A NOTE ON THE ASYMPTOTIC EXPANSION OF THE T-TEST FOR NONLINEAR HYPOTHESIS TESTS

Kazumitsu Nawata
Department of Social and International Relations
University of Tokyo

Abstract

The t-test is the most widely used test for the single equation hypothesis. When the hypothesis is nonlinear and given by H_0: $h(\beta) = 0$, the test statistic is calculated as $t = h(\hat{\beta})$ / (estimated standard deviation). However, since the standard deviation is calculated from the asymptotic distribution, the t-test does not behave well for some nonlinear hypotheses. Gregory and Veall [1985] compared the hypotheses: H_0: $\beta_1 \beta_2 = 1$ and H_0^*: $\beta_1 = 1/\beta_2$. Although the two hypotheses are mathematically identical, the t-test based on the second hypothesis was shown to perform poorly. Lafontaine and White [1986] considered H_0: $\beta_1^k = 1$. Since the standard deviation is estimated by $\{k \cdot \hat{\beta}_1^{k-1} \cdot V(\hat{\beta}_1)\}^{-1/2}$, the test statistic can be made any arbitrary value by appropriate selection of k. Phillips and Park (1988) used the $O(1/n)$ expansion of the Wald test statistic, which squares the *t*-test statistic, and proposed a modification of the test. They showed that the $O(1/\sqrt{n})$ terms can be ignored for the Wald test. One of the problems of the Wald is that it cannot be used for one-tailed tests, which are often important in single equation hypothesis testing.

In this paper, the asymptotic expansion of the t-test is considered. Unlike the Wald test cases, $O(1/\sqrt{n})$ terms cannot be ignored and correction of $O(1/\sqrt{n})$ is necessary for the t-test. A simple formula of the $O(1/\sqrt{n})$ correction of the t-test is proposed. The correction formula is obtained without using the inverse of the characteristic function. The calculation and inversion of the characteristic function are quite complicated. The method described in this paper is surprisingly easy.

1. Introduction

The t-test is the most widely used test for the single equation hypothesis. When the hypothesis is nonlinear and given by H_0: $h(\beta) = 0$, the test statistic is calculated as

$t = h(\hat{\beta})$ / (estimated standard deviation).

However, since the standard deviation is calculated from the asymptotic distribution, the t-test does not behave well for some nonlinear hypothses.

Gregory and Veall [1985] compared the hypotheses:

H_0: $\beta_1 \beta_2 = 1$ and H_0^*: $\beta_1 = 1/\beta_2$.

Although the two hypotheses are mathematically identical, the t-test based on the second hypothesis was shown to perform poorly. Lafontaine and White [1986] considered H_0: $\beta_1^k = 1$. Since the standard deviation is estimated by $\{k \cdot \hat{\beta}_1^{k-1} \cdot V(\hat{\beta}_1)\}^{-1/2}$, the test statistic can be made anyarbitrary value by appropriate selection of k.

Phillips and Park [1988] used the $O(1/n)$ expansion of the Wald test statistic, which squares the *t*-test statistic, and proposed a modification of the test. They showed that the $O(1/\sqrt{n})$ terms can be ignored for the Wald test. One of the problems of the Wald is that it cannot be used for one-tailed tests, which are often important in single equation hypothesis testing.

In this paper, the asymptotic expansion of the t-test is considered. Unlike the Wald test $O(1/\sqrt{n})$ terms cannot be ignored and correction of $O(1/\sqrt{n})$ is necessary for the t-test. The performance of the t-test and its $O(1/\sqrt{n})$ correction are analyzed using a simple example.

