This is to certify that the

dissertation entitled

SOME APPLICATIONS OF THE LAGRANGE MULTIPLIER TEST IN ECONOMETRICS

presented by

Tsai-Fen Lin

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Economics

Major professor

Date August 6, 1982

SOME APPLICATIONS OF THE LAGRANGE MULTIPLIER TEST IN ECONOMETRICS

BY

Tsai-Fen Lin

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Economics

1982

ABSTRACT

SOME APPLICATIONS OF THE LAGRANGE MULTIPLIER TEST IN ECONOMETRICS

By

Tsai-Fen Lin

There are three kinds of tests for model specification - the Wald test, the likelihood ratio test and the Lagrange multiplier test. They have the same asymptotic power. Therefore, the choice among them depends on computational convenience. Since the Lagrange multiplier test is based on the restricted estimates, we choose the Lagrange multiplier test when estimation is easier in the restricted model than in the unrestricted model.

Since the Lagrange multiplier test is not well known, and the derivation for the test statistic is complicated, in this thesis, I develop the Lagrange multiplier test statistic for some commonly used econometric models so that they can be used readily by applied economists. These models include distributed lag models, qualitative and limited dependent variable models, and stochastic production and cost frontiers. In the distributed lag models, the Lagrange multiplier test statistic is shown to be asymptotically equivalent to the F statistic in testing the coefficients of the lagged explanatory variables when they are added to the restricted model. In Heckman's sample selection bias model, the Lagrange multiplier test statistic is asymptotically equal to the square of the t test statistic in testing the coefficient of the correction term for the sample selection bias when this correction term is added to the restricted model. In Poirier's partial observability model, the Lagrange multiplier test statistic is equivalent to the explained sum of squares in a regression of residuals on a set of regressors. In the stochastic production and cost frontiers models, the

Lagrange multiplier test fails in some cases, and alternative tests are suggested.

In summary, the Lagrange multiplier test, except in a few cases, can be used to test the adequency of the simple models. Since the simple model usually involves a simple estimation method or less computational cost than the more complicated alternative, the Lagrange multiplier test can be useful.

ACKNOWLEDGMENTS

I am deeply indebted to my advisor, Professor Peter Schmidt, for sparking my interest in these topics and for all his guidance and help. I also thank other members of my dissertation committee, William Quinn, John Goddeeris, and Paul Menchik.

TABLE OF CONTENTS

CHAPTER I

INTRODUCTION

1.1 Introduction

In statistics and econometrics, there are three basic principles for the construction of test statistics for model specification. They are the Wald test (Wald(1943)), the likelihood ratio (LR) test and the Lagrange Multiplier (LM) test. Suppose there are two possible model specifications, one of which is a special case of the other one under some restrictions. Let's call the special case the restricted model, and the generalized case the unrestricted model. The Wald test is based on the estimates from the unrestricted model, while the LM test is based on the estimates from the restricted model, and the LR test is based on both sets of estimates. These three principles yield tests which are equivalent in large samples when the restrictions are true (see Silvey(1959)). Their small sample properties are unknown, except in special cases. Therefore the choice among them will often be based on computational convenience. The LM test is very useful in cases in which the restricted model is easier to estimate than the unrestricted model. This will often be the case when one is testing the adequacy of a particular model. Then the null hypothesis is that a relatively simple model is adequate, while the alternative is that a more complicated model is necessary. The LM test permits a test of this hypothesis without having to estimate the more complicated model.

Although the LM test was suggested by Aitchison and Silvey in 1958, it did not receive much attention from econometricians until recent years. Therefore, not many economists are aware of the LM test and its computational advantages in many cases. It is the responsibility of the econometricians to introduce the LM test to applied economists by developing LM test statistics for common models in econometrics. The LM test has been applied successfully in testing for a liquidity trap, autocorrelation, the error components model, seemingly unrelated equation systems and various non-nested hypotheses. (See Breusch and Pagan(1980) for a survey and references.) In this thesis, I report the successful application of the LM test to distributed lag models in chapter 2, to some qualitative and limited dependent variable models in chapter 3, and to stochastic production/cost frontiers in chapter 4.

1.2 The LM Test

Let $\vartheta = (\vartheta_1, \cdots, \vartheta_s)'$ be a set of parameters, $L(\vartheta)$ be the log-likelihood fuction, $h(\vartheta) = [h_1(\vartheta), \cdots, h_r(\vartheta)]' = 0$ be a set of r restrictions, $\lambda = [\lambda_1, \cdots, \lambda_r]'$ be a set of Lagrange Multipliers, $I(\vartheta)$ be the information matrix and n be sample size. Define the Lagrangian function for the maximization of the likelihood subject to the restrictions as

$$L_R(\vartheta,\lambda) = L(\vartheta) + \lambda' h(\vartheta)$$

A constrained maximum of $L(\vartheta)$ is obtained at a stationary point of $L_R(\vartheta)$. By differentiating $L_R(\vartheta,\lambda)$ with respect to ϑ and λ , we have the first order conditions:

$$\begin{aligned} D(\vartheta) + H_\vartheta \lambda &= 0 \\ h(\vartheta) &= 0 \end{aligned} \tag{1.1}$$

where $D(\vartheta)$ is the $s \times 1$ vector, $\left\{\frac{\partial L(\vartheta)}{\partial \vartheta_i}\right\}$, and H_ϑ is the $s \times r$ matrix , $\left\{\frac{\partial h_j(\vartheta)}{\partial \vartheta_i}\right\}$. By solving eq.(1.1), we obtain the restricted MLE $\tilde{\vartheta}$ and $\tilde{\lambda}$. When the restrictions are in fact true ($h(\vartheta) = 0$), the restricted estimates $\tilde{\vartheta}$ will tend to be near the unrestricted estimates, and $D(\tilde{\vartheta})$ and $\tilde{\lambda}$ will tend to be near zero. It seems reasonable to decide that $h(\vartheta) = 0$ is true if $\tilde{\lambda}$ is in some sense near enough zero. Aitchison and Silvey (1958) proved that under the null hypothesis that $h(\vartheta) = 0$, $\sqrt{n}\tilde{\lambda}$ is asymptotically distributed as normal with mean zero and covariance matrix $\left[H_{\tilde{\vartheta}}' n [I(\tilde{\vartheta})]^{-1} H_{\tilde{\vartheta}}\right]^{-1}$ where $H_{\tilde{\vartheta}}$ and $I(\tilde{\vartheta})$ are H_ϑ and $I(\vartheta)$ evaluated at $\tilde{\vartheta}$ respectively. They suggested a test statistic which is based on the estimated Lagrange Multipliers ($\tilde{\lambda}$) and called this the Lagrange Multiplier test statistic:

$$LM\ test\ statistic = \tilde{\lambda}' H_{\tilde{\vartheta}}' [I(\tilde{\vartheta})]^{-1} H_{\tilde{\vartheta}} \tilde{\lambda} \tag{1.2}$$

This statistic asymptotically follows a chi-square distribution with r degrees of freedom when $h(\vartheta) = 0$ is true. The region of acceptance of the null hypothesis

$h(\vartheta) = 0$ is $\tilde{\lambda}' H_{\tilde{\vartheta}}' [I(\tilde{\vartheta})]^{-1} H_{\tilde{\vartheta}} \tilde{\lambda} \le K$, where K is determined by $Prob(\chi_r^2 \le K) = 1 - significance\ level.$.

Note that from eq.(1.1), $H_{\tilde{\vartheta}} \tilde{\lambda} = -D(\tilde{\vartheta})$, so eq.(1.2) can be rewritten as

$$LM\ test\ statistic = [D(\tilde{\vartheta})]'[I(\tilde{\vartheta})]^{-1}[D(\tilde{\vartheta})] \tag{1.3}$$

The right hand side of eq.(1.3) is just Rao's score statistic (Rao(1947)). Hence the LM test statistic is the same as Rao's score statistic. Since eq.(1.3) is easier to use, in the following chapters eq.(1.3) instead of eq.(1.2) will be used.

When ϑ is partitioned into 2 subsets, ϑ_1 and ϑ_2, and the restriction under test is that one of the subsets of parameters equals particular values, i.e. $H_0: \vartheta_1 = \vartheta_{10}$, then we can establish a simpler form of the LM test statistic. From eq.(1.1), $D(\tilde{\vartheta}) = -H_{\tilde{\vartheta}} \tilde{\lambda}$, therefore,

$$\begin{bmatrix} \frac{\partial L}{\partial \tilde{\vartheta}_1} \\ \frac{\partial L}{\partial \tilde{\vartheta}_2} \end{bmatrix} = - \begin{bmatrix} \frac{\partial h_1(\vartheta)}{\partial \tilde{\vartheta}_1} & \cdots & \frac{\partial h_r(\vartheta)}{\partial \tilde{\vartheta}_1} \\ \frac{\partial h_1(\vartheta)}{\partial \tilde{\vartheta}_2} & \cdots & \frac{\partial h_r(\vartheta)}{\partial \tilde{\vartheta}_2} \end{bmatrix} \begin{bmatrix} \tilde{\lambda}_1 \\ \cdot \\ \cdot \\ \tilde{\lambda}_r \end{bmatrix}$$

$$= - \begin{bmatrix} \sum_{j=1}^{r} \frac{\partial h_j(\vartheta)}{\partial \tilde{\vartheta}_1} \tilde{\lambda}_j \\ 0 \end{bmatrix} \tag{1.4}$$

where $\frac{\partial L}{\partial \tilde{\vartheta}_i}$ and $\frac{\partial h_j(\vartheta)}{\partial \tilde{\vartheta}_i}$ are $\frac{\partial L}{\partial \vartheta_i}$ and $\frac{\partial h_j(\vartheta)}{\partial \vartheta_i}$ evaluated at $\tilde{\vartheta}$ respectively. $\frac{\partial L}{\partial \tilde{\vartheta}_2} = 0$ because $h_j(\vartheta)$ is not a function of ϑ_2. Partitioning $I(\tilde{\vartheta})$ conformably, eq.(1.3) becomes

$$LM\ test\ statistic = \begin{bmatrix} \frac{\partial L}{\partial \tilde{\vartheta}_1} \\ 0 \end{bmatrix}' \begin{bmatrix} \tilde{I}_{11} & \tilde{I}_{12} \\ \tilde{I}_{21} & \tilde{I}_{22} \end{bmatrix}^{-1} \begin{bmatrix} \frac{\partial L}{\partial \tilde{\vartheta}_1} \\ 0 \end{bmatrix}$$

$$= \left(\frac{\partial L}{\partial \tilde{\vartheta}_1} \right)' \left[\tilde{I}_{11} - \tilde{I}_{12} \tilde{I}_{22}^{-1} \tilde{I}_{21} \right]^{-1} \left(\frac{\partial L}{\partial \tilde{\vartheta}_1} \right) \tag{1.5}$$

If $I(\tilde{\vartheta})$ is block diagonal, the LM test statistic can be further simplified as

$$LM\ test\ statistic = \left(\frac{\partial L}{\partial \tilde{\vartheta}_1} \right)' \tilde{I}_{11}^{-1} \left(\frac{\partial L}{\partial \tilde{\vartheta}_1} \right) \tag{1.6}$$

When $I(\vartheta)$ is difficult to calculate, we can use the negative of the Hessian (matrix of second derivatives) or its limiting form to construct the LM statistic because in many cases, *plim* $\left\{\left[I(\tilde{\vartheta})\right]^{-1}\left[-\frac{\partial^2 L}{\partial\tilde{\vartheta}\partial\tilde{\vartheta}'}\right]\right\} = I_s$, where $\frac{\partial^2 L}{\partial\tilde{\vartheta}\partial\tilde{\vartheta}'}$ is $\frac{\partial^2 L}{\partial\vartheta\partial\vartheta'}$ evaluated at $\tilde{\vartheta}$. Also note that whenever the usual regularity conditions hold, $I(\vartheta)$ can be obtained from the first partial derivatives of the log-likelihood func-

-

σ

tion. That is, $I(\vartheta) = E\left[-\frac{\partial^2 L}{\partial\vartheta\partial\vartheta'}\right] = E\left[\left(\frac{\partial L}{\partial\vartheta}\right)\left(\frac{\partial L}{\partial\vartheta}\right)'\right]$. Besides these, there is an indirect approach using the scoring algorithm (Newton-Raphson algorithm) to compute the LM statistic indirectly (Breusch and Pagan (1980)).

If $I(\tilde{\vartheta})$ is not of full rank, say, *rank* $[I(\tilde{\vartheta})] = s - t < s$, then $I(\tilde{\vartheta})$ is singular and is therefore not invertible. Silvey (1959) assumes there exists a $s \times t$ submatrix H_1 of H_ϑ such that $\frac{1}{n}I(\tilde{\vartheta}) + H_1H_1'$ is positive definite. Then, he proposed a modified LM test statistic

$$LM^* = \frac{1}{n}\left[D(\tilde{\vartheta})\right]'\left[\frac{1}{n}I(\tilde{\vartheta}) + H_1H_1'\right]^{-1}\left[D(\tilde{\vartheta})\right] \tag{1.7}$$

which asymptotically follows a chi-square distribution with (s - t) degrees of freedom. This case arises in one of our analyses of chapter II.

CHAPTER II

DISTRIBUTED LAG MODELS

2.1 Introduction

A distributed lag model describes how the lagged independent variable affects the dependent variable over time. The length of the lag may sometimes be known a priori, but usually it is unknown and in many cases it is assumed to be infinite. Thus we consider a distributed lag model of the general form

$$y_t = \sum_{i=0}^{\infty} \beta_i x_{t-i} + \varepsilon_t$$

where y_t is the dependent variable, x_{t-i} is the lagged independent variable, β_i is the distributed lag weight, ε_t is a disturbance term. Infinite lag distributions involve an infinite number of unknown parameters, and thus it is impossible to estimate all these parameters. To make estimation possible, it is necessary to make some reasonable assumption about the pattern of the distributed lag weights. The earliest distributed lag model is the geometric lag model proposed by Koyck (1954). He assumes that the lag weights decline geometrically, i.e.

$$\beta_i = \beta\lambda^i, \quad \text{for } i = 0,1,2,\ldots$$

where $0 \le \lambda < 1$. Since the lag weights of the geometric lag model decline monotonically, and this may not always be reasonable, various alternative models have been proposed. For example, the Pascal lag model proposed by Solow (1960) permits a hump in the lag weight distribution curve. In 1966, Jorgenson proposed a more general rational lag model

$$y_t = \frac{A(L)}{B(L)} x_t + u_t$$

where $A(L)$ and $B(L)$ are polynomials in the lag operator of order μ and ν, respectively. He also proved that any arbitary lag model can be approximated to any desired degree of accuracy by a rational lag model with sufficiently high values of μ and ν. If we take $A(L) = \beta(1 - \lambda)$ and $B(L) = 1 - \lambda L$, the rational lag model is Koyck's geometric lag model. If we take $A(L) = \beta(1 - \lambda)^r$ and $B(L) = (1 - \lambda L)^r$, the result is the Pascal lag model.

In this chapter, two distributed lag models are discussed. The geometric lag model is discussed in section 2.2, and rational lag model is discussed in section 2.3.

2.2 The Geometric Lag Model

Following Klein (1958), the geometric lag model can be expressed as

$$y_t = \beta \sum_{i=0}^{\infty} \lambda^i x_{t-i} + \varepsilon_t$$
$$= \beta w_t + \eta_0 \lambda^t + \varepsilon_t \quad , \quad t = 1, \cdots, T, \tag{2.2.1}$$

where $0 \leq \lambda < 1$, $w_t = \sum_{i=0}^{t-1} \lambda^i x_{t-i}$, $\eta_0 = \beta \sum_{i=0}^{\infty} \lambda^i x_{-i}$, ε_t *iid* $N(0, \sigma^2)$. If $\lambda = 0$, or if we know the value of λ, we could estimate eq.(2.2.1) by OLS. Usually,we don't know the value of λ and we use a search procedure to estimate eq.(2.2.1). Since the search procedure is not simple, it may be useful to test whether $\lambda = 0$ before we start the search procedure. The restriction $\lambda = 0$ is easy to impose and the restricted model can be estimated by OLS of y_t on x_t only. Therefore, the LM test is very suitable in this case.

It is well known that the parameter η_0 can not be estimated consistently, and that indeed the information matrix is singular asymptotically when η_0 is included in the list of parameters to be estimated. See Appendix A for details. However, Schmidt and Guilkey (1976) showed that it makes no difference asymptotically whether one drops or estimates the truncation remainder term in the maximum likelihood estimation of distributed lag models. Maximum likelihood estimation of eq.(2.2.1) amounts to minimizing the sum of squares

$$\sum_{t=1}^{T} (y_t - \beta w_t - \eta_0 \lambda^t)^2$$

with respect to λ, β, and η_0. Since $\eta_0 \lambda^t$ disappers asymptotically, this is equivalent to minimizing $\sum_{t=1}^{T} (y_t - \beta w_t)^2$; that is, to setting $\eta_0 = 0$, and applying OLS to the model

$$y_t = \beta w_t + \varepsilon_t \quad , \quad t = 1, \cdots, T, \tag{2.2.2}$$

where $w_t = \sum_{i=0}^{t-1} \lambda^i x_{t-i}$, $0 \le \lambda < 1$, ε_t *iid* $N\,(0, \sigma^2)$. Also, the estimated variances of λ and β resulting from estimation of eq.(2.2.2) are asymptotically the same as ones from eq.(2.2.1). This is so because after deleting the row and column corresponding to η_0, the resulting submatrix of the inverse of the information matrix corresponding to eq.(2.2.1) is asymptotically the same as the inverse of the information matrix corresponding to eq.(2.2.2). Therefore, we can construct our LM test statistic based on eq.(2.2.2) instead of eq.(2.2.1).

The log-likelihood function for eq.(2.2.2) is

$$L = constant - \frac{T}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{t=1}^{T}(y_t - \beta w_t)^2$$

The first partial derivatives are

$$\frac{\partial L}{\partial \lambda} = \frac{1}{\sigma^2}\sum_{t=1}^{T}\beta R_t \varepsilon_t \quad where\ R_t = \frac{dw_t}{d\lambda} = \sum_{i=1}^{t-1} i\lambda^{i-1}x_{t-i}$$

$$\frac{\partial L}{\partial \beta} = \frac{1}{\sigma^2}\sum_{t=1}^{T} w_t \varepsilon_t$$

$$\frac{\partial L}{\partial \sigma^2} = -\frac{T}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{t=1}^{T}\varepsilon_t^2$$

The elements of the information matrix are

$$I_{\lambda\lambda} = -E\left[\frac{\partial^2 L}{\partial \lambda^2}\right] = \frac{1}{\sigma^2}\sum_{t=1}^{T}(\beta R_t)^2$$

$$I_{\beta\beta} = -E\left[\frac{\partial^2 L}{\partial \beta^2}\right] = \frac{1}{\sigma^2}\sum_{t=1}^{T} w_t^2$$

$$I_{\sigma^2\sigma^2} = -E\left[\frac{\partial^2 L}{\partial (\sigma^2)^2}\right] = \frac{T}{2\sigma^4}$$

$$I_{\lambda\beta} = -E\left[\frac{\partial^2 L}{\partial \lambda \partial \beta}\right] = \frac{1}{\sigma^2}\sum_{t=1}^{T}\beta R_t w_t$$

$$I_{\lambda\sigma^2} = -E\left[\frac{\partial^2 L}{\partial \lambda \partial \sigma^2}\right] = 0$$

$$I_{\beta\sigma^2} = -E\left[\frac{\partial^2 L}{\partial \beta \partial \sigma^2}\right] = 0$$

The restricted model is

$$y_t = x_t\beta + \varepsilon_t \quad , \quad t=1,\cdots,T. \tag{2.2.3}$$

Let $\tilde{\vartheta} = (\tilde{\lambda},\tilde{\beta},\tilde{\sigma}^2)$ where $\tilde{\lambda} = 0$, $\tilde{\beta} =$ OLS estimate, $\tilde{\varepsilon}_t = y_t - \tilde{\beta}x_t$ and $\tilde{\sigma}^2 = \dfrac{\sum_{t=1}^{T}\tilde{\varepsilon}_t^2}{T}$.

Thererfore, $\frac{\partial L}{\partial \lambda}$, $I_{\lambda\lambda}$, $I_{\lambda\beta}$, $I_{\beta\beta}$ evaluated at $\tilde{\vartheta}$ are

$$\frac{\partial L}{\partial \tilde{\lambda}} = \frac{\tilde{\beta}}{\tilde{\sigma}^2}\sum_{t=1}^{T} x_{t-1}\tilde{\varepsilon}_t = \frac{\tilde{\beta}}{\tilde{\sigma}^2}X_{t-1}'\,\mathbf{e}_t$$

$$\tilde{I}_{\lambda\lambda} = \frac{\tilde{\beta}^2}{\tilde{\sigma}^2}\sum_{t=1}^{T} x_{t-1}^2 = \frac{\tilde{\beta}^2}{\tilde{\sigma}^2}X_{t-1}'X_{t-1}$$

$$\tilde{I}_{\lambda\beta} = \frac{\tilde{\beta}}{\tilde{\sigma}^2}\sum_{t=1}^{T} x_t x_{t-1} = \frac{\tilde{\beta}}{\tilde{\sigma}^2}X_t'X_{t-1}$$

$$\tilde{I}_{\beta\beta} = \frac{1}{\tilde{\sigma}^2}\sum_{t=1}^{T} x_t^2 = \frac{1}{\tilde{\sigma}^2}X_t'X_t$$

where

$$X_{t-1} = (x_0, x_1, \cdots, x_{T-1})'$$
$$X_t = (x_1, \cdots, x_T)'$$
$$\mathbf{e}_t = (\tilde{\varepsilon}_1, \cdots, \tilde{\varepsilon}_T)'$$

Let $\vartheta_1 = \lambda$, $\vartheta_2 = (\beta,\sigma^2)$. Then we can use eq.(1.5),

$$LM\ test\ statistic = \left[\frac{\partial L}{\partial \tilde{\lambda}}\right]'\left\{\tilde{I}_{\lambda\lambda} - \begin{bmatrix}\tilde{I}_{\lambda\beta}\\0\end{bmatrix}'\begin{bmatrix}\tilde{I}_{\beta\beta}^{-1} & 0\\ 0 & \tilde{I}_{\sigma^2\sigma^2}^{-1}\end{bmatrix}\begin{bmatrix}\tilde{I}_{\lambda\beta}\\0\end{bmatrix}\right\}^{-1}\left[\frac{\partial L}{\partial \tilde{\lambda}}\right]$$

$$= \left[\frac{\partial L}{\partial \tilde{\lambda}}\right]'\left\{\tilde{I}_{\lambda\lambda} - \tilde{I}_{\lambda\beta}'\tilde{I}_{\beta\beta}^{-1}\tilde{I}_{\lambda\beta}\right\}^{-1}\left[\frac{\partial L}{\partial \tilde{\lambda}}\right]$$

$$= \frac{\mathbf{e}_t'X_{t-1}(X_{t-1}'M_1X_{t-1})^{-1}X_{t-1}'\,\mathbf{e}_t}{\tilde{\sigma}^2}$$

$$= \frac{\left[(X_{t-1}'M_1X_{t-1})^{-1}(X_{t-1}'M_1Y_t)\right]^2}{\tilde{\sigma}^2(X_{t-1}'M_1X_{t-1})^{-1}} \quad (Note\ 1) \tag{2.2.4}$$

where $M_1 = I - X_t(X_t'X_t)^{-1}X_t'$ and $Y_t = [y_1, \cdots, y_T]'$. The last equality holds because

$$M_1Y_t = Y_t - X_t(X_t'X_t)^{-1}X_t'Y_t = Y_t - X_t\tilde{\beta} = \mathbf{e}_t.$$

Note that this LM test statistic can be expressed as the square of the t statistic for the coefficient of X_{t-1} in regression of Y_t on (X_t, X_{t-1}). This point can be clarified from the following discussion. Consider the regression

$$Y_t = X_t\beta + X_{t-1}c + u_t \quad , \quad u_t \ iid \ N(0,\sigma^2) \qquad (2.2.5)$$

OLS estimate for c is $\hat{c} = (X_{t-1}'M_1X_{t-1})^{-1}(X_{t-1}'M_1Y_t)$ and OLS estimate for σ^2 is $\hat{\sigma}^2 = \frac{\sum_{t=1}^{T}\hat{u}_t^2}{T}$ where $\hat{u}_t$ is the OLS residual. The OLS estimate for $\sigma_{\hat{c}}^2$ is equal to $\hat{\sigma}^2(X_{t-1}'M_1X_{t-1})^{-1}$. The t test statistic for $c = 0$ is

$$t = \frac{\hat{c} - 0}{\hat{\sigma}_{\hat{c}}^2}$$

$$= \frac{(X_{t-1}'M_1X_{t-1})^{-1}(X_{t-1}'M_1Y_t)}{[\hat{\sigma}^2(X_{t-1}'M_1X_{t-1})^{-1}]^{1/2}}$$

Therefore, $t^2 = \frac{[(X_{t-1}'M_1X_{t-1})^{-1}(X_{t-1}'M_1Y_t)]^2}{\hat{\sigma}^2(X_{t-1}'M_1X_{t-1})^{-1}}$ which differs from eq.(2.2.4) only in one term, namely the estimate for σ^2. The test of $c = 0$ in eq.(2.2.5) is asymptotically equivalent to the test of $\lambda = 0$ in eq.(2.2.2), since when $c = 0$ is true , $\hat{\sigma}^2$ is near $\tilde{\sigma}^2$.

This is an interesting result. We can test the existence of a lag ($\lambda = 0$) in the geometric lag model by testing the significance of the single lagged term, X_{t-1}, in the OLS regression of y_t on (X_t, X_{t-1}). This provides an asymptotically optimal test, despite the fact that the geometric lag is a lag of infinite order.

2.3 The Rational Lag Model

The rational lag model is a rather general distributed lag model. It can be expressed as follows:

$$y_t = \frac{A(L)}{B(L)} x_t + u_t, \quad , \quad t = 1, \cdots, T, \tag{2.3.1}$$

where L is the lag operator defined as

$$L^k x_t = x_{t-k} \quad , \quad k = 0,1, \cdots \quad , \quad L^0 = I \quad , \quad I x_t = x_t$$

and $A(L) = \sum_{i=0}^{\mu} a_i L^i$, $B(L) = \sum_{j=0}^{\nu} b_j L^j$, $b_0 = 1$, $\mu < \nu$. The independent variable x_t is assumed to be nonstochastic, or if stochastic, uncorrelated with the random term u_t. We also assume u_t is independently, identically distributed as $N(0,\sigma^2)$. Dhrymes, Klein and Steiglitz (1970) suggested that this model can be estimated by maximum likelihood methods through a search procedure (search a_i, given b_j), or through an iterative procedure for all of the parameter estimates simultaneously. Then, using the estimates for a_i and b_j, one can estimate σ^2 easily from the first order conditions.

Since the estimation of a_i and b_j is not easy, we have two alternative model specifications which can be estimated by OLS. The test of $B(L) = 1$ is given in section 2.3.1, and the test of $A(L) = a_0$ and $B(L) = 1$ is given in section 2.3.2.

2.3.1 Test of $B(L) = 1$

The restriction $B(L) = 1$ can be written as

$$b_j = 0 \quad , \quad j = 1, \cdots, \nu. \tag{2.3.2}$$

Under the restrictions, eq.(2.3.1) becomes

$$y_t = A(L)x_t + u_t = \sum_{i=0}^{\mu} a_i L^i x_t + u_t \quad , \quad t = 1, \cdots, T. \tag{2.3.3}$$

Despite the fact that there might exist high multicollinearity among x's, we still can use the OLS method to estimate this restricted model, and indeed OLS provides MLE'S subject to the restriction. Let $\vartheta_1 = (b_1, \cdots, b_\nu)'$, $\vartheta_2 = (a_0, a_1, \cdots, a_\mu, \sigma^2)'$ and $\widetilde{\vartheta} = (\widetilde{b}_1, \cdots, \widetilde{b}_\nu, \widetilde{a}_0, \widetilde{a}_1, \cdots, \widetilde{a}_\mu, \widetilde{\sigma}^2)'$ where $\widetilde{b}_1 = \cdots = \widetilde{b}_\nu = 0$, $\widetilde{a}_i$ = *OLS estimates* for a_i, $i = 0, \cdots, \mu$, and $\widetilde{\sigma}^2 = \dfrac{\sum_{t=1}^{T} \widetilde{u}_t^2}{T}$ where $\widetilde{u}_t = y_t - \widetilde{A}(L)x_t$ and $\widetilde{A}(L) = \sum_{i=0}^{\mu} \widetilde{a}_i L^i$. We can use eq.(1.5) to construct the LM test statistic.

