FEA Working Paper No. 2001-8

WALD OR NULL WALD TEST? A DECISION GUIDE FOR TESTING PARAMETER SIGNIFICANCE

Kim-Leng Goh

Department of Applied Statistics
Faculty of Economics & Administration
University of Malaya
50603 Kuala Lumpur
MALAYSIA

E-Mail: g7klgoh@umcsd.um.edu.my

July 2001

WALD OR NULL WALD TEST? A DECISION GUIDE FOR TESTING PARAMETER SIGNIFICANCE

Abstract

The Wald test suffers from the problem of non-monotonic power function in some finite-sample applications. As an alternative, Goh and King (2000) proposed the null Wald (NW) test that retains the computational convenience of the Wald test while behaving monotonically. Monte Carlo results show that even in situations where the Wald test suffers power non-monotonicity, regions of the parameter space where its power performance is more desirable than the NW test still exist, suggesting preference for the use of the Wald test. This paper proposes a simple diagnostic check for discerning if the Wald test behaves monotonically in the neighbourhood of the parameter space from which the sample data are drawn. Monte Carlo simulations show that this method works rather well for detecting situations where the Wald test can be safely applied. Based on the outcome of this diagnostic check, the paper further proposes a two-stage Wald (2SW) procedure. Although not a clear winner in all circumstances, simulation results demonstrate that the 2SW procedure is a better alternative than relying on either the Wald or NW test totally.

1. Introduction

It is a common practice to examine the significance of a parameter using a t-test, or its equivalent, the Wald test in the case of non-linear models. Unfortunately, the power function of the Wald test can behave non-monotonically for small-sample applications that involve typically non-linear models. Among others, this evidence is reported by Hauck and Donner (1977), Vaeth (1985), Nelson and Savin (1988, 1990), and Goh and King (2000). With this difficulty, the test performs expectably with an increasing power function at a sequence of local alternatives, but contrary to its asymptotic behaviour, the function declines eventually to zero with further departure from the null hypothesis. This puts a researcher in a quandary when a non-rejection is found. It could indicate that the data do not provide sufficient evidence to reject the null hypothesis, but it could just as well imply that the data are grossly inconsistent with the null.

As a solution, Laskar and King (1997) modified the Wald test for single-parameter testing problems and named it as the null Wald (NW) test. This was extended by Goh and King (2000) for multiparameter situations involving nuisance parameters in the test.

The Wald test is popular for testing significance of parameters as it requires only information from the unrestricted model, whereas an alternative such as the likelihood ratio (LR) test uses estimations from both the restricted and unrestricted models. While the NW test was developed to retain the computational convenience of the Wald test, it does not always yield better power properties. How well or poorly the Wald test performs in terms of power depends not only on the testing problem, but also which region of the parameter space the sample data are drawn. For the same testing problem, the Wald test may suffer power non-monotonicity in some regions of the parameter space, but it can have desirable power properties in other regions where its power is monotonic (see Goh, 1999). Therefore to test for the significance of a parameter, an immediate challenge is to discern whether the Wald test has good power behaviour before an alternative test is considered. This paper proposes a quick and easy diagnostic check for non-monotonicity in the neighbourhood of the parameter space from which the sample data are drawn. A simulation study is conducted to examine the performance of this check and an application is given.

2. Diagnostic Check for Non-Monotonicity

Let θ, a k x 1 vector, represent the parameter space. Suppose n independent observations, y_t, t = 1, 2, ..., n, are from a stochastic process with the density function $\phi(y_t | x_t, \theta)$ where x_t is a vector of exogenous variables. The log-likelihood function is given by:

$$l(\theta) = \Sigma \ln \phi(y_t | x_t, \theta) \qquad (1)$$

The parameter space is partitioned into two sub-vectors $\theta = (\beta', \gamma')'$ where β is r x 1. The hypotheses of interest are

$$H_0: \beta = 0 \text{ against } H_a: \beta \neq 0 \qquad (2)$$

where the nuisance parameter vector γ is left unconstrained in the testing problem. Let $V(\theta)$ be a consistent estimator of the limiting variance-covariance matrix of the maximum likelihood (ML) estimator evaluated at θ. The Wald test rejects H_0 if the test statistic

$$W = \hat{\beta}'(RV(\hat{\theta})R')^{-1}\hat{\beta} \qquad (3)$$

is larger than $\chi^2_{r,\alpha}$ where all parameters denoted with a ^ refer to their unconstrained ML estimators, $R = (I_r : 0)$ is an $r \times k$ matrix, I_r is an r-dimensional identity matrix, χ^2_r represents a central chi-square distribution with r degrees of freedom and α is the level of significance.

