. Some tests may have an inferior power performance compared to other tests. For the inferior tests the power upper bound will be a less satisfactory approximation because the power upper bound is valid for any test. In the example, the LR test is not an uniformly most powerful test.

## 5. NONMONOTONIC POWER

In this section the power upper bound is used to detect a nonmonotonic power function. This is illustrated with a LM test of a zero slope coefficient in the exponential regression model. For this model the displacement function decreases to zero for alternatives far away from the null hypothesis. A similar result was obtained by Savin and Würtz (1996) in the context of a binary response model though they used a different technique.

The exponential model has been extensively investigated by Gallant (1987). Consider a simple version of the exponential model

$$Y_i \sim iid\ N(\beta_1 e^{-\beta_2 x_i}, \sigma^2).$$

Let the composite null hypothesis be $\beta_2 = 0$ against the alternative $\beta_2 \neq 0$. With the restriction imposed on $\beta_2$, the exponential model reduces to a linear model which is easy to estimate. Therefore, the practical choice is the Lagrange Multiplier (LM) test since it only requires estimation of the restricted model. The LM test based on the expected information matrix is

$$LM(b^0) = \frac{\partial l(b^0)}{\partial \beta'}\left[E\frac{-\partial^2 l(b^0)}{\partial\beta\partial\beta'}\right]^{-1}\frac{\partial l(b^0)}{\partial\beta}$$

where $\beta = (\beta_1,\beta_2)'$,

$$\frac{\partial l}{\partial \beta}(\beta) = \begin{pmatrix} \frac{1}{\sigma^2}\Sigma_{i=1}^{n}(y_i - \beta_1 e^{-\beta_2 x_i})e^{-\beta_2 x_i} \\ \frac{-1}{\sigma^2}\Sigma_{i=1}^{n}(y_i - \beta_1 e^{-\beta_2 x_i})\beta_1 x_i e^{-\beta_2 x_i} \\ \frac{-n}{2\sigma^2} + \frac{1}{2\sigma^4}\Sigma_{i=1}^{n}(y_i - \beta_1 e^{-\beta_2 x_i}) \end{pmatrix}$$

is the score vector and

$$E\frac{-\partial^2 l}{\partial\beta\partial\beta'}(\beta) = \frac{1}{\sigma^2}\sum_{i=1}^{n}\begin{bmatrix} e^{-2\beta_2 x_i} & -\beta_1 x_i e^{-2\beta_2 x_i} & 0 \\ -\beta_1 x_i e^{-2\beta_2 x_i} & (\beta_1)^2 x_i^2 e^{-2\beta_2 x_i} & 0 \\ 0 & 0 & \frac{n}{2\sigma^4} \end{bmatrix}$$

is the expected information matrix. The LM test statistic is evaluated at $b^0$, which is the estimate of β in the restricted model and $(s^0)^2$, which is the constrained estimate of $\sigma^2$.

To find the power of the LM test a Monte Carlo experiment was performed. The Monte Carlo experiment is conducted with 2000 replications. The regressor x is the perfect normal N(2.5,1) given by $x_i = 2.5 + \Phi^{-1}(i/(n+1))$, i = 1,..,n, where Φ is the standard normal cdf. The number of observations is n = 50 which implies that all the values of the regressor are positive. The standard deviation σ is set to 1. The alternatives of interests are increasing values of $\beta_2$ with $\beta_1$ and $\sigma^2$ being held fixed. The LM test is size-corrected for $\beta_1 \in [1,3]$ and $\sigma^2 \in [0.5,2]$. Under the alternative, $\beta_1$ is held fixed at 2 and $\sigma^2$ at 1.

The power of the LM test is nonmonotonic in the alternatives of interest. This is illustrated in Figure 3. The power of the LM test peaks at $\beta_2 = 0.5$. At the peak, the power is approximately 0.6. As $\beta_2$ increases beyond 0.6 the power of the LM test declines towards 0.05, the size of the test.

The lowest power upper bound, $PUB^*$, is calculated easily in this example. Using the Corollary, the minimum $d(\beta^0,\beta;\sigma^2 I)$ can be calculated from the regression of $\beta_1 e^{-\beta_2 x_i}$, as the dependent variable, on $\beta_1^0 e^{-\beta_2^0 x_i}$ with the restriction $\beta_2^0 = 0$. The estimate of $\beta_1^0$ equals the average of $\beta_1 e^{-\beta_2 x_i}$, for i = 1,..,n. From this regression, the square root of the sum of squared residuals equals the minimum value of $d(\beta^0,\beta;\sigma^2 I)$.