2. Nonlinear Hypothesis Tests

This paper considers a single equation nonlinear hypothesis given by

(1) H_0: $h(\beta) = 0$,

where β is a k dimensional vector. Following Phillips and Park (1988), the distribution of $\hat{\beta}$ is given by

$$(2)\quad \sqrt{n}(\hat{\beta} - \beta_0) \sim N(0, I),$$

where n is the number of observations and β_0 is the true parameter values of β. Since $\hat{\beta}$ can be normalized by multiplying $\Sigma^{-1/2}$ if the covariance matrix of $\hat{\beta}$ is a nonidentity matrix Σ, we can assume the covariance matrix is the identity matrix without loss of generality.

The test statistic is calculated as

$$(3)\quad t = \sqrt{n}\,h(\hat{\beta})/\sqrt{a(\hat{\beta})},$$

where $$a(\beta) = \frac{\partial h'}{\partial \beta}\frac{\partial h}{\partial \beta}.$$

It is well known that

$$(4)\quad t \Rightarrow N(0, 1).$$

Single equation nonlinear hypothesis testing is usually based on (4).

3. Correction of the Test Statistic

Since

$$(5)\quad h(\hat{\beta}) = \frac{\partial h'}{\partial \beta}\Big|_{\beta_0}(\hat{\beta} - \beta_0)$$

$$+ \frac{1}{2}(\hat{\beta} - \beta_0)'\frac{\partial^2 h}{\partial \beta \partial \beta'}\Big|_{\beta_0}(\hat{\beta} - \beta_0)$$

$$+ o_p(1/\sqrt{n}),$$

$$\frac{\partial h}{\partial \beta}\Big|_{\hat{\beta}} = \frac{\partial h}{\partial \beta}\Big|_{\beta_0} + \frac{\partial^2 h}{\partial \beta \partial \beta'}\Big|_{\beta_0}(\hat{\beta} - \beta_0)$$

$$+ o_p(1/\sqrt{n}),$$

the t-statistic defined in (4) can be written as

$$(6)\quad t = \sqrt{n/a(\beta_0)}\,\{1 - \frac{1}{a(\beta_0)}$$

$$\times \frac{\partial h'}{\partial \beta}\Big|_{\beta_0}\frac{\partial^2 h}{\partial \beta \partial \beta'}\Big|_{\beta_0}(\hat{\beta} - \beta_0)\}$$

$$\cdot \{\frac{\partial h'}{\partial \beta}\Big|_{\beta_0}(\hat{\beta} - \beta_0)$$

$$+ \frac{1}{2}(\hat{\beta} - \beta_0)'\frac{\partial^2 h}{\partial \beta \partial \beta'}\Big|_{\beta_0}(\hat{\beta} - \beta_0)\}$$

$$+ o_p(1/\sqrt{n})$$

$$= u - \frac{1}{\sqrt{n}}u\gamma'\epsilon + \frac{1}{\sqrt{n}}\epsilon' A \epsilon$$

$$+ o_p(\frac{1}{\sqrt{n}}),$$

where

$$\epsilon = \sqrt{n}(\hat{\beta} - \beta_0),$$

$$u = \frac{1}{\sqrt{a(\beta_0)}}\frac{\partial h'}{\partial \beta}\Big|_{\beta_0}\epsilon,$$

$$\gamma = \frac{1}{a(\beta_0)}\frac{\partial^2 h}{\partial \beta \partial \beta'}\Big|_{\beta_0}\frac{\partial h}{\partial \beta}\Big|_{\beta_0},$$

and

$$A = \frac{1}{2}\frac{1}{\sqrt{a(\beta_0)}}\frac{\partial^2 h}{\partial \beta \partial \beta'}\Big|_{\beta_0}.$$

u follows the standard normal distribution.

Let

$$(7)\quad t^{*} = u - \frac{1}{\sqrt{n}}u\gamma'\epsilon + \frac{1}{\sqrt{n}}\epsilon' A \epsilon.$$

t^* is the $O_p(1/\sqrt{n^*})$ approximation of t. The distribution of t^* can be obtained without using the inverse of the characteristic function. (The calculation and inversion of the characteristic function are quite complicated. The method described in this paper is surprisingly easy. For details and examples of the asymptotic expansion, see Hayakawa [1977], Takeuchi and Morimune [1985], Bhattachaya and Denker [1990], Hosoya [1990], Taniguchi [1987, 1991], Yoshida [1992, 1993], and Ghosh [1994].)

