REGRESSION ANALYSIS UNDER DIFFERENTIAL NON-RESPONSE

Gad Nathan, Hebrew University, Jerusalem

1. Introduction

We consider the estimation of the regression coefficient β_{12} in a linear regression:

$$E(X_{1i}|X_{2i}) = \mu_1 + \beta_{12}X_{2i}, \tag{1.1}$$

on the basis of a random sample of observations, some of which are missing due to non-response. We assume that response depends, in a way to be specified, on values of a third variable, X_3, which is observable for the whole sample and which is related to the variables X_1 and X_2.

Two alternative approaches to this problem have been considered. The econometric approach, summarized by Heckman (1979), assumes response to be determined by an unobservable variable, X_4, which is the dependent variable in a linear regression relationship with X_3. Heckman (1976) proposes a two-stage (probit and OLS) estimator, which is asymptotically unbiased, under certain assumptions, within a conditional framework of inference (given the values of X_2 and X_3), and for which asymptotic expressions of variance are given by Lee, Maddala and Trost (1980).

The sampling theory approach, proposed by Nathan and Holt (1980) and by Holt and Smith (1979), to deal with the effect of sample design on regression analysis can be applied to this case. This is done by considering the response as obtained by sub-sampling from the main sample, via a non-informative design, with the sample distribution determined, in an unknown way, by the values of X_3. An overall MLE, under assumptions of multivariate normality, which is asymptotically unbiased, is proposed by Holt and Smith (1979) and some of its properties are shown to hold under less stringent assumptions by Nathan and Holt (1980).

In order to compare the estimators proposed, an overall model is set up, within which both approaches can be imbedded as special cases (section 2). The implications of each of the approaches with respect to the assumptions are examined in sections 3 and 4 and expressions for the conditional expectations and variances of the estimators are given in section 5. In section 6 an estimator based on the overall model (without simplifying assumptions) is proposed and in section 7 the results of some simulation comparisons between the performances of the estimators under the full model assumptions and under certain relaxations of the assumptions are presented.

2. Overall Model

We formulate the model for univariate variables, for simplicity, but the extension to the multivariate case is immediate. We assume a finite sample of N values of four variables with a multivariate normal distribution, i.e.

$$(X_{1i}, X_{2i}, X_{3i}, X_{4i}) \sim N(\underline{\mu};\Sigma) \tag{2.1}$$

$(i=1, \ldots, N)$

where: $\underline{\mu}' = (\mu_1, \mu_2, \mu_3, \mu_4)$

and:

$$\Sigma_{jk} = \sigma_{jk};\ \sigma_{jj} = \sigma_j^2 ,$$

and define β_{ab}, $\beta_{ab.c}$, $\sigma^2_{a.b}$, $\sigma_{ab.c}$, $\sigma^2_{a.bc}$, $\sigma_{ab.cd}$ as usual. The conditional distribution of X_{1i}, given X_{2i}:

$$X_{1i}|X_{2i} \sim N(\mu_1+\beta_{12}(X_{2i}-\mu_2);\ \sigma^2_{1.2}), \tag{2.2}$$

defines the regression relationship of interest and β_{12} is the parameter of interest. Observations on X_{1i} are available only for a sub-sample of size n, defined by the indicator variable, d_i, which is determined by the variable X_{4i}:

$$d_i = \begin{cases} 1: X_{4i} \geq 0 \\ 0: \text{otherwise.} \end{cases} \tag{2.3}$$

Observations on X_{4i} are not available for any unit (except via d_i), while observations on X_{3i} are available for all units of the main sample $(i=1, \ldots, N)$. Observations on X_{2i} may be available for all $i=1, \ldots, N$, but those for which $d_i=0$ are not in general useful for estimating β_{12}.

