International Journal of Pure and Applied Mathematics
Volume 30 No. 3 2006, 387-392

THE DOUBLY NONCENTRAL F DISTRIBUTION IN A REGRESSION MODEL WITH PROXY VARIABLES

Deniz Ünal

Department of Statistics
Faculty of Arts and Sciences
University of Çukurova
Adana, 01330, TURKEY
e-mail: denizu@alumni.bilkent.edu.tr

Abstract: In this paper a way of computing expectations for the ratios of different powers of two independent noncentral chi-squared distributions is proposed in a regression model with proxy variables. For the same powers of this ratio, expectation of doubly noncentral F and expectation of its powers can be reached.

AMS Subject Classification: 62J05
Key Words: chi-squared distribution, F distribution

1. Introduction

A linear regression model with some unobservable relevant regressors may be estimated by using the proxy variables [6] as regressors. In such models ratio of quadratic forms may be required. The most important example for the ratio of quadratic forms is the noncentral F distribution. It has an important role especially for the calculation of power for tests of linear hypothesis. Beside of this, numerous examples exist where the doubly noncentral F is required for example in two way cross classification ANOVA [5]. In econometrics, it arises naturally in testing linear models with proxy variables [2] and in signal processing and pattern recognition applications [4], [1].

In this paper we find the expectation of the fraction of the powers of quadratic forms in a regression model with proxy variables. This result gives a general formula, this means that for special values of the powers we can reach

Received: June 22, 2006

the expectation formula for the noncentral F, doubly noncentral F and some special case for the ratios of quadratic forms.

The distribution of $\frac{\nu_1/\lambda_1^*}{\nu_2/\lambda_2^*}$ for two independently distributed noncentral chi-squared variables $\nu_1 \sim \chi_k'^2(\lambda_1^*)$ and $\nu_2 \sim \chi_\nu'^2(\lambda_2^*)$ describes the doubly noncentral F distribution. For $\lambda_1^* = \lambda_2^* = 0$ this becomes the central F distribution and for $\lambda_2^* = 0$ with $\lambda_1^* \neq 0$ it becomes the noncentral F distribution. In this study we aim to find the expectation of the ratio of different powers of two independently distributed noncentral chi-squared variables that is we aim to obtain an explicit formula for $E(\frac{(\nu_1/\lambda_1^*)^b}{(\nu_2/\lambda_2^*)^a})$ in regression model with proxy variables. This paper is organized as follows. The model is defined and the estimators are given in Section 2. In Section 3, we execute the expectation formula for the ratio of non-central chi-squared distributions and its powers.

2. The Model and the Estimators

Let us first consider the partitioned linear regression model

$$\boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim N(0, \sigma^2 I_n), \tag{1}$$

where $\boldsymbol{y} : n \times 1$ observation vector of a dependent variable, $\boldsymbol{X}_1 : n \times k_1$ and $\boldsymbol{X}_2 : n \times k_2$ matrices of observations of non-stochastic independent variables, $\boldsymbol{\beta}_1 : k_1 \times 1$ and $\boldsymbol{\beta}_2 : k_2 \times 1$ vectors of parameters and $\boldsymbol{\epsilon} : n \times 1$ vector of normal disturbance terms. The usual OLSE (ordinary least squares estimator) for the parameter vector $\boldsymbol{\beta} = [\boldsymbol{\beta}_1', \boldsymbol{\beta}_2']'$ is

$$\boldsymbol{b} = \boldsymbol{S}^{-1}\boldsymbol{X}'\boldsymbol{y}\,,$$

where $\boldsymbol{S} = (\boldsymbol{X}'\boldsymbol{X})$

We assume the existence of $\boldsymbol{X}_2^*$ as the matrix of the proxy variables though $\boldsymbol{X}_2$ is unobservable. Now, let us consider the linear regression model [3] with the proxy variable $\boldsymbol{X}_2^*$ in place of $\boldsymbol{X}_2$:

$$\boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2^*\boldsymbol{\beta}_2^* + \boldsymbol{u}^*,$$

