青森公立大学経営経済学研究
第15巻第２号（2010）

A Practical Way of Executing the F Test in the Framework of Multiple Linear Regression Equations

Hiroshi Murao ※

In linear regression analysis, hypothesis testing of linear restrictions on regression coefficients is usually done with the F test, provided that some assumptions are met. It is easy to carry out the F test in the case of single linear regression equation. However, it is not easy for the case of multiple linear regression equations. If we make the straight forward application of textbook knowledge to the F test in the framework of multiple linear regression equations, we face the problems of dealing with huge matrixes. Even though it depends on the numbers of regression coefficients and so on, the problems become serial enough to think about another way of performing the test. For example, if the number of regression coefficients is 822, then we need to write computer codes to specify the elements of a symmetric matrix in size of 822×822, and hence the number of the elements to be computed is 822×(822+1)/2=338, 253. It is just a small part of the whole task, and similar difficulties or problems occur here and there in the process of executing the F test. Nobody wants to write computer codes for such huge and tedious work. Also, we might face to the capacity limitation of statistical software or computer for dealing with huge matrices.

This paper shows a practical way of executing the F test in the framework of multiple linear regression equations. It is based on the technique of residual regressions. In this way we don't face to the problems of dealing with huge metrics. Our example shows that the F test can be carried out using a symmetric matrix in 10×10 rather 822×822. This kind of dimensional deduction makes us easy to perform the F test.

1. Introduction

In linear regression analysis, hypothesis testing of linear restrictions on regression coefficients is usually done with the F test, provided that some assumptions including the normality of error terms are met. For the case of single linear regression equation, the F test is easy to carry out and popular, reflecting the fact that a test procedure without the use of matrices is available.

On the other hand, for the case of multiple linear regression equations, the F test involves the use of matrices and seems to be less popular. The less popularity reflects difficulties or problems associated with the use of matrices. Some researchers might feel difficulties for writing computer codes to deal with matrices. We face to more serious difficulties or problems if we deal with huge matrices. For example, if the number of regression coefficients in the system is 800, then the straight forward application of textbook knowledge involves the use of the variance-covariance matrix in size of 800×800. This means that we need to write computer codes to specify each element of the symmetric matrix in 800×800, and hence the number of elements to

※ Aomori Public College Associate Professor

be computed is $800\times(800+1)/2=320,400$. This is just a small part of the whole task. Similar difficulties or problems occur here and there in the process of executing the F test. Nobody wants to write computer codes for such huge and tedious work. Also, we might face to the capacity limitation of statistical software or computer due to the use of huge matrices.

There should be a way to avoid these problems. This paper shows a practical way of executing the F test in the framework of multiple linear regression equations. It is based on the technique of residual regression. For our example case, we need to specify each element of the variance-covariance matrix in 10×10, and hence the number of elements to be computed is $10\times(10+1)/2=55$. The dimensional deduction, say from 800×800 to 10×10, makes us easy to perform the F test.

The rest of the paper is organized as follows. Section 2 introduces our example model in concrete context. Section 3 describes our estimation procedure for the model, which is closely related to our test hypotheses in Section 4. Section 4 states our test hypotheses which can be done with the F test. Section 5 reviews theoretical knowledge about the F test in the framework of multiple linear regression equations. Section 6 describes a practical method of how to execute the F test using the technique of residual regressions. Section 7 provides computational aspects of our residual regressions. Section 8 shows the results of the F tests for our example case. Section 9 concludes.

2. Our Example of the System of Multiple Linear Regression Equations

In order to make our concern clear in concrete context, we introduce a panel vector autoregression (VAR) model developed by Iwata and Murao (2007) for evaluating the effectiveness of the International Monetary Fund (IMF) lending programs. First we provide brief background information of the model, and then we show the model in details.

Developing countries often face to external payment problems rooted in underlying macroeconomic disequilibrium. Such imbalances typically show us that the country's output growth is low, inflation rate is high, and current account deficit is large. As a member of the IMF, the country may ask the IMF for financial assistance. The IMF lending program is two folds: loan and advice. The IMF's advice for the country is called as "IMF conditionality."Accepting the IMF conditionality is a key to make successive arrangements for loan.

There are a quite large number of empirical literatures that attempt to evaluate the effectiveness of the IMF lending programs. Previous studies show mixed results. For example, one group shows negative or zero impacts of the IMF lending programs on the country's output growth while the other group shows positive impacts. Beside such results, a problem of previous studies is that they don't provide a dynamic feature of evaluation over the wide range of time horizon. Another problem is that previous studies attempt to estimate only the total effect of the IMF lending programs.

With regard to the IMF lending programs, we need to deal with a statistical problem called as "self-selection problem" or "sample selection problem." The program participation by a country is a product of interactions between the desire of the country for financial help and the IMF's willingness to lend, which in turns are related to the country's underlying macroeconomic situations. In this sense it is no doubt that the program participation is endogenously determined. Ignoring such endogeneity, the usual OLS estimation provides neither unbiased estimation nor consistent estimation. In order to fix this kind of problem, we can utilize sample selection models such as the type Ⅱ and Ⅲ Tobit models. Some estimation procedures are also developed for correcting sample selection bias. Heckman (1976) developed a popular two-step estimation procedure for correcting sample selection bias and his procedure provides consistent estimation. With regard to the hypothesis test of sample selection bias in censored regression models, Vella (1992) suggested that the test can be done with the usual t test using a modified linear regression equation. Wooldridge (1998) showed that Vella's procedure also solves the sample selection problem.

Considering the problems of previous approaches as well as the sample selection problem, Iwata and Murao (2007) have developed a new approach for evaluating the IMF lending programs. Unlike previous approaches, it is based on a VAR model with regime switching so that it provides a dynamic feature of evaluation over the wide range of time horizon. Unlike previous approaches, it provides a way to estimate not only the total effect of IMF programs, but also the loan effect and the policy advice effect. This kind of separation is often very important in policy discussion.

