Some Theorems in the Linear Probability Model
Author(s): Takeshi Amemiya
Source: *International Economic Review*, Vol. 18, No. 3 (Oct., 1977), pp. 645-650
Published by: Wiley for the Economics Department of the University of Pennsylvania and Institute of Social and Economic Research -- Osaka University
Stable URL: http://www.jstor.org/stable/2525953
Accessed: 06/11/2014 17:06

INTERNATIONAL ECONOMIC REVIEW
Vol. 18, No. 3, October, 1977

SOME THEOREMS IN THE LINEAR PROBABILITY MODEL*

BY TAKESHI AMEMIYA[1]

1. INTRODUCTION

In this paper we prove some theorems about relationships between several weighted least squares estimators in the linear probability model.[2] These theorems are concerned with both finite-sample and asymptotic results and are given in Section 2. We obtain, among other things, the necessary and sufficient condition for the least squares estimator and the maximum likelihood estimator to be identical.

As an application of these theorems, we consider in Section 3 a special case of the linear probability model in which all the variables that appear in both sides of a regression equation are binary (called binary variable multiple regression) and show how some of the theorems of Amundsen [**1**] follow immediately from the general theorems of Section 2.

2. LINEAR PROBABILITY MODEL

Let $\{y_t\}$, $t=1, 2, \ldots, T$, be a sequence of independent binary random variables such that

$$P(y_t = 1) = \alpha' w_t \tag{2.1}$$

where w_t is a q-component column vector of known constants such that $w_t \neq 0$. Then we can write

$$y_t = \alpha' w_t + v_t \tag{2.2}$$

where $Ev_t = 0$ and $Vv_t \equiv \lambda_t = \alpha' w_t(1-\alpha' w_t)$, or in obvious vector notation

$$y = W\alpha + v. \tag{2.3}$$

We define $\Lambda = Evv'$, so that Λ is a diagonal matrix whose t-th element is λ_t. We assume that the rank of W is equal to q.

DEFINITION 1: Let D be any positive diagonal matrix of size T, stochastic or nonstochastic, such that whenever $w_t = w_s$ the t-th and s-th diagonal elements of D are the same. Then, we define the class of weighted least squares estimators

* Manuscript received December, 1975; revised July, 1976.

[1] This work was supported by National Science Foundation Grant SOC73-05453-A02 at the Institute for Mathematical Studies in the Social Sciences, Stanford University, and by a fellowship from the John Simon Guggenheim Memorial Foundation.

[2] For the various other aspects of the model, see Goldberger [**2**, (250–51)].

of α, denoted $\tilde{\alpha}(D)$, by

$$\tilde{\alpha}(D) = (W'D^{-1}W)^{-1}W'D^{-1}y. \tag{2.4}$$

Note that $\tilde{\alpha}(I)$, the least squares estimator, and $\tilde{\alpha}(\Lambda)$, the Gauss-Markov estimator, belong to this class.

Next we will define the maximum likelihood estimator. The logarithmic likelihood function is given by

$$L = \sum_{t=1}^{T} [y_t \log \alpha' w_t + (1 - y_t) \log (1 - \alpha' w_t)]. \tag{2.5}$$

We have

$$\frac{\partial L}{\partial \alpha} = \Sigma [y_t (\alpha' w_t)^{-1} - (1 - y_t)(1 - \alpha' w_t)^{-1}] w_t \tag{2.6}$$

and

$$\frac{\partial^2 L}{\partial \alpha \partial \alpha'} = - \Sigma [y_t (\alpha' w_t)^{-2} + (1 - y_t)(1 - \alpha' w_t)^{-2}] w_t w_t', \tag{2.7}$$

where the range of the summation is $t = 1$ to T unless otherwise indicated. Therefore we have

$$E \frac{\partial^2 L}{\partial \alpha \partial \alpha'} = - \Sigma [\alpha' w_t (1 - \alpha' w_t)]^{-1} w_t w_t'. \tag{2.8}$$

DEFINITION 2: The maximum likelihood estimator of α, denoted $\hat{\alpha}$, is the value of α that is implicitly determined by setting the right-hand side of (2.6) equal to zero.