Let $\theta_\lambda = (\lambda\beta', \gamma')'$ where λ is a scalar greater than one. If H_0 is false, θ_λ is in a region that is further away from the null space than the actual sample point itself. The alternative point θ_λ lies along a ray that projects from the null space to the sample data point, on a plane cutting through the parameter space at given γ values. The true parameter values are not known, but θ_λ can be approximated by $\hat{\theta}_\lambda = (\lambda\hat{\beta}', \hat{\gamma}')'$. Let the Wald statistic evaluated at $\hat{\theta}_\lambda$ be

$$W_\lambda = \lambda\hat{\beta}'(RV(\hat{\theta}_\lambda)R')^{-1}\lambda\hat{\beta} \qquad (4)$$

The non-monotonic behaviour of the Wald power arises from a test statistic that declines as we move further away from H_0. Thus, if W_λ is smaller than W, the Wald test might suffer from non-monotonicity in the neighbourhood of θ. The diagnostic check is given by:

$$\Delta W = W_\lambda - W \qquad (5)$$

where a negative ΔW indicates that the Wald test may exhibit power non-monotonicity. On the contrary, a positive value of ΔW indicates that the sample data are in the region of parameter space where the Wald power is monotonic.

3. Monte Carlo Experiment

A simulation experiment based on the logit and Tobit models is conducted to evaluate how well ΔW performs. Consider a two-regressor logit model defined as follows:

$$P(y_t = 1) = \Lambda(x_t'\theta) = \left[1 + \exp(-x_t'\theta)\right]^{-1} \qquad (6)$$

where $\theta = (\beta, \gamma)'$. The log-likelihood function is:

$$l(\theta) = \sum_t \left[y_t \ln \Lambda(x_t'\theta) + (1 - y_t) \ln \Lambda(-x_t'\theta)\right] \qquad (7)$$

Consider a two-regressor Tobit model given by:

$$y_t = x_t'\varphi + \varepsilon_t \text{ if } x_t'\varphi + \varepsilon_t > 0, \text{ and } y_t = 0 \text{ otherwise} \qquad (8)$$

where $\varepsilon_t \sim IN(0, \sigma^2)$ and $\varphi = (\beta, \gamma)'$. In this case, $\theta = (\varphi', \sigma^2)'$. The log-likelihood function is:

$$l(\theta) = -\frac{1}{2}\sum_{y>0}\left[\ln(2\pi) + \ln\sigma^2 + (y_t - x_t'\varphi)^2/\sigma^2\right] + \sum_{y=0}\ln\left[1 - F_t\right] \qquad (9)$$

where F_t is the standard normal distribution function evaluated at $z_t = \dfrac{x_t'\varphi}{\sigma}$.

The NW and LR tests are also included in the experiment for power comparison. The NW test statistic is:

$$NW = \hat{\beta}'(RV(\hat{\theta}_0)R')^{-1}\hat{\beta} \qquad (10)$$

where $\hat{\theta}_0 = (0, \hat{\gamma})'$ for the logit model and $\hat{\theta}_0 = (0, \hat{\gamma}, \hat{\sigma}^2)'$ for the Tobit model. The LR test statistic is:

$$LR = 2[l(\hat{\theta}) - l(\tilde{\theta})] \qquad (11)$$

where $\tilde{\theta}$ is the constrained ML estimator. Under H_0, both the test statistics follow a χ^2_r distribution asymptotically.

Two design matrices are used for each of the model. Those for the logit model are:

X1: Two independent series of independent N(0, 1) random drawings, and
X2: Monthly Australian official total capital in the balance of payments and reserve assets in the Australian Reserve Bank beginning September 1987 ($'000m).

The design matrices for the Tobit model are:

X3: Two independent series of independent [0, 1] uniform random drawings, and
X4: Monthly Australian exports of fresh, chilled or frozen beef with bone (10^7 kg) and live sheep ('000m) beginning September 1988.