Our VAR model consists of three groups of equations. The first two groups of equations describe the behavior of what we call the policy variables and the target variables, while the third equation describes the selection mechanism for the IMF loan program.

The policy variables are a set of policy instruments that the government can control through its monetary, fiscal and exchange rate policies. The target variables are the country's macroeconomic performance variables such as output growth, inflation, and the balance of payments. The third equation mimics the IMF's selection criterion that determines whether a given country in a given period is "in" or "out" of the IMF program.

As pointed out previously, the IMF program has two major aspects: the loan provision and the policy advice. To distinguish the two different channels through which the IMF program possibly operates, we make the following specification. Since the policy advice often aims at a systematic change in the pattern of the policy reaction of the government, the policy equation is assumed to switch between two regimes: in-program and out-program.

To formally introduce our VAR model, we define the following variables.

$\mathbf{y}_{it}^P =$ an $m^P \times 1$ vector of policy variables in country i in period t

$\mathbf{y}_{it}^T =$ an $m^T \times 1$ vector of target variables in country i in period t

$d_{it} =$ a program dummy taking on the value 1 or 0 depending on the IMF program is in effect in country i in period t

$L_{it} =$ the amount of the IMF loan to country i in period t.

Let $\mathbf{y}_{it} = [\mathbf{y}_{it}^{P\prime}, \mathbf{y}_{it}^{T\prime}]'$, which is an $m \times 1$ vector with $m = m^P + m^T$. Then the structural model for $\mathbf{y}_{it}^P$ and $\mathbf{y}_{it}^T$ is given by

$$\mathbf{A}_{11}\mathbf{y}_{it}^P = \mathbf{a}_i^{Pd} + \mathbf{A}_{y1}^{Pd}\mathbf{y}_{i,t-1} + \cdots + \mathbf{A}_{yp}^{Pd}\mathbf{y}_{i,t-p} + \boldsymbol{\varepsilon}_{it}^P \tag{1a}$$

$$\mathbf{A}_{22}\mathbf{y}_{it}^T = \mathbf{a}_i^T + \mathbf{A}_{12}\mathbf{y}_{it}^P + \mathbf{A}_{y1}^T\mathbf{y}_{i,t-1} + \cdots + \mathbf{A}_{yp}^T\mathbf{y}_{i,t-p} + \mathbf{a}_{L0}^T L_{it} + \cdots + \mathbf{a}_{Lp}^T L_{i,t-p} + \boldsymbol{\varepsilon}_{it}^T \tag{1b}$$

for $d = 1,0$, where $\mathbf{A}_{11}$, $\mathbf{A}_{22}$ and $\mathbf{A}_{12}$ are $m^P \times m^P$, $m^T \times m^T$ and $m^T \times m^P$ matrices, $\mathbf{A}_{yj}^{Pd}$ and $\mathbf{A}_{yj}^T$ are $m^P \times m$, $m^T \times m$ matrices of regression coefficients for j=1,. . . , p, and $\mathbf{a}_{Lj}^T$ is an $m^T\times 1$ vector of regression coefficients for j=0,. . . , p. The first equation above is a set of policy reaction functions that switch between two regimes: in-program (d= 1) and out-program (d= 0). The second equation describes each target variable as a function of contemporaneous policy variables, the amount of IMF loan, and all lagged variables. Note that the contemporaneous target variables and the IMF loan variables are excluded from the policy equation (1a), while the contemporaneous policy variables and the IMF loan variables are included in the target equation (1b). These points of specification are based on highly standardized facts even though we don't explain in detail since they are not related to our concern in this paper.

We now turn to the selection mechanism. We need to consider the endogeneity of the program participation, as pointed out previously. This mechanism is usually characterized by a probit model. That is, the program participation (d= 0 or 1) is formulated as a function of underlying macroeconomic and other variables. However, the amount of loan (L_{it}) is observable and the IMF program dummy (d_{it}) is 1 only when L_{it} is positive. A more efficient way to model the selection process is, therefore, to use a Tobit model. Let L_{it}^* be the potential loan amount implied by the IMF standardized formula, which can be observed only when L_{it}^* is positive. We then postulate that the selection process is governed by

$$L_{it}^* = \mathbf{w}_i^{t-1}{}'\boldsymbol{\theta} + v_{it}, \tag{2}$$

where $\mathbf{w}_i^{t-1}$ is a vector of historical records on the target, policy and other variables. The actual observed loan L_{it} and the IMF program dummy d_{it} are given by

$$L_{it} = L_{it}^* d_{it} \quad \text{and} \quad d_{it} = 1(L_{it}^* > 0) \tag{3}$$

Rewriting (1a) and (1b) yields

$$\mathbf{y}_{it}^P = \mathbf{b}_i^{Pd} + \mathbf{B}_{y1}^{Pd}\mathbf{y}_{i,t-1} + \cdots + \mathbf{B}_{yp}^{Pd}\mathbf{y}_{i,t-p} + \mathbf{u}_{it}^P \tag{4a}$$

$$\mathbf{y}_{it}^T = \mathbf{b}_i^T + \mathbf{B}_{12}\mathbf{y}_{it}^P + \mathbf{B}_{y1}^T\mathbf{y}_{i,t-1} + \cdots + \mathbf{B}_{yp}^T\mathbf{y}_{i,t-p} + \mathbf{b}_{L0}^T L_{it} + \cdots + \mathbf{b}_{Lp}^T L_{i,t-p} + \mathbf{u}_{it}^T \tag{4b}$$