$$\boldsymbol{u}^* \sim N(\boldsymbol{X}_2\boldsymbol{\beta}_2 - \boldsymbol{X}_2^*\boldsymbol{\beta}_2^*, \sigma^2 I_n), \tag{2}$$

where $\boldsymbol{u}^* = \boldsymbol{X}_2\boldsymbol{\beta}_2 - \boldsymbol{X}_2^*\boldsymbol{\beta}_2^* + \boldsymbol{\epsilon}$. Assume $\boldsymbol{X} = [\boldsymbol{X}_1, \boldsymbol{X}_2]$ and $\boldsymbol{X}^* = [\boldsymbol{X}_1, \boldsymbol{X}_2^*]$ have full column rank and consider the estimators for $\boldsymbol{\beta}^* = [\boldsymbol{\beta}_1', \boldsymbol{\beta}_2'^*]'$ based on the model (2). The usual OLSE for the parameter vector $\boldsymbol{\beta}^* = [\boldsymbol{\beta}_1', \boldsymbol{\beta}_2'^*]'$ is

$$\boldsymbol{b}^* = \boldsymbol{S}^{*-1}\boldsymbol{X}^{*'}\boldsymbol{y}\,,$$

where $\boldsymbol{S}^* = (\boldsymbol{X}^{*'}\boldsymbol{X}^*)$.

3. Derivation of the Expectation Formula

Let us take the model (2) into consideration. Since $\nu_1 = \frac{\boldsymbol{b}^{*'}\boldsymbol{S}^*\boldsymbol{b}^*}{\sigma^2}$ and $\nu_2 = \frac{\boldsymbol{e}^{*'}\boldsymbol{e}^*}{\sigma^2}$ have the independent noncentral chi-square distributions with non-centrality parameters

$$\lambda_1^* = \frac{\boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{X}^*\boldsymbol{S}^{*-1}\boldsymbol{X}^{*'}\boldsymbol{X}\boldsymbol{\beta}}{\sigma^2} \quad \text{and} \quad \lambda_2^* = \frac{\boldsymbol{\beta}'\boldsymbol{X}'(I - \boldsymbol{X}^*\boldsymbol{S}^{*-1}\boldsymbol{X}^{*'})\boldsymbol{X}\boldsymbol{\beta}}{\sigma^2},$$

respectively. It is known that

$$f(\nu_1) = e^{-\lambda_1^*}\sum_{i=0}^{\infty}\frac{\lambda_1^{*i}}{i!}\frac{\nu_1^{k/2+i-1}e^{-\nu_1/2}}{2^{k/2+i}\Gamma(\frac{k}{2}+i)}, \tag{3}$$

and

$$f(\nu_2) = e^{-\lambda_2^*}\sum_{j=0}^{\infty}\frac{\lambda_2^{*j}}{j!}\frac{\nu_2^{\nu/2+j-1}e^{-\nu_2/2}}{2^{\nu/2+j}\Gamma(\frac{\nu}{2}+j)}. \tag{4}$$

Since ν_1 and ν_2 are independent (see Appendix), then the joint probability density function becomes,

$$f(\nu_1,\nu_2) = e^{-\lambda_1^*}e^{-\lambda_2^*}\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\frac{\lambda_1^{*i}}{i!}\frac{\lambda_2^{*j}}{j!}\frac{\nu_1^{k/2+i-1}\nu_2^{\nu/2+j-1}e^{-\nu_1/2}e^{-\nu_2/2}}{2^{k/2+i}2^{\nu/2+j}\Gamma(\frac{k}{2}+i)\Gamma(\frac{\nu}{2}+j)}. \tag{5}$$

