Approximations of the Generalized Wilks' Distribution

Raats, V.M.

Publication date:
2004

Citation for published version (APA):
Raats, V. M. (2004). *Approximations of the Generalized Wilks' Distribution*. (CentER Discussion Paper; Vol. 2004-85). Econometrics.

CentER

Discussion Paper

No. 2004–85

APPROXIMATIONS OF THE GENERALIZED WILKS' DISTRIBUTION

By V.M. Raats

September 2004

ISSN 0924-7815

Tilburg University

Approximations of the generalized Wilks' distribution

V.M. Raats*

Abstract

Wilks' lambda and the corresponding Wilks' distribution are well known concepts in testing in multivariate regression models. The topic of this paper is a generalization of the Wilks' distribution. This generalized Wilks' distribution is relevant for testing in multivariate regression models with monotone missing data. Just as the (usual) Wilks' distribution can be approximated by the χ^2-distribution of Bartlett (1947), the generalized Wilks' distribution can be approximated by χ^2-distributions in more or less the same way. We use Box transformations to derive the χ^2-approximations and compare them by simulation.

Keywords: approximating distributions, Box transformations, generalized Wilks' distribution, monotone missing data

JEL code: C16

1 Introduction and definition

A well known test statistic for (likelihood ratio) tests in multivariate regression models is Wilks' lambda, Λ. In case of simultaneously normally distributed errors, Λ has a Wilks' distribution under the null hypothesis. For this distribution we use the same notation as Rencher (1998) *e.g.*:

$$\Lambda \sim \Lambda_{d,t,s} \quad ,$$

where $\Lambda_{d,t,s}$ denotes the Wilks' distribution with parameters d (dimension), t (degrees of freedom of the null hypothesis) and s (degrees of freedom of the errors).

*V.M.Raats@uvt.nl

This distribution is analytically rather complex and therefore it is usually approximated. Two well known approximations are the χ^2-approximation of Bartlett (1947) and the F-approximation of Rao (1952).

In this paper we study several χ^2-approximations of a generalization of the Wilks' distribution. We define the generalized Wilks' distribution $\Lambda_{A,D,T,S}$ with parameter vectors A, D, T and $S \in I\!R^{1\times r}$ as follows: let $\Lambda_i \sim \Lambda_{d_i,t_i,s_i}$ be independent and $a_i \in [0,1]$ with $a_1 = 1$. Then, by definition,

$$\prod_{i=1}^{r} \Lambda_i^{a_i} \sim \Lambda_{A,D,T,S}. \tag{1.1}$$

The vector A contains the exponents a_i of the separate factors as elements, D the d_i, T the t_i and S the s_i $(i = 1, \ldots, r)$.

The generalized Wilks' distribution is relevant for testing in missing data problems; we will elaborate on this in Section 2. In Section 3 several χ^2-approximations of the generalized Wilks' distribution are derived by means of Box transformations. The different approximations are compared by means of a simulation study in Section 4. The final Section 5 contains our conclusions.

2 Motivation

The generalized Wilks' distribution is relevant for the likelihood ratio test in multivariate linear regression with monotone missing observations of the dependent variables, or equivalently, with consecutively added dependent variables. This regression model is an important generalization of the model with the constant term as the sole explanatory variable. The latter model is used very frequently in missing data problems and it has been discussed extensively in literature (see Fujisawa (1995) *e.g.*).

Since the likelihood ratio test in the model for multivariate regression with monotone missing observations of the dependent variables is an important application of the generalized Wilks' distribution, we will describe it in more detail.

Consider the multivariate linear regression model with M dependent variables and k (deterministic) explanatory variables; observations are gathered for N cases. Let $X_{tj} \in I\!R$ be the observed value of the j^{th} explanatory variable ($j = 1, \ldots, k$) for the t^{th} case; complete data are available for the explanatory variables, so $t = 1, \ldots, N$ for all j.

