Testing General Linear Constraints on the Regression Coefficient Vector: A Note

Ki-Jun Jeong*

Consider a linear model with n observations and k explanatory variables:

(1) $\boldsymbol{y}=\boldsymbol{X\beta}+\boldsymbol{u},\ \boldsymbol{u}\sim N(\boldsymbol{0},\ \sigma^2\boldsymbol{I}_n)$.

We assume that the model satisfies the ideal conditions. Consider the general linear constraints on regression coefficient vector:

(2) $\boldsymbol{R\beta}=\boldsymbol{r}$,

where $\boldsymbol{R}$ and $\boldsymbol{r}$ are known matrices of orders $q\times k$ and $q\times 1$ respectively, and the rank of $\boldsymbol{R}$ is $q<k$. We also assume $n>k+q$. Without loss of generality $\boldsymbol{R}$ can be partitioned as $\boldsymbol{R}=(\boldsymbol{R}_1\ \boldsymbol{R}_2)$, where $\boldsymbol{R}_2$ is a $q\times q$ nonsingular matrix so that

$$\boldsymbol{R\beta}=(\boldsymbol{R}_1\ \boldsymbol{R}_2)\begin{pmatrix}\boldsymbol{\beta}_1\\ \boldsymbol{\beta}_2\end{pmatrix}=\boldsymbol{R}_1\boldsymbol{\beta}_1+\boldsymbol{R}_2\boldsymbol{\beta}_2=\boldsymbol{r},$$

or,

(3) $\boldsymbol{\beta}_2=-\boldsymbol{R}_2^{-1}\boldsymbol{R}_1\boldsymbol{\beta}_1+\boldsymbol{R}_2^{-1}\boldsymbol{r}$.

Therefore, under the constraints (2) or (3), (1) can be written as

$$\boldsymbol{y}=\boldsymbol{X\beta}+\boldsymbol{u}=(\boldsymbol{X}_1\ \boldsymbol{X}_2)\begin{pmatrix}\boldsymbol{\beta}_1\\ \boldsymbol{\beta}_2\end{pmatrix}+\boldsymbol{u}=\boldsymbol{X}_2\boldsymbol{R}_2^{-1}\boldsymbol{r}+\boldsymbol{XR}^*\boldsymbol{\beta}_1+\boldsymbol{u},$$

or,

(4) $\boldsymbol{y}^*=\boldsymbol{X}^*\boldsymbol{\beta}_1+\boldsymbol{u}$,

where

$$\boldsymbol{R}^*=\begin{pmatrix}\boldsymbol{I}_{k-q}\\ -\boldsymbol{R}_2^{-1}\boldsymbol{R}_1\end{pmatrix}$$

$$\boldsymbol{y}^*=\boldsymbol{y}-\boldsymbol{X}_2\boldsymbol{R}_2^{-1}\boldsymbol{r}$$

and $\boldsymbol{X}^*=\boldsymbol{XR}^*$.

Now, we have the following lemma:

*Department of Economics, Seoul National University

Lemma: Let $\boldsymbol{R}^*$ be a $k \times k_1$ matrix with full column rank $k_1 < k$, and let $\boldsymbol{X}^* = \boldsymbol{X}\boldsymbol{R}^*$, where $\boldsymbol{X}$ is an $n \times k$ matrix with full column rank. Define $\boldsymbol{M} = \boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$ and $\boldsymbol{M}^* = \boldsymbol{I} - \boldsymbol{X}^*(\boldsymbol{X}^{*\prime}\boldsymbol{X}^*)^{-1}\boldsymbol{X}^{*\prime}$. Then $\boldsymbol{M}$ and $\boldsymbol{M}^* - \boldsymbol{M}$ are symmetric idempotent matrices with ranks $n-k$ and $k-k_1$, respectively, and $\boldsymbol{M} \cdot (\boldsymbol{M}^* - \boldsymbol{M}) = 0$. (See Fisher [1].)

Proof: Notice

$$\boldsymbol{X}^*(\boldsymbol{X}^{*\prime}\boldsymbol{X}^*)^{-1}\boldsymbol{X}^{*\prime}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}' = \boldsymbol{X}^*(\boldsymbol{X}^{*\prime}\boldsymbol{X}^*)^{-1}\boldsymbol{R}^{*\prime}\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$$
$$= \boldsymbol{X}^*(\boldsymbol{X}^{*\prime}\boldsymbol{X}^*)^{-1}\boldsymbol{R}^{*\prime}\boldsymbol{X}' = \boldsymbol{X}^*(\boldsymbol{X}^{*\prime}\boldsymbol{X}^*)^{-1}\boldsymbol{X}^{*\prime},$$

then the proof is straightforward.

We can easily verify that $\boldsymbol{X}$, $\boldsymbol{R}^*$ and $\boldsymbol{X}^*$ in (4) satisfy the conditions of the lemma with $k_1 = k - q$, or,
$q = k - k_1$. Therefore,

(5) $\boldsymbol{y}^{*\prime}\boldsymbol{M}\boldsymbol{y}^* = \boldsymbol{y}'\boldsymbol{M}\boldsymbol{y} = \boldsymbol{u}'\boldsymbol{M}\boldsymbol{u}$

is distributed as $\sigma^2\chi^2(n-k)$, and

(6) $\boldsymbol{y}^{*\prime}(\boldsymbol{M}^* - \boldsymbol{M})\boldsymbol{y}^* = \boldsymbol{u}'(\boldsymbol{M}^* - \boldsymbol{M})\boldsymbol{u}$

is independently distributed as $\sigma^2\chi^2(q)$ under the constraints (2). (The equalities in (5) and (6) follow from the fact that $\boldsymbol{M}\boldsymbol{X} = \boldsymbol{0}$, and $\boldsymbol{M}\boldsymbol{X}_2 = \boldsymbol{0}$.) Therefore

(7) $F = \dfrac{\boldsymbol{y}^{*\prime}(\boldsymbol{M}^* - \boldsymbol{M})\boldsymbol{y}^*/q}{\boldsymbol{y}^{*\prime}\boldsymbol{M}\boldsymbol{y}^*/(n-k)}$

is distributed as $F(q, n-k)$ under the constraints (2). If we denote the LS residual vectors, $\boldsymbol{M}\boldsymbol{y}$ of (1) and $\boldsymbol{M}^*\boldsymbol{y}^*$ of (4), by $\boldsymbol{e}$ and $\boldsymbol{e}^*$, respectively, then (7) becomes

(7)′ $F = \dfrac{(e^{*\prime}e^* - e'e)/q}{e'e/(n-k)}$

Therefore, using (7) or (7)′, we can test any hypothesis given by, or reducible to, the constraints (2).

REFERENCES

[1] Fisher, F.M., "Tests of Equality between Sets of Coefficients in Two Linear Regressions: An Expository Note," *Econometrica*, 38 (1970), 361-366.

[2] Theil, H., *Principles of Econometrics*. New York; Wiley, 1971.