The power upper bound can be used in advance to restrict the shape of the power function of the LM test. In Figure 3 the lowest power upper bound is illustrated together with the power of the LM test. The power upper bound first increases, peaks and then declines. The power upper bound peaks approximately at the same alternative as the power function of the LM test. The power upper bound and the power of the LM test behave qualitatively the same.

The power upper bound defines a two-dimensional surface in $\Re^3$. The power upper bound shown in Figure 3 is a slice of this surface along the ray ($\beta_1 = 2$, $\beta_2 \in [0,4]$). The contours of the power upper bound are the same as the contours of the displacement function adjusted by the significance level $\alpha$. The contours of the displacement surface are shown in Figure 4. The contours are denoted by isodisplacement curves. The isodisplacement curves show that the power of any test, not just the LM test, cannot have higher power than the significance level for alternatives far away from the null hypothesis. The isodisplacement curves also illustrate the influence of the value of the nuisance parameter $\beta_1$. If the true $\beta_1$ is close to zero, no test can have a power gain larger than, say, 20% even for alternatives far away from the null hypothesis.

## 6. CONCLUSION

The finite sample power function of a test was shown to be bounded by a nontrivial power upper bound. The power upper bound depends on what here is called the displacement. The displacement measures the difference between two densities specified by the model, independent of the test. This implies that the power upper bound is valid for any test.

In a normal regression model, the power upper bound has a simple form. The simple form makes the power upper bound easy to apply to tests of composite hypotheses. It also provides which may detect nonmonotonic power functions.

## APPENDIX. PROOF OF THE COROLLARY

The proof is established by transforming the displacement into a function of standard normal densities.

The displacement is

$$D(\beta^0,\beta;\Sigma) = \frac{1}{2}\int_{\Re^n} |f(y|\beta^0,\Sigma,X) - f(y|\beta,\Sigma,X)|\ dy$$

Decompose Σ into Σ = PP', where P is a nonsingular matrix. Then

$$f(y|\beta,\Sigma,X) = |P^{-1}|\ g(P^{-1}y - P^{-1}G(\beta,X)). \qquad \textbf{(9)}$$

where $g(z) = \Pi^n_{i=1}\ \phi(z_i)$ and φ is the standard normal density. Using (9) the displacement can be written as

$$D(\beta^0,\beta;\Sigma) = \frac{1}{2}\int_{\Re^n} |P^{-1}||g(P^{-1}y - P^{-1}G(\beta^0,X)) - g(P^{-1}y - P^{-1}G(\beta,X))|dy$$

By the change of variables $y = Pz + G(\beta^0,X)$ then

$$D(\beta^0,\beta;\Sigma) = \frac{1}{2}\int_{\Re^n} |g(z) - g(z - P^{-1}[G(\beta,X) - G(\beta^0,X)])|\ dz$$

The displacement can be simplified by transforming the variables such that the two densities in (10) are evaluated at the same coordinate z except for the first coordinate of z. Let H be a full rank n×n orthonormal matrix with the first column of H equal to the vector $P^{-1}(G(\beta,X)-G(\beta^0,X))/\|P^{-1}(G(\beta,X)-G(\beta^0,X))\|$, where ‖•‖ is the Euclidean norm. Then $H'P^{-1}(G(\beta,X)-G(\beta^0,X)) = (d(\beta^0,\beta;\Sigma),0,..,0)'$ with $d(\beta^0,\beta;\Sigma) = \|P^{-1}(G(\beta,X)-G(\beta^0,X))\|$. Using H to change variables from z to u = H'z and substituting for g, the displacement is

$$D(\beta^0,\beta;\Sigma) = \frac{1}{2}\int_{\Re^n} |\ \phi(u_1) - \phi(u_1 - d(\beta^0,\beta;\Sigma))\ |\ \Pi^n_{i=2}\ \phi(u_i)\ du$$

The displacement can be further simplified by using an equivalent expression for the displacement. The displacement (11) is written according to the definition (i). The displacement can also be expressed by (ii). To apply expression (ii) the set U must be derived. The set U is given by

$$U = \{\ u\ |\ g(u) - g(u-(d(\beta^0,\beta;\Sigma),0,..,0)') > 0\}$$
$$= \{\ u\ |\ u_1 < d(\beta^0,\beta;\Sigma)/2,\ u_i \in \Re \text{ for } i = 2,..,n\ \}.$$