The observations of the dependent variables are incomplete; the dependent variables are ordered such that later added variables come last. So their data are divided into r ordered groups according to the pattern of increasingly missing data. Group i contains m_i variables for which exactly the first N_i observations are available:

$$N = N_1 \geq N_2 \geq \ldots \geq N_r; \quad M_i = \sum_{j=1}^{i} m_j \quad (i = 1, \ldots, r,\ M_r = M).$$

The vector $Y_{ti} \in \mathbb{R}^{m_i}$ contains the values of these m_i dependent variables for case t. So Y_{ti} is observable for $t = 1, \ldots, N_i$ and missing for $t = N_i + 1, \ldots, N$. The special case $N = N_1 = \ldots = N_r$ gives the usual complete model.

The r (multivariate) regression equations can be written as

$$Y_{ti} = \mu_{ti} + \varepsilon_{ti}, \quad \mu_{ti} = \sum_{j=1}^{k} X_{tj}\beta_{ji}, \quad i = 1, \ldots, r, \quad t = 1, \ldots, N_i, \tag{2.1}$$

where $\beta_{ji} \in \mathbb{R}^{m_i}$ denotes a vector of unknown regression coefficients. For the errors we assume

$$E\{\varepsilon_{ti}\} = 0, \quad Cov(\varepsilon_{ti}, \varepsilon_{sj}) = \delta_{ts}\sigma_{ij}, \tag{2.2}$$

with (completely unknown) non-singular $\Sigma = (\sigma_{ij}) \in \mathbb{R}^{M \times M}$ not depending on the β_{ji}. We write $\Sigma > 0$ for positive definiteness.

In Raats *et al.* (2002) we showed that the likelihood ratio (to the power $\frac{2}{N}$) for testing homogeneous linear restrictions on the regression coefficients in the described model has a generalized Wilks' distribution under the null hypothesis.

3 Box transformations

The approximations of the generalized Wilks' distribution in (1.1) can be derived by means of transformations which were introduced in Box (1949); we have used the main result of the transformations as presented in Muirhead (1982) Section 8.2.4.

Theorem 3.1. *A second order approximation of the distribution of*

$$Q = -2log(\prod_{i=1}^{r} \Lambda_i^{a_i})$$

is

$$P(Q \leq q) = (1 - \omega_2)P(\chi_f^2 \leq \rho q) + \omega_2 P(\chi_{f+4}^2 \leq \rho q) + O(N^{-3}) \tag{3.1}$$

with

$$\begin{aligned} f &= \sum_{i=1}^{r}\sum_{j=1}^{d_i} t_i, \\ \rho &= \frac{1}{4f}\sum_{i=1}^{r}\sum_{j=1}^{d_i}\frac{t_i}{a_i}(2s_i - 2j + t_i) \\ \omega_2 &= -\frac{1}{4f} + \frac{1}{48\rho^2}\sum_{i=1}^{r}\sum_{j=1}^{d_i}\frac{t_i}{a_i^2}\left(3(s_i+j+1)(s_i+j+t_i-1)+(t_i-2)(t_i-1)\right) \end{aligned}$$

Proof. According to Muirhead (1982) Theorem 10.5.3 the Wilks' distribution can be rewritten as the following product of independent beta distributions:

$$\Lambda_{d,t,s} = \prod_{j=1}^{d} Beta(\frac{1}{2}(s-j+1), \frac{1}{2}t).$$

Accordingly, the moments of the generalized Wilks' distribution follow from the independence and the moments of Λ_i:

$$E\{\prod_{i=1}^{r}\Lambda_i^{a_i h}\} = K\prod_{i=1}^{r}\prod_{j=1}^{d_i}\frac{\Gamma\left[\frac{1}{2}a_i(1+h) - \frac{1}{2}(s_i - j + 1 - 2a_i)\right]}{\Gamma\left[\frac{1}{2}a_i(1+h) - \frac{1}{2}(s_i + t_i - j + 1 - 2a_i)\right]}, \qquad (3.2)$$

Box transformations lead, after algebraic manipulations, to the approximating distribution (3.1) with parameters f, ρ and ω_2 (see the Appendix). □

We call (3.1) the Box approximation. From (3.1) the first order approximation follows

$$P(Q \leq q) = P(\chi_f^2 \leq \rho q) + O(N^{-2}). \qquad (3.3)$$

Since (3.3) coincides with Bartlett's approximation in case of complete data, we will call (3.3) Bartlett's approximation even in this more general situation.

In the context of the model of Section 2, in case of only the constant as explanatory variable, our parameters in (3.1) reduce to the ones derived in Bhargava (1962).