The log-likelihood function is

$$L = -\frac{T}{2} ln(2\pi) - \frac{T}{2} \ln\sigma^2 - \frac{1}{2\sigma^2} \sum_{t=1}^{T} (y_t - A(L)x_t)^2$$

The first partial derivative with respect to ϑ_2 evaluated at $\widetilde{\vartheta}$ is zero (see eq.(1.4)). The first partial derivative with respect to ϑ_1 evaluated at $\widetilde{\vartheta}$ is

$$\begin{aligned}
\frac{\partial L}{\partial \widetilde{\vartheta}_1} &= \left[\frac{\partial L}{\partial \widetilde{b}_1}, \cdots, \frac{\partial L}{\partial \widetilde{b}_\nu} \right]' \\
&= -\frac{1}{\widetilde{\sigma}^2} \left[\sum_{t=1}^{T} \widetilde{u}_t \widetilde{A}(L) x_{t-1}, \cdots, \sum_{t=1}^{T} \widetilde{u}_t \widetilde{A}(L) x_{t-\nu} \right]' \\
&= -\frac{1}{\widetilde{\sigma}^2} X' \widetilde{U}
\end{aligned}$$

where

$$X = \begin{bmatrix} A(L)x_{1-1} & \cdots & A(L)x_{1-\nu} \\ \vdots & \ddots & \vdots \\ A(L)x_{T-1} & \cdots & A(L)x_{T-\nu} \end{bmatrix}$$

and

$$\widetilde{U} = [\widetilde{u}_1, \cdots \widetilde{u}_T]'$$

The information matrix evaluated at $\tilde{\vartheta}$ is

$$I(\tilde{\vartheta}) = \begin{bmatrix} \tilde{I}_{11} & \tilde{I}_{12} \\ \tilde{I}_{21} & \tilde{I}_{22} \end{bmatrix} = \begin{bmatrix} \tilde{I}_{bb} & \tilde{I}_{ba} & 0 \\ \tilde{I}_{ab} & \tilde{I}_{aa} & 0 \\ 0 & 0 & \tilde{I}_{\sigma^2\sigma^2} \end{bmatrix}$$

where

$$\tilde{I}_{bb} = \begin{bmatrix} \tilde{I}_{b_1b_1} & \cdots & \tilde{I}_{b_1b_\nu} \\ \vdots & & \vdots \\ \tilde{I}_{b_\nu b_1} & \cdots & \tilde{I}_{b_\nu b_\nu} \end{bmatrix}$$

$$= \frac{1}{\tilde{\sigma}^2} \begin{bmatrix} \sum_{t=1}^{T} [\tilde{A}(L)x_{t-1}][\tilde{A}(L)x_{t-1}] & \cdots & \sum_{t=1}^{T} [\tilde{A}(L)x_{t-1}][\tilde{A}(L)x_{t-\nu}] \\ \vdots & \cdots & \vdots \\ \sum_{t=1}^{T} [\tilde{A}(L)x_{t-\nu}][\tilde{A}(L)x_{t-1}] & \cdots & \sum_{t=1}^{T} [\tilde{A}(L)x_{t-\nu}][\tilde{A}(L)x_{t-\nu}] \end{bmatrix}$$

$$= \frac{1}{\tilde{\sigma}^2} X'X$$

$$\tilde{I}_{ba} = \begin{bmatrix} \tilde{I}_{b_1a_0} & \cdots & \tilde{I}_{b_1a_\mu} \\ \vdots & & \vdots \\ \tilde{I}_{b_\nu a_0} & \cdots & \tilde{I}_{b_\nu a_\mu} \end{bmatrix}$$

$$= -\frac{1}{\tilde{\sigma}^2} \begin{bmatrix} \sum_{t=1}^{T} [\tilde{A}(L)x_{t-1}]x_t & \cdots & \sum_{t=1}^{T} [\tilde{A}(L)x_{t-1}]x_{t-\mu} \\ \vdots & \cdots & \vdots \\ \sum_{t=1}^{T} [\tilde{A}(L)x_{t-\nu}]x_t & \cdots & \sum_{t=1}^{T} [\tilde{A}(L)x_{t-\nu}]x_{t-\mu} \end{bmatrix}$$

$$= -\frac{1}{\sigma^2}X'X_{\bullet}$$

$$\tilde{I}_{aa} = \begin{bmatrix} \tilde{I}_{a_0a_0} & \cdots & \tilde{I}_{a_0a_\mu} \\ \vdots & & \vdots \\ \tilde{I}_{a_\mu a_0} & \cdots & \tilde{I}_{a_\mu a_\mu} \end{bmatrix}$$

$$= \frac{1}{\tilde{\sigma}^2}\begin{bmatrix} \sum_{t=1}^{T} x_t x_t & \cdots & \sum_{t=1}^{T} x_t x_{t-\mu} \\ \vdots & & \vdots \\ \sum_{t=1}^{T} x_{t-\mu} x_t & \cdots & \sum_{t=1}^{T} x_{t-\mu} x_{t-\mu} \end{bmatrix}$$

$$= \frac{1}{\tilde{\sigma}^2}X_{\bullet}'X_{\bullet}$$

with

$$X_{\bullet} = \begin{bmatrix} x_1 & x_{1-1} & \cdots & x_{1-\mu} \\ \vdots & \vdots & & \vdots \\ x_T & x_{T-1} & \cdots & x_{T-\mu} \end{bmatrix}$$

and

$$\tilde{I}_{\sigma^2\sigma^2} = \frac{T}{2\tilde{\sigma}^4}.$$

From eq.(1.5),

$$LM\ test\ statistic = \left(\frac{\partial L}{\partial \tilde{\vartheta}_1}\right)'\left\{\tilde{I}_{bb} - \begin{bmatrix}\tilde{I}_{ba} & 0\end{bmatrix}\begin{bmatrix}\tilde{I}_{aa}^{-1} & 0 \\ 0 & \tilde{I}_{\sigma^2\sigma^2}\end{bmatrix}\begin{bmatrix}\tilde{I}_{ab} \\ 0\end{bmatrix}\right\}^{-1}\left(\frac{\partial L}{\partial \tilde{\vartheta}_1}\right)$$

$$= \left(-\frac{\tilde{U}'X}{\tilde{\sigma}^2}\right)\left\{\frac{X'X}{\tilde{\sigma}^2} - \left(-\frac{X'X_\bullet}{\tilde{\sigma}^2}\right)\left(\frac{X_\bullet'X_\bullet}{\tilde{\sigma}^2}\right)^{-1}\left(-\frac{X_\bullet'X}{\tilde{\sigma}^2}\right)\right\}^{-1}\left(-\frac{X'\tilde{U}}{\tilde{\sigma}^2}\right)$$

$$= \frac{\tilde{U}'X\left[X'M_3X\right]^{-1}X'\tilde{U}}{\tilde{\sigma}^2} \tag{2.3.4}$$

where $M_3 = I - X_\bullet(X_\bullet'X_\bullet)^{-1}X_\bullet'$. Note that in this case, the LM test is essentially the F test of significance when X is added to the regression, i.e.

$$Y = X_\bullet a + Xc + \varepsilon \tag{2.3.5}$$

where ε is distributed as $N(0, \sigma^2 I)$. The OLS estimate for c is $\hat{c} = (X'M_3X)^{-1}X'M_3Y$, and $\hat{\sigma}^2 = \frac{\hat{\varepsilon}'\hat{\varepsilon}}{T}$ where $\hat{\varepsilon}$ is OLS residual. The F test statistic for the null hypothesis $c = 0$ is

$$F = \frac{\hat{c}'(X'M_1X)\hat{c}/\nu}{\hat{\varepsilon}'\hat{\varepsilon}/[T-(\mu+\nu+1)]} \tag{2.3.6}$$

The restriction of $B(L) = 1$ in eq.(2.3.1) is equivalent to the restriction of $c = 0$ in eq.(2.3.5). When the restriction is true, $\hat{\sigma}^2$ should be close to $\tilde{\sigma}^2$. Therefore, eq.(2.3.4) can be rewritten as

$$\begin{aligned} LM\ test\ statistic &= \frac{Y'M_3X(X'M_3X)^{-1}X'M_3Y}{\tilde{\sigma}^2} \quad because\ M_3Y = \tilde{U}. \\ &= \frac{\hat{c}'(X'M_3X)\hat{c}}{\tilde{\sigma}^2} \\ &\approx \frac{\hat{c}'(X'M_3X)\hat{c}}{\hat{\sigma}^2} \end{aligned} \tag{2.3.7}$$

From eq.(2.3.6) and eq.(2.3.7), $F \approx \frac{LM/\nu}{T/[T-(\mu+\nu+1)]}$. In a large sample, $\frac{T}{T-(\mu+\nu+1)} \to 1$ and the F test is asymptotically equivalent to the LM test. This gives the justification of doing the F test in this case.

2.3.2 Test of $A(L) = a_0$ and $B(L) = 1$

This section might seem to be a special case of last section. In fact, in this case a singularity problem arises and needs special discussion.

The restrictions $A(L) = a_0$ and $B(L) = 1$ can be expressed as

$$a_1 = a_2 = \cdots = a_\mu = b_1 = b_2 = \cdots = b_\nu = 0 \tag{2.3.8}$$

Under the restrictions, eq.(2.3.1.) becomes

$$y_t = a_0 x_t + u_t \quad , \quad t = 1, \cdots, T. \tag{2.3.9}$$

We can apply the OLS method to eq.(2.3.9) and obtain the restricted estimates

$$\tilde{\vartheta} = [\tilde{a}_1, \cdots, \tilde{a}_\mu, \tilde{b}_1, \cdots, \tilde{b}_\nu, \tilde{a}_0, \tilde{\sigma}^2]'$$

where $\tilde{a}_1 = \cdots = \tilde{a}_\mu = \tilde{b}_1 = \cdots = \tilde{b}_\nu = 0$, $\tilde{a}_0 = OLS\ estimate\ of\ a_0$, and $\tilde{\sigma}^2 = \frac{\sum_{t=1}^{T} \tilde{u}_t^2}{T}$, with $\tilde{u}_t = y_t - \tilde{a}_0 x_t$. The information matrix evaluated at $\tilde{\vartheta}$ is singular because

$$I(\tilde{\vartheta}) = \begin{bmatrix} \tilde{I}_{a_* a_*} & \tilde{I}_{a_* b} & \tilde{I}_{a_* a_0} & 0 \\ \tilde{I}'_{a_* b} & \tilde{I}_{bb} & \tilde{I}_{b a_0} & 0 \\ \tilde{I}'_{a_* a_0} & \tilde{I}'_{b a_0} & \tilde{I}_{a_0 a_0} & 0 \\ 0 & 0 & 0 & \tilde{I}_{\sigma^2 \sigma^2} \end{bmatrix}$$

where

$$\tilde{I}_{a_* a_*} = \begin{bmatrix} \tilde{I}_{a_1 a_1} & \cdots & \tilde{I}_{a_1 a_\mu} \\ \vdots & & \vdots \\ \tilde{I}_{a_\mu a_1} & \cdots & \tilde{I}_{a_\mu a_\mu} \end{bmatrix} = \frac{1}{\tilde{\sigma}^2} X_*' X_*$$

$$\tilde{I}_{a_* b} = \begin{bmatrix} \tilde{I}_{a_1 b_1} & \cdots & \tilde{I}_{a_1 b_\nu} \\ \vdots & & \vdots \\ \tilde{I}_{a_\mu b_1} & \cdots & \tilde{I}_{a_\mu b_\nu} \end{bmatrix} = -\frac{\tilde{a}_0}{\tilde{\sigma}^2} X_*' X_{**}$$

$$\tilde{I}_{bb} = \begin{bmatrix} \tilde{I}_{b_1 b_1} & \cdots & \tilde{I}_{b_1 b_\nu} \\ \vdots & & \vdots \\ \tilde{I}_{b_\nu b_1} & \cdots & \tilde{I}_{b_\nu b_\nu} \end{bmatrix} = \frac{\tilde{a}_0^2}{\tilde{\sigma}^2} X_{**}' X_{**}$$

$$\tilde{I}_{a_\bullet a_0} = \begin{bmatrix} \tilde{I}_{a_1 a_0} \\ \cdot \\ \cdot \\ \tilde{I}_{a_\mu a_0} \end{bmatrix} = \frac{1}{\tilde{\sigma}^2} X_\bullet' X$$

$$\tilde{I}_{b a_0} = \begin{bmatrix} \tilde{I}_{b_1 a_0} \\ \cdot \\ \cdot \\ \tilde{I}_{b_\nu a_0} \end{bmatrix} = - \frac{\tilde{a}_0}{\tilde{\sigma}^2} X_{\bullet\bullet}' X$$

$$\tilde{I}_{a_0 a_0} = \frac{1}{\tilde{\sigma}^2} X' X$$

where

$$X = \begin{bmatrix} x_1 \\ \cdot \\ \cdot \\ \cdot \\ x_T \end{bmatrix}$$

$$X_\bullet = \begin{bmatrix} x_{1-1} & \cdots & x_{1-\mu} \\ \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot \\ x_{T-1} & \cdots & x_{T-\mu} \end{bmatrix}$$

and

$$X_{\bullet\bullet} = \begin{bmatrix} x_{1-1} & \cdots & x_{1-\mu} & \cdots & x_{1-\nu} \\ \cdot & \cdots & \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot & \cdots & \cdot \\ x_{T-1} & \cdots & x_{T-\mu} & \cdots & x_{T-\nu} \end{bmatrix}$$

Note that , in $I(\tilde{\vartheta})$, the $(\mu + k)$th column is just $(-\tilde{a}_0)$ times kth column, where $k = 1, \cdots, \mu$. Hence $I(\tilde{\vartheta})$ is singular with a rank equal to $(\nu + 2)$. In order to utilize eq.(1.5), we have to find some good reason to reduce the size of $I(\tilde{\vartheta})$ so that the resultant information matrix is of full rank. Since

$$\frac{\partial L}{\partial \tilde{a}_i} = \frac{1}{\tilde{\sigma}^2} \sum_{t=1}^{T} \tilde{u}_t x_{t-i} \quad , \quad i = 1, \cdots, \mu$$

and

$$\frac{\partial L}{\partial \tilde{b}_j} = - \frac{\tilde{a}_0}{\tilde{\sigma}^2} \sum_{t=1}^{T} \tilde{u}_t x_{t-j} \quad ; \quad j = 1, \cdots, \nu.$$

we have

$$\frac{\partial L}{\partial \tilde{b}_j} = -\tilde{\alpha}_0 \frac{\partial L}{\partial \tilde{\alpha}_j} \ , \quad j = 1, \cdots, \mu.$$

That is $\frac{\partial L}{\partial \tilde{\alpha}_i}$, $i = 1, \ldots, \mu$ do not contain any information not contained in $\frac{\partial L}{\partial \tilde{b}_j}$, $j = 1, \cdots, \nu$. Thus we can drop $\frac{\partial L}{\partial \tilde{\alpha}_i}$, $i = 1, \cdots, \mu$ from the vector of the first partials and drop the rows and columns corresponding to α_i, $i = 1, \cdots, \mu$ in the information matrix without sacrificing any information.

Define

$$\left[D(\tilde{\vartheta})\right]_{*} = \left[\frac{\partial L}{\partial \tilde{b}_1}, \cdots, \frac{\partial L}{\partial \tilde{b}_\nu}, \frac{\partial L}{\partial \tilde{\alpha}_0}, \frac{\partial L}{\partial \tilde{\sigma}^2}\right]' = \left[-\frac{\tilde{\alpha}_0}{\tilde{\sigma}^2}\tilde{U}'X_{**}, 0, 0\right]'$$

$$I(\tilde{\vartheta})_{*} = \begin{bmatrix} \tilde{I}_{bb} & \tilde{I}_{b\alpha_0} & 0 \\ \tilde{I}_{b\alpha_0} & \tilde{I}_{\alpha_0\alpha_0} & 0 \\ 0 & 0 & \tilde{I}_{\sigma^2\sigma^2} \end{bmatrix}$$

The resulting LM statistic is

$$\begin{aligned} LM_{*} &= \left[D(\tilde{\vartheta})\right]_{*}'\left[I(\tilde{\vartheta})_{*}\right]^{-1}\left[D(\tilde{\vartheta})\right]_{*} \\ &= \left[-\frac{\tilde{\alpha}_0}{\tilde{\sigma}^2}\tilde{U}'X_{**}\right]\left\{\tilde{I}_{bb} - \tilde{I}_{b\alpha_0}\tilde{I}^{-1}_{\alpha_0\alpha_0}\tilde{I}_{\alpha_0 b}\right\}^{-1}\left[-\frac{\tilde{\alpha}_0}{\tilde{\sigma}^2}X_{**}'\tilde{U}\right] \\ &= \frac{\tilde{U}'X_{**}\left(X_{**}'M_4X_{**}\right)^{-1}X_{**}'\tilde{U}}{\tilde{\sigma}^2} \end{aligned} \qquad (2.3.10)$$

where $M_4 = I - X(X'X)^{-1}X'$. By the same reasoning as in section 2.3.1, LM_{*} is asymptotically equivalent to the F statistic in testing the coefficients of the lagged x's in X_{*} when they are added to eq.(2.3.9)

This way of dealing the singularity problem may seem too simple and without theoretical support. But eq.(2.3.10) turns out to be exactly the same as Silvey's modified LM test statistic eq.(1.7). To see this, note that in this case,

$$H_{\vartheta} = \begin{bmatrix} I_\mu & 0 \\ 0 & I_\nu \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

$$H_1 = \begin{bmatrix} I_\mu \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Therefore

$$H_1 H_1' = \begin{bmatrix} I_\mu & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

and

$$\left[\frac{1}{T} I(\tilde{\vartheta}) + H_1 H_1'\right]^{-1} = \begin{bmatrix} \frac{1}{T}\tilde{I}_{\alpha_*\alpha_*} + I_\mu & \frac{1}{T}\tilde{I}_{\alpha_* b} & \frac{1}{T}\tilde{I}_{\alpha_*\alpha_0} & 0 \\ \frac{1}{T}\tilde{I}'_{\alpha_* b} & \frac{1}{T}\tilde{I}_{bb} & \frac{1}{T}\tilde{I}_{b\alpha_0} & 0 \\ \frac{1}{T}\tilde{I}'_{\alpha_*\alpha_0} & \frac{1}{T}\tilde{I}'_{b\alpha_0} & \frac{1}{T}\tilde{I}_{\alpha_0\alpha_0} & 0 \\ 0 & 0 & 0 & \frac{1}{T}\tilde{I}_{\sigma^2\sigma^2} \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} A & F & K & 0 \\ B & G & M & 0 \\ C & J & N & 0 \\ 0 & 0 & 0 & T\tilde{I}_{\sigma^2\sigma^2}^{-1} \end{bmatrix}$$

where

$$\begin{bmatrix} A & F & K \\ B & G & M \\ C & J & N \end{bmatrix} = \begin{bmatrix} \frac{1}{T}\tilde{I}_{\alpha_*\alpha_*} + I_\mu & \frac{1}{T}\tilde{I}_{\alpha_* b} & \frac{1}{T}\tilde{I}_{\alpha_*\alpha_0} \\ \frac{1}{T}\tilde{I}'_{\alpha_* b} & \frac{1}{T}\tilde{I}_{bb} & \frac{1}{T}\tilde{I}_{b\alpha_0} \\ \frac{1}{T}\tilde{I}'_{\alpha_*\alpha_0} & \frac{1}{T}\tilde{I}'_{b\alpha_0} & \frac{1}{T}\tilde{I}_{\alpha_0\alpha_0} \end{bmatrix}^{-1}$$

From eq.(1.7),

$$LM' = \frac{1}{T} \begin{bmatrix} \frac{\partial L}{\partial \tilde{\alpha}_*} \\ \frac{\partial L}{\partial \tilde{b}} \\ 0 \\ 0 \end{bmatrix}' \begin{bmatrix} A & F & K & 0 \\ B & G & M & 0 \\ C & J & N & 0 \\ 0 & 0 & 0 & T\tilde{I}_{\sigma^2\sigma^2}^{-1} \end{bmatrix} \begin{bmatrix} \frac{\partial L}{\partial \tilde{\alpha}_*} \\ \frac{\partial L}{\partial \tilde{b}} \\ 0 \\ 0 \end{bmatrix}$$

$$= \frac{\left(\frac{\partial L}{\partial \tilde{a}_\bullet}\right)' A \left(\frac{\partial L}{\partial \tilde{a}_\bullet}\right) + \left(\frac{\partial L}{\partial \tilde{b}}\right)' B \left(\frac{\partial L}{\partial \tilde{a}_\bullet}\right) + \left(\frac{\partial L}{\partial \tilde{a}_\bullet}\right)' F \left(\frac{\partial L}{\partial \tilde{b}}\right) + \left(\frac{\partial L}{\partial \tilde{b}}\right)' G \left(\frac{\partial L}{\partial \tilde{b}}\right)}{T} \qquad (2.3.11)$$

where

$$\frac{\partial L}{\partial \tilde{a}_\bullet} = \left[\frac{\partial L}{\partial \tilde{a}_1}, \quad \cdots \quad , \frac{\partial L}{\partial \tilde{a}_\mu}\right]' = \frac{1}{\tilde{\sigma}^2} X_\bullet' \tilde{U}$$

$$\frac{\partial L}{\partial \tilde{b}} = \left[\frac{\partial L}{\partial \tilde{b}_1}, \quad \cdots \quad , \frac{\partial L}{\partial \tilde{b}_\nu}\right]' = -\frac{\tilde{a}_0}{\tilde{\sigma}^2} X_{\bullet\bullet}' \tilde{U}$$

and A, B, F, G, can be found by some manipulations involving the inverse of a partitioned matrix (see Appendix B) :

$$A = I_\mu$$

$$B = \frac{1}{\tilde{a}_0} \begin{bmatrix} I_\mu \\ 0 \end{bmatrix}$$

$$F = \frac{1}{\tilde{a}_0} \begin{bmatrix} I_\mu, & 0 \end{bmatrix}$$

$$G = \frac{1}{\tilde{a}_0^2} \left\{ T\tilde{\sigma}^2 \left(X_{\bullet\bullet}' M_4 X_{\bullet\bullet}\right)^{-1} + \begin{bmatrix} I_\mu & 0 \\ 0 & 0 \end{bmatrix} \right\}$$

Therefore eq.(2.3.11) becomes

$$LM' = \frac{\tilde{U}' \left[X_\bullet A X_\bullet' - \tilde{a}_0 X_{\bullet\bullet} B X_\bullet' - \tilde{a}_0 X_\bullet F X_{\bullet\bullet}' + \tilde{a}_0^2 X_{\bullet\bullet} G X_{\bullet\bullet}'\right] \tilde{U}}{T\tilde{\sigma}^4}$$

$$= \frac{\tilde{U}' \left[X_\bullet X_\bullet' - X_\bullet X_\bullet' - X_\bullet X_\bullet' + T\tilde{\sigma}^2 X_{\bullet\bullet} \left(X_{\bullet\bullet}' M_4 X_{\bullet\bullet}\right)^{-1} X_{\bullet\bullet}' + X_\bullet X_\bullet'\right] \tilde{U}}{T\tilde{\sigma}^4}$$

$$= \frac{\tilde{U}' X_{\bullet\bullet} \left(X_{\bullet\bullet}' M_4 X_{\bullet\bullet}\right)^{-1} X_{\bullet\bullet}' \tilde{U}}{\tilde{\sigma}^2}$$

which is exactly the same as eq.(2.3.10).

2.4

rat

lag

sta

va

te

pr

si

si

a

i

2.4 Conclusion

The LM test statistic for the geometric lag model (eq.(2.2.3)) and for the rational lag model (eq.(2.3.4) and eq.(2.3.10)) are similar because the geometric lag model is a special case of the rational lag model. In both models, the LM statistic can be constructed by adding some lagged values of the explanatory variable to the restricted model, and testing their significance using the usual F test (Note 2). This is a favorable result, since in practice many researchers may prefer to estimate under the restrictions (using the OLS method), and to consider more complicated estimation methods only if the LM test provides significant evidence of the existence of a more complicated lag pattern.

Although we have considered a model with only a single explanatory variable, the results do not depend on this assumption (Note 1). They would still hold if there were additional regressors (not subject to the distributed lag).

3.1

de

us

ar

th

th

ti

b

a

CHAPTER III

QUALITATIVE AND LIMITED DEPENDENT VARIABLE MODELS

3.1 Introduction

In many cases, economic studies have to deal with situations in which the dependent variable is dichotomous ; that is, it is observed by its sign only. The usual least squares method will cause many problems (see Judge, Griffiths, Hill and Lee (1980) p.586). The most serious problem is that the predicted value of the dependent variable will not be in the unit interval. One of the solutions to this problem is the "probit model" (see Finney (1971)) which uses the cumulative normal distribution function to transform the dependent variable into a probability. This model takes the form

$$y_t^* = x_t\beta + \varepsilon_t, \qquad \varepsilon_t \;\; i.i.d. \;\; N(0, \sigma^2),$$

and

$$y_t = \begin{cases} 1 & \text{if } y_t^* > 0 \\ 0 & \text{if } y_t^* \leq 0 \end{cases}$$

Then

$$Prob.(y_t = 1) = Prob.(y_t^* > 0) = Prob.(x_t\beta + \varepsilon_t > 0) = Prob.(\varepsilon_t < x_t\beta) = \Phi\left(\frac{x_t\beta}{\sigma}\right)$$

where $\Phi(\cdot)$ is cumulative distribution function of N (0,1). Therefore,

$$Prob.(y_t = 0) = 1 - Prob.(y_t = 1) = 1 - \Phi\left(\frac{x_t\beta}{\sigma}\right).$$

The probit estimate for $\frac{\beta}{\sigma}$ can be obtained by maximizing the following likelihood function:

$$L = \prod_{y_t=1} \Phi\left(\frac{x_t\beta}{\sigma}\right) \prod_{y_t=0} \left[1 - \Phi\left(\frac{x_t\beta}{\sigma}\right)\right]$$

Sin

iden

of t

mo

whi

sid

one

tion

two

Since we cannot identify β, σ separately, we choose the normalization $\sigma=1$ to identify β. The classical example of the probit analysis in economics is the study of the consumer's decision of buying a durable good.

In this chapter, we consider three models which are extensions of the probit model. Section 3.2 considers the Tobit model, and Cragg's extension of it, in which the dependent variable is observable in a limited range. Section 3.3 considers Heckman's sample selection bias model which consists of two equations, one of which is a probit equation representing the rule for sample selection. Section 3.4 considers Poirier's partial observability probit model, which consists of two probit equations with a condition of partial observability.

3.2

3.2

ab

lin

ar

th

w

T

d

d

s

o

f

a

3.2 Test of the Tobit Specification against Cragg's Extension of the Tobit Model

3.2.1 The Tobit Model and Cragg's Extension

Tobin(1958) considered a case in which the dependent variable is observable in a limited range and the analyst is not only interested in the probability of limit and non-limit responses, but also in the value of non-limit responses. Probit analysis is not suitable for this purpose. He proposed the following model, called the Tobit model:

$$y_t^* = x_t\beta + \varepsilon_t \quad , \quad \varepsilon_t \ \ iid \ \ N\,(0,\, \sigma^2)\,,$$
$$y_t = \begin{cases} y_t^* & \text{if } y_t^* > 0 \\ 0 & \text{if } y_t^* \leq 0 \end{cases} \tag{3.2.1}$$
$$t = 1,\ldots,T$$

where y_t^* is unobservable and y_t is observable. y_t has a lower limit which is zero. That is, there is an event which at each observation may or may not occur. If it does occur, associated with it will be a continuous positive random variable. If it does not occur, this variable has a zero value. An example is an individual's decision whether or not to buy a new car, and the amount he spends if he does buy one.