The maximum likelihood estimator may be iteratively computed by the method of scoring defined by

$$\hat{\alpha}_2 = \{\Sigma [\hat{\alpha}_1' w_t (1 - \hat{\alpha}_1' w_t)]^{-1} w_t w_t'\}^{-1} \Sigma [\hat{\alpha}_1' w_t (1 - \hat{\alpha}_1' w_t)]^{-1} w_t y_t \tag{2.9}$$

when $\hat{\alpha}_2$ is the second-round estimator when $\hat{\alpha}_1$ is used as the initial estimator.[3] If the iteration converges, it converges to the maximum likelihood estimator $\hat{\alpha}$[4] so that we have

$$\hat{\alpha} = \tilde{\alpha}(\hat{\Lambda}) \tag{2.10}$$

where $\hat{\Lambda}$ is the diagonal matrix whose t-th element is equal to $\hat{\alpha}' w_t (1 - \hat{\alpha}' w_t)$. Thus, $\hat{\alpha}$ is also a member of the class of weighted least squares estimators defined in Definition 1.

In the remainder of this section we will study relationships between the mem-

[3] The method of scoring is the iteration defined by $\hat{\alpha}_2 = \hat{\alpha}_1 - [E(\partial^2 L/\partial\alpha\partial\alpha')]^{-1} \partial L/\partial\alpha$ where both derivatives are evaluated at $\hat{\alpha}_1$. See Rao [4, (366)]. Using (2.6) and (2.8), this can be shown to be equivalent to (2.9).

[4] Since $\partial^2 L/\partial\alpha\partial\alpha'$ is negative definite by (2.7), there is a unique maximum for (2.5).

bers of the class of weighted least squares estimators and prove three theorems and state one without proof. Theorems 1 and 2 pertain to finite sample results and 3 and 4 to asymptotic results.

THEOREM 1: *Let W_0 consist of the maximal distinct rows of W. Then, $\tilde{\alpha}(I)=\tilde{\alpha}(\Lambda)$ for all α if and only if W_0 is nonsingular.*

PROOF: By a well-known theorem (see, for example, Rao [**3**]) $\tilde{\alpha}(I)=\tilde{\alpha}(\Lambda)$ for all α if and only if

$$\Lambda W = WB \quad \text{for some nonsingular} \quad B. \tag{2.11}$$

Clearly, (2.11) is equivalent to

$$\Lambda_0 W_0 = W_0 B \quad \text{for some nonsingular} \quad B, \tag{2.12}$$

where Λ_0 is the sub-diagonal block of Λ that corresponds to the position of W_0 in W. If W_0 is nonsingular, (2.12) obviously holds because we can put $B = W_0^{-1}\Lambda_0 W_0$. Suppose (2.12) is true and yet W_0 is singular (that is to say, if there are more rows than columns), let W^* be a $q \times q$ nonsingular submatrix of W_0 and let w_t' be an arbitrary row vector in W_0 but not in W^*. Then we must have

$$\Lambda^* W^* = W^* B, \tag{2.13}$$

where Λ^* is the sub-diagonal block of Λ corresponding to the position of W^*, so that B is uniquely determined by $B = W^{*-1}\Lambda^* W^*$. We can write $w_t' = c'W^*$ for some vector c, so from (2.12) we have

$$\lambda c' W^* = c' W^* B, \tag{2.14}$$

where the position of λ in Λ_0 corresponds to that of w_t' in W_0. But from (2.13) we have

$$c' \Lambda^* W^* = c' W^* B. \tag{2.15}$$

Therefore from (2.14) and (2.15) we have

$$\lambda c' = c' \Lambda^*. \tag{2.16}$$

Since at least one element of c is nonzero by our assumption, (2.16) implies that for some row vector w_s' in W^* we have

$$w_t'\alpha = w_s'\alpha. \tag{2.17}$$

But since $w_t \neq w_s$ by definition of W_0, (2.17) cannot hold for all α. Q.E.D.

THEOREM 2: *Let D be as defined in Definition 1. Then, $\tilde{\alpha}(D)=\tilde{\alpha}(\Lambda)$ for all α and for all D if and only if W_0 is nonsingular.*

PROOF: $\tilde{\alpha}(D)$ is the least squares estimator applied to equation $D^{-1/2}y = D^{-1/2}W\alpha + D^{-1/2}v$ and $\tilde{\alpha}(\Lambda)$ is the Gauss-Markov estimator applied to the same equation. Therefore, using Theorem 1, $\tilde{\alpha}(D)=\tilde{\alpha}(\Lambda)$ for all α and for any D if and

only if the submatrix consisting of the maximal distinct rows of $D^{-1/2}W$ is nonsingular. But, because of the condition on D, this submatrix is equal to $D^{-1/2}W_0$.
Q. E. D.

