THE EGYPTIAN STATISTICAL JOURNAL
ISSR, CAIRO UNIV., Vol.39 No.2, 1994

GOODNESS OF FIT FOR MULTIVARIATE LINEAR REGRESSION MODELS

By

Ahmed H. Haroun
Cairo University

1- Introduction

This paper presents a test statistics for goodness of fit for the multivariate linear regression models. This test statistics has an approximate central F distribution. In section 2, the specification and estimation of the model are presented. Derivation of the appropriate test statistics and its approximate distribution is given in sections 3 and 4, respectively.

2- The Multivariate linear regression Model: Specification and estimation

The multivariate linear regression model is a direct extension of the linear regression model to the case where the dependent variable is an mx1 random variable y_t . That is

[I] The statistical generating mechanism (GM) takes the form

$$y_t = B'x_t + u_t \qquad t \in T \tag{2-1}$$

where y_t : mx1, B: kxm, x_t : kx1, and u_t : mx1

In direct analogy with the m=1 the multivariate linear regression model will be derived from first principles based on the joint distribution the observable random variables involved , $D(Z_t;\Psi)$ where $Z_t = (y_t' \; x_t')'$, (m+k)x1. Assuming that Z_t is an IID normally distributed vector , i.e,

$$\begin{pmatrix} y_t \\ x_t \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}\right) \quad \text{for all } t \in T \tag{2-2}$$

We can proceed to define

[1] The systematic and non systematic component are:

$$\eta_t = E(y_t \mid X_t = x_t) = B'x_t , \quad B = \Sigma_{22}^{-1}\Sigma_{21} \tag{2-3}$$

and

$$u_t = y_t - E(y_t \mid X_t = x_t) , \quad t \in T \tag{2-4}$$

Moreover, by construction, u_t and η_t satisfy the following properties :

(i) $E(u_t) = E\{E(u_t | X_t = x_t)\} = 0, \quad t \in T$

(ii) $E(u_t u_s') = E\{E(u_t u_s' | X_t = x_t)\}$

$$= \begin{cases} \Omega & t = s \\ 0 & t \neq s \end{cases}$$

(iii) $E(\eta_t u_t') = E\{E(\eta_t u_t' | X_t = x_t)\}$

$$= E\{\eta_t E(u_t' | X_t = x_t)\}$$

$$= 0 \qquad t \in T$$

where $\Omega = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$

[2] The statistical parameters of interest are $\Theta = (B, \Omega)$

[3] X_t is assumed to be weakly exogenous with respect to Θ.

[4] No a prior information on Θ.

[5] Rank(X) = k, $X = (x_1, x_2, \dots, x_T)'$: Txk, for T > k

(II) Probability Model [1]:

$$\Phi = D(y_t | X_t ; \Theta) = \frac{|\Omega|^{-\frac{1}{2}}}{(2\pi)^{m/2}} e^{-\frac{1}{2}(y_t - B'x_t)'\Omega^{-1}(y_t - B'x_t)}, \Theta \in R^{mk} x C^m, t \in T \quad (2\text{-}5)$$

that is

[6] (i) $D(y_t | X_t ; \Theta)$ is normal,

(ii) $E(y_t | X_t = x_t) = B'x_t$ linear in x_t,

(iii) $Cov(y_t | X_t = x_t) = \Omega$ homoskedastic (free of X_t),

[7] Θ is time invariant.

(III) Sampling model :

[8] $Y = (y_1, y_2, \dots, y_T)'$ is an independent sample drawn from $D(y_t | X_t ; \Theta)$, t = 1,2,...,T. and T ≥ m+k .