The nuisance parameters are fixed at $\gamma = 0.1$ for both the models and $\sigma^2 = 1$ for the Tobit model. The simulation is based on 1,000 replications and the sample size is set at 30. All the tests are performed at the 5 per cent significance level. Two λ values, 2.5 and 5.0, are used. Only single parameter restrictions are considered in this experiment. (See Goh, 1999 for more extensive testing problems).

4. Results

The power curves of the Wald, NW and LR tests for the logit model are reported in Table 1. The power curve of the Wald test increases initially but declines eventually with further departure from H_0 and this happens at both tails. On contrary, the NW and LR tests have monotonic power functions. In regions of the parameter space where the Wald test is monotonic, the proportion of ΔW that is negative is very low. This increases to one or close to one in the regions where the Wald test is non-monotonic, suggesting that the diagnostic test is almost certain to pick up the problem when there is one. Using a larger λ value produces a greater tendency to obtain a negative ΔW. This is expected as the distance between $\hat{\theta}_\lambda$ and H_0 increases with larger λ values, and the test statistic is displaced sooner into the region of non-monotonicity if the problem exists when λ is factored in. Thus, the diagnostic check is more conservative with the use of a larger λ value.

To further assess the effectiveness of the diagnostic check, a two-stage Wald (2SW) test is used where a Wald test is performed if $\Delta W > 0$, and a NW test is performed if $\Delta W < 0$. The results show that the power curve of the 2SW test mimics that of the Wald test in the regions where the Wald power is monotonic, but inherits the properties of the NW test otherwise.

The results for the Tobit model are given in Table 2. The Wald test is only non-monotonic in power at the left tail. At alternatives to the right of H_0, the Wald test has the highest power and is the most desirable test. In this region, ΔW almost never indicates a negative sign. The right-tail power function of the 2SW test is exactly similar to that of the Wald test. To the left of H_0, the proportion of negative ΔW is close to one at the alternatives where the Wald power is non-monotonic. Here, the 2SW power curve picks up the properties of the NW test.

Table 1: The Power Function of the Wald, NW, LR and 2SW Test and the Proportion of Negative ΔW for Testing H_0: β = 0 in the Two-Regressor Logit Model

β	Asymptotic Tests Wald	NW	LR	Proportion of negative ΔW λ=2.5	λ=5.0	Two-stage Wald (2SW) Test λ=2.5	λ=5.0
Design Matrix: X1							
0.00	0.040	0.070*	0.058	0.001	0.006	0.040	0.040
-0.25	0.100	0.161	0.132	0.007	0.024	0.100	0.100
-0.40	0.186	0.265	0.226	0.016	0.058	0.186	0.186
-0.70	0.440	0.572	0.520	0.078	0.213	0.440	0.440
-1.00	0.729	0.841	0.792	0.228	0.469	0.729	0.729
-1.50	0.954	0.980	0.975	0.591	0.837	0.956	0.956
-4.00	0.793	1.000	1.000	0.999	1.000	1.000	1.000
-5.00	0.645	1.000	1.000	1.000	1.000	1.000	1.000
-6.00	0.525	1.000	1.000	1.000	1.000	1.000	1.000
0.25	0.079	0.139	0.111	0.005	0.019	0.079	0.079
0.40	0.155	0.245	0.200	0.017	0.057	0.155	0.155
0.70	0.421	0.550	0.497	0.070	0.211	0.421	0.421
1.00	0.674	0.808	0.752	0.210	0.436	0.674	0.674
1.50	0.937	0.980	0.964	0.543	0.810	0.938	0.938
4.00	0.825	1.000	1.000	1.000	1.000	1.000	1.000
5.00	0.672	0.999	1.000	1.000	1.000	0.999	0.999
6.00	0.533	1.000	1.000	1.000	1.000	1.000	1.000
Design Matrix: X2							
0.00	0.021*	0.079*	0.057	0.000	0.003	0.021*	0.021*
-0.25	0.047	0.134	0.096	0.001	0.011	0.047	0.047
-0.45	0.121	0.262	0.200	0.004	0.030	0.121	0.121
-0.85	0.301	0.519	0.436	0.047	0.147	0.301	0.301
-1.50	0.687	0.886	0.810	0.236	0.487	0.687	0.687
-2.50	0.924	0.985	0.981	0.654	0.866	0.932	0.932
-5.50	0.861	0.999	1.000	0.996	1.000	0.999	0.999
-6.00	0.818	0.998	1.000	0.997	1.000	0.998	0.998
-6.20	0.806	0.998	1.000	0.997	1.000	0.998	0.998
0.25	0.041	0.131	0.091	0.004	0.017	0.041	0.041
0.45	0.089	0.215	0.157	0.009	0.037	0.089	0.089
0.85	0.265	0.485	0.392	0.041	0.125	0.268	0.266
1.50	0.620	0.864	0.782	0.224	0.457	0.622	0.621
2.50	0.918	0.987	0.977	0.681	0.873	0.929	0.928
5.50	0.817	0.997	1.000	0.997	1.000	0.997	0.997
6.00	0.777	0.998	1.000	1.000	1.000	0.998	0.998
6.20	0.766	0.998	1.000	1.000	1.000	0.998	0.998