where $\mathbf{b}_i^{Pd} = \mathbf{A}_{11}^{-1}\mathbf{a}_i^{Pd}$, $\mathbf{B}_{yj}^{Pd} = \mathbf{A}_{11}^{-1}\mathbf{A}_{yj}^{Pd}$, $\mathbf{b}_i^T = \mathbf{A}_{22}^{-1}\mathbf{a}_i^T$, $\mathbf{B}_{12} = \mathbf{A}_{22}^{-1}\mathbf{A}_{12}$, $\mathbf{B}_{yj}^T = \mathbf{A}_{22}^{-1}\mathbf{A}_{yj}^T$ for j=1,. . . , p, and $\mathbf{b}_{Lj}^T = \mathbf{A}_{22}^{-1}\mathbf{a}_{Lj}^T$ for j=0,. . . , p. The error term v_{it} in (2) is assumed to be correlated with the shocks to the system $\mathbf{u}_{it} = [\mathbf{u}_{it}^{P\prime}, \mathbf{u}_{it}^{T\prime}]'$. We can write the error terms $\mathbf{u}_{it}$ and v_{it} jointly as:

$$\begin{bmatrix} \mathbf{u}_{it}^P \\ \mathbf{u}_{it}^T \\ v_{it} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{11}^{-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{22}^{-1} & \mathbf{0} \\ \mathbf{c}_P' & \mathbf{c}_T' & c_L \end{bmatrix} \begin{bmatrix} \boldsymbol{\varepsilon}_{it}^P \\ \boldsymbol{\varepsilon}_{it}^T \\ \varepsilon_{it}^L \end{bmatrix} \equiv \boldsymbol{\Gamma}\boldsymbol{\varepsilon}_{it} \tag{5}$$

where $\boldsymbol{\varepsilon}_{it} \sim \text{iid}(\mathbf{0},\mathbf{1})$ is a vector of structural shocks. The term $\boldsymbol{\varepsilon}_{it}^L$ stands for an exogenous shock to the IMF loan. The zero restrictions on matrix $\boldsymbol{\Gamma}$ in (5) imply that $\mathbf{y}_{it}$ is not affected contemporaneously by this shock. This assumption simply reflects the notion that the effect of the loan would not be materialized immediately. This obviously does not rule out the possibility that the shock will have influence on all variables in $\mathbf{y}_{it}$ with delay. On the other hand, the IMF loan decision is affected by all contemporaneous exogenous shocks to all variables in $\mathbf{y}_{it}$.

3. Estimation

We can write the likelihood function of the model and maximize it with respect to its parameters to get the maximum likelihood estimates. We do not follow this approach, however, because it does not work well with a large number of parameters. The number of parameters is around 900 including the parameters in the variance-covariance matrix for our panel VAR analysis.

Instead of using the maximum likelihood method, we employ the recursive estimation procedure based on the partial likelihood approach (Vella 1992). Note that the conditional expectations of $\mathbf{y}_{it}^{P}$ and $\mathbf{y}_{it}^{T}$ in (4) given v_{it} and d_{it} together with the lagged $\mathbf{y}_{it}$ are given by

$$E(\mathbf{y}_{it}^{P} \mid \mathbf{y}_{i}^{t-1}, v_{it}, d_{it}) = \mathbf{b}_{i}^{Pd} + \mathbf{B}_{y1}^{Pd}\mathbf{y}_{i,t-1} + \cdots + \mathbf{B}_{yp}^{Pd}\mathbf{y}_{i,t-p} + \boldsymbol{\gamma}_{P}^{d} v_{it} \tag{6a}$$

$$E(\mathbf{y}_{it}^{T} \mid y_{it}^{P}, \mathbf{y}_{i}^{t-1}, v_{it}, \mathbf{u}_{it}^{P}, d_{it}) = \mathbf{b}_{i}^{T} + \mathbf{B}_{12}\mathbf{y}_{it}^{P} + \mathbf{B}_{y1}^{T}\mathbf{y}_{i,t-1} + \cdots + \mathbf{B}_{yp}^{T}\mathbf{y}_{i,t-p} + \mathbf{b}_{L0}^{T} L_{it} + \ldots + \mathbf{b}_{Lp}^{T} L_{i,t-p} + \boldsymbol{\gamma}_{T} v_{it} + \boldsymbol{\delta}\mathbf{u}_{it}^{P} \tag{6b}$$

where $\mathbf{y}_{i}^{t-1} = \{\mathbf{y}_{i,t-1}, \mathbf{y}_{i,t-2}, \ldots\}$, $\boldsymbol{\gamma}_P = \mathbf{c}_P/\sigma_{vv}$, $\boldsymbol{\gamma}_T = \mathbf{c}_P/\sigma_{vv}$, $\delta = \mathbf{B}_{12}\mathbf{A}_{22}^{-1}$ and $\sigma_{vv} = \mathbf{c}_P'\mathbf{c}_P + \mathbf{c}_T'\mathbf{c}_T + \mathbf{c}_L^2$. In other words, we can obtain consistent estimates of the parameters of equation (4a), free of selectivity bias, by running separate regressions for $d=0$ and $d=1$ when including v_{it} as an additional regressor. Also, we can avoid the endogeneity problem of $\mathbf{y}_{it}^{P}$ and L_{it} for the regression (4b), if we include v_{it} and $\mathbf{u}_{it}^{P}$ as additional regressors. Of course v_{it} and $\mathbf{u}_{it}^{P}$ are unobservable, but they can be estimated. First, we estimate θ from the selection equation (2) by maximizing the Tobit likelihood. Then we obtain $\hat{v}_{it} = L_{it} - \mathbf{w}_{i}^{t-1\prime}\hat{\theta}$. Next, we run the OLS regression of $\mathbf{y}_{it}^{P}$ on $\mathbf{y}_{i}^{t-1}$ and $\hat{v}_{it}$, and then compute:

$$\hat{\mathbf{u}}_{it}^{P} = \mathbf{y}_{it}^{P} - \hat{\mathbf{b}}_{i}^{Pd} - \hat{\mathbf{B}}_{y1}^{Pd}\mathbf{y}_{i,t-1} - \ldots - \hat{\mathbf{B}}_{yp}^{Pd}\mathbf{y}_{i,t-p} - \hat{\boldsymbol{\gamma}}_{p}^{d}\hat{v}_{it} \tag{7}$$