4 A simulation study

In this section we compare approximations (3.1), (3.3) and in addition the standard approximation *i.e.* $-Nlog(\prod_{i=1}^{r}\Lambda_i^{a_i}) \sim \chi_f^2$. This standard approximation is derived from the equivalence of the likelihood ratio LR_0 with $\prod_{i=1}^{r}\Lambda_i^{\frac{Na_i}{2}}$ in the multivariate regression model with monotone missing observations (see Section 2), and the standard asymptotic distribution of the likelihood ratio : $-2log(LR_0) \sim \chi^2_f$.

In order to compare the approximations, we first simulate (runsize 1,000,000) the critical value of $\prod_{i=1}^{r} \Lambda_i^{a_i}$ (with significance level α). Then the probability that $\prod_{i=1}^{r} \Lambda_i^{a_i}$ exceeds this critical value is determined according to the three different approximations.

In the context of the model of Section 2, the simulations have been performed for four explanatory variables, three groups ($r = 3$), t linear constraints per group ($t_i = t$ for $i = 1, \ldots, 3$). We study different values of the significance level α, number of cases (N), number of constraints t, fractions of (missing) data ($A = [a_1\ a_2\ a_3]$ with $a_i = N_i/N$) and different number of variables per group ($D = [d_1\ d_2\ d_3]$). For the model of Section 2 the degrees of freedom of the errors can be shown to be $s_i = N(1 - a_i) - d_{i-1} - 4$ (with $d_0 = 0$).

Table 4.1 contains the results for $D = [1\ 2\ 1]$.

$D = [1\quad 2\quad 1]$		$A = [1\quad 0.9\quad 0.8]$			$A = [1\quad 0.8\quad 0.6]$		
$\alpha = 0.05$		Standard	Bartlett	Box	Standard	Bartlett	Box
$N = 20$	$t = 1$	.009	.047	.050	.004	.040	.048
	$t = 2$	.012	.047	.050	.007	.042	.049
	$t = 4$	.037	.047	.050	.032	.045	.049
$N = 200$	$t = 1$	.044	.050	.050	.044	.050	.050
	$t = 2$	.045	.050	.050	.044	.050	.050
	$t = 4$	.049	.050	.050	.048	.050	.050
$N = 2000$	$t = 1$	.049	.050	.050	.049	.050	.050
	$t = 2$	.049	.050	.050	.049	.050	.050
	$t = 4$	.050	.050	.050	.050	.050	.050
$\alpha = 0.10$							
$N = 20$	$t = 1$	.026	.096	.100	.015	.085	.098
	$t = 2$	.031	.095	.100	.020	.087	.098
	$t = 4$	.078	.095	.100	.070	.092	.099
$N = 200$	$t = 1$	.091	.100	.100	.090	.100	.100
	$t = 2$	.092	.100	.100	.090	.100	.100
	$t = 4$	.098	.100	.100	.097	.100	.100
$N = 2000$	$t = 1$	.099	.100	.100	.099	.100	.100
	$t = 2$	.099	.100	.100	.099	.100	.100
	$t = 4$	.100	.100	.100	.100	.100	.100

Table 4.1: Simulated approximations for $D = [1\quad 2\quad 1]$

As can be expected, the accuracy of the approximations increases with the sample sizes. Approximation (3.1) outperforms the other ones. The standard approximation is quite bad for small sample sizes. Only for $N = 2000$, this approximation gives good results. Approximation (3.3) performs well for big sample sizes ($N = 200(0)$), but is not as accurate as approximation (3.1) for small sample sizes ($N = 20$). All the approximations seem to improve with the degrees of freedom of the null hypothesis (t). As the fraction of missing observations increases (*i.e.* a_i decreases), the approximations become less accurate.

To study the effect of the number of variables per group (d_i) on the quality of the approximations, we also did a simulation for $D = [1\ 3\ 2]$. Table 4.2 contains the results.