This follows from normality. Then the displacement can be written as

$$D(\beta^0,\beta;\Sigma) = \int_{-\infty}^{d(\beta^0,\beta;\Sigma)/2}\phi(u_1)\int_{\Re}\phi(u_2)\cdots\int_{\Re}\phi(u_n)du_n\cdots du_2du_1$$

$$- \int_{-\infty}^{d(\beta^0,\beta;\Sigma)/2}\phi(u_1 - d(\beta^0,\beta;\Sigma))\int_{\Re}\phi(u_2)\cdots\int_{\Re}\phi(u_n)du_n\cdots du_2du_1$$

which simplifies to

$$D(\beta^0,\beta;\Sigma) = \Phi(\frac{d(\beta^0,\beta;\Sigma)}{2}) - \Phi(\frac{-d(\beta^0,\beta;\Sigma)}{2})$$

$$= 2\Phi(\frac{d(\beta^0,\beta;\Sigma)}{2}) - 1$$

Now, apply the Theorem to get the power upper bound.

If the null hypothesis is composite and the displacement depends on $\beta^0 \in H^0_\beta$, then the lowest of all the possible PUB's can be found. In order to do that, the displacement needs to be minimized for a given $(\beta,\Sigma)$ and a $\beta^0 \in H^0_\beta$. This is equivalent to minimizing $d(\beta^0,\beta;\Sigma)$ with respect to $\beta^0 \in H^0_\beta$. The function $d(\beta^0,\beta;\Sigma)$ measures the Euclidean distance between the two n-vectors $P^{-1}G(\beta^0,X)$ and $P^{-1}G(\beta,X)$. Only $P^{-1}G(\beta^0,X)$ changes with $\beta^0$ so the problem is to find the smallest distance to the fixed vector $P^{-1}G(\beta,X)$. But this is exactly what the nonlinear least squares estimator does. Hence, the optimal $\beta^{0*} \in H^0_\beta$ is the nonlinear least squares estimate $\beta^0$ in the regression of $P^{-1}G(\beta,X)$ on $P^{-1}G(\beta^0,X)$. Since the residuals from this regression equal $[P^{-1}G(\beta,X) - P^{-1}G(\beta^{0*},X)]$, the square root of the sum of squared residuals is the minimum value of $d(\beta^0,\beta;\Sigma)$.

Q.E.D.

## REFERENCES

Amemiya, T. (1985): *Advanced Econometrics.* Harvard University Press, Cambridge.

Devroye, L. and L. Györfi (1985): *Nonparametric Density Estimation. The $L_1$ View.* John Wiley and Sons, New York.

Dufour, J. M. (1989): "Nonlinear hypotheses, inequality restrictions, and non-nested hypotheses: Exact simultaneous tests in linear regressions," *Econometrica* 57, 335-355.

Dufour, J. M. and M. L. King (1991): "Optimal invariant tests for the autocorrelation coefficient in linear regressions with stationary or nonstationary AR(1) errors," *Journal of Econometrics* 47, 115-143.

Elliott, G, T. J. Rothenberg and J. H. Stock (1996): "Efficient tests for an autoregressive unit root", *Econometrica* 64, 763-812.

Gallant, R. (1987): *Nonlinear statistical models*. John Wiley and Sons, New York.

Gregory, A. W. and M. R. Veall (1985): "Formulating Wald tests of nonlinear restrictions," *Econometrica,* 53, 1465-1468.

Keuzenkamp, H. A. and J. R. Magnus (1995): "On tests and significance in econometrics," *Journal of Econometrics,* 67, 5-24.

King, M. L. (1987): "Towards a theory of point optimal testing," *Econometric Reviews* 6, 169-255.

Lehmann, E. L. (1959): *Testing statistical hypotheses*. John Wiley and Sons, New York.

McCloskey, D. and S. T. Ziliak (1996): "The standard error of regressions," *The Journal of Economic Literature,* March 1996, 97-114.

Nelson, F.D. and N. E. Savin (1990): "The danger of extrapolating asymptotic local power," *Econometrica* 58, 977-981.

Neyman, J and E. S. Pearson (1933a): "On the problem of the most efficient tests of statistical hypotheses," *Phil. Trans. Roy. Soc. A*, 231, 289-337.