Now, let

$$(8)\quad t^{**} = E(t^* \mid u).$$

Since $\epsilon \sim N(0, I)$, it is easy to show that $E(t^* - t^{**})^2 = O(1/n)$.

Therefore,

$$(9)\quad E[\exp(i\lambda t^*)]$$

$$= E[\exp\{i\lambda(t^* - t^{**}) + i\lambda t^{**}\}]$$

$$= E[\exp(i\lambda t^{**})\exp\{i\lambda(t^* - t^{**})]$$

$$= E[\exp(i\lambda t^{**})\{1 + i\lambda(t^* - t^{**})\}]$$

$$+ o(1/\sqrt{n})$$

$$= E[\exp(i\lambda t^{**})]$$

$$+ E[\{i\lambda(t^* - t^{**})\}\exp(i\lambda t^{**})]$$

$$+ o(1/\sqrt{n}).$$

Since

$$E[i\lambda(t^* - t^{**})\exp(i\lambda t^{**})]$$

$$= E_u E[i\lambda(t^* - t^{**})\exp(i\lambda t^{**})|u]$$

$$= 0,$$

the asymptotic distribution function of t is approximated by t^{**} up to the order of $1/\sqrt{n}$.

Let H be the (k x k) matrix such that

$$(10)\quad H = (h_1, h_2, \ldots, h_k),$$

$$h_1 = \frac{1}{\sqrt{a(\beta_0)}}\frac{\partial h}{\partial \beta}\Big|_{\beta_0},$$

$$h_i' h_j = 0 \quad for\ i \neq j, \text{ and}$$

$$\|h_i\| = 1 \quad for\ any\ i.$$

Define the k-dimensional vector as

$$(11)\quad \nu = H'\epsilon, \text{ and}$$

$$\nu' = (\nu_1, \nu_2, \ldots, \nu_k).$$

Since $H'H = I$,

$$(12)\ t^* = u - \frac{1}{\sqrt{n}}u\gamma' HH'\epsilon$$

$$+ \frac{1}{\sqrt{n}}\epsilon' HH'AHH'\epsilon$$

$$= u - \frac{1}{\sqrt{n}}u\gamma' H\nu + \frac{1}{\sqrt{n}}\nu' H'AH\nu.$$

Here, $\nu_1 = u$, $\{\nu_i\}$ are independent, and $E\nu_i^2 = 1$ for any i. Therefore,

$$(13)\quad t^{**} = E(t^* | u)$$

$$= u - \frac{1}{\sqrt{n}}u\gamma' HE(\nu | u)$$

$$- \frac{1}{\sqrt{n}}tr\{H'AHE(\nu\nu' | u)\}$$

$$= u - \frac{1}{\sqrt{n}}b(\beta_0)u^2 + \frac{1}{\sqrt{n}}tr\{B\}$$

$$= u + \frac{1}{\sqrt{n}}g(\beta_0, u),$$

$$b(\beta_0) = a(\beta_0)^{-3/2} \frac{\partial h'}{\partial \beta}\Big|_{\beta_0} \frac{\partial^2 h}{\partial \beta \partial \beta'}\Big|_{\beta_0} \frac{\partial h}{\partial \beta}\Big|_{\beta_0},$$

$$B = H' A H C,$$

$$C = \begin{bmatrix} u^2, 0, 0, \cdots, 0 \\ 0, 1, 0, \cdots, 0 \\ 0, 0, 1, \cdots, 0 \\ \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\ \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\ 0, 0, 0, \cdots, 1 \end{bmatrix}, \quad \text{and}$$

$$g(\beta_0, u) = -b(\beta_0)u^2 + tr(B).$$

Since $P[t^{**}$ is not a monotonically increasing function of $u] = o(n^{-d})$ for any $d > 0$,

$$(14)\quad P[u<z] = P[t^{**} < z + \frac{1}{\sqrt{n}} g(\beta_0, z)] + o(1/\sqrt{n}).$$