$D = [1 \quad 3 \quad 2]$		$A = [1 \quad 0.9 \quad 0.8]$			$A = [1 \quad 0.8 \quad 0.6]$		
$\alpha = 0.05$		Standard	Bartlett	Box	Standard	Bartlett	Box
$N = 20$	$t = 1$	.003	.040	.049	.000	.022	.040
	$t = 2$	.003	.041	.049	.001	.027	.043
	$t = 4$	.017	.042	.049	.001	.035	.046
$N = 200$	$t = 1$	.042	.050	.050	.040	.050	.050
	$t = 2$	.046	.050	.050	.041	.050	.050
	$t = 4$	.049	.050	.050	.045	.050	.050
$N = 2000$	$t = 1$	.049	.050	.050	.049	.050	.050
	$t = 2$	.049	.050	.050	.049	.050	.050
	$t = 4$	.049	.050	.050	.050	.050	.050
$\alpha = 0.10$							
$N = 20$	$t = 1$	.009	.085	.098	.002	.054	.086
	$t = 2$	.011	.086	.099	.003	.063	.091
	$t = 4$	.041	.087	.099	.026	.077	.096
$N = 200$	$t = 1$	.088	.100	.100	.085	.100	.100
	$t = 2$	.087	.100	.100	.085	.100	.100
	$t = 4$	.094	.100	.100	.092	.100	.100
$N = 2000$	$t = 1$	.098	.100	.100	.098	.100	.100
	$t = 2$	.098	.100	.100	.099	.100	.100
	$t = 4$	.100	.100	.100	.099	.100	.100

Table 4.2: Simulated approximations for $D = [1 \quad 3 \quad 2]$

The previous conclusions about the effect of the different parameters still remain valid. However, in comparison to Table 4.1, the quality of the approximations is worse if there is only a small number of observations ($N = 20$) available.

5 Summary and conclusions

The generalized Wilks' distribution occurs in the likelihood ratio test for multivariate regression with monotone missing observations. In case of complete data, this distribution coincides with the (usual) Wilks distribution.

We derived several χ^2-approximations for the generalized Wilks' distribution. As was to be expected, the highest (second) order approximation outperforms the other ones. In case of complete data the first order approximation coincides with the well known χ^2-approximation of Bartlett (1947). Hence the latter can be improved by taking the second order approximation in (3.1).

In this paper we have solely focussed on χ^2-approximations for the generalized Wilks' distribution; it would also be interesting to look at F-approximations similar to the one of Rao (1952) for the usual Wilks' distribution.

Acknowledgements

Very useful remarks of Ben van der Genugten and Hans Moors are gratefully acknowledged.

Appendix

To approximate the generalized Wilks' distribution, we have used the main result of Box transformations as presented in Muirhead (1982) Section 8.2.4:

Consider a random variable Z $(0 \leq Z \leq 1)$ with moments:

$$E\{Z^h\} = K \left[\frac{\prod_{j=1}^{p} y_j^{y_j}}{\prod_{k=1}^{q} x_k^{x_k}} \right]^h \frac{\prod_{k=1}^{q} \Gamma\left[x_k(1+h)+\xi_k\right]}{\prod_{j=1}^{p} \Gamma\left[y_j(1+h)+\eta_j\right]},$$

where

$$\sum_{j=1}^{p} y_j = \sum_{k=1}^{q} x_k$$

and K is a constant such that E$\{Z^0\}$=1. Then

$$P(-2\rho log(Z) \leq x) =$$

$$P(\chi^2_f \leq x) + \omega_2 \left[P(\chi^2_{f+4} \leq x) - P(\chi^2_f \leq x)\right] + O(N^{-3}),$$

where

$$f = -2\left[\sum_{k=1}^{q}\xi_k - \sum_{j=1}^{p}\eta_j - \frac{1}{2}(q-p)\right]$$

and

$$\rho = 1 - \frac{1}{f}\left[\sum_{k=1}^{q} x_k^{-1}\left(\xi_k^2 - \xi_k + \frac{1}{6}\right) - \sum_{j=1}^{p} y_j^{-1}\left(\eta_j^2 - \eta_j + \frac{1}{6}\right)\right]$$

and

$$\begin{aligned}\omega_2 &= -\frac{1}{6\rho^2}\left\{\sum_{k=1}^{q} x_k^{-2}\left[(\beta_k+\xi_k)^3 - \frac{3}{2}(\beta_k+\xi_k)^2 + \frac{1}{2}(\beta_k+\xi_k)\right]\right.\\ &\quad \left. - \sum_{j=1}^{p} y_j^{-2}\left[(\epsilon_j+\eta_j)^3 - \frac{3}{2}(\epsilon_j+\eta_j)^2 + \frac{1}{2}(\epsilon_j+\eta_j)\right]\right\},\end{aligned}$$