The statistical GM (2-1) for the sample period t = 1,2,..., T can be written as:

$$Y = XB + U \qquad (2\text{-}6)$$

where

$$\underset{(Txm)}{Y} = \begin{pmatrix} y_{11} & y_{12} & \cdots & y_{1m} \\ y_{21} & y_{22} & \cdots & y_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ y_{T1} & y_{T2} & \cdots & y_{Tm} \end{pmatrix}, \quad \underset{(Txk)}{X} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \vdots \\ x_{T1} & x_{T2} & \cdots & x_{Tk} \end{pmatrix} \qquad (2\text{-}7)$$

and

---

[1] R^{mk} denotes the mk- dimensional space, and C^m denotes the space of all real positive definite symmetric matrices of rank m .

$$\underset{(k\times m)}{B} = \begin{pmatrix} \beta_{11} & \beta_{12} & \cdots & \beta_{1m} \\ \beta_{21} & \beta_{22} & \cdots & \beta_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ \beta_{k1} & \beta_{k2} & \cdots & \beta_{km} \end{pmatrix} , \quad \underset{(T\times m)}{U} = \begin{pmatrix} u_{11} & u_{12} & \cdots & u_{1m} \\ u_{21} & u_{22} & \cdots & u_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ u_{T1} & u_{T2} & \cdots & u_{Tm} \end{pmatrix} \qquad (2\text{-}8)$$

It is clear that the system in (2-1) can be viewed as the t th row of (2-6).

It is usually in the measures of goodness of fit to operate entirely in terms of deviations from sample means, thus eliminating the constant terms . Hence let

$$\underset{(T\times m)}{Y}^{*} = (I_T - \frac{1}{T}\ell\ell')Y \;, \quad \underset{T\times(k-1)}{X}^{*} = (I_T - \frac{1}{T}\ell\ell')X^{o}$$
$$\underset{(T\times m)}{U}^{*} = (I_T - \frac{1}{T}\ell\ell')U \;, \quad B^{*} = (\beta_{ij}) \quad \begin{matrix} i = 2,3,\dots,k \\ j = 1,2,\dots,m \end{matrix} \qquad (2\text{-}9)$$

where $\underset{(T\times 1)}{\ell} = (1,1,\dots,1)'$ and X^{o} is the sub matrix of X resulting when we delete the first column of the latter . The model (2-6) may now be written as

$$\underset{(T\times m)}{Y^{*}} = \underset{T\times(k-1)}{X^{*}} \underset{(k-1)\times m}{B^{*}} + \underset{T\times m}{U^{*}} \qquad (2\text{-}10)$$

From assumptions [6] to [8] we can deduce [see, Muirhead,1982,p.432] that likelihood function takes the form :

$$L(B^{*},\Omega^{*}|Y^{*}) = \prod_{t=1}^{T} D(y_t^{*}|X_t^{*};B^{*},\Omega^{*})$$

and the log likelihood is

$$\log L = -\frac{mT}{2}\log 2\pi - \frac{T}{2}\log|\Omega^{*}| - \frac{1}{2}\sum_{t=1}^{T}(y_t^{*} - B^{*'}x_t^{*})'\Omega^{*-1}(y_t^{*} - B^{*'}x_t^{*})$$
$$= \text{const.} - \frac{1}{2}\{T\log|\Omega^{*}| + \operatorname{tr}\Omega^{*-1}(y_t^{*} - X^{*}B^{*'})'(y_t^{*} - X^{*}B^{*'})\} \qquad (2\text{-}11)$$

using the differentiation rules

$$\frac{\partial\log|A|}{\partial A} = A'^{-1} \;, \quad \frac{\partial\log|A|}{\partial A^{-1}} = A' \;, \quad \frac{\partial \operatorname{tr}(AB)}{\partial A} = B' \qquad (2\text{-}12)$$

we find

$$\frac{\partial \log L}{\partial B} = (X^{*'}Y^{*} - X^{*'}X^{*}B^{*})\Omega^{*-1} = 0 \qquad (2\text{-}13)$$

$$\frac{\partial \log L}{\partial \Omega^{*-1}} = \frac{T}{2}\Omega^{*} - \frac{1}{2}(Y_t^{*} - X^{*}B^{*})'(Y_t^{*} - X^{*}B^{*}) = 0 \qquad (2\text{-}14)$$