Finally, we run the OLS regression of $\mathbf{y}_{it}^{T}$ on $\mathbf{y}_{it}^{P}$, $\mathbf{y}_{i}^{t-1}$, $\hat{v}_{it}$ and $\hat{\mathbf{u}}_{it}^{P}$. The resulting estimates of $\mathbf{b}_{i}^{P}$, $\mathbf{b}_{i}^{T}$ $\mathbf{B}_{12}\mathbf{B}_{yj}^{Pd}$, $\mathbf{B}_{Lj}^{T}$ and $\mathbf{b}_{Lj}^{T}$ are all consistent (Rivers and Vuong 1998).

4. Our Test Hypotheses for the F Test

The estimation procedure in the previous section can be used for testing the selectivity bias as well as the endogeneity of $\mathbf{y}_{it}^{P}$ and L_{it}. We can test the selectivity bias in (6a) by testing a hypothesis that each component of $\boldsymbol{\gamma}_P$ equal zero. Also the endogeneity of $\mathbf{y}_{it}^{P}$ and L_{it} in (6b) can be tested by testing hypotheses $\boldsymbol{\delta} = \mathbf{0}$ and $\boldsymbol{\gamma}_T = \mathbf{0}$.

As an example, we focus on testing the sample selectivity bias. If the components of $\boldsymbol{\gamma}_P$ are tested separately, then we can use the usual t test as suggested by Vella (1992). However, our interest is whether or not sample selection bias presents in the system of multiple equations. Hence our test would be the F test. We can think of three null hypotheses for our interest.

$$H_0 : \boldsymbol{\gamma}_P = \mathbf{0} \tag{8a}$$

$$H_0 : \boldsymbol{\gamma}_T = \mathbf{0} \tag{8b}$$

$$H_0 : \boldsymbol{\gamma} = \mathbf{0} \tag{8c}$$

where $\boldsymbol{\gamma} = [\boldsymbol{\gamma}_P', \boldsymbol{\gamma}_T']'$. Our VAR analysis uses three policy variables: fiscal deficits, a change in domestic credit, and a change in exchange rate. Each policy equation has two regimes: in-program and out-program. This means that $\boldsymbol{\gamma}_P = \mathbf{0}$ has 6 linear restrictions on the regression coefficients across the six different policy equations in (6a). Our VAR analysis also uses three target variables: output growth, inflation rate, and the balance of payments. This means that $\boldsymbol{\gamma}_T = \mathbf{0}$ has 3 restrictions on the regression coefficients across the three different target equations in (6b). Putting $\boldsymbol{\gamma}_P = \mathbf{0}$ and $\boldsymbol{\gamma}_T = \mathbf{0}$ together, $\boldsymbol{\gamma} = \mathbf{0}$ has 9 restrictions.

5. Theoretical Review for the F Test

This section provides theoretical basic knowledge for the F test in the framework of linear multiple regression equations. In the connection with (6) and (8), we consider the reduced form of multiple equations. We suppose that there are m dependent variables indexed by j and there are T observations on each dependent variable. Let y_j denote the vector of T observations on the j-th dependent variable, $\mathbf{X}_j$ denote the $T \times k_j$ matrix of observations on regressors in the j-th equation, β_j denote the k_j-vector of regression coefficients in the j-th equation, and u_j denote the T-vector of error terms for the j-th equation. The j-th equation in the system of multiple linear regression equations can be written as

$$y_j = \mathbf{X}_j \beta_j + u_j \tag{9}$$

By making appropriate definition, we can write the entire system of m linear equations as

$$\boldsymbol{y}_\bullet = \mathbf{X}_\bullet \beta_\bullet + \boldsymbol{u}_\bullet \tag{10}$$

where $\boldsymbol{y}_\bullet$ is the mT-vector consisting of the T-vector y_1 through y_m stacked vertically, $\boldsymbol{u}_\bullet$ is similarly the vector of u_1 through u_m stacked vertically. The matrix $\mathbf{X}_\bullet$ is a $mT \times K$ block-diagonal matrix, where K is equal to $\sum_{j=1}^{m} k_j$. If m=3, then the system looks

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} \mathbf{X_1} & \mathbf{O} & \mathbf{O} \\ \mathbf{O} & \mathbf{X_2} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} & \mathbf{X_3} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} \tag{11}$$

where each of the $\mathbf{O}$ blocks has T rows and as many columns as the $\mathbf{X}_j$ block that it shares those columns with. To be conformable with $X_\bullet$, the vector $\boldsymbol{\beta}_\bullet$ is a K-vector consisting of the vector β_1 through β_m. By making some assumptions on the error terms, the variance-covariance matrix of the error vector $\boldsymbol{u}_\bullet$ is written as

$$\begin{aligned} E(\boldsymbol{u}_\bullet \boldsymbol{u}_\bullet') &= \begin{bmatrix} E(u_1 u_1') & E(u_1 u_2') & \cdots & E(u_1 u_m') \\ E(u_2 u_1') & E(u_2 u_2') & \cdots & E(u_2 u_m') \\ \vdots & \vdots & \vdots & \vdots \\ E(u_m u_1') & E(u_2 u_m') & \cdots & E(u_m u_m') \end{bmatrix} \\ &= \begin{bmatrix} \sigma_{11}\mathbf{I} & \sigma_{12}\mathbf{I} & \cdots & \sigma_{1m}\mathbf{I} \\ \sigma_{21}\mathbf{I} & \sigma_{22}\mathbf{I} & \cdots & \sigma_{1m}\mathbf{I} \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{m1}\mathbf{I} & \sigma_{m2}\mathbf{I} & \cdots & \sigma_{mm}\mathbf{I} \end{bmatrix} \\ &= \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1m} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{m1} & \sigma_{m2} & \cdots & \sigma_{mm} \end{bmatrix} \otimes \mathbf{I} = \mathbf{\Sigma} \otimes \mathbf{I} \equiv \mathbf{\Sigma}_\bullet \end{aligned} \tag{12}$$

where

$$\mathbf{\Sigma} \equiv \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1m} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{m1} & \sigma_{m2} & \cdots & \sigma_{mm} \end{bmatrix}$$