Now, the expectation formula can be given as

$$\begin{aligned}
E(\frac{(\nu_1/k)^n}{(\nu_2/\nu)^m}) &= \frac{\nu^m}{k^n}\int_0^\infty\int_0^\infty\frac{\nu_1^n}{\nu_2^m}f(\nu_1,\nu_2)d\nu_1 d\nu_2 = \frac{\nu^m}{k^n}\int_0^\infty\int_0^\infty\frac{\nu_1^n}{\nu_2^m} \\
&\quad\times e^{-\lambda_1^*}e^{-\lambda_2^*}\sum_{i=0}^{\infty}\frac{\lambda_1^{*i}}{i!}\frac{\nu_1^{k/2+i-1}e^{-\nu_1/2}}{2^{k/2+i}\Gamma(\frac{k}{2}+i)}\sum_{j=0}^{\infty}\frac{\lambda_2^{*j}}{j!}\frac{\nu_2^{\nu/2+j-1}e^{-\nu_2/2}}{2^{\nu/2+j}\Gamma(\frac{\nu}{2}+j)}d\nu_1 d\nu_2, \\
&= \frac{\nu^m}{k^n}\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\frac{e^{-\lambda_1^*}e^{-\lambda_2^*}\lambda_1^{*i}\lambda_2^{*j}}{2^{k/2+i}2^{\nu/2+j}i!j!\Gamma(\frac{k}{2}+i)\Gamma(\frac{\nu}{2}+j)} \\
&\quad\times\int_0^\infty\int_0^\infty \nu_1^{k/2+n+i-1}\nu_2^{\nu/2-m+j-1}e^{-\nu_1/2}e^{-\nu_2/2}d\nu_1 d\nu_2.
\end{aligned} \tag{6}$$

Making use of the change of variables $\nu_3 = \nu_1/\nu_2$ and $\nu_2 = \nu_4$ and taking

$$A_{ij} = \frac{e^{-\lambda_1^*}e^{-\lambda_2^*}\lambda_1^{*i}\lambda_2^{*j}}{2^{k/2+i}2^{\nu/2+j}i!j!\Gamma(\frac{k}{2}+i)\Gamma(\frac{\nu}{2}+j)},$$

the integral in (6) reduces to

$$E(\frac{(\nu_1/k)^n}{(\nu_2/\nu)^m})$$

$$= \frac{\nu^m}{k^n}\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}A_{ij}\int_0^{\infty}\int_0^{\infty}(\nu_4\nu_3)^{k/2+n+i-1}\nu_4^{\nu/2-m+j-1}e^{-\frac{\nu_4(\nu_3+1)}{2}}\nu_4 d\nu_3 d\nu_4,$$

$$= \frac{\nu^m}{k^n}\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}A_{ij}\int_0^{\infty}\int_0^{\infty}(\nu_3)^{k/2+n+i-1}\nu_4^{\frac{k+\nu}{2}+n-m+i+j-1} \times e^{-\frac{\nu_4(\nu_3+1)}{2}}d\nu_3 d\nu_4. \quad (7)$$

Again making use of the change of variable $\nu_3 = \frac{\nu_6}{1-\nu_6}$ and $\nu_4 = 2\nu_5(1-\nu_6)$, the equation (7) reduces to,

$$E(\frac{(\nu_1/k)^n}{(\nu_2/\nu)^m}) = \frac{\nu^m}{k^n}\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}A_{ij}\int_0^{1}\int_0^{\infty}\nu_6^{\frac{k}{2}+n+i-1}(1-\nu_6)^{\frac{\nu}{2}-m+j} \times 2^{\frac{k+\nu}{2}+n-m+i+j-1}\nu_5^{\frac{k+\nu}{2}+n-m+i+j-1}e^{-\nu_5}\frac{2}{1-\nu_6}d\nu_5 d\nu_6$$

$$= \frac{\nu^m}{k^n}\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}A_{ij}2^{\frac{k+\nu}{2}+n-m+i+j}\int_0^{1}\nu_6^{\frac{k}{2}+n+i-1}(1-\nu_6)^{\frac{\nu}{2}-m+j-1}d\nu_6 \times \int_0^{\infty}\nu_5^{\frac{k+\nu}{2}+n-m+i+j-1}e^{-\nu_5}d\nu_5. \quad (8)$$