These two equations lead to the following maximum likelihood estimators (MLE's) :

$$\hat{B}^{*} = (X^{*'}X^{*})^{-1}X^{*'}Y^{*} \qquad (2\text{-}15)$$

$$\hat{\Omega}^{*} = \frac{1}{T}\hat{U}^{*'}\hat{U}^{*} \qquad (2\text{-}16)$$

where

$$\hat{U}^* = Y^* - X^*\hat{B}^* \qquad (2\text{-}17)$$

is the matrix of residuals in the entire system . For $\hat{\Omega}^*$ to be positive definite we need to assume that $T \geq m+k$ (see Dykstra 1970) . Moreover, the MLE's $\hat{B}^*$ and $\hat{\Omega}^*$, given by (2-15) and (2-16), are independently distributed , $\hat{B}^*$ is $MN(B^*, \Omega^* \otimes (X'X)^{-1})$ and $T\ \hat{\Omega}^*$ is Wishart distribution with T-k+1 degrees of freedom and scale matrix Ω^*, that is $W(\Omega^*, T\text{-}k+1)$ [see Muirhead ,1982, pp 431].

3 - Testing Goodness of Fit for The Multivariate Linear Regression Model

In the general linear model [2]

$$y^* = X^*\beta^* + u^* \qquad (3\text{-}1)$$

if we wish to test the null hypothesis that the dependent variable y^* does not depend on $X^* = [x_2^*, x_3^*, \ldots, x_k^*]$ (or, alternatively , that the fit does not improve by introducing them as explanatory variables), this can be carried out as a test of the null hypothesis :

$$H_o : \beta^* = 0 \qquad (3\text{-}2)$$

But under this null hypothesis

$$\frac{R^2}{1-R^2} \times \frac{T-k}{k-1} = \frac{\beta^{*\prime} X^{*\prime} X^* \beta^*}{\hat{u}^{*\prime} \hat{u}^*} \cdot \frac{T-k}{k-1} = F_{k-1,T-k} \qquad (3\text{-}3)$$

where $R^{*} = 1 - \frac{\hat{u}^{*\prime}\hat{u}^*}{\hat{y}^{*\prime}\hat{y}^*}$ is the familiar coefficient of determination , $\hat{\beta}^*$ is the ML (or OLS) estimator , and $\hat{u}^*$ is the residual vector . Thus , a test of the hypothesis in (3-2) can be based on the central F distribution, which is well tabulated .

As in the case of $\hat{\beta}^*$ in the linear regression model, The MLE $\hat{B}^*$ (2-15) preserves the original orthogonality between the systematic and the non-systematic component . That is, for the estimated systematic and non-systematic components

$$\hat{Y}^* = X^*\hat{B}^* \text{ and } \hat{U}^* = Y^* - \hat{Y}^* = Y^* - X^*\hat{B}^* \qquad (3\text{-}4)$$

we have

$$Y^{*\prime}Y^* = \hat{Y}^{*\prime}\hat{Y}^* + \hat{U}^{*\prime}\hat{U}^* \qquad (3\text{-}5)$$

---

2 This is corresponding to the multivariate linear regression model (2-10) .Therefore y^* (of dimension Tx1) and $X^*$of dimension Tx(k-1)) in (3-1) are in terms of deviations from the sample mean. That is, the constant term in (3-1) is eliminated.

and $\hat{y}' \perp \hat{y}$. This orthogonality can be used to extend R^2 and $1-R^2$, respectively, to

$$\underset{(m\times m)}{G} = I_m - \left(\hat{U}^{*'}\hat{U}^*\right)\left(Y^{*'}Y^*\right)^{-1}$$

$$= \left[Y^{*'}Y^* - \hat{U}^{*'}\hat{U}^*\right]\left(Y^{*'}Y^*\right)^{-1} \qquad (3\text{-}6)$$

$$\underset{(m\times m)}{H} = I - G = \left(\hat{U}^{*'}\hat{U}^*\right)\left(Y^{*'}Y^*\right)^{-1} \qquad (3\text{-}7)$$