Note that $\mathbf{\Sigma}$ is a symmetric $m \times m$ matrix and $\mathbf{\Sigma}_\bullet \equiv \mathbf{\Sigma} \otimes \mathbf{I}$ is a $mT \times mT$ matrix. By making the assumption of normality for the error vector $\boldsymbol{u}_\bullet$, the whole system can be written as

$$\boldsymbol{y}_\bullet = X_\bullet \boldsymbol{\beta}_\bullet + \boldsymbol{u}_\bullet \tag{13a}$$

$$\boldsymbol{u}_\bullet \sim N(\mathbf{0}, \mathbf{\Sigma} \otimes \mathbf{I}) \tag{13b}$$

Now we consider a set of J linear restrictions on the form

$$H_0 : \mathbf{R}\boldsymbol{\beta}_\bullet = \mathbf{c} \tag{14}$$

where $\mathbf{R}$ is a matrix of know constants, and $\mathbf{c}$ is a vector of know constants. The linear equation $\mathbf{R}\beta_\bullet=\mathbf{c}$ is used to specify any linear restrictions on the regression coefficients. Textbooks such as Greene (2000, p620) show the following F test statistic for testing the null hypothesis.

$$F=\frac{\dfrac{(\mathbf{R}\hat{\beta}_\bullet-c)'[\mathbf{R}(\mathbf{X}_\bullet'(\Sigma\otimes\mathbf{I})^{-1}\mathbf{X}_\bullet)^{-1}\mathbf{R}']^{-1}(\mathbf{R}\hat{\beta}_\bullet-c)}{J}}{\dfrac{\hat{u}_\bullet'(\Sigma\otimes\mathbf{I})^{-1}\hat{u}_\bullet}{mT-K}}\sim F(J,mT-K) \tag{15}$$

Note that $(\mathbf{X}_\bullet'(\Sigma\otimes\mathbf{I})^{-1}\mathbf{X}_\bullet)^{-1}=Var(\hat{\beta}_\bullet)$. Using a consistent estimate $\hat{\Sigma}$, the F test statistic reduces to

$$F=\frac{1}{J}(\mathbf{R}\hat{\beta}_\bullet-c)'[\hat{V}ar(\mathbf{R}\hat{\beta}_\bullet)]^{-1}(\mathbf{R}\hat{\beta}_\bullet-c)\sim F(J,mT-K) \tag{16}$$

Because the above F test statistic uses the estimated Σ, the F distribution is only asymptotically valid.

For our panel VAR analysis, the stacked $\hat{\beta}_\bullet$ consists of 822 regression coefficients, and hence the dimension of $\hat{V}ar(\hat{\beta}_\bullet)$ is 822×822. This means that we need to write computer codes to calculate each element of the symmetric matrix in 822×822, and hence the number of the elements to be computed is 822×(822+1)/2=338, 253. Moreover, it is just a small part of the whole task of executing the F test. Nobody wants to write computer codes for such huge and tedious work.

Another problem is related to the matrix $\hat{\Sigma}\otimes\mathbf{I}$ if we use it. We have m=10 including the selection equation for our panel VAR analysis. If we use a data set of 2000 observations and m=10, then the dimension of $\hat{\Sigma}\otimes\mathbf{I}$ is 20,000×20,000. No body wants to deal with such a huge matrix for executing the F test. We need to consider the capacity limit of statistical software or computer in order to invert the matrix of size in 20,000×20,000.

It is clear that we are facing to the problems of dealing with huge matrices. There should be a way to avoid these problems. The next section provides an idea of practical way of executing the F test so that we don't need to deal with huge matrices.

6. A Practical Way of Executing the F Test

This section shows how to apply the previous basic knowledge to the standard model specification that is similar to equation (6). It also provides theoretical explanation of our practical way of executing the F test. We consider again the stacked from in (13).

$$\boldsymbol{y}_\bullet=X_\bullet\beta_\bullet+\boldsymbol{u}_\bullet$$

$$\boldsymbol{u}_\bullet\sim N(\mathbf{0},\Sigma\otimes\mathbf{I})$$

First we make a partition of $\beta_\bullet$ as $[\beta,\boldsymbol{\gamma}]'$ where $\boldsymbol{\gamma}$ denotes the part of our interest or the part for testing, and β denotes the rest of $\beta_\bullet$. The stacked $X_\bullet$ is also partitioned as $[\mathbf{X},\mathbf{Z}]$ so that it is conformable with $[\beta,\boldsymbol{\gamma}]'$. In the connection with equation (6) and $H_0:\boldsymbol{\gamma}=\mathbf{0}$, each of the m equations is assumed to have one γ_j. Then equation (13a) can be written as

$$\boldsymbol{y}_\bullet=\mathbf{X}\beta+\mathbf{Z}\boldsymbol{\gamma}+\boldsymbol{u}_\bullet \tag{17}$$