According to eq.(3.2.1), for $y_t > 0$, the probability density function (p.d.f.) for y_t is

$$f\,(y_t) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{1}{2\sigma^2}(y_t - x_t\beta)^2\right\} \tag{3.2.2}$$

and for $y_t = 0$, the probability of observing $y_t = 0$ is

$$\begin{aligned} Prob.\,(y_t = 0) &= Prob.\,(y_t^* \leq 0) \\ &= Prob.\,(x_t\beta + \varepsilon_t \leq 0) \\ &= Prob.\,(\varepsilon_t \leq -x_t\beta) \\ &= \int_{-\infty}^{-x_t\beta} \frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{1}{2\sigma^2}s^2\right\}ds \end{aligned}$$

wh

ob

3

$$= \Phi\left(-\frac{x_t\beta}{\sigma}\right) \tag{3.2.3}$$

where $\Phi(\cdot)$ is the c.d.f. of the standard normal distribution. The probability of observing $y_t = 0$ is represented by the shaded area in Fig.3.2.1.

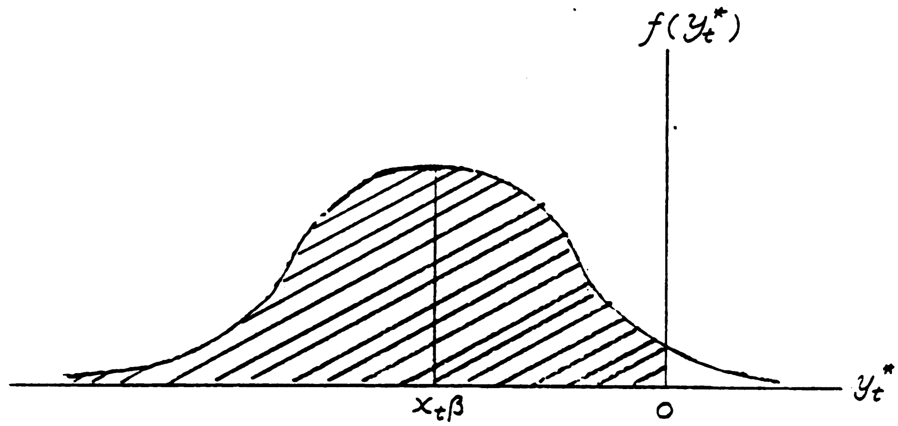

Fig.3.2.1

Note that there is one and only one β to determine both the probability of $y_t = 0$ and the shape of the probability distribution for $y_t > 0$. That is, in the example of purchases of a durable good, the decisions on whether to acquire and on how much to spend if acquisition occurs are basically the same in this model, in the sense that the same variables and parameters occur in eq.(3.2.2) and eq.(3.2.3). Cragg(1971) argues that "In some situations the decison to acquire and the amount of the acquisition may not be so intimately related. In particular, even when the probability of a non-zero value is less than one half, one might not feel that values close to zero are more probable than ones near some larger value, given that a positive value will occur." In the case of buying a new car, this argument is certainly true. The probability of buying a new car for an individual in a particular year is probably less than one half. From Fig.3.2.1, the Tobit model implies that, if a new car is purchased, smaller expenditures(e.g. 5 dollars) are more likely than larger expenditures(e.g. 5000 dollars). This foolish imp

min

$y_t >$

ter

the

Tob

Fir

S

implication is due to the fact that there is only one set of parameters to determine the probability of $y_t = 0$ and the shape of the probability distribution for $y_t > 0$.

Cragg(1971) proposed a more general model which uses two sets of parameters. One set determines the probability of $y_t = 0$, and the other set determines the shape of the probability distribution for $y_t > 0$. Cragg's extension of the Tobit model can be written as a two-stage decision process.

First-stage — decision on whether to acquire

The probability of not buying a durable good is

$$Prob.(y_t=0) = Prob.(x_t\beta_1+u_1<0) = \Phi(-x_t\beta_1) \tag{3.2.4}$$

and the probability of buying a durable good is

$$Prob.(y_t>0) = 1 - Prob.(y_t=0) = 1 - \Phi(-x_t\beta_1) = \Phi(x_t\beta_1) \tag{3.2.5}$$

where σ_1 is normalized as 1 because we can not identify β_1 and σ_1 separately in a probit model.

Second-stage — decision on how much to acquire if acquisition occurs

The probability density function for y_t, given acquisition occurs, is

$$f(y_t \mid y_t > 0) = N(x_t\beta_2, \sigma_2^2) \textit{ truncated at zero}$$

$$= \frac{\frac{1}{\sqrt{2\pi}\sigma_2}\exp\left[-\frac{(y_t-x_t\beta_2)^2}{2\sigma_2^2}\right]}{\int_0^\infty \frac{1}{\sqrt{2\pi}\sigma_2}\exp\left[-\frac{(y_t-x_t\beta_2)^2}{2\sigma_2^2}\right]dy_t}$$

$$= \frac{\frac{1}{\sqrt{2\pi}\sigma_2}\exp\left[-\frac{(y_t-x_t\beta_2)^2}{2\sigma_2^2}\right]}{\Phi\left(\frac{x_t\beta_2}{\sigma_2}\right)} \tag{3.2.6}$$

since

$$\int_0^\infty \frac{1}{\sqrt{2\pi}\sigma_2}\exp\left[-\frac{(y_t-x_t\beta_2)^2}{2\sigma_2^2}\right]dy_t = \Phi\left(\frac{x_t\beta_2}{\sigma_2}\right).$$

The unrestricted estimates for β_1, β_2 and σ_2 can be obtained by maximizing the f

whe

lik

3

w

(

the following likelihood function:

$$L^*(\beta_1,\beta_2,\sigma_2) = \prod_{y_t=0}\left[Prob.\,(y_t=0)\right]\prod_{y_t>0}\left[Prob.\,(y_t>0)\cdot f\,(y_t \mid y_t>0)\right]$$
$$= \prod_{t=1}^{T}\left[Prob.\,(y_t=0)\right]^{1-d_t}\left[Prob.\,(y_t>0)\cdot f\,(y_t \mid y_t>0)\right]^{d_t}$$
$$= \prod_{t=1}^{T}\left[\Phi(-x_t\beta_1)\right]^{1-d_t}\left\{\frac{\Phi(x_t\beta_1)}{\Phi\left(\frac{x_t\beta_2}{\sigma_2}\right)}\frac{1}{\sqrt{2\pi}\sigma_2}\exp\left[-\frac{(y_t-x_t\beta_2)^2}{2\sigma_2^2}\right]\right\}^{d_t} \quad (3.2.7)$$

where $d_t=1$ if $y_t>0$, $d_t=0$ if $y_t=0$. Equivalently, one can maximize the log-likelihood function

$$L(\beta_1,\beta_2,\sigma_2) = \ln L^*(\beta_1,\beta_2,\sigma_2)$$
$$= \sum_{t=1}^{T}\left[(1-d_t)\ln[\Phi(-x_t\beta_1)]\right. \quad (3.2.8)$$
$$\left.+ d_t\left\{\ln\Phi(x_t\beta_1)-\ln\Phi\left(\frac{x_t\beta_2}{\sigma_2}\right)-\ln(\sqrt{2\pi}\sigma_2)-\frac{(y_t-x_t\beta_2)^2}{2\sigma_2^2}\right\}\right]$$

3.2.2 The LM Test

In order to derive the LM test statistic, it is convenient to reparametize in a way similar to that suggested by Olsen(1978). That is, letting

$$\beta = \frac{\beta_2}{\sigma_2}$$
$$\xi = \beta_1 - \beta \quad (3.2.9)$$
$$h = \frac{1}{\sigma_2}$$

eq.(3.2.8) becomes

$$L(\xi,\beta,h) = \sum_{t=1}^{T}\left\{(1-d_t)\ln\Phi[-x_t(\xi+\beta)]\right. \quad (3.2.10)$$
$$\left.+ d_t\left[\ln\Phi[x_t(\xi+\beta)]-\ln\Phi(x_t\beta)-\frac{1}{2}\ln(2\pi)+\ln h-\frac{1}{2}(hy_t-x_t\beta)^2\right]\right\}$$

Note that when $\xi=0$, eq.(3.2.10) reduces to

$$L_1 = \sum_{t=1}^{T}\left\{(1-d_t)\ln\Phi(-x_t\beta)+d_t\left[-\frac{1}{2}\ln(2\pi)+\ln h-\frac{1}{2}(hy_t-x_t\beta)^2\right]\right\} \quad (3.2.11)$$

which is the log-likelihood function for the Tobit model. The restricted MLE is $\tilde{\vartheta} = (\tilde{\xi},\tilde{\beta},\tilde{h})$ where $\tilde{\xi} = 0$ and $\tilde{\beta}$, $\tilde{h}$ are obtained by maximizing eq.(3.2.11), i.e. $\tilde{\beta}$ and $\tilde{h}$ satisfy the following first order condition:

$$\frac{\partial L_1}{\partial \beta} = \sum_{t=1}^{T} -(1-d_t)\cdot m(-x_t\tilde{\beta})x_t' + d_t(\tilde{h}y_t - x_t\tilde{\beta})x_t' = 0 \qquad (3.2.12)$$

where $m(\cdot) = \frac{\varphi(\cdot)}{\Phi(\cdot)}$, with $\varphi(\cdot)$ being the p.d.f. of N(0,1).

From eq.(3.2.10), the first partial derivatives of the unrestricted likelihood are

$$\frac{\partial L}{\partial \xi} = \sum_{t=1}^{T}\left\{-(1-d_t)\cdot m[-x_t(\xi+\beta)]\cdot x_t' + d_t\cdot m[x_t(\xi+\beta)]\cdot x_t'\right\}$$

$$\frac{\partial L}{\partial \beta} = \sum_{t=1}^{T}\left\{-(1-d_t)\cdot m[-x_t(\xi+\beta)]\cdot x_t' + d_t\left[m[x_t(\xi+\beta)] - m[x_t\beta] + hy_t - x_t\beta\right]x_t'\right\}$$

$$\frac{\partial L}{\partial h} = \sum_{t=1}^{T} d_t\left\{\frac{1}{h} - (hy_t - x_t\beta)y_t\right\}$$

and the second partial derivatives are

$$\frac{\partial^2 L}{\partial \xi \partial \xi'} = \sum_{t=1}^{T}\Big\{x_t'x_t\Big[x_t(\xi+\beta)\cdot m[-x_t(\xi+\beta)] - m^2[-x_t(\xi+\beta)]\Big]$$
$$+d_t\Big[-x_t(\xi+\beta)\cdot m[-x_t(\xi+\beta)] + m^2[-x_t(\xi+\beta)] - x_t(\xi+\beta)\cdot m[x_t(\xi+\beta)]$$
$$- m^2[x_t(\xi+\beta)]\Big]\Big\}$$

$$\frac{\partial^2 L}{\partial \xi \partial \beta'} = \sum_{t=1}^{T}\Big\{x_t'x_t\Big[m[-x_t(\xi+\beta)]\cdot x_t(\xi+\beta) - m^2[-x_t(\xi+\beta)]\Big]$$
$$-d_t\Big[m[-x_t(\xi+\beta)]\cdot x_t(\xi+\beta) - m^2[-x_t(\xi+\beta)] - m[x_t(\xi+\beta)]x_t(\xi+\beta)$$
$$- m^2[x_t(\xi+\beta)]\Big]\Big\}$$
$$= \frac{\partial^2 L}{\partial \xi \partial \xi'}$$

$$\frac{\partial^2 L}{\partial \xi \partial h} = 0$$

$$\frac{\partial^2 L}{\partial \beta \partial \beta'} = \frac{\partial^2 L}{\partial \xi \partial \xi'} - \sum_{t=1}^{T} x_t'x_t d_t[-x_t\beta\cdot m(x_t\beta) - m^2(x_t\beta) + 1]$$

$$\frac{\partial^2 L}{\partial\beta\partial h} = \sum_{t=1}^{T} d_t x_t' y_t$$

$$\frac{\partial^2 L}{\partial h^2} = -\sum_{t=1}^{T} d_t \left(\frac{1}{h^2} + y_t^2\right)$$

The elements of the information matrix (see Appendix C) are

$$I_{\xi\xi} = -E\left[\frac{\partial^2 L}{\partial\xi\partial\xi'}\right]$$

$$= \sum_{t=1}^{T} x_t' x_t \cdot m[-x_t(\xi+\beta)] \cdot m[x_t(\xi+\beta)]$$

$$I_{\xi\beta'} = -E\left[\frac{\partial^2 L}{\partial\xi\partial\beta'}\right]$$

$$= I_{\xi\xi}$$

$$I_{\xi h} = -E\left[\frac{\partial^2 L}{\partial\xi\partial h}\right] = 0$$

$$I_{\beta\beta'} = -E\left[\frac{\partial^2 L}{\partial\beta\partial\beta'}\right]$$

$$= I_{\xi\xi} + \sum_{t=1}^{T} x_t' x_t \cdot \Phi[x_t(\xi+\beta)] \cdot [-x_t\beta \cdot m(x_t\beta) - m^2(x_t\beta) + 1]$$

$$I_{\beta h} = -E\left[\frac{\partial^2 L}{\partial\beta\partial h}\right]$$

$$= -\sum_{t=1}^{T} x_t' \Phi[x_t(\xi+\beta)] \frac{1}{h}[x_t\beta + m(x_t\beta)]$$

$$I_{hh} = -E\left[\frac{\partial^2 L}{\partial h^2}\right]$$

$$= \sum_{t=1}^{T} \frac{1}{h^2}[2 + (x_t\beta)^2 + x_t\beta \cdot m(x_t\beta)] \cdot \Phi[x_t(\xi+\beta)]$$

If we let $\vartheta_1 = \xi$, $\vartheta_2 = (\beta, h)$, then

$$\frac{\partial L}{\partial\tilde{\vartheta}_1} = \frac{\partial L}{\partial\xi} \textit{ evaluated at } \tilde{\vartheta}$$

$$= \sum_{t=1}^{T}\left[-(1-d_t)\cdot m(-x_t\tilde{\beta})x_t' + d_t \cdot m(x_t\tilde{\beta})x_t'\right] \tag{3.2.13}$$

$$= -\sum_{t=1}^{T} d_t[(\tilde{h}y_t - x_t\tilde{\beta}) - m(x_t\tilde{\beta})]x_t' \quad \textit{by eq.}(3.2.11).$$

Also

$$\frac{\partial L}{\partial\tilde{\vartheta}_2} = \begin{bmatrix} \frac{\partial L}{\partial\beta} \\ \frac{\partial L}{\partial h} \end{bmatrix} \textit{ evaluated at } \tilde{\vartheta}$$

$$= \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad by \;\; eq.\,(1.4)$$

The information matrix evaluated at $\tilde{\vartheta}$ is

$$I(\tilde{\vartheta}) = \begin{bmatrix} \tilde{I}_{\xi\xi} & \tilde{I}_{\xi\beta} & \tilde{I}_{\xi h} \\ \tilde{I}_{\beta\xi} & \tilde{I}_{\beta\beta} & \tilde{I}_{\beta h} \\ \tilde{I}_{h\xi} & \tilde{I}_{h\beta} & \tilde{I}_{hh} \end{bmatrix}$$

$$= \begin{bmatrix} X'AX & X'AX & 0 \\ X'AX & X'(A+B)X & (CX)' \\ 0 & CX & D \end{bmatrix}$$

where

$$X = \begin{bmatrix} x_{11} & . & . & . & x_{1k} \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ x_{T1} & . & . & . & x_{Tk} \end{bmatrix} = \begin{bmatrix} x_1 \\ . \\ . \\ . \\ x_T \end{bmatrix}$$

$$A = \begin{bmatrix} a_1 & 0 & 0 & . & . & 0 \\ 0 & a_2 & 0 & . & . & 0 \\ 0 & 0 & a_3 & . & . & 0 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 0 & 0 & 0 & . & . & a_T \end{bmatrix} \quad where \;\; a_t = m(-x_t\tilde{\beta})\cdot m(x_t\tilde{\beta}), \;\; t=1,\dots T.$$

$$B = \begin{bmatrix} b_1 & 0 & 0 & . & . & 0 \\ 0 & b_2 & 0 & . & . & 0 \\ 0 & 0 & b_3 & . & . & 0 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 0 & 0 & 0 & . & . & b_T \end{bmatrix} \quad where \;\; b_t = \Phi(x_t\tilde{\beta})[1 - x_t\tilde{\beta}\cdot m(x_t\tilde{\beta}) - m^2(x_t\tilde{\beta})]$$

$$C = \frac{-1}{\tilde{h}}\left[\Phi(x_1\tilde{\beta})[x_1\tilde{\beta}+m(x_1\tilde{\beta})], \dots, \Phi(x_T\tilde{\beta})[x_T\tilde{\beta}+m(x_T\tilde{\beta})]\right]$$

$$D = \sum_{t=1}^{T} \frac{1}{\tilde{h}^2}[2+(x_t\tilde{\beta})^2 + x_t\tilde{\beta}\cdot m(x_t\tilde{\beta})]\cdot\Phi(x_t\tilde{\beta})$$

Let the inverse of $I(\tilde{\vartheta})$ be

$$[I(\tilde{\vartheta})]^{-1} = \begin{bmatrix} \tilde{I}^{\xi\xi} & \tilde{I}^{\xi\beta} & \tilde{I}^{\xi h} \\ \tilde{I}^{\beta\xi} & \tilde{I}^{\beta\beta} & \tilde{I}^{\beta h} \\ \tilde{I}^{h\xi} & \tilde{I}^{h\beta} & \tilde{I}^{hh} \end{bmatrix}$$

Then, from eq.(1.3) and eq.(3.2.14), (Note 3)

$$LM\ statistic = \begin{bmatrix} \frac{\partial L}{\partial \tilde{\vartheta}_1} \\ 0 \\ 0 \end{bmatrix}' \begin{bmatrix} \tilde{I}^{\xi\xi} & \tilde{I}^{\xi\beta} & I^{\xi h} \\ \tilde{I}^{\beta\xi} & \tilde{I}^{\beta\beta} & I^{\beta h} \\ \tilde{I}^{h\xi} & \tilde{I}^{h\beta} & I^{hh} \end{bmatrix} \begin{bmatrix} \frac{\partial L}{\partial \tilde{\vartheta}_1} \\ 0 \\ 0 \end{bmatrix}$$

$$= \left[\frac{\partial L}{\partial \tilde{\vartheta}_1}\right]' \tilde{I}^{\xi\xi} \left[\frac{\partial L}{\partial \tilde{\vartheta}_1}\right] \tag{3.2.15}$$

We now look for an explicit formula for $\tilde{I}^{\xi\xi}$. Let us rewrite $I(\tilde{\vartheta})$ as

$$I(\tilde{\vartheta}) = \begin{bmatrix} P & R \\ R' & Q \end{bmatrix}$$

where

$$P = \begin{bmatrix} X'AX & X'AX \\ X'AX & X'(A+B)X \end{bmatrix}$$

$$R = \begin{bmatrix} 0 \\ (CX)' \end{bmatrix}$$

$$Q = D$$

Therefore,

$$\begin{bmatrix} \tilde{I}^{\xi\xi} & \tilde{I}^{\xi\beta} \\ \tilde{I}^{\beta\xi} & \tilde{I}^{\beta\beta} \end{bmatrix} = (P - RQ^{-1}R')^{-1}$$

$$= \begin{bmatrix} M & N \\ N' & S \end{bmatrix}^{-1}$$

where

$$M = N = X'AX$$

$$S = X'[(A+B)-C'D^{-1}C]X$$

Note that $M^{-1}N = I$. Thus,

$$\tilde{I}^{\xi\xi} = M^{-1}+M^{-1}N(S-N'M^{-1}N)^{-1}N'M^{-1}$$

$$= (X'AX)^{-1}+\left\{X'[(A+B)-C'D^{-1}C]X-X'AX\right\}^{-1}$$

$$= (X'AX)^{-1}+\left\{X'BX-\frac{(CX)'(CX)}{D}\right\}^{-1} \tag{3.2.16}$$

Substituting eq.(3.2.13) and eq.(3.2.16) back into eq.(3.2.15), we have the LM test statistic.

Although $\widetilde{I}^{\xi\xi}$ has no obvious interpretation, it is easily calculated from the Tobit estimates. On the other hand, $\frac{\partial L}{\partial \tilde{\vartheta}_1}$ is both easily calculated and also easiy interpreted as a vector of cross products between the explanatory variables and the Tobit "residuals" for the non-limit observations. That is so in the sense that

$$E[(\tilde{h}y_t - x_t\tilde{\beta}) \mid y_t > 0] = m(x_t\tilde{\beta}) \quad by\ eq.(C-1)$$

and thus the term in brackets in eq.(3.2.13) can be regarded as the Tobit "residual."

3.3 Test of Sample Selection Bias

3.3.1 Heckman's Sample Selection Bias Model and His λ Test

In some cases, the dependent variable is unobservable while the corresponding independent variables are still available. That is, we have an incomple sample(or censored sample). Heckman(1976,1979) proposed a two-equation model to deal with this situation:

$$y_{1i} = x_{1i}\beta_1 + u_{1i} \tag{3.3.1}$$
$$y_{2i} = x_{2i}\beta_2 + u_{2i} \tag{3.3.2}$$

with

$$\begin{bmatrix} u_{1i} \\ u_{2i} \end{bmatrix} \quad iid \quad N(0, \Sigma), \quad i=1,\dots,n,$$

where

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1 \\ \rho\sigma_1 & 1 \end{bmatrix}$$

since eq.(3.3.2) is a probit equation. We observe the sign of y_{2i} and we observe y_{1i} if and only if $y_{2i} > 0$. That is, $y_{2i} > 0$ is the sample selection rule and we have a nonrandomly selected sample.

This model can be estimated by the maximum likelihood method. But if we use least squares for eq.(3.3.1) and probit analysis for eq.(3.3.2) instead , the resulting estimates of β_1 will be biased. This is so because

$$\begin{aligned} E(y_{1i} \mid x_{1i}, \textit{sample selection rule}) &= E(y_{1i} \mid x_{1i}, y_{2i} > 0) \\ &= x_{1i}\beta_1 + E(u_{1i} \mid y_{2i} > 0) \\ &= x_{1i}\beta_1 + \rho\sigma_1\lambda_i \end{aligned}$$

(See Johnson and Kotz (1970),p.81), where the inverse of the Mill's ratio is

$$\lambda_i = \frac{\varphi(-x_{2i}\beta)}{1-\Phi(-x_{2i}\beta_2)}$$

with $\varphi(\cdot)$ and $\Phi(\cdot)$ being the p.d.f. and c.d.f. of a standard normal distribution respectively.

Thus, the expectation of y_{1i} for the nonrandom sample is not equal to the expectation of y_{1i} for the complete (random) sample unless $E(u_{1i} \mid y_{2i} > 0)$ equals zero; that is, unless $\rho=0$. Therefore, by using eq.(3.3.1),we have the equivalent of an omitted variable problem which will result in a bias in the estimate. This bias is called "sample selection bias." This bias will be eliminated if the conditional mean of u_{1i} is included as a regressor. However, since β_2 is a parameter to be estimated, the λ_i's are unknown. Heckman proposed a simple two-stage estimation procedure to estimate the parameters.

Step 1 — Probit Analysis

Let

$$d_i = \begin{cases} 1 & \text{if } y_{2i} > 0 \\ 0 & \text{if } y_{2i} \leq 0 \end{cases}$$

$$\sum_{i=1}^{n} d_i = n_1$$

$$G_i = Prob.(y_{2i} \leq 0 \mid x_{2i}) = \Phi(-x_{2i}\beta_2)$$

The probit estimate β_2 is obtained by maximizing the following likelihood function:

$$L_1 = \prod_{i=1}^{n} (1-G_i)^{d_i} G_i^{1-d_i}$$

Step 2 — Least Squares

Let

$$\tilde{\lambda}_i = \frac{\varphi(-x_{2i}\tilde{\beta}_2)}{1-\Phi(-x_{2i}\tilde{\beta}_2)}$$

then apply OLS method to the following equation

$$y_{1i} = x_{1i}\beta_1 + \tilde{\lambda}_i c + error, \quad i=,\ldots,n \tag{3.3.3}$$

where $c = \rho\sigma_1$. The OLS estimates are

$$\hat{\beta}_1 = (X_1'X_1)^{-1}X_1'Y_1 - (X_1'X_1)^{-1}X_1'\tilde{\lambda}(\tilde{\lambda}'M_1\tilde{\lambda})^{-1}\tilde{\lambda}'M_1Y_1$$

$$\hat{c} = (\tilde{\lambda}'M_1\tilde{\lambda})^{-1}\tilde{\lambda}'M_1Y_1$$

where X_1 is a $n_1 \times k$ matrix which consists of x_{1i}, Y_1 is a $n_1 \times 1$ vector which consists of y_{1i}, $\tilde{\lambda}$ is a $n_1 \times 1$ vector which consists of $\tilde{\lambda}_i$ corresponding to observed y_{1i}, $M_1 = I - X_1(X_1'X_1)^{-1}X_1'$. Note that $\hat{\beta}_1$ is a consistent estimate of β_1.

If $\rho=0$, $E(y_{1i} \mid x_{1i}, y_{2i} > 0) = x_{1i}\beta_1$. Then, there is no sample selection bias even if we apply OLS to eq.(3.3.1). Therefore, the test of sample selection bias is equivalent to the test of $\rho = 0$. Since $c = \rho\sigma_1$ in eq.(3.3.3), the test of $\rho = 0$ is equivalent to the test of $c = 0$. Heckman uses the standard t test to test the hypothesis $c = 0$.(We will refer to it as the "λ test.") The t statistic is

$$t = \frac{(\tilde{\lambda}'M_1\tilde{\lambda})^{-1}\tilde{\lambda}'M_1Y_1}{[\hat{\sigma}_1^2(\tilde{\lambda}'M_1\tilde{\lambda})^{-1}]^{1/2}} \tag{3.3.4}$$

where $\hat{\sigma}_1^2$ is the usual variance estimate (SSE divided by n_1, or degrees of freedom) from OLS to eq.(3.3.3). This model and the λ test have been widely used, especially in labor economics. Many applications have reported an insignificant value for λ test statistic. One possible conjecture is that the λ test is not a very powerful test of sample selection bias. However, this turns out to be a false conjecture. The λ test is asymptotically equivalent to the LM test, as is shown in the next section, and thus has good asymptotic power properties.

3.3.2 The LM Test of Sample Selection Bias

Let

$$F_i = Prob.(y_{1i}, y_{2i} > 0 \mid x_{1i}, x_{2i}) = \int_{-x_{2i}\beta_2}^{\infty} h(y_{1i} - x_{1i}\beta_1, u_{2i})\,du_{2i} \tag{3.3.5}$$

where $h(\cdot,\cdot)$ is the p.d.f. of $N(0, \Sigma)$. Then the log-likelihood function for eq.(3.3.1) and eq.(3.3.2) is

$$L = \sum_{i=1}^{n}[d_i \ln F_i + (1-d_i)\ln G_i] \tag{3.3.6}$$

The restricted model is the one in which $\rho = 0$ is imposed. When $\rho = 0$, eq.(3.3.5) becomes

$$F_i = \int_{-x_{2i}\beta_2}^{\infty} h_1(y_{1i}-x_{1i}\beta_1)\cdot\varphi(u_{2i})du_{2i}$$
$$= h_1(y_{1i}-x_{1i}\beta_1)\cdot[1-\Phi(-x_{2i}\beta_2)] \quad (3.3.7)$$

where $h_1(\cdot)$ is p.d.f. of $N(0, \sigma_1^2)$. Hence, when $\rho = 0$, eq.(3.3.6) becomes

$$L^* = \sum_{i=1}^{n}\left\{d_i \ln h_1(y_{1i}-x_{1i}\beta_1) + d_i \ln[1-G_i] + (1-d_i)\ln G_i\right\} \quad (3.3.8)$$

The restricted MLE $\tilde{\vartheta} = (\tilde{\rho},\tilde{\beta}_1,\tilde{\beta}_2,\tilde{\sigma}_1)$ is obtained by maximizing L^* w.r.t. β_1, β_2 and σ_1; i.e., $\tilde{\rho} = 0$, $\tilde{\beta}_1$ = OLS estimate from eq.(3.3.1), $\tilde{\beta}_2$ = probit MLE of eq.(3.3.2), $\tilde{\sigma}_1^2 = \frac{1}{n_1}e_1'e_1$ where $e_1 = Y_1 - X_1\tilde{\beta}_1$. That is, when $\rho = 0$, we can estimate β_1, σ_1^2 from eq.(3.3.1) by OLS and estimate β_2 from eq.(3.3.2) by probit analysis.