with

$$\beta_k = (1-\rho)x_k, \quad \epsilon_j = (1-\rho)y_j.$$

Using the specific shape of (3.2) the parameters can be written as

$$f = \sum_{i=1}^{r}\sum_{j=1}^{d_i} t_i = \sum_{i=1}^{r} d_i t_i$$

and

$$\begin{aligned}\rho &= 1 - \frac{1}{f}\sum_{i=1}^{r}\sum_{j=1}^{d_i} a_i^{-1}\left[\frac{1}{2}t_i - \frac{1}{4}t_i^2 - \frac{1}{2}(s_i - j + 1 - 2a_i)t_i\right]\\ &= \frac{1}{4f}\sum_{i=1}^{r}\sum_{j=1}^{d_i}\frac{t_i}{a_i}(2s_i - 2j + t_i)\end{aligned}$$

and

$$\omega_2 = -\frac{1}{6\rho^2}\sum_{i=1}^{r}\sum_{j=1}^{d_i}\frac{1}{a_i^2}\Big[-\frac{3}{4}t_i^2(\frac{1}{2}(s_i-j+1)-\rho a_i)-\frac{3}{2}t_i(\frac{1}{2}(s_i-j+1)-\rho a_i)^2 -\frac{1}{8}t_i^3+\frac{3}{2}t_i(\frac{1}{2}(s_i-j+1)-\rho a_i)+\frac{3}{8}t_i^2-\frac{1}{4}t_i\Big]$$

$$= \frac{1}{4}f-\frac{1}{6\rho}\sum_{i=1}^{r}\sum_{j=1}^{d_i}\frac{1}{a_i}(\frac{3}{4}t_i^2+\frac{3}{2}(s_i-j+1)t_i-\frac{3}{2}t_i) -\frac{1}{48\rho^2}\sum_{i=1}^{r}\sum_{j=1}^{d_i}\frac{t_i}{a_i^2}(-3(s_i-j+1)(-t_i-(s_i-j+1)+2)-t_i^2+3t_i-2)$$

$$= \frac{f}{4}-\frac{1}{8\rho}\sum_{i=1}^{r}\sum_{j=1}^{d_i}\frac{t_i}{a_i}(2s_i-2j+t_i) +\frac{1}{48\rho^2}\sum_{i=1}^{r}\sum_{j=1}^{d_i}\frac{t_i}{a_i^2}(3(s_i-j+1)(s_i-j+t_i-1)+(t_i-2)(t_i-1))$$

$$= -\frac{1}{4}f+\frac{1}{48\rho^2}\sum_{i=1}^{r}\sum_{j=1}^{d_i}\frac{t_i}{a_i^2}(3(s_i-j+1)(s_i-j+t_i-1)+(t_i-2)(t_i-1))$$

References

Bartlett, M.S. (1947). Multivariate Analysis. *Journal of the Royal Statistical Society B*, **9**, 176–197.

Bhargava, R.P. (1962). *Multivariate Tests of Hypotheses with Incomplete Data*. Ph.d. dissertation, Stanford University.

Box, G.E.P. (1949). A General Distribution Theory for a Class of Likelihood Criteria. *Biometrika*, **36**, 317–346.

Fujisawa, H. (1995). A Note on the Maximum Likelihood Estimators for Multivariate Normal Distribution with Monotone Data. *Communications in Statistics: Theory and Methods*, **24**, 1377–1382.

Muirhead, R.J. (1982). *Aspects of Multivariate Statistical Theory*. Wiley.

Raats, V.M., B.B. van der Genugten, and J.J.A. Moors (2002). Multivariate Regression with Monotone Missing Observations of the Dependent Variables. CentER Discussion Paper 63, Tilburg University.

Rao, C.R. (1952). *Advanced Statistical Methods in Biometric Research*. Wiley.

Rencher, A.C. (1998). *Multivariate Statistical Inference and Applications*. John Wiley and Sons.