If m=3 , then it looks

$$\begin{bmatrix}y_1\\y_2\\y_3\end{bmatrix}=\begin{bmatrix}\mathbf{X}_1&\mathbf{O}&\mathbf{O}\\\mathbf{O}&\mathbf{X}_2&\mathbf{O}\\\mathbf{O}&\mathbf{O}&\mathbf{X}_3\end{bmatrix}\begin{bmatrix}\beta_1\\\beta_2\\\beta_3\end{bmatrix}+\begin{bmatrix}\mathbf{Z}_1&\mathbf{O}&\mathbf{O}\\\mathbf{O}&\mathbf{Z}_2&\mathbf{O}\\\mathbf{O}&\mathbf{O}&\mathbf{Z}_3\end{bmatrix}\begin{bmatrix}\gamma_1\\\gamma_2\\\gamma_3\end{bmatrix}+\begin{bmatrix}u_1\\u_2\\u_3\end{bmatrix} \tag{18}$$

Notice that Z_j can be interpreted as the vector of v_{it} in equation (6). The system (17) has the following dimensions.

$$mT \times 1 \quad \text{vs.}$$
$$(mT \times K^*)(K^* \times 1) + (mT \times m)(m \times 1) + (mT \times 1)$$

where $K^* = \sum_{j=1}^{m} k_j$, and k_j is the number of coefficients for X_j in the j-th equation.

Let σ_{ij} be the (i,j) element of the variance-covariance matrix Σ, and σ^{ij} be the (i,j) element of the inverse matrix Σ^{-1}. Then, the variance-covariance matrix of $\hat{\beta}_\bullet$ is given by

$$Var(\hat{\beta}_\bullet) = \left[\mathbf{X}_\bullet{}'(\Sigma^{-1} \otimes \mathbf{I})\mathbf{X}_\bullet\right]^{-1}$$

$$Var(\hat{\beta}_\bullet) = \begin{bmatrix} \sigma^{11}\mathbf{X}_{\bullet 1}{}'\mathbf{X}_{\bullet 1} & \sigma^{12}\mathbf{X}_{\bullet 1}{}'\mathbf{X}_{\bullet 2} & \cdots & \sigma^{1m}\mathbf{X}_{\bullet 1}{}'\mathbf{X}_{\bullet m} \\ \sigma^{12}\mathbf{X}_{\bullet 2}{}'\mathbf{X}_{\bullet 1} & \sigma^{22}\mathbf{X}_{\bullet 2}{}'\mathbf{X}_{\bullet 2} & \cdots & \sigma^{2m}\mathbf{X}_{\bullet 2}{}'\mathbf{X}_{\bullet m} \\ \vdots & \vdots & & \vdots \\ \sigma^{1m}\mathbf{X}_{\bullet m}{}'\mathbf{X}_{\bullet 1} & \sigma^{2m}\mathbf{X}_{\bullet m}{}'\mathbf{X}_{\bullet 2} & \cdots & \sigma^{mm}\mathbf{X}_{\bullet m}{}'\mathbf{X}_{\bullet m} \end{bmatrix}^{-1} \tag{19}$$

The estimator of $\boldsymbol{\gamma}$, which is a part of $\hat{\beta}_\bullet$, is given by

$$\hat{\boldsymbol{\gamma}} = [\mathbf{Z}'\mathbf{M}_X \mathbf{Z}]^{-1}\mathbf{Z}'\mathbf{M}_X y_\bullet \tag{20}$$

where $\mathbf{M}_X = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. Hence we have

$$\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma} = [\mathbf{Z}'\mathbf{M}_X \mathbf{Z}]^{-1}\mathbf{Z}'\mathbf{M}_X u_\bullet \tag{21}$$

The variance-covariance matrix of $\hat{\boldsymbol{\gamma}}$ is given by

$$\begin{aligned} Var(\hat{\boldsymbol{\gamma}}) &= E[(\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma})(\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma})'] = E[[\mathbf{Z}'\mathbf{M}_X\mathbf{Z}]^{-1} \\ &\quad (\mathbf{Z}'\mathbf{M}_X u_\bullet u_\bullet' \mathbf{Z}\mathbf{M}_X)[\mathbf{Z}'\mathbf{M}_X\mathbf{Z}]^{-1}] \\ &= [\mathbf{Z}'\mathbf{M}_X\mathbf{Z}]^{-1}\mathbf{Z}'\mathbf{M}_X \Sigma \otimes \mathbf{I}\mathbf{Z}\mathbf{M}_X[\mathbf{Z}'\mathbf{M}_X\mathbf{Z}]^{-1} \end{aligned}$$

$$Var(\hat{\boldsymbol{\gamma}}) = \begin{bmatrix} \sigma^{11}\mathbf{Z}_1{}'\mathbf{M}_X\mathbf{Z}_1 & \sigma^{12}\mathbf{Z}_1{}'\mathbf{M}_X\mathbf{Z}_2 & \cdots & \sigma^{1m}\mathbf{Z}_1{}'\mathbf{M}_X\mathbf{Z}_m \\ \sigma^{12}\mathbf{Z}_2{}'\mathbf{M}_X\mathbf{Z}_1 & \sigma^{22}\mathbf{Z}_2{}'\mathbf{M}_X\mathbf{Z}_2 & \cdots & \sigma^{2m}\mathbf{Z}_2{}'\mathbf{M}_X\mathbf{Z}_m \\ \vdots & \vdots & & \vdots \\ \sigma^{1m}\mathbf{Z}_m{}'\mathbf{M}_X\mathbf{Z}_1 & \sigma^{2m}\mathbf{Z}_m{}'\mathbf{M}_X\mathbf{Z}_2 & \cdots & \sigma^{mm}\mathbf{Z}_m{}'\mathbf{M}_X\mathbf{Z}_m \end{bmatrix}^{-1} \tag{22}$$

We define the following terms.

$\mathbf{M}_X y_j \equiv \tilde{y}_j$ = residuals of y_j regressed on $\mathbf{X}_j$ for the j-th equation.

$\mathbf{M}_X \mathbf{Z}_j \equiv \tilde{\mathbf{Z}}_j$ = residuals of $\mathbf{Z}_j$ regressed on $\mathbf{X}_j$ for the j-th equation.