Using the following Gamma and Beta functions

$$\Gamma(\alpha) = \int_0^{\infty}y^{\alpha-1}e^{-y}dy, \quad \forall\alpha > 0\,,$$

$$B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} = \int_0^{1}y^{\alpha-1}(1-y)^{\beta-1}dy, \quad \forall\alpha > 0 \text{ and } \beta > 0\,,$$

and using A_{ij}, the equation (8) reduces to

$$E(\frac{(\nu_1/k)^n}{(\nu_2/\nu)^m}) = \frac{\nu^m}{k^n} \times \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\frac{e^{-\lambda_1^*}e^{-\lambda_2^*}\lambda_1^{*i}\lambda_2^{*j}2^{\frac{k+\nu}{2}+n-m+i+j}}{2^{k/2+i}2^{\nu/2+j}i!j!\Gamma(\frac{k}{2}+i)\Gamma(\frac{\nu}{2}+j)}\Gamma(k/2+n+i)\Gamma(\nu/2-m+j)$$

$$= \frac{\nu^m}{k^n}2^{n-m}e^{-\lambda_1^*}e^{-\lambda_2^*}\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\frac{\lambda_1^{*i}\lambda_2^{*j}}{i!j!}\frac{\Gamma(\frac{k}{2}+i+n)\Gamma(\frac{\nu}{2}+j-m)}{\Gamma(\frac{k}{2}+i)\Gamma(\frac{\nu}{2}+j)}\,. \quad (9)$$

Then,

$$E(\frac{{\nu_1}^n}{{\nu_2}^m}) = e^{-\lambda_1^*}e^{-\lambda_2^*}2^{n-m}\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\frac{\lambda_1^{*i}\lambda_2^{*j}\Gamma(\frac{k}{2}+i+n)\Gamma(\frac{\nu}{2}+j-m)}{i!j!\Gamma(\frac{k}{2}+i)\Gamma(\frac{\nu}{2}+j)}\,. \quad (10)$$

Finally for $\nu_1 \sim \chi_k^{\prime 2}(\lambda_1^*)$ and $\nu_2 \sim \chi_\nu^{\prime 2}(\lambda_2^*)$, the equation (10) becomes,

$$E(\frac{(\boldsymbol{b}^{*'}\boldsymbol{S}^{*}\boldsymbol{b}^{*})^{n}}{(\boldsymbol{e}^{*'}\boldsymbol{e}^{*})^{m}}) = \sigma^{2(n-m)}e^{-\lambda_1^*}e^{-\lambda_2^*}2^{n-m} \times \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\frac{\lambda_1^{*i}\lambda_2^{*j}\Gamma(\frac{k}{2}+i+n)\Gamma(\frac{\nu}{2}+j-m)}{i!j!\Gamma(\frac{k}{2}+i)\Gamma(\frac{\nu}{2}+j)}, \quad (11)$$

where $\lambda_1^* = \frac{\boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{X}^*\boldsymbol{S}^{*-1}\boldsymbol{X}^{*'}\boldsymbol{X}\boldsymbol{\beta}}{\sigma^2}$ and $\lambda_2^* = \frac{\boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{M}^*\boldsymbol{X}\boldsymbol{\beta}}{\sigma^2}$ with $\boldsymbol{M}^* = I - \boldsymbol{X}^*\boldsymbol{S}^{*-1}\boldsymbol{X}^{*'}$, and m and n are constants.

Here for special values of m, n, k and ν we obtain different expressions. For example, if $m = n = 1$ this result becomes the expectation, and if $m = n = r$ the result becomes the r-th moment of the doubly noncentral F in a regression model with proxy variables.

4. Conclusion

As mentioned before, in regression model with proxy variables we present a way for the expectation formula of the doubly noncentral F distribution and ratio of the quadratic forms which are used quite often in statistics and econometrics. The result which is found for the regression model with proxy variables can be extended to general conditions.

Acknowledgments

I wish to thank Professor Dr. Fikri Akdeniz for the useful comments and suggestions.