Let z_α be the critical value of the standard normal distribution at the significance level α. Then the corresponding critical value is given by

$$(15)\quad z_\alpha^* = z_\alpha + \frac{1}{\sqrt{n}} g(\beta_0, z_\alpha).$$

The test can be done by (15). Since $\hat{\beta} - \beta_0 = O_p(1/\sqrt{n})$, we can still get the $1/\sqrt{n}$ correction if we substitute $\hat{\beta}$ in (15).

Note that since $g(\beta_0, z)$ is a function of z^2, $g(\beta_0, z_\alpha) = g(\beta_0, -z_\alpha)$. Therefore,

$$(16)\quad P[-z_\alpha < t < z_\alpha] - P[-z_\alpha^* < t < z_\alpha^*] = o(1/\sqrt{n}).$$

This means that when we use the two-tailed test, which is equivalent to the Wald test using the chi-square distribution with one degree of freedom, the size of the test is correct up to the order of $1/\sqrt{n}$ as suggested by Phillips and Park [1988]. However, even in such a case, the critical values are underestimated at one tail and overestimated at the other tail.

4. Performance of the Tests

In this section, the t-test and the correction formula are analyzed. $\epsilon = \sqrt{n}(\hat{\beta} - \beta_0)$, $\beta_0 = 1$ is assumed to be the standard normal distribution. The null hypothesis considered is

$$(17)\quad H_0: \beta^2 = 1.$$

The alternative hypotheses are:

(*i*) $H_1: \beta^2 < 1$,

(*ii*) $H_1: \beta^2 > 1$, and

(*iii*) $H_1: \beta^2 \neq 1$.

Cases where $n = 4, 9, 16$ are considered. The number of trials is 100 million for each case.

When the alternative hypotheses are given by (i) - (iii) and the significant levels are 5% and 1%, the sizes of the test without and with correction are given in Tables 1 - 3.

Table 1 Sizes of the T-Test $H_1: \beta^2 < 1$

Significant Level	1%	5%
n = 4 Without Correction With Correction	 8.33% 1.29%	 12.53% 2.79%
n = 9 Without Correction With Correction	 6.26% 0.83%	 11.11% 4.92%
n = 16 Without Correction With Correction	 4.48% 0.93%	 9.46% 5.02%

Table 2 Sizes of the T-Test $H_1: \beta^2 > 1$

Significant Level	1%	5%
n = 4 Without Correction With Correction	 1.99% 3.43%	 3.31% 6.39%
n = 9 Without Correction With Correction	 0.16% 1.03%	 1.98% 5.12%
n = 16 Without Correction With Correction	 0.12% 0.94%	 2.40% 5.02%

Table 3 Sizes of the T-Test H_1: $\beta^2 \neq 1$

Significant Level	1%	5%
n = 4 Without Correction With Correction	 9.24% 7.81%	 12.53% 2.90%
n = 9 Without Correction With Correction	 5.29% 0.86%	 9.02% 4.92%
n = 16 Without Correction With Correction	 3.46% 0.88%	 7.40% 4.91%

When H_1: $\beta^2 < 1$, the t-test rejects the correct null hypothesis too frequently. When the significance level is 1%, the t-test rejects the null hypothesis 8.33%, 6.26%, and 4.48% of the time for n = 4, 9, 16; when the significance level is 5%, the t-test rejects the null hypothesis 12.53%, 11.11%, and 9.46% of the time. The correction of the critical value by (15) works well. The sizes with correction are 1.29%, 0.83%, and 0.93% for the 1% level, and 2.79%, 4.92%, and 5.02% for the 5% level.