The matrix G varies between the identity matrix when $\hat{U}^* = 0$ and zero matrix when $y' = \hat{U}^*$. In order to reduce these matrices (G and H) to scalars we can use [see: Hooper(1959), (1962)] the determinants

$$d_1 = |G|$$

$$= \frac{|Y^{*'}Y^* - \hat{U}^{*'}\hat{U}^*|}{|Y^{*'}Y^*|} = \frac{|\hat{B}^{*'}X^{*'}X^*\hat{B}^*|}{|Y^{*'}Y^*|} \qquad (3\text{-}8)$$

$$d_2 = |H| = \frac{|\hat{U}^{*'}\hat{U}^*|}{|Y^{*'}Y^*|} \qquad (3\text{-}9)$$

Now suppose that we wish to test the null hypothesis

$$H_o : B' = 0 \qquad (3\text{-}10)$$

strictly speaKing, this a hypothesis on the goodness of fit of the multiveriate linear regression model .

In the next section we shall prove that the statistic

$$F_{c_1,c_2} = \left(\frac{d_1}{d_2}\right)^{\frac{1}{m}} \frac{c_2}{c_1} \qquad (3\text{-}11)$$

where c_1 and c_2 are constant defined in (4-14) and (4-17), is approximately distributed as central F distribution with c_1 and c_2 degrees of freedom. Therefore, we may test the null hypothesis (3-10) - or the goodness of fit of the multivariate linear regression model - by the following procedure:

(1) Select a level of significance α

(2) Determine the critical value F^*_α by the condition

$$P[F_{c_1,c_2} \geq F^*_\alpha] = \alpha$$

where F_{c_1,c_2} has the central F distribution with c_1 and c_2 degrees of freedom

(3) Compute $\left(\frac{d_1}{d_2}\right)^{\frac{1}{m}} \frac{c_2}{c_1}$ and if it is less than or equal to F^* accept H_o in (3-11), otherwise reject it.

4- The Approximate Distribution of the Test Statistic in (3-11)

In deriving the approximate distribution of the test statistic in (3-11) the following two lemmas are required.

Lemma (1): [Generalization of Cochran's theorem]

Let V be a Txm matrix such that it's i th row v_i obeys $v_i' \sim MN(0, \Psi)$ Suppose that we can write

$$V'V = \sum_{i=1}^{p} V'A^i V$$

where the A^i 's are matrices with $\text{rank}(A^i) = r_i$. Then a necessary and sufficient condition that,

$$V'A^i V = \sum_{\alpha=r_1+\dots+r_{i-1}+1}^{r_1+\dots+r_i} \xi_\alpha \xi_\alpha'$$

where the ξ_α's are mutually independent, each with $MN(0, \Psi)$, $\alpha = 1,\dots,\sum_{i=1}^{p} r_i$

is that

$$\sum_{i=1}^{p} r_i = T$$

Moreover, if $r_i \geq m$, then

$$V'A^i V \sim W(\Psi, r_i)$$

Proof : see Anderson [1958,p. 165].

Lemma (2) :

If the m x m matrix A is $W(\Psi, n)$, where n ≥ m is an integer , then $|A| / |\Psi|$ has the same distribution as

$$\prod_{i=0}^{m-1} \chi^2_{n-i}$$

where the χ^2_{n-i} for i = 0,1,...,m-1, denote independent chi-square random variables.

Proof : see Muirhead [1982,pp. 100-101].