3. The Sampling Theory Approach

The sampling theory model considers the marginal distribution of (X_{1i}, X_{2i}, X_{3i}) and derives the overall MLE of β_{12} under the assumption of non-informativeness of the sample design, as defined by:

$$(X_{1i},X_{2i})|(X_{3i},d_i=1) \sim (X_{1i},X_{2i})|X_{3i}. \tag{3.1}$$

This is equivalent to the condition:

$$\sigma^2_{14.3} = \sigma_{24.3} = 0 \tag{3.2}$$

The MLE estimator is obtained by DeMets and Halperin (1977) as:

$$\hat{\beta}_{12} = \frac{s_{12} + \dfrac{s_{13}s_{23}}{s_3^2}\left(\dfrac{\hat{\sigma}_3^2}{s_3^2} - 1\right)}{s_2^2 + \dfrac{s_{23}^2}{s_3^2}\left(\dfrac{\hat{\sigma}_3^2}{s_3^2} -1\right)}, \tag{3.3}$$

where: $s_{ab} = \frac{1}{n}\sum_{i=1}^{N} d_i(X_{ai}-\bar{x}_a)(X_{bi}-\bar{x}_b)$

$$s_a^2 = \frac{1}{n}\sum_{i=1}^{N} d_i(X_{ai}-\bar{x}_a)^2 ;$$

$$\bar{x}_a = \frac{1}{n}\sum_{i=1}^{N} d_iX_{ai};$$

$$\hat{\sigma}_a^2 = \frac{1}{N}\sum_{i=1}^{N} (X_{3i}-\bar{X}_3)^2 ;\ \bar{X}_a = \frac{1}{N}\sum_{i=1}^{N} X_{3i}$$

$(a,b = 1,2,3)$.

4. The Econometric Approach

Heckman (1976) considers the joint conditional distribution of (X_{1i},X_{4i}), given (X_{2i},X_{3i}), induced by (2.1), in the form of the regression relationships:

$$X_{1i} = \mu_1 + \beta_{12}(X_{2i}-\mu_2) + u_{1i}$$
$$X_{4i} = \mu_4 + \beta_{43}(X_{3i}-\mu_3) + u_{4i}\ , \qquad (4.1)$$

where:

$$(u_{1i},u_{4i})|(X_{2i},X_{3i}) \sim N(\underline{0},\Sigma^*) \qquad (4.2)$$

$$\Sigma^* = \begin{pmatrix} \sigma^2_{1.23} & \sigma^2_{14.23} \\ \sigma_{14.23} & \sigma^2_{4.23} \end{pmatrix}.$$

The derivation of the estimator is made under the implied assumptions:

$$X_{1i}|(X_{2i},X_{3i}) \sim X_{1i}|X_{2i}$$
$$X_{4i}|(X_{2i},X_{3i}) \sim X_{4i}|X_{3i}\ , \qquad (4.3)$$

which is equivalent to the conditions:

$$\sigma_{13.2} = \sigma_{24.3} = 0. \qquad (4.4)$$

Note that this implies that the parameters of (4.2) become:

$$\sigma^2_{1.23} = \sigma^2_{1.2};\ \sigma^2_{4.23} = \sigma^2_{4.3}\ ;$$
$$\sigma^2_{14.23} = \sigma_{14.2} = \sigma_{14.3} \qquad . \qquad (4.5)$$

Heckman (1976) obtains an estimator of β_{12} in two stages. At the first stage, probit analysis is used to estimate the parameters of the conditional probability that the unit is sampled, given X_{3i},

$$P(d_i=1|X_{3i}) = \Phi(-Z_i)\ , \qquad (4.6)$$

where $Z_i = -[\mu_4 + \beta_{43}(X_{3i}-\mu_3)]/\sigma_{4.3}$ and Φ is the standard normal d.f. Let $\hat{Z}_i$ be the estimate of Z_i obtained at this stage. At the second stage, β_{12} is estimated by OLS from the regression relationship:

$$E(X_{1i}|X_{2i},d_i=1) = \mu_1 + \beta_{12}(X_{2i}-\mu_2) + \frac{\sigma_{14.3}}{\sigma_{4.3}}\hat{\lambda}_i, \qquad (4.7)$$

where $\hat{\lambda}_i = \phi(\hat{Z}_i)/\Phi(-\hat{Z}_i)$ is estimated at the first stage (and ϕ is the standard normal p.d.f.).