Neyman, J and E. S. Pearson (1933b): "The testing of statistical hypotheses in relation to probabilities *a priori*," *Proc. Camb. Phil. Soc*, 24, 492-510.

Pitman, E. J. G. (1979): *Some basic theory for statistical inference.* Chapman and Hall, London.

Plott, C. R., J. Wit and W. C. Yang (1996): "Parimutuel betting markets as information aggregation devices: Experimental results," *Mimeo,* California Institute of Technology, Pasadena, CA.

Rudin, Walter (1987): *Real and complex analysis*. McGraw-Hill, Inc. New York.

Savin, N. E. and A. H. Würtz (1996): "Power of tests in binary response models," *Working Paper Series #96-06*, University of Iowa, Iowa.

Zacks, S. (1971): *The theory of statistical inference.* John Wiley & Sons, Inc., New York.

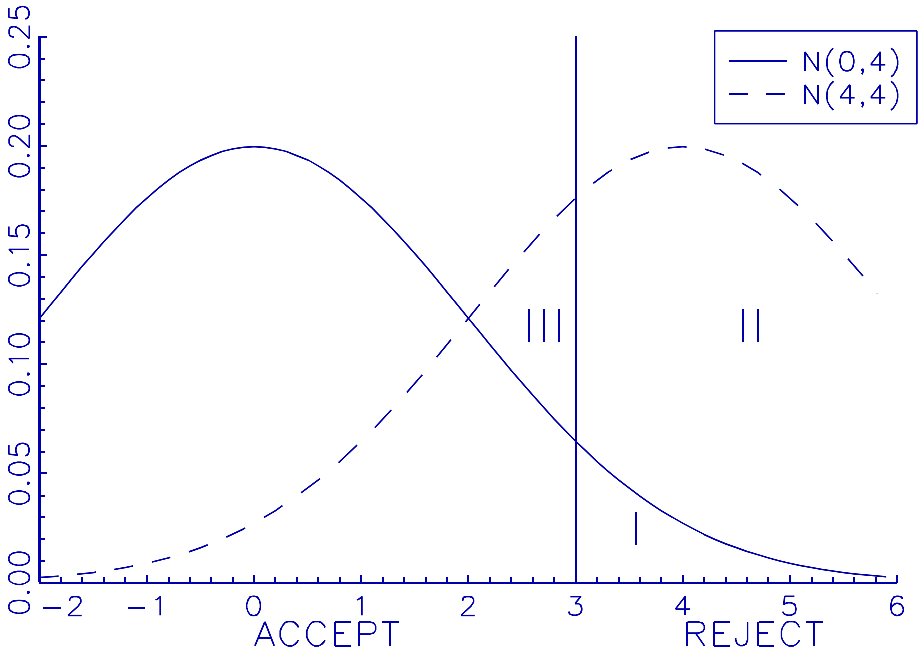


Figure 1. Displacement and change in rejection probability. The two densities over the sample space are N(0,4) and N(4,4). The sample size is n = 1. The hypothesis is the mean $\mu = 0$. The test rejects the hypothesis if the observation is larger than 3.