The next two theorems are concerned with asymptotic results as T approaches infinity. For this purpose we need additional notation and assumptions. Without loss of generality assume that W_0 consists of the first T_0 row vectors of W. Define the set of integers $C_t = \{i | 1 \leqq i \leqq T, w_i = w_t\}$ for $t = 1, 2, \ldots, T_0$. Define n_t as the number of elements in C_t. Throughout this paper we assume that as T goes to infinity, W_0 and hence T_0 are fixed and n_t goes to infinity in such a way that $\lim_{T\to\infty} n_t T^{-1} = r_t \neq 0$ for $t = 1, 2, \ldots, T_0$.

DEFINITION 3: The weighted least squares estimator $(W'D^{-1}W)^{-1}W'D^{-1}y$ is called the consistently-weighted least squares estimator if D is as defined in Definition 1 and in addition D converges to Λ in probability as T goes to infinity.

DEFINITION 4: Suppose $\hat{\theta}$ is a consistent estimator of θ and $\sqrt{T}(\hat{\theta} - \theta)$ converges in distribution to $N(0, A)$. Then, we say that the asymptotic distribution of $\hat{\theta}$ is normal with mean θ and the asymptotic variance-covariance matrix given by A.

We state the following theorem without proof as it must be well-known and at any rate is very easy to prove.

THEOREM 3: *The asymptotic distributions of the maximum likelihood estimator and the consistently-weighted least squares estimator are the same and equal to* $N(0, [\sum_{t=1}^{T_0} r_t \lambda_t^{-1} w_t w_t']^{-1})$.

Next we will prove

THEOREM 4: *The asymptotic distributions of the maximum likelihood estimator and the least squares estimator are the same if and only if the two estimators are identical.*

PROOF: Since the "if" part is obvious, we will prove the "only if" part. Moreover, since the asymptotic normality is obvious, we need only to compare the asymptotic variance-covariance matrix, denoted A, of the least squares estimator with that of the maximum likelihood estimator given in Theorem 3, denoted B. We have

$$A = [\sum_{t=1}^{T_0} r_t w_t w_t']^{-1} \sum_{t=1}^{T_0} \alpha' w_t (1 - \alpha' w_t) r_t w_t w_t' [\sum_{t=1}^{T_0} r_t w_t w_t']^{-1} \tag{2.18}$$

or in vector notation

$$A = (W_0' R W_0)^{-1} W_0' R^{\frac{1}{2}} \Lambda_0 R^{\frac{1}{2}} W_0 (W_0' R W_0)^{-1}, \tag{2.19}$$

where R is the $T_0 \times T_0$ diagonal matrix whose t-th diagonal element is equal to r_t.

We have

$$B = (W_0' R^{\frac{1}{2}} \Lambda_0^{-1} R^{\frac{1}{2}} W_0)^{-1}. \tag{2.20}$$

Therefore $A = B$ if and only if

$$\Lambda_0 R^{\frac{1}{2}} W_0 = R^{\frac{1}{2}} W_0 B \quad \text{for some nonsingular} \quad B. \tag{2.21}$$

But clearly (2.21) is equivalent to (2.12). Therefore the theorem follows from Theorem 1 and Theorem 2. Q.E.D.

3. BINARY VARIABLES MULTIPLE REGRESSION

In this section we will consider the following model. Suppose there are $K+1$ binary random variables $y_t, x_{1t}, x_{2t}, \ldots, x_{Kt}$, $t=1, 2, \ldots, T$, each taking values 0 or 1 and independent through t. Define $x_t = (x_{1t}, x_{2t}, \ldots, x_{Kt})'$. Then x_t can take $S=2^K$ distinct values. Let f be an arbitrary one-to-one mapping from these S values to integers $1, 2, \ldots, S$. The unknown parameters to estimate are $P_s = \text{Prob.}[y_t = 1 | f(x_t) = s]$, $s = 1, 2, \ldots, S$.