Denoting sample variances and covariances involving $\hat{\lambda}_i$ by

$$s^2_{\hat{\lambda}} = \frac{1}{n}\sum_{i=1}^{N} d_i(\hat{\lambda}_i-\hat{\lambda})^2, \text{ where } \hat{\lambda} = \frac{1}{n}\sum_{i=1}^{N} d_i\hat{\lambda}_i;$$

$$s_{a\hat{\lambda}} = \frac{1}{n}\Sigma d_i\hat{\lambda}_i(X_{ai}-\bar{x}_a) \qquad (a=1,2,3) \qquad (4.8)$$

and the conditional sample variances and covariances by:

$$s_{ab.c} = s_{ab} - s_{ac}\,s_{bc}/s^2_c\ ;$$

$$s^2_{a.b} = s^2_a - s^2_{ab}/s^2_b, \quad (a,b,c = 1,2,3,\hat{\lambda}),$$

the resulting estimator can be written as:

$$\hat{\beta}^*_{12} = s_{12.\hat{\lambda}}/s^2_{2.\hat{\lambda}} = \frac{s_{12}-s_{1\hat{\lambda}}s_{2\hat{\lambda}}/s^2_{\hat{\lambda}}}{s^2_2-s^2_{2\hat{\lambda}}/s^2_{\hat{\lambda}}}. \qquad (4.9)$$

5. Conditional Expectations and Variances of the Estimators

Under the general model of section 2, the conditional expectation of $\hat{\beta}_{12}$, given the values of $\underline{X}_2 = (X_{21}, \ldots, X_{2N})$, $\underline{X}_3 = (X_{31}, \ldots, X_{3N})$ and $\underline{d} = (d_1, \ldots, d_N)$, can be derived as:

$$E(\hat{\beta}_{12}|\underline{X}_2,\underline{X}_3,\underline{d}) = \beta_{12.3} + \frac{\beta_{13.2}s_{23}\hat{\sigma}^2_3/s^2_3 + (s_{23}/s^2_3)(\hat{\sigma}^2_3/s^2_3)\ s_{3\lambda*} + s_{2\lambda*.3}}{s^2_{2.3} + (s^2_{23}/s^2_3)(\hat{\sigma}^2_3/s^2_3)}, \qquad (5.1)$$

where:

$$\lambda^*_i = (\sigma_{14.23}/\sigma_{4.23})\ \phi(Z^*_i)/\Phi(-Z^*_i),$$

$$Z^*_i = \frac{-[\mu_4 + \beta_{42.3}(X_{2i}-\mu_2) + \beta_{43.2}(X_{3i}-\mu_3)]}{\sigma_{4.23}}$$

and $s_{a\lambda*}$, $s_{a\lambda*.b}$ $(a,b = 1,2,3,\hat{\lambda})$ are defined as in (4.8) with $\hat{\lambda}_i$ replaced by λ^*_i. Similarly the conditional expectation of $\hat{\beta}^*_{12}$ can be derived as:

$$E(\hat{\beta}^*_{12}|\underline{X}_2,\underline{X}_3,\underline{d}) = \beta_{12.3} + \frac{\beta_{13.2}s_{23.\hat{\lambda}} + s_{2\lambda*.\hat{\lambda}}}{s^2_{2.\hat{\lambda}}}. \qquad (5.2)$$

The expectations under special conditions are as follows:

1) If $\hat{\sigma}^2_3 = s^2_3$:

$$E(\hat{\beta}_{12}|\underline{X}_2,\underline{X}_3,\underline{d}) = \beta_{12.3} + \frac{\beta_{13.2}s_{23} + s_{2\lambda*}}{s^2_2}, \qquad (5.3)$$

while there is no change in (5.2).