From (3-8), (3-9), and the test statistic in (3-11) we have

$$\frac{d_1}{d_2} = \frac{\left|\hat{B}^{*\prime} X^{*\prime} X^* \hat{B}^*\right|}{\left|\hat{U}^{*\prime} \hat{U}^*\right|} \tag{4-1}$$

we now show that $\hat{B}^{*\prime} X^{*\prime} X^* \hat{B}^*$ is distributed independently of $\hat{U}^{*\prime} \hat{U}^*$. But notice that

$$\hat{U}^{*'}\hat{U}^{*} = \hat{U}^{*'}\left(I - N^{*}\right)U^{*} \qquad (4\text{-}2)$$

where N^* is the symmetric idempotant matrix

$$N^{*} = X^{*}\left(X^{*'}X^{*}\right)^{-1}X^{*'} \qquad (4\text{-}3)$$

Furthermore ,

$$\hat{U}^{*'}\left(I - N^{*}\right)U^{*} = U'\left(I - \frac{1}{T}\ell\ell'\right)\left(I - N^{*}\right)\left(I - \frac{1}{T}\ell\ell'\right)U$$

$$= U'\left(I - N^{*} - \frac{1}{T}\ell\ell'\right)U \qquad (4\text{-}4)$$

This is so because

$$N^{*}\left(\frac{1}{T}\ell\ell'\right) = 0 \qquad (4\text{-}5)$$

Thus

$$U'U = U'\left(I - N^{*} - \frac{1}{T}\ell\ell'\right)U + U'N^{*}U + U'\left(\frac{1}{T}\ell\ell'\right)U$$

$$= U'M_1U + U'M_2U + U'M_3U \qquad (4\text{-}6)$$

Now it is apparent that M_i , i=1,2,3 are symmetric ,mutually orthogonal, idempotent matrices and thus

$$\left.\begin{aligned} \operatorname{rank}(M_1) &= \underline{tr}\left(I - N^{*} - \frac{1}{T}\ell\ell'\right) = T - (k-1) - 1 = T - k \\ \operatorname{rank}(M_2) &= \underline{tr}\, M_2 = k - 1 \\ \operatorname{rank}(M_3) &= \underline{tr}\left(\frac{1}{T}\ell\ell'\right) = 1 \end{aligned}\right\} \qquad (4-7)$$

But the T rows of U are mutually independent, identically distributed, each with $MN(0,\Omega\ )$ .Thus, by lemma(1), we conclude that $U'M_iU$, i = 1,2,3, are mutually independently distributed. Moreover, if

$$T\text{-}k \geq m \quad , k\text{-}1 \geq m \qquad (4\text{-}8)$$

then the two matrices $U'M_1U$ and $U'M_2U$ are distributed as $W(\Omega,T\text{-}k)$ and $W(\Omega,k\text{-}1)$ respectively. On the other hand, it is immediately obvious that

$$U'M_1U = U^{*'}\left(I - N^{*}\right)U^{*} = \hat{U}^{*'}\hat{U}^{*} \qquad (4\text{-}9)$$

$$U'M_2U = U^{*'}N^{*}U^{*} = \left(\hat{B}^{*} - B^{*}\right)'X^{*'}X^{*}\left(\hat{B}^{*} - B^{*}\right) \qquad (4\text{-}10)$$

Hence from (3-8) and (3-9) we see that, under the null hypothesis $H_o : B^* = 0$

$$\frac{d_1}{d_2} = \frac{\left|\hat{B}^{*'}X^{*'}X^{*}\hat{B}^{*}\right|}{\left|\hat{U}^{*'}\hat{U}^{*}\right|} = \frac{|U'M_2U|}{|U'M_1U|} \qquad (4\text{-}11)$$

and by lemma(1) and lemma(2) we have

$$\frac{d_1}{d_2} = \frac{\prod_{i=0}^{m-1} \chi^2_{k-1-i}}{\prod_{i=0}^{m-1} \chi^2_{T-k-i}} \tag{4-12}$$