Using the technique of residual regressions, $\hat{\gamma}_j$ can be obtained from the regression of $\tilde{y}_j$ on $\tilde{\mathbf{Z}}_j$. Using $\tilde{\mathbf{Z}}_j$ notation, the (j,k) element of $Var(\hat{\boldsymbol{\gamma}})^{-1}$ is given by

$$c_{j,k} = \sigma^{ik}\tilde{\mathbf{Z}}_j{}'\tilde{\mathbf{Z}}_k \tag{23}$$

The matrix $Var(\hat{\boldsymbol{\gamma}})$ is now written as

$$Var(\hat{\boldsymbol{\gamma}}) = \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1m} \\ c_{21} & c_{22} & \cdots & c_{2m} \\ \vdots & \vdots & & \vdots \\ c_{m1} & c_{m2} & \cdots & c_{mm} \end{bmatrix}^{-1} \tag{24}$$

This matrix is estimated as $\hat{Var}(\hat{\boldsymbol{\gamma}})$ using the estimate of σ^{ij}. The F test statistic for H_0: $\boldsymbol{\gamma} = \mathbf{0}$ is given by

$$F = \frac{1}{J}\hat{\boldsymbol{\gamma}}'[\hat{Var}(\hat{\boldsymbol{\gamma}})]^{-1}\hat{\boldsymbol{\gamma}} \sim F(J, n_2) \tag{25}$$

where $J=m$ is the number of restrictions, $n_2 = mT-K$, mT is the number of observations on m endogenous variables, and K is the number of all regression coefficients including γ_j in the whole system.

Notice the following points. In the course of computing the F test statistic, we don't use huge elements such as $\hat{\beta}_\bullet$ and $\hat{Var}(\hat{\beta}_\bullet)$. The elements we use are $\hat{\boldsymbol{\gamma}}$ and $\hat{Var}(\hat{\boldsymbol{\gamma}})$. The dimension of $\hat{\boldsymbol{\gamma}}$ is 10 and $\hat{Var}(\hat{\boldsymbol{\gamma}})$ is 10×10 if $J=m=10$, while the dimension of $\hat{\beta}_\bullet$ is 822 and $\hat{Var}(\hat{\beta}_\bullet)$ is 822×822 in our example case.

So far we assume that each equation has one γ_j in $H_0: \boldsymbol{\gamma} = \mathbf{0}$ so that we have $J=m$. Our idea can be extended to other cases as well. The important point is that we need to stack the coefficients (γ_j) of our interest vertically as shown in (18). Then β_j denotes all other coefficients in the equation and (X_j, Z_j) is defined accordingly. Then we use the technique of residual regressions so that we can estimate γ_j for $j=1,2,...,J$ and get $\hat{V}ar(\hat{\boldsymbol{\gamma}})^{-1}$. The rest is basically the same as above. Even J becomes larger than m, it would be still much smaller than K, which is the dimension of the whole coefficient vector $\boldsymbol{\beta}_{\bullet}$.

7. Computational Aspects of the F Test for Sample Selection Bias

The null hypothesis in (8c) has 9 restrictions for testing the presence of sample selection bias in the whole system of multiple equations. Thus, we consider the case of $J=9$ in order to explain computational aspects. We make the following stacked from.

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_9 \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{O} & \cdots & \mathbf{O} \\ \mathbf{O} & \mathbf{X}_2 & \cdots & \mathbf{O} \\ \vdots & \vdots & & \vdots \\ \mathbf{O} & \mathbf{O} & \cdots & \mathbf{X}_9 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_9 \end{bmatrix} + \hat{v} \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_9 \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_9 \end{bmatrix} \tag{26}$$

where the first through sixth equations represent the three policy equations under two regimes $d=0$ and $d=1$, and the seventh through ninth equations represent the three target equations. Note that β_j denotes for all regression coefficients expect γ_j in the j-th equation, and X_j is defined accordingly. Note also that the selection equation (2) does not appear in the above stacked form since it does not have a coefficient for the test. Of course all of the equations including (2) are used when we compute $\hat{\Sigma}$.

The above stacked from is written as

$$\boldsymbol{y}_{\bullet} = \mathbf{X}\boldsymbol{\beta} + \hat{v}\boldsymbol{\gamma} + \boldsymbol{u}_{\bullet} \tag{27}$$

Since our data set is a panel data set of many countries with annual frequency, we let N be the number of countries, T be the number of time periods, mNT be the number of observations on m dependent variables. Then the stacked $y_{\bullet}$ has the dimension of $mNT \times 1$.

Next we show the computational steps of the F statistic. As the first step, we obtain all regression residuals in the following order. We run the selection equation (2) first and obtain its residual $\hat{v}_{it}$. Following equation (6), we run the policy equation with $\hat{v}_{it}$, and then we obtain $\hat{u}_{it}^{P}$. We run the target equation with $\hat{v}_{it}$ and $\hat{u}_{it}^{P}$, obtaining $\hat{u}_{it}^{T}$. In this way we can get all regression residuals. The second step is to obtain the estimated variance-covariance matrix of error terms as follows.

$$\hat{\sigma}_{jk} = \frac{1}{NT} \sum_{t=1}^{NT} \hat{u}_{jt} \hat{u}_{kt} \tag{28}$$

$$\hat{\boldsymbol{\Sigma}} = (\hat{\sigma}_{jk}) \tag{29}$$

where $\hat{v}_{jt}$ is denoted as $\hat{u}_{jt}$ and equation (29) means that $\hat{\Sigma}$ consists of $\hat{\sigma}_{jk}$ for $j,k=1,2,3,...,10$. The third step is to invert $\hat{\Sigma}$ and obtain $\hat{\sigma}^{jk}$ in $\hat{\Sigma}^{-1}$.