Letting $\vartheta_1 = \rho$, $\vartheta_2 = (\beta_1,\beta_2,\sigma_1)$, we can use eq.(1.5) to construct the LM statistic. From eq.(3.3.6), the first partial derivatives evaluated at $\tilde{\vartheta}$ are (see Appendix D)

$$\frac{\partial L}{\partial \tilde{\vartheta}} = \begin{bmatrix} \frac{\partial L}{\partial \tilde{\rho}} \\ \frac{\partial L}{\partial \tilde{\beta}_1} \\ \frac{\partial L}{\partial \tilde{\beta}_2} \\ \frac{\partial L}{\partial \tilde{\sigma}_1} \end{bmatrix} = \begin{bmatrix} \frac{\tilde{\lambda}'e_1}{\tilde{\sigma}_1} \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad by \;\; eq.(1.4) \quad (3.3.9)$$

Note that $\tilde{\lambda}$ is a $n_1\times 1$ vector, not a $n\times 1$ vector. The information matrix evaluated at $\tilde{\vartheta}$ (see Appendix D) is

$$I(\tilde{\vartheta}) = \begin{bmatrix} \tilde{I}_{\rho\rho} & \tilde{I}_{\rho\beta_1} & \tilde{I}_{\rho\beta_2} & \tilde{I}_{\rho\sigma_1} \\ . & \tilde{I}_{\beta_1\beta_1} & \tilde{I}_{\beta_1\beta_2} & \tilde{I}_{\beta_1\sigma_1} \\ . & . & \tilde{I}_{\beta_2\beta_2} & \tilde{I}_{\beta_2\sigma_1} \\ . & . & . & \tilde{I}_{\sigma_1\sigma_1} \end{bmatrix}$$

$$
= \begin{bmatrix} \sum_{i=1}^{n}(1-\tilde{G}_i)\tilde{\lambda}_i^2 & \dfrac{\sum_{i=1}^{n}(1-\tilde{G}_i)\tilde{\lambda}_i x_{1i}}{\tilde{\sigma}_1} & 0 & 0 \\ . & \dfrac{\sum_{i=1}^{n}(1-\tilde{G}_i)x_{1i}'x_{1i}}{\tilde{\sigma}_1^2} & 0 & 0 \\ . & . & \sum_{i=1}^{n}x_{2i}'x_{2i}\tilde{\lambda}_i\tilde{m}_i & 0 \\ . & . & . & \dfrac{2\sum_{i=1}^{n}(1-\tilde{G}_i)}{\tilde{\sigma}_1^2} \end{bmatrix} \qquad (3.3.10)
$$

where

$$
\tilde{G}_i = \Phi(-x_{2i}\tilde{\beta}_2)
$$

$$
\tilde{m}_i = \frac{\varphi(-x_{2i}\tilde{\beta}_2)}{\Phi(-x_{2i}\tilde{\beta}_2)}.
$$

Since $plim(1-\tilde{G}_i) = 1 - G_i$, $E(d_i) = 1 - G_i$, and $plim\frac{1}{n}\sum_{i=1}^{n}[d_i - (1-G_i)]\cdot H_i = 0$ with H_i being a nonstochastic variable, we can replace $(1-\tilde{G}_i)$ by d_i in $I(\tilde{\vartheta})$ without affecting $plim\frac{1}{n}I(\tilde{\vartheta})$. Thus, eq.(3.3.10) becomes

$$
I(\tilde{\vartheta}) \approx \begin{bmatrix} \tilde{\lambda}'\tilde{\lambda} & \dfrac{\tilde{\lambda}'X_1}{\tilde{\sigma}_1} & 0 & 0 \\ \dfrac{X_1\tilde{\lambda}}{\tilde{\sigma}_1} & \dfrac{X_1'X_1}{\tilde{\sigma}_1^2} & 0 & 0 \\ 0 & 0 & \sum_{i=1}^{n}x_{2i}'x_{2i}\tilde{\lambda}_i\tilde{m}_i & 0 \\ 0 & 0 & 0 & \dfrac{2n_1}{\tilde{\sigma}_1^2} \end{bmatrix} \qquad (3.3.11)
$$

From eq.(3.3.9), eq.(3.3.11) and eq.(1.5), we have

$$LM\ statistic \approx \left(\frac{\tilde{\lambda}'e_1}{\tilde{\sigma}_1}\right)' \left\{ \tilde{\lambda}'\tilde{\lambda} - \left[\frac{\tilde{\lambda}'X_1}{\tilde{\sigma}_1} \ 0 \ 0\right] \begin{bmatrix} \left(\frac{X_1'X_1}{\tilde{\sigma}_1^2}\right)^{-1} & 0 & 0 \\ 0 & \left(\sum_{i=1}^{n} x_{2i}'x_{2i}\tilde{\lambda}_i\tilde{m}_i\right)^{-1} & 0 \\ 0 & 0 & \frac{\tilde{\sigma}_1^2}{2n_1} \end{bmatrix} \begin{bmatrix} \frac{X_1'\tilde{\lambda}}{\tilde{\sigma}_1} \\ 0 \\ 0 \end{bmatrix} \right\}^{-1}$$

$$\cdot \left(\frac{\tilde{\lambda}'e_1}{\tilde{\sigma}_1}\right)$$

$$= \left(\frac{\tilde{\lambda}'e_1}{\tilde{\sigma}_1}\right)' \left(\tilde{\lambda}'M_1\tilde{\lambda}\right)^{-1} \left(\frac{\tilde{\lambda}'e_1}{\tilde{\sigma}_1}\right)$$

$$= \frac{[(\tilde{\lambda}'M_1\tilde{\lambda})^{-1}\tilde{\lambda}'e_1]^2}{\tilde{\sigma}_1^2(\tilde{\lambda}'M_1\tilde{\lambda})^{-1}} \qquad (3.3.12)$$

where

$$M_1 = I - X_1(X_1'X_1)^{-1}X_1'.$$

Comparing eq.(3.3.12) and eq.(3.3.4), we see that the LM test statistic is almost the square of the t test statistic used to test the coefficient of $\tilde{\lambda}$ when it is added to eq.(3.3.1). The only difference is the difference between $\hat{\sigma}_1^2$ and $\tilde{\sigma}_1^2$, which is asymptotically negligible when $\rho = 0$. In other words, Heckman's λ test is almost the LM test; the simple λ test therefore has desirable large sample properties (Note 4).

3.4 Test of Independence in Poirier's Partial Observability Probit Model

3.4.1 Poirier's Partial Observability Probit Model

In a recent paper Poirier(1980) has proposed the partial observability probit model:

$$\begin{aligned}
y_{i1}^* &= x_i\beta_1 + v_{i1} \\
y_{i2}^* &= x_i\beta_2 + v_{i2} \\
y_{i1} &= \begin{cases} 1 & \text{if } y_{i1}^* > 0 \\ 0 & \text{if} y_{i1}^* \le 0 \end{cases} \\
y_{i2} &= \begin{cases} 1 & \text{if } y_{i2}^* > 0 \\ 0 & \text{if } y_i^* \le 0 \end{cases} \\
z_i &= y_{i1}y_{i2} \\
i &= 1,\ldots n. \\
\begin{bmatrix} v_{i1} \\ v_{i2} \end{bmatrix} &\ iid \quad N(0,\Sigma) \quad where \quad \Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}
\end{aligned} \tag{3.4.1}$$

Here y_{i1}^*, y_{i2}^*, y_{i1} and y_{i2} are unobservable. We observe only x_i and z_i. We observe $z_i = 1$ if and only if $y_{i1} = y_{i2} = 1$, and $z_i = 0$ if $y_{i1} = 0$ or $y_{i2} = 0$ or both.

Some examples of this model are

1) Retention of trainees (see Gunderson (1974))

2) Two-member committee voting anonymously under a unanimity rule (see Poirier (1980)).

3) Colletive bargaining between cities and municipal employees' unions in Michigan; binding arbitration is imposed if either side asks for it (see Connally (1982)).

If y_{i1} and y_{i2} were individually observed, we would simply have a system of two probit equations. Instead we observe only the product of y_{i1} and y_{i2}, and estimation is correspondingly more difficult. If we define

$$p_i = Prob.(z_i = 1) = Prob.(y_{i1} = 1 \text{ and } y_{i2} = 1) = F(x_i\beta_1, x_i\beta_2; \rho) \tag{3.4.2}$$

$$1-p_i = Prob.(z_i=0) = Prob.(y_{i1}=0 \text{ or } y_{i2}=0) = 1-F(x_i\beta_1, x_i\beta_2; \rho) \tag{3.4.3}$$

where F is the bivariate standard normal cumulative distribution function, then, the log-likelihood function for this model is

$$L(\rho,\beta_1,\beta_2) = \sum_{i=1}^{n} z_i \ln p_i + (1-z_i)\ln(1-p_i) \tag{3.4.4}$$

It can be maximized numerically with respect to the parameters ρ, β_1, β_2. The main numerical difficulty involved is the accurate evaluation of the bivariate normal c.d.f. for arbitary ρ. Furthermore, there is some (limited) experience with the model which indicates that ρ is rather hard to estimate. These problems would be avoided if the restriction $\rho = 0$ were imposed. Then the bivariate standard normal c.d.f. factors into the product of two univariate standard normal c.d.f.'s:

$$F(x_i\beta_1, x_i\beta_2; 0) = \Phi(x_i\beta_1)\Phi(x_i\beta_2) \tag{3.4.5}$$

Since univariate normal c.d.f.'s are fairly easy to evaluate, and since the parameter ρ need no longer be estimated, the cost savings from the restriction $\rho = 0$ can be substantial. Given that $\rho = 0$ is a potentially valuable restriction, and the estimation in the restricted model is easier, the LM test can be used to test the hypothesis $\rho = 0$.

3.4.2 The LM Test

In order to construct the LM test statistic, we need the first partial derivatives and the information matrix, both evaluated at the restricted MLE, $\tilde{\vartheta} = (\tilde{\rho}, \tilde{\beta}_1, \tilde{\beta}_2)$ where $\tilde{\rho} = 0$, $\tilde{\beta}_1$ and $\tilde{\beta}_2$ are obtained by maximizing eq.(3.4.4) with $\rho = 0$ being imposed. From eq.(3.4.4), the first partial derivatives of the unrestricted likelihood are

$$\frac{\partial L}{\partial \rho} = \sum_{i=1}^{n} \frac{z_i - p_i}{p_i(1-p_i)} \frac{\partial p_i}{\partial \rho}$$

$$\frac{\partial L}{\partial \beta_1} = \sum_{i=1}^{n} \frac{z_i - p_i}{p_i(1-p_i)} \frac{\partial p_i}{\partial \beta_1}$$

$$\frac{\partial L}{\partial \beta_2} = \sum_{i=1}^{n} \frac{z_i - p_i}{p_i(1-p_i)} \frac{\partial p_i}{\partial \beta_2}.$$

The information matrix is

$$I(\vartheta) = C'C$$

where C is the $n \times (2k+1)$ matrix with ith row equalling

$$c_i = [p_i(1-p_i)]^{-\frac{1}{2}}[f(a_i, b_i; \rho), \varphi(a_i)\Phi(A_i)x_i, \varphi(b_i)\Phi(B_i)x_i]$$

with $\varphi(\cdot)$ and $f(\cdot,\cdot;\cdot)$ being the univariate and bivariate standard normal densities respectively, and where

$$a_i = x_i\beta_1 \quad , \quad A_i = (1-\rho^2)^{-\frac{1}{2}}(b_i - \rho a_i)$$

$$b_i = x_i\beta_2 \quad , \quad B_i = (1-\rho^2)^{-\frac{1}{2}}(a_i - \rho b_i)$$

The first partial derivatives evaluated at $\tilde{\vartheta}$ are:

$$\frac{\partial L}{\partial \tilde{\rho}} = \sum_{i=1}^{n} \frac{z_i - \tilde{p}_i}{\tilde{p}_i(1-\tilde{p}_i)} d_1 d_2$$

where

$$\tilde{p}_i = \Phi(x_i\tilde{\beta}_1)\cdot\Phi(x_i\tilde{\beta}_2) \quad by \quad eq.\,(3.4.2) \text{ and } eq.\,(3.4.5)$$

$$= \Phi(\tilde{a}_i)\cdot\Phi(\tilde{b}_i) \quad with \ \tilde{a}_i = x_i\tilde{\beta}_1 \ , \quad \tilde{b}_i = x_i\tilde{\beta}_2.$$

$$d_1 = \int_{-\infty}^{\tilde{a}_i} v_{i1}\cdot\varphi(v_{i1})dv_{i1}$$

$$= \Phi(\tilde{a}_i)\cdot E(v_{i1} \mid v_{i1} < \tilde{a}_i)$$

$$= \Phi(\tilde{a}_i)\cdot\left[-\frac{\varphi(\tilde{a}_i)}{\Phi(\tilde{a}_i)}\right] \quad \text{(see } Johnson \text{ and } Kotz \text{ (1970), } p.\,83)$$

$$= -\varphi(\tilde{a}_i)$$

$$d_2 = \int_{-\infty}^{\tilde{b}_i} v_{i2}\cdot\varphi(v_{i2})dv_{i2}$$

$$= -\varphi(\tilde{b}_i) \quad by\ the\ same\ reasoning.$$

$$\frac{\partial L}{\partial \tilde{\beta}_1} = \sum_{i=1}^{n} \frac{z_i - \tilde{p}_i}{\tilde{p}_i(1-\tilde{p}_i)} \varphi(\tilde{a}_i)\cdot\Phi(\tilde{b}_i)x_i = 0 \quad by \quad eq.\,(1.4)$$

$$\frac{\partial L}{\partial \tilde{\beta}_2} = \sum_{i=1}^{n} \frac{z_i - \tilde{p}_i}{\tilde{p}_i(1-\tilde{p}_i)} \Phi(\tilde{a}_i)\cdot \varphi(\tilde{b}_i)\cdot x_i = 0 \quad by \quad eq.(1.4)$$

In matrix form,

$$\frac{\partial L}{\partial \tilde{\vartheta}} = \begin{bmatrix} \sum_{i=1}^{n} \frac{z_i - \tilde{p}_i}{\tilde{p}_i(1-\tilde{p}_i)} \varphi(\tilde{a}_i)\cdot\varphi(\tilde{b}_i) \\ \sum_{i=1}^{n} \frac{z_i - \tilde{p}_i}{\tilde{p}_i(1-\tilde{p}_i)} \varphi(\tilde{a}_i)\cdot\Phi(\tilde{b}_i)\cdot x_i \\ \sum_{i=1}^{n} \frac{z_i - \tilde{p}_i}{\tilde{p}_i(1-\tilde{p}_i)} \Phi(\tilde{a}_i)\varphi(\tilde{b}_i)\cdot x_i \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{i=1}^{n} \frac{z_i - \tilde{p}_i}{\tilde{p}_i(1-\tilde{p}_i)} \varphi(\tilde{a}_i)\varphi(\tilde{b}_i) \\ 0 \\ 0 \end{bmatrix} \tag{3.4.6}$$

The information matrix evaluated at $\tilde{\vartheta}$ is

$$I(\tilde{\vartheta}) = \tilde{C}'\tilde{C} \tag{3.4.7}$$

where

$$C = \begin{bmatrix} \frac{\varphi(\tilde{a}_1)\varphi(\tilde{b}_1)}{\sqrt{\tilde{p}_1(1-\tilde{p}_1)}} & \frac{\varphi(\tilde{a}_1)\Phi(\tilde{b}_1)}{\sqrt{\tilde{p}_1(1-\tilde{p}_1)}} x_1 & \frac{\Phi(\tilde{a}_1)\varphi(\tilde{b}_1)}{\sqrt{\tilde{p}_1(1-\tilde{p}_1)}} x_1 \\ \vdots & \vdots & \vdots \\ \frac{\varphi(\tilde{a}_n)\varphi(\tilde{b}_n)}{\sqrt{\tilde{p}_n(1-\tilde{p}_n)}} & \frac{\varphi(\tilde{a}_n)\Phi(\tilde{b}_n)}{\sqrt{\tilde{p}_n(1-\tilde{p}_n)}} x_n & \frac{\Phi(\tilde{a}_n)\varphi(\tilde{b}_n)}{\sqrt{\tilde{p}_n(1-\tilde{p}_n)}} x_n \end{bmatrix} \tag{3.4.8}$$

From eq.(3.4.6) - eq.(3.4.8) and eq.(1.3), we have

$$\begin{aligned} LM\ statistic &= \left[\frac{\partial L}{\partial \tilde{\rho}}\right]^2 (\tilde{C}'\tilde{C})^{-1}_{11} \\ &= \left[\sum_{i=1}^{n} \frac{z_i - \tilde{p}_i}{\tilde{p}_i(1-\tilde{p}_i)} \varphi(\tilde{a}_i)\varphi(\tilde{b}_i)\right]^2 (\tilde{C}'\tilde{C})^{-1}_{11} \end{aligned}$$

where $(\tilde{C}'\tilde{C})^{-1}_{11}$ is the upper left corner element of $(\tilde{C}'\tilde{C})^{-1}$.

This expression can be further simplified as following :

Let

$$\tilde{Q} = \left[\frac{z_1 - \tilde{p}_1}{\sqrt{\tilde{p}_1(1-\tilde{p}_1)}} \cdots \frac{z_n - \tilde{p}_n}{\sqrt{\tilde{p}_n(1-\tilde{p}_n)}} \right]'$$

then

$$\tilde{C}'\tilde{Q} = \begin{bmatrix} \sum_{i=1}^{n} \frac{z_i - \tilde{p}_i}{\tilde{p}_i(1-\tilde{p}_i)} \varphi(\tilde{a}_i)\varphi(\tilde{b}_i) \\ \sum_{i=1}^{n} \frac{z_i - \tilde{p}_i}{\tilde{p}_i(1-\tilde{p}_i)} \varphi(\tilde{a}_i)\Phi(\tilde{b}_i)x_i \\ \sum_{i=1}^{n} \frac{z_i - \tilde{p}_i}{\tilde{p}_i(1-\tilde{p}_i)} \Phi(\tilde{a}_i)\varphi(\tilde{b}_i)x_i \end{bmatrix} \qquad (3.4.9)$$

$$= \frac{\partial L}{\partial \tilde{\vartheta}} \quad by \quad eq.(3.4.6)$$

Therefore,

$$LM \; statistic = \left[\frac{\partial L}{\partial \tilde{\vartheta}}\right]' \left[I(\tilde{\vartheta})\right]^{-1} \left[\frac{\partial L}{\partial \tilde{\vartheta}}\right]$$

$$= (\tilde{C}'\tilde{Q})'(\tilde{C}'\tilde{C})^{-1}(\tilde{C}'\tilde{Q}) \quad by \quad eq.(3.4.7) \text{ and } eq.(3.4.9)$$

$$= \tilde{Q}'\tilde{C}\ddot{\beta} \qquad (3.4.10)$$

where $\ddot{\beta}$ is the OLS estimate for the coefficient of $\tilde{C}$ in the following regression:

$$\tilde{Q} = \tilde{C}\beta + \varepsilon$$

Define

$$\ddot{Q} = \tilde{C}\ddot{\beta}$$
$$e = \tilde{Q} - \ddot{Q}$$

Eq.(3.4.10) becomes

$$LM \; statistic = (\ddot{Q}' + e')\tilde{C}\ddot{\beta}$$
$$= \ddot{Q}'\tilde{C}\ddot{\beta} \quad by \quad e'\tilde{C} = 0$$
$$= \ddot{Q}'\ddot{Q}$$

which equals explained sum of squares in a regression of $\tilde{Q}$ on $\tilde{C}$.

Note that Q can be interpreted as a vector of standardized residuals from the restricted model. This is so since

$$E(z_i) = 1 \cdot Prob.(z_i = 1) = p_i$$

and

$$\begin{aligned} Var.(z_i) &= E\left[z_i - E(z_i)\right]^2 \\ &= (1-p_i)^2 \cdot Prob.(z_i=1) + (0-p_i)^2 \cdot Prob.(z_i=0) \\ &= (1-p_i)^2 \cdot p_i + p_i^2 \cdot (1-p_i) \\ &= p_i(1-p_i). \end{aligned}$$

3.5 Conclusions

In section 3.2 and section 3.4, the LM test statistics are not very simple. However, they are performed from the results of estimating the simpler, restricted models, and they have the reasonable property that they are based on the "residuals" from the estimated restricted model. This avoids estimation of the more complicated alternative models, at least in cases in which the restricted models are not rejected.

In section 3.3, the LM test statistic is almost equal to the square of the t statistic for the "λ test". Therefore, we can use the Heckman's two-stage procedure to construct the LM statistic without estimating the whole system jointly.

CHAPTER IV

STOCHASTIC PRODUCTION/COST FRONTIERS

4.1 Introduction

The theoretical definition of a production function expresses the maximum amount of output obtainable from given input bundles with fixed technology. On the other hand, the traditional econometric methods (least squares of one kind or another) for estimating a production function allow points above the fitted line. Therefore, the resulting fitted function just represents the "average" relationship between inputs and output, and does not necessarily reflect the "frontier" relationship (the production function).

In 1957, Farrell first explored the possibility of estimating the frontier production function in order to bridge the gap between theory and empirical work. Later, other work was done by the use of mathematical programming techniques under a deterministic frontier assumption (see Førsund, Lovell and Schmidt (1980) for references). However, mathematical programming techniques do not lead to estimates with known statistical properties, since no statistical assumptions are made in those models. Schmidt (1976) explicitly added a one-sided disturbance to the traditional production function, which yields the model

$$y_t = f(x_t;\beta) + \varepsilon_t, \quad t = 1, \cdots, T.$$

where y_t is the observed output, $f(x_t;\beta)$ is the maximum output obtainable from inputs x_t, β is an unknown parameter vector to be estimated, and the disturbance term ε_t is non-positive. Although, given a distribution assumption for the disturbance term, the model can be estimated by maximum-likelihood techniques, the asymptotic distribution of the parameter estimates is not known since the usual "regularity conditions" for the application of maximum likelihood are violated (since $y_t \le f(x_t;\beta)$, the range of y depends on the parameters to be estimated). In order to avoid this difficulty, Aigner, Lovell and Schmidt (*ALS*)(1977) proposed a stochastic production frontier, $f(x_t;\beta) + v_t$, with v_t being a symmetric random disturbance -- v_t is assumed to be *i.i.d* as $N(0, \sigma_v^2)$. Thus, the frontier itself can vary randomly across firms, or over time for the same firm due to favorable or unfavorable external events which are beyond the control of the firm. Errors of observation and measurement on output constitute another source of variation in the frontier. Also, the *ALS* model allows the firms to be technically inefficient relative to their own frontier rather than to some sample norm. In summary, the *ALS* model is as follows (eq.(4.1.1) to eq.(4.1.4))

$$y_t = f(x_t;\beta) + v_t - u_t, \quad t=1,\cdots,T \tag{4.1.1}$$

where u_t is a non-negative disturbance term representing the deviation from the stochastic frontier as a result of technical inefficiency. The following assumptions are made:

u_t *is i.i.d. as* $N(0,\sigma_u^2)$ *truncated at zero.* (4.1.2)

u_t and v_t *are independent.* (4.1.3)

v_t *is i.i.d. as* $N(0, \sigma_u^2)$. (4.1.4)

In 1980, Stevenson extended the *ALS* model by allowing a nonzero mode for technical inefficiency. A test of zero mode is considered in Section 4.2. Besides this, Schmidt and Lovell (1979, 1980) extended the *ALS* model in another direction. That is, in *SL* models, not only technical inefficiency but also allocative inefficiency are considered. The simplest *SL* model (*SL* I) which allows technical and allocative inefficiency is in Section 4.3. A more general model (*SL* II) which generalizes *SL* I model to allow systematic allocative inefficiency is in Section 4.4. Lastly, a model (*SL* III) which generalizes *SL* II model and allows correlation between these two inefficiencies is considered in Section 4.5.

4.2 Test of Zero Mode for the Technical Inefficiency in Stevenson's Extension of the ALS Model

4.2.1 Stevenson's Extension of the ALS Model

Stevenson (1980) pointed out that the *ALS* specification about the level of inefficiency (eq.(4.1.2)) implies that the likelihood of inefficient behavior monotonically decreases for increasing levels of inefficiency. This point can be seen clearly in Fig.4.2.1:

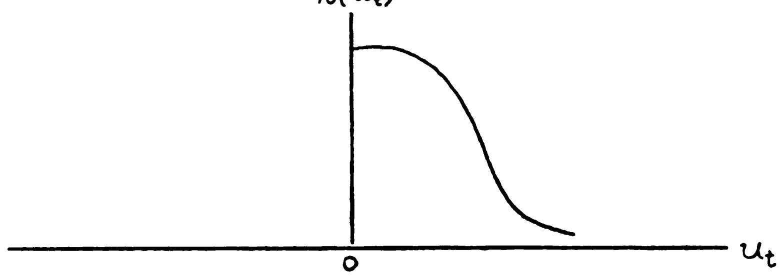

Fig. 4.2.1

where u_t is the level of inefficiency and $k(u_t)$ is the (half-normal) density function of u_t. According to eq.(4.1.2), the mode is at zero, and the normal distribution is truncated at 0, therefore, $k(u_t)$ is a monotonically decreasing function of u_t. Stevenson argues that some characteristics are not likely distributed with such a monotonically declining density function over the population. The possibility of a non-zero mode for the density function of u_t would seem a more reasonable assumption. He, thus, generalizes the *ALS* model by permitting a non-zero mode for the density function of u_t:

$$u_t \sim N(\mu, \sigma_u^2) \textit{ truncated at zero.}$$

Note that *ALS* model is a special case of a zero mode ($\mu = 0$) in Stevenson's model. Since the restriction of $\mu = 0$ can be easily imposed in Stevenson's model and estimation for the *ALS* model is easier, the *LM* test is a suitable one.