2) If the conditions of the sampling theory model (3.2) hold:

$$E(\hat{\beta}_{12}|\underline{X}_2,\underline{X}_3,\underline{d}) = \beta_{12.3} + \frac{\beta_{13.2}s_{23}\hat{\sigma}^2_3/s^2_3}{s^2_{2.3} + (s^2_{23}/s^2_3)(\hat{\sigma}_3/s^2_3)} \qquad (5.4)$$

$$E(\hat{\beta}^*_{12}|\underline{X}_2,\underline{X}_3,\underline{d}) = \beta_{12.3} + \frac{\beta_{13.2}s_{23.\hat{\lambda}}}{s^2_{2.\hat{\lambda}}} \qquad (5.5)$$

3) If the conditions of the econometric model (4.4) hold:

$$E(\hat{\beta}_{12}|\underline{X}_2,\underline{X}_3,\underline{d}) = \beta_{12} + \frac{(s_{23}/s^2_3)(\hat{\sigma}^2_3/s^2_3)s_{3\lambda*} + s_{2\lambda*.3}}{s^2_{2.3} + (s^2_{23}/s^2_3)(\hat{\sigma}^2_3/s^2_3)} \qquad (5.6)$$

$$E(\hat{\beta}^*_{12}|\underline{X}_2,\underline{X}_3,\underline{d}) = \beta_{12} + \frac{s_{2\lambda*.\hat{\lambda}}}{s^2_{2.\hat{\lambda}}} \qquad (5.7)$$

Thus $\hat{\beta}_{12}$ is conditionally unbiased if and only if both the conditions of the sampling theory model, (3.2), and those of the econometric model (4.4), hold.

The conditional variance of $\hat{\beta}_{12}$ can be derived, under the general model, as:

$$V(\hat{\beta}_{12}|\underline{X}_2,\underline{X}_3,\underline{d}) =$$

$$\sigma^2_{1.23}\ \{s^2_{2.3}+(s^2_{23}/s^2_3)(\hat{\sigma}^4_3/s^4_3) + \frac{\rho^2}{n}\sum_{i=1}^{N} d_i\lambda_i(Z^*_i-\lambda_i)$$

$$\left[(X_{2i}-\bar{x}_2) + (\frac{s_{23}}{s^2_3})(\frac{\hat{\sigma}^2_3}{s^2_3}-1)(X_{3i}-\bar{x}_3)\right]^2\}/$$

$$\{n[s^2_{2.3}+(s^2_{23}/s^2_3)(\sigma^2_3/s^2_3)]^2\} \qquad (5.8)$$

while that of $\hat{\beta}^*_{12}$ can be derived as:

$$V(\hat{\beta}^*_{12}|\underline{X}_2,\underline{X}_3,\underline{d}) =$$

$$\sigma^2_{1.23}\ \{s^2_{2.\hat{\lambda}} + \frac{\rho^2}{n}\sum_{i=1}^{N} d_i\lambda_i(Z^*_i-\lambda_i)$$

$$\left[(X_{2i}-\bar{x}_2) - (s_{2\hat{\lambda}}/s^2_{\hat{\lambda}})(\hat{\lambda}_i-\hat{\lambda})\right]^2\}/(ns^4_{2.\hat{\lambda}}). \qquad (5.9)$$

where: $\rho^2 = \sigma_{14.23}/(\sigma^2_{1.23}\ \sigma^2_{4.23})$.

For special conditions we obtain:

1) <u>If $\hat{\sigma}^2_3 = s^2_3$:</u>

$$V(\hat{\beta}_{12}|\underline{X}_2,\underline{X}_3,\underline{d}) =$$

$$= \frac{\sigma^2_{1.23}\ [s^2_{23}+\frac{\rho^2}{n}\ \Sigma d_i\lambda_i(Z^*_i-\lambda_i)(X_{2i}-\bar{x}^2_2)]}{ns^4_{23}}, \qquad (5.10)$$

while (5.9) does not change.