We can write

$$y_t = \beta_0 + \sum_{i=1}^{K} \beta_i x_{it} + \sum_i \sum_j \beta_{ij} x_{it} x_{jt} + \cdots + \beta_{12\cdots K} x_{1t} x_{2t} \cdots x_{Kt} + u_t, \tag{3.1}$$

where $u_t = y_t - E(y_t | x_t)$. Write (3.1) as

$$y_t = z_t' \gamma + u_t \tag{3.2}$$

where z_t and γ are defined in an obvious way as column vectors with S components. We further write (3.2) in obvious vector notation as

$$y = Z\gamma + u. \tag{3.3}$$

We assume that the rank of Z is S. This is equivalent to assuming that the matrix Z_0 consisting of the maximal distinct rows of Z is nonsingular.[5] We also assume that $P_s \neq 0$ for $s = 1, 2, \ldots, S$.[6] Defining a vector $P = (P_1, P_2, \ldots, P_S)'$, we can establish a one-to-one correspondence between γ and P by

[5] We will show that this assumption involves no loss of generality. Let H be the $S \times S$ nonsingular matrix which Z would assume if we had $f(x_t) = t$, $t = 1, 2, \ldots, T$, and $T = S$. Partition $H' = (H_1', H_2')$ and P' accordingly as (P_1', P_2') and suppose $Z_0 = H_1$. Then we can estimate only $P_1 (\equiv P^*)$ or $H_1 \gamma (\equiv \gamma^*)$. We can find Z^* such that $Z\gamma = Z^* \gamma^*$ where the matrix Z_0^* consisting of the maximal distinct rows of Z^* is nonsingular (in fact, equal to an identity matrix). Thus we have (3.3)′ $y = Z^* \gamma^* + u$ and (3.4)′ $P^* = Z_0^* \gamma^*$ instead of (3.3) and (3.4), for which the assumption is satisfied.

[6] This assumption likewise involves no loss of generality. Let H be as defined in footnote 5. Partition $H' = (H_1', H_2')$ and $P' = (P_1', P_2')$ accordingly and suppose $P_2 = 0$. Then all the row vectors of Z that are the same as those in H_2 and the corresponding elements of y may be discarded from the sample. But, then, this is exactly the situation considered in footnote 5.

$$P = Z_0\gamma. \tag{3.4}$$

Because of (3.4) we can consider the estimation of γ rather than that of P. Thus, this model may be regarded as a special case of the linear probability model considered in the previous section.

Amundsen [1] considered this model and obtained several interesting results. We will show below in Theorems 5 and 6 that some of the theorems Amundsen states with or without proof follow immediately from the more general theorems of the previous section.[7]

THEOREM 5: *The least squares estimator of γ is identical with the weighted least squares estimator (see Definition 1) of γ.*

PROOF: Follows immediately from Theorems 1 and 2. Q.E.D.

Theorem 5 implies Amundsen's Theorem 1(i) and his statement in the first seven lines on page 65. Moreover, because we have shown in footnotes 4 and 5 that the assumption of the full rank of Z and the assumption of nonzero P involve no loss of generality, Theorem 5 actually implies the relevant part of Amundsen's Theorems 2 and 3 as well.

THEOREM 6: *Partition $\gamma'=(\gamma_1', \gamma_2')$ and suppose $\gamma_2=0$.*[8] *Then, the asymptotic distribution of the maximum likelihood estimator of γ_1 is (i) the same as that of the consistently-weighted least squares estimator but (ii) different from that of the least squares estimator.*

PROOF: (i) follows immediately from Theorem 3. Partition Z accordingly as (Z_1, Z_2). Then, Z_{10}, the matrix consisting of the maximal distinct rows of Z_1, is singular. Therefore, (ii) follows from Theorems 1, 2, and 4. Q.E.D.

Theorem 6 implies the main results of Section 4.1 of Amundsen [1].

Stanford University

REFERENCES

[1] AMUNDSEN, H. T., "Binary Variable Multiple Regressions," *Scandinavian Journal of Statistics*, Vol. 1 (1974), 59–70.
[2] GOLDBERGER, A. S., *Econometric Theory* (New York: John Wiley & Sons, 1964).
[3] RAO, C. R., "The Theory of Least Squares when the Parameters are Stochastic and its Application to the Analysis of Growth Curves," *Biometrika*, Vol. 52 (1965), 447–458.
[4] ———, *Linear Statistical Inference and its Applications,*, Second edition (New York: John Wiley & Sons, 1973).

[7] It should be pointed out that Amundsen obtained a few other interesting results which we do not discuss in this paper.

[8] We can assume without loss of generality that $\gamma_2=0$ is consistent with our assumption $P_s \neq 0$, $s=1, 2, \ldots, S$. For, if some of the equations in $\gamma_2=0$ imply $P_2=0$ where P_2 is a part of P, we can eliminate the corresponding rows of Z as indicated in footnote 6.