4.2.2 The LM Test

Before we derive the *LM* test statistic, we have to derive the likelihood function for Stevenson's model. With a linear model, (e.g. Cobb-Douglas) , eq.(4.1.1) becomes

$$y_t = x_t\beta + v_t - u_t, \quad t=1,\cdots,T.$$

where v_t is i.i.d. as $N(0,\sigma_v^2)$, or explicitly

$$g(v_t) = \frac{1}{\sqrt{2\pi}\sigma_v}\exp\left[-\frac{v_t^2}{2\sigma_v^2}\right] \quad \text{for } all\ v_t$$

and u_t is i.i.d. as $N(\mu,\sigma_u^2)$ truncated at zero, or explicitly,

$$k(u_t) = \begin{cases} \dfrac{1}{\left[1-\Phi\left(-\dfrac{\mu}{\sigma_u}\right)\right]\sqrt{2\pi}\sigma_u}\exp\left[-\dfrac{(u_t-\mu)^2}{2\sigma_u^2}\right] & \text{for } u_t>0 \\ 0 & otherwise \end{cases}$$

with $\Phi(\cdot)$ being the c.d.f. of $N(0,1)$. Therefore, the joint density function for $w_t = v_t - u_t$ is

$$h(w_t) = \int_0^\infty \frac{1}{\left[1-\Phi\left(-\dfrac{\mu}{\sigma_u}\right)\right]2\pi\sigma_u\sigma_v}\exp\left\{-\frac{1}{2}\left[\left(\frac{u_t-\mu}{\sigma_u}\right)^2+\left(\frac{w_t+u_t}{\sigma_v}\right)^2\right]\right\}du_t \qquad (4.2.1)$$

$$= \frac{1}{\sigma}\varphi\left(\frac{w_t+\mu}{\sigma}\right)\cdot\left[1-\Phi\left[\frac{1}{\sigma}\left(-\frac{\mu}{\lambda}+w_t\lambda\right)\right]\right]\cdot\left[1-\Phi\left(-\frac{\mu}{\sigma_u}\right)\right]^{-1}$$

where $\sigma = (\sigma_u^2+\sigma_v^2)^{1/2}$, $\lambda = \frac{\sigma_u}{\sigma_v}$, $\varphi(\cdot) =$ *standard normal p.d.f.* Eq.(4.2.1) utilizes the following formula for integration:

$$\int_0^\infty \exp[-(au^2+bu+c)]du = \frac{1}{2}\sqrt{\frac{\pi}{a}}\exp\left[\frac{(b^2-4ac)}{4a}\right]\cdot erfc\left[\frac{b}{2\sqrt{a}}\right]$$

where $erfc(p) = \frac{2}{\sqrt{\pi}}\int_p^\infty \exp(-u^2)du$

The log-likelihood function for Stevenson's model is

$$L(\mu,\beta,\lambda,\sigma^2) = \sum_{t=1}^{T} \ln h(w_t)$$

$$= -\frac{T}{2}\ln(2\pi) - \frac{T}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_{t=1}^{T}[(y_t - x_t\beta)+\mu]^2$$

$$+ \sum_{t=1}^{T}\ln[1-\Phi(a_t)] - T\ln[1-\Phi(b)]$$

$$where \quad a_t = \frac{1}{\sigma}\left[-\frac{\mu}{\lambda} + (y_t - x_t\beta)\lambda\right]$$

$$b = -\frac{\mu}{\sigma}(\lambda^{-2}+1)^{1/2}$$

The first partial derivatives of the likelihood function are

$$\frac{\partial L}{\partial \mu} = -\frac{1}{\sigma^2}\sum_{t=1}^{T}[(y_t - x_t\beta)+\mu] + \frac{1}{\sigma\lambda}\sum_{t=1}^{T} m(a_t) - \frac{T}{\sigma}(\lambda^{-2}+1)^{1/2}\cdot m(b)$$

$$\frac{\partial L}{\partial \beta} = \frac{1}{\sigma^2}\sum_{t=1}^{T}[(y_t - x_t\beta)+\mu]x_t' + \frac{\lambda}{\sigma}\sum_{t=1}^{T} m(a_t)\cdot x_t'$$

$$\frac{\partial L}{\partial \lambda} = -\frac{1}{\sigma}\sum_{t=1}^{T}\left\{\left[\frac{\mu}{\lambda^2} + (y_t - x_t\beta)\right]\cdot m(a_t)\right\} + \frac{T\mu}{\sigma\lambda^3}(\lambda^{-2}+1)^{-1/2}\cdot m(b)$$

$$\frac{\partial L}{\partial \sigma^2} = -\frac{N}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{t=1}^{T}[(y_t - x_t\beta)+\mu]^2 + \frac{1}{2\sigma^2}\sum_{t=1}^{T} a_t\cdot m(a_t) - \frac{Tb}{2\sigma^2}m(b)$$

$$where \quad m(\cdot) = \frac{\varphi(\cdot)}{1-\Phi(\cdot)}$$

The second partial derivatives are

$$\frac{\partial^2 L}{\partial \mu^2} = -\frac{T}{\sigma^2} + \frac{1}{\sigma^2\lambda^2}\sum_{t=1}^{T} z(a_t) - \frac{T}{\sigma^2}(\lambda^{-2}+1)\cdot z(b)$$

where $z(s) = s\cdot m(s) - m^2(s)$

$$\frac{\partial^2 L}{\partial \mu \partial \beta} = \frac{1}{\sigma^2}\sum_{t=1}^{T} x_t[1+z(a_t)]$$

$$\frac{\partial^2 L}{\partial\mu\partial\lambda} = -\frac{1}{\sigma\lambda^2}\sum_{t=1}^{T} m(a_t) - \frac{1}{\sigma^2\lambda}\sum_{t=1}^{T}\left[\frac{\mu}{\lambda^2} + (y_t - x_t\beta)\right]\cdot z(a_t)$$
$$+ \frac{T}{\sigma\lambda^3}(\lambda^{-2}+1)^{-½}\cdot m(b) + \frac{T\mu}{\sigma^2\lambda^3} z(b)$$

$$\frac{\partial^2 L}{\partial\mu\partial\sigma^2} = \frac{1}{\sigma^4}\sum_{t=1}^{T}[(y_t - x_t\beta)+\mu] - \frac{1}{2\lambda\sigma^2}\sum_{t=1}^{T} m(a_t) + \frac{1}{2\lambda\sigma^3}\sum_{t=1}^{T} a_t\cdot z(a_t)$$
$$+ \frac{T}{2\sigma^3}(\lambda^{-2}+1)^{½}\cdot m(b) - \frac{T}{2\sigma^3}(\lambda^{-2}+1)^{½} b\cdot z(b)$$

$$\frac{\partial^2 L}{\partial\beta\partial\beta'} = \frac{1}{\sigma^2}\sum_{t=1}^{T} x_t' x_t[-1+\lambda^2\cdot z(a_t)]$$

$$\frac{\partial^2 L}{\partial\beta\partial\lambda} = \frac{1}{\sigma}\sum_{t=1}^{T} x_t'\left\{m(a_t) - \frac{1}{\sigma}\left[\frac{\mu}{\lambda} + (y_t - x_t\beta)\lambda\right]\cdot z(a_t)\right\}$$

$$\frac{\partial^2 L}{\partial\beta\partial\sigma^2} = -\frac{1}{\sigma^4}\sum_{t=1}^{T}[(y_t - x_t\beta)+\mu]\cdot x_t' + \frac{\lambda}{2\sigma^3}\sum_{t=1}^{T} x_t' Q(a_t)$$

where $Q(s) = s\cdot z(s) - m(s)$

$$\frac{\partial^2 L}{\partial\lambda^2} = -\frac{1}{\sigma}\sum_{t=1}^{T}\left\{-\frac{2\mu}{\lambda^3} m(a_t) - \frac{1}{\sigma}\left[\frac{\mu}{\lambda^2} + (y_t - x_t\beta)\right]^2\cdot z(a_t)\right\}$$
$$+ \frac{T\mu}{\sigma\lambda^4}(\lambda^{-2}+1)^{-½}\left\{[-3+\lambda^{-2}(\lambda^{-2}+1)^{-1}]\cdot m(b) - \frac{\mu}{\sigma\lambda^2}(\lambda^{-2}+1)^{-½} z(b)\right\}$$

$$\frac{\partial^2 L}{\partial\lambda\partial\sigma^2} = \frac{1}{2\sigma^3}\sum_{t=1}^{T}\left[\frac{\mu}{\lambda^2} + (y_t - x_t\beta)\right]\cdot m(a_t) - \frac{1}{2\sigma^3}\sum_{t=1}^{T} a_t\left[\frac{\mu}{\lambda^2} + (y_t - x_t\beta)\right]\cdot z(a_t)$$
$$+ \frac{T\mu}{2\sigma^3\lambda^3}(\lambda^{-2}+1)^{-½}\cdot Q(b)$$

$$\frac{\partial^2 L}{\partial\sigma^2} = \frac{T}{2\sigma^4} - \frac{1}{\sigma^6}\sum[(y_t - x_t\beta)+\mu]^2 + \frac{1}{4\sigma^4}\sum_{t=1}^{T}[-3a_t\cdot m(a_t) + a_t^2 z(a_t)]$$
$$+ \frac{T}{4\sigma^4} b\cdot m(b)\cdot[3-b^2+b\cdot m(b)]$$

Since it is not possible to calculate analytically $E[z(a_t)]$, $E[a_t \cdot z(a_t)]$, $E[a_t^2 \cdot z(a_t)]$, $E[Q(a_t)]$,..., we can use the negative of the Hessian instead of the informaion matrix to construct the *LM* statistic. Let $\tilde{\vartheta} = (\tilde{\mu}, \tilde{\beta}, \tilde{\lambda}, \tilde{\sigma}^2)$ where $\tilde{\mu} = 0$, $\tilde{\beta}$, $\tilde{\lambda}$, $\tilde{\sigma}^2$ are the *MLE* estimates from *ALS* model. The first partial derivatives evaluated at $\tilde{\vartheta}$ are

$$\frac{\partial L}{\partial \tilde{\mu}} = -\frac{1}{\tilde{\sigma}^2}\sum_{t=1}^{T}(y_t - x_t\tilde{\beta}) + \frac{1}{\tilde{\sigma}\tilde{\lambda}}\sum_{t=1}^{T} m(\tilde{a}_t) - \frac{T}{\tilde{\sigma}}(\tilde{\lambda}^{-2}+1)^{1/2} \cdot \left[\frac{2}{\pi}\right]^{1/2} \qquad (4.2.2)$$

where $\tilde{a}_t = \frac{1}{\tilde{\sigma}}(y_t - x_t\tilde{\beta})\tilde{\lambda}$

$$\frac{\partial L}{\partial \tilde{\beta}} = 0$$

$$\frac{\partial L}{\partial \tilde{\lambda}} = 0 \qquad \textit{by eq. (1.4)}$$

$$\frac{\partial L}{\partial \tilde{\sigma}^2} = 0$$

The elements of negative Hessian evaluated at $\tilde{\vartheta}$ can be used as substitutes for the elements of $I(\tilde{\vartheta})$. They are :

$$\tilde{H}_{\mu\mu} = \frac{T}{\tilde{\sigma}^2} - \frac{1}{\tilde{\sigma}^2\tilde{\lambda}^2}\sum_{t=1}^{T}\tilde{z}_t - \frac{2}{\pi}\frac{T}{\tilde{\sigma}^2}(\tilde{\lambda}^{-2} + 1)$$

where $\tilde{z}_t = \tilde{a}_t \cdot m(\tilde{a}_t) - m^2(\tilde{a}_t)$

$$\tilde{H}_{\mu\beta} = -\frac{1}{\tilde{\sigma}^2}\sum_{t=1}^{T} x_t(1+\tilde{z}_t)$$

$$\tilde{H}_{\mu\lambda} = \frac{1}{\tilde{\sigma}\tilde{\lambda}^2}\sum_{t=1}^{T} m(\tilde{a}_t) + \frac{1}{\tilde{\sigma}^2\tilde{\lambda}}\sum_{t=1}^{T}(y_t - x_t\tilde{\beta})\tilde{z}_t - \frac{T}{\tilde{\sigma}\tilde{\lambda}^3}(\tilde{\lambda}^{-2}+1)^{-2}\left[\frac{2}{\pi}\right]^{1/2}$$

$$\tilde{H}_{\mu\sigma^2} = -\frac{1}{\tilde{\sigma}^4}\sum_{t=1}^{T}(y_t - x_t\tilde{\beta}) - \frac{1}{2\tilde{\lambda}\tilde{\sigma}^3}\sum_{t=1}^{T}\tilde{Q}_t - \frac{T}{2\tilde{\sigma}^3}(\tilde{\lambda}^{-2}+1)^{1/2}\left[\frac{2}{\pi}\right]^{1/2}$$

where $\tilde{Q}_t = \tilde{a}_t\tilde{z}_t - m(\tilde{a}_t)$

$$\tilde{H}_{\beta\beta} = \frac{1}{\tilde{\sigma}^2}\sum_{t=1}^{T} x_t' x_t(1-\tilde{\lambda}^2\tilde{z}_t)$$

$$\tilde{H}_{\beta\lambda} = \frac{1}{\tilde{\sigma}^2}\sum_{t=1}^{T} x_t' \tilde{Q}_t$$

$$\tilde{H}_{\beta\sigma^2} = \frac{1}{\tilde{\sigma}^4}\sum_{t=1}^{T}(y_t - x_t\tilde{\beta})x_t' - \frac{\tilde{\lambda}}{2\tilde{\sigma}^3}\sum_{t=1}^{T} x_t'\tilde{Q}_t$$

$$\tilde{H}_{\lambda\lambda} = -\frac{1}{\tilde{\sigma}^2}\sum_{t=1}^{T}(y_t - x_t\tilde{\beta})^2\tilde{z}_t$$

$$\tilde{H}_{\lambda\sigma^2} = \frac{1}{2\tilde{\sigma}^3}\sum_{t=1}^{T}(y_t - x_t\tilde{\beta})\tilde{Q}_t$$

$$\tilde{H}_{\sigma^2\sigma^2} = -\frac{T}{2\tilde{\sigma}^4} + \frac{1}{\tilde{\sigma}^6}\sum_{t=1}^{T}(y_t - x_t\tilde{\beta})^2 - \frac{1}{4\tilde{\sigma}^4}\sum_{t=1}^{T}[-3\tilde{a}_t \cdot m(\tilde{a}_t) + \tilde{a}_t^2 \cdot \tilde{z}_t]$$

The negative of Hessian evaluated at $\tilde{\vartheta}$ is

$$H(\tilde{\vartheta}) = \begin{bmatrix} \tilde{H}_{\mu\mu} & \tilde{H}_{\mu\beta} & \tilde{H}_{\mu\lambda} & \tilde{H}_{\mu\sigma^2} \\ \cdot & \tilde{H}_{\beta\beta} & \tilde{H}_{\beta\lambda} & \tilde{H}_{\beta\sigma^2} \\ \cdot & \cdot & \tilde{H}_{\lambda\lambda} & \tilde{H}_{\lambda\sigma^2} \\ \cdot & \cdot & \cdot & \tilde{H}_{\sigma^2\sigma^2} \end{bmatrix}$$

Let the inverse of $H(\tilde{\vartheta})$ be

$$[H(\tilde{\vartheta})]^{-1} = \begin{bmatrix} \tilde{H}^{\mu\mu} & \tilde{H}^{\mu\beta} & \tilde{H}^{\mu\lambda} & \tilde{H}^{\mu\sigma^2} \\ \cdot & \tilde{H}^{\beta\beta} & \tilde{H}^{\beta\lambda} & \tilde{H}^{\beta\sigma^2} \\ \cdot & \cdot & \tilde{H}^{\lambda\lambda} & \tilde{H}^{\lambda\sigma^2} \\ \cdot & \cdot & \cdot & \tilde{H}^{\sigma^2\sigma^2} \end{bmatrix}$$

then,

$$LM\ statistic \approx \begin{bmatrix} \frac{\partial L}{\partial \tilde{\mu}} \\ 0 \\ 0 \\ 0 \end{bmatrix}' \begin{bmatrix} \hat{H}^{\mu\mu} & \tilde{H}^{\mu\beta} & \tilde{H}^{\mu\lambda} & \tilde{H}^{\mu\sigma^2} \\ \cdot & \tilde{H}^{\beta\beta} & \tilde{H}^{\beta\lambda} & \tilde{H}^{\beta\sigma^2} \\ \cdot & \cdot & \tilde{H}^{\lambda\lambda} & \tilde{H}^{\lambda\sigma^2} \\ \cdot & \cdot & \cdot & \tilde{H}^{\sigma^2\sigma^2} \end{bmatrix} \begin{bmatrix} \frac{\partial L}{\partial \tilde{\mu}} \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

$$= \left(\frac{\partial L}{\partial \tilde{\mu}}\right)' \tilde{H}^{\mu\mu} \left(\frac{\partial L}{\partial \tilde{\mu}}\right)$$

where $\frac{\partial L}{\partial \tilde{\mu}}$ is given in eq.(4.2.2) and $\tilde{H}^{\mu\mu}$ is the left corner of the inverse of $H(\tilde{\vartheta})$.

4.3 Test of Allocative Inefficiency in SL I Model

4.3.1 SL I Model

Although the *ALS* model and Stevenson's extension can deal with technical inefficiency, there is another kind of inefficiency which is not considered in the above models. A production process can be inefficient in two ways:

1) Technical inefficiency

   It fails to produce maximum output from a given input bundle. This results in an equiproportionate overutilization of all inputs.

2) Allocative inefficiency

   The marginal revenue product of an input is not equal to the marginal cost of that input. This results in utilization of inputs in the wrong proportions, given input prices.

Since *ALS* model and Stevenson's extension do not relate to input prices, they cannot detect allocative inefficiency.

Schmidt and Lovell (SL)(1979) extend the *ALS* model to permit allocative inefficiency. They assume that the firm seeks to minimize the cost of producing its desired rate of output, subject to a stochastic production frontier constraint. If the firm is technically inefficient, it operates beneath its stochastic production frontier, and if the firm is allocatively inefficient it operates off its least cost expansion path. The least cost expansion path can be derived as follows:
Let x_i, $i=1,\cdots,n$ be the n inputs, p_i, $i=1,\cdots,n$ be the prices of the inputs, y_t be the firm's output. We assume the firm's production technology is characterized by a Cobb-Douglas production function of the form:

$$y_t = e^{A}\prod_{i=1}^{n} x_i^{a_i} e^{w_t} \qquad (4.3.1)$$

where $w_t = v_t - u_t$ is a random disturbance, v_t is a random disturbance due to external shocks, u_t is a random disturbance due to technical inefficiency, A is a constant term. The minimum cost input combination can be obtained by minimizing the cost function

$$c_t = \sum_{i=1}^{n} p_{it} x_{it}$$

subject to eq.(4.3.1). From the first order condition, we have

$$\frac{x_{1t}}{x_{it}} = \frac{p_{it}\alpha_1}{p_{1t}\alpha_i}, \quad i=2,\cdots,n \tag{4.3.2}$$

This can be rewritten as

$$\ln x_{1t} - \ln x_{it} = B_{it}, \quad i=2,\cdots,n \tag{4.3.3}$$

where $B_{it} = \ln\left(\frac{p_{it}\alpha_1}{p_{1t}\alpha_i}\right)$

Eq.(4.3.3) represents the least cost expansion path. The deviation from the least cost expansion path can be expressed as a disturbance term, ε_i, added to the right hand side of eq.(4.3.3). That is, ε_{it} measures the percent by which the chosen $\frac{x_{1t}}{x_{it}}$ ratio deviates from its cost minimizing value.

In summary, the *SL* I model

$$\ln y_t = A + \sum_{i=1}^{n} \alpha_i \ln x_{it} + w_t, \quad t=1,\cdots,T \quad (Note\ 5) \tag{4.3.4}$$

where $w_t = v_t - u_t$, v_t *is i.i.d. as* $N(0, \sigma_v^2)$, u_t *is i.i.d. as* $N(0, \sigma_u^2)$ truncated at zero and

$$\ln x_{1t} - \ln x_{it} = B_{it} + \varepsilon_{it}, \quad i=2,\cdots,n \tag{4.3.5}$$

where

$$B_{it} = \ln\left(\frac{p_{it}\alpha_1}{p_{1t}\alpha_i}\right),$$

$$\varepsilon_t = (\varepsilon_{2t},\cdots,\varepsilon_{nt})' \quad i.i.d. \quad N(0, \Sigma_{\varepsilon\varepsilon}),$$

$$\Sigma_{\varepsilon\varepsilon} = \begin{bmatrix} \sigma_{22} & \cdot & \cdot & \sigma_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \sigma_{n2} & \cdot & \cdot & \sigma_{nn} \end{bmatrix},$$

ε_t *is independent of* v_t and u_t.

Note that, when $\Sigma_{\varepsilon\varepsilon} = 0$, this *SL* I model reduces to the *ALS* model in which there is no allocative inefficiency. That is, we can test the null hypothesis of exact cost minimization by testing whether $\Sigma_{\varepsilon\varepsilon}$ equals 0. However, the *LM* test fails as shown in Section 4.3.2. A simple test is suggested in Section 4.3.3.

4.3.2 The LM Test

From eq.(4.3.4), the p.d.f. for w_t is

$$h(w_t) = \frac{2}{\sigma}\varphi\left[\frac{w_t}{\sigma}\right]\left[1-\Phi(a_t)\right] \tag{4.3.6}$$

where

$$a_t = \frac{w_t\lambda}{\sigma}$$
$$\sigma = (\sigma_u^2 + \sigma_v^2)^{1/2}$$
$$\lambda = \frac{\sigma_u}{\sigma_v}$$

From eq.(4.3.5), the p.d.f. for ε_t is

$$\psi(\varepsilon_t) = (2\pi)^{-\frac{n-1}{2}}|\Sigma_{\varepsilon\varepsilon}|^{-1/2}\exp\left[-\frac{1}{2}\varepsilon_t'\Sigma_{\varepsilon\varepsilon}^{-1}\varepsilon_t\right] \tag{4.3.7}$$

Since ε_t is independent of w_t , the joint density function is

$$\begin{aligned} f(w_t,\varepsilon_t) &= h(w_t)\cdot\psi(\varepsilon_t) \\ &= 2(2\pi)^{-\frac{2}{n}}\cdot\frac{1}{\sigma}|\Sigma_{\varepsilon\varepsilon}|^{-1/2}[1-\Phi(a_t)]\exp\left[-\frac{1}{2}\left[\varepsilon_t'\Sigma_{\varepsilon\varepsilon}^{-1}\varepsilon_t + \frac{w_t^2}{\sigma^2}\right]\right] \end{aligned} \tag{4.3.8}$$

The Jacobian of the transformation from $(w_t,\varepsilon_t')'$ to $[lnx_{1t},\cdots,\ln x_{nt}]'$ is $r = \sum_{i=1}^{n}\alpha_i$. Therefore, eq.(4.3.7) becomes

$$g(\ln x_{1t},\cdots,\ln x_{nt}) = 2(2\pi)^{-\frac{n}{2}}\frac{r}{\sigma}|\Sigma_{\varepsilon\varepsilon}|^{-1/2}[1-\Phi(a_t)]\exp\left[-\frac{1}{2}\left[\varepsilon_t'\Sigma_{\varepsilon\varepsilon}^{-1}\varepsilon_t + \frac{w_t^2}{\sigma^2}\right]\right]$$

where

$$w_t = \ln y_t - A - \sum_{i=1}^{n} \alpha_i \ln x_{it}$$
$$\varepsilon_t = (\varepsilon_{2t}, \cdots, \varepsilon_{nt})'$$
$$\varepsilon_{it} = \ln x_{1t} - \ln x_{it} - B_{it}$$
$$B_{it} = \ln\left[\frac{p_{it}\alpha_1}{p_{1t}\alpha_i}\right]$$

Hence the log-likelihood function for *SL* I model (eq.(4.3.4) and (4.3.5)) is

$$L(\vartheta) = \sum_{t=1}^{T} \ln g(\ln x_{1t}, \cdots, \ln x_{nt}) \qquad (4.3.9)$$
$$= T\ln 2 - \frac{Tn}{2}\ln 2\pi + T\ln r - T\ln\sigma - \frac{T}{2}\ln|\Sigma_{\varepsilon\varepsilon}| + \sum_{t=1}^{T}\ln[1-\Phi(a_t)]$$
$$- \frac{1}{2}\sum_{t=1}^{T}\left[\varepsilon_t'\Sigma_{\varepsilon\varepsilon}^{-1}\varepsilon_t + \frac{w_t^2}{\sigma^2}\right]$$

where $\vartheta = (A, \lambda, \sigma, \Sigma_{\varepsilon\varepsilon}, \alpha_1, \cdots, \alpha_n)$.

In order to construct the *LM* test, we need the first partial derivatives with respect to elements of $\Sigma_{\varepsilon\varepsilon}$, evaluated at the restricted *MLE* $\tilde{\vartheta} = (\tilde{A}, \tilde{\lambda}, \tilde{\sigma}, 0, \tilde{\alpha}_1, \cdots, \tilde{\alpha}_n)$, where $\tilde{A}, \tilde{\lambda}, \tilde{\sigma}, \tilde{\alpha}_1, \cdots, \tilde{\alpha}_n$ are the *MLE's* subject to the restriction. The first partial derivative with respect to $\Sigma_{\varepsilon\varepsilon}$ is

$$\frac{\partial \ln L}{\partial \Sigma_{\varepsilon\varepsilon}} = -\frac{T}{2}\Sigma_{\varepsilon\varepsilon}^{-1} - \frac{1}{2}\sum_{t=1}^{T}\frac{\partial \varepsilon_t'\Sigma_{\varepsilon\varepsilon}^{-1}\varepsilon_t}{\partial \Sigma} \qquad (4.3.10)$$

The first term in the right hand side of eq.(4.3.10) does not exist when $\frac{\partial \ln L}{\partial \Sigma_{\varepsilon\varepsilon}}$ is evaluated at $\tilde{\vartheta}$ since $\Sigma_{\varepsilon\varepsilon}^{-1}$ does not exist when $\Sigma_{\varepsilon\varepsilon} = 0$. Therefore, the *LM* test can not be used in this model to test H_0: $\Sigma_{\varepsilon\varepsilon} = 0$.

4.3.3 An Alternative Test

Since the *LM* test fails in *SL* I model, an alternative test is suggested in this section. Recall that exact cost minimization is attained when eq.(4.3.2) holds, i.e., when the following equation holds:

$$\frac{p_{it}x_{it}}{p_{1t}x_{1t}} = \frac{\alpha_i}{\alpha_1}$$

A simple test of exact cost minimization is to see if $\frac{p_{it}x_{it}}{p_{1t}x_{1t}}$ are the same for all observations; they should be exactly equal to $\frac{\alpha_i}{\alpha_1}$ for each observation. Therefore this test has a power which is equal to 1 (i.e. under H_0, Prob.(type I error) = 0, under H_A, Prob.(type II error) = 0). Since

$$\frac{p_{it}x_{it}}{p_{1t}x_{1t}} = \frac{\frac{p_{it}x_{it}}{c_t}}{\frac{p_{1t}x_{1t}}{c_t}} = \frac{factor\ share\ \text{for}\ input\ i\ \text{for}\ observation\ t}{factor\ share\ \text{for}\ input\ 1\ \text{for}\ observation\ t}$$

$$= ratio\ of\ factor\ shares\ \text{for}\ observation\ t \quad ,$$

this test is based on the ratios of factor shares.

4.4 Test of Systematic Allocative Inefficiency in SL II Model

4.4.1 SL II Model

The *SL* II model is composed of eq.(4.3.4) and

$$\ln x_{1t} - \ln x_{it} = B_{it} + \varepsilon_{it}, \quad i=2, \cdots, n \tag{4.4.1}$$

where

$$B_{it} = \ln\left(\frac{p_{it}\alpha_1}{p_{1t}\alpha_i}\right)$$

$$\varepsilon_t = (\varepsilon_{2t}, \cdots, \varepsilon_{nt})' \text{ i.i.d. } N(\xi, \Sigma_{\varepsilon\varepsilon}), \text{ with } \Sigma_{\varepsilon\varepsilon} = \begin{pmatrix} \sigma_{22} & \cdot & \cdot & \sigma_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \sigma_{n2} & \cdot & \cdot & \sigma_{nn} \end{pmatrix}$$

ε_t *is independent of* v_t and u_t.

Note that $E(\varepsilon_t) = \xi$. If $\xi = 0$, this reduces to *SL* I model and implies that there is no systematic tendency to over- or under-utilize any input relative to any other input. There is a well-known argument (the Averch-Johnson hypothesis) that suggests that firms which are subject to rate of return regulation will tend to use higher ratios of capital to other inputs than cost minimization would dictate. Schmidt and Lovell (1979) supported this hypothesis by using data on steam-electric generating plants. In their paper, they used the likelihood ratio test for the null hypothesis H_0: $\xi = 0$. Since the *SL* II model is easier to estimate when $\xi = 0$ is imposed, we can use the *LM* test to test this null hypothesis.