Notes: 1,000 replications are used in the simulation.
The significance level is 0.05 and sample size is 30.
*Size is significantly different from 5 per cent at the 0.01 level.

Table 2: The Power Function of the Wald, NW, LR and 2SW Test and the Proportion of Negative ΔW for Testing H_0: $\beta = 0$ in the Two-Regressor Tobit Model

β	Asymptotic Tests			Number of rep.(a)	Proportion of negative ΔW		Two-stage Wald (2SW) Test	
	Wald	NW	LR		λ=2.5	λ=5.0	λ=2.5	λ=5.0
Design Matrix: X3								
0.00	0.049	0.054	0.050	1000	0.013	0.056	0.053	0.068*
-0.35	0.063	0.134	0.106	997	0.064	0.199	0.081	0.141
-0.50	0.068	0.195	0.140	997	0.109	0.318	0.120	0.198
-0.80	0.162	0.399	0.310	998	0.293	0.549	0.299	0.399
-1.00	0.219	0.510	0.411	997	0.451	0.698	0.443	0.510
-1.30	0.301	0.695	0.593	996	0.671	0.854	0.654	0.695
-1.80	0.537	0.899	0.830	998	0.884	0.960	0.879	0.899
-2.00	0.286	0.917	0.848	984	0.937	0.980	0.914	0.917
-3.20	0.100	0.986	0.933	881	0.998	1.000	0.986	0.986
0.35	0.157	0.119	0.129	1000	0.002	0.012	0.157	0.162
0.50	0.231	0.184	0.198	1000	0.000	0.005	0.231	0.232
0.80	0.455	0.390	0.407	1000	0.000	0.000	0.455	0.455
1.00	0.629	0.568	0.581	1000	0.000	0.000	0.629	0.629
1.30	0.820	0.793	0.796	1000	0.000	0.000	0.820	0.820
1.80	0.977	0.972	0.973	1000	0.000	0.000	0.977	0.977
2.00	0.990	0.988	0.988	1000	0.000	0.000	0.990	0.990
3.20	1.000	1.000	1.000	1000	0.000	0.000	1.000	1.000
Design Matrix: X4								
0.00	0.060	0.067	0.062	1000	0.014	0.052	0.063	0.082*
-0.35	0.070	0.152	0.116	999	0.065	0.218	0.091	0.156
-0.50	0.094	0.238	0.168	999	0.124	0.333	0.146	0.238
-0.80	0.182	0.420	0.341	999	0.327	0.603	0.336	0.421
-1.00	0.235	0.573	0.464	999	0.493	0.751	0.487	0.573
-1.20	0.283	0.689	0.588	997	0.661	0.855	0.645	0.689
-2.00	0.253	0.932	0.817	977	0.964	0.992	0.932	0.932
-2.50	0.232	0.982	0.892	946	0.996	0.997	0.982	0.982
-3.00	0.175	0.992	0.934	862	0.999	0.999	0.991	0.991
0.35	0.149	0.121	0.128	1000	0.000	0.005	0.149	0.152
0.50	0.227	0.181	0.193	1000	0.000	0.002	0.227	0.229
0.80	0.478	0.421	0.435	1000	0.000	0.000	0.478	0.478
1.00	0.647	0.594	0.606	1000	0.000	0.000	0.647	0.647
1.20	0.807	0.763	0.771	1000	0.000	0.000	0.807	0.807
2.00	0.991	0.988	0.990	1000	0.000	0.000	0.991	0.991
2.50	1.000	1.000	1.000	1000	0.000	0.000	1.000	1.000
3.00	1.000	1.000	1.000	1000	0.000	0.000	1.000	1.000

Notes: 1,000 replications are used in the simulation.
The significance level is 0.05 and sample size is 30.
*Size is significantly different from 5 per cent at the 0.01 level.