although the exact distribution of $\frac{d_1}{d_2}$ cannot be expressed in a convenient form and be tabulated, we may approximate[3] the m th root of the numerator $\left(\prod_{i=0}^{m-1} \chi^2_{k-1-i}\right)^{\frac{1}{m}}$, by a chi-square variable whose parameter would be chosen so as to coincide with the first moment of $\left(\prod_{i=0}^{m-1} \chi^2_{k-1-i}\right)^{\frac{1}{m}}$. Thus we have

$$\left(\prod_{i=0}^{m-1} \chi^2_{k-1-i}\right)^{\frac{1}{m}} \sim \chi^2_{c_1} \tag{4-13}$$

where

$$c_1 = 2\,\frac{\prod_{i=0}^{m-1} \Gamma\left(\frac{k-1-i}{2} + \frac{1}{m}\right)}{\prod_{i=0}^{m-1} \Gamma\left(\frac{k-1-i}{2}\right)} \tag{4-14}$$

this constant is chosen so that

$$c_1 = E\left(\chi^2_{c_1}\right) = E\left(\prod_{i=0}^{m-1} \chi^2_{k-1-i}\right)^{\frac{1}{m}} = \prod_{i=0}^{m-1} E\left(\chi^2_{k-1-i}\right)^{\frac{1}{m}}$$

$$= \prod_{i=0}^{m-1} \int_0^{\infty} \frac{1}{\Gamma\left(\frac{k-1-i}{2}\right) 2^{\frac{k-1-i}{2}}}\, e^{-\frac{x}{2}}\, x^{\frac{k-1-i}{2} + \frac{1}{m} - 1}\, dx$$

$$= \prod_{i=0}^{m-1} \left[\frac{2^{\frac{1}{m}}\,\Gamma\left(\frac{k-1-i}{2} + \frac{1}{m}\right)}{\Gamma\left(\frac{k-1-i}{2}\right)}\right] \tag{4-15}$$

Similarly, we may approximate the m th root of the denominator in (4-12) by

$$\left(\prod_{i=0}^{m-1} \chi^2_{T-k-i}\right)^{\frac{1}{m}} \sim \chi^2_{c_2} \tag{4-16}$$

where

[3] For other approximation see Hoel (1937).

$$c_2 = 2\frac{\prod_{i=0}^{m-1}\Gamma\left(\frac{T-k-i}{2}+\frac{1}{m}\right)}{\prod_{i=0}^{m-1}\Gamma\left(\frac{T-k-i}{2}\right)} \tag{4-17}$$

But then it is apparent that the m th root of $\frac{d_1}{d_2}$ may be approximated, under the null hypothesis, by a central F distribution with c_1 and c_2 degrees of freedom , for the numerator and the denominator of the fraction are mutually independent. Thus we can write , approximately,

$$\left(\frac{d_1}{d_2}\right)^{\frac{1}{m}}\frac{c_2}{c_1}\sim F_{c_1,c_2} \tag{4-18}$$

and we may test H_o - or the goodness of fit of the multivariate regression model - by the procedure presented at the end of the preceding section.

References

[1] Anderson, T.W.," An Introduction to Multivariate Statistical Analysis ", John Wiely & Sons, New York, 1958.

[2] Dykstra, R.L., " Establishing the positive definiteness of the sample covariance matrix ", Annals of Mathematical Statistics, 1970, pp. 2153 - 4.

[3] Hoel, P.G., " A significance test for component analysis ", Annals of Mathematical Statistics , 1937, pp. 149-158.

[4] Hooper, J.W., "Simultaneous equations and canonical correlation theory " , Econometrica , 1959, pp. 245-250.

[5] Hooper, J.W., " Partial trace correlations " , Econometrica , 1962, pp. 324-331.

[6] Muirhead, Robb J.," Aspect of Multivariate Statistical Theory , " John Wiely & Sons, New York, 1982.