4.4.2 The LM Test

From eq.(4.4.1), the p.d.f. for ε_t is

$$\psi(\varepsilon_t) = (2\pi)^{-\frac{n-1}{2}} |\Sigma_{\varepsilon\varepsilon}|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\varepsilon_t-\xi)'\Sigma_{\varepsilon\varepsilon}^{-1}(\varepsilon_t-\xi)\right] \tag{4.4.2}$$

Since ε_t is independent of w_t from eq.(4.3.6) and eq.(4.4.2), the joint density function is

$$\begin{aligned} f(w_t,\varepsilon_t) &= h(w_t)\cdot\psi(\varepsilon_t) \qquad (4.4.3)\\ &= 2(2\pi)^{-\frac{n}{2}}\cdot\frac{1}{\sigma}|\Sigma_{\varepsilon\varepsilon}|^{-\frac{1}{2}}\cdot[1-\Phi(a_t)]\exp\left[-\frac{1}{2}(\varepsilon_t-\xi)'\Sigma_{\varepsilon\varepsilon}^{-1}(\varepsilon_t-\xi)-\frac{1}{2}\left(\frac{w_t}{\sigma}\right)^2\right] \end{aligned}$$

The Jacobian of the transformation from $(w_t,\varepsilon_t)'$ to $[\ln x_{1t},\cdots,\ln x_{nt}]'$ is $r = \sum_{i=1}^{n}\alpha_i$. Therefore eq.(4.4.3) becomes

$$\begin{aligned} g(\ln x_{1t},\cdots,\ln x_{nt}) &= (2\pi)^{-\frac{n}{2}}\cdot\frac{r}{\sigma}|\Sigma_{\varepsilon\varepsilon}|^{-\frac{1}{2}}\cdot 2[1-\Phi(a_t)]\\ &\cdot\exp\left[-\frac{1}{2}(\varepsilon_t-\xi)'\Sigma_{\varepsilon\varepsilon}^{-1}(\varepsilon_t-\xi)-\frac{1}{2}\left(\frac{w_t}{\sigma}\right)^2\right] \end{aligned}$$

where

$$\begin{aligned} w_t &= \ln y_t - A - \sum_{i=1}^{n}\alpha_i\ln x_{it},\\ \varepsilon_t &= (\varepsilon_{2t},\cdots,\varepsilon_{nt})',\\ \varepsilon_{it} &= \ln x_{1t} - \ln x_{it} - B_{it},\\ B_{it} &= \left(\frac{p_{it}\alpha_1}{p_{1t}\alpha_i}\right). \end{aligned}$$

Hence the log-likelihood function for *SL* II model (eq.(4.3.4) and eq.(4.4.1)) is

$$\begin{aligned} L(\vartheta) &= \sum_{t=1}^{T}\ln g(\ln x_{1t},\cdots,\ln x_{nt}) \qquad (4.4.4)\\ &= -\frac{Tn}{2}\ln 2\pi + T\ln r - T\ln\sigma - \frac{T}{2}\ln|\Sigma_{\varepsilon\varepsilon}| + \sum_{t=1}^{T}\ln[1-\Phi(a_t)]\\ &- T\ln\frac{1}{2} - \frac{1}{2}\sum_{t=1}^{T}(\varepsilon_t-\xi)'\Sigma_{\varepsilon\varepsilon}^{-1}(\varepsilon_t-\xi) - \frac{1}{2}\sum_{t=1}^{T}\left(\frac{w_t}{\sigma}\right)^2 \end{aligned}$$

where $\vartheta = (\xi,A,\lambda,\sigma,\Sigma_{\varepsilon\varepsilon},\alpha_1,\cdots,\alpha_n)$.

The restricted *MLE* is $\tilde{\vartheta} = (\tilde{\xi},\tilde{A},\tilde{\lambda},\tilde{\sigma},\tilde{\Sigma}_{\varepsilon\varepsilon},\tilde{\alpha}_1,\cdots,\tilde{\alpha}_n)$ where $\tilde{\xi} = 0$ and $\tilde{A},\tilde{\lambda},\tilde{\sigma},\tilde{\Sigma}_{\varepsilon\varepsilon},\tilde{\alpha}_1,\cdots,\tilde{\alpha}_n$ are the *MLE*'s from *SL* I model when $\xi = 0$ is imposed.

Then, from eq.(E-1) in Appendix E, and eq.(1.4), the first partial derivatives evaluated at $\tilde{\vartheta}$ are

$$D(\tilde{\vartheta}) = \begin{bmatrix} \frac{\partial L}{\partial \xi} \\ 0 \end{bmatrix} = \begin{bmatrix} \tilde{\Sigma}_{\varepsilon\varepsilon}^{-1}[\sum_{t=1}^{T} \tilde{\varepsilon}_t] \\ 0 \end{bmatrix} \tag{4.4.5}$$

where $\tilde{\Sigma}_{\varepsilon\varepsilon}^{-1}$, $\tilde{\varepsilon}_t$ are $\Sigma_{\varepsilon\varepsilon}^{-1}$, ε_t evaluated at $\tilde{\vartheta}$ respectively. Also, from Appendix E, the information matrix evaluated at $\tilde{\vartheta}$ is

$$I(\tilde{\vartheta}) = \begin{bmatrix} \tilde{I}_{\xi\xi} & 0 & 0 & \tilde{I}_{\xi\alpha} \\ 0 & \tilde{I}_{22} & 0 & \tilde{I}_{24} \\ 0 & 0 & \tilde{I}_{\Sigma_{\varepsilon\varepsilon}\Sigma_{\varepsilon\varepsilon}} & 0 \\ \tilde{I}'_{\xi\alpha} & \tilde{I}'_{24} & 0 & \tilde{I}_{\alpha\alpha} \end{bmatrix} \tag{4.4.6}$$

where

$$\tilde{I}_{\xi\xi} = \begin{bmatrix} \tilde{I}_{\xi_2\xi_2} & \cdot & \cdot & \tilde{I}_{\xi_2\xi_n} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \tilde{I}_{\xi_n\xi_2} & \cdot & \cdot & \tilde{I}_{\xi_n\xi_n} \end{bmatrix} = T\tilde{\Sigma}_{\varepsilon\varepsilon}^{-1}$$

$$\tilde{I}_{\xi\alpha} = \begin{bmatrix} \tilde{I}_{\xi_2\alpha_1} & \cdot & \cdot & \tilde{I}_{\xi_2\alpha_n} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \tilde{I}_{\xi_n\alpha_1} & \cdot & \cdot & \tilde{I}_{\xi_n\alpha_n} \end{bmatrix}$$

with

$$\tilde{I}_{\xi_l\alpha_1} = T\tilde{\alpha}_1^{-1}\sum_{i=2}^{n}\tilde{\sigma}^{il}, \qquad l = 2, \cdots, n.$$

$$\tilde{I}_{\xi_l\alpha_h} = -T\tilde{\alpha}_h^{-1}\tilde{\sigma}^{hl}, \qquad l, h = 2, \cdots, n.$$

$$\tilde{I}_{22} = \begin{bmatrix} \tilde{I}_{AA} & \tilde{I}_{A\lambda} & \tilde{I}_{A\sigma} \\ \cdot & \tilde{I}_{\lambda\lambda} & \tilde{I}_{\lambda\sigma} \\ \cdot & \cdot & \tilde{I}_{\sigma\sigma} \end{bmatrix}$$

(see eq.(E-2)' to (E-9)')

$$\tilde{I}_{24} = \begin{bmatrix} \tilde{I}_{A\alpha_1} & \cdot & \cdot & \cdot & \tilde{I}_{A\alpha_n} \\ \tilde{I}_{\lambda\alpha_1} & \cdot & \cdot & \cdot & \tilde{I}_{\lambda\alpha_n} \\ \tilde{I}_{\sigma\alpha_1} & \cdot & \cdot & \cdot & \tilde{I}_{\sigma\alpha_n} \end{bmatrix}$$

(see eq.(E-6)', (E-10)', (E-13)', and (E-15)')

$$\tilde{I}_{\Sigma_{\varepsilon\varepsilon}\Sigma_{\varepsilon\varepsilon}} = \begin{bmatrix} \tilde{I}_{\sigma_{22}\sigma_{22}} & \cdots & \tilde{I}_{\sigma_{22}\sigma_{nn}} \\ \vdots & & \vdots \\ \tilde{I}_{\sigma_{nn}\sigma_{22}} & \cdots & \tilde{I}_{\sigma_{nn}\sigma_{nn}} \end{bmatrix}$$

(see eq.(E-11)' to (E-13)')

$$\tilde{I}_{aa} = \begin{bmatrix} \tilde{I}_{a_1a_1} & \cdots & \tilde{I}_{a_1a_n} \\ \vdots & & \vdots \\ \tilde{I}_{a_na_1} & \cdots & \tilde{I}_{a_na_n} \end{bmatrix}$$

(see eq.(E-14)')

From eqs.(4.4.5), (4.4.6), and (1.5),

$$LM\ statistic = \left(\frac{\partial L}{\partial \tilde{\xi}}\right)' \left\{ \tilde{I}_{\xi\xi} - [0,\ 0,\ \tilde{I}_{\xi a}] \begin{bmatrix} \tilde{I}_{22} & 0 & \tilde{I}_{24} \\ 0 & \tilde{I}_{\Sigma_{\varepsilon\varepsilon}\Sigma_{\varepsilon\varepsilon}} & 0 \\ \tilde{I}_{24}' & 0 & \tilde{I}_{aa} \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 0 \\ \tilde{I}_{\xi a}' \end{bmatrix} \right\}^{-1} \left(\frac{\partial L}{\partial \tilde{\xi}}\right)$$

$$= \left(\frac{\partial L}{\partial \tilde{\xi}}\right)' \left\{ \tilde{I}_{\xi\xi} - \tilde{I}_{\xi a}[\tilde{I}_{aa} - \tilde{I}_{24}'\tilde{I}_{22}^{-1}\tilde{I}_{24}]^{-1}\tilde{I}_{\xi a}' \right\}^{-1} \left(\frac{\partial L}{\partial \tilde{\xi}}\right)$$

Note that the average of the ε_t, $\bar{\varepsilon}$, is distributed as normal with mean ξ and varianve-covariance matrix $\frac{1}{T}\Sigma_{\varepsilon\varepsilon}$. Therefore, $T\bar{\varepsilon}\Sigma_{\varepsilon\varepsilon}^{-1}\bar{\varepsilon}$ follows a χ^2 distribution with $n-1$ degrees of freedom. Since $\bar{\varepsilon}$ is near 0 if $E(\varepsilon)=\xi=0$, then we could accept H_0: $E(\varepsilon)=\xi=0$ if $T\bar{\varepsilon}\Sigma_{\varepsilon\varepsilon}^{-1}\bar{\varepsilon}$ is near zero. However, we do not observe $\bar{\varepsilon}$ and $\Sigma_{\varepsilon\varepsilon}$. We could try to use their estimated values, say $\tilde{\bar{\varepsilon}}$ and $\tilde{\Sigma}_{\varepsilon\varepsilon}$, to construct the test statistic $T\tilde{\bar{\varepsilon}}'\tilde{\Sigma}_{\varepsilon\varepsilon}^{-1}\tilde{\bar{\varepsilon}}$. However, this does not work; this test statistic does not have the same asymptotic distribution (χ^2_{n-1}) as the test based on $\bar{\varepsilon}$ and $\Sigma_{\varepsilon\varepsilon}$. The reason is that the difference between $\tilde{\bar{\varepsilon}}$ and $\bar{\varepsilon}$ does not go to zero any faster than $\bar{\varepsilon}$ does. Explicitly,

$$\tilde{\bar{\varepsilon}}_i = \bar{\varepsilon}_i + (\ln\tilde{\alpha}_i - \ln\alpha_i) - (\ln\tilde{\alpha}_1 - \ln\alpha_1)$$

so that *plim* $\sqrt{T}(\tilde{\bar{\varepsilon}} - \bar{\varepsilon}) \neq 0$. A test can be based on the value of $\tilde{\bar{\varepsilon}}$, and indeed it is clear from (4.4.5) that the LM test itself is based on $\tilde{\bar{\varepsilon}}$, but the correct covariance matrix is more complicated than just $\Sigma_{\varepsilon\varepsilon}^{-1}$.

4.5 Test of Independence between Technical Inefficiency and Allocative Inefficiency in SL III Model

4.5.1 SL III Model

The *SL* III model allows correlation between technical inefficiency and allocative inefficiency, i.e. correlation between u_t and the absolute value of ε_t. This can be formulated as follows:

$$\ln y_t = A + \sum_{i=1}^{n} \alpha_i \ln x_{it} + w_t, \quad t=1,\cdots,T. \qquad (\textit{Note } 5)$$

$$\ln x_{1t} - \ln x_{it} = B_{it} + \varepsilon_{it}, \quad i=2,\cdots,n$$

$$\begin{bmatrix} u_t^* \\ \varepsilon_t \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ \xi \end{bmatrix}, \Sigma\right)$$

where

$$w_t = v_t - u_t$$

$$B_{it} = \ln\left(\frac{p_{it}\alpha_1}{p_{1t}\alpha_i}\right)$$

$$\varepsilon_t = (\varepsilon_{2t},\cdots,\varepsilon_{nt})'$$

$$\xi = (\xi_2,\cdots,\xi_n)'$$

$$u_t = |u_t^*|$$

$$\Sigma = \begin{bmatrix} \sigma_u^2 & \Sigma_{u\varepsilon} \\ \Sigma_{u\varepsilon}' & \Sigma_{\varepsilon\varepsilon} \end{bmatrix}$$

$$\Sigma_{u\varepsilon} = (\sigma_{u2},\cdots,\sigma_{un})$$

$$\Sigma_{\varepsilon\varepsilon} = \begin{bmatrix} \sigma_{22} & \cdots & \sigma_{2n} \\ \vdots & & \vdots \\ \sigma_{n2} & \cdots & \sigma_{nn} \end{bmatrix}$$

v_t *is independent of* u_t^* and ε_t,

v_t *i.i.d.* $N(0, \sigma_v^2)$

Note that $cov(u_t, \varepsilon_{it}) = 0$ but that

$$cov(u_t, |\varepsilon_{it}|) = (2\sigma_u\sqrt{\sigma_{ii}}/\pi)[\sqrt{1-\rho_i^2}+\rho_i \arcsin(\rho_i)-1]$$
$$> 0, \text{ if } \rho_i \neq 0$$

where ρ_i is the correlation of u_t^* and ε_{it}, $-1<\rho_i<1$. (See Schmidt and Lovell (1980) p.87.) That is, u_t is positively correlated with $|\varepsilon_{it}|$ as long as $\rho_i \neq 0$, or equivalently, as long as $\sigma_{ui} \neq 0$. Thus, the *SL* III model allows a non-zero correlation between technical and allocative inefficiency. When $\Sigma_{u\varepsilon} = 0$, the *SL* III model reduces to the *SL* II model.

Since the *SL* II model is easier to estimate than the *SL* III model, it seems that we might use the *LM* test to test the null hypothesis H_0: $\Sigma_{u\varepsilon} = 0$. But as we will see in Section 4.5.2, the *LM* test fails again. An alternative test is suggested in section 4.5.3.

4.5.2 The LM Test

The log-likelihood function for *SL* III model is

$$L = T\ln r + \sum_{t=1}^{T}\ln(f_{1t} + f_{2t}) \tag{4.5.1}$$

where

$$r = \sum_{i=1}^{n}\alpha_i$$

$$f_{jt} = [1-\Phi(c_{jt})](2\pi)^{-\frac{n}{2}}|G_1|^{-\frac{1}{2}}\exp(D_{jt})$$

$$c_{jt} = \frac{w_t+(-1)^{j+1}\sigma_v^2\Sigma^{u\varepsilon}(\varepsilon_t-\xi)}{\sigma_v}\sqrt{\frac{|\Sigma|}{|G_1|}}$$

$$D_{jt} = -\frac{1}{2}[w_t, (\varepsilon_t-\xi)']G_j^{-1}\begin{bmatrix} w_t \\ \varepsilon_t-\xi \end{bmatrix}$$

$$G_j = \begin{bmatrix} \sigma_u^2+\sigma_v^2 & (-1)^{j+1}\Sigma_{u\varepsilon} \\ (-1)^{j+1}\Sigma_{u\varepsilon}' & \Sigma_{\varepsilon\varepsilon} \end{bmatrix}$$

$$G_j^{-1} = \begin{bmatrix} G^{NW} & (-1)^{j+1}G^{NE} \\ (-1)^{j+1}G^{SW} & G^{SE} \end{bmatrix}$$

$j=1,2$

and

$$\Sigma^{u\varepsilon} = -\frac{1}{R_\Sigma}\Sigma_{u\varepsilon}\Sigma_{\varepsilon\varepsilon}^{-1}$$

$$R_\Sigma = \sigma_u^2 - \Sigma_{u\varepsilon}\Sigma_{\varepsilon\varepsilon}^{-1}\Sigma_{u\varepsilon}'$$

$$G^{NW} = \frac{1}{R_G}$$

$$G^{SE} = \Sigma_{\varepsilon\varepsilon}^{-1} + \Sigma_{\varepsilon\varepsilon}^{-1}\Sigma_{u\varepsilon}'\frac{1}{R_G}\Sigma_{u\varepsilon}\Sigma_{\varepsilon\varepsilon}^{-1}$$

$$R_G = (\sigma_u^2+\sigma_v^2) - \Sigma_{u\varepsilon}\Sigma_{\varepsilon\varepsilon}^{-1}\Sigma_{u\varepsilon}'$$

$$|\Sigma| = \sigma_u^2|\Sigma_{\varepsilon\varepsilon}| + \sum_{i=2}^{n}\sigma_{ui}\ cofactor(\sigma_{ui})$$

$$|G_1| = (\sigma_u^2+\sigma_v^2)|\Sigma_{\varepsilon\varepsilon}| + \sum_{i=2}^{n}\sigma_{ui}\ cofactor(\sigma_{ui})$$

Since the null hypothesis is $\Sigma_{u\varepsilon} = 0$, we need the first partial derivative of the log-likelihood function with respect to $\Sigma_{u\varepsilon}$ evaluated at the restricted *MLE* $\tilde{\vartheta} = (\tilde{\Sigma}_{u\varepsilon}, \tilde{A}, \tilde{\alpha}_i, \tilde{\sigma}_u^2, \tilde{\sigma}_v^2, \tilde{\Sigma}_{\varepsilon\varepsilon}, \tilde{\xi})$, where $\tilde{\Sigma}_{u\varepsilon} = 0$, and $\tilde{A}, \tilde{\alpha}_i, \tilde{\sigma}_u^2, \tilde{\sigma}_v^2, \tilde{\Sigma}_{\varepsilon\varepsilon}, \tilde{\xi}$ are the *MLS*'s from the *SL* II model. From eq.(4.5.1),

$$\frac{\partial L}{\partial \Sigma_{u\varepsilon}} = \begin{bmatrix} \frac{\partial L}{\partial \sigma_{u2}} \\ \cdot \\ \cdot \\ \cdot \\ \frac{\partial L}{\partial \sigma_{un}} \end{bmatrix}$$

where

$$\frac{\partial L}{\partial \sigma_{ui}} = \sum_{t=1}^{T}\frac{1}{f_{1t}+f_{2t}}\left[\frac{\partial f_{1t}}{\partial \sigma_{ui}} + \frac{\partial f_{2t}}{\partial \sigma_{ui}}\right] \tag{4.5.2}$$

$$\frac{\partial f_{jt}}{\partial \sigma_{ui}} = \left[-\varphi(c_{jt})\frac{\partial c_{jt}}{\partial \sigma_{ui}}\right](2\pi)^{-\frac{n}{2}}|G_1|^{-\frac{1}{2}}\exp(D_{jt}) \tag{4.5.3}$$

$$+ [1-\Phi(c_{jt})](2\pi)^{-\frac{n}{2}}\left[-\frac{1}{2}|G_1|^{-\frac{3}{2}}\frac{\partial |G_1|}{\partial \sigma_{ui}}\right]\exp(D_{jt}) + f_{jt}\frac{\partial D_{jt}}{\partial \sigma_{ui}},$$

$$\frac{\partial c_{jt}}{\partial \sigma_{ui}} = (-1)^{j+1} \frac{\partial \Sigma^{ut}}{\partial \sigma_{ui}} (\varepsilon_t - \xi) \sqrt{\frac{|\Sigma|}{|G_1|}}$$

$$+ \frac{w_t + (-1)^{j+1} \sigma_v^2 \Sigma^{u\varepsilon} (\varepsilon_t - \xi)}{2\sigma_v} \sqrt{\frac{|G_1|}{|\Sigma|}} \frac{|G_1| \frac{\partial |\Sigma|}{\partial \sigma_{ui}} - |\Sigma| \frac{\partial |G_1|}{\partial \sigma_{ui}}}{|G_1|^2}$$

$$\frac{\partial D_{jt}}{\partial \sigma_{ui}} = -\frac{1}{2} \frac{\partial [w_t G^{NW} w_t + (-1)^{j+1} (\varepsilon_t - \xi)' G^{SW} w_t + (-1)^{j+1} w_t G^{NE} (\varepsilon_t - \xi)]}{\partial \sigma_{ui}}$$
$$- \frac{1}{2} \frac{\partial [(\varepsilon_t - \xi)' G^{SE} (\varepsilon_t - \xi)]}{\partial \sigma_{ui}}$$

$$\frac{\partial |\Sigma|}{\partial \sigma_{ui}} = \frac{\partial |G_1|}{\partial \sigma_{ui}} = cofactor(\sigma_{ui})$$

$$\frac{\partial G^{NW}}{\partial \sigma_{ui}} = \frac{1}{R_G^2} \frac{\partial (\Sigma_{u\varepsilon} \Sigma_{\varepsilon\varepsilon}^{-1} \Sigma'_{u\varepsilon})}{\partial \sigma_{ui}}$$

$$\frac{\partial G^{SE}}{\partial \sigma_{ui}} = \frac{R_G \frac{\partial [\Sigma_{\varepsilon\varepsilon}^{-1} \Sigma'_{u\varepsilon} \Sigma_{u\varepsilon} \Sigma_{\varepsilon\varepsilon}^{-1}]}{\partial \sigma_{ui}} - \Sigma_{\varepsilon\varepsilon}^{-1} \Sigma'_{u\varepsilon} \Sigma_{u\varepsilon} \Sigma_{\varepsilon\varepsilon}^{-1} \frac{\partial R_G}{\partial \sigma_{ui}}}{R_G^2}$$

Note that if $\widetilde{\Sigma}_{u\varepsilon} = 0$ then $c_{1t} = c_{2t}$, $D_{1t} = D_{2t}$, $f_{1t} = f_{2t}$ and $\frac{\partial |\Sigma|}{\partial \sigma_{ui}} = \frac{\partial |G_1|}{\partial \sigma_{ui}} = 0$ since cofactor (σ_{ui}) has $\Sigma_{u\varepsilon}$ as its first column. Thus, when $\widetilde{\Sigma}_{u\varepsilon} = 0$, from eq.(4.5.2),

$$\frac{\partial f_{1t}}{\partial \sigma_{ui}} + \frac{\partial f_{2t}}{\partial \sigma_{ui}} = -\varphi(c_{1t})(2\pi)^{-\frac{n}{2}} |G_1|^{-\frac{1}{2}} \exp(D_{1t}) \cdot \left[\frac{\partial c_{1t}}{\partial \sigma_{ui}} + \frac{\partial c_{2t}}{\partial \sigma_{ui}}\right]_{\widetilde{\Sigma}_{u\varepsilon}=0} \qquad (4.5.4)$$
$$+ f_{1t} \cdot \left[\frac{\partial D_{1t}}{\partial \sigma_{ui}} + \frac{\partial D_{2t}}{\partial \sigma_{ui}}\right]_{\widetilde{\Sigma}_{u\varepsilon}=0}$$
$$= 0,$$

since

$$\left[\frac{\partial c_{1t}}{\partial \sigma_{ui}} + \frac{\partial c_{2t}}{\partial \sigma_{ui}}\right]_{\widetilde{\Sigma}_{u\varepsilon}=0} = (-1+1) \frac{\partial \Sigma^{u\varepsilon}}{\partial \sigma_{ui}} (\varepsilon_t - \xi) \sqrt{\frac{|\Sigma|}{|G_1|}} = 0$$

and

$$\left[\frac{\partial D_{1t}}{\partial \sigma_{ui}}+\frac{\partial D_{2t}}{\partial \sigma_{ui}}\right]_{\widetilde{\Sigma}_{u\varepsilon}=0} = \left[-\frac{\partial[w_t G^{NW} w_t+(\varepsilon_t-\xi)' G^{SE}(\varepsilon_t-\xi)]}{\partial \sigma_{ui}}\right]_{\widetilde{\Sigma}_{u\varepsilon}=0} = 0,$$

$$(because \left[\frac{\partial G^{NW}}{\partial \sigma_{ui}}\right]_{\widetilde{\Sigma}_{u\varepsilon}=0} = \left[\frac{\partial G^{SE}}{\partial \sigma_{ui}}\right]_{\widetilde{\Sigma}_{u\varepsilon}=0} = 0)$$

Substituting eq.(4.5.4) into eq.(4.5.2), we have

$$\left[\frac{\partial L}{\partial \sigma_{ui}}\right]_{\widetilde{\Sigma}_{u\varepsilon}=0} = \sum_{t=1}^{T} \frac{1}{2f_{1t}}\left[\frac{\partial f_{1t}}{\partial \sigma_{ui}}+\frac{\partial f_{2t}}{\partial \sigma_{ui}}\right]_{\widetilde{\Sigma}_{u\varepsilon}=0} = 0 \qquad (4.5.5)$$

Eq.(4.5.5) holds not just at the restricted *MLE* $\widetilde{\vartheta}$, but everywhere when $\Sigma_{u\varepsilon} = 0$ is imposed. Therefore, the *LM* test fails in this case.

4.5.3 Alternative Tests

Since v_t is independent of ε_{it} and u_t^*, we can test the correlation between u_t and $|\varepsilon_t|$ by seeing whether $w_t (\equiv v_t - u_t)$ correlates with $|\varepsilon_{it}|$. If we observed w_t and ε_t, we could calculate the correlation of w_t with $|\varepsilon_{it}|$, i=2, $\cdots$, n and do $(n-1)$ univariate t tests, or regress w_t on a constant term plus $|\varepsilon_{2t}|, \cdots, |\varepsilon_{nt}|$ and do an F test. Since w_t and $|\varepsilon_{it}|$ are non-normal, these tests would hold only asymptotically.

Because w_t and $|\varepsilon_{it}|$ are not observed, they are replaced by the estimates $\widetilde{w}_t$ and $|\widetilde{\varepsilon}_{it}|$. The above tests still hold asymptotically since the estimation error part of $\widetilde{w}_t$ and $\widetilde{\varepsilon}_{it}$ goes to zero as $T \to \infty$ while the w_t and ε_{it} part does not.

4.6 Conclusions

There are several extensions of the basic *ALS* model. Stevenson's non-zero mode for technical inefficiency is discussed in Section 4.2. The *SL* I model which allows allocative inefficiency in addition to technical inefficiency is discussed in Section 4.3. The *SL* II model which allows systematic allocative inefficiency is discussed in Section 4.4. The *SL* III model which allows correlation between technical and allocative inefficiency is discussed in Section 4.5.

The *SL* III model reduces to the *SL* II model when $\Sigma_{u\varepsilon} = 0$; the *SL* II model reduces to the *SL* I model when $\xi = 0$; and the *SL* I model reduces to the *ALS* model when $\Sigma_{\varepsilon\varepsilon} = 0$. Also, Stevenson's model reduces to the ALS model when $\mu = 0$. Since the estimation for the more general models is more complicated, it is reasonable to test the simpler models based on the estimates which impose these restrictions. However, the *LM* test fails in testing $\Sigma_{\varepsilon\varepsilon}$ in Section 4.3. A simple test based on the ratio of factor shares is suggested. The *LM* test fails again in testing $\Sigma_{u\varepsilon} = 0$ in Section 4.5. Tests based on the correlation between w_t and $|\varepsilon_t|$ are suggested.

CHAPTER V

SUMMARY AND CONCLUSIONS

There are three kinds of tests for model specification -- the Wald test, the likelihood ratio test and the Lagrange multiplier test. They have the same asymptotic power. Therefore, the choice among them depends on computational convenience. Since the *LM* test is based on the restricted estimates, we choose the *LM* test when estimation is easier in the restricted model than in the unrestricted model.

In Chapter 2, the LM test is applied to distributed lag models to test different alternative specifications. The test of the geometric lag specification against the alternative of no lag is discussed in Section 2.2. The corresponding LM test statistic is equivalent to the square of the t statistic for the coefficient of the lagged independent variable when it is added to the restricted model. In Section 2.3, the rational lag specification is tested against two simpler alternative specifications. The results are similar -- the LM test is essentially the F test of significance when lagged independent variables are added to the restricted model. For example, comparing eq.(2.2.4) and eq.(2.3.10), the *LM* statistics are similar (note that M_1 in eq.(2.2.4) is the same as M_4 in eq.(2.3.10)) because the geometric lag model is a special case of the rational lag model and both models have the same alternative specification (no lag). Note that the *LM* test statistics in Chapter 2 can be constructed from *OLS* residuals without running an *MLE* procedure, since the estimation of the restricted model requires only *OLS*.