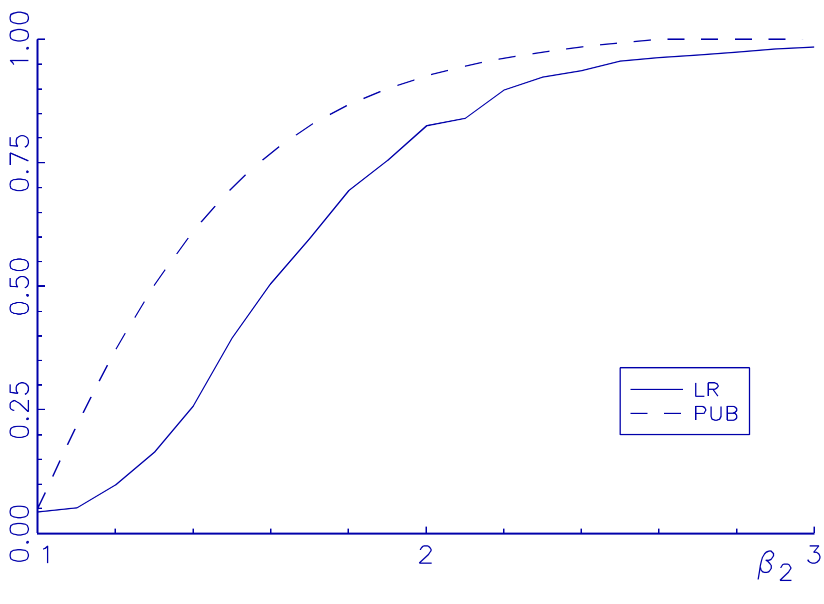


Figure 2. Power of a 0.05 size LR test and the lowest power upper bound. The hypothesis is $\beta_2\beta_3 = 1$ against $\beta_2\beta_3 \neq 1$ in the linear normal regression model. The regressors $x_2$ and $x_3$ are perfect standard normals and n = 50, $\beta_1 = 1$, $\beta_3 = 1$ and $\sigma^2 = 1$.

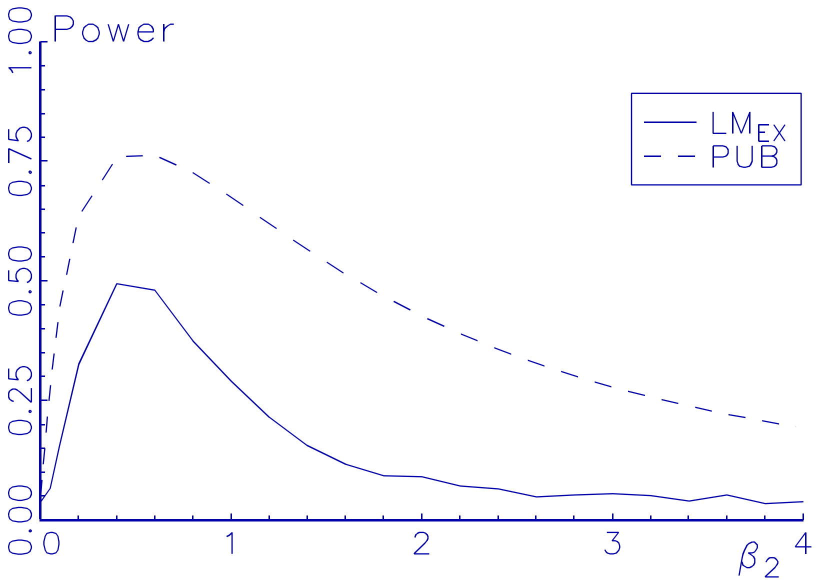


Figure 3. Power of a 0.05 size LM test and the power upper bound. The hypothesis is $\beta_2 = 0$ against $\beta_2 \neq 0$ in the exponential model with a perfect normal N(2.5,1) regressor, n = 50, $\beta_1 = 2$ and $\sigma^2 = 1$.

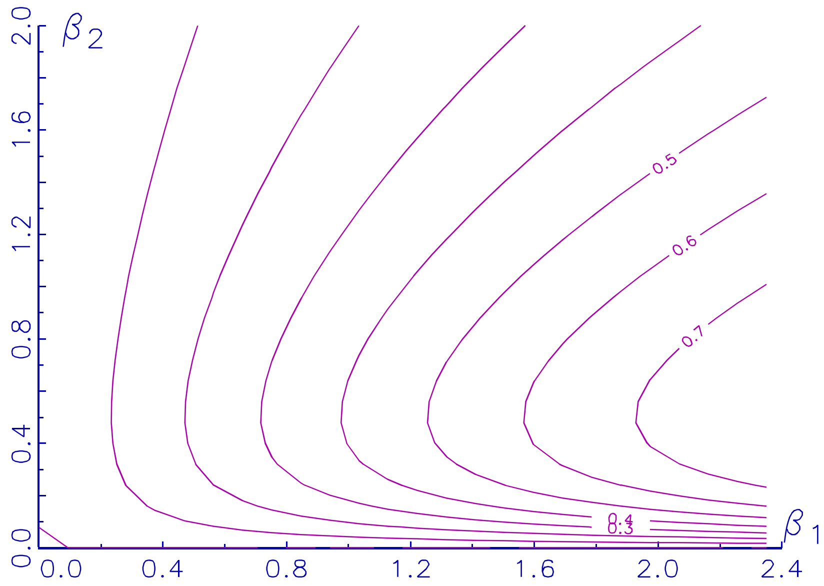


Figure 4. Isodisplacement curves for the null hypothesis $\beta_2 = 0$ against $\beta_2 \neq 0$ in the exponential model with a perfect normal N(2.5,1) regressor and n = 50.