(a) refers to the number of replications not affected by 100% censoring in the experiment.

5. An Application

A logit model for selected variables (defined in Table 3) from McManus's (1985) study on the deterrent effect of capital punishment is used to illustrate the non-monotonicity problem of the Wald test and the application of the diagnostic check. The data is for 44 states in US in 1950. The dependent variable is whether capital punishment was introduced in a state. The results are shown in Table 3. The significance of the slope for each of the explanatory variables as well as the joint significance of the three slopes is tested. Based on the Wald test, none of the null hypotheses is rejected. The diagnostic check, however, indicated a negative sign over the entire range of selected λ values when H_0: $\beta_3 = 0$ and H_0: $\beta_1=\beta_2 =\beta_3=0$ are tested. Turning to the alternative tests, both these hypotheses are strongly rejected using the NW and LR tests. In the other cases where ΔW is positive for all or majority of the λ values, non-monotonicity is not a problem and the three tests yield consistent results.

Table 3: The Logit Model for determination of propensity to have capital punishment

$$P(y_t = 1) = 1/[1+\exp(-(\beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \beta_3 X_{3t}))]$$

	Constant	X_1	X_2	X_3	
Coefficient	0.388	-2.759	0.008	20.809	
s.e.	1.371	3.017	0.007	9864.313	
Test Results (a)					
H_0	$\beta_0 = 0$	$\beta_1 = 0$	$\beta_2 = 0$	$\beta_3 = 0$	$\beta_1=\beta_2 =\beta_3=0$
Wald	0.777	0.361	0.273	0.998	1.000
NW	0.766	0.466	0.235	0.000	0.000
LR	0.778	0.364	0.258	0.002	0.013
ΔW					
$\lambda = 1.2$	0.032	0.394	0.291	-0.000	-0.000
$\lambda = 1.5$	0.087	1.053	0.567	-0.000	-0.000
$\lambda = 2.0$	0.192	2.033	0.611	-0.000	-0.000
$\lambda = 2.5$	0.304	2.571	0.385	-0.000	-0.000
$\lambda = 3.0$	0.412	2.736	0.101	-0.000	-0.000
$\lambda = 5.0$	0.701	2.170	-0.716	-0.000	-0.000

Notes: (a) p-values are reported.
$Y = 1$ if the state has capital punishment, 0 otherwise.
X_1= ratio of number of convictions to number of murders in 1950.
X_2= median time served in months of convicted murderers released in 1951.
X_3= 1 if southern states, 0 otherwise.

6. Conclusion

A few asymptotic tests are available for testing parameter significance in non-linear models, but the Wald test is the easiest to apply. However, the Wald test can lead to misleading inferences when its power function is non-monotonic. This paper proposes a simple diagnostic check for pretesting if this problem is present before the Wald test is applied. Simulation results show that the check is effective in detecting the problem and useful for avoiding the pitfall of power non-monotonicity in the Wald test. Incorporating an additional stage of pretesting for non-monotonicity in hypothesis testing procedures for deciding on the appropriate choice of test can lead to more desirable power properties.

References

Goh, K.L., 1999, Some solutions to small-sample problems of Wald tests in econometrics, Unpublished PhD thesis, Monash University, Australia.

Goh, K.L. and K.L. King, 2000, A solution to power problems of the Wald test in non-linear models, *Pakistan Journal of Statistics* (Saleh-Aly Special Edition) 16(2), 195-205.

Hauck, Jr., W.W. and A. Donner, 1977, Wald's test as applied to hypotheses in logit analysis, *Journal of the American Statistical Association* 72, 851-853, and, 1980, Corrigendum, 75, 482.

Laskar, M.R. and M.L. King, 1997, Modified Wald test for regression disturbances, *Economics Letters* 56, 5-11.

McManus, W.S., 1985, Estimates of the deterrent effect of capital punishment: The importance of the researcher's prior beliefs, *Journal of Political Economy* 93, 417-425.

Nelson, F.D. and N.E. Savin, 1988, The non-monotonicity of the power function of the Wald test in non-linear models, Working Paper Series No. 88-7, Department of Economics, University of Iowa.

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

Vaeth, M., 1985, On the use of Wald's test in exponential families, *International Statistical Review* 53, 199-214.