References

[1] C.W. Helstrom, J.A. Ritcey, Evaluation of the non-central F distribution by numerical contour integration, *SIAM Journal on Scientific and Statistical Computing*, **8**, No. 3 (1985), 505-514.

[2] H. Krumai, K. Ohtani, The exact distribution and density functions of a pre-test estimator of the error variance in a linear regression model with proxy variables, *Statistical Papers*, **39** (1998), 163-177.

[3] A. Namba, K. Ohtani, PMSE performance of the Stein-rule and positive part Stein-rule estimators in a linear regression model with or without proxy variables, *Statistics and Probability Letters* (2006), To Appear.

[4] R. Price, Some non-central F distributions expressed in closed form, *Biometrica*, **51**, No. 1, 2 (1964), 107-122.

[5] M.L. Tiku, A note on the distribution of the doubly noncentral F distribution, *Australian Journal of Statistics*, **14**, No. 1 (1972), 37-40.

[6] M.R. Wickens, A note on the use of proxy variables, *Econometrica*, **40** (1972), 759-761.

Appendix

Since $\boldsymbol{b}^* = \boldsymbol{S}^{*-1}\boldsymbol{X}^{*'}\boldsymbol{y}$,

$$\nu_1 = \frac{\boldsymbol{b}^{*'}\boldsymbol{S}^*\boldsymbol{b}^*}{\sigma^2} = \frac{\boldsymbol{y}'\boldsymbol{X}^*\boldsymbol{S}^{*-1}\boldsymbol{S}^*\boldsymbol{S}^{*-1}\boldsymbol{X}^{*'}\boldsymbol{y}}{\sigma^2} = \frac{\boldsymbol{y}'\boldsymbol{X}^*\boldsymbol{S}^{*-1}\boldsymbol{X}^{*'}\boldsymbol{y}}{\sigma^2} \tag{12}$$

and since $\boldsymbol{e}^* = \boldsymbol{y} - \boldsymbol{X}^*\boldsymbol{b}^*$,

$$\begin{aligned}\nu_2 &= \frac{\boldsymbol{e}^{*'}\boldsymbol{e}^*}{\sigma^2} = \frac{(\boldsymbol{y}' - \boldsymbol{b}^{*'}\boldsymbol{X}^{*'})(\boldsymbol{y} - \boldsymbol{X}^*\boldsymbol{b}^*)}{\sigma^2} \\ &= \frac{\boldsymbol{y}'\boldsymbol{y} - \boldsymbol{y}'\boldsymbol{X}^*\boldsymbol{b}^* - \boldsymbol{b}^{*'}\boldsymbol{X}^{*'}\boldsymbol{y} + \boldsymbol{b}^{*'}\boldsymbol{X}^{*'}\boldsymbol{X}^*\boldsymbol{b}^*}{\sigma^2} = \frac{\boldsymbol{y}'(I - \boldsymbol{X}^*\boldsymbol{S}^{*-1}\boldsymbol{X}^{*'})\boldsymbol{y}}{\sigma^2}. \end{aligned} \tag{13}$$

The quadratic forms in (12) and (13) have the noncentral chi-square distributions with k and ν degrees of freedom having $\lambda_1^* = \frac{\boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{X}^*\boldsymbol{S}^{*-1}\boldsymbol{X}^{*'}\boldsymbol{X}\boldsymbol{\beta}}{\sigma^2}$ and $\lambda_2^* = \frac{\boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{M}^*\boldsymbol{X}\boldsymbol{\beta}}{\sigma^2}$ as the non-centrality parameters, respectively and $\boldsymbol{M}^* = I - \boldsymbol{X}^*\boldsymbol{S}^{*-1}\boldsymbol{X}^{*'}$. Here ν_1 and ν_2 are independent since $(I - \boldsymbol{X}^*\boldsymbol{S}^{*-1}\boldsymbol{X}^{*'})\cdot(\boldsymbol{X}^*\boldsymbol{S}^{*-1}\boldsymbol{X}^{*'}) = 0$ and $\boldsymbol{y} \sim N(\boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2, \sigma^2 I_n)$.