In Chapter 3, the *LM* test is applied to qualitative and limited dependent variable models. Since the estimation of the restricted models can not utilize the *OLS* method, the *LM* statistic can not be expected to be constructed from

OLS results. The simplest example is in testing for sample selection bias in Heckman's sample selection bias model (Section 3.3). The restricted model can be estimated first by probit analysis, and then by *OLS*. Therefore, the corresponding *LM* test statistic can be constructed from the result of this two stage estimation procedure. Actually, this *LM* test statistic is equivalent to the square of the t statistic for the coefficient of the probit estimate for the inverse of the Mill's ratio when it is added to the regression. This means the simple "λ test" proposed by Heckman is asymptotically the same as the *LM* test. Therefore, the *LM* test gives the justification for using the λ test in large samples. The *LM* tests in the other two examples (Section 3.2 and Section 3.4) do not have such clear interpretations. In the test of the Tobit specification against Cragg's generalization of the Tobit model (Section 3.2), the *LM* statistic itself is not very complicated, but not much can be said about it except that it is indeed based on the Tobit residuals. In the test of independence in Poirier's partial observability model (Section 3.4), the *LM* statistic is equivalent to the explained sum of squares in a regression of residuals on a set of regressors. The regressors are related to the terms in the information matrix, but otherwise have no clear interpretation.

In Chapter 4, the *LM* test is applied to stochastic production and cost frontiers. A basic model proposed by Aigner, Lovell and Schmidt (*ALS* model) is presented in Section 4.1. There are several extensions of the *ALS* model in the literature. Stevenson considers a non-zero mode of technical inefficiency, while Schmidt and Lovell consider the possibility of allocative inefficiency (*SL* I model), of systematic allocative inefficiency (*SL* II model), and of correlation between technical and allocative inefficiency (*SL* III model). Stevenson's model reduces to the *ALS* model when the mode of technical inefficiency is zero, and the *LM* test of zero mode is presented in Section 4.2. The *SL* I model reduces to the *ALS* model when the variance-covariance matrix of the allocative inefficiency errors is zero. In Section 4.3, the *SL* I model is presented and the *LM* test is shown to fail in testing the zero variance-covariance matrix of the allocative inefficiency errors. A simple test based on the ratios of factor shares is suggested in Section 4.3.3. The *SL* II model reduces to the *SL* I model when the mean of allocative inefficiency is zero. The *LM* test of zero means is presented in Section 4.4. The *SL* III model reduces to the *SL* II model when technical and allocative inefficiency are independent. In Section 4.5, the *LM* test is shown to fail in testing the independence of technical and allocative inefficiency. Alternative tests based on the restricted estimates are suggested in Section 4.5.3.

In summary, the *LM* test, except in a few cases, can be used to test the adequency of the simple models which we discuss. Since the simpler model usually involves a simpler estimation method or less computational cost than the more complicated alternative, the *LM* test can be useful.

Note 1:

If the geometric lag model is

$$y_t = \beta w_t + \gamma z_t + \varepsilon_t, \textit{ where } w_t = \sum_{i=0}^{t-1} \lambda^i x_{t-i}, \quad t=1, \cdots, T$$

then the LM test statistic equals

$$\frac{\left[(X'_{t-1}M_2X_{t-1})^{-1}X'_{t-1}M_2Y_t\right]^2}{\tilde{\sigma}^2(X'_{t-1}M_2X_{t-1})}$$

which is the square of the t statistic for the coefficient of X_{t-1} in the regression of Y_t on (X_t, z_t, X_{t-1}). Here $M_2 = I - (X_t, z_t)[(X_t, z_t)'(X_t, z_t)]^{-1}(X_t, z_t)'$ and $\tilde{\sigma}^2 = \frac{\sum_{t=1}^{T}\tilde{\varepsilon}_t^2}{T}$ with $\tilde{\varepsilon}_t = y_t - \tilde{\beta}X_t - \tilde{\gamma}z_t$, and $\tilde{\beta}, \tilde{\gamma}$ are the OLS estimates.

Note 2:

Godfrey (1978) applies the LM test on an autoregressive model (AR model) and has a similar result. After all, both the AR model and the rational lag model are special cases of the $ARMAX$ model (see Nicholls, Pagan, and Terrell (1975)).

Note 3:

We can use eq.(1.5) directly, but the resulting LM statistic looks messy.

Note 4:

The same conclusion had been derived independently by Angelo Melino of NBER.

His paper will appear in *Review of Economic Studies*.

Note 5:

y_t is exogenous, while x_{it}'s are endogenous.

APPENDICES

APPENDIX A

INFORMATION MATRIX FOR THE GEOMETRIC MODEL

The log-likelihood function for the geometric lag model is

$$L = constant - \frac{T}{2}\log\sigma^2 - \frac{1}{2}\sigma^2\sum_{t=1}^{T}(y_t - \beta w_t - \eta_0\lambda^t)^2.$$

The first partial derivatives are

$$\frac{\partial L}{\partial \lambda} = \frac{1}{\sigma^2}\sum_{t=1}^{T}(\beta R_t + t\eta_0\lambda^{t-1})\varepsilon_t$$

$$\frac{\partial L}{\partial \beta} = \frac{1}{\sigma^2}\sum_{t=1}^{T} w_t\varepsilon_t$$

$$\frac{\partial L}{\partial \eta_0} = \frac{1}{\sigma^2}\sum_{t=1}^{T}\lambda^t\varepsilon_t$$

$$\frac{\partial L}{\partial \sigma^2} = -\frac{T}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{t=1}^{T}\varepsilon_t^2 ,$$

where $R_t = \dfrac{dw_t}{d\lambda} = \sum_{i=1}^{t-1} i\lambda^{i-1}x_{t-i}$.

The second partial derivatives are

$$\frac{\partial^2 L}{\partial \lambda^2} = -\frac{1}{\sigma^2}\sum_{t=1}^{T}\left\{(\beta R_t + t\eta_0\lambda^{t-1})^2 + \varepsilon_t\cdot something\right\}$$

$$\frac{\partial^2 L}{\partial \beta^2} = -\frac{1}{\sigma^2}\sum_{t=1}^{T} w_t^2$$

$$\frac{\partial^2 L}{\partial \eta_0^2} = -\frac{1}{\sigma^2}\sum_{t=1}^{T}\lambda^{2t}$$

$$\frac{\partial^2 L}{\partial (\sigma^2)^2} = \frac{T}{2\sigma^4} - \frac{1}{\sigma^6}\sum_{t=1}^{T}\varepsilon_t^2$$

$$\frac{\partial^2 L}{\partial \lambda \partial \beta} = -\frac{1}{\sigma^2}\sum_{t=1}^{T}\left\{w_t(\beta R_t + t\eta_0\lambda^{t-1}) - R_t\varepsilon_t\right\}$$

$$\frac{\partial^2 L}{\partial \lambda \partial \eta_0} = -\frac{1}{\sigma^2}\sum_{t=1}^{T}\left\{(\beta R_t + t\eta_0\lambda^{t-1})\lambda^t - t\lambda^{t-1}\varepsilon_t\right\}$$

$$\frac{\partial^2 L}{\partial \lambda \partial \sigma^2} = -\frac{1}{\sigma^4}\sum_{t=1}^{T}(\beta R_t + t\eta_0\lambda^{t-1})\varepsilon_t$$

$$\frac{\partial^2 L}{\partial\beta\partial\eta_0} = -\frac{1}{\sigma^2}\sum_{t=1}^{T} w_t \lambda^t$$

$$\frac{\partial^2 L}{\partial\beta\partial\sigma^2} = -\frac{1}{\sigma^4}\sum_{t=1}^{T} w_t \varepsilon_t$$

$$\frac{\partial^2 L}{\partial\eta_0\partial\sigma^2} = -\frac{1}{\sigma^4}\sum_{t=1}^{T} \lambda^t \varepsilon_t$$

The elements of information matrix are

$$I_{\lambda\lambda} = -E\left[\frac{\partial^2 L}{\partial\lambda^2}\right] = \frac{1}{\sigma^2}\sum_{t=1}^{T}(\beta R_t + t\eta_0\lambda^{t-1})^2$$

$$I_{\beta\beta} = -E\left[\frac{\partial^2 L}{\partial\beta^2}\right] = \frac{1}{\sigma^2}\sum_{t=1}^{T} w_t^2$$

$$I_{\eta_0\eta_0} = -E\left[\frac{\partial^2 L}{\partial\eta_0}\right] = -\frac{1}{\sigma^2}\sum_{t=1}^{T}\lambda^{2t}$$

$$I_{\sigma^2\sigma^2} = -E\left[\frac{\partial^2 L}{\partial(\sigma^2)^2}\right] = -\frac{T}{2\sigma^4} + \frac{T}{\sigma^4} = \frac{T}{2\sigma^4}$$

$$I_{\lambda\beta} = -E\left[\frac{\partial^2 L}{\partial\lambda\partial\beta}\right] = \frac{1}{\sigma^2}\sum_{t=1}^{T}\left\{w_t(\beta R_t + t\eta_0\lambda^{t-1})\right\}$$

$$I_{\lambda\eta_0} = -E\left[\frac{\partial^2 L}{\partial\lambda\partial\eta_0}\right] = \frac{1}{\sigma^2}\sum_{t=1}^{T}(\beta R_t + t\eta_0\lambda^{t-1})\lambda^t$$

$$I_{\lambda\sigma^2} = -E\left[\frac{\partial^2 L}{\partial\lambda\partial\sigma^2}\right] = 0$$

$$I_{\beta\eta_0} = -E\left[\frac{\partial^2 L}{\partial\beta\partial\eta_0}\right] = \frac{1}{\sigma^2}\sum_{t=1}^{T} w_t\lambda^t$$

$$I_{\beta\sigma^2} = -E\left[\frac{\partial^2 L}{\partial\beta\partial\sigma^2}\right] = 0$$

$$I_{\eta_0\sigma^2} = -E\left[\frac{\partial^2 L}{\partial\eta_0\partial\sigma^2}\right] = 0$$

The information matrix is

$$I(\vartheta) = \begin{bmatrix} I_{\lambda\lambda} & I_{\lambda\beta} & I_{\lambda\eta_0} & 0 \\ I_{\beta\lambda} & I_{\beta\beta} & I_{\beta\eta_0} & 0 \\ I_{\eta_0\lambda} & I_{\eta_0\beta} & I_{\eta_0\eta_0} & 0 \\ 0 & 0 & 0 & I_{\sigma^2\sigma^2} \end{bmatrix}$$

Under reasonable assumptions about the explanatory variables, $\frac{1}{T}I_{\eta_0\eta_0}$, $\frac{1}{T}I_{\lambda\eta_0}$ and $\frac{1}{T}I_{\beta\eta_0}$ all converge in probability to zero. Thus the information matrix is singular asymptotically, if η_0 is estimated as a parameter.

APPENDIX B

INVERSE OF A PARTITIONED MATRIX FOR THE RATIONAL LAG MODEL

Step 1

$$\begin{bmatrix} A & F & K \\ B & G & M \\ C & J & N \end{bmatrix} = \begin{bmatrix} P_1 & R_1 \\ R_1' & Q_1 \end{bmatrix}^{-1}$$

where

$$P_1 = \begin{bmatrix} \frac{1}{T}\tilde{I}_{a_*a_*} + I_\mu & \frac{1}{T}\tilde{I}_{a_*b} \\ \frac{1}{T}\tilde{I}_{ba_*} & \frac{1}{T}\tilde{I}_{bb} \end{bmatrix} = \begin{bmatrix} \frac{1}{T}\frac{1}{\tilde{\sigma}^2}X_*'X_* + I_\mu & -\frac{1}{T}\frac{\tilde{\alpha}_0}{\tilde{\sigma}^2}X_*'X_{**} \\ -\frac{1}{T}\frac{\tilde{\alpha}_0}{\tilde{\sigma}^2}X_{**}'X_* & \frac{1}{T}\frac{\tilde{\alpha}_0}{\tilde{\sigma}^2}X_{**}'X_{**} \end{bmatrix}$$

$$R_1 = \begin{bmatrix} \frac{1}{T}\tilde{I}_{a_*a_0} \\ \frac{1}{T}\tilde{I}_{ba_0} \end{bmatrix} = \begin{bmatrix} \frac{1}{T}\frac{1}{\tilde{\sigma}^2}X_*'X \\ -\frac{1}{T}\frac{\tilde{\alpha}_0}{\tilde{\sigma}^2}X_{**}'X \end{bmatrix}$$

$$Q_1 = \frac{1}{T}\tilde{I}_{a_0a_0} = \frac{1}{T}\frac{1}{\tilde{\sigma}^2}X'X$$

Step 2

$$\begin{bmatrix} A & F \\ B & G \end{bmatrix} = [P_1 - R_1 Q_1^{-1} R_1']^{-1}$$

$$= \begin{bmatrix} I_\mu + \frac{1}{T\tilde{\sigma}^2}X_*'M_4X_* & -\frac{\tilde{\alpha}_0}{T\tilde{\sigma}^2}X_*'M_4X_{**} \\ -\frac{\tilde{\alpha}_0}{T\tilde{\sigma}^2}X_{**}'M_4X_* & \frac{\tilde{\alpha}_0^2}{T\tilde{\sigma}^2}X_{**}'M_4X_{**} \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} P_2 & R_2 \\ R_2' & Q_2 \end{bmatrix}^{-1}$$

Hence,

$$A = (P_2 - R_2 Q_2^{-1} R_2')^{-1}$$

$$= \left\{ I_\mu + \frac{1}{T\tilde{\sigma}^2} X_\bullet' M_4 X_\bullet - \frac{1}{T\tilde{\sigma}^2} X_\bullet' M_\bullet X_{\bullet\bullet} (X_{\bullet\bullet}' M_4 X_{\bullet\bullet})^{-1} X_{\bullet\bullet}' M_4 X_\bullet \right\}^{-1}$$

$$= I_\mu$$

Since

$$X_\bullet' M_4 X_\bullet - X_\bullet' M_4 X_{\bullet\bullet} (X_{\bullet\bullet}' M_4 X_{\bullet\bullet})^{-1} X_{\bullet\bullet}' M_4 X_\bullet = \underline{X}_\bullet' [I - \underline{X}_{\bullet\bullet} (\underline{X}_{\bullet\bullet}' \underline{X}_{\bullet\bullet})^{-1} \underline{X}_{\bullet\bullet}] \underline{X}_\bullet$$
$$= 0$$

where $\underline{X}_\bullet = M_4 X_\bullet$, $\underline{X}_{\bullet\bullet} = M_4 X_{\bullet\bullet}$. The second equality holds because $\underline{X}_\bullet$ is contained in $\underline{X}_{\bullet\bullet}$.

$$B = -Q_2^{-1} R_2' (P_2 - R_2 Q_2^{-1} R_2')^{-1}$$
$$= -Q_2^{-1} R_2' A$$
$$= \frac{1}{\tilde{\alpha}_0} (X_{\bullet\bullet}' M_4 X_{\bullet\bullet})^{-1} X_{\bullet\bullet}' M_4 X_\bullet$$
$$= \frac{1}{\tilde{\alpha}_0} \begin{bmatrix} I_\mu \\ 0 \end{bmatrix}$$

$$F = -(P_2 - R_2 Q_2^{-1} R_2')^{-1} R_2 Q_2^{-1}$$
$$= -A R_2 Q_2^{-1}$$
$$= \frac{1}{\tilde{\alpha}_0} [I_\mu, 0]$$

$$G = Q_2^{-1} + Q_2^{-1} R_2' (P_2 - R_2 Q_2^{-1} R_2')^{-1} R_2 Q_2^{-1}$$
$$= \left[\frac{\tilde{\alpha}_0^2}{T\tilde{\sigma}^2} X_{\bullet\bullet}' M_4 X_{\bullet\bullet} \right]^{-1} + (\tilde{\alpha}_0^2)^{-1} (X_{\bullet\bullet}' M_4 X_{\bullet\bullet})^{-1} (X_{\bullet\bullet}' M_4 X_\bullet)(X_\bullet' M_4 X_{\bullet\bullet})(X_{\bullet\bullet}' M_4 X_{\bullet\bullet})^{-1}$$
$$= \frac{1}{\tilde{\alpha}_0^2} \left\{ T\tilde{\sigma}^2 (X_{\bullet\bullet}' M_4 X_{\bullet\bullet})^{-1} + \begin{bmatrix} I_\mu \\ 0 \end{bmatrix} [I_\mu, 0] \right\}$$

APPENDIX C

SOME EXPECTATIONS FOR CRAGG'S EXTENSION OF THE TOBIT MODEL

$$
\begin{aligned}
E(d_t) &= 1 \cdot Prob\,(d_t = 1) + 0 \cdot Prob\,(d_t = 0) \\
&= Prob\,(y_t > 0) \\
&= \Phi[x_t(\xi + \beta)]
\end{aligned}
$$

$$
\begin{aligned}
E(d_t y_t) &= E(y_t \mid y_t > 0) \cdot Prob\,(y_t > 0) \\
&= \left[x_t\beta_2 + m\left(\frac{x_t\beta_2}{\sigma_2}\right)\cdot\sigma_2\right]\cdot\Phi(x_t\beta_1) \\
&= \frac{1}{h}[x_t\beta + m(x_t\beta)]\cdot\Phi[x_t(\xi + \beta)] \qquad by\ eq.\,(3.2.9)
\end{aligned}
$$

$$
\begin{aligned}
E(d_t y_t^2) &= E(y_t^2 \mid y_t > 0)\cdot Prob\,(y_t > 0) \\
&= Var\,(y_t \mid y_t > 0) + [E(y_t \mid y_t > 0)]^2 \\
&= \left[1 - \frac{x_t\beta_2}{\sigma_2} m\left(\frac{x_t\beta_2}{\sigma_2}\right) - m^2\left(\frac{x_t\beta_2}{\sigma_2}\right)\right]\sigma_2^2 + \left\{(x_t\beta_2)^2 + 2(x_t\beta_2)\sigma_2\cdot m\left(\frac{x_t\beta_2}{\sigma_2}\right) + \sigma_2^2\cdot m^2\left(\frac{x_t\beta_2}{\sigma_2}\right)\right\} \\
&= \sigma_2^2 + (x_t\beta_2)^2 + \sigma_2\cdot(x_t\beta_2)\cdot m\left(\frac{x_t\beta_2}{\sigma_2}\right) \\
&= \frac{1}{h^2}[1 + (x_t\beta)^2 + x_t\beta\cdot m(x_t\beta)]\cdot\Phi[x_t(\xi + \beta)] \qquad by\ eq.\,(3.2.9)
\end{aligned}
$$

$$
\begin{aligned}
E(hy_t - x_t\beta \mid y_t > 0) &= E\left(\frac{y_t - x_t\beta_2}{\sigma_2} \mid y_t > 0\right) \qquad \text{(C-1)} \\
&= m\left(\frac{x_t\beta_2}{\sigma_2}\right),
\end{aligned}
$$

since

$$
E(y_t \mid y_t > 0) = x_t\beta_2 + m\left(\frac{x_t\beta_2}{\sigma_2}\right)\cdot\sigma_2.
$$

APPENDIX D

INFORMATION MATRIX FOR THE SAMPLE SELECTION BIAS MODEL

The log-likelihood function is

$$L = \sum_{i=1}^{n} d_i \ln F_i + (1 - d_i) \ln G_i$$

The first partial derivatives are

$$\frac{\partial L}{\partial \rho} = \sum_{i=1}^{n} d_i \left[\frac{\rho}{1-\rho^2} - \frac{\rho}{(1-\rho^2)^2} \frac{(y_{1i} - x_{1i}\beta_1)^2}{\sigma_1^2} + \frac{(1+\rho^2)}{(1-\rho^2)^2} \frac{(y_{1i} - x_{1i}\beta_1)}{\sigma_1} \frac{A_i}{F_i} - \frac{\rho}{(1-\rho^2)^2} \frac{B_i}{F_i} \right]$$

$$\frac{\partial L}{\partial \beta_1} = \sum_{i=1}^{n} d_i \left[\frac{(y_{1i} - x_{1i}\beta_1) x_{1i}'}{\sigma_1^2 (1-\rho^2)} - \frac{\rho}{1-\rho^2} \frac{x_{1i}'}{\sigma_1} \frac{A_i}{F_i} \right]$$

$$\frac{\partial L}{\partial \beta_2} = \sum_{i=1}^{n} \left[d_i x_{2i}' \frac{h_i^*}{F_i} - (1 - d_i) x_{2i}' m_i \right]$$

$$\frac{\partial L}{\partial \sigma_1} = \sum_{i=1}^{n} d_i \left[-\frac{1}{\sigma_1} + \frac{(y_{1i} - x_{1i}\beta_1)^2}{(1-\rho^2)\sigma_1^3} - \frac{\rho}{1-\rho^2} \frac{(y_{1i} - x_{1i}\beta_1)}{\sigma_1^2} \frac{A_i}{F_i} \right]$$

where

$$h_i^* = h(y_{1i} - x_{1i}\beta_1, \; -x_{2i}\beta_2),$$

$$A_i = \int_{-x_{2i}\beta_2}^{\infty} u_{2i} \cdot h(y_{1i} - x_{1i}\beta_1, \; u_{2i}) \, du_{2i},$$

$$B_i = \int_{-x_{2i}\beta_2}^{\infty} u_{2i}^2 \cdot h(y_{1i} - x_{1i}\beta_1, \; u_{2i}) \, du_{2i},$$

$$F_i = \int_{-x_{2i}\beta_2}^{\infty} h(y_{1i} - x_{1i}\beta_1, \; u_{2i}) \, du_{2i},$$

$$m_i = \frac{\varphi(-x_{2i}\beta_2)}{\Phi(-x_{2i}\beta_2)}.$$

When $\rho = 0$,

$$\frac{A_i}{F_i} = \frac{\int_{-x_{2i}\beta_2}^{\infty} u_{2i}\cdot\varphi(u_{2i})du_{2i}}{\int_{-x_{2i}\beta_2}^{\infty} \varphi(u_{2i})du_{2i}} = E(u_{2i}|u_{2i}>-x_{2i}\beta_2) = \lambda_i \tag{D-1}$$

Hence the first partial derivatives evaluated at ϑ are

$$\frac{\partial L}{\partial\tilde{\rho}} = \sum_{i=1}^{n} d_i \frac{y_{1i}-x_{1i}\tilde{\beta}_1}{\tilde{\sigma}_1}\lambda_i = \frac{\tilde{\lambda}' e_1}{\tilde{\sigma}_1} \quad where \ \tilde{\lambda}_i = \frac{\varphi(-x_{2i}\tilde{\beta}_2)}{1-\Phi(-x_{2i}\tilde{\beta}_2)},$$

$$\frac{\partial L}{\partial\tilde{\beta}_1} = \frac{\partial L}{\partial\tilde{\beta}_2} = \frac{\partial L}{\partial\tilde{\sigma}_1} = 0 \qquad by\ eq.(1.4)$$

The second partial derivatives are

$$\begin{aligned}\frac{\partial^2 L}{\partial\rho^2} = {} & \frac{1+\rho^2}{(1-\rho^2)^2}\sum_{i=1}^{n} d_i - \frac{(1+3\rho^2)}{(1-\rho^2)^3}\frac{1}{\sigma_1^2}\sum_{i=1}^{n} d_i(y_{1i}-x_{1i}\beta_1)^2 + \frac{2\rho(3+\rho^2)}{(1-\rho^2)^3}\frac{1}{\sigma_1}\sum_{i=1}^{n} d_i(y_{1i}-x_{1i}\beta_1)\frac{A_i}{F_i} \\ & + \frac{(1+\rho^2)^2}{(1-\rho^2)^4}\frac{1}{\sigma_1^2}\sum_{i=1}^{n} d_i(y_{1i}-x_{1i}\beta_1)^2\left[\frac{B_i}{F_i} - \frac{A_i^2}{F_i^2}\right] - \frac{\rho^2}{(1-\rho^2)^4}\sum_{i=1}^{n}\left[\frac{B_i^2}{F_i^2} - \frac{D_i}{F_i}\right] \\ & - \frac{(1+3\rho^2)}{(1-\rho^2)^3}\sum_{i=1}^{n} d_i\frac{B_i}{F_i} - \frac{2\rho(1+\rho^2)}{(1-\rho^2)^4}\frac{1}{\sigma_1}\sum_{i=1}^{n} d_i(y_{1i}-x_{1i}\beta_1)\left[\frac{C_i}{F_i} - \frac{B_iA_i}{F_i^2}\right]\end{aligned}$$

$$\begin{aligned}\frac{\partial^2 L}{\partial\rho\partial\beta_1} = \sum_{i=1}^{n} d_i \Bigg[& \frac{2\rho x'_{1i}(y_{1i}-x_{1i}\beta)}{(1-\rho^2)^2\sigma_1^2} - \frac{x'_{1i}\left[(1+\rho^2)\frac{A_i}{F_i} + \frac{\rho(1+\rho^2)}{(1-\rho^2)}\frac{(y_{1i}-x_{1i}\beta_1)}{\sigma_1}\left[\frac{B_i}{F_i} - \frac{A_i^2}{F_i^2}\right]\right]}{(1-\rho^2)^2\sigma_1} \\ & - \frac{x'_{1i}\left[\frac{\rho^2}{1-\rho^2}\left[\frac{C_i}{F_i} - \frac{A_iB_i}{F_i^2}\right]\right]}{(1-\rho^2)^2\sigma_1}\Bigg]\end{aligned}$$

$$\begin{aligned}\frac{\partial^2 L}{\partial\rho\partial\beta_2} = {} & -\frac{(1+\rho^2)}{(1-\rho^2)^2}\frac{1}{\sigma_1}\sum_{i=1}^{n} d_i x'_{2i}(y_{1i}-x_{1i}\beta_1)\frac{h_i^*}{F_i}\left[x_{2i}\beta_2 + \frac{A_i}{F_i}\right] \\ & - \frac{\rho}{(1-\rho^2)^2}\sum_{i=1}^{n} d_i x'_{2i}\frac{h_i^*}{F_i}\left[(x_{2i}\beta_2)^2 - \frac{B_i}{F_i}\right]\end{aligned}$$

$$\frac{\partial^2 L}{\partial\rho\partial\sigma_1} = \frac{2\rho}{(1-\rho^2)^2}\frac{1}{\sigma_1^3}\sum_{i=1}^{n} d_i(y_{1i}-x_{1i}\beta_1)^2 - \frac{(1+\rho^2)}{(1-\rho^2)^2}\frac{1}{\sigma_1^2}\sum_{i=1}^{n} d_i(y_{1i}-x_{1i}\beta_1)\frac{A_i}{F_i}$$

$$\frac{\partial^2 L}{\partial\beta_1\partial\beta_1'} = \sum_{i=1}^{n} d_i\left[\frac{-x_{1i}'x_{1i}}{\sigma_1^2(1-\rho^2)} - \frac{\rho^2}{(1-\rho^2)^2}\frac{x_{1i}'x_{1i}}{\sigma_1^2}\frac{A_i^2-B_iF_i}{F_i^2}\right]$$

$$\frac{\partial^2 L}{\partial\beta_1\partial\beta_2'} = -\sum_{i=1}^{n} d_i\frac{\rho}{1-\rho^2}\frac{x_{1i}'x_{2i}}{\sigma_1}\left[\frac{-x_{2i}\beta_2 h_i^*}{F_i} - \frac{A_i}{F_i}\frac{h_i^*}{F_i}\right]$$

$$\frac{\partial^2 L}{\partial\beta_1\partial\sigma_1} = \sum_{i=1}^{n} d_i\left\{\frac{-2(y_{1i}-x_{1i}\beta_1)x_{1i}'}{(1-\rho^2)\sigma_1^3} + \frac{\rho}{1-\rho^2}\frac{x_{1i}'}{\sigma_1^2}\left[\frac{A_i}{F_i} + \frac{\rho}{1-\rho^2}\frac{(y_{1i}-x_{1i}\beta_1)}{\sigma_1}\left[\frac{B_i}{F_i} - \frac{A_i^2}{F_i^2}\right]\right]\right\}$$

$$\frac{\partial^2 L}{\partial\beta_2\partial\beta_2'} = -\frac{\rho}{1-\rho^2}\frac{1}{\sigma_1}\sum_{i=1}^{n} d_i x_{2i}'x_{2i}(y_{1i}-x_{1i}\beta_1)\frac{h_i^*}{F_i} - \frac{1}{1-\rho^2}\sum_{i=1}^{n} d_i x_{2i}'x_{2i}(x_{2i}\beta_2)\frac{h_i^*}{F_i}$$
$$-\sum_{i=1}^{n} d_i x_{2i}'x_{2i}\frac{h_i^{*2}}{F_i^2} + \sum_{i=1}^{n}(1-d_i)x_{2i}'x_{2i}(x_{2i}\beta_2)m_i - \sum_{i=1}^{n}(1-d_i)x_{2i}'x_{2i}m_i^2$$

$$\frac{\partial^2 L}{\partial\beta_2\partial\sigma_1} = \frac{\rho}{1-\rho^2}\frac{1}{\sigma_1}\sum_{i=1}^{n} d_i x_{2i}'(y_{1i}-x_{1i}\beta_1)\frac{h_i^*}{F_i}\left[x_{2i}\beta_2 + \frac{A_i}{F_i}\right]$$

$$\frac{\partial^2 L}{\partial\sigma_1^2} = \frac{1}{\sigma_1^2}\sum_{i=1}^{n} d_i - \frac{3}{(1-\rho^2)\sigma_1^4}\sum_{i=1}^{n} d_i(y_{1i}-x_{1i}\beta_1)^2 - \frac{\rho^2}{(1-\rho^2)^2}\frac{1}{\sigma_1^4}\sum_{i=1}^{n} d_i(y_{1i}-x_{1i}\beta_1)^2\left[\frac{A_i^2}{F_i^2} - \frac{B_i}{F_i}\right]$$
$$+\frac{2\rho}{1-\rho^2}\frac{1}{\sigma_1^3}\sum_{i=1}^{n} d_i(y_{1i}-x_{1i}\beta_1)\frac{A_i}{F_i}$$

where

$$C_i = \int_{-x_{2i}\beta_2}^{\infty} u_{2i}^3\cdot h(y_{1i}-x_{1i}\beta_1,\, u_{2i})du_{2i}$$

$$D_i = \int_{-x_{2i}\beta_2}^{\infty} u_{2i}^4\cdot h(y_{1i}-x_{1i}\beta_1,\, u_{2i})du_{2i}$$

When $\rho = 0$,

$$\frac{A_i}{F_i} = \lambda_i \tag{D-1}$$

$$\frac{h_i^*}{F_i} = \frac{\varphi(-x_{2i}\beta_2)}{1 - \Phi(-x_{2i}\beta_2)} = \lambda_i \tag{D-2}$$

$$\frac{B_i}{F_i} = \frac{\int\limits_{-x_{2i}\beta_2}^{\infty} u_{2i}^2 \varphi(u_{2i}) du_{2i}}{\int\limits_{-x_{2i}\beta_2}^{\infty} \varphi(u_{2i}) du_{2i}} = E(u_{2i}^2 \mid u_{2i} > -x_{2i}\beta_2) \tag{D-3}$$

$$= [E(u_{2i} \mid u_{2i} > -x_{2i}\beta_2)]^2 + Var(u_{2i} \mid u_{2i} > -x_{2i}\beta_2)$$

$$= \lambda_i^2 + (1 + z_i\lambda_i - \lambda_i^2)$$

$$= 1 + z_i\lambda_i$$

where $z_i = -x_{2i}\beta_2$.