The forth step is the time to consider the stack form (26) for the technique of residual regressions. Here we regress y_j on X_j including the country dummies but excluding $\hat{v}$, and obtain its residuals $\tilde{y}_j$ for $j=1,2,...,9$. We also regress $\hat{v}$ on X_j including the country dummies, and obtain its residual $\tilde{v}_j$ for $j=1,2,...,9$. The fifth step is to compute

$$\hat{c}_{jk} = \hat{\sigma}^{jk} \tilde{\boldsymbol{v}}_j' \tilde{\boldsymbol{v}}_k = \hat{\sigma}^{jk} \sum_{t=1}^{NT} \tilde{v}_{jt} \tilde{v}_{kt} \quad \text{for} \quad j,k = 1,2,\cdots,9 \,. \tag{30}$$

The sixth step is to form $\hat{V}ar(\hat{\boldsymbol{\gamma}})$ from $\hat{c}_{jk}$'s.

$$\hat{V}ar(\hat{\boldsymbol{\gamma}}) = \begin{bmatrix} \hat{c}_{11} & \hat{c}_{12} & \cdots & \hat{c}_{19} \\ \hat{c}_{21} & \hat{c}_{22} & \cdots & \hat{c}_{29} \\ \vdots & \vdots & & \vdots \\ \hat{c}_{91} & \hat{c}_{92} & \cdots & \hat{c}_{99} \end{bmatrix}^{-1} \tag{31}$$

The seventh step is to compute the value of the F test statistic.

$$F = \frac{1}{J}\hat{\boldsymbol{\gamma}}'[\hat{V}ar(\hat{\boldsymbol{\gamma}})]^{-1}\hat{\boldsymbol{\gamma}} \tag{32}$$

This F test statistic is asymptotically distributed as $F(J, n_2)$, where $n_2=mNT-K$, mNT is the number of observations on m endogenous variables, and K is the number of regression coefficients including γ_j's in the whole system of multiple linear equations.

8. Test Results of Sample Selection Bias

Our data set is an unbalanced panel data set with 79 countries over 28 years of time periods form 1976 to 2003. If it were a balanced panel data set, then the number of observations would be 79×28=2,212. For our unbalanced panel data set, we can say that each country has observations somewhere between 10 years and 24 years for estimations. The deduction in the maximum time length from 28 years to 24 years is due to the use of lags for our panel VAR analysis.

With regard to testing the sample selectivity bias in our panel VAR analysis, $\boldsymbol{\gamma}_P=\mathbf{0}$ has 6 restrictions, $\boldsymbol{\gamma}_T=\mathbf{0}$ has 3 restrictions, and $\boldsymbol{\gamma}=[\boldsymbol{\gamma}'_P, \boldsymbol{\gamma}'_T]'$ has 9 restrictions. Counting regression coefficients and so on, n_2=8,787 is used with K=822. The following table reports the results of the F tests.

Table 1: Tests for Sample Selectivity

Null hypothesis (H_0)	F-value	5% critical value	p-value
$\boldsymbol{\gamma}_P=\mathbf{0}$	10.68	2.10	0.00000
$\boldsymbol{\gamma}_T=\mathbf{0}$	0.49	2.61	0.69
$\boldsymbol{\gamma}=\mathbf{0}$	7.21	1.88	0.00000

The test results suggest that the sample selection bias appears quite serious in our panel VAR analysis for evaluating the effectiveness of the IMF lending programs.

9. Conclusions

If we make the straight forward application of textbook knowledge to the F test in the framework of multiple linear regression equations, we probably face to the problems of dealing with huge matrices. Even though it depends on the numbers of regression coefficients and so on, the problems become serious enough to think about another way of performing the test. For example, if the number of regression coefficients is 822, then we need to write computer codes to specify the elements in a symmetric matrix in 822×822, and the number of the elements to be computed is 822×(822+1)/2=338,253. It is just a small part of the whole task of executing the F test. Nobody wants to write computer codes for such huge and tedious work. Also, we might face to the capacity limitation of statistical software or computer due to the use of huge matrices.

This paper shows a practical way of executing the F test in the framework of multiple linear regression equations by using the technique of residual regressions. In this way we don't face to the problems of dealing with huge metrics. Our example shows that the F test can be carried out using a symmetric matrix in 10×10 rather than 822×822.

Even though we focus on the F test, the idea illustrated in this paper can be applied to other tests as well without any difficulties.

(Received : November 26, 2009, Accepted : January 13, 2010)

References

Davidson, Russell, and James G. MacKinnon (2004), *Econometric Theory and Methods*, New York: Oxford University Press.

Greene, William H. (2000), *Econometric Analysis* (Forth Edition), New Jersey: Prentice Hall.

Goldstein, Morris, and Peter Montiel (1986), "Evaluating Fund Stabilization Programs with Multicountry Data: Some Methodological Pitfalls," *IMF Staff Papers* 33, 304-344.

Heckman, J. J. (1976), "The Common Structure of Statistical Models of Truncation, Sample Selection, and Limited Dependent Variables and a Simple Estimator for Such Models," *Annals of Economic and Social Measurement* 5, 475-492.

Iwata, Shigeru, and Hiroshi Murao (2007), "Are IMF-Supported Stabilization Program Effective? A Panel VAR Approach," written for *2007 Far Eastern Meeting of the Econometric Society* in Taipei, Taiwan.

Rivers, D., and Q. H. Vuong (1988), "Limited Information Estimation and Exogeneity Tests for Simultaneous Probit Models," *Journal of Econometrics* 39, 347-66.

Vella, F. (1992), "Simple Tests for Sample Selection Bias in Censored and Discrete Choice Models," *Journal of Applied Econometrics* 7, 413-421.

Wooldridge, J.M. (1998), "Selection Corrections with a Censored Selection Variables," mimeo, Michigan State University Department of Economics.

Wooldridge, J.M. (2002), *Econometric Analysis of Cross Section and Panel Data*, Cambridge MA: MIT Press.