On the other hand, the sizes of the t-test are too small when H_1: $\beta^2 > 1$. The sizes are 1.98%, 0.16%, and 0.12% for the 1% level, and 3.31%, 1.98%, 2.39% for the 5% level. The sizes with correction are 3.43%, 1.03%, and 0.94% for the 1% level, and 6.39%, 5.12%, and 5.02% for the 5% level. Except for the n = 4 and 1% case, the correction significantly improves the sizes of the t-test.

When H_1: $\beta^2 \neq 1$, the sizes of the t-tests are 9.24%, 5.28%, and 3.46% for the 1% level, and 12.53%, 9.02%, and 7.39% for the 5% level. Although the t-test may not require the $1/\sqrt{n}$ correction under this alternative hypothesis, it still rejects the null hypothesis too often. The sizes with correction are 7.81%, 0.86%, and 0.88% for the 1 % level, and 2.90%, 4.91%, and 4.92% for the 5% level. As the previous cases show, the correction method works well and improves the t-test.

5. Conclusion

This paper considers the performance of the t-test for the nonlinear hypothesis and its correction. The results are:

1)Unlike the Wald test, the t-test requires the $O(1/\sqrt{n})$ correction.
2) The performance of the t-test is quite poor even if n is relatively large.
3) The correction of the critical value of the test works well and improves the t-test even for the two-tailed test, which may not require the $O(1/\sqrt{n})$ correction.

Acknowledgement

I would like to thank Michael McAleer for helpfull comments and suggestions.

References

Bhattacharya, R. and M. Denker, *Asymptotic Statistics*, Birkhauser, Basel, 1990.

Ghosh, J. K., 1994, *Higher Order Asymptotic*, NSF- CBMS Regional Conference Series in Probability and Statistics, Vol. 4, IMS, Hayward, CA, 1994.

Gregory, A. and M. R. Veall, "On Formulating Wald Tests of Nonlinear Restrictions," *Econometrica*, 53, 1465-68, 1985.

Griffiths, W.E., R. C. Hill, and P. J. Pope, "Small Sample Properties of Probit Model Estimators," *Journal of the American Statistical Association*, 82, 929-937, 1987.

Hayakawa, T., "The Likelihood Ratio and the Asymptotic Expansion of its Distribution", *Annals of the Institute of Statistical Mathematics*, 29, 359-378, 1997.

Hosoya, Y., "Information Amount and Higher-Order Efficiency in Estimation", *Annals of the Institute of Statistical Mathematics*, 42, 37-49, 1990.

Lanfontaine, F., and K. J. White, "Obtaining Any Wald Statistic You Want", *Economics Letters*, 21, 35-40, 1987.

Lehmann, E. L., *Testing Statistical Hypothesis*, John Wiely, New York, 1986.

Pfanzagl, J., *Asymptotic Expansions for General Statistical Models*, Lecture Notes in Statistics, Vol. 35, Springer-Verlag, Berlin, 1985.

Phillips, P. C. B., and J. Y. Park, "On Formulation of Wald Tests of Nonlinear Restrictions,"*Econometrica*,56, 1065-1083, 1988.

Taniguchi, M., "Third Order Asymptotic Properties of BLUE and LSE for Regression Models with ARMA residuals," *Journal of Time Series Analysis*, 8, 111-114, 1987.

______, *Higher Order Asymptotic Theory for Time Series Analysis*, Lecture Notes in Statistics, Vol. 68, Springer-Verlag, Berlin, 1991.

Veall, M. R., "Application of Resampling Methods in Econometrics," *Mimeo*, Department of Economics, McMaster University, 1993.

Yoshida, N., " Asymptotic Expansion of Maximum Likelihood Estimators for Small Diffusions via the Theory of Malliavin-Watanabe", *Probability Theory Related Fields*, 92, 275-311, 1992.

__________, "Asymptotic Expansion of Bayes Estimators for Small Diffusions, *Probability Theory Related Fields*, 95, 429-450, 1993.