The elements of the information matrix evaluated at $\tilde{\vartheta}$ are

$$\tilde{I}_{\rho\rho} = -E\left[\frac{\partial^2 L}{\partial \rho^2}\right] \quad \textit{evaluated at } \tilde{\vartheta}$$

$$= -\sum_{i=1}^{n}(1-\tilde{G}_i) + \frac{1}{\tilde{\sigma}_1^2}\sum_{i=1}^{n}\tilde{\sigma}_1^2(1-\tilde{G}_i) - \frac{1}{\tilde{\sigma}_1^2}\sum_{i=1}^{n}\tilde{\sigma}_1^2(1-\tilde{G}_i)[(1+\tilde{z}_i\tilde{\lambda}_i)-\tilde{\lambda}_i^2] + \sum_{i=1}^{n}(1-\tilde{G}_i)(1+\tilde{z}_i\tilde{\lambda}_i)$$

$$= \sum_{i=1}^{n}(1-\tilde{G}_i)\tilde{\lambda}_i^2 \quad \textit{with } \tilde{z}_i = -x_{2i}\tilde{\beta}_2,$$

since eq.(D-1), (D-3), and

$$E(d_i) = 1 \cdot Prob(d_i=1) + 0 \cdot Prob(d_i=0) = 1 - \Phi(-x_{2i}\beta_2) = 1 - G_i \tag{D-4}$$

and

$$E(\sum_{i=1}^{n} d_i u_{1i}^2) = \sum_{i=1}^{n} E(u_{1i}^2 \mid d_i=1) \cdot Prob(d_i=1) \tag{D-5}$$

$$= \sum_{i=1}^{n}(1+\rho^2 z_i\lambda_i)\sigma_1^2(1-G_i).$$

$$\widetilde{I}_{\rho\beta_1} = \sum_{i=1}^{n}(1-\widetilde{G}_i)\frac{x_{1i}'\widetilde{\lambda}_i}{\widetilde{\sigma}_1} \qquad by\ eq.(D-1) \text{ and } eq.(D-4).$$

$\widetilde{I}_{\rho\beta_2} = 0$, since

$$E(d_i u_{1i}) = E(u_{1i}\,|\,d_i=1)\cdot Prob(d_i=1) \tag{D-6}$$
$$= \sigma_{12}\lambda_i\cdot Prob(d_i=1)$$
$$= \rho\sigma_1\lambda_i\cdot Prob(d_i=1)$$

$$\widetilde{I}_{\rho\sigma_1} = 0 \qquad by\ eq.(D-6).$$

$$\widetilde{I}_{\beta_1\beta_1} = \sum_{i=1}^{n}\frac{(1-\widetilde{G}_i)x_{1i}'x_{1i}}{\widetilde{\sigma}_1^2} \qquad by\ eq.(D-4).$$

$$\widetilde{I}_{\beta_1\beta_2} = 0$$

$$\widetilde{I}_{\beta_1\sigma_1} = 0 \qquad by\ eq.(D-6)$$

$$\widetilde{I}_{\beta_2\beta_2} = \sum_{i=1}^{n}\varphi(-x_{2i}\widetilde{\beta}_2)x_{2i}'x_{2i}(x_{2i}\widetilde{\beta}_2) + \sum_{i=1}^{n}\frac{[\varphi(-x_{2i}\widetilde{\beta}_2)]^2}{[1-\Phi(-x_{2i}\widetilde{\beta}_2)]}x_{2i}'x_{2i}$$
$$- \sum_{i=1}^{n}x_{2i}'x_{2i}(x_{2i}\widetilde{\beta}_2)\varphi(-x_{2i}\widetilde{\beta}_2) + \sum_{i=1}^{n}\frac{[\varphi(-x_{2i}\widetilde{\beta}_2)]^2}{\Phi(-x_{2i}\widetilde{\beta}_2)}x_{2i}'x_{2i}$$
$$= \sum_{i=1}^{n}\widetilde{\lambda}_i\widetilde{m}_i x_{2i}'x_{2i}, \qquad by\ eq.(D-2) \text{ and } eq.(D-4)$$

where $\widetilde{m}_i = \dfrac{\varphi(-x_{2i}\widetilde{\beta}_2)}{\Phi(-x_{2i}\widetilde{\beta}_2)}$

$$\widetilde{I}_{\beta_2\sigma_1} = 0$$

$$\widetilde{I}_{\sigma_1\sigma_1} = -\frac{1}{\widetilde{\sigma}_1^2}\sum_{i=1}^{n}(1-\widetilde{G}_i) + \frac{3}{\widetilde{\sigma}_1^4}\sum_{i=1}^{n}\widetilde{\sigma}_1^2(1-\widetilde{G}_i)$$
$$= \frac{2}{\widetilde{\sigma}_1^2}\sum_{i=1}^{n}(1-\widetilde{G}_i) \qquad by\ eq.(D-4) \text{ and } eq.(D-5)$$

APPENDIX E

INFORMATION MATRIX FOR SL II MODEL

From eq.(4.4.4), the first partial derivatives are

$$\frac{\partial L}{\partial \xi_l} = \sum_{t=1}^{T}\sum_{i=2}^{n}(\varepsilon_{it} - \xi_i)\sigma^{il},\ l = 2, \cdots, n \tag{E-1}$$

$$\frac{\partial L}{\partial \mu} = \sum_{t=1}^{T}\frac{m(a_t)}{\sigma\lambda} - \frac{Tm(b)}{\sigma}(\lambda^{-2} + +1)^{1/2} - \sum_{t=1}^{T}\left[\frac{w_t}{\sigma^2}\right]$$

$$\frac{\partial L}{\partial A} = \sum_{t=1}^{T}\left[m(a_t)\cdot\frac{\lambda}{\sigma} + \frac{w_t}{\sigma^2}\right]$$

$$\frac{\partial L}{\partial \lambda} = -\frac{1}{\sigma}\left[\sum_{t=1}^{T} m(a_t)\cdot w_t\right]$$

$$\frac{\partial L}{\partial \sigma} = -\frac{T}{\sigma} + \frac{1}{\sigma^2}\sum_{t=1}^{T} w_t\lambda\cdot m(a_t) + \frac{1}{\sigma^3}\sum_{t=1}^{T} w_t^2$$

$$\frac{\partial L}{\partial \sigma_{hk}} = -\frac{T}{2}\frac{\partial \ln|\Sigma_{\varepsilon\varepsilon}|}{\partial \sigma_{hk}} + \frac{1}{2}\sum_{t=1}^{T}\sum_{j=2}^{n}\sum_{i=2}^{n}(\varepsilon_{it} - \xi_i)(\varepsilon_{jt} - \xi_j)\sigma^{ih}\sigma^{kj}$$

where

$$\frac{\partial \ln|\Sigma_{\varepsilon\varepsilon}|}{\partial \sigma_{hk}} = \begin{cases}\sigma^{hh} & \text{if } h=k \\ 2\sigma^{hk} & \text{if } h\neq k\end{cases}$$

$$h = 2, \cdots, n, \quad k = 2, \cdots n$$

$$\frac{\partial L}{\partial \alpha_1} = \frac{T}{r} + \frac{\lambda}{\sigma}\sum_{t=1}^{T} m(a_t)\cdot \ln x_{1t} + \frac{1}{2\alpha_1}\sum_{t=1}^{T}\sum_{j=2}^{n}\sum_{i=2}^{n}(\varepsilon_{it} + \varepsilon_{jt} - \xi_i - \xi_j)\sigma^{ij}$$
$$+ \frac{1}{\sigma^2}\sum_{t=1}^{T} w_t \ln x_{1t}$$

$$\frac{\partial L}{\partial \alpha_h} = \frac{T}{r} + \frac{\lambda}{\sigma}\sum_{t=1}^{T} m(a_t)\cdot \ln x_{ht} - \frac{1}{\alpha_h}\sum_{t=1}^{T}\sum_{i=2}^{n}(\varepsilon_{it} - \xi_i)\sigma^{ih} + \frac{1}{\sigma^2}\sum_{t=1}^{T} w_t \ln x_{ht}$$

$h = 2, \ldots, n.$

where σ^{il} is the (i,l)th element of $\Sigma_{\varepsilon\varepsilon}^{-1}$, $m(\cdot) = \frac{\varphi(\cdot)}{1-\Phi(\cdot)}$.

The second partial derivatives are

$$\frac{\partial^2 L}{\partial \xi_l \partial \xi_m} = -T\sigma^{ml}, \quad l = 2, \cdots, n, \quad m = 2, \cdots, n$$

$$\frac{\partial^2 L}{\partial \xi_l \partial \mu} = 0 \quad l = 2, \cdots, n$$

$$\frac{\partial^2 L}{\partial \xi_l \partial A} = 0, \quad l = 2, \cdots, n$$

$$\frac{\partial^2 L}{\partial \xi_l \partial \lambda} = 0, \quad l = 2, \cdots, n$$

$$\frac{\partial^2 L}{\partial \xi_l \partial \sigma} = 0, \quad l = 2, \cdots, n$$

$$\frac{\partial^2 L}{\partial \xi_l \partial \sigma_{hk}} = -\sum_{t=1}^{T}\sum_{i=2}^{n}(\varepsilon_{it} - \xi_i)\sigma^{ih}\sigma^{kl}, \quad l, h, k = 2, \cdots, n$$

$$\frac{\partial^2 L}{\partial \xi_l \partial \alpha_1} = -\frac{T}{\alpha_1}\sum_{i=2}^{n}\sigma^{il}, \quad l = 2, \cdots, n$$

$$\frac{\partial^2 L}{\partial \xi_l \partial \alpha_h} = \frac{T}{\alpha_h}\sigma^{hl}, \quad l = 2, \cdots, n$$

$$\frac{\partial^2 L}{\partial A^2} = \frac{\lambda^2}{\sigma^2}\sum_{t=1}^{T} z(a_t) - \frac{T}{\sigma^2} \tag{E-2}$$

$$\frac{\partial^2 L}{\partial A \partial \lambda} = \frac{1}{\sigma}\sum_{t=1}^{T}\left\{m(a_t) - \frac{\lambda}{\sigma} w_t \cdot z(a_t)\right\} \tag{E-3}$$

$$\frac{\partial^2 L}{\partial A \partial \sigma} = \frac{\lambda}{\sigma^2}\sum_{t=1}^{T} \dot{Q}(a_t) - \frac{2}{\sigma^3}\sum_{t=1}^{T} w_t \tag{E-4}$$

$$\frac{\partial^2 L}{\partial A \partial \sigma_{ij}} = 0$$

$$\frac{\partial^2 L}{\partial A \partial \alpha_i} = \frac{\lambda^2}{\sigma^2}\sum_{t=1}^{T} z(a_t)\ln x_{it} - \frac{1}{\sigma^2}\sum_{t=1}^{T} \ln x_{it}, \quad i = 1, \cdots, n \tag{E-5}$$

$$\frac{\partial^2 L}{\partial \lambda^2} = \frac{1}{\sigma^2}\sum_{t=1}^{T}\left\{w_t^2 \cdot z(a_t)\right\} \tag{E-6}$$

$$\frac{\partial^2 L}{\partial \lambda \partial \sigma} = -\frac{1}{\sigma^2}\sum_{t=1}^{T} w_t Q(a_t) \tag{E-7}$$

$$\frac{\partial^2 L}{\partial \lambda \partial \sigma_{ij}} = 0$$

$$\frac{\partial^2 L}{\partial \lambda \partial \alpha_i} = -\frac{1}{\sigma}\sum_{t=1}^{T}\left\{\frac{\lambda}{\sigma} w_t z(a_t)\ln x_{it} - m(a_t)\ln x_{it}\right\} \tag{E-8}$$

$$\frac{\partial^2 L}{\partial \sigma^2} = \frac{T}{\sigma^2} + \frac{1}{\sigma^2}\sum_{t=1}^{T} a_t P(a_t) - \frac{T}{\sigma^2} b P(b) - \frac{3}{\sigma^4}\sum_{t=1}^{T} w_t^2 \tag{E-9}$$

$$\frac{\partial^2 L}{\partial \sigma \partial \sigma_{ij}} = 0$$

$$\frac{\partial^2 L}{\partial\sigma\partial\alpha_t} = \frac{1}{\sigma^2}\sum_{t=1}^{T}\lambda Q(a_t)\ln x_{it} - \frac{1}{\sigma^3}\sum_{t=1}^{T}2w_t\ln x_{it} \tag{E-10}$$

$$\frac{\partial^2 L}{\partial\sigma_{hh}\partial\sigma_{hh}} = \sum_{t=1}^{T}\sum_{j=2}^{n}\sum_{i=2}^{n}(\varepsilon_{it}-\xi_i)(\varepsilon_{jt}-\xi_j)\sigma^{ih}\sigma^{hh}\sigma^{hj} - \frac{T}{2}(\sigma^{hh})^2,\ h=2,\cdots,n \tag{E-11}$$

$$\frac{\partial^2 L}{\partial\sigma_{hh}\partial\sigma_{lm}} = -\frac{T}{2}\sigma^{hl}\sigma^{mh} + \frac{1}{2}\sum_{t=1}^{T}\sum_{j=2}^{n}\sum_{i=2}^{n}(\varepsilon_{it}-\xi_i)(\varepsilon_{jt}-\xi_j)(\sigma^{il}\sigma^{mh}\sigma^{hj}+\sigma^{ih}\sigma^{hl}\sigma^{mj}) \tag{E-12}$$

$$h,l,m = 2,\cdots,n$$

$$\frac{\partial^2 L}{\partial\sigma_{hk}\partial\sigma_{lm}} = -T\sigma^{hl}\sigma^{mk} + \frac{1}{2}\sum_{t=1}^{T}\sum_{j=2}^{n}\sum_{i=2}^{n}(\varepsilon_{it}-\xi_i)(\varepsilon_{jt}-\xi_j)(\sigma^{il}\sigma^{mh}\sigma^{kj}+\sigma^{ih}\sigma^{kl}\sigma^{mj})$$

$$h\neq k,\quad h,k,l,m = 2,\cdots,n$$

$$\frac{\partial^2 L}{\partial\sigma_{hk}\partial\alpha_1} = -\frac{1}{2\alpha_1}\sum_{t=1}^{T}\sum_{j=2}^{n}\sum_{i=2}^{n}\left[(\varepsilon_{it}-\xi_i) + (\varepsilon_{jt}-\xi_j)\right]\sigma^{ih}\sigma^{kj}$$

$$h,k = 2,\cdots,n$$

$$\frac{\partial^2 L}{\partial\sigma_{hk}\partial\alpha_l} = \frac{1}{2\alpha_l}\sum_{t=1}^{T}\left[\sum_{j=2}^{n}(\varepsilon_{jt}-\xi_j)\sigma^{lh}\sigma^{kj} + \sum_{i=2}^{n}(\varepsilon_{it}-\xi_i)\sigma^{ih}\sigma^{kl}\right]$$

$$h,k,l = 2,\cdots,n$$

$$\frac{\partial^2 L}{\partial\alpha_1^2} = -\frac{T}{r^2} + \frac{\lambda^2}{\sigma^2}\sum_{t=1}^{T}z(a_t)\cdot(\ln x_{1t})^2 - \frac{1}{\sigma^2}\sum_{t=1}^{T}(\ln x_{1t})^2 - \frac{1}{2\alpha_1^2}\sum_{t=1}^{T}\sum_{j=2}^{n}\sum_{i=2}^{n}\sigma^{ij}(2+\varepsilon_{it}+\varepsilon_{jt}-\xi_i-\xi_j) \tag{E-13}$$

$$\frac{\partial^2 L}{\partial\alpha_1\partial\alpha_h} = -\frac{T}{r^2} + \frac{\lambda^2}{\sigma^2}\sum_{t=1}^{T}z(a_t)(\ln x_{1t})(\ln x_{ht}) - \frac{1}{\sigma^2}\sum_{t=1}^{T}(\ln x_{1t})(\ln x_{ht}) + \frac{1}{\alpha_1\alpha_h}\sum_{t=1}^{T}\sum_{i=2}^{n}\sigma^{ih},\quad h = 2,\cdots,n \tag{E-14}$$

$$\frac{\partial^2 L}{\partial\alpha_h\partial\alpha_k} = -\frac{T}{r^2} + \frac{\lambda^2}{\sigma^2}\sum_{t=1}^{T}z(a_t)(\ln x_{ht})(\ln x_{kt}) - \frac{1}{\sigma^2}\sum_{t=1}^{T}(\ln x_{ht})(\ln x_{kt}) \tag{E-15}$$

$$- \frac{T}{\alpha_h \alpha_k} \sigma^{hk}, \quad h, k = 2, \cdots, n$$

where

$$z(s) = s \cdot m(s) - m^2(s)$$

$$Q(s) = -m(s) + s^2 \cdot m(s) - s \cdot m^2(s)$$

$$P(s) = -2m(s) + s^2 \cdot m(s) - s \cdot m^2(s)$$

The elements of the information matrix are

$$I_{\xi_l \xi_m} = T \sigma^{ml}$$

$$I_{\xi_l A} = 0$$

$$I_{\xi_l \lambda} = 0$$

$$I_{\xi_l \sigma} = 0$$

$$I_{\xi_l \sigma_{hk}} = 0$$

$$I_{\xi_l \alpha_1} = \frac{T}{\alpha_1} \sum_{i=2}^{n} \sigma^{il}$$

$$I_{\xi_l \alpha_h} = -\frac{T}{\alpha_h} \sigma^{hl}$$

$$I_{AA} \approx -\frac{\partial^2 L}{\partial A \partial A} \quad \text{see eq.} (E-2) \qquad \text{(E-2)'}$$

$$I_{A\lambda} \approx -\frac{\partial^2 L}{\partial A \partial \lambda} \quad see\ eq.(E-3) \tag{E-3)'}$$

$$I_{A\sigma} \approx -\frac{\partial^2 L}{\partial A \partial \sigma} \quad see\ eq.(E-4) \tag{E-4)'}$$

$$I_{A\sigma_{ij}} = 0$$

$$I_{A\alpha_i} \approx -\frac{\partial^2 L}{\partial A \partial \alpha_i} \quad see\ eq.(E-5) \tag{E-5)'}$$

$$I_{\lambda\lambda} \approx -\frac{\partial^2 L}{\partial \lambda \partial \lambda} \quad see\ eq.(E-6) \tag{E-6)'}$$

$$I_{\lambda\sigma} \approx -\frac{\partial^2 L}{\partial \lambda \partial \sigma} \quad see\ eq(E-7) \tag{E-7)'}$$

$$I_{\lambda\sigma_{ij}} = 0$$

$$I_{\lambda\alpha_i} \approx -\frac{\partial^2 L}{\partial \lambda \partial \alpha_i} \quad see\ eq.(E-8) \tag{E-8)'}$$

$$I_{\sigma\sigma} \approx -\frac{\partial^2 L}{\partial \sigma \partial \sigma} \quad see\ eq.(E-9) \tag{E-9)'}$$

$$I_{\sigma\sigma_{ij}} = 0$$

$$I_{\sigma\alpha_i} \approx -\frac{\partial^2 L}{\partial \sigma \partial \alpha_i} \quad see\ eq.(E-19) \tag{E-10)'}$$

$$I_{\sigma_{hh}\sigma_{hh}} = \frac{T}{2}(\sigma^{hh})^2 \tag{E-11)'}$$

$$I_{\sigma_{hh}\sigma_{lm}} = \frac{T}{2}\sigma^{hl}\sigma^{mh} \tag{E-12)'$$

$$I_{\sigma_{hk}\sigma_{lm}} = T\sigma^{hl}\sigma^{mk} \tag{E-13)'$$

$$I_{\sigma_{hk}\alpha_l} = 0, \quad l = 1, \cdots, n$$

$$I_{\alpha_h\alpha_k} \approx -\frac{\partial^2 L}{\partial\alpha_h\partial\alpha_k}, \quad h, k = 1, \cdots, n. \qquad \textit{see eq.}\,(E-13) - (E-15) \tag{E-14)'$$

Some elements of the information matrix are difficult to find and are approximated by the negative of the second partial derivatives since this will not affect the probability limit of the resulting "information matrix."

BIBLIOGRAPHY

BIBLIOGRAPHY

Aigner, D. J., C. A. K. Lovell and P. Schmidt (1977), "Formulation and Estimation of Stochastic Frontier Production Function Models," *Journal of Econometrics*, 6, 21-37.

Aitchison, J. and S. D. Silvey (1958), "Maximum-likelihood Estimation of Parameters Subject to Restraints," *Annals of Mathematical Statistics*, 29, 813-828.

--------- (1960), "Maximum-likelihood Estimation Procedures and Associated Tests of Significance," *Journal of the Royal Statistical Society*, Series B, 22, 154-171.

Breusch, T. S. and A. R. Pagan (1980), "The Lagrange Multiplier Test and Its Application to Model Specification in Econometrics," *Review of Economic Studies*, XLVII, 239-253.

Connally, M. (1982), "The Impact of Final Offer Arbitration on the Bargaining Process and Wage Outcomes," unpublished Ph.D. thesis, Michigan State University.

Cragg, J. G. (1971), "Some Statistical Models for Limited Dependent Variables with Application to the Demand for Double Goods," *Econometrica*, 39, 829-844.

Dhrymes, P. J., L. R. Klein and K. Steiglitz (1970), "Estimation of Distributed Lags," *International Economic Review*, 11, 235-250.

Farrell, M. J. (1957), "The Measurement of Productive Efficiency," *Journal of the Royal Statistical Society* (A. Goneral), 120, 253-281.

Finney, D. J. (1971), *Probit Analysis*; Third edition, Cambridge University Press.

Forsund, F. R., C. A. K. Lovell and P. Schmidt (1980), "A Survey of Frontier Production Functions and of Their Relationship to Efficiency Measurement," *Journal of Econometrics*, 13, 5-25.

Godfrey, L. G. (1978), "Testing Against General Autoregressive and Moving Average Error Models When the Regressors Include Lagged Dependent Variables," *Econometrica*, 46, 1293-1301.

Gunderson, M. (1974), "Retention of Trainees : A Study with Dichotomous Dependent Variables," *Journal of Econometrics*, 2, 79-93.

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.

-------- (1979), "Sample Selection Bias as a Specification Error," *Econometrica*, 47, 153-161.

Johnson, N. L. and S. Kotz (1970), *Continuous Univariate Distribution –I*, Boston : Houghton Miffin Company.

Jorgenson, D. W. (1966), "Rational Distributed Lag Functions," *Econometrica*, 34, 135-149.

Judge, G. G., W. E. Griffiths, R. C. Hill and T. C. Lee (1980), *The Theory* and *Practice of Econometrics*, New York : Wiley.

Klein, L. R. (1958), "The Estimation of Distributed Lags," *Econometrica*, 26, 553-565.

Koyck, L. M. (1954), *Distributed Lags* and *Investment Analysis*, Amsterdam : North-Holland.

Melino, A. (1982), "Testing for Sample Selection Bias," *Review of Economic Studies*, forthcoming.

Nicholls, D. F., A. R. Pagan and R. D. Terrell (1975), "The Estimation and Use of Models with Moving Average Disturbance Terms : A Survey," *International Economic Review*, 16, 113-133.

Olsen, R. J. (1978), "Note on the Uniqueness of th Maximum Likelihood Estimator for the Tobit Model," *Econometrica*, 46, 1121-1215.

Poirier, D. J. (1980), "Partial Observability in Bivariate Probit Models," *Journal of Econometrics*, 12, 209-217.

Rao, C. R. (1947), "Large Sample Tests of Statistical Hypothesis Concerning Several Parameters with Applications to Problems of Estimation," *Proceedings of the Cambridge Philosophical Society*, 44, 50-57.

Schmidt, P. (1976), "On the Statistical Estimation of Parametric Frontier Production Functions," *Review of Economics* and *Statistics*, 58, 238-239.

------ and D. K. Guilkey (1976), "The Effects of Various Treatments of Truncation Remainders on Tests of Hypotheses in Distributed Lag Models," *Journal of Econometrics*, 4, 211-230.

------ and C. A. K. Lovell (1979), "Estimating Technical and Allocative Inefficiency Relative to Stochastic Production and Cost Frontiers," *Journal of Econometrics*, 9, 343-366.

------ and C. A. K. Lovell (1980), "Estimating Stochastic Production and Cost Frontiers when Technical and Allocative Inefficiency Are Correlated," *Journal of Econometrics*, 13, 83-100.

Silvey, S. D. (1959), "The Lagrangian Multiplier Test," *Annals of Mathematical Statistics*, 30, 389-407.

Solow, R. M. (1960), "On a Family of Lag Distributions," *Econometrica*, 28, 393-406.

Stevenson, R. E. (1980), "Likelihood Functions for Generalized Stochastic Frontier Estimation," *Journal of Econometrics*, 13, 57-66.

Tobin, J. (1958), "Estimation of Relationships for Limited Dependent Variables," *Econometrica*, 26, 24-36.

Wald, A. (1943), "Tests of Statistical Hypotheses Concerning Several Parameters when the Number of Observations is Large," *Trans. Am. Math. Soc.*, 54, 426-482.