2) <u>If the conditions of the sampling theory model (3.2) hold:</u>

$$V(\hat{\beta}_{12}|\underline{X}_2,\underline{X}_3,\underline{d}) =$$

$$\frac{\sigma^2_{1.23}[s^2_{2.3} + (s^2_{23}/s^2_3)(\hat{\sigma}^4_3/s^2_3)]}{n[s^2_{2.3} + (s^2_{23}.s^2_3)(\hat{\sigma}^2_3/s^2_3)]^2} \qquad (5.11)$$

$$V(\hat{\beta}^*_{12}|\underline{X},\underline{X}_3,\underline{d}) = \frac{\sigma^2_{1.23}}{ns^2_{2.\hat{\lambda}}}. \qquad (5.12)$$

3) <u>If the conditions of the econometric model (4.4) hold:</u>

The only changes in (5.8) and (5.9) are in the definitions of ρ^2, λ_i and Z^*_i.

6. An Estimator Based on the Overall Model

Consider the joint conditional distribution of (X_{1i},X_{4i}), given (X_{2i},X_{3i}), under the overall model (1.2). This distribution is completely specified as follows:

$$X_{1i} = \mu_1 + \beta_{12.3}(X_{2i}-\mu_2) + \beta_{13.2}(X_{3i}-\mu_3) + v_{1i}$$
$$X_{4i} = \mu_4 + \beta_{42.3}(X_{2i}-\mu_2) + \beta_{43.2}(X_{3i}-\mu_3) + v_{2i}, \qquad (6.1)$$

where:

$$(v_{1i},v_{2i})|(X_{2i},X_{3i}) \sim N(\underline{0},\Sigma^*) \qquad (6.2)$$

and Σ^* is defined by (4.2).

Note that this reduces to (4.1)-(4.2) under the assumptions (4.3).

An estimator based on the more general model above can be derived in an analogous manner to that used in the econometric approach, described in section 4 as follows:

a) Use probit analysis to estimate the parameters of the conditional probability of a unit's being observed, given both X_{3i} and X_{2i}:

$$P(d_i=1|X_{2i},X_{3i}) = \Phi(-Z^*_i), \qquad (6.3)$$

where Z^*_i is defined by (5.1). Note that, under this model, the probit estimator of Z^*_i is just $\hat{Z}_i$ – the probit estimator of Z_i obtained under the simplifying assumptions.

b) The parameters of the following regression relationship:

$$E[X_{1i}|X_{2i},X_{3i},d_i=1] = \mu_1+\beta_{12.3}(X_{2i}-\mu_2)$$

$$+ \beta_{13.2}(X_{3i}-\mu_3) + \frac{\sigma_{14.23}}{\sigma_{4.23}}\frac{\phi(Z^*_i)}{\Phi(-Z^*_i)} \qquad (6.4)$$

are estimated by OLS after substituting $\hat{Z}_i$ instead of Z^*_i. Let $\hat{\beta}_{12.3}$ and $\hat{\beta}_{13.2}$ be the estimators of $\beta_{12.3}$ and $\beta_{13.2}$ thus obtained.

c) Let $\hat{\beta}_{32}$ be the OLS estimator of β_{32} obtained from the observations $\{(X_{2i},X_{3i});\ i=1,\ ...,\ N\}$, i.e.

$$\hat{\beta}_{32} = \hat{\sigma}_{23}/\hat{\sigma}^2_2 = \frac{\sum_{i=1}^{N}(X_{2i}-\bar{X}_2)(X_{3i}-\bar{X}_3)}{\sum_{i=1}^{N}(X_{2i}-\bar{X}_2)^2}. \qquad (6.5)$$

Then the estimator of β_{12} –$\hat{\beta}^o_{12}$– is obtained from the relationship:

$$\beta_{12} = \beta_{12.3} + \beta_{32}\ \beta_{13.2} \qquad (6.6)$$

as:

$$\hat{\beta}^o_{12} = \hat{\beta}_{12.3} + \hat{\beta}_{32}\ \hat{\beta}_{13.2}\ . \qquad (6.7)$$

The conditional expectation of $\hat{\beta}^o_{12}$, given $\underline{X}_2,\underline{X}_3$ and $\underline{d}$ can be shown to be:

$$E(\hat{\beta}^o_{12}|\underline{X}_2,\underline{X}_3,\underline{d}) = \beta_{12.3} + \hat{\beta}_{32}\ \beta_{13.2}$$

$$+ \frac{s_{2\lambda^*.3\hat{\lambda}}}{s^2_{2.3\hat{\lambda}}} + \hat{\beta}_{32}\ \frac{s_{3\lambda^*.2\hat{\lambda}}}{s^2_{3.2\hat{\lambda}}} \qquad (6.8)$$

Since $\hat{\beta}_{32}$ is a consistent estimator of β_{32} and since the last two terms of (6.8) vanish as $N\to\infty$ (because $\hat{\lambda}_i$ is a consistent estimator of λ_i), $\hat{\beta}^o_{12}$ is a consistent estimator of $\hat{\beta}_{12}$. This can also be seen from Heckman's (1979) result, if both $\underline{X}_2$ and $\underline{X}_3$ are used as explanatory variables in both equations.

The conditional variance of $\hat{\beta}^o_{12}$, given $\underline{X}_2, \underline{X}_3$ and $\underline{d}$ can be shown to be:

$$
\begin{aligned}
V(\hat{\beta}^o_{12} | \underline{X}_2, \underline{X}_3, \underline{d}) = \frac{\sigma^2_{1.23}}{n s^4_{2.3\hat{\lambda}}} \{ & s^2_{2.3\hat{\lambda}} (1+\hat{\beta}_{32}\ s^2_{2.\hat{\lambda}}/s^2_{3.\hat{\lambda}})^2 \\
& + \frac{\rho^2}{n} \sum_{i=1}^{N} d_i \lambda_i (Z^*_i - \lambda_i) \\
& [(1-\hat{\beta}_{32}\ s_{23.\hat{\lambda}}/s^2_{3.\hat{\lambda}})(X_{2i}-\bar{x}_2) \\
& - \frac{s_{23.\hat{\lambda}} - \hat{\beta}_{32}\ s^2_{2.\hat{\lambda}}}{s^2_{3.\hat{\lambda}}} (X_{3i}-\bar{x}_3) - \\
& \frac{s_{2\hat{\lambda}.3} + \hat{\beta}_{32}\ s^2_2/s^2_3}{s^2_{\hat{\lambda}.3}} (\hat{\lambda}_i - \hat{\lambda})]^2 \} \qquad (6.9)
\end{aligned}
$$

7. Comparisons of the Estimators

The proposed estimators can be computed without the assumptions implied by the models under which they are derived. Their conditional biases and variances can be computed both under the general model assumptions and under sub-models. However, the results cannot be compared analytically due to the fact that analytic expressions for the probit estimators $\hat{\lambda}_i$ are not available. Thus simulation comparisons have to be used.

An initial small simulation study was carried out with respect to the estimators $\hat{\beta}_{12}$ and $\hat{\beta}^*_{12}$. Four sets of population parameters, specified in Table 1 by their correlation matrices, were used. Under the first neither the econometric model assumptions (4.4) nor the sampling model assumptions (3.2) hold. The second set represents conditions close to those of the econometric model: $|\sigma_{13.2}|, |\sigma_{24.3}| < .01$, while for the remaining two sets the exact econometric model and the exact sampling model assumptions hold, respectively.

Table 1: Parameter Sets for Simulation Study

Parameter Set	Correlation Matrix			
1. General	1	.762	.626	.418
		1	.590	.288
			1	.319
				1
($\rho_{13.2} = .177$; $\rho_{24.3} = .100$; $\rho_{14.3} = .218$)				
2. Approximate Econometric Model Conditions	1	.762	.626	.418
		1	.704	.274
			1	.319
				1
($\rho_{13.2} = .009$; $\rho_{24.3} = .005$; $\rho_{14.3} = .218$)				
3. Exact Econometric Model Conditions	1	.762	.626	.418
		1	.822	.262
			1	.319
				1
($\rho_{13.2} = s_{24.3} = 0$; $\rho_{14.3} = .218$)				
4. Exact Sampling Model Conditions	1	.762	.626	.200
		1	.590	.188
			1	.319
				1
($\rho_{13.2} = .177$; $\rho_{24.3} = \rho_{14.3} = 0$)				

For all sets of parameters the value of the regression coefficient is the same: $\beta_{12} = .762$. The mean of X_4 was determined so that $E(n)/N = P(X_{4i} \geq 0) = .6$. For each set of parameters $N = 2{,}000$ values of (X_{2i}, X_{3i}, X_{4i}) were generated by the marginal trivariate normal distribution and the conditional biases and variances of $\hat{\beta}_{12}$ and β^*_{12} were computed, on the basis of the expressions in section 5. Ten repetitions were sufficient to attain the small standard errors of the estimates of the relative bias and of the estimates of the relative standard error, given in Table 2.

Table 2: Simulation Study Results

	Estimator	Parameter Set: General	Approx. Econom. Condit.	Exact Econom. Condit.	Exact Sample Condit.
		1	2	3	4
Relative Bias	$\hat{\beta}_{12}$	.035 (.0025)	.041 (.0011)	.046 (.0003)	-.002 (.0028)
	$\hat{\beta}^*_{12}$	.226 (.0004)	.176 (.1115)	-.004 (.0001)	.206 (.0008)
Relative Standard Error	$\hat{\beta}_{12}$	.024 (.0002)	.024 (.0002)	.024 (.0001)	.023 (.0002)
	$\hat{\beta}^*_{12}$	.028 (.0001)	.033 (.0001)	.041 (.0003)	.028 (.0003)
Relative Root M.S.E.	$\hat{\beta}_{12}$	.043 (.0022)	.048 (.0011)	.052 (.0003)	.024 (.0009)
	$\hat{\beta}^*_{12}$	.228 (.0004)	.179 (.0004)	.041 (.0003)	.208 (.0008)

The results, though limited in their scope, give an indication that the sample model estimator, $\hat{\beta}_{12}$, may be more robust than the econometric model estimator, $\hat{\beta}^*_{12}$. The latter has high mean square error (due to high bias) when the econometric model conditions do not hold, while $\hat{\beta}_{12}$ maintains a low level of mean square error even when the conditions for its consistency do not hold. Since $\hat{\beta}_{12}$ is an explicit single-stage estimator, this might make it preferable to $\hat{\beta}^*_{12}$, however further studies on the performances of the estimators for different departures from model assumptions have to be carried out.

8. Acknowledgements

The author is grateful to Sheila Ben-Tuvya and to Esther Sharon for programming and running the simulations of section 7. Research for this paper was partially supported by research agreements between the Israel Central Bureau of Statistics and the U.S.-Israel Binational Agricultural Research and Development Fund and the Israel National Council of Research and Development.

9. References

DeMets, D. and Halperin, M. (1977). Estimation of a simple regression coefficient in samples arising from a sub-sampling procedure. *Biometrics, 33,* 47-56.

Heckman, J. (1979). Sample selection bias as a specification error. *Econometrica, 47,* 153-161.

Heckman, J. (1976). The common structure of statistical models of truncation, sample selection and limited dependent variables and simple estimator for such models. *Annals of Economic and Social Measurement, 5,* 475-492.

Holt, D. and Smith, T.M.F. (1979). *Regression Analysis of Data from Complex Surveys.* Paper presented at Royal Statistical Society Conference, Oxford.

Lee, L., Maddala, G.S. and Trost, R.P. (1980). Asymptotic covariance matrices of two-stage probit and two-stage tobit methods for simultaneous equation models with selectivity. *Econometrica, 48,* 491-503.

Nathan, G. and Holt, D. (1980). The effect of survey design on regression analysis. *Journal of the Royal Statistical Society, B